Additive and partially dominant effects from genomic variation contribute to rice heterosis

Zhiwu Dan[1†], Yunping Chen[1†], 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 430072, China.

[†] Zhiwu Dan and Yunping Chen contributed equally to this article.

*Correspondence: wenchaoh@whu.edu.cn

Abstract

Background: The molecular mechanisms of heterosis remain debated due to discrepancies in factors such as species, phenotypes, hybrid populations, developmental stages or growth environments.

Results: Here, we screened heterosis-associated genes and metabolites from molecular differences in rice inbred lines and found that the phenylpropanoid biosynthesis was involved in heterosis. The main inheritance patterns of the heterosis-associated molecules were additive and partially dominant effects, namely, at the mid-parent levels and between the parental and mid-parent levels. The additive and partially dominant effects contributed to heterosis of 17 agronomic traits in rice, including grain yield and plant height across developmental stages, and to heterosis of F_1 hybrids from different populations and growth environments in both rice and maize. Furthermore, the two inheritance patterns were associated with parental genomic variants, which were correlated with heterosis of agronomic traits.

Conclusions: Our study systematically integrates parental genetic variation and highlights the predominant effects of additive and partially dominant in rice heterosis.

Keywords: Heterosis, Genomic variants, Additive and partially dominant effects, Grain yield, Plant height, Rice

Background

Heterosis describes the universal phenomenon of heterozygous F_1 hybrids outperforming their parents [1-3]. The exploitation of heterosis to increase crop yield and alleviate food shortages has been treated as a landmark innovation in modern agriculture [4]. The dominance (pseudo-overdominance) [5-9], overdominance [10, 11], and epistasis models [12-14], which emphasize complementation, heterozygosity, and interactions of loci, respectively, have been proposed to interpret heterosis. However, the molecular mechanisms of crop heterosis are still under debate because the conclusions differ across species, phenotypes, hybrid populations, developmental stages or growth environments.

Four terms have been used to quantitatively describe the inheritance patterns of hybrid phenotypes or molecules [15, 16], including additive (mid-parent values, MPVs), dominant (parental values), partially dominant (between parental and MPVs), and overdominant (outside parental values) effects. The potential connection between additive effect and heterosis has been inferred in several studies. *SINGLE FLOWER TRUSS*, the expression level of which exhibits a near-additive effect, drives strong yield

49 heterosis in tomato [17]. The additive effect of transcripts, proteins, metabolites, and
50 component traits has been suggested to be the genetic basis of heterosis in maize, rice,
51 and tomato [18-25]. Moreover, partially dominant effect has also been supposed to
52 underlie heterosis [6, 14, 18, 26].

53 Through the systematic identification of heterosis-associated genes and
54 metabolites, the proportions and contributions of the four types of inheritance patterns
55 were analysed in this study. The finding of the main genetic effects of additive and
56 partially dominant on heterosis is reliable without considering discrepancies in
57 agronomic traits, hybrid populations, developmental stages, growth environments,
58 detection platforms of omics data, and even species.

59 Results

60 Heterosis-associated genes and metabolites from molecular differences in parents

61 The seedling length and weight of the rice cultivars Yuetai B (YB, *indica*), Balilla
62 (*japonica*), and their reciprocal F_1 hybrids were measured at 5, 10, and 15 days after
63 sowing (DAS) (Fig. 1A, Fig. S1A, and Table S1). Better-parent heterosis (BPH) for
64 seedling length was observed in both F_1 hybrids at 15 DAS (Fig. 1A). Moreover, the

65 variance in seedling length was almost completely explained by genotype, stage, and
66 their interaction (Fig. S1B).

67 Principal component analysis (PCA) of global gene expression levels showed that
68 the F_1 hybrids were located between both parents and closer to one parent, Balilla (Fig.
69 1B, Fig. S1C, Table S2, and Data S1). Differentially expressed genes (DEGs) were
70 identified for the two inbred lines (ILs) and F_1 hybrids, and the number of DEGs
71 increased along with development (Fig. S1D-E). Unlike the pooled non-DEGs (NDEGs)
72 (Fig. S1F), 66.6% of the sample variance in genotypes was explained by principal
73 component 1 (PC1), which was based on the DEGs (Fig. S1G). Furthermore, the F_1
74 hybrids deviated slightly from the mimicked "additive hybrids" (calculated as MPVs),
75 suggesting the contributions of additive and other genetic effects to hybrid
76 transcriptomics.

77 The greatest number of DEGs existed between the two parents, and most of the
78 DEGs (83.5% to 97.1%, average 92.4%) between the F_1 hybrids and the parents were
79 parental DEGs (Fig. 1C and Fig. S1H-J). Moreover, the majority of the significantly
80 enriched Gene Ontology terms between the F_1 hybrids and the parents were among the

81 differentially expressed terms between Balilla and YB (Fig. S1K-P), showing that the
82 reconstruction of the hybrid transcriptomes was predominantly determined by parental
83 expression differences.

84 Next, 626 heterosis-associated genes (HAGs) related specifically to seedling length
85 were then identified (Fig. S2A-P), and these HAGs indicated that the F_1 hybrids were
86 closer to "additive hybrids" (Fig. 1D). A total of 37 HAGs were enriched in 12 terms,
87 and these terms formed a complete network with the bridge of heme binding (Fig. 1E-
88 F), indicating that the genes involved in heterosis for seedling length had tight
89 connections.

90 Moreover, the PCA score plot based on untargeted metabolite profiles revealed that
91 the F_1 hybrids were located between the corresponding parents and presented a closer
92 relationship with YB (Fig. 1G, Fig. S3A, and Data S2). Although the percentages were
93 lower than those of the transcriptomic data, most of the metabolomic differences
94 between F_1 hybrids and parents were due to differences in the parents (Fig. S3B-J,
95 average 84.0%). Meanwhile, most of the significantly enriched metabolic pathways
96 between the F_1 hybrids and the parents also differed between the two parents (Fig. 1H

97 and Fig. S3K-U). Thus, both the transcriptomic and metabolomic variation between F_1

98 hybrids and parents were determined mainly by the molecular differences in

99 corresponding parents.

100 When the differential metabolites (DMs) were further screened for heterosis-

101 associated metabolites (HAMs) related to seedling length (Fig. S4A-M), 645 HAMs

102 were identified and indicated that the F_1 hybrids were also closer to "additive hybrids"

103 (Fig. 1I). Eight HAMs annotated with the *mummichog* algorithm [27] were significantly

104 enriched in three pathways, and other chemically identified metabolites from these

105 pathways were correlated with seeding length (Fig. 1J and Fig. S5A). Furthermore,

106 three of six accurately annotated metabolites from the phenylpropanoid biosynthesis

107 were significantly correlated with the seedling length of F_1 hybrids (Fig. S4N-V).

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

109 five pathways were found to be significantly enriched, among which the

110 phenylpropanoid biosynthesis was most significant with eight metabolites and ten

111 genes matched (Fig. 1K). Among the ten genes, nine genes were peroxidases, and seven

112 genes were among the 37 HAGs (Fig. S5B), confirming the combined contributions of

genes and metabolites involved in phenylpropanoid biosynthesis to seedling length heterosis.

The explanation of variance in heterosis by differential molecules of parents

Another nine pairs of rice 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) (Fig. S6A-D). The 645 HAMs did not classify the three groups according to PCA (Fig. S6E), but they separated the high BPH group from the other two groups without overfitting according to PLS-DA (Fig. S6F-H). Thus, the HAGs or HAMs for a single hybrid combination manifested limited 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 (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 F_1 hybrids (Fig. S7A-C). Unexpectedly, the F_1 hybrids were divided into distinct groups

129 in PLS-DA, and the predictabilities of the models varied, with that of Module_2 being
130 as high as 0.97 (Fig. S7D-I). Thus, the integration of HAGs and HAMs into network
131 modules increased the predictability of heterosis for seedling length.

132 A total of 967 DEGs and 990 DMs were subsequently identified for six ILs of the
133 ten pairs of reciprocal F_1 hybrids. The 967 DEGs displayed better separations in PLS-
134 DA than in PCA, and the predictability was comparable to that of the 990 DMs (Fig.
135 S8A-H). The combination of DEGs and DMs separated the three groups according to
136 PC1 and PC4, but the model was overfitted in PLS-DA (Fig. S8I-L). Next, another
137 network (Network_HLN) integrating the 967 DEGs and 990 DMs was constructed and
138 the largest subnetwork (subnetwork_1) contained 868 edges (Fig. 2D). Although the
139 top three modules of subnetwork_1 from Network_HLN had fewer nodes than those of
140 Network_51015 did (Fig. 2E), the predictabilities of the top three modules of
141 Network_HLN were comparable to those of Network_51015 (Fig. 2f and Fig. S9A-F).
142 And the variance explained by Component 1 were larger in Network_HLN (Fig. 2G).
143 The combination of Module_1 and Module_2 further increased the predictability to
144 0.94, whereas the addition of Module_3 did not (Fig. S9G-J). Thus, the integration of

145 parental transcriptomic and metabolomic differences into network modules explained

146 the variance in heterosis for seedling length.

147 **Heterosis-associated molecules with additive and partially dominant effects**

148 Heterosis-associated molecules specific to seedling length were further screened

149 from Module_1 and Module_2 of Network_HLN, and the 68 molecules (29 metabolites

150 and 39 genes) with the greatest contribution separated the three groups with high

151 predictability (Fig. 2H-I, Q^2 = 0.90). The contributions of the molecules were

152 determined (Fig. 2J and Fig. S9K), but the 68 molecules did not exhibit continuously

153 changing patterns at the molecular level across groups. Moreover, F_1 hybrids from the

154 same group formed distinct clusters (Fig. 2K), displaying heterogeneity at the

155 molecular level. In addition, the hybrid profiles were intermediate between the levels

156 of the parents or close to those of one parent in the heatmap. And the F_1 hybrids were

157 located near the middle regions of corresponding parents on the PCA score plot (Fig.

158 2L). These results highlighted the quantitative connections of heterosis-associated

159 molecules between parents and F_1 hybrids.

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

161 by comparing the molecular levels of heterosis-associated molecules (Fig. 2M and Fig.
162 S9L). Additive and partially dominant effects accounted for more than 70% of the
163 molecular levels in most of the F_1 hybrids (Fig. 2N and Fig. S9M-O). The predictability
164 of the additive effect for heterosis of seedling length was 0.63 (Fig. 2O, $P < 5\times10^{-4}$).
165 Two types of partially dominant (PD) effects, namely, PD_0.3 and PD_0.7, were
166 identified, and their predictabilities were 0.37 ($P < 5\times10^{-4}$) and 0.29 ($P < 5\times10^{-4}$),
167 respectively (Fig. 2P). The predictabilities of the dominant effects of the female and
168 male parents were 0.14 ($P = 0.40$) and 0.07 ($P = 0.015$), respectively (Fig. 2Q).
169 Importantly, the predictability of the additive effect was not significantly different from
170 those of hybrid profiles, and the addition of additive, partially dominant, or dominant
171 effects to the hybrid profiles did not improve the predictability (Fig. 2Q). Furthermore,
172 the sum percentages of additive and partially dominant effects had a significant positive
173 correlation with seedling length heterosis (Fig. 2R). The above results demonstrated the
174 predominant roles of additive and partially dominant effects on seedling length
175 heterosis.

176 **Contributions of additive and partially dominant effects to heterosis**

177 To investigate contributions of the predominant genetic effects to heterosis of field
178 agronomic traits, we then first measured plant height for the above rice F_1 hybrids at
179 four developmental stages and detected significant positive correlations between
180 seedling length heterosis and plant height heterosis (Fig. S10A-B). Meanwhile, a hybrid
181 population with a complete diallel cross design was constructed for 17 representative
182 rice ILs (Figs. 3A-B, Fig. S10C-D, and Data S3). Seventeen agronomic traits were
183 measured (Fig. S10E-H), and the average BPH values for only four traits were less than
184 zero (Figs. 3C-D), indicating increased fitness of the F_1 hybrids to ILs.
185 After metabolite profiles of the 17 ILs were obtained (Fig. S11A and Data S4), five
186 chemically identified metabolites, including the previously reported 4-
187 hydroxycinnamic acid [22], from the phenylpropanoid biosynthesis displayed
188 significant correlations with heterosis for plant height (Fig. 3E and Fig. S11B-G).
189 Importantly, additive and partially dominant effects were more strongly correlated with
190 plant height heterosis than the dominant effects (Fig. S11H), even after the inclusion of
191 another 13 traits (Fig. 3F).
192 Next, the 1,306 differential metabolites of the 17 ILs were used to predict heterosis

193 for each trait by PLS. We found that the additive effect had higher R^2 (adjusted) values

194 than the dominant and partially dominant effects (Fig. 3G and Fig. S11I-X). Meanwhile,

195 the greatest R^2 values of the regression models for the 17 traits with additive effect were

196 significantly greater than those with dominant effects but comparable to those with

197 partially dominant effects (Fig. 3H). Moreover, eight traits of the 17 ILs and 36 F_1

198 hybrids were recollected, and significant differences between environments were

199 observed for most of the traits (Fig. S12A-B). The greatest R^2 values of the eight traits

200 with additive effect were greater than those with dominant effects, and were close to

201 those with partially dominant effects (Fig. 3I and Fig. S13A-H). Therefore, the additive

202 and partially dominant effects were the main inheritance patterns for heterosis of 17

203 agronomic traits.

204 A test cross population was then designed with a male sterile line and 211

205 recombinant inbred lines (RILs) (Fig. S14A). The F_1 hybrids exhibited negative BPH

206 only for heading date of the four traits (Fig. S14B-F). The metabolite profiles of 107

207 representative RILs and Yuetai A were collected (Fig. S14G-H), and the R^2 values were

208 greater than 0.85 with five latent factors (Fig. 3J). Thus, the differential metabolites of

representative rice ILs were robust biomarkers for the heterosis of F_1 hybrids from recombinant inbred lines.

To further investigate the contributions of additive and partially dominant effects to maize heterosis, the field fresh weight of two hybrid populations[28], which significantly differed under distinct growth environments (Fig. S15A-B), and 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 (Fig. S15C-J). For the greatest R^2 values, additive and partially dominant effects were comparable to those of the hybrid profiles and were greater than the dominant effects (Fig. 3K). Unexpectedly, additive and partially dominant effects also 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 [29] (Fig. 3L and Fig. S15K-N). Moreover, the grain yield of another hybrid population [30], which exhibited strong yield heterosis, and the expression levels of 84 DEGs of 12 maize ILs were obtained (Fig. S15O-Q). The R^2 values of the additive and partially dominant effects were greater than those of the dominant effects (Figs. 3M-N and Fig. S15R-S). These results showed the predominant

225 roles of additive and partially dominant effects in maize heterosis. In addition,
226 phenotypic and transcriptomic data of two tissues at successive developmental stages
227 were obtained from Col-0, Per-1, and their F_1 hybrids in *Arabidopsis* [31]. The F_1
228 hybrids exhibited robust heterosis for both cotyledon biomass and first true leaf size.
229 And showed tight associations with additive and partially dominant effects based on
230 the screened heterosis-associated genes (Fig. S16A-T).

231 **The association between genomic variants and the main inheritance patterns**

232 Genomic sequencing was performed on seedlings of the 17 rice ILs, and these ILs were
233 classified into two groups based on genomic variants, including single-nucleotide
234 polymorphisms (SNPs), insertions and deletions (InDels), copy number variations
235 (CNVs), and structural variations (SVs) (Figs. 4A-F, Fig. S17-19, and Table S3).
236 The genomic variants of each two ILs were divided into common and unique types, and
237 the number of these genomic variants had significant correlations with trait heterosis
238 (Fig. 5A and Fig. S20A-B). The parental unique variants had more correlations with
239 heterosis than the common ones did (Fig. 5B). Meanwhile, unique SNPs and InDels
240 exhibited more correlations with heterosis than the unique CNVs and SVs did (Fig. 5C).

241 The parental unique SNPs and InDels were significantly correlated with heterosis for
242 grain yield, but CNVs and SVs were not. With more parental unique SNPs or InDels,
243 the heterosis for grain number increased, but that for the seed setting rate decreased
244 (Fig. 5D and Fig. S20C-E), revealing the barrier of reproductive isolation in the
245 utilization of yield heterosis between *indica* and *japonica* varieties [32]. Finally, the
246 inverse relationship between heterosis for grain number and the seed setting rate caused
247 the heterosis for grain yield to form a parabola (Fig. 5E and Fig. S20F-H).

248 The heterosis-associated metabolites of each trait were screened to top 50 and used
249 to perform correlation analysis with heterosis and the number of parental unique
250 variants. Unexpectedly, the additive and partially dominant effects exhibited more
251 correlations with both heterosis and parental unique variants (Fig. 5F-I and Fig. S20I).
252 Finally, a model integrating parental genetic variation in genomics, transcriptomics, and
253 metabolomics was proposed to show the contribution of additive effect to positive
254 better-parent heterosis of a complex trait in rice (Fig. 6).

255 Discussion

256 The genetic mechanisms underlying heterosis have been debated for more than a

257 century, and no consensus has been reached. The mysterious molecular mechanisms of
258 heterosis may be due to the large number of heterosis-associated molecules, including
259 genes [1, 3, 15, 33]. Heterosis-associated genes with strong functions are rarely reported
260 [17], and the regulation of gene expression levels for one pathway or network module
261 is worthy of further study.

262 Models of heterosis based solely on gene interactions may hinder the interpretation
263 of heterosis for complex traits because interactions among component traits occur at
264 the phenotypic level [18]. The heterosis of complex traits, e.g., grain yield, is a product
265 of several component traits, which are always negatively correlated. The component
266 traits also include component traits. It is therefore necessary to decompose the
267 components of a complex trait, and the investigation of heterosis for component traits
268 may be fruitful.

269 Additive and partially dominant effects are the main inheritance patterns in our
270 analyses, irrespective of changes in agronomic traits, hybrid populations,
271 developmental stages, growth environments, types of omics data, and even species. The
272 additive effect, namely, the mid-parent values, has no significant difference in

273 predictability from that of hybrid profiles, and the additive effect has been used to
274 predict multiple agronomic traits of F_1 hybrids [22-25], all of which contribute to the
275 precise pairing of male and female parents in hybrid breeding. To build robust
276 prediction models for crop heterosis, a collection of genetically divergent accessions
277 whose genetic diversity can be assessed by untargeted metabolite profiles should first
278 be implemented. For heterosis of traits with low predictability, e.g., tiller number,
279 obtaining metabolite profiles from corresponding tissues or mixed tissues [21], other
280 types of omics data [34-37], and phenotypic data from distinct environments [38] may
281 explain more of the variance.

282 Methods

283 Plant materials and field growth conditions

284 According to the *indica–japonica* indices of 223 inbred lines (ILs) evaluated by 33 pairs
285 of InDel markers [39, 40], 17 ILs, ranging from *indica* to intermediate types to *japonica*,
286 were selected as parents of a hybrid population using a complete diallel cross design
287 [25]. Moreover, 211 recombinant inbred lines (RILs, F_5 followings) derived from F_1
288 individuals of the 223 ILs were crossed with the Honglian-type male sterile line Yuetai

289 A. Plants of the diallel and test cross populations were planted at the Hybrid Rice
290 Experimental Base of Wuhan University in Ezhou (Hubei Province, China) in 2012 and
291 2015, respectively.

292 For the diallel cross population (245 F_1 hybrids), ten experimental plants were
293 planted with 16.5 cm between plants and 26.4 cm between rows, and two plants of the
294 sterile line, namely Yuetai A, were planted on two sides of each row. To obtain F_1
295 hybrids, the young panicles of the female parents that would flower in the next few
296 days were immersed in warm water at 44 °C (for *indica* or near-*indica* ILs) or 46 °C
297 (for *japonica* or near-*japonica* ILs) for 5 min to sterilize the pollen. For the hand-
298 pollination procedure, three to four panicles were combined in a paper bag, and one to
299 two bags from each female parent without pollination were used as controls. For female
300 parents with a high self-fertilization rate, such as R465 and Qianlijing, the pistils in
301 each spikelet were manually removed and pollinated by hand. Three biological
302 replicates of each genotype were randomly distributed in the field. The middle plants
303 of each row, ranging from three to ten plants, were selected for the collection of
304 phenotypic data for 17 agronomic traits. Plant height, measured from the base of the

305 plant to the highest part of the plant using meter rulers, was recorded from the seedling
306 to mature stages. The four time points included 28 days after transplanting (8 July,
307 2012), 42 days after transplanting (22 July, 2012), the time of flowering, and the time
308 of maturation, which were defined as the seedling stage plant height, elongation stage
309 plant height, flowering stage plant height, and maturation stage plant height,
310 respectively. The tiller number per plant was also counted together with the plant height
311 at four different stages, which were defined as the seedling stage tiller number,
312 elongation stage tiller number, flowering stage tiller number, and maturation stage tiller
313 number. The heading date was the number of days between the heading time when a
314 plant from a biological replicate started flowering and the date of sowing (13 May,
315 2012). At maturity, all panicles from each plant were harvested and dried in the sunshine
316 for approximately one week and then dried in ovens at 42 °C for two days before yield
317 and yield-related traits were measured. Panicle length was measured from the panicle
318 neck to the top for all panicles from each plant. The primary branch number and
319 secondary branch number per panicle were counted manually for all panicles. Panicles
320 from each plant were counted and treated as the tiller number per plant. After manually

321 removing seeds (with or without grains) from panicles for each individual plant, empty
322 seeds and seeds with grains were mechanically separated and counted using seed
323 counters (PME-1, Shanke Equipment, Shanghai). The grain number per panicle was
324 calculated by dividing the total seed number (sum of empty seeds and seeds with grains)
325 of a plant by the tiller number per plant. The seed setting rate was calculated as the ratio
326 of the number of seeds with grains to the total number of seeds. The weight of 1,000
327 grains from a plant was defined as the thousand-grain weight, and the weight of the
328 seeds with grains per plant was defined as the grain yield per plant.

329 For the test cross population (211 F_1 hybrids), each biological replicate contained
330 ten plants (per row) spaced at 16.5 cm × 16.5 cm apart. Three biological replicates of
331 each genotype were placed in three consecutive rows and randomly planted in the field.
332 The phenotypic data of four agronomic traits, namely, heading date, thousand-grain
333 weight, maturation stage plant height, and yield per plant were collected. The heading
334 date was recorded for one biological replicate. The middle five plants per row were
335 selected for measurement of plant height at the maturation stage. Panicles from the five
336 plants were bagged and used to collect data on grain yield per plant and thousand-grain

337 weight. Moreover, eight agronomic traits, namely, heading date, thousand grain weight,

338 panicle length, seed setting rate, grain number per panicle, maturation stage plant height,

339 grain yield per plant, and tiller number per plant, of the 17 ILs and 36 F_1 hybrids from

340 the diallel cross population were recollected along with the test crosses. Parts of trait

341 values of the diallel and test cross populations have been reported in previous studies

342 [19, 22-25, 32], and the complete datasets for all investigated traits are provided in

343 source data.

344 To evaluate the degree of heterosis for each trait, better-parent heterosis was

345 calculated using the equation: BPH_trait = $(F_1 - P_{high})/P_{high}$, where P_{high} represents the

346 high trait value of both parents. In the calculation of better-parent heterosis for heading

347 date (BPH_HD), parents with earlier heading dates were selected as P_{high}. The

348 calculated values were multiplied by -1 to represent the heterosis for heading date.

349 **Planting of rice seedlings for metabolomics analysis**

350 Two batches of rice seedlings were planted for untargeted metabolite profiling analysis.

351 The first batch in 2016 (Batch_2016) included 17 ILs, 107 RILs selected from the 211

352 RILs, and the male sterile line Yuetai A. As previously described [23], the germinated

353 seeds were sown in soil with a spacing of 2 × 2 cm. The seedlings were grown at 30 °C
354 under a 16 h light/8 h dark photoperiod in a light incubator. Two biological replicates
355 were established for each genotype. On the 15th day, the aboveground parts, excluding
356 approximately 0.5 cm of tissue close to the soil, of approximately ten plants were
357 harvested and washed with ddH_2O three times. Two biological replicates of the 17 ILs
358 were harvested as two experimental samples for each genotype. Two biological
359 replicates of the 107 RILs and Yuetai A were pooled as one experimental sample for
360 each genotype. The seedlings were first placed in 5 mL tubes (Eppendorf #0030119460)
361 and snap frozen in liquid nitrogen.

362 The second batch in 2018 (Batch_2018) consisted of ten pairs of hybrid
363 combinations. Seeds of 6 ILs and ten corresponding reciprocal F_1 hybrids were first
364 dehulled and sterilized with 75% ethanol (1 min) and 0.15% $HgCl_2$ (10 min). The seeds
365 were then transferred to sterilized half-strength Murashige and Skoog rooting medium
366 (pH = 5.8) supplemented with phytogel. Ten seeds were sown on the surface of 30 mL
367 of medium per glass tube (diameter: 3 cm and height: 20 cm). Three to six biological
368 replicates were set up for each genotype or the same genotype at different time points.

369 Seedlings were grown at 28 °C under 24-hour light in an artificial climate chamber
370 (Zhu Laboratory at the College of Life Sciences, Wuhan University). The seedling
371 length and weight of a pair of hybrid combination containing two ILs (Balilla and
372 Yuetai B) and the corresponding reciprocal F_1 hybrids were recorded on 5, 10, and 15
373 days after sowing. Five seedlings from each glass tube were removed from the medium
374 and photographed for subsequent measurement of seedling length using ImageJ
375 (version 1.5.3). After excluding the roots, the seedling bases were washed three times
376 with ddH_2O and the water on the surface was removed with absorbent paper before
377 weighing. The five seedlings were snap frozen in liquid nitrogen, and photographs and
378 weights were taken and used for subsequent experiments. The remaining nine pairs of
379 hybrid combinations were harvested for seedling length measurements after 15 days of
380 growth. The sealing films on the glass tubes were removed on day 11 and 5 mL of
381 ddH_2O was added to each tube for normal growth. All the experimental samples were
382 stored at -80 °C.

383 **RNA sequencing**

384 Seedlings from each biological replicate were ground into fine powder using mortar

385 and pestle with liquid nitrogen. After total RNA was extracted with TRIzol (Invitrogen),
386 the RNA was first assessed by electrophoresis in 1% agarose gels to check for
387 degradation and contamination states. The concentrations (Qubit2.0 Fluorometer) and
388 integrity (Agilent 2100 bioanalyzer) of each RNA sample were analyzed. DNA library
389 preparation (NEBNext Ultra RNA Library Prep Kit, NEB) with index codes and
390 sequencing experiments were performed on the sequencing platform of Novogene Co.,
391 Ltd (Beijing, China). Briefly, mRNA was first purified from 3 μg of total RNA using
392 magnetic beads attached to poly(T) oligo. The mRNA with poly(A) tails was randomly
393 fragmented (NEBNext First Strand Synthesis Reaction Buffer) and used to synthesize
394 the first strand of cDNA with the addition of hexamer primers and M-MuLV reverse
395 transcriptase. After the RNA was degraded with RNase H, the second strand of cDNA
396 was synthesized using DNA polymerase I with dNTPs. The strands were then converted
397 to blunt ends using the activities of T4 DNA polymerase and DNA polymerase I. The
398 products were subjected to adenylation, and NEBNext adaptors were added to the
399 products. Approximately 200 bp of cDNA was purified using AMPure XP beads
400 (Beckman Coulter) and treated with 3 μL of USER Enzyme (NEB) at 37 °C for 15

401 minutes and 95 °C for 5 minutes. PCR was then performed using Phusion High-Fidelity
402 DNA polymerase with universal and index primers. The reaction products were
403 repurified with AMPure XP beads, and the quality of the library was assessed using an
404 Agilent Bioanalyzer 2100. After clustering of the index-coded samples using the cBot
405 Cluster Generation System (TruSeq PE Cluster Kit v3-cBot-HS, Illumina), the libraries
406 were sequenced on the Illumina Hiseq PE150 platform, and 150 bp paired-end reads
407 were acquired with more than 6 G per sample.

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

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

414 Normalization and differential analysis of the transcriptomics data (42,341 genes) for
415 the hybrid combination of Balilla and YB at 5, 10, and 15 DAS were performed using
416 NetworkAnalyst 3.0 [43] with the following parameters: specify organism, not

417 specified; data type, bulk RNA-seq data (counts); ID type, not specified; gene-level
418 summarization, sum; variance filter, 30; low abundance, 30; filter unannotated genes,
419 yes; normalization, Log2-counts per million; statistical method, DESeq2; Log2 fold
420 change, 2; and adjusted *p*-value, 0.0001. As the numbers of differentially expressed
421 genes between reciprocal F_1 hybrids of Balilla and YB at 5, 10, and 15 DAS were only
422 1, 13, and 29, respectively, the differences between reciprocal F_1 hybrids were not
423 included in the analysis. To analyse the differentially expressed genes of the six ILs,
424 the data matrix containing 42,341 genes was separately normalized (sum, none, and
425 autoscaling) with MetaboAnalyst 4.0 [44] and analysed with ANOVA (adjusted *p*-value
426 = 0.05, post-hoc analysis = Fisher's LSD). Principal component analysis (PCA) of the
427 transcriptomic and metabolomic data, which were normalized data from
428 NetworkAnalyst 3.0 [43], was performed at ClustVis (https://biit.cs.ut.ee/clustvis) [45].
429 Two-way ANOVA of the 5,844 DEGs was performed using the "Time Series
430 Analysis" (time series and one factor) module of MetaboAnalyst 4.0 [44] with the
431 following parameters: ANOVA type, type I; interaction, yes; adjusted *p*-value cut-off,
432 0.05; and correction for multiple testing, false discovery rate.

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

434 The ClueGO (version 2.5.8) [46] and CluePedia (version 1.5.8) [47] plugins of
435 Cytoscape (version 3.9.0) [48] were used for pathway enrichment analysis of the
436 differentially expressed genes. Functional analysis was selected as the analysis mode.
437 The up- and downregulated genes were represented by diamonds (shape 1) and triangles
438 (shape 2), respectively. The GO (ellipse), KEGG (hexagon), and WikiPathways (V)
439 were selected as ontologies/pathways. The GO term fusion was selected and only
440 pathways with pV ≤ 0.01 were shown. The GO tree interval was between 3 (minimum
441 level) and 5 (maximum level). The minimum number and percentage of genes were set
442 to 3 and 4%, respectively, for both clusters. The genes accounting for more than 60%
443 of a cluster were defined as cluster-specific genes. The kappa score of network
444 connectivity was set to 0.4, and enrichment/depletion (two-sided hypergeometric test)
445 with Benjamini–Hochberg pV correction was set as the statistical option. GO term
446 grouping was selected in the group options, and the leading term was based on the
447 highest significance with fixed colours. The kappa score was selected for term grouping.
448 The initial group size, percentage of genes for group merging and percentage of terms

449 for group merging were set to 1, 50% and 50%, respectively.

450 **Untargeted metabolite profiling**

451 For the seedling samples in Batch_2016, liquid chromatography-mass spectrometry
452 (LC–MS) analysis procedures have been previously described [23, 49]. In brief, all
453 tissues in each EP tube were first ground into homogenized powder by mortar and pestle
454 with the addition of liquid nitrogen. Eighty milligrams of powder was then transferred
455 to 2 mL EP tubes and the tubes were vortexed for 1 min after the addition of 1 mL of
456 precooled extraction solution (methanol/acetonitrile/water, 2/2/1). A quality control
457 sample was prepared by mixing 10 μL of each experimental sample. The experimental
458 and quality control samples were then subjected to two rounds of ultrasonic treatments
459 (30 min each) and kept at -20 °C for 60 min. The samples were then centrifuged at
460 14,000 g for 15 min at 4 °C and the supernatants were transferred to clean tubes. After
461 drying the supernatants in a vacuum concentrator, 100 μL of acetonitrile
462 (acetonitrile/water, 1/1) was added to the tubes. The tubes were vortexed and
463 centrifuged at 14,000 g for 15 minutes at 4 °C. The supernatants were transferred to
464 sample tubes for LC–MS analysis on the platform of Shanghai Applied Protein

465 Technology Co., Ltd.

466 The experimental and quality control samples were separated on an Agilent 1290
467 Infinity LC system (Agilent Technologies) equipped with a reverse-phase T3 column
468 (ACQUITY UPLC HSS T3, 1.8 μm, 2.1 mm × 100 mm, Waters). The injection volume
469 was 2 μL per sample. The flow rate was 0.3 mL/min and the column temperature was
470 25 °C. Regents A and B in positive ion mode were 0.1% formic acid (in water) and 0.1%
471 formic acid (in acetonitrile), respectively. In addition, 0.5 mM ammonium fluoride
472 (Sigma–Aldrich) and acetonitrile (Merck) were used in negative ion mode. The gradient
473 elution steps were as follows: 0-1 min, 1% B; 1-8 min, linearly increased to 100% B;
474 8-10 min, held at 100% B; 10-10.1 min, reduced to 1% B; and 10.1-12 min, held at 1%
475 B. Experimental samples were analysed in a random order, and the quality control
476 sample was injected every nine experimental samples. To obtain MS profiles in both
477 positive and negative electron spray ion modes, the LC system was coupled to an
478 Agilent 6550 iFunnel QTOF mass spectrometer (Agilent Technologies). The MS data
479 were collected from 60 to 1,200 Da. The acquisition rate was four spectra per second,
480 and the cycle time was 250 microseconds. The temperatures of the drying gas and the

481 sheath gas were 250 °C and 400 °C, respectively. The flow rates of the drying gas and
482 the sheath gas were 16 L/min and 12 L/min, respectively. The nebulizer pressure was
483 set at 20 pounds per square inch (psi). The capillary, nozzle, and fragment voltages were
484 3,000, 0, and 175 V, respectively. Tandem mass spectrometry data were then acquired
485 on the quality control sample using a TripleTOF 6600 mass spectrometer (AB SCIEX).
486 The mass range was between 60 and 1,200 Da. To increase the data acquisition rate of
487 the quality control sample, the sample was injected twice and the mass range was
488 further divided into four sequential windows: 60-200, 190-400, 390-600, 590-1,200
489 (full range, 60-1,200) Da (for chemical standard library) and 200-600, 590-750, 740-
490 900, 890-1,200 (full range, 200-1,200) Da (for lipid MS/MS library). The mass
491 spectrometry accumulation time was 200 ms per spectrum, and the tandem mass spectra
492 were obtained in an information-dependent manner with high sensitivity. Ion source
493 gas 1, ion source gas 2, and the curtain gas were set at 40, 80, and 30 psi, respectively.
494 The source temperature was 650 °C. The ion spray voltage floating were +5,000 V and
495 -4,500 V, respectively. The declustering potential and collision energy were ± 60 V and
496 35 ± 15 eV, respectively. The number of candidate ions for monitoring per cycle was

497 ten and the isotope isolation width was 4 Da. The MS/MS accumulation time was set
498 to 50 ms.

499 The raw MS files (.d) of the experimental and quality control samples were
500 converted to mzXML format using ProteoWizard (version 3.0.6150) [50], and the
501 transformed files were uploaded to XCMS (version 1.46.0) [51] for feature detection,
502 retention time (RT) correction and feature alignment with previously reported
503 parameters [52]. The 17 ILs formed one group, and the 107 RILs plus Yuetai A formed
504 another group. The features with more than 50% nonzero values in at least one group
505 were retained for further analysis. The error thresholds for m/z and RT were set at 15
506 ppm and 20 s for the alignment of MS and MS/MS data, respectively. The MS/MS data
507 of the quality control sample were matched against both the local chemical standard
508 library [53] and the lipid MS/MS library [54], and metabolites were annotated with a
509 mass accuracy of less than 25 ppm. For the seedling samples in Batch_2018, the
510 metabolomics analyses were the same as those in Batch_2016, except that the four
511 sequential windows of the quality control sample were changed to three sequential
512 windows: 60-300, 290-600, and 590-1,200 (full range, 60-1,200) Da. The Metlin library

513 [55] was also used for feature annotation.

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

515 The normalization of the metabolomics data for the hybrid combination of Balilla and
516 YB at 5, 10, and 15 DAS followed the same parameters as those used for the
517 transcriptomics data. ANOVA was used to analyse differential metabolites for the six
518 ILs with a data matrix of 15,083 analytes. For ease of interpretation, the word
519 "metabolite" was used in this manuscript to represent analytes with or without
520 annotations. The "statistical analysis (metadata table)" module of MetaboAnalyst 5.0
521 [56] was used to perform two-way ANOVA on the 2,331 differential metabolites with
522 genotype and stage as metadata. One of the three biological replicates of each genotype
523 at each stage was used as a subject. Multiple factors were chosen for the study design.
524 Further data filtering, sample normalization, data transformation, and data scaling were
525 not performed. Type I ANOVA was used for two-way independent ANOVA, and two-
526 way repeated ANOVA was used to calculate interactions. The false discovery rate was
527 used to correct for multiple testing, and the adjusted *P* value cut-off was set at 0.05.

528 To normalize the metabolomics data of the 17 ILs, Yuetai A, and 107 RILs, the

529 intensities of the two biological replicates of the 17 ILs were first averaged and then
530 normalized using MetaboAnalyst 5.0 [56] (sum, none, and autoscaling for sample
531 normalization, data transformation, and data scaling). Metabolites with RSD greater
532 than 25% in the QC samples were removed and the interquantile range was selected to
533 filter the features (40%). For the differential metabolite analysis of the 17 ILs, the
534 feature intensities of two biological replicates for each IL were used and ANOVA was
535 performed after filtering 40% of the features and data normalization (sum, none, and
536 autoscaling) using MetaboAnalyst 5.0 [56].

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

538 To determine the inheritance patterns for heterosis-associated molecules,
539 transcriptomic and metabolic levels were compared among F_1 hybrids, the
540 corresponding parents, and the average levels of the corresponding parents via ANOVA
541 using MetaboAnalyst 4.0 [44]. Three biological replicates of reciprocal F_1 hybrids from
542 two ILs were combined into a hybrid genotype. An additive effect was assigned when
543 the molecular levels of the hybrid genotype were not significantly different from the
544 average levels of the corresponding parents but were significantly different from those

545 of both parents. A partially dominant effect was assigned when the hybrid levels were
546 significantly different from both the average parental levels and the molecular levels of
547 the parent on the same side. A dominant effect was when the hybrid levels were not
548 significantly different from those of the parent but were significantly different from the
549 average levels of the two parents. An overdominant effect was the pattern when the
550 hybrid levels were significantly higher or lower than those of the parent on the same
551 side. The patterns of hybrid levels other than the above four types were defined as
552 unclassified and marked as NA. To quantitatively describe the partially dominant effect
553 (PD), the degrees for the molecules with partially dominant effect of the six hybrid
554 combinations was calculated with the equation $PD = (F_1-P_2)/(P_1-P_2)$, where F_1, P_1, and
555 P_2 represent the molecular levels of the F_1 hybrids, parent 1, and parent 2, respectively.
556 Several molecules that had a partially dominant effect in multiple F_1 hybrids were
557 selected multiple times. The degree of the partially dominant effect was divided into
558 two parts, with 0.5 as the threshold. The averages of the two groups were calculated
559 and defined as two types of partially dominant effects (PD_0.3 and PD_0.7). Moreover,
560 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

561 represent partially dominant effects based on parental molecular levels.

562 **Pathway enrichment analysis of differential metabolites**

563 The metabolomics data of the hybrid combination of Balilla and YB at 15 DAS were
564 first used for pathway enrichment analysis. The table with normalized feature
565 intensities and retention times (in seconds) was uploaded to the "functional analysis"
566 module of MetaboAnalyst 5.0 [56]. Both positive and negative modes were selected for
567 analysis. In positive mode, the mass tolerance was set to 5 ppm. The interquantile range
568 was selected for data filtering because the number of peaks was greater than 5,000. No
569 data were used for sample normalization, data transformation, or data scaling. Both
570 *mummichog* [27] (*P* value cut-off: default top 10% peaks, version 2.0) and GSEA were
571 used for enrichment analysis. As no pathway was significantly enriched for the
572 differential metabolites between Balilla×Yuetai B and its parent Balilla with the default
573 *P* value cut-off (0.01) in positive mode, the *P* value cut-off was increased to 0.05. The
574 pathway library used was *Oryza sativa japonica* (Japanese rice) [KEGG]. Metabolite
575 currency and adducts were set with the default sets, and pathways or metabolite sets
576 containing at least 3 entries were set. The parameters for the negative mode analysis

577 were the same as those for the positive mode analysis, except for that no data filtering
578 was performed and the *P* value cut-off was set to 0.05. Putatively annotated metabolites
579 and enriched pathways were obtained in each analysis. For the pathway analysis of the
580 645 DMs using the "Pathway Analysis" module of MetaboAnalyst 5.0 [56], feature
581 annotation information was obtained from the results of differential metabolites
582 between Balilla and YB at 15 DAS. The parameters were set as follows: scatter plot,
583 visualization method; hypergeometric test, enrichment method; relative-betweenness
584 centrality, topology analysis; all compounds in the selected pathway library, reference
585 metabolome; and *Oryza sativa japonica* (Japanese rice) (KEGG), pathway library.
586 Three pathways were significantly enriched based on the annotation of eight peaks in
587 positive mode. Moreover, 20 metabolites (23 peaks from both ion modes), which were
588 annotated with local chemical standards, lipids, and Metlin libraries, from the three
589 significantly enriched pathways were also used for analysis. Correlations between the
590 metabolite levels of 26 peaks and seedling length were determined. One peak
591 (M147T118_POS) was annotated as 4-hydroxycinnamic acid in both the local standard
592 library and the result of the pathway enrichment analysis, and four peaks had no

593 concentration in the normalized metabolomics data.

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

595 The normalized transcriptomics and metabolomics data were used to construct
596 networks via OmicsAnalyst [57]. No organism or ID type was specified in the
597 annotation procedure. Features with 50% missing values were excluded and the
598 remaining missing values were replaced by LoDs. The filter threshold for variance and
599 abundance was set to 0. In the comparison step, the genotype was used for grouping,
600 and no data transformation was performed. ANOVA was used for the differential
601 analysis. The fold change cut-off was set at 2.0, and the *P* value cut-off was set at 0.01.
602 To make the data suitable for inter-omics analysis, both the transcriptomics and
603 metabolomics data underwent normalization by sum, variance stabilizing normalization,
604 quantile normalization and range scaling. The statistically significant features were then
605 used to construct an interomics correlation network using the Spearman method. The
606 correlation threshold was set at 0.8, and both positive and negative correlations were
607 included. For Network_51015, 626 genes and 645 metabolites were used, and a total of
608 213 genes and 198 metabolites were statistically significant. The network contained

three subnetworks, and the largest subnetwork with 2,138 edges (subnetwork_1, 61.8% and 38.2% positive and negative correlations, respectively) had 211 genes and 196 metabolites as nodes. The other two subnetworks had two interactions. For Network_HLN, 967 genes and 990 metabolites were used for analysis, and 103 genes and 158 metabolites were retained in the network. Network_HLN contained seven subnetworks, and the largest subnetwork with 868 edges (subnetwork_1) contained 97 genes and 149 metabolites. The remaining six subnetworks had nine edges. The walktrap algorithm was used for module detection. Subnetwork_1 of Network_51015 contained seven modules. The module size and *P* value of each module were 150 ($P = 2.34\times10^{-21}$, Module_1), 105 ($P = 1.01\times10^{-14}$, Module_2), 80 ($P = 2.29\times10^{-12}$, Module_3), 21 ($P = 0.00303$, Module_4), 13 ($P = 0.452$, Module_5), 7 ($P = 1$, Module_6), and 6 ($P = 0.318$, Module_7), respectively. Subnetwork_1 of Network_HLN contained eight modules. The module size and *P* value of each module were 112 ($P = 6.6\times10^{-35}$, Module_1), 37 ($P = 2.61\times10^{-7}$, Module_2), 35 ($P = 1.39\times10^{-6}$, Module_3), 16 ($P = 3.8\times10^{-5}$, Module_4), 11 ($P = 0.111$, Module_5), 7 ($P = 0.0592$, Module_6), 6 ($P = 0.0176$, Module_7), and 6 ($P = 0.0309$, Module_8), respectively.

625 The subnetworks were imported into Cytoscape [48] for display using the compound

626 spring embedder layout. The node size was determined by the node degree.

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

628 To integrate the transcriptomics and metabolomics results, the MSU IDs of the 626

629 DEGs were first converted to gene symbols using the NCBI "Batch Entrez" tool

630 (https://www.ncbi.nlm.nih.gov/sites/batchentrez). Genes with gene symbols in the

631 *Oryza sativa japonica* group (340 genes) were uploaded to the "Joint pathway" analysis

632 module of MetaboAnalyst 5.0 [56]. Moreover, the annotation information of the 645

633 peaks was extracted from the results of the pathway enrichment analysis between

634 Balilla and YB at 15 DAS. The KEGG IDs for 29 compounds in positive mode and 44

635 compounds in negative mode were uploaded as metabolomics data. "All pathways

636 (integrated)" was selected as the pathway database. The parameters for enrichment

637 analysis, topology measure, and integration method were set as hypergeometric test,

638 degree centrality, and combine queries, respectively.

639 **Analyses of phenotypic and transcriptomic/metabolic data for inbred lines and F_1**

640 **hybrids in maize and *Arabidopsis***

641 The field fresh weight, grain dry matter yield, and grain dry matter content of maize
642 were obtained from published data in two previous studies [28, 29]. Metabolite profiles
643 of 3.5-day-old roots and leaves at the third leaf stage were also collected for 210 Dent
644 and Flint lines and 363 F_1 hybrids in the two studies [28, 29]. The root and leaf
645 metabolic datasets contained levels of 165 and 81 metabolites, respectively. Since
646 metabolic data were not available for the parents of some of the F_1 hybrids, the field
647 fresh weight of 326 F_1 hybrids, which were crossed between 24 Dent lines and 25 Flint
648 lines, were used for analysis. No further data normalization was performed on the
649 metabolic data. As reported in the study [28], the field fresh weight of the 326 F_1 hybrids
650 was recorded in two distinct field tests in 2010 (146 F_1 hybrids) and 2012 (180 F_1
651 hybrids). In the PLS analysis, the number was set to 100 to ensure a sufficient number
652 of latent factors, and the number was defined for each regression analysis when the R^2
653 value (adjusted) peaked. The R^2 values (adjusted) corresponding to different numbers
654 of latent factors and the greatest R^2 values for hybrid profiles and different inheritance
655 patterns were used for comparison analysis via ANOVA. Among the 326 F_1 hybrids,
656 291 F_1 hybrids had recorded data on grain dry matter yield and grain dry matter content

657 in approximately seven agro-ecologically diverse environments [29]. The root
658 metabolic data were used for PLS analysis, and the R^2 values (adjusted) were used for
659 comparison analysis.

660 The grain yield per plant (in three environments) of 12 maize inbred lines and 66
661 F_1 hybrids from the inbred lines with a half-diallel design reported in a previous study
662 [30] was used for analysis. The expression levels of 27,608 genes, which were pan and
663 core genes in Zheng58, from the 12 inbred lines were also obtained from the study [30].
664 As described in the study [30], each genotype of the 12 maize ILs had two sets (not
665 biological replicate) of RNA-seq data, with one set containing equally mixed RNAs
666 from eight tissues and the other from five tissues at different developmental stages. The
667 expression data were processed with MetaboAnalyst 5.0 [56], with 40% set as the
668 percentage for filtering. No data transformation, normalization, or scaling was
669 performed. In total, 84 DEGs were identified in the 12 inbred lines by ANOVA, with
670 Fisher's LSD used in post hoc analysis (*P* value cut-off = 0.05). The average values of
671 the two sets of RNA sequencing data were used to represent the levels of each inbred
672 line. The 84 genes were classified according to the previously reported results [30]. The

673 number of latent factors was set to 50 in the PLS analysis of better-parent heterosis for
674 grain yield per plant, mid-parent heterosis for grain yield per plant, and grain yield per
675 plant.

676 The phenotypes of cotyledon and the first true leaf and transcriptomics data in
677 *Arabidopsis* were obtained from a previously published study [31]. The transcriptomics
678 data were normalized with above mentioned parameters in NetworkAnalyst 3.0.
679 Differential expression genes between Col-0 and Per-1 at different stages were analyzed
680 with the limma method, with Log2 fold change at 2 and adjusted *p*-value at 0.0001. The
681 pooled differential expression genes across stages were treated as differential
682 expression genes for each tissue. Two-way ANOVA on the normalized levels of DEGs
683 with genotype and stage as metadata was implemented with MetaboAnalyst 5.0 [56].
684 The overlapped DEGs of genotype, stage, and their interaction were used for further
685 analysis. Pearson correlation between expression levels of the DEGs and phenotypes
686 of cotyledon and the first true leaf was performed to screen the genes and the
687 significance level was set at 0.05.

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

689 Seedlings from the 17 rice ILs were ground into fine powder using liquid nitrogen. The
690 genomic DNA was extracted, and the quality and concentration were assessed by 1%
691 agarose gel electrophoresis and Qubit 3.0 fluorometer. The DNA libraries (with index
692 codes) were prepared and sequenced on the Illumina sequencing platform of Novogene
693 Co., Ltd (Beijing, China). In brief, genomic DNA was randomly fragmented into ~350
694 bp size via sonication (Covaris). The strands were then added with adapters and used
695 for PCR amplification. Products were purified with AMPure XP beads (Beckman
696 Coulter) and the concentrations were measured with Agilent 2100 Bioanalyzer.
697 Libraries were sequenced on the Illumina PE150 platform, and the raw data were
698 recorded in FASTQ format (> 15 G per sample). After the clean reads were obtained,
699 the sequences were aligned to the reference genome using the Burrows–Wheeler
700 Aligner (mem -t 4 -k 32 -M, version 0.7.8-r455) [58]. Dislodged duplications in the
701 BAM files were removed via SAMtools (rmdup, version 1.3.1) [59]. SNPs and InDels
702 in each sample were called with the parameters "-C 50 - mpileup -m 2 -F 0.002 -d 1000"
703 and further filtered to reduce errors (depth of variant > 4 and mapping quality > 20).
704 CNVs were detected by CNVnator (-call 100, version 0.3) [60], and SVs were detected

705 by BreakDancer (depth of variant > 2, version 1.4.4) [61]. Variant annotation was
706 implemented with ANNOVAR [62]. Genomic variants on chromosomes were
707 visualized by Circos (0.69-9) [63]. To analyse the population structures of the 17 ILs,
708 GATK (version 4.2.0.0) was first used to filter SNPs (QD < 2, FS > 60, MQ < 40,
709 MQRankSum < -12.5, ReadPosRankSum < -8) and InDels (QD < 2, FS > 200,
710 ReadPosRankSum < -20). SNPs and InDels were further filtered with VCFtools
711 (maximum missing value < 10%, minus allele: 0.05, minimum read depth > 4, multiple
712 alleles removed, version 0.1.17) and analysed with Plink (version 1.90) and Admixture
713 (version 1.3.0). When comparing the SNPs and InDels between each pair of samples of
714 the 17 ILs, the variants were defined as common variants if they shared the same values
715 for five parameters, including CHROM (chromosome), POS (position of the reference),
716 REF (reference base), ALT (alternative base) and GT (genotype). Differential variants
717 were those that were unique to each IL. For CNVs and SVs, variants were defined as
718 common variants if they shared the same values of CHROM, POS1, POS2, TYPE and
719 SIZE. The numbers of common and unique variants were counted, and the sums of
720 unique genomic variants between each pair of ILs were used to represent the genetic

721 distance. Phylogenetic trees (neighbour-joining method) based on one or multiple types

722 of genomic variants in the 17 ILs were generated with MEGA (version 10.2.6) [64].

723 **Statistical analyses**

724 Analysis of variance using SPSS 20.0 was performed with the least significant

725 difference used in the post hoc test to compare the seedling length of the F_1 hybrids

726 and corresponding parents in Fig. 1A, the seedling length of the F_1 hybrids and

727 corresponding better parent in Fig. 2A, the predictabilities in Fig. 2Q, the heterosis for

728 plant height and tiller number at four developmental stages in Fig. 3C, the R^2 values

729 of heterosis for yield per plant with different inheritance patterns in Fig. 3G, the

730 greatest R^2 values of heterosis for 17 traits with different inheritance patterns in Fig.

731 3H, the greatest R^2 values of heterosis for eight traits with different inheritance

732 patterns in Fig. 3I, the numbers (weighted) of correlations for four types of genomic

733 variants in Fig. 5C and Fig. S20I, and correlation numbers (weighted) of the three

734 inheritance patterns in Fig. 5F-I. An independent samples *t* test was performed to

735 compare the numbers (weighted) of differentially expressed genes between F_1 hybrids

736 and corresponding parents in Fig. S1C, percentages of differential molecules between

737 F_1 hybrids and corresponding parents in parental differences in Fig. S3J, the node

738 numbers (weighted), predictability, and percentage of Component 1 of three modules

739 from Subnetwork_1 between Network_51015 and Network_HLN in Fig. 2E-G, the

740 numbers (weighted) of correlations between five metabolites from phenylpropanoid

741 biosynthesis and heterosis in Figs. 3F and Fig. S11H, trait values and heterosis for

742 four traits between the diallel and test cross populations in Fig. S14B and Fig. S14F,

743 and the numbers (weighted) of correlations for four types of genomic variants in Fig.

744 5B. Spearman rank correlation was performed between the numbers of genetic

745 variants and heterosis in Fig. 5A, Figs. 5F-I, Figs. 20A-F, and Figs. 20H. Pearson

746 correlation analysis (two-tailed) with significance level set at 0.05 was conducted

747 between gene expression (1,201 genes)/metabolite (1,249 metabolites) levels and

748 seedling length to screen for heterosis-associated molecules in Fig. S2J and Fig. S4H,

749 between seedling length and heterosis-associated metabolites from three enriched

750 metabolic pathways in Fig. 5A, among better-parent heterosis for 17 traits in Fig. 3D,

751 between three types of inheritance patterns of five metabolites and heterosis in Fig.

752 3E, and between three types of inheritance patterns for the 50 top contributing

753 analytes and heterosis in Fig. S20I. Partial least squares regression between gene
754 expression or metabolite levels and seedling length in Fig. S2J and Fig. S4H, between
755 parental differential metabolites and heterosis in Fig. 3G-N, Fig. S11I-X, and Fig.
756 S13A-H were implemented with SPSS. The maximum number of latent factors was
757 initially set at 100 to extract enough latent factors. The regression model for each trait
758 was determined when the value of the adjusted R^2 was the highest. The means of
759 variable importance in the projection of all latent factors were used to estimate the
760 contributions of each variable, and threshold was set to 1 in Fig. 2I. Partial least
761 squares-discriminant analysis (PLS-DA) in Fig. 2I, Fig. 2O, and Fig. 2Q was
762 performed with class order taken into account in MetaboAnalyst 5.0 [56]. The
763 maximum number of components to be searched was 5 and tenfold cross validation
764 was chosen. The separation distance (B/W) method was used for permutation analysis
765 (2,000 times) to assess whether the model was overfitting. Q^2 was chosen as the
766 predictability in PLS-DA models. Venn diagrams were drawn with the tool from
767 http://bioinformatics.psb.ugent.be/webtools/Venn/. The phylogenetic tree of the 223
768 ILs shown in Fig. 3B was based on genotypes and the tree was constructed using

MEGA [64] 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 [65] (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. The statistical source data have been provided in Data S5.

Acknowledgements

The authors thank members of the Zhu lab for assistance in the collection of field phenotypic data. The authors thank Professor Baorong Lu (Fudan University) for the design of the diallel cross population. The authors thank Mr. Caibin Zhu for the fund of rice research. The authors thank Fulin Luo for the planting of rice. The authors thank Dr. Xiaofei Zeng for suggestions.

Authors' 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.

Funding

This work is supported by National Natural Science Foundation of China (31771746 to W.H., 31801439 to Z.D., and 32101667 to Yunping Chen), the China Postdoctoral Science Foundation (2018M632910, 2019M660186, and 2022T150500 to Z.D., and 2023T160497 to Yunping Chen), 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..

Competing interests

The authors declare no competing interests.

Availability of data and materials

All data are available in the main text or the materials. 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 and MTBL1344. Source data are available with this paper.

Supplementary Information

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

Fig. S1 Transcriptomic differences in two inbred lines (Balilla and Yuetai B) and their reciprocal F_1 hybrids at three developmental stages.

Fig. S2 The identification of heterosis-associated genes related to seedling length from the differentially expressed genes between F_1 hybrids and their corresponding parents.

Fig. S3 Differential metabolites and pathways between parents and between F_1 hybrids and corresponding parents.

Fig. S4 The identification of heterosis-associated metabolites related to seedling length from differential metabolites between F_1 hybrids and their corresponding

822 parents.
823 Fig. S5 Heterosis-associated molecules related to seedling length.
824 Fig. S6 Validation of 626 heterosis-associated genes and 645 heterosis-associated
825 metabolites for seedling length.
826 Fig. S7 Validation of network modules to heterosis for seedling length.
827 Fig. S8 Score plots of inbred lines, F_1 hybrids, and/or quality controls based on
828 transcriptomic and/or metabolomic data.
829 Fig. S9 Inheritance patterns of heterosis-associated molecules for seedling length.
830 Fig. S10 Phenotypes of seventeen agronomic traits of the diallel cross population.
831 Fig. S11 Predictive models for heterosis of 16 traits based on three types of inheritance
832 patterns of 1,306 differential metabolites.
833 Fig. S12 Phenotypes of eight traits across environments.
834 Fig. S13 PLS-based predictive models for heterosis of eight traits.
835 Fig. S14 Cross design and four traits of the test cross population.
836 Fig. S15 Additive and partially dominant effects contribute to maize heterosis.
837 Fig. S16 Additive and partially dominant effects contribute to heterosis in *Arabidopsis*.
838 Fig. S17 SNP and InDel density on 12 chromosomes of 16 rice ILs.
839 Fig. S18 Genomic variants of 16 rice ILs.
840 Fig. S19 Population structure and neighbor-joining trees of the 17 rice ILs based on
841 genomic variants.
842 Fig. S20 Correlations of genomic variants, heterosis, and different inheritance patterns.
843 Table S1 Seedling length of six rice inbred lines and corresponding F1 hybrids.
844 Table S2 Summary of RNA-seq data of rice seedling samples.
845 Table S3. Summary of genomic data of 17 rice inbred lines.
846 Data S1 RNA-sequencing result (read count) of six pairs of reciprocal rice F1 hybrids
847 at seedling stage.
848 Data S2 Untargeted metabolite profiling result (intensity) of ten pairs of reciprocal rice
849 F1 hybrids at seedling stage.
850 Data S3 Genotype of 223 rice inbred lines estimated by 33 pairs of InDel markers.
851 Data S4 Metabolite profiles of rice inbred lines, recombinant inbred lines, and quality
852 control samples.
853 Data S5 Statistical source data.

854 # References

855 1. Birchler JA, Yao H, Chudalayandi S. Unraveling the genetic basis of hybrid vigor. Proc
856 Natl Acad Sci USA. 2006;103:12957-12958.
857 2. Shull GH. What is "heterosis"? Genetics. 1948;33:439-446.
858 3. Birchler JA, Yao H, Chudalayandi S, Vaiman D, Veitia RA. Heterosis. Plant Cell.
859 2010;22:2105-2112.
860 4. Hochholdinger F, Baldauf JA. Heterosis in plants. Curr Biol. 2018;28:R1089-R1092.
861 5. Bruce AB. The Mendelian theory of heredity and the augmentation of vigor. Science.
862 1910;32:627-628.
863 6. Jones DF. Dominance of linked factors as a means of accounting for heterosis. Proc Natl
864 Acad Sci U S A. 1917;3:310-312.
865 7. Li X, Li X, Fridman E, Tesso TT, Yu J. Dissecting repulsion linkage in the dwarfing gene
866 *Dw3* region for sorghum plant height provides insights into heterosis. Proc Natl Acad Sci
867 USA. 2015;112:11823-11828.
868 8. Davenport CB. Degeneration, albinism and inbreeding. Science. 1908;28:454-455.
869 9. Xiao Y, Jiang S, Cheng Q, Wang X, Yan J, Zhang R, Qiao F, Ma C, Luo J, Li W, et al. The
870 genetic mechanism of heterosis utilization in maize improvement. Genome Biol.
871 2021;22:148.
872 10. East EM. Heterosis. Genetics. 1936;21:375-397.
873 11. Crow JF. Alternative hypotheses of hybrid vigor. Genetics. 1948;33:477–487.
874 12. Yu SB, Li JX, Xu CG, Tan YF, Gao YJ, Li XH, Zhang Q, Saghai Maroof MA. Importance of
875 epistasis as the genetic basis of heterosis in an elite rice hybrid. Proc Natl Acad Sci USA.
876 1997;94:9226-9231.
877 13. Li Z, Pinson S, Park W, Paterson A, Stansel J. Epistasis for three grain yield components
878 in rice (*Oryza sativa* L.). Genetics. 1997;145:453-465.
879 14. Powers L. An expansion of Jones's theory for the explanation of heterosis. Am Nat.
880 1944;78:275-280.
881 15. Schnable PS, Springer NM. Progress toward understanding heterosis in crop plants.
882 Annu Rev Plant Biol. 2013;64:71-88.
883 16. Lisec J, Romisch-Margl L, Nikoloski Z, Piepho HP, Giavalisco P, Selbig J, Gierl A,
884 Willmitzer L. Corn hybrids display lower metabolite variability and complex metabolite
885 inheritance patterns. Plant J. 2011;68:326-336.

886 17. Krieger U, Lippman ZB, Zamir D. The flowering gene *SINGLE FLOWER TRUSS* drives
887 heterosis for yield in tomato. Nat Genet. 2010;42:459-463.
888 18. Williams W. Heterosis and the genetics of complex characters. Nature. 1959;184:527-
889 530.
890 19. Dan Z, Hu J, Zhou W, Yao G, Zhu R, Huang W, Zhu Y. Hierarchical additive effects on
891 heterosis in rice (*Oryza sativa* L.). Front Plant Sci. 2015;6:738.
892 20. Wang Z, Xue Z, Wang T. Differential analysis of proteomes and metabolomes reveals
893 additively balanced networking for metabolism in maize heterosis. J Proteome Res.
894 2014;13:3987-4001.
895 21. Zhou P, Hirsch CN, Briggs SP, Springer NM. Dynamic patterns of gene expression
896 additivity and regulatory variation throughout maize development. Mol Plant.
897 2019;12:410-425.
898 22. Dan Z, Chen Y, Li H, Zeng Y, Xu W, Zhao W, He R, Huang W. The metabolomic
899 landscape of rice heterosis highlights pathway biomarkers for predicting complex
900 phenotypes. Plant Physiol. 2021;187:1011-1025.
901 23. Dan Z, Chen Y, Zhao W, Wang Q, Huang W. Metabolome-based prediction of yield
902 heterosis contributes to the breeding of elite rice. Life Sci Alliance. 2020;3:e201900551.
903 24. Dan Z, Chen Y, Xu Y, Huang JR, Huang JS, Hu J, Yao G, Zhu Y, Huang W. A
904 metabolome-based core hybridisation strategy for the prediction of rice grain weight
905 across environments. Plant Biotechnol J. 2019;17:906-913.
906 25. Dan Z, Hu J, Zhou W, Yao G, Zhu R, Zhu Y, Huang W. Metabolic prediction of important
907 agronomic traits in hybrid rice (*Oryza sativa* L.). Sci Rep. 2016;6:21732.
908 26. Sun Z, Peng J, Lv Q, Ding J, Chen S, Duan M, He Q, Wu J, Tian Y, Yu D, et al. Dissecting
909 the genetic basis of heterosis in elite super-hybrid rice. Plant Physiol. 2023;192:307-325.
910 27. Li S, Park Y, Duraisingham S, Strobel FH, Khan N, Soltow QA, Jones DP, Pulendran B.
911 Predicting network activity from high throughput metabolomics. PLoS Comput Biol.
912 2013;9:e1003123.
913 28. de Abreu ELF, Westhues M, Cuadros-Inostroza Á, Willmitzer L, Melchinger AE, Nikoloski
914 Z. Metabolic robustness in young roots underpins a predictive model of maize hybrid
915 performance in the field. Plant J. 2017;90:319-329.
916 29. Schrag T, Westhues M, Schipprack W, Seifert F, Thiemann A, Scholten S, Melchinger AE.
917 Beyond genomic prediction: combining different types of *omics* data can improve
918 prediction of hybrid performance in maize. Genetics. 2018;208:1373-1385.

30. Wang B, Hou M, Shi J, Ku L, Song W, Li C, Ning Q, Li X, Li C, Zhao B, et al. De novo genome assembly and analyses of 12 founder inbred lines provide insights into maize heterosis. Nat Genet. 2023;55:312-323.
31. Liu W, He G, Deng XW. Biological pathway expression complementation contributes to biomass heterosis in Arabidopsis. Proc Natl Acad Sci U S A. 2021;118:
32. Dan Z, Liu P, Huang W, Zhou W, Yao G, Hu J, Zhu R, Lu B, Zhu Y. Balance between a higher degree of heterosis and increased reproductive isolation: a strategic design for breeding inter-subspecific hybrid rice. PLoS ONE. 2014;9:e93122.
33. Huang X, Yang S, Gong J, Zhao Q, Feng Q, Zhan Q, Zhao Y, Li W, Cheng B, Xia J, et al. Genomic architecture of heterosis for yield traits in rice. Nature. 2016;537:629-633.
34. Price ND, Magis AT, Earls JC, Glusman G, Levy R, Lausted C, McDonald DT, Kusebauch U, Moss CL, Zhou Y, et al. A wellness study of 108 individuals using personal, dense, dynamic data clouds. Nat Biotechnol. 2017;35:747-756.
35. Zhang M, Wang Y, Hu Y, Wang H, Liu Y, Zhao B, Zhang J, Fang R, Yan Y. Heterosis in root microbiota inhibits growth of soil-borne fungal pathogens in hybrid rice. J Integr Plant Biol. 2023;65:1059-1076.
36. Xu Y, Zhao Y, Wang X, Ma Y, Li P, Yang Z, Zhang X, Xu C, Xu S. Incorporation of parental phenotypic data into multi-omic models improves prediction of yield-related traits in hybrid rice. Plant Biotechnol J. 2021;19:261-272.
37. Kakoulidou I, Piecyk RS, Meyer RC, Kuhlmann M, Gutjahr C, Altmann T, Johannes F. Mapping parental DMRs predictive of local and distal methylome remodeling in epigenetic F1 hybrids. Life Sci Alliance. 2024;7:
38. Liu N, Du Y, Warburton ML, Xiao Y, Yan J. Phenotypic plasticity contributes to maize adaptation and heterosis. Mol Biol Evol. 2021;38:1262-1275.
39. Lu B, Cai X, Jin X. Efficient *indica* and *japonica* rice identification based on the InDel molecular method:its implication in rice breeding and evolutionary research. Progress in Nature Science. 2009;19:1241-1252.
40. Yao G, Huang W. Identifying *indica-japonica* characteristic of rice breeding materials with *indica-japonica* differentiation InDel markers. Hybrid Rice. 2013;28:53-57.
41. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12:357-360.
42. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923-930.

43. Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. Nucleic Acids Res. 2019;47:W234-W241.
44. Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, Wishart DS, Xia J. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids Res. 2018;46:W486-W494.
45. Metsalu T, Vilo J. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. Nucleic Acids Res. 2015;43:W566-570.
46. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pages F, Trajanoski Z, Galon J. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics. 2009;25:1091-1093.
47. Bindea G, Galon J, Mlecnik B. CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. Bioinformatics. 2013;29:661-663.
48. Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, Lotia S, Pico AR, Bader GD, Ideker T. A travel guide to Cytoscape plugins. Nat Methods. 2012;9:1069-1076.
49. Chen Y, Dan Z, Li S. *GROWTH REGULATING FACTOR 7*-mediated arbutin metabolism enhances rice salt tolerance. Plant Cell. 2024;
50. Chambers MC, MacLean B, Burke R, Amode D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt B, Egertson J, et al. A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotechnol. 2012;30:918-920
51. Gowda H, Ivanisevic J, Johnson CH, Kurczy ME, Benton HP, Rinehart D, Nguyen T, Ray J, Kuehl J, Arevalo B, et al. Interactive XCMS Online: simplifying advanced metabolomic data processing and subsequent statistical analyses. Anal Chem. 2014;86:6931-6939.
52. Jia H, Shen X, Guan Y, Xu M, Tu J, Mo M, Xie L, Yuan J, Zhang Z, Cai S, et al. Predicting the pathological response to neoadjuvant chemoradiation using untargeted metabolomics in locally advanced rectal cancer. Radiother Oncol. 2018;128:548-556.
53. Shen X, Wang R, Xiong X, Yin Y, Cai Y, Ma Z, Liu N, Zhu Z-J. Metabolic reaction network-based recursive metabolite annotation for untargeted metabolomics. Nat Commun. 2019;10:1516.
54. Tu J, Yin Y, Xu M, Wang R, Zhu ZJ. Absolute quantitative lipidomics reveals lipidome-wide alterations in aging brain. Metabolomics. 2017;14:5.
55. Domingo-Almenara X, Montenegro-Burke JR, Ivanisevic J, Thomas A, Sidibe J, Teav T,

Guijas C, Aisporna AE, Rinehart D, Hoang L, et al. XCMS-MRM and METLIN-MRM: a cloud library and public resource for targeted analysis of small molecules. Nat Methods. 2018;15:681-684.

56. Pang Z, Zhou G, Ewald J, Chang L, Hacariz O, Basu N, Xia J. Using MetaboAnalyst 5.0 for LC-HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. Nat Protoc. 2022;
57. Zhou G, Ewald J, Xia J. OmicsAnalyst: a comprehensive web-based platform for visual analytics of multi-omics data. Nucleic Acids Res. 2021;
58. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754-1760.
59. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078-2079.
60. Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 2011;21:974-984.
61. Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, McGrath SD, Wendl MC, Zhang Q, Locke DP, et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat Methods. 2009;6:677-681.
62. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.
63. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19:1639-1645.
64. Stecher G, Tamura K, Kumar S. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. Mol Biol Evol. 2020;37:1237-1239.
65. Spitzer M, Wildenhain J, Rappsilber J, Tyers M. BoxPlotR: a web tool for generation of box plots. Nat Methods. 2014;11:121-122.

Figures

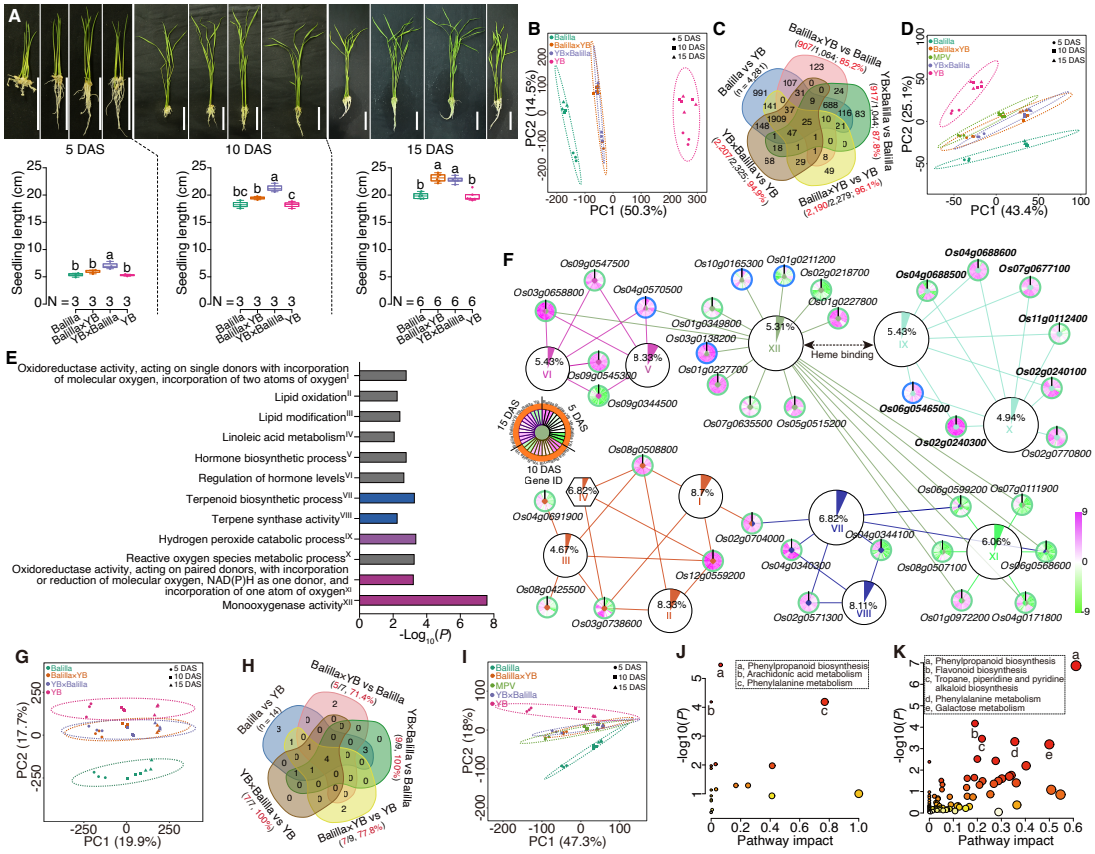

Fig. 1 Heterosis-associated genes and metabolites related to rice seedling length. (**A**) Seedlings and seedling length of Balilla, Yuetai B (YB), and their reciprocal F_1 hybrids at 5, 10, and 15 days after sowing (DAS). The significance level was set at 0.05. Scale bar = 5 cm. (**B**) PCA score plot 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.

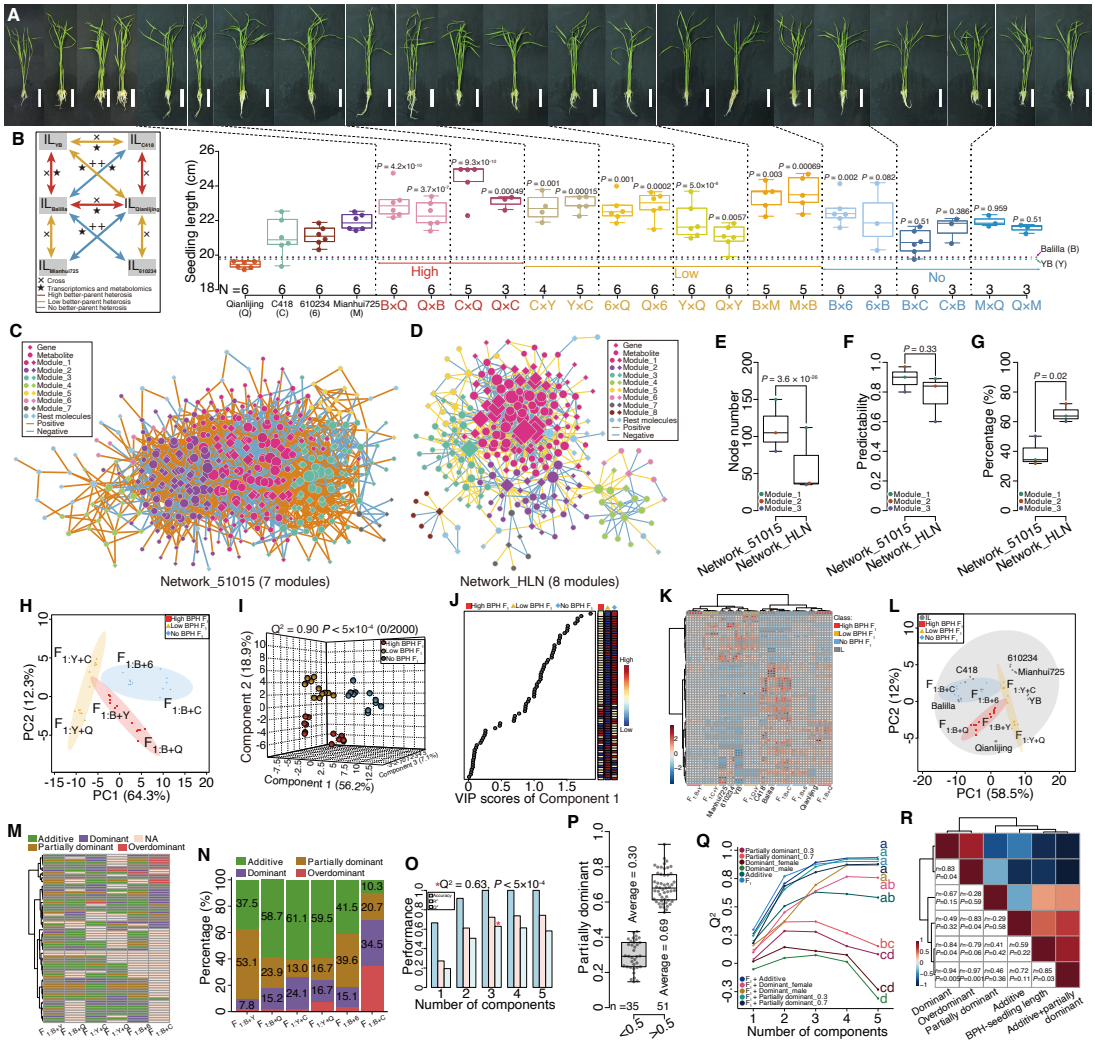

Fig. 2 Inheritance patterns of heterosis-associated genes and metabolites. (**A**) Seedlings and seedling length of nine pairs of reciprocal F_1 hybrids. *P* values indicate the least significant difference in the post hoc test for the seedling length of the F_1 hybrid compared with that of its better 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.

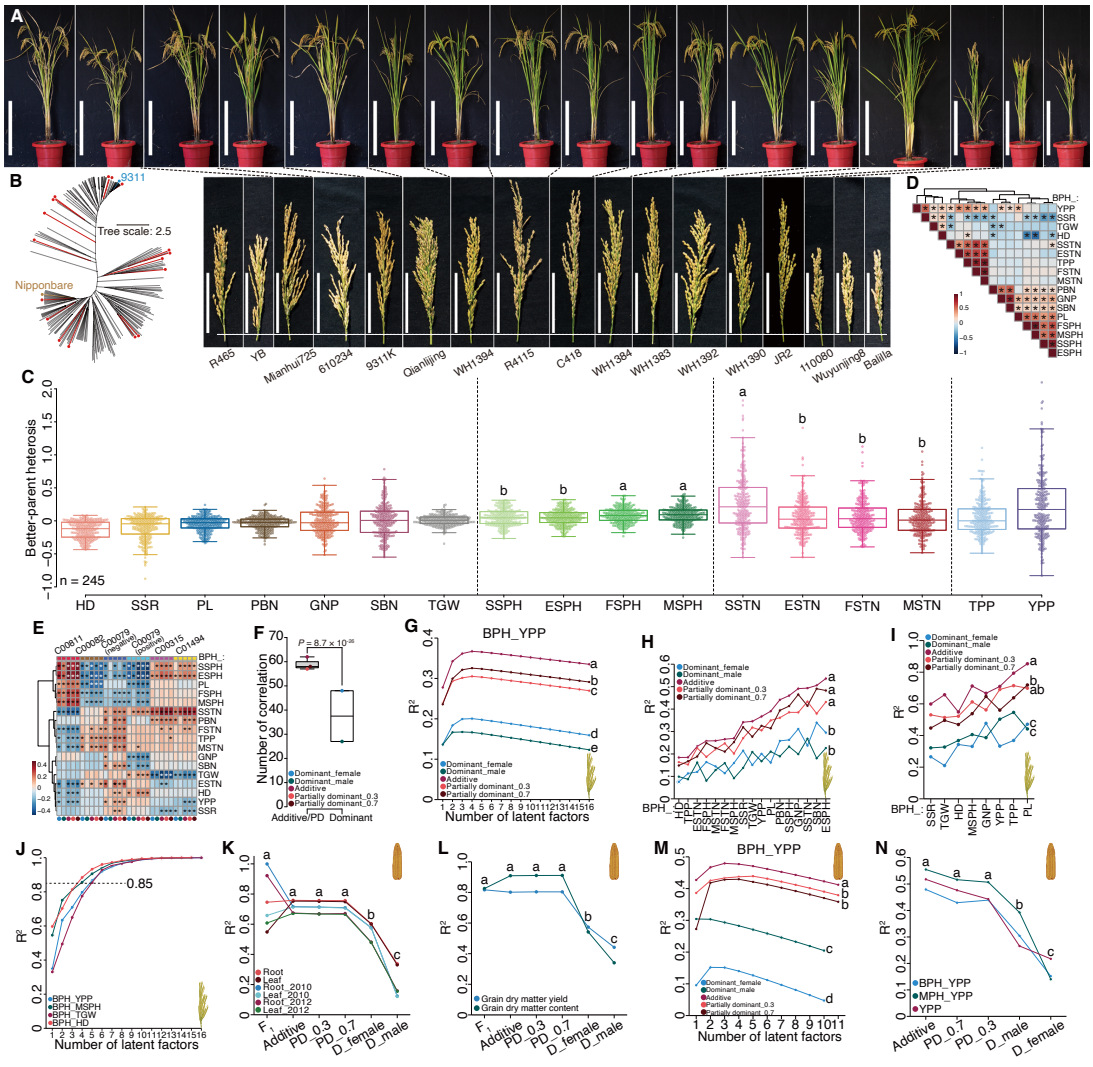

Fig. 3 Additive and partially dominant effects contribute to heterosis of agronomic traits in both rice and maize. (**A**) Plants and panicles of 17 rice inbred lines (ILs). Scale bars represent 50 cm and 10 cm for plants and panicles, respectively. (**B**) Neighbour-joining tree of 223 rice ILs. (**C**) 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**) Top R^2 values for heterosis of eight traits. (**J**) R^2 values for heterosis of four 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 (mid-parent heterosis), and YPP in maize. The R^2 are with adjusted values.

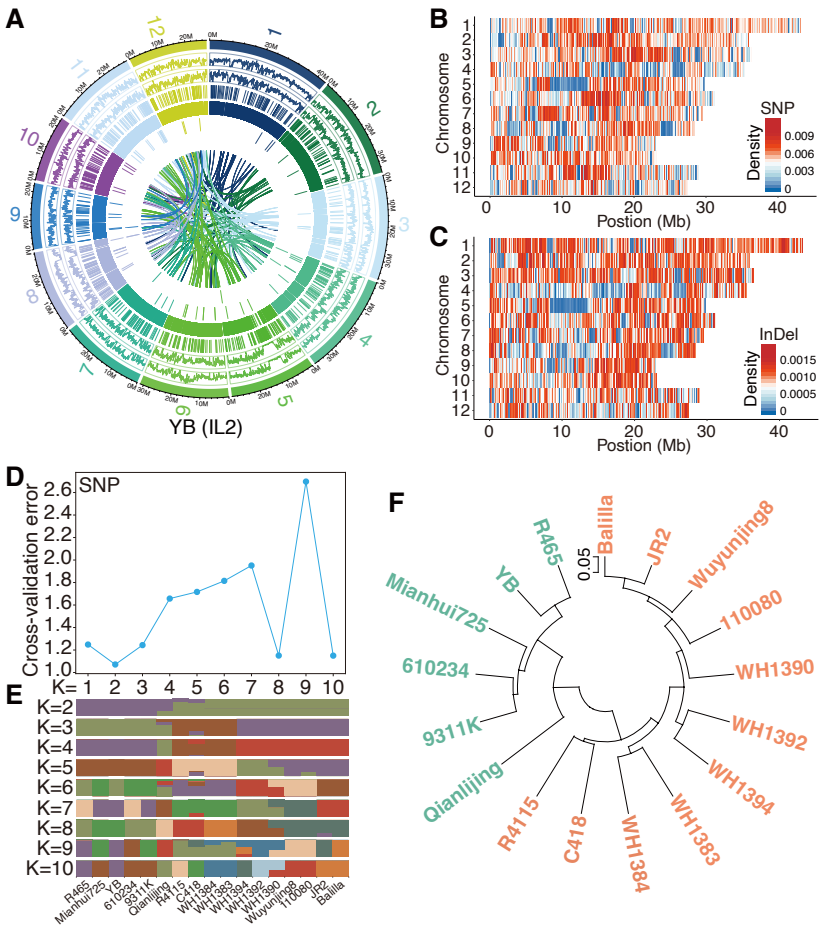

Fig. 4 Genomic variants of 17 rice inbred lines. (**A**) Genomic variants of Yuetai B. Chromosome position, SNP density, InDel density, CNV duplication position, CNV deletion position, SV insertion position, SV deletion position, and SV inversion position are listed from the outer to inner circles. (**B-C**) 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.

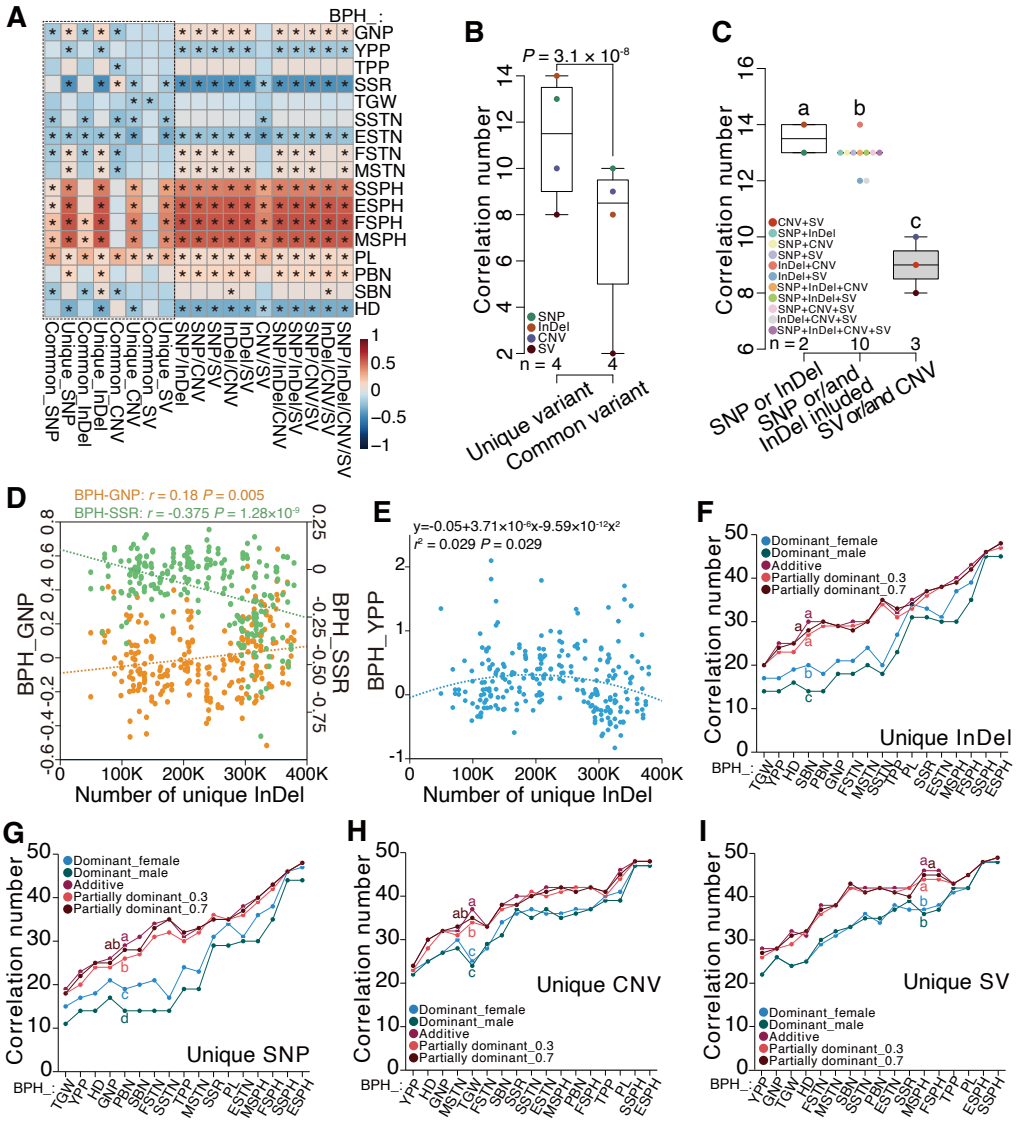

Fig. 5 Parental genomic variants are associated with additive and partially dominant effects. **(A)** Heatmap of correlations between parental genomic variants and heterosis of 17 traits. **(B)** Correlation numbers of parental unique and common genomic variants. *P* value indicates independent samples *t* test. **(C)** Correlation numbers of different types of parental unique genomic variants. **(D)** 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. **(E)** Quadratic regression between the number of parental unique InDels and heterosis for grain yield per plant. **(F-I)** Correlation numbers of parental unique InDels (**F**), SNPs (**G**), CNVs (**H**), and SVs (**I**).

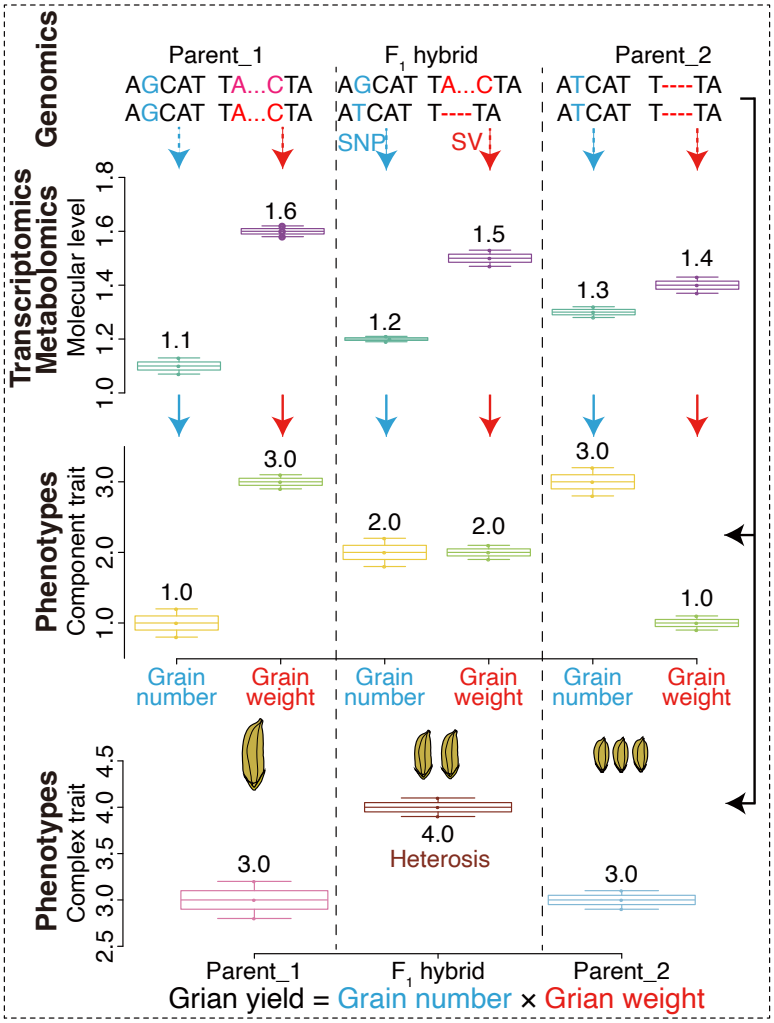

Fig. 6 Model for the contribution of additive effect to positive heterosis of a complex trait in rice. The unique SNP and SV in parents regulate grain number and grain weight, respectively. Meanwhile, these genomic variants are associated with transcriptomic and metabolomic additive effects in F_1 hybrids. The grain number and grain weight of F_1 hybrids are at the mid-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.