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.

[†]These authors contributed equally to this article.

*Correspondence: Wenchao Huang (wenchaoh@whu.edu.cn)

Summary

Heterosis, or hybrid vigor, describes the superior performance of F_1 hybrids relative to their parents. Despite its significant importance in crop breeding, the molecular mechanisms underlying heterosis remain debated, mainly attributable to discrepancies across genotypes, traits, tissues, populations, developmental stages, growth environments, and species. In this study, we systematically identified heterosis-associated genes and metabolites derived from parental molecular differences for seedling length in rice and incorporated them into network modules to explain the variance of heterosis. The predominant inheritance patterns of these heterosis-associated molecules were additive and partially dominant effects, namely at mid-parent levels or values between mid-parent and parental levels, respectively. These two genetic effects contributed to heterosis of 17 agronomic traits in rice, including grain yield and plant height across four developmental stages. They also explained yield heterosis in diverse hybrid populations and distinct growth environments in both rice and maize, as well as biomass heterosis in *Arabidopsis*. Notably, additive and partially dominant effects were associated with parental genomic variants in rice, and the number of these variants correlated significantly with heterosis of agronomic traits. Furthermore, we demonstrated the significant impact of parental genomic, transcriptomic, and metabolomic variation in phenylpropanoid biosynthesis on heterosis for seedling length/plant height. Unlike classical heterosis models primarily focused on genomic sequence variation, our findings provide quantitative insights from genomic downstream information into the molecular mechanisms of plant heterosis,

highlighting their potential for improving breeding efficiency of hybrid crops.

Keywords: heterosis, additive effect, heterosis-associated genes, heterosis-associated metabolites, genomic variant, yield, rice

Introduction

Heterosis describes the phenomenon where heterozygous F_1 hybrids outperform their parents (Birchler et al., 2006; Birchler et al., 2010; Shull, 1948). The exploitation of heterosis to enhance crop yield represents a landmark innovation in modern agriculture for mitigating food shortages (Hochholdinger and Baldauf, 2018). Several well-known models have been proposed to explain heterosis, including dominance (pseudo-overdominance) (Bruce, 1910; Davenport, 1908; Jones, 1917; Li et al., 2015; Xiao et al., 2021), overdominance (Crow, 1948; East, 1936), and epistasis (Li et al., 1997; Powers, 1944; Yu et al., 1997), emphasizing complementation, heterozygosity, and interactions of loci at the genomic level, respectively. However, the molecular mechanisms underpinning crop heterosis remain controversial, largely due to diversity in factors such as genotypes, phenotypes, tissues, hybrid populations, developmental stages, growth environments, and species.

To quantitatively describe the inheritance patterns of hybrid phenotypes and molecular levels of F_1 hybrids, geneticists used four key terms: additive effect (mid-parent value, MPV), partially dominant effect (between MPV and parental values), dominant effect (parental values), and overdominant effect (beyond parental values) (Lisec et al., 2011; Schnable and Springer, 2013). Accumulating evidence suggests a

connection between additive effect and heterosis. For example, the near-additive expression of *SINGLE FLOWER TRUSS* drives strong yield heterosis in tomato (Krieger et al., 2010). Moreover, additive effect of transcripts, proteins, metabolites, and component traits has been implicated as the genetic basis of heterosis in maize, rice, and tomato (Dan et al., 2021; Dan et al., 2019; Dan et al., 2020; Dan et al., 2015; Dan et al., 2016; Wang et al., 2014; Williams, 1959; Zhou et al., 2019b). Partially dominant effect has also been suggested to play a role in heterosis (Jones, 1917; Powers, 1944; Sun et al., 2023; Williams, 1959).

In this study, we systematically identified heterosis-associated genes and metabolites for seedling length in rice and analyzed the proportions and contributions of the four inheritance patterns. Our results demonstrate that additive and partially dominant effects predominantly account for plant heterosis, a finding consistent across diverse traits, hybrid populations, developmental stages, growth environments, omics data types, and even species.

Results

Heterosis-associated genes identified from parental transcriptomic differences

To identify heterosis-associated genes (HAGs) in rice, we cultivated cultivars Yuetai B (YB, *indica*), Balilla (*japonica*), and their reciprocal F_1 hybrids under sterile conditions. Seedling length and weight were measured at 5, 10, and 15 days after sowing (DAS) (Figure 1a and Table S1). The two traits were positively correlated in parents and F_1 hybrids and seedling length was selected for subsequent analysis (Figure S1a).

86 Unexpectedly, one F_1 hybrid (YB×Balilla) exhibited positive better-parent heterosis
87 (BPH) as early as 5 DAS (Figure 1a). Both F_1 hybrids exhibited positive BPH at 15
88 DAS. Moreover, genotype, stage, and their interaction explained nearly all variance in
89 seedling length (adjusted R^2 = 0.99, Figure S1b), confirming minimal environmental
90 influence on phenotypic changes.

91 We subsequently performed transcriptomic analysis on these seedling samples and
92 used the normalized expression levels of 22,194 genes for principal component analysis
93 (PCA; Table S2 and Dataset S1). Though the two parental inbred lines (ILs) had no
94 difference in seedling length at each stage, they located at different regions of the plot,
95 inferring the existence of gene expression heterogeneity among rice ILs. The F_1 hybrids
96 occupied an intermediate position between the parents, with a closer alignment to
97 Balilla and greater number of differentially expressed genes (DEGs) to YB (Figure 1b
98 and Figure S1c). The number of DEGs among the two ILs and F_1 hybrids increased
99 over time (Figure S1d-e), implicating the involvements of more genes during
100 development. Unlike the pooled 16,350 non-DEGs (Figure S1f), the 5,844 DEGs
101 explained 66.6% of sample variance in genotypes along principal component 1 (PC1;
102 Figure S1g). Furthermore, the F_1 hybrids deviated slightly from "additive hybrids"
103 (calculated as MPVs), suggesting contributions from additive and non-additive effects
104 to hybrid transcriptomes.

105 The DEGs were then treated as candidates underlying phenotypic changes between
106 F_1 hybrids and their parents. The parental ILs harbored the most DEGs at each stage.
107 Crucially, 83.5-97.1% (mean 92.38%) of hybrid-parent DEGs originated from parental

108 expression differences (Figure 1c and Figure S2a-b). With Balilla showing higher
109 expression levels than that of YB across stages, the F_1 hybrids showed more up-
110 regulated than down-regulated DEGs relative to each parent (Figure S2c). Furthermore,
111 the significantly enriched Gene Ontology (GO) terms for hybrid-parent comparisons
112 largely overlapped with those for parental ILs (Figure S2d-i), indicating that parental
113 expression differences predominantly shape hybrid transcriptomes.

114 To focus on genes mainly affected by genotype rather than developmental stage,
115 we grouped DEGs using analysis of variance (ANOVA; Figure S3a). We identified
116 1,476 DEGs responsive to genotype, stage, and their interaction. After excluding 275
117 DEGs specific to ILs or reciprocal F_1 hybrids (Figure S3b), 1,201 DEGs remained.
118 These DEGs retained the sample distribution patterns of the full gene set (Figure S3c),
119 with 79.27% up-regulated in F_1 hybrids (Figure S3d). These genes participated in 17
120 GO terms, ten of which overlapped with parental differential terms (Figure S3e).

121 Correlation analysis between expression levels of the 1,201 DEGs and seedling
122 length identified 626 genes with significant correlations ($P < 0.05$). These genes also
123 contributed substantially to seedling length in partial least squares (PLS) analysis
124 (Figure S3f-g). Moreover, correlation coefficients strengthened in F_1 hybrids versus
125 parents when shifting from the full gene set to these core genes (Figure S3h-j),
126 validating them as HAGs.

127 Unexpectedly, PCA of the 626 HAGs positioned F_1 hybrids near MPVs (Figure 1d).
128 Among these, 483 HAGs were up-regulated in F_1 hybrids at one or more stages (Figure
129 S3k), inferring phased up-regulation drives heterosis. Most up-regulated HAGs

correlated positively with seedling length, and about 50% of down-regulated correlated negatively (Figure S3l). GO enrichment analysis of 499 positively and 127 negatively correlated HAGs revealed 37 HAGs in 12 terms (Figure 1e). These formed a tightly connected network bridged by heme binding (Figure S4), demonstrating functional coherence of genes involved in seedling length heterosis.

Heterosis-associated metabolites from parental metabolomic variation

Accordingly, we performed untargeted metabolite profiling analysis on these seedling samples. PCA of metabolite profiles revealed that the F_1 hybrids clustered intermediately between their parents but positioned closer to YB (Figure 1f, Figure S5a, and Dataset S2). Differential metabolite (DM) analysis of ILs and F_1 hybrids showed that, unlike DEGs, the number of DMs decreased during development (Figure S5b-c). Genotypic variation was clearly indicated along PC1 based on pooled DMs, consistent with the transcriptomic patterns (Figure S5d-e). Meanwhile, the DMs were generally up-regulated in reciprocal F_1 hybrids, with no significant difference between YB and Balilla (Figure S5f). Although the proportions were lower than in transcriptomic data (Figure S5g), most metabolomic differences between F_1 hybrids and their parents (average 84.0%) originated from parental variation (Figure S5h-j). Pathway enrichment analysis of DMs showed that the majority of significantly enriched metabolic pathways distinguishing F_1 hybrids from parents were also divergent between the two parents (Figure 1g and Figure S6), further supporting that those parental molecular differences primarily determined hybrid-parent molecular variation.

Combining ANOVA, PLS, and correlation analyses identified 645 heterosis-

152 associated metabolites (HAMs) of seedling length (Figure S7). With these metabolites,
153 the F_1 hybrids exhibited metabolite levels closely resembling "additive hybrids"
154 (Figure 1h). Based on annotation information with *mummichog* algorithm (Li et al.,
155 2013), we found eight HAMs were significantly enriched in three pathways:
156 phenylpropanoid biosynthesis, arachidonic acid metabolism, and phenylalanine
157 metabolism (Figure 1i). After chemically annotated metabolites from these pathways
158 were added for correlation analysis, we found that they were positively or negatively
159 correlated with seedling length (Figure S8), indicating distinct contributions from
160 metabolites within the same pathway. Notably, three chemically annotated metabolites
161 from phenylpropanoid biosynthesis (4-hydroxycinnamic acid, L-phenylalanine, and
162 trans-ferulic acid) all showed significant correlations with seedling length in F_1 hybrids,
163 but not in their parents (Figure S9).

164 Joint pathway analysis integrating the 626 HAGs and 645 HAMs identified five
165 significantly enriched pathways (Figure 1j). Phenylpropanoid biosynthesis was the
166 most significant pathway, containing eight metabolites and ten genes. Among these
167 genes, nine encoded peroxidases, seven of which were included in the 37 previously
168 identified HAGs (Figure S10). This finding underscored the combined contributions of
169 transcriptomic and metabolomic changes in phenylpropanoid biosynthesis to seedling
170 length heterosis.

171 **Integrating parental differential molecules into network modules explains the**
172 **variance of heterosis**

173 To validate contributions of the identified HAGs and HAMs to heterosis, we analyzed

174 phenotypic and omics data from nine additional reciprocal F_1 hybrid pairs, each
175 exhibiting varying level of BPH for seedling length (Figure 2a-b). Neither PCA nor
176 partial least squares-discriminant analysis (PLS-DA) successfully separated hybrids
177 into high-, low-, and no-BPH groups using the 626 HAGs (Figure S11a-d). The 645
178 HAMs clearly distinguished the high-BPH group from others without overfitting but
179 failed to classify low- and no-BPH groups (Figure S11e-h). This indicated that HAGs
180 or HAMs specific to a single hybrid pair had limited predictive power for heterosis in
181 F_1 hybrids with divergent genetic background.

182 We next integrated HAGs and HAMs into Network_51015, whose largest
183 subnetwork (subnetwork_1) comprised seven modules and 2,138 edges (Figure S12A).
184 Unlike the decreasing trend in parents for the top three modules, coefficients of module
185 levels (average node levels) and seedling length maintained greater than 0.97 in F_1
186 hybrids (Figure S12b-d). Notably, PLS-DA separated F_1 hybrids into distinct groups,
187 with Module_2 showing high predictability (Q^2 = 0.97; Figure S12e-j). This result
188 indicated that integrating HAGs and HAMs into network modules enhanced the
189 predictability of seedling length heterosis.

190 We identified 967 DEGs and 990 DMs across six ILs of the ten reciprocal F_1 hybrid
191 pairs. The DEGs displayed better separation in PLS-DA than PCA, with predictability
192 comparable to DMs (Figure S13a-h). Combining gene expression and metabolite levels
193 separated the three groups along PC1 and PC4 but caused overfitted in PLS-DA (Figure
194 S13i-l). We then constructed Network_HLN through integrating these DEGs and DMs,
195 with the largest subnetwork (subnetwork_1) contained eight modules and 868 edges

196 (Figure 2c). Though the top three modules in subnetwork_1 from Network_HLN had
197 fewer nodes than those in Network_51015 (Figure 2d), their predictabilities were
198 comparable (Figure 2e and Figure S14a-f), and Component 1 explained greater variance
199 in Network_HLN (Figure 2f). Combining Module_1 and Module_2 (149 nodes)
200 increased predictability to 0.94, but adding Module_3 provided no further improvement
201 (Figure S14g-j). Thus, integrating parental transcriptomic and metabolomic differences
202 into network modules effectively explained the variance in seedling length heterosis.

203 **Heterosis-associated molecules primarily exhibit additive and partially dominant**
204 **effects**

205 We further identified heterosis-associated molecules specific to seedling length by
206 applying PLS analysis to Module_1 and Module_2 of Network_HLN. This yielded 68
207 high-contribution molecules (29 metabolites and 39 genes) that effectively separated
208 the three groups with high predictability (Q^2 = 0.90; Figure S14k-l). However, these
209 molecules did not exhibit continuous expression or abundance across the high-, low-,
210 and no-BPH groups along either Component 1 or Component 2, which together
211 explained 75.1% of group variance (Figure 2g and Figure S15). Additionally, F_1 hybrids
212 within the same heterotic group clustered distinctly and exhibited divergent molecular
213 levels (Figure S16), highlighting molecular heterogeneity among F_1 hybrids with
214 similar degrees of heterosis. Heatmap and PCA analyses further revealed that hybrid
215 molecular profiles were typically intermediate between the parent levels or aligned with
216 one parent (Figure 2h), demonstrating quantitative inheritance of heterosis-associated
217 molecules from parents to hybrids.

We next characterized the inheritance patterns of these molecules, observing all four genetic types: additive, partially dominant, dominant, and overdominant effects (Figure 2i). Additive and partially dominant effects collectively accounted for over 70% of molecular-level variation in most F_1 hybrids (Figure 2j and Figure S17a). This predominance persisted when genes and metabolite were analyzed separately and when restricted to the top 50 contributing molecules (Figure S17b-d). The additive effect showed outstanding predictability for seedling length heterosis ($Q^2 = 0.63$, $P < 5\times10^{-4}$; Figure 2k). Partially dominant (PD) effects were subdivided into PD_0.3 and PD_0.7 subtypes (Figure S18), with respective predictability of 0.37 ($P < 5\times10^{-4}$) and 0.29 ($P < 5\times10^{-4}$; Figure 2l). In contrast, predictabilities of dominant effects were non-significant for the female parent ($Q^2 = 0.14$, $P = 0.40$) and marginal for the male parent ($Q^2 = 0.07$, $P = 0.015$; Figure 2l). Importantly, the predictability of additive effect was statistically indistinguishable from that of hybrid profiles, and the additions of additive, partially dominant, or dominant effects to hybrid profiles did not improve predictability (Figure 2l). Furthermore, the combined proportion of additive and partially dominant effects positively correlated with seedling length heterosis (Figure S19). Together, these results underscored the predominant roles of additive and partially dominant effects in heterosis at both transcriptomic and metabolomic levels.

Additive and partially dominant effects contribute to heterosis in rice, maize, and *Arabidopsis*

Since seedling length and plant height represent the same trait measured at different developmental stages in rice, we quantified plant height for above F_1 hybrids across

240 four stages (28 days post-transplanting to maturation). Significant positive correlations
241 existed between heterosis for seedling length and heterosis for plant height (Figure
242 S20a-b), inferring shared genetic mechanisms underlying heterosis along development.
243 Additionally, correlation coefficients declined from early to mature stages, suggesting
244 increasing environmental influences on heterosis manifestation.

245 To evaluate the contribution of different inheritance patterns to heterosis for plant
246 height and other agronomic traits, we then constructed a hybrid population using a
247 complete diallel cross design. The 17 parental ILs represented genetic diversity within
248 a panel of 223 accessions encompassing *indica*, intermediate types, and *japonica*
249 (Figure 3a-b, Figure S20c-d, and Dataset S3). We measured 17 agronomic traits (yield,
250 yield components, and yield-related traits) for the parental ILs and their 245 F_1 hybrids.
251 Fourteen traits, including yield per plant, tiller number, and plant height, displayed
252 higher values in F_1 hybrids compared to parents (Figure S20e-f). We calculated BPH
253 and revealed negative average BPH for only four traits (Figure 3c), confirming superior
254 fitness in F_1 hybrids. Trait correlation patterns differed between parents and hybrids:
255 yield per plant tightly clustered with seed setting rate, grain weight, and heading date
256 in parental ILs, but with plant height in F_1 hybrids (Figure S20g-h). Notably, heterosis
257 for the 17 traits displayed correlation patterns similar to those of the parental ILs (Figure
258 3d). These results indicated that crossing genetically divergent ILs achieved positive
259 heterosis for most agronomic traits, with heterosis correlations mirroring parental
260 phenotypic correlations.

261 Using metabolite profiles from the 17 ILs, we selected five chemically annotated

262 metabolites from phenylpropanoid biosynthesis to investigate associations between
263 their inheritance patterns and heterosis for plant height (Figure S21a-f and Dataset S4).
264 Significant correlations emerged between plant height heterosis and four inheritance
265 patterns of these metabolites, except for the male dominant effect of L-phenylalanine
266 (negative ion mode) (Figure S21g). Among the five metabolites, 4-hydroxycinnamic
267 acid, which has been linked to seedling length heterosis in our above result and plant
268 height heterosis in previously reported findings (Dan et al., 2021), correlated
269 significantly with plant height heterosis across stages with all inheritance patterns.
270 Meanwhile, correlation coefficients of additive levels of 4-hydroxycinnamic acid and
271 heterosis for plant height decreased during later developmental stages, coinciding with
272 significant increases in heterosis at these stages (Figure S21h). Importantly, additive
273 and partially dominant effects of these metabolites correlated more strongly with plant
274 height heterosis than dominant effects (Figure S21i), a pattern consistent across an
275 additional 13 traits (Figure 3e). These results showed that phenylpropanoid
276 biosynthesis was associated with heterosis for plant height at the population level and
277 additive and partially dominant effects might be the predominant contributors for
278 heterosis of multiple agronomic traits.

279 We next analyzed DMs among the 17 ILs. Using 1,306 DMs with different
280 inheritance patterns, we predicted heterosis for each trait via PLS. Additive effect
281 yielded higher adjusted R^2 values in predictive models compared to dominant and
282 partially dominant effects (Figure 3f and Figure S22). The maximum R^2 values for
283 predicting the 17 traits were significantly greater for additive effect than dominant

284 effects and comparable to partially dominant effects (Figure 3g). When re-measuring
285 eight representative traits for the 17 ILs and 36 F_1 hybrids from the diallel crosses under
286 different growth conditions, we observed significant differences in almost all traits
287 influenced by environmental effects (Figure S23a). However, heterosis remained stable
288 for six of these traits (Figure S23b). Consistent with the above result, the highest R^2
289 values for predicting heterosis of these traits were greater for additive effect than
290 dominant effects and similar to partially dominant effects (Figure 3h and Figure S24a-
291 h). These findings confirmed additive and partially dominant effects as the main
292 inheritance patterns underlying heterosis of the 17 agronomic traits, independent of
293 environmental variation.

294 To assess the application potential of these findings in hybrid breeding programs,
295 we created a test-cross population, with a Honglian-type cytoplasmic male sterile line
296 (Yuetai A) as female and 211 recombinant inbred lines (RILs, derived from hybrids of
297 the above-mentioned 223 ILs) as male parents (Figure S25a-b). Trait values (yield per
298 plant, maturation stage plant height, and heading date) differed significantly between
299 test crosses and parents (Figure S25c), and trait correlation patterns also differed
300 (Figure S25d-f). For heterosis of these four traits, BPH was negative only for heading
301 date (Figure S25g). Metabolite profiles from 107 RILs (selected for varied yield
302 heterosis) and Yuetai A, which were profiled alongside the 17 ILs, were used for PLS
303 modeling (Figure S25h-i). Despite significant differences in heterosis for yield per plant
304 and heading date between the diallel and test-cross populations (Figure S25g), the
305 values of predictive models using additive DMs achieved $R^2 > 0.85$ with five latent

306 factors (Figure 3i). Thus, the DMs from representative ILs served as robust biomarkers
307 for predicting heterosis in test-cross hybrids.

308 To further explore the contributions of additive and partially dominant effects to
309 maize heterosis, we analyzed field fresh weight data from two hybrid populations
310 exhibiting significant different heterosis (Figure S26a-b) (de Abreu et al., 2017). PLS
311 analyses utilized root and leave metabolite profiles from 24 Dent lines, 25 Flint lines,
312 and 326 Dent × Flint F_1 hybrids (Figure S26c-j). Unexpectedly, additive and partially
313 dominant effects yielded R^2 values (maximum) comparable to models using hybrid
314 metabolite profiles directly and significantly higher than dominant effects (Figure
315 S26k). Similarly, for grain dry matter yield and grain dry matter content in 291 maize
316 F_1 hybrids (Schrag et al., 2018), additive and partially dominant effects showed R^2
317 values comparable to hybrid profiles and higher than dominant effects (Figure S27).
318 Additionally, we analyzed maize grain yield heterosis for a third hybrid population
319 (Figure S28a-b), with the 66 F_1 hybrids showing strong heterosis (Wang et al., 2023).
320 We used expression data of 84 DEGs, which were primarily core genes in pan-genome
321 analysis (Figure S28c), from 12 representative maize ILs. Consistent with metabolite
322 results, additive and partially dominant effects produced greater R^2 values than
323 dominant effects for predicting yield heterosis (Figure S28d-g). These findings further
324 supported the predominant roles of additive and partially dominant effects in maize
325 heterosis.

326 Similarly, in *Arabidopsis*, phenotypic data from Col-0, Per-1, and their F_1 hybrids
327 across developmental stages revealed robust heterosis for cotyledon biomass and first

328 true leaf size (Liu et al., 2021b) (Figure S29a and Figure S29k). Using our HAG
329 screening strategy in rice, we identified HAGs for each phenotype (Figure S29b-i and
330 Figure S29l-s). Unexpectedly, the hybrid profiles strongly associated with additive and
331 partially dominant effects throughout development (Figure S29j and Figure S29t).

332 **Associations of parental genomic variants with additive/partially dominant effects**
333 **and heterosis**

334 Because both transcriptomic and metabolomic differences in parental ILs were
335 connected to heterosis, we next performed genomic sequencing of the 17 rice ILs to
336 investigate associations between parental genomic variants and heterosis. Genetic
337 diversity analysis revealed two distinct groups of these ILs based on variants including
338 single-nucleotide polymorphisms (SNPs), insertions and deletions (InDels), copy
339 number variants (CNVs), and structural variants (SVs) (Figure 4a-f, Figure S30-S32,
340 and Table S3). Using the Nipponbare genome as a reference, genomic variants in each
341 pair of ILs were categorized as common or unique, whose number correlated
342 significantly with trait heterosis (Figure 4g). The unique variants showed stronger
343 correlations with heterosis than common variants (Figure 4h and Figure S33a-b).
344 Specifically, the unique variants significantly correlated with heterosis for seed setting
345 rate, tiller number (elongation stage), plant height (across four stages), and panicle
346 length. Meanwhile, unique SNPs and InDels exhibited stronger heterosis correlation
347 than CNVs and SVs (Figure 4i), likely reflecting limitations of short-read sequencing
348 in resolving large structural variants. Notably, the unique SNPs and InDels correlated
349 significantly with yield heterosis, while CNVs and SVs did not. For heterosis, increased

350 number of unique SNPs or InDels correlated with higher heterosis for grain number but
351 lower heterosis for seed setting rate (Figure 4j and Figure S33c-e), revealing barriers of
352 reproductive isolation in the exploitation of yield heterosis between *indica* and *japonica*
353 subspecies (Dan et al., 2014). Consequently, the inverse relationship between grain
354 number and seed setting rate heterosis produced a parabolic pattern in yield heterosis
355 (Figure 4k and Figure S33f-h). These results showed that parental genomic variants
356 were tightly linked to heterosis and highlighted the genomic complexity underlying
357 heterosis of polygenic traits like yield.

358 Compared to heterosis for yield, yield components, and yield-related traits,
359 heterosis for plant height, which was less affected by reproductive isolation of
360 subspecies, showed greater correlation coefficients with parental unique variants
361 (Figure 4g). To study the effect of variants in a specific pathway on trait heterosis, we
362 focused on sequences of six HAGs from phenylpropanoid biosynthesis (previously
363 linked to seedling length) in 12 of the 17 ILs. Sequence variants were detected in all
364 the six genes between YB and Balilla. With YB as the reference, SNPs dominated the
365 variants, with several InDels (Figure 5a and Figure S34-S38). Unexpectedly, presence
366 and absence variants occurred only in *Os09g0262000*, whose function is putatively
367 annotated as cinnamoyl CoA reductase. Sequence variants of each gene clustered the
368 12 ILs into two or three groups (Figure 5b and Figure S34-S38), which were close to
369 *indica* and *japonica* groups, implicating these genes in rice subspeciation. Notably,
370 sequence variants in *Os09g0262000*, particularly two presence variants in the fifth exon,
371 shifted Balilla into the *indica* group (Figure 5b). Using all six genes, the 12 ILs

372 contained two haplotypes (Figure 5c-d). Due to *Os09g0262000*, Balilla was grouped
373 with six *indica* ILs in haplotype-1. Unexpectedly, the average plant height heterosis of
374 F_1 hybrids with heterozygous haplotypes (haplotype-1_haplotype-2) were significantly
375 higher than their homozygous counterparts (haplotype-1 and haplotype-2; Figure 5e).
376 These hybrids also exhibited significantly higher heterosis for grain number and
377 primary/secondary branch number, lower heterosis for heading date, and no yield
378 heterosis (Figure 5f-g and Figure S39). In combination of results from transcriptomic
379 and metabolomic data, we concluded that phenylpropanoid biosynthesis contributed to
380 heterosis for seedling length/plant height in rice.

381 To investigate associations of parental genomic variants and different inheritance
382 patterns, we used the top 50 HAMs per trait based on the established predictive models.
383 Additive and partially dominant effects correlated more strongly with heterosis of the
384 17 traits than dominant effects (Figure S40). Importantly, additive and partially
385 dominant effects showed stronger associations with the number of parental unique
386 variants across all variant types (Figure 6a-d). Overall, parental genomic variants were
387 tightly associated with both heterosis and additive/partially dominant effects.

388 Finally, with consistent transcriptomic evidence for rice heterosis (Dan et al.,
389 2024b), we proposed a model integrating genomic, transcriptomic, and metabolomic
390 variation to illustrate how the main inheritance patterns (e.g., additive effect) contribute
391 to heterosis of component and complex traits in rice (Figure 6e). For simplicity, grain
392 yield was modeled as a function of grain number (SNP-associated) and grain weight
393 (SV-associated), assuming equal parental yield. Genomic variants drove differential

394 gene expression and metabolite levels in parents (Parent_1 and Parent_2). Due to
395 additive effect, the F_1 hybrids yielded intermediate transcriptomic and metabolomic
396 levels. Since trait values scaled linearly with molecular levels, grain number and grain
397 weight in F_1 hybrids also displayed mid-parent values. The additive effects of two yield
398 components, which differed reciprocally in parents, then generated positive yield
399 heterosis in F_1 hybrids.

400 **Discussion**

401 The genetic mechanisms underlying heterosis have been debated for over a century
402 without definitive consensus. The mysterious molecular bases of heterosis may stem
403 from the multitude of molecules associated with heterosis, including genes, metabolites,
404 and proteins (Birchler et al., 2006; Birchler et al., 2010; Huang et al., 2016; Schnable
405 and Springer, 2013). Accurate identification of these molecules remains challenging,
406 hinders mechanistic insights; consequently, few heterosis-associated genes with strong
407 function have been reported by far (Krieger et al., 2010; Sun et al., 2023).

408 To highlight genetic factors in rice heterosis, we minimized environmental effects
409 by employing uniform growth conditions for phenotyping seedling length. This allowed
410 precise dissection of genotype- and developmental stage-dependent effects on gene
411 expression and metabolite levels, facilitating the successful screening of heterosis-
412 associated molecules from parental differential molecules. These molecules form
413 tightly connected networks, and specific modules within these networks explain a
414 significant proportion of heterosis variance. Thus, modulating entire pathways (e.g.,

415 phenylpropanoid biosynthesis) or key network modules (e.g., Module_1 and Module_2
416 of Network_HLN), rather than targeting single molecule, through genetic engineering
417 technology may enable the precision establishment of heterosis. Furthermore, external
418 application of specific metabolites to F_1 hybrids at appropriate developmental stages
419 may enhance the degrees of heterosis without genomic alteration, given the significant
420 correlations between the levels of heterosis-associated metabolite and heterosis
421 magnitude.

422 Additive and partially dominant effects represent the predominant inheritance
423 patterns for heterosis-associated molecules. Notably, the additive effect, corresponding
424 to mid-parent values, exhibits predictability comparable to hybrid profiles in both rice
425 and maize, highlighting its application potential for predicting heterosis of F_1 hybrids
426 in breeding programs. Indeed, additive effect has predicted trait values and heterosis of
427 diverse agronomic traits for F_1 hybrids across hybrid populations, developmental stages,
428 and growth environments in rice (Dan et al., 2024a; Dan et al., 2021; Dan et al., 2019;
429 Dan et al., 2020; Dan et al., 2016). However, building robust predictive models for
430 heterosis of complex traits like yield remains challenging. Several key issues should be
431 considered to ensure accuracy of predictive models. Firstly, the accessions used for
432 modeling must capture sufficient genetic diversity to avoid reduced predictability from
433 distant genetic relatedness. Then, for low heritability traits (e.g., tiller number, yield),
434 omics data derived from relevant tissues (e.g., young tillers, young panicles) will ensure
435 accurate predictions (Zhou et al., 2019b). Apart from transcriptomic and metabolomic
436 profiles used in this study, adding other quantitative omics data, including proteomics,

437 ionomics, and metagenomics (Kakoulidou et al., 2024; Price et al., 2017; Xu et al., 2021;
438 Zhang et al., 2023), can improve predictability. Accounting for the important roles of
439 environmental effects on heterosis, predictive models need to integrate phenotypic data
440 from multiple environments (Liu et al., 2021a). Additionally, rapid advances in data
441 processing ability from machine learning and artificial intelligence will further enable
442 robust model development using single or combined strategies.

443 Unlike classical heterosis models emphasizing modes of genomic loci (Bruce,
444 1910; Crow, 1948; Davenport, 1908; East, 1936; Jones, 1917; Li et al., 2015; Li et al.,
445 1997; Powers, 1944; Xiao et al., 2021; Yu et al., 1997), our approach focuses on
446 downstream omics data to provide quantitative insight into heterosis. Beyond
447 demonstrating contributions of additive/partially dominant effects at
448 transcriptomic/metabolomic levels, we bridge associations among parental genomic
449 variants, inheritance patterns of heterosis-associated molecules, and heterosis of
450 component/complex traits. Using heterosis for plant height as an example, we detect
451 sequence variants (particularly SVs) in phenylpropanoid biosynthesis genes that cluster
452 parents into distinct haplotypes. The F_1 hybrids with heterozygous haplotypes exhibit
453 stronger heterosis than those with homozygous haplotypes, suggesting an
454 overdominance model underlying rice heterosis.

455 The genetic mechanisms of heterosis for yield or other complex traits are more
456 intricate than to heterosis for seedling length/plant height. Given that the F_1 hybrids
457 with heterozygous haplotypes enhanced heterosis not only for plant height but also
458 grain number (while reducing heterosis for heading date), we speculate that

459 phenylpropanoid biosynthesis genes may interact with known heterosis-regulating
460 genes like *Ghd7* (Wang et al., 2022), *Ghd8* (Li et al., 2016; Sun et al., 2023), and *Hd1*
461 (Zhou et al., 2021b) to influence yield heterosis. Moreover, large-size genomic variants,
462 especially presence variants, play pivotal roles in haplotype divergence and affect
463 heterosis of multiple traits. Thus, assembling parental genomes and accurately
464 identifying SVs will provide a more comprehensive genetic landscape of plant heterosis.
465 In summary, our study provides multi-omics evidence that additive and partially
466 dominant effects constitute the predominant inheritance patterns underlying heterosis,
467 irrespective of genotype, trait/phenotype, hybrid population, developmental stage,
468 growth environment, omics data type, or even species. Future efforts should leverage
469 accurately assembled pan-genomes and multi-omics integration to dissect the interplay
470 between large-effect genomic variants and the main inheritance patterns in heterosis of
471 complex traits, which will improve the breeding efficiencies of hybrid crops.

472 **Methods**

473 **Plant materials and field growth conditions**

474 According to the *indica–japonica* indices of 223 inbred lines (ILs) evaluated by 33 pairs
475 of InDel markers (Lu et al., 2009; Yao and Huang, 2013), 17 ILs, ranging from *indica*
476 to intermediate types to *japonica*, were selected as parents of a hybrid population using
477 a complete diallel cross design (Dan et al., 2016). Moreover, 211 recombinant inbred
478 lines (F_5 followings) derived from F_1 individuals of the 223 ILs were crossed with a
479 Honglian-type cytoplasmic male sterile line Yuetai A. Plants of the diallel and test cross

480 populations were planted with different density as previously described at the Hybrid
481 Rice Experimental Base of Wuhan University in Ezhou (Hubei Province, China) in
482 2012 and 2015 (Dan et al., 2019), respectively.

483 For the diallel cross population, ten experimental plants were planted with 16.5 cm
484 between plants and 26.4 cm between rows, and two plants of the sterile line, namely
485 Yuetai A, were planted on two sides of each row. To obtain F_1 hybrids, the young
486 panicles of the female parents that would flower in the next few days were immersed
487 in warm water at 44 °C (for *indica* or near-*indica* ILs) or 46 °C (for *japonica* or near-
488 *japonica* ILs) for 5 min to sterilize the pollens. For the hand-pollination procedure,
489 three to four panicles were combined in a paper bag, and one to two bags from each
490 female parent without pollination were used as controls. For female parents with a high
491 self-fertilization rate, such as R465 and Qianlijing, the pistils in each spikelet were
492 manually removed and pollinated by hand. Three biological replicates of each genotype
493 were randomly distributed in the field. The middle plants of each row, ranging from
494 three to ten plants, were selected for the collection of phenotypic data for 17 agronomic
495 traits. Plant height, measured from the base of the plant to the highest part of the plant
496 using meter rulers, was recorded from the seedling to mature stages. The four time
497 points included 28 days after transplanting (8 July, 2012), 42 days after transplanting
498 (22 July, 2012), the time of flowering, and the time of maturation, which were defined
499 as the seedling stage plant height, elongation stage plant height, flowering stage plant
500 height, and maturation stage plant height, respectively. The tiller number per plant was
501 also counted together with the plant height at four different stages, which were defined

502 as the seedling stage tiller number, elongation stage tiller number, flowering stage tiller
503 number, and maturation stage tiller number. The heading date was the number of days
504 between the heading time when a plant from a biological replicate started flowering
505 and the date of sowing (13 May, 2012). At maturity, all panicles from each plant were
506 harvested and dried in the sunshine for approximately one week and then dried in ovens
507 at 42 °C for two days before yield and yield-related traits were measured. Panicle length
508 was measured from the panicle neck to the top for all panicles from each plant. The
509 primary branch number and secondary branch number per panicle were counted
510 manually for all panicles. Panicles from each plant were counted and treated as the tiller
511 number per plant. After manually removing seeds (with or without grains) from
512 panicles for each individual plant, empty seeds and seeds with grains were
513 mechanically separated and counted using seed counters (PME-1, Shanke Equipment,
514 Shanghai). The grain number per panicle was calculated by dividing the total seed
515 number (sum of empty seeds and seeds with grains) of a plant by the tiller number per
516 plant. The seed setting rate was calculated as the ratio of the number of seeds with
517 grains to the total number of seeds. The weight of 1,000 grains from a plant was defined
518 as the thousand-grain weight, and the weight of the seeds with grains per plant was
519 defined as the grain yield per plant. Due to low germination rate of several hybrid seeds,
520 phenotypic data of 245 (expected 272) F_1 hybrids were used for analyses.

521 For the test cross population (211 F_1 hybrids), each biological replicate contained
522 ten plants (per row) spaced at 16.5 cm × 16.5 cm apart. Three biological replicates of
523 each genotype were placed in three consecutive rows and randomly planted in the field.

524 The phenotypic data of four agronomic traits, namely, heading date, thousand-grain
525 weight, maturation stage plant height, and yield per plant were collected. The heading
526 date was recorded for one biological replicate. The middle five plants per row were
527 selected for measurement of plant height at the maturation stage. Panicles from the five
528 plants were bagged and used to collect data on grain yield per plant and thousand-grain
529 weight. Moreover, eight agronomic traits, namely, heading date, thousand grain weight,
530 panicle length, seed setting rate, grain number per panicle, maturation stage plant height,
531 grain yield per plant, and tiller number per plant, of the 17 ILs and 36 F_1 hybrids from
532 the diallel cross population were recollected along with the test crosses. Parts of trait
533 values of the diallel and test cross populations have been reported in previous studies
534 (Dan et al., 2021; Dan et al., 2019; Dan et al., 2020; Dan et al., 2015; Dan et al., 2016;
535 Dan et al., 2014), and the complete datasets for all investigated traits are provided in
536 source data.

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

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

543 Two batches of rice seedlings were planted for untargeted metabolite profiling analysis.
544 The first batch in 2016 (Batch_2016) included 17 ILs, 107 RILs selected from the 211
545 RILs, and the cytoplasmic male sterile line Yuetai A. As previously described (Dan et

546 al., 2020), the germinated seeds were sown in soil with a spacing of 2 × 2 cm. The
547 seedlings were grown at 30 °C under a 16 h light/8 h dark photoperiod in a light
548 incubator. Two biological replicates were established for each genotype. On the 15th
549 day, the aboveground parts, excluding approximately 0.5 cm of tissue close to the soil,
550 of approximately ten plants were harvested and washed with ddH_2O three times. Two
551 biological replicates of the 17 ILs were harvested as two experimental samples for each
552 genotype. Two biological replicates of the 107 RILs and Yuetai A were pooled as one
553 experimental sample for each genotype. The seedlings were first placed in 5 mL tubes
554 (Eppendorf #0030119460) and snap frozen in liquid nitrogen.

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

568 (version 1.5.3). After excluding the roots, the seedling bases were washed three times
569 with ddH_2O and the water on the surface was removed with absorbent paper before
570 weighing. The five seedlings were snap frozen in liquid nitrogen, and photographs and
571 weights were taken and used for subsequent experiments. The remaining nine pairs of
572 hybrid combinations were harvested for seedling length measurements after 15 days of
573 growth. The sealing films on the glass tubes were removed on day 11 and 5 mL of
574 ddH_2O was added to each tube for normal growth. All the experimental samples were
575 stored at -80 °C.

576 **RNA sequencing**

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

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

601 The raw reads were first processed by removing adapters, poly N and low-quality
602 reads. The Q20, Q30, and GC contents were calculated for the clean reads. The clean
603 reads were then aligned to the Nipponbare reference genome
604 (Oryza_sativa_Ensembl_43) using Hisat2 (v2.0.5) (Kim et al., 2015) and the numbers
605 of gene reads were counted using featureCounts (v1.5.0-p3) (Liao et al., 2014).

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

607 Normalization and differential analysis of the transcriptomics data (42,341 genes) for
608 the hybrid pair of Balilla and YB at 5, 10, and 15 DAS were performed using
609 NetworkAnalyst 3.0 (Zhou et al., 2019a) with the following parameters: specify
610 organism, not specified; data type, bulk RNA-seq data (counts); ID type, not specified;
611 gene-level summarization, sum; variance filter, 30; low abundance, 30; filter

612 unannotated genes, yes; normalization, Log2-counts per million; statistical method,
613 DESeq2; Log2 fold change, 2; and adjusted *p*-value, 0.0001. As the numbers of
614 differentially expressed genes between reciprocal F_1 hybrids of Balilla and YB at 5, 10,
615 and 15 DAS were only 1, 13, and 29, respectively, the differences between reciprocal
616 F_1 hybrids were not included in the analysis. To analyze the differentially expressed
617 genes of the six ILs, the data matrix containing 42,341 genes was separately normalized
618 (sum, none, and autoscaling) with MetaboAnalyst 4.0 (Chong et al., 2018) and analyzed
619 with ANOVA (adjusted *p*-value = 0.05, post-hoc analysis = Fisher's LSD). Principal
620 component analysis (PCA) of the transcriptomic and metabolomic data, which were
621 normalized data from NetworkAnalyst 3.0 (Zhou et al., 2019a), was performed at
622 ClustVis (https://biit.cs.ut.ee/clustvis) (Metsalu and Vilo, 2015).

623 Two-way ANOVA of the 5,844 DEGs was performed using the "Time Series
624 Analysis" (time series and one factor) module of MetaboAnalyst 4.0 (Chong et al., 2018)
625 with the following parameters: ANOVA type, type I; interaction, yes; adjusted *p*-value
626 cut-off, 0.05; and correction for multiple testing, false discovery rate.

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

628 The ClueGO (version 2.5.8) (Bindea et al., 2009) and CluePedia (version 1.5.8) (Bindea
629 et al., 2013) plugins of Cytoscape (version 3.9.0) (Saito et al., 2012) were used for
630 pathway enrichment analysis of the differentially expressed genes. Functional analysis
631 was selected as the analysis mode. The up- and downregulated genes were represented
632 by diamonds (shape 1) and triangles (shape 2), respectively. The GO (ellipse), KEGG
633 (hexagon), and WikiPathways (V) were selected as ontologies/pathways. The GO term

634 fusion was selected and only pathways with pV ≤ 0.01 were shown. The GO tree
635 interval was between 3 (minimum level) and 5 (maximum level). The minimum number
636 and percentage of genes were set to 3 and 4%, respectively, for both clusters. The genes
637 accounting for more than 60% of a cluster were defined as cluster-specific genes. The
638 kappa score of network connectivity was set to 0.4, and enrichment/depletion (two-
639 sided hypergeometric test) with Benjamini–Hochberg pV correction was set as the
640 statistical option. GO term grouping was selected in the group options, and the leading
641 term was based on the highest significance with fixed colors. The kappa score was
642 selected for term grouping. The initial group size, percentage of genes for group
643 merging and percentage of terms for group merging were set to 1, 50% and 50%,
644 respectively.

645 **Untargeted metabolite profiling**

646 For the seedling samples in Batch_2016, liquid chromatography-mass spectrometry
647 (LC–MS) analysis procedures have been previously described (Chen et al., 2024; Dan
648 et al., 2020). In brief, all tissues in each EP tube were first ground into homogenized
649 powder by mortar and pestle with the addition of liquid nitrogen. Eighty milligrams of
650 powder were then transferred to 2 mL EP tubes and the tubes were vortexed for 1 min
651 after the addition of 1 mL of precooled extraction solution (methanol/acetonitrile/water,
652 2/2/1). A quality control sample was prepared by mixing 10 μL of each experimental
653 sample. The experimental and quality control samples were then subjected to two
654 rounds of ultrasonic treatments (30 min each) and kept at -20 °C for 60 min. The
655 samples were then centrifuged at 14,000 g for 15 min at 4 °C and the supernatants were

656 transferred to clean tubes. After drying the supernatants in a vacuum concentrator, 100
657 μL of acetonitrile (acetonitrile/water, 1/1) was added to the tubes. The tubes were
658 vortexed and centrifuged at 14,000 g for 15 minutes at 4 °C. The supernatants were
659 transferred to sample tubes for LC–MS analysis on the platform of Shanghai Applied
660 Protein Technology Co., Ltd.

661 The experimental and quality control samples were separated on an Agilent 1290
662 Infinity LC system (Agilent Technologies) equipped with a reverse-phase T3 column
663 (ACQUITY UPLC HSS T3, 1.8 μm, 2.1 mm × 100 mm, Waters). The injection volume
664 was 2 μL per sample. The flow rate was 0.3 mL/min and the column temperature was
665 25 °C. Regents A and B in positive ion mode were 0.1% formic acid (in water) and 0.1%
666 formic acid (in acetonitrile), respectively. In addition, 0.5 mM ammonium fluoride
667 (Sigma–Aldrich) and acetonitrile (Merck) were used in negative ion mode. The gradient
668 elution steps were as follows: 0-1 min, 1% B; 1-8 min, linearly increased to 100% B;
669 8-10 min, held at 100% B; 10-10.1 min, reduced to 1% B; and 10.1-12 min, held at 1%
670 B. Experimental samples were analyzed in a random order, and the quality control
671 sample was injected every nine experimental samples. To obtain MS profiles in both
672 positive and negative electron spray ion modes, the LC system was coupled to an
673 Agilent 6550 iFunnel QTOF mass spectrometer (Agilent Technologies). The MS data
674 were collected from 60 to 1,200 Da. The acquisition rate was four spectra per second,
675 and the cycle time was 250 microseconds. The temperatures of the drying gas and the
676 sheath gas were 250 °C and 400 °C, respectively. The flow rates of the drying gas and
677 the sheath gas were 16 L/min and 12 L/min, respectively. The nebulizer pressure was

678 set at 20 pounds per square inch (psi). The capillary, nozzle, and fragment voltages were
679 3,000, 0, and 175 V, respectively. Tandem mass spectrometry data were then acquired
680 on the quality control sample using a TripleTOF 6600 mass spectrometer (AB SCIEX).
681 The mass range was between 60 and 1,200 Da. To increase the data acquisition rate of
682 the quality control sample, the sample was injected twice and the mass range was
683 further divided into four sequential windows: 60-200, 190-400, 390-600, 590-1,200
684 (full range, 60-1,200) Da (for chemical standard library) and 200-600, 590-750, 740-
685 900, 890-1,200 (full range, 200-1,200) Da (for lipid MS/MS library). The mass
686 spectrometry accumulation time was 200 ms per spectrum, and the tandem mass spectra
687 were obtained in an information-dependent manner with high sensitivity. Ion source
688 gas 1, ion source gas 2, and the curtain gas were set at 40, 80, and 30 psi, respectively.
689 The source temperature was 650 °C. The ion spray voltage floating were +5,000 V and
690 -4,500 V, respectively. The declustering potential and collision energy were ± 60 V and
691 35 ± 15 eV, respectively. The number of candidate ions for monitoring per cycle was
692 ten and the isotope isolation width was 4 Da. The MS/MS accumulation time was set
693 to 50 ms.

694 The raw MS files (.d) of the experimental and quality control samples were
695 converted to mzXML format using ProteoWizard (version 3.0.6150) (Chambers et al.,
696 2012), and the transformed files were uploaded to XCMS (version 1.46.0) (Gowda et
697 al., 2014) for feature detection, retention time (RT) correction and feature alignment
698 with previously reported parameters (Jia et al., 2018). The 17 ILs formed one group,
699 and the 107 RILs plus Yuetai A formed another group. The features with more than 50%

700 nonzero values in at least one group were retained for further analysis. The error
701 thresholds for m/z and RT were set at 15 ppm and 20 s for the alignment of MS and
702 MS/MS data, respectively. The MS/MS data of the quality control sample were matched
703 against both the local chemical standard library (Shen et al., 2019) and the lipid MS/MS
704 library (Tu et al., 2017), and metabolites were annotated with a mass accuracy of less
705 than 25 ppm. For the seedling samples in Batch_2018, the metabolomics analyses were
706 the same as those in Batch_2016, except that the four sequential windows of the quality
707 control sample were changed to three sequential windows: 60-300, 290-600, and 590-
708 1,200 (full range, 60-1,200) Da. The Metlin library (Domingo-Almenara et al., 2018)
709 was also used for feature annotation.

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

711 The normalization of the metabolomics data for the hybrid combination of Balilla and
712 YB at 5, 10, and 15 DAS followed the same parameters as those used for the
713 transcriptomics data. ANOVA was used to analyze differential metabolites for the six
714 ILs with a data matrix of 15,083 analytes. For ease of interpretation, the word
715 "metabolite" was used in this manuscript to represent analytes with or without
716 annotations. The "statistical analysis (metadata table)" module of MetaboAnalyst 5.0
717 (Pang et al., 2022) was used to perform two-way ANOVA on the 2,331 differential
718 metabolites with genotype and stage as metadata. One of the three biological replicates
719 of each genotype at each stage was used as a subject. Multiple factors were chosen for
720 the study design. Further data filtering, sample normalization, data transformation, and
721 data scaling were not performed. Type I ANOVA was used for two-way independent

722 ANOVA, and two-way repeated ANOVA was used to calculate interactions. The false
723 discovery rate was used to correct for multiple testing, and the adjusted *P* value cut-off
724 was set at 0.05.

725 To normalize the metabolomics data of the 17 ILs, Yuetai A, and 107 RILs, the
726 intensities of the two biological replicates of the 17 ILs were first averaged and then
727 normalized using MetaboAnalyst 5.0 (Pang et al., 2022) (sum, none, and autoscaling
728 for sample normalization, data transformation, and data scaling). Metabolites with RSD
729 greater than 25% in the QC samples were removed and the interquantile range was
730 selected to filter the features (40%). For the differential metabolite analysis of the 17
731 ILs, the feature intensities of two biological replicates for each IL were used and
732 ANOVA was performed after filtering 40% of the features and data normalization (sum,
733 none, and autoscaling) using MetaboAnalyst 5.0 (Pang et al., 2022).

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

735 To determine the inheritance patterns for heterosis-associated molecules,
736 transcriptomic and metabolic levels were compared for F_1 hybrids, the corresponding
737 parents, and the average levels of the corresponding parents via ANOVA using
738 MetaboAnalyst 4.0 (Chong et al., 2018). Three biological replicates of reciprocal F_1
739 hybrids from two ILs were combined into a hybrid genotype. An additive effect was
740 assigned when the molecular levels of the hybrid genotype were not significantly
741 different from the average levels of the corresponding parents but were significantly
742 different from those of both parents. A partially dominant effect was assigned when the
743 hybrid levels were significantly different from both the average parental levels and the

molecular levels of the parent on the same side. A dominant effect was when the hybrid levels were not significantly different from those of the parent but were significantly different from the average levels of the two parents. An overdominant effect was the pattern when the hybrid levels were significantly higher or lower than those of the parent on the same side. The patterns of hybrid levels other than the above four types were defined as unclassified and marked as NA. To quantitatively describe the partially dominant effect (PD), the degrees for the molecules with partially dominant effect of the six hybrid combinations was calculated with the equation $PD = (F_1-P_2)/(P_1-P_2)$, where F_1, P_1, and P_2 represent the molecular levels of the F_1 hybrids, parent_1, and parent_2, respectively. Several molecules that had a partially dominant effect in multiple F_1 hybrids were used multiple times. The degree of the partially dominant effect was divided into two parts, with 0.5 as the threshold. The averages of the two groups were calculated and defined as two types of partially dominant effects (PD_0.3 and PD_0.7). Moreover, 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 represent partially dominant effects based on parental molecular levels.

Pathway enrichment analysis of differential metabolites

The metabolomics data of the hybrid combination of Balilla and YB at 15 DAS were first used for pathway enrichment analysis. The table with normalized feature intensities and retention times (in seconds) was uploaded to the “functional analysis” module of MetaboAnalyst 5.0 (Pang et al., 2022). Both positive and negative modes were selected for analysis. In positive mode, the mass tolerance was set to 5 ppm. The

766 interquantile range was selected for data filtering because the number of peaks was
767 greater than 5,000. No data were used for sample normalization, data transformation,
768 or data scaling. Both *mummichog* (Li et al., 2013) (*P* value cut-off: default top 10%
769 peaks, version 2.0) and GSEA were used for enrichment analysis. As no pathway was
770 significantly enriched for the differential metabolites between Balilla×Yuetai B and its
771 parent Balilla with the default *P* value cut-off (0.01) in positive mode, the *P* value cut-
772 off was increased to 0.05. The pathway library used was *Oryza sativa japonica*
773 (Japanese rice) [KEGG]. Metabolite currency and adducts were set with the default sets,
774 and pathways or metabolite sets containing at least 3 entries were set. The parameters
775 for the negative mode analysis were the same as those for the positive mode analysis,
776 except for that no data filtering was performed and the *P* value cut-off was set to 0.05.
777 Putatively annotated metabolites and enriched pathways were obtained in each analysis.
778 For the pathway analysis of the 645 DMs using the "Pathway Analysis" module of
779 MetaboAnalyst 5.0 (Pang et al., 2022), feature annotation information was obtained
780 from the results of differential metabolites between Balilla and YB at 15 DAS. The
781 parameters were set as follows: scatter plot, visualization method; hypergeometric test,
782 enrichment method; relative-betweenness centrality, topology analysis; all compounds
783 in the selected pathway library, reference metabolome; and *Oryza sativa japonica*
784 (Japanese rice) (KEGG), pathway library. Three pathways were significantly enriched
785 based on the annotation of eight peaks in positive mode. Moreover, 20 metabolites (23
786 peaks from both ion modes), which were annotated with local chemical standards, lipids,
787 and Metlin libraries, from the three significantly enriched pathways were also used for

788 analysis. Correlations between the metabolite levels of 26 peaks and seedling length
789 were determined. One peak (M147T118_POS) was annotated as 4-hydroxycinnamic
790 acid in both the local standard library and the result of the pathway enrichment analysis,
791 and four peaks had no concentration in the normalized metabolomics data.

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

793 The normalized transcriptomics and metabolomics data were used to construct
794 networks via OmicsAnalyst (Zhou et al., 2021a). No organism or ID type was specified
795 in the annotation procedure. Features with 50% missing values were excluded and the
796 remaining missing values were replaced by LoDs. The filter threshold for variance and
797 abundance was set to 0. In the comparison step, the genotype was used for grouping,
798 and no data transformation was performed. ANOVA was used for the differential
799 analysis. The fold change cut-off was set at 2.0, and the *P* value cut-off was set at 0.01.
800 To make the data suitable for inter-omics analysis, both the transcriptomics and
801 metabolomics data underwent normalization by sum, variance stabilizing normalization,
802 quantile normalization and range scaling. The statistically significant features were then
803 used to construct an interomics correlation network using the Spearman method. The
804 correlation threshold was set at 0.8, and both positive and negative correlations were
805 included. For Network_51015, 626 genes and 645 metabolites were used, and a total of
806 213 genes and 198 metabolites were statistically significant. The network contained
807 three subnetworks, and the largest subnetwork with 2,138 edges (subnetwork_1, 61.8%
808 and 38.2% positive and negative correlations, respectively) had 211 genes and 196
809 metabolites as nodes. The other two subnetworks had two interactions. For

810 Network_HLN, 967 genes and 990 metabolites were used for analysis, and 103 genes
811 and 158 metabolites were retained in the network. Network_HLN contained seven
812 subnetworks, and the largest subnetwork with 868 edges (subnetwork_1) contained 97
813 genes and 149 metabolites. The remaining six subnetworks had nine edges. The
814 walktrap algorithm was used for module detection. Subnetwork_1 of Network_51015
815 contained seven modules. The module size and *P* value of each module were 150 (P =
816 2.34×10^{-21}, Module_1), 105 (P = 1.01×10^{-14}, Module_2), 80 (P = 2.29×10^{-12},
817 Module_3), 21 (P = 0.00303, Module_4), 13 (P = 0.452, Module_5), 7 (P = 1,
818 Module_6), and 6 (P = 0.318, Module_7), respectively. Subnetwork_1 of
819 Network_HLN contained eight modules. The module size and *P* value of each module
820 were 112 (P = 6.6×10^{-35}, Module_1), 37 (P = 2.61×10^{-7}, Module_2), 35 (P = $1.39\times10^{-}$
821 ^{6}, Module_3), 16 (P = 3.8×10^{-5}, Module_4), 11 (P = 0.111, Module_5), 7 (P = 0.0592,
822 Module_6), 6 (P = 0.0176, Module_7), and 6 (P = 0.0309, Module_8), respectively.
823 The subnetworks were imported into Cytoscape (Saito et al., 2012) for display using
824 the compound spring embedder layout. The node size was determined by the node
825 degree.

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

827 To integrate the transcriptomics and metabolomics results, the MSU IDs of the 626
828 DEGs were first converted to gene symbols using the NCBI "Batch Entrez" tool
829 (https://www.ncbi.nlm.nih.gov/sites/batchentrez). Genes with symbols (340 genes) in
830 the *Oryza sativa japonica* group were uploaded to the "Joint pathway" analysis module
831 of MetaboAnalyst 5.0 (Pang et al., 2022). Moreover, the annotation information of the

832 645 peaks was extracted from the results of the pathway enrichment analysis between
833 Balilla and YB at 15 DAS. The KEGG IDs for 29 compounds in positive mode and 44
834 compounds in negative mode were uploaded as metabolomics data. "All pathways
835 (integrated)" was selected as the pathway database. The parameters for enrichment
836 analysis, topology measure, and integration method were set as hypergeometric test,
837 degree centrality, and combine queries, respectively.

838 **Analyses of transcriptomic/metabolic/phenotypic data for inbred lines and F_1**
839 **hybrids in maize and *Arabidopsis***

840 The field fresh weight, grain dry matter yield, and grain dry matter content of maize
841 were obtained from published data in two previous studies (de Abreu et al., 2017;
842 Schrag et al., 2018). Metabolite profiles of 3.5-day-old roots and leaves at the third leaf
843 stage were also collected for 210 Dent and Flint lines and 363 F_1 hybrids in the two
844 studies (de Abreu et al., 2017; Schrag et al., 2018). The root and leaf metabolic datasets
845 contained levels of 165 and 81 metabolites, respectively. Since metabolic data were not
846 available for the parents of some of the F_1 hybrids, the field fresh weight of 326 F_1
847 hybrids, which were crossed between 24 Dent lines and 25 Flint lines, were used for
848 analysis. No further data normalization was performed on the metabolic data. As
849 reported in the study (de Abreu et al., 2017), the field fresh weight of the 326 F_1 hybrids
850 was recorded in two distinct field tests in 2010 (146 F_1 hybrids) and 2012 (180 F_1
851 hybrids). In the PLS analysis, the number was set to 100 to ensure a sufficient number
852 of latent factors, and the number was defined for each regression analysis when the R^2
853 value (adjusted) peaked. The R^2 values (adjusted) corresponding to different numbers

854 of latent factors and the greatest R^2 values for hybrid profiles and different inheritance
855 patterns were used for comparison analysis via ANOVA. Among the 326 F_1 hybrids,
856 291 F_1 hybrids had recorded data on grain dry matter yield and grain dry matter content
857 in approximately seven agro-ecologically diverse environments (Schrag et al., 2018).
858 The root metabolic data were used for PLS analysis, and the R^2 values (adjusted) were
859 used for comparison analysis.

860 The grain yield per plant (in three environments) of 12 maize inbred lines and 66
861 F_1 hybrids from the inbred lines with a half-diallel design reported in a previous study
862 (Wang et al., 2023) was used for analysis. The expression levels of 27,608 genes, which
863 were pan and core genes in Zheng58, from the 12 inbred lines were also obtained from
864 the study (Wang et al., 2023). As described in the study (Wang et al., 2023), each
865 genotype of the 12 maize ILs had two sets (not biological replicate) of RNA-seq data,
866 with one set containing equally mixed RNAs from eight tissues and the other from five
867 tissues at different developmental stages. The expression data were processed with
868 MetaboAnalyst 5.0 (Pang et al., 2022), with 40% set as the percentage for filtering. No
869 data transformation, normalization, or scaling was performed. In total, 84 DEGs were
870 identified in the 12 inbred lines by ANOVA, with Fisher's LSD used in post hoc
871 analysis (*P* value cut-off = 0.05). The average values of the two sets of RNA sequencing
872 data were used to represent the levels of each inbred line. The 84 genes were classified
873 according to the previously reported results (Wang et al., 2023). The number of latent
874 factors was set to 50 in the PLS analysis of better-parent heterosis for grain yield per
875 plant, mid-parent heterosis for grain yield per plant, and grain yield per plant.

The phenotypes of cotyledon and the first true leaf and transcriptomics data in *Arabidopsis* were obtained from a previously published study (Liu et al., 2021b). The transcriptomics data were normalized with above mentioned parameters in NetworkAnalyst 3.0. Differential expression genes between Col-0 and Per-1 at different stages were analyzed with the limma method, with Log2 fold change at 2 and adjusted *p*-value at 0.0001. The pooled differential expression genes across stages were treated as differential expression genes for each tissue. Two-way ANOVA on the normalized levels of DEGs with genotype and stage as metadata was implemented with MetaboAnalyst 5.0 (Pang et al., 2022). The overlapped DEGs of genotype, stage, and their interaction were used for further analysis. Pearson correlation between expression levels of the DEGs and phenotypes of cotyledon and the first true leaf was performed to screen the genes and the significance level was set at 0.05.

Genome sequencing and analysis of genomic data

Seedlings from the 17 rice ILs were ground into fine powder using liquid nitrogen. The genomic DNA was extracted, and the quality and concentration were assessed by 1% agarose gel electrophoresis and Qubit 3.0 fluorometer. The DNA libraries (with index codes) were prepared and sequenced on the Illumina sequencing platform of Novogene Co., Ltd (Beijing, China). In brief, genomic DNA was randomly fragmented into ~350 bp size via sonication (Covaris). The strands were then added with adapters and used for PCR amplification. Products were purified with AMPure XP beads (Beckman Coulter) and the concentrations were measured with Agilent 2100 Bioanalyzer. Libraries were sequenced on the Illumina PE150 platform, and the raw data were

898 recorded in FASTQ format (> 15 G per sample). After the clean reads were obtained,
899 the sequences were aligned to the reference genome using the Burrows–Wheeler
900 Aligner (mem -t 4 -k 32 -M, version 0.7.8-r455) (Li and Durbin, 2009). Dislodged
901 duplications in the BAM files were removed via SAMtools (rmdup, version 1.3.1) (Li
902 et al., 2009). SNPs and InDels in each sample were called with the parameters "-C 50
903 - mpileup -m 2 -F 0.002 -d 1000" and further filtered to reduce errors (depth of variant >
904 4 and mapping quality > 20). CNVs were detected by CNVnator (-call 100, version 0.3)
905 (Abyzov et al., 2011), and SVs were detected by BreakDancer (depth of variant > 2,
906 version 1.4.4) (Chen et al., 2009). Variant annotation was implemented with
907 ANNOVAR (Wang et al., 2010). Genomic variants on chromosomes were visualized
908 by Circos (0.69-9) (Krzywinski et al., 2009). To analyze the population structures of
909 the 17 ILs, GATK (version 4.2.0.0) was first used to filter SNPs (QD < 2, FS > 60, MQ
910 < 40, MQRankSum < -12.5, ReadPosRankSum < -8) and InDels (QD < 2, FS > 200,
911 ReadPosRankSum < -20). SNPs and InDels were further filtered with VCFtools
912 (maximum missing value < 10%, minus allele: 0.05, minimum read depth > 4, multiple
913 alleles removed, version 0.1.17) and analyzed with Plink (version 1.90) and Admixture
914 (version 1.3.0). When comparing the SNPs and InDels between each pair of samples of
915 the 17 ILs, the variants were defined as common variants if they shared the same values
916 for five parameters, including CHROM (chromosome), POS (position of the reference),
917 REF (reference base), ALT (alternative base) and GT (genotype). Differential variants
918 were those that were unique to each IL. For CNVs and SVs, variants were defined as
919 common variants if they shared the same values of CHROM, POS1, POS2, TYPE and

920 SIZE. The numbers of common and unique variants were counted, and the sums of
921 unique genomic variants between each pair of ILs were used to represent the genetic
922 distance. Phylogenetic trees (neighbour-joining method) based on one or multiple types
923 of genomic variants in the 17 ILs were generated with MEGA (version 10.2.6) (Stecher
924 et al., 2020).

925 **Phylogenetic analysis of phenylpropanoid biosynthesis genes**

926 Using genome assemblies from 12 rice inbred lines and available gene family
927 information (Dan et al., 2024b), we extracted sequences for eight of the ten heterosis-
928 associated genes of seedling length, except for *Os04g0688500* and *Os07g0677100*.
929 Among these eight genes, *Os01g0293900* and *Os02g0240300* were not detected in
930 some inbred lines and thus excluded. Consequently, sequences spanning the start-to-
931 stop codons of the six remaining genes were used for analyses. With sequences of YB
932 as the reference, sequence variants were identified for each accession. Exon-intron
933 structure of each gene was visualized using the Exon-Intron Graphic Maker
934 (http://wormweb.org/exonintron) with variants annotated. Phylogenetic trees for
935 individual and concatenated genes were built in Jalview (version 2.11.4.1) (Waterhouse
936 et al., 2009), using E-INS in mafft method. Haplotypes of the 12 inbred lines were
937 classified based on tree topology.

938 **Partial least squares regression analysis**

939 Partial least squares regression (PLS), which has inherent ability to handle noisy,
940 collinear, and incomplete variables in both predictors and responses (Eriksson et al.,
941 1995; Wold et al., 2001), was applied to build multivariate models for screening

942 heterosis-associated molecules and predicting heterosis of agronomic traits or hybrid
943 phenotypes. To extract enough latent factors, the maximum number was initially set to
944 100 in SPSS 20. The optimum model for each trait was determined by selecting the
945 highest adjusted R^2 value, which also served as the measure of predictability. The mean
946 variable importance in projection across all latent factors were used to estimate
947 contributions of predictors, applying a threshold of 1 for screening heterosis-associated
948 molecules of seedling length in rice. With class order taken into account, partial least
949 squares-discriminant analysis (PLS-DA) was performed on F_1 hybrids with different
950 degrees of seedling length heterosis using MetaboAnalyst 5.0 (Pang et al., 2022). The
951 analysis searched a maximum of five components and employed tenfold cross-
952 validation. To assess model overfitting, the separation distance (B/W) method was used
953 for permutation analysis (2,000 times). The predictability of PLS-DA model was
954 evaluated using the Q^2.

955 **Statistical analyses**

956 Analysis of variance (least significant difference used in post hoc test), independent
957 samples *t* test, paired samples *t* test, Spearman rank correlation, Pearson correlation
958 (two-tailed, P value < 0.05) were performed using SPSS 20.0. Venn diagrams were
959 drawn with the tool from http://bioinformatics.psb.ugent.be/webtools/Venn/. The
960 phylogenetic tree of 223 inbred lines was based on genotypes and the tree was
961 constructed using MEGA (Stecher et al., 2020) via the neighbour-joining method. The
962 *indica* genotype, heterozygous genotype, and *japonica* genotype based on InDel
963 markers were represented by 3, 2, and 1, respectively. A similarity matrix of the 223

964 inbred lines was then derived from MORPHEUS
965 (https://software.broadinstitute.org/morpheus/) using Euclidean distance. For each box
966 generated with BoxPlotR (Spitzer et al., 2014) (http://shiny.chemgrid.org/boxplotr/),
967 the central line represents the median, the upper and lower edges indicate the 75th and
968 25th percentiles, respectively, and the whiskers are 1.5× the interquartile range away
969 from the 75th and 25th percentiles. The statistical source data have been provided in
970 Dataset S5.

971 **Abbreviations**

972 ANOVA: analysis of variance; BPH: better-parent heterosis; CNV: copy number
973 variation; DAS: days after sowing; DEG: differentially expressed gene; DM:
974 differential metabolite; GO: gene ontology; HAG: heterosis-associated gene; HAM:
975 heterosis-associated metabolite; IL: inbred line; InDel: insertions and deletion; MPV:
976 mid-parent value; PC: principal component; PCA: principal component analysis; PLS:
977 partial least squares; PLS-DA: partial least squares-discriminant analysis; RIL:
978 recombinant inbred line; SNP: single-nucleotide polymorphism; SV: structural
979 variation; YA: Yuetai A; YB: Yuetai B; YPP: yield per plant; SBN: secondary branch
980 number; GNP: grain number per panicle; TPP: tiller number per plant; SSR: seed
981 setting rate; HD: heading date; TGW: thousand grain weight; SSTN: seedling stage
982 tiller number; ESTN: elongation stage tiller number; FSTN: flowering stage tiller
983 number; MSTN: maturation stage tiller number; SSPH: seedling stage plant height;
984 ESPH: elongation stage plant height; FSPH: flowering stage plant height; MSPH:
985 maturation stage plant height; PL: panicle length; and PBN: primary branch number.

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..

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.

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.

Conflict of interest

The authors declare no competing interests.

Data availability

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 (Dan et al., 2020) and MTBL1344. Source data are available with this paper.

Supplemental information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1-S40.

Table S1-S3.

Dataset S1-S5.

References

Abyzov, A., Urban, A.E., Snyder, M. and Gerstein, M. (2011) CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* **21**, 974-984.

Bindea, G., Galon, J. and Mlecnik, B. (2013) CluePedia Cytoscape plugin: pathway insights

1030 using integrated experimental and in silico data. *Bioinformatics* **29**, 661-663.
1031 Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H.*,*
1032 *et al.* (2009) ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology
1033 and pathway annotation networks. *Bioinformatics* **25**, 1091-1093.
1034 Birchler, J.A., Yao, H. and Chudalayandi, S. (2006) Unraveling the genetic basis of hybrid
1035 vigor. *Proc. Natl. Acad. Sci. USA* **103**, 12957-12958.
1036 Birchler, J.A., Yao, H., Chudalayandi, S., Vaiman, D. and Veitia, R.A. (2010) Heterosis. *Plant*
1037 *Cell* **22**, 2105-2112.
1038 Bruce, A.B. (1910) The Mendelian theory of heredity and the augmentation of vigor. *Science*
1039 **32**, 627-628.
1040 Chambers, M.C., MacLean, B., Burke, R., Amode, D., Ruderman, D.L., Neumann, S., Gatto,
1041 L.*, et al.* (2012) A cross-platform toolkit for mass spectrometry and proteomics. *Nat.*
1042 *Biotechnol.* **30**, 918-920
1043 Chen, K., Wallis, J.W., McLellan, M.D., Larson, D.E., Kalicki, J.M., Pohl, C.S., McGrath, S.D.*,*
1044 *et al.* (2009) BreakDancer: an algorithm for high-resolution mapping of genomic structural
1045 variation. *Nat. Methods* **6**, 677-681.
1046 Chen, Y., Dan, Z. and Li, S. (2024) *GROWTH REGULATING FACTOR 7*-mediated arbutin
1047 metabolism enhances rice salt tolerance. *Plant Cell* **36**, 2834-2850.
1048 Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., Wishart, D.S.*, et al.* (2018)
1049 MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis.
1050 *Nucleic Acids Res.* **46**, W486-W494.
1051 Crow, J.F. (1948) Alternative hypotheses of hybrid vigor. *Genetics* **33**, 477-487.
1052 Dan, Z., Chen, Y. and Huang, W. (2024a) The metabolic pathways underlying dynamic
1053 heterosis for plant height are robust biomarkers in rice. *bioRxiv*,
1054 10.1101/2024.1111.1109.622823.
1055 Dan, Z., Chen, Y. and Huang, W. (2024b) Structural variations contribute to genetic diversity
1056 and heterosis in rice. *bioRxiv*, 10.1101/2024.1111.1103.621459.
1057 Dan, Z., Chen, Y., Li, H., Zeng, Y., Xu, W., Zhao, W., He, R.*, et al.* (2021) The metabolomic
1058 landscape of rice heterosis highlights pathway biomarkers for predicting complex
1059 phenotypes. *Plant Physiol.* **187**, 1011-1025.
1060 Dan, Z., Chen, Y., Xu, Y., Huang, J.R., Huang, J.S., Hu, J., Yao, G.*, et al.* (2019) A
1061 metabolome-based core hybridisation strategy for the prediction of rice grain weight
1062 across environments. *Plant Biotechnol. J.* **17**, 906-913.

1063 Dan, Z., Chen, Y., Zhao, W., Wang, Q. and Huang, W. (2020) Metabolome-based prediction
1064 of yield heterosis contributes to the breeding of elite rice. *Life Sci. Alliance* **3**,
1065 e201900551.
1066 Dan, Z., Hu, J., Zhou, W., Yao, G., Zhu, R., Huang, W. and Zhu, Y. (2015) Hierarchical
1067 additive effects on heterosis in rice (*Oryza sativa* L.). *Front. Plant Sci.* **6**, 738.
1068 Dan, Z., Hu, J., Zhou, W., Yao, G., Zhu, R., Zhu, Y. and Huang, W. (2016) Metabolic
1069 prediction of important agronomic traits in hybrid rice (*Oryza sativa* L.). *Sci. Rep.* **6**,
1070 21732.
1071 Dan, Z., Liu, P., Huang, W., Zhou, W., Yao, G., Hu, J., Zhu, R.*, et al.* (2014) Balance between
1072 a higher degree of heterosis and increased reproductive isolation: a strategic design for
1073 breeding inter-subspecific hybrid rice. *PLoS ONE* **9**, e93122.
1074 Davenport, C.B. (1908) Degeneration, albinism and inbreeding. *Science* **28**, 454-455.
1075 de Abreu, E.L.F., Westhues, M., Cuadros-Inostroza, Á., Willmitzer, L., Melchinger, A.E. and
1076 Nikoloski, Z. (2017) Metabolic robustness in young roots underpins a predictive model of
1077 maize hybrid performance in the field. *Plant J.* **90**, 319-329.
1078 Domingo-Almenara, X., Montenegro-Burke, J.R., Ivanisevic, J., Thomas, A., Sidibe, J., Teav,
1079 T., Guijas, C.*, et al.* (2018) XCMS-MRM and METLIN-MRM: a cloud library and public
1080 resource for targeted analysis of small molecules. *Nat. Methods* **15**, 681-684.
1081 East, E.M. (1936) Heterosis. *Genetics* **21**, 375-397.
1082 Eriksson, L., Hermens, J., L. M., Johansson, E., Verhaar, H., J. M. and Wold, S. (1995)
1083 Multivariate analysis of aquatic toxicity data with PLS. *Aquat. Sci.* **57**, 217-241.
1084 Gowda, H., Ivanisevic, J., Johnson, C.H., Kurczy, M.E., Benton, H.P., Rinehart, D., Nguyen,
1085 T.*, et al.* (2014) Interactive XCMS Online: simplifying advanced metabolomic data
1086 processing and subsequent statistical analyses. *Anal. Chem.* **86**, 6931-6939.
1087 Hochholdinger, F. and Baldauf, J.A. (2018) Heterosis in plants. *Curr. Biol.* **28**, R1089-R1092.
1088 Huang, X., Yang, S., Gong, J., Zhao, Q., Feng, Q., Zhan, Q., Zhao, Y.*, et al.* (2016) Genomic
1089 architecture of heterosis for yield traits in rice. *Nature* **537**, 629-633.
1090 Jia, H., Shen, X., Guan, Y., Xu, M., Tu, J., Mo, M., Xie, L.*, et al.* (2018) Predicting the
1091 pathological response to neoadjuvant chemoradiation using untargeted metabolomics in
1092 locally advanced rectal cancer. *Radiother Oncol* **128**, 548-556.
1093 Jones, D.F. (1917) Dominance of linked factors as a means of accounting for heterosis. *Proc.*
1094 *Natl. Acad. Sci. USA* **3**, 310-312.
1095 Kakoulidou, I., Piecyk, R.S., Meyer, R.C., Kuhlmann, M., Gutjahr, C., Altmann, T. and

1096 Johannes, F. (2024) Mapping parental DMRs predictive of local and distal methylome
1097 remodeling in epigenetic F1 hybrids. *Life Sci. Alliance* **7**, e202402599.
1098 Kim, D., Langmead, B. and Salzberg, S.L. (2015) HISAT: a fast spliced aligner with low
1099 memory requirements. *Nat. Methods* **12**, 357-360.
1100 Krieger, U., Lippman, Z.B. and Zamir, D. (2010) The flowering gene *SINGLE FLOWER*
1101 *TRUSS* drives heterosis for yield in tomato. *Nat. Genet.* **42**, 459-463.
1102 Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J.*, et*
1103 *al.* (2009) Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**,
1104 1639-1645.
1105 Li, D., Huang, Z., Song, S., Xin, Y., Mao, D., Lv, Q., Zhou, M.*, et al.* (2016) Integrated analysis
1106 of phenome, genome, and transcriptome of hybrid rice uncovered multiple heterosis-
1107 related loci for yield increase. *Proc Natl Acad Sci U S A* **113**, E6026-E6035.
1108 Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler
1109 transform. *Bioinformatics* **25**, 1754-1760.
1110 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G.*, et al.* (2009)
1111 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079.
1112 Li, S., Park, Y., Duraisingham, S., Strobel, F.H., Khan, N., Soltow, Q.A., Jones, D.P.*, et al.*
1113 (2013) Predicting network activity from high throughput metabolomics. *PLoS Comput.*
1114 *Biol.* **9**, e1003123.
1115 Li, X., Li, X., Fridman, E., Tesso, T.T. and Yu, J. (2015) Dissecting repulsion linkage in the
1116 dwarfing gene *Dw3* region for sorghum plant height provides insights into heterosis. *Proc.*
1117 *Natl. Acad. Sci. USA* **112**, 11823-11828.
1118 Li, Z., Pinson, S., Park, W., Paterson, A. and Stansel, J. (1997) Epistasis for three grain yield
1119 components in rice (*Oryza sativa* L.). *Genetics* **145**, 453-465.
1120 Liao, Y., Smyth, G.K. and Shi, W. (2014) featureCounts: an efficient general purpose program
1121 for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930.
1122 Lisec, J., Romisch-Margl, L., Nikoloski, Z., Piepho, H.P., Giavalisco, P., Selbig, J., Gierl, A.*, et*
1123 *al.* (2011) Corn hybrids display lower metabolite variability and complex metabolite
1124 inheritance patterns. *Plant J.* **68**, 326-336.
1125 Liu, N., Du, Y., Warburton, M.L., Xiao, Y. and Yan, J. (2021a) Phenotypic plasticity contributes
1126 to maize adaptation and heterosis. *Mol Biol Evol* **38**, 1262-1275.
1127 Liu, W., He, G. and Deng, X.W. (2021b) Biological pathway expression complementation
1128 contributes to biomass heterosis in Arabidopsis. *Proc. Natl. Acad. Sci. USA* **118**,

1129 e2023278118.
1130 Lu, B., Cai, X. and Jin, X. (2009) Efficient *indica* and *japonica* rice identification based on the
1131 InDel molecular method:its implication in rice breeding and evolutionary research.
1132 *Progress in Nature Science* **19**, 1241-1252.
1133 Metsalu, T. and Vilo, J. (2015) ClustVis: a web tool for visualizing clustering of multivariate
1134 data using Principal Component Analysis and heatmap. *Nucleic Acids Res.* **43**, W566-
1135 570.
1136 Pang, Z., Zhou, G., Ewald, J., Chang, L., Hacariz, O., Basu, N. and Xia, J. (2022) Using
1137 MetaboAnalyst 5.0 for LC-HRMS spectra processing, multi-omics integration and
1138 covariate adjustment of global metabolomics data. *Nat. Protoc.*
1139 Powers, L. (1944) An expansion of Jones's theory for the explanation of heterosis. *Am. Nat.*
1140 **78**, 275-280.
1141 Price, N.D., Magis, A.T., Earls, J.C., Glusman, G., Levy, R., Lausted, C., McDonald, D.T.*, et*
1142 *al.* (2017) A wellness study of 108 individuals using personal, dense, dynamic data
1143 clouds. *Nat. Biotechnol.* **35**, 747-756.
1144 Saito, R., Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L., Lotia, S., Pico, A.R.*, et al.*
1145 (2012) A travel guide to Cytoscape plugins. *Nat. Methods* **9**, 1069-1076.
1146 Schnable, P.S. and Springer, N.M. (2013) Progress toward understanding heterosis in crop
1147 plants. *Annu. Rev. Plant Biol.* **64**, 71-88.
1148 Schrag, T., Westhues, M., Schipprack, W., Seifert, F., Thiemann, A., Scholten, S. and
1149 Melchinger, A.E. (2018) Beyond genomic prediction: combining different types of *omics*
1150 data can improve prediction of hybrid performance in maize. *Genetics* **208**, 1373-1385.
1151 Shen, X., Wang, R., Xiong, X., Yin, Y., Cai, Y., Ma, Z., Liu, N.*, et al.* (2019) Metabolic reaction
1152 network-based recursive metabolite annotation for untargeted metabolomics. *Nat.*
1153 *Commun.* **10**, 1516.
1154 Shull, G.H. (1948) What is "heterosis"? *Genetics* **33**, 439-446.
1155 Spitzer, M., Wildenhain, J., Rappsilber, J. and Tyers, M. (2014) BoxPlotR: a web tool for
1156 generation of box plots. *Nat. Methods* **11**, 121-122.
1157 Stecher, G., Tamura, K. and Kumar, S. (2020) Molecular Evolutionary Genetics Analysis
1158 (MEGA) for macOS. *Mol Biol Evol* **37**, 1237-1239.
1159 Sun, Z., Peng, J., Lv, Q., Ding, J., Chen, S., Duan, M., He, Q.*, et al.* (2023) Dissecting the
1160 genetic basis of heterosis in elite super-hybrid rice. *Plant Physiol.* **192**, 307-325.
1161 Tu, J., Yin, Y., Xu, M., Wang, R. and Zhu, Z.J. (2017) Absolute quantitative lipidomics reveals

lipidome-wide alterations in aging brain. *Metabolomics* **14**, 5.

Wang, B., Hou, M., Shi, J., Ku, L., Song, W., Li, C., Ning, Q.*, et al.* (2023) De novo genome assembly and analyses of 12 founder inbred lines provide insights into maize heterosis. *Nat. Genet.* **55**, 312-323.

Wang, K., Li, M. and Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164.

Wang, X., Zhou, T., Li, G., Yao, W., Hu, W., Wei, X., Che, J.*, et al.* (2022) A Ghd7-centered regulatory network provides a mechanistic approximation to optimal heterosis in an elite rice hybrid. *Plant J.* **112**, 68-83.

Wang, Z., Xue, Z. and Wang, T. (2014) Differential analysis of proteomes and metabolomes reveals additively balanced networking for metabolism in maize heterosis. *J. Proteome Res.* **13**, 3987-4001.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. and Barton, G.J. (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191.

Williams, W. (1959) Heterosis and the genetics of complex characters. *Nature* **184**, 527-530.

Wold, S., Sjöström, M. and Eriksson, L. (2001) PLS-regression: a basic tool of chemometrics. *Chemometrics Intellig. Lab. Syst.* **58**, 109-130.

Xiao, Y., Jiang, S., Cheng, Q., Wang, X., Yan, J., Zhang, R., Qiao, F.*, et al.* (2021) The genetic mechanism of heterosis utilization in maize improvement. *Genome Biol.* **22**, 148.

Xu, Y., Zhao, Y., Wang, X., Ma, Y., Li, P., Yang, Z., Zhang, X.*, et al.* (2021) Incorporation of parental phenotypic data into multi-omic models improves prediction of yield-related traits in hybrid rice. *Plant Biotechnol. J.* **19**, 261-272.

Yao, G. and Huang, W. (2013) Identifying *indica-japonica* characteristic of rice breeding materials with *indica-japonica* differentiation InDel markers. *Hybrid Rice* **28**, 53-57.

Yu, S.B., Li, J.X., Xu, C.G., Tan, Y.F., Gao, Y.J., Li, X.H., Zhang, Q.*, et al.* (1997) Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proc. Natl. Acad. Sci. USA* **94**, 9226-9231.

Zhang, M., Wang, Y., Hu, Y., Wang, H., Liu, Y., Zhao, B., Zhang, J.*, et al.* (2023) Heterosis in root microbiota inhibits growth of soil-borne fungal pathogens in hybrid rice. *J. Integr. Plant Biol.* **65**, 1059-1076.

Zhou, G., Ewald, J. and Xia, J. (2021a) OmicsAnalyst: a comprehensive web-based platform for visual analytics of multi-omics data. *Nucleic Acids Res.*

Zhou, G., Soufan, O., Ewald, J., Hancock, R.E.W., Basu, N. and Xia, J. (2019a) NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res.* **47**, W234-W241.

Zhou, P., Hirsch, C.N., Briggs, S.P. and Springer, N.M. (2019b) Dynamic patterns of gene expression additivity and regulatory variation throughout maize development. *Mol. Plant* **12**, 410-425.

Zhou, X., Nong, C., Wu, B., Zhou, T., Zhang, B., Liu, X., Gao, G.*, et al.* (2021b) Combinations of *Ghd7*, *Ghd8*, and *Hd1* determine strong heterosis of commercial rice hybrids in diverse ecological regions. *J. Exp. Bot.* **72**, 6963-6976.

Figure and figure legends

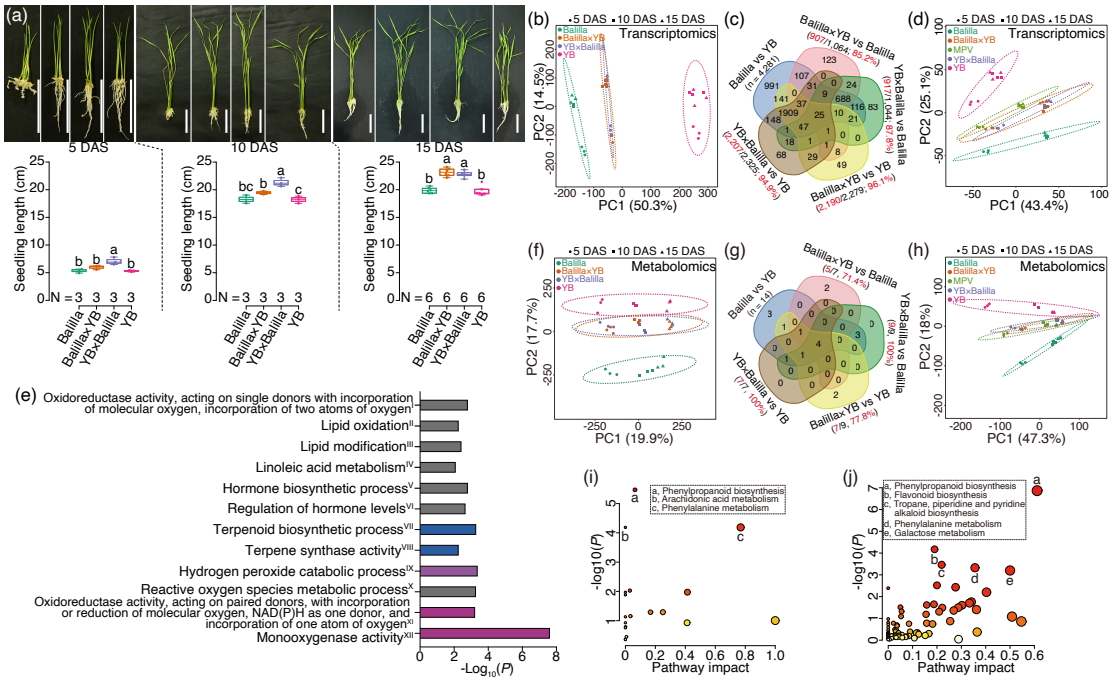

Figure 1 Heterosis-associated genes and metabolites for seedling length in rice. (a) Seedlings and seedling length of Balilla, Yuetai B (YB), and their reciprocal F_1 hybrids at 5, 10, and 15 days after sowing (DAS). Analysis of variance was implemented to compare differences. The significance level was set at 0.05. Scale bar = 5 cm. (b) PCA score plot of Balilla, YB, and F_1 hybrids using normalized expression levels of 22,194 genes. (c) Venn diagram for differentially expressed genes of parental inbred lines and F_1 hybrids at 15 DAS. Red texts indicate the number and percentage of DEGs derived from parental differences. (d) PCA score plot of Balilla, YB, F_1 hybrids, and mid-parent values (MPVs) using levels of 626 heterosis-associated genes. (e) Twelve significantly enriched Gene Ontology (GO) terms for the 626 heterosis-associated genes. Colored terms denote those overlapping with parental differential terms. (f) PCA score plot of Balilla, YB, and F_1 hybrids based on normalized levels of 10,522 metabolites. (g) Venn diagram for differential metabolic pathways (positive ion mode) of parental inbred lines and F_1 hybrids at 15 DAS. (h) PCA score plot based on 645 heterosis-associated metabolites. (i) Enriched pathways of heterosis-associated metabolites (positive ion mode). Three significantly enriched pathways are labeled. (j) Integrated pathway enrichment analysis of heterosis-associated genes and metabolites. Five significantly enriched pathways are indicated.

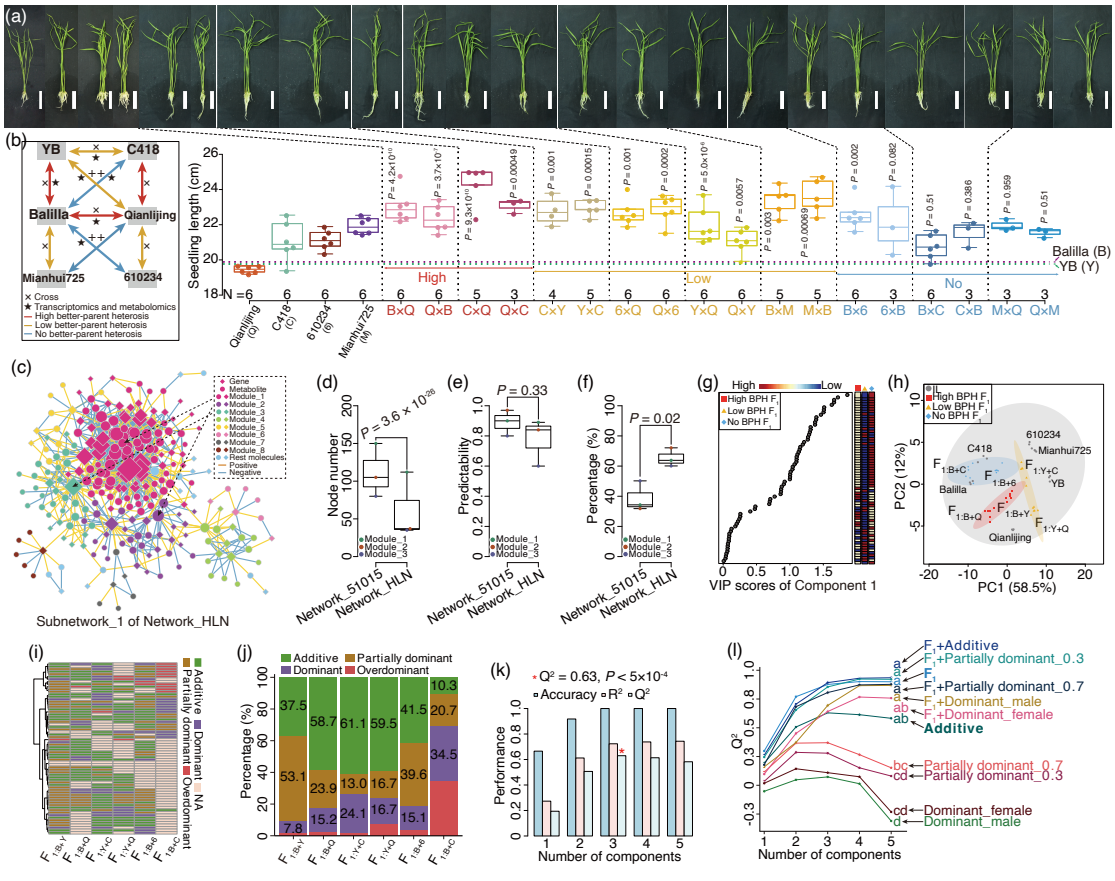

Figure 2 Inheritance patterns and predictabilities of heterosis-associated molecules for seedling length. (a) Seedlings and seedling length of four inbred lines and nine pairs of reciprocal F_1 hybrids. Seedling length was compared between F_1 hybrid and its better-parent. *P* values indicate least significant difference in post hoc test of analysis of variance. (b) Cross scheme of six inbred lines. (c) Subnetwork_1 of Network_HLN integrating heterosis-associated genes and metabolites. The subnetwork contains eight modules and 868 edges. Node size reflects node degree. (d-f) Comparison of Subnetwork_1 between Network_51015 and Network_HLN. Node number (d; weighted), predictability (e), and percentage of Component 1 (f) of top three modules are compared. *P* values are for independent samples *t* test. (g) Scores of variable importance in projections (VIP) of Component 1 for 68 heterosis-associated molecules across F_1 hybrids with high-, low-, or no better-parent heterosis (BPH). (h) PCA score plot of six inbred lines and their reciprocal F_1 hybrids based on 68 heterosis-associated molecules. (i) Inheritance patterns of the 68 molecules in six pairs of reciprocal F_1 hybrids. NA = unclassifiable inheritance patterns. (j) Percentages of four inheritance patterns. (k) Performance of additive effect. (l) Predictabilities (Q^2) of hybrid metabolite profiles (F_1), individual inheritance patterns, and their combinations for seedling length heterosis. Analysis of variance was performed with the least significant difference used in post hoc test.

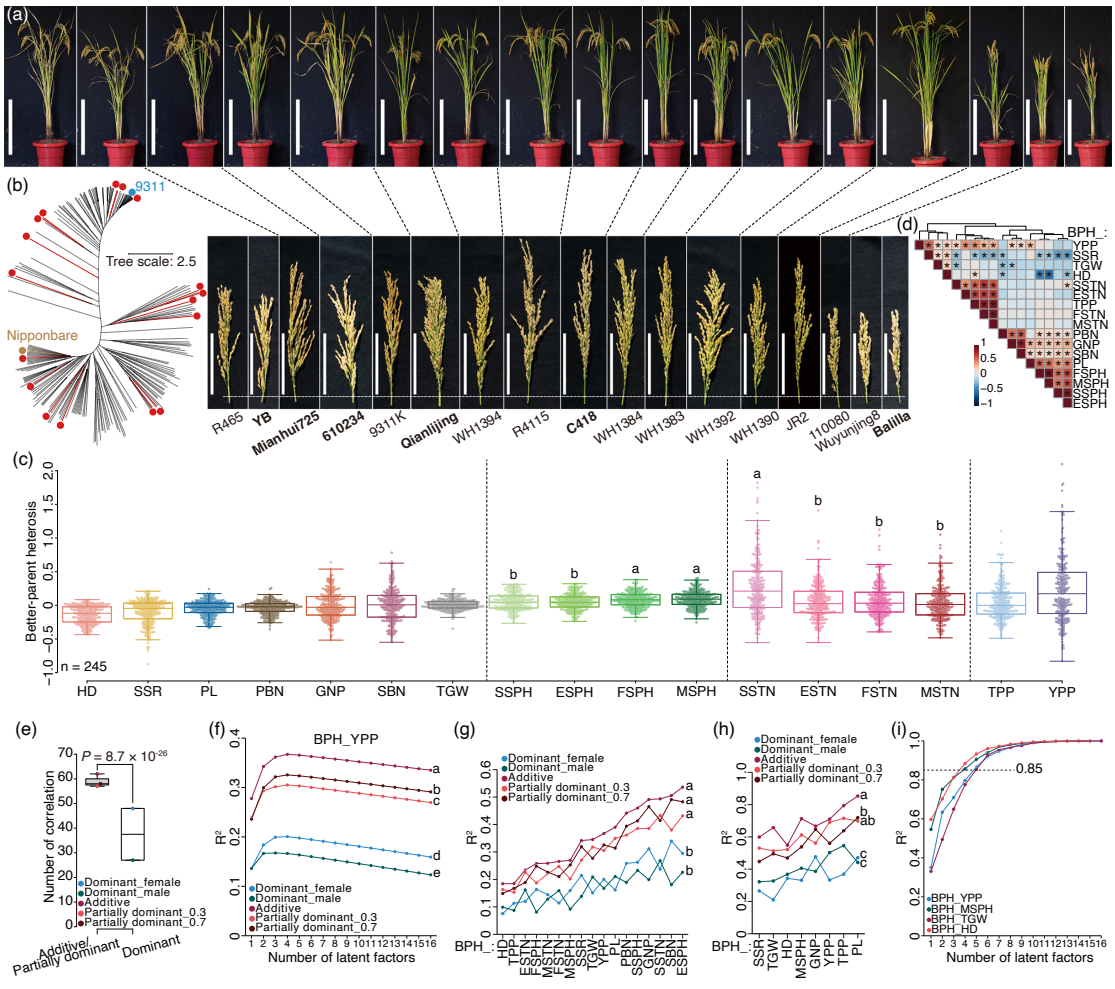

Figure 3 Additive and partially dominant effects contribute to heterosis of 17 agronomic traits in rice. (a) Plants and panicles of 17 inbred lines (ILs) used for complete diallel-crosses. Scale bars represent 50 cm and 10 cm for plants and panicles, respectively. The six ILs used in Figure 2b are in bold. (b) Neighbour-joining tree of 223 ILs based on InDels. The 17 selected ILs are with red dots. (c) Better-parent heterosis (BPH) of 17 traits in 245 F_1 hybrids. (d) Correlation heatmap for BPH of 17 traits. Asterisks indicate significant Pearson correlations ($P < 0.05$). (e) Number of significant correlations between different inheritance patterns of five chemically annotated metabolites and heterosis of 17 traits. *P* value indicates independent samples *t* tests (weighted). (f) R^2 values (adjusted) with varying number of latent factors in partial least squares analysis on better-parent heterosis for grain yield per plant (BPH_YPP). The 1,306 differential metabolites analyzed from 17 ILs were used as predictor with different inheritance patterns. (g) Top R^2 values of models for predicting heterosis of 17 traits. (h) Top R^2 values for predicting heterosis of eight traits across environments. (i) R^2 values for predicting heterosis of four traits in test-cross F_1 hybrids. Analysis of variance was performed with the least significant difference used in post hoc test in C, F, G, and H. Abbreviations: secondary branch number = SBN, grain number per panicle = GNP, tiller number per plant = TPP, seed setting rate = SSR, heading date = HD, thousand grain weight = TGW, seedling stage tiller number = SSTN, elongation stage tiller number = ESTN, flowering stage tiller number = FSTN, maturation stage tiller number = MSTN, seedling stage plant height = SSPH, elongation stage plant height = ESPH, flowering stage plant height = FSPH, maturation stage plant height = MSPH, panicle length = PL, and primary branch number = PBN.

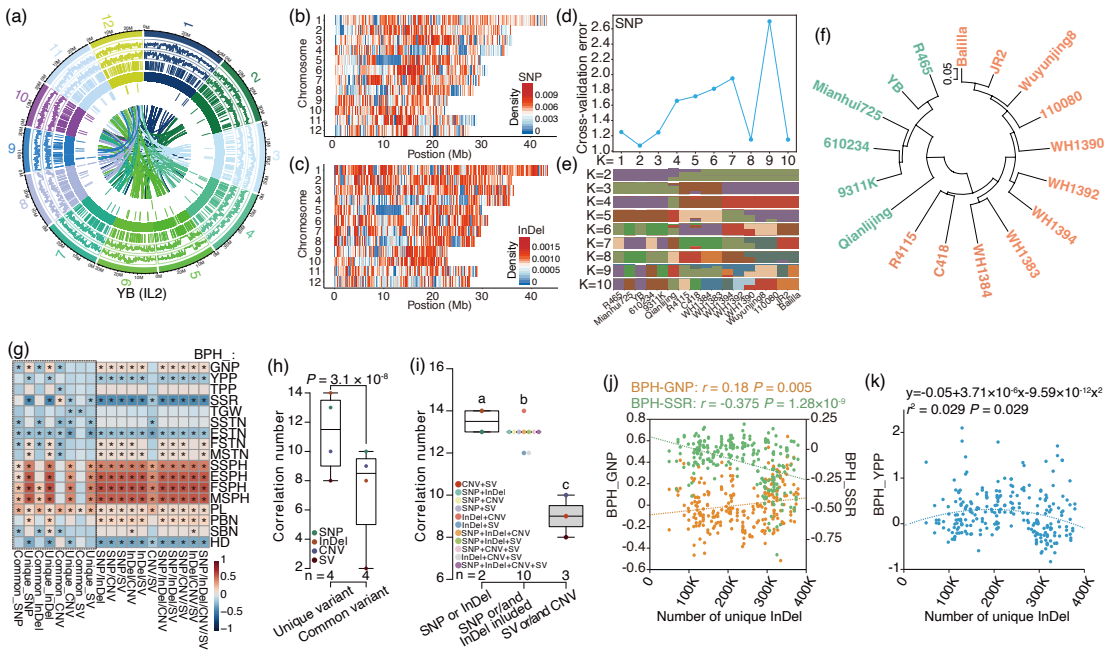

Figure 4 Parental genomic variants correlate with heterosis of agronomic traits in rice. (a) Circos visualization of genomic variants in Yuetai B (YB). Chromosome position, SNP density, InDel density, CNV duplication position, CNV deletion position, SV insertion position, SV deletion position, and SV inversion position are listed from outer to inner circles. (b-c) Heatmaps for SNP (b) and InDel (c) density of Yuetai B. (d) Cross-validation errors with K values ranging from 1 to 10 using SNPs from 17 inbred lines (ILs). (e) Population structure of 17 ILs based on SNPs. (f) Neighbour-joining tree of 17 ILs based on SNPs. (g) Correlation heatmap between parental genomic variants (common/unique) and heterosis of 17 traits. Asterisks indicate significant Spearman correlations ($P < 0.05$). (h) Number of significant correlations between parental genomic variants and trait heterosis. P value indicates independent samples t test (weighted). (i) Number of significant correlations between parental unique variants and trait heterosis. Analysis of variance was performed with the least significant difference used in post hoc test (weighted). (j) Correlations between parental unique InDels and heterosis for grain number per plant/seed setting rate. P value indicates Spearman rank correlation. (k) Quadratic regression between parental unique InDels and heterosis for grain yield per plant. Abbreviations: better-parent heterosis = BPH, yield per plant = YPP, secondary branch number = SBN, grain number per panicle = GNP, tiller number per plant = TPP, seed setting rate = SSR, heading date = HD, thousand grain weight = TGW, seedling stage tiller number = SSTN, elongation stage tiller number = ESTN, flowering stage tiller number = FSTN, maturation stage tiller number = MSTN, seedling stage plant height = SSPH, elongation stage plant height = ESPH, flowering stage plant height = FSPH, maturation stage plant height = MSPH, panicle length = PL, and primary branch number = PBN.

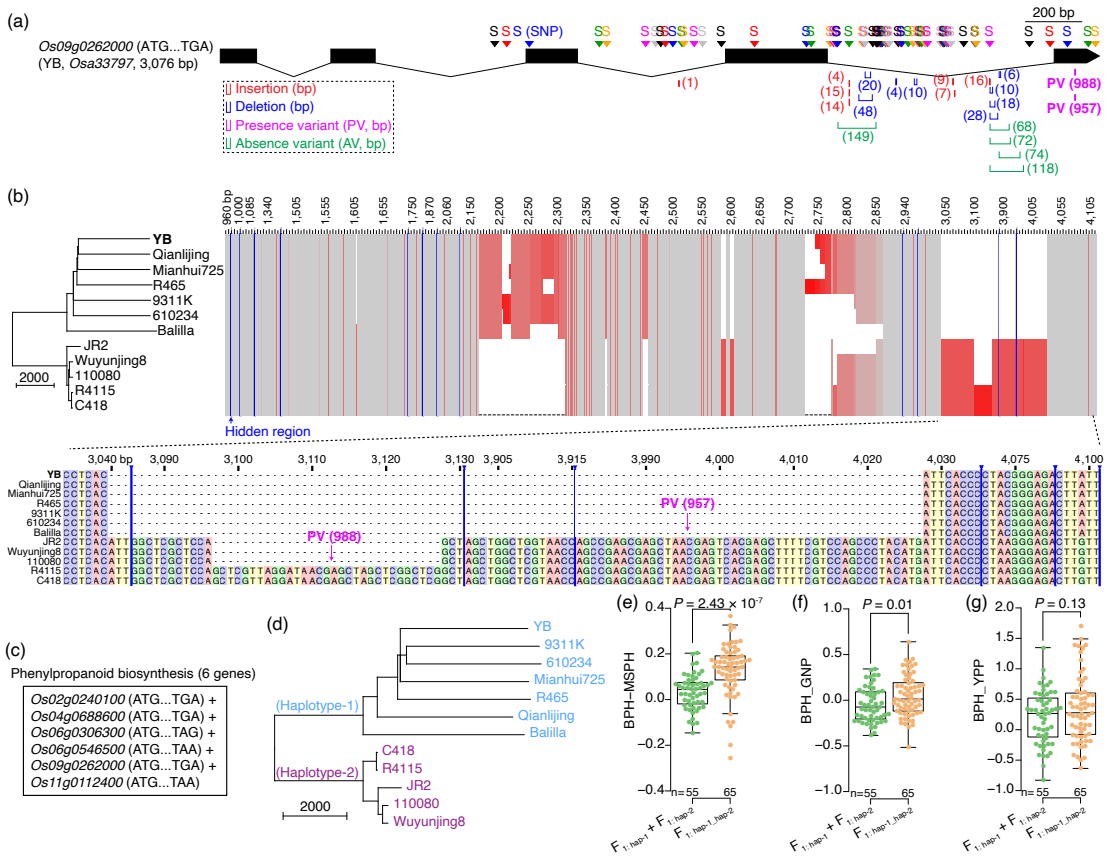

Figure 5 Genomic variants in phenylpropanoid biosynthesis genes contribute to rice heterosis. (a) Gene structure of *Os09g0262000*. With Yuetai B (YB) as the reference, genomic variants (SNPs, insertions, deletions, presence variants (PVs), and absence variants (AVs)) detected in 11 inbred lines are indicated. (b) Clustering of 12 inbred lines based on sequence variants of *Os09g0262000*. Red blocks represent the presence of variants. Detailed sequences of two PVs and two SNPs in the fifth exon are shown. (c) Six phenylpropanoid biosynthesis genes. (d) Haplotypes of 12 inbred lines based on sequence variants of the six genes. (e-g) Comparisons of better-parent heterosis for maturation stage plant height (BPH_MSPH; e), grain number per panicle (BPH_GNP; f), and grain yield per plant (BPH_YPP; g) between F_1 hybrids with heterozygous (haplotype-1_haplotype-2) versus homozygous haplotypes (haplotype-1 and haplotype-2). *P* values indicate independent samples *t* test.

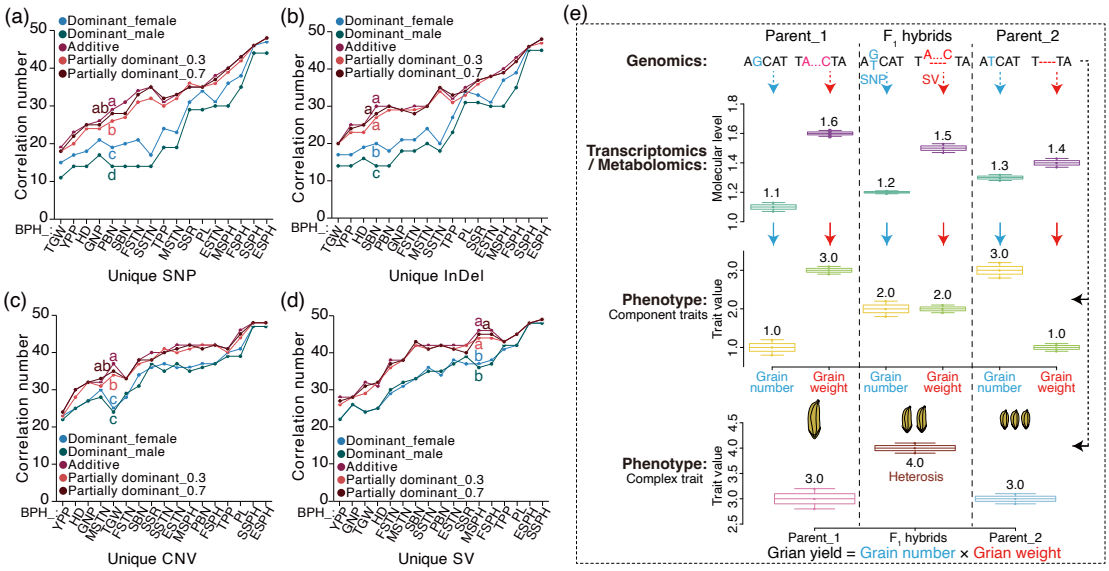

Figure 6 Associations of parental multi-omics variation, different inheritance patterns, and heterosis of component/complex traits in rice. (a-d) Number of significant correlations between parental unique variants and the top 50 heterosis-associated metabolites per trait (with different inheritance patterns). Analysis of variance (weighted) was performed with the least significant difference used in post hoc test. (e) A model of additive effect from genomic variants on yield heterosis in rice. To simplify interpretation, grain yield is determined by grain number and grain weight, with SNP and SV related with each trait, respectively. Due to additive effect from genomic variants in parents, the F_1 hybrids showed intermediate transcriptomic and metabolomic levels. Accordingly, in F_1 hybrids displayed mid-parent values for grain number and grain weight. With multiplicative interaction of the two yield components, positive yield heterosis occurred in F_1 hybrids, with equal yield in parents. Abbreviations: better-parent heterosis = BPH, yield per plant = YPP, secondary branch number = SBN, grain number per panicle = GNP, tiller number per plant = TPP, seed setting rate = SSR, heading date = HD, thousand grain weight = TGW, seedling stage tiller number = SSTN, elongation stage tiller number = ESTN, flowering stage tiller number = FSTN, maturation stage tiller number = MSTN, seedling stage plant height = SSPH, elongation stage plant height = ESPH, flowering stage plant height = FSPH, maturation stage plant height = MSPH, panicle length = PL, and primary branch number = PBN.