Machine learning-based prediction of dynamic height heterosis with pathway biomarkers in rice

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

[1]State Key Laboratory of Hybrid Rice, Key Laboratory for Research and Utilization of Heterosis in Indica Rice, the Ministry of Agriculture, The Yangtze River Valley Hybrid Rice Collaboration & Innovation Center, College of Life Sciences, Wuhan University, Wuhan 430072, China.

[†]These authors contributed equally to this work.

[*]Author for correspondence: wenchaoh@whu.edu.cn

Abstract

The development of robust biomarkers enables accurate prediction of complex phenotypes. However, the dynamic nature of biomarkers is often underestimated since their quantitative changes during development are directly connected to phenotypic transformations, influencing both crop agronomic traits and human diseases. Here, we performed network analysis of untargeted metabolite profiles to investigate height heterosis in rice, which is dynamic that varies during development and is a key determinant of yield heterosis. We found that the levels of pyruvaldehyde were predictive of height heterosis specific at the seedling stage, while 4-hydroxycinnamic acid positively correlated with height heterosis across four developmental stages. We identified metabolic pathways associated with height heterosis and found that metabolomic changes during the elongation stage had a greater impact than those in other stages. Finally, 11 heterosis-associated pathways were developed into metabolomic biomarkers through random forest analysis, successfully predicting height heterosis in an independent population under different growth conditions. This study elucidates the metabolomic landscape of dynamic height heterosis in rice and develops pathway biomarkers for complex phenotypes, demonstrating robustness across diverse populations, environments, and developmental stages.

Keywords

dynamic heterosis, pathway biomarker, metabolomics, plant height, rice

Introduction

Complex phenotypes are mainly determined by genotypes, environmental factors, and their interactions (Hunter, 2005; Crossa et al., 2017; Scheres and van der Putten, 2017). The influence of genetic and environmental factors on phenotypic changes has been extensively studied in both crop agronomic traits and human diseases (Fave et al., 2018; Li et al., 2018; de Los Campos et al., 2020). To accurately characterize the states of complex phenotypes, various types of biomarkers have been developed, and predictive models have been constructed to facilitate accurate predictions across populations and environments (Riedelsheimer et al., 2012; Sarkizova et al., 2020; Tong et al., 2020; Zhao et al., 2021). However, the dynamic nature of the processes shaping complex phenotypes is often underestimated (Lesterhuis et al., 2017; Miao et al., 2020). Quantitative changes in specific molecules during development have been shown to directly linked to phenotypic transformations, including plant agronomic traits (Obermeyer et al., 2013; Ko et al., 2016; Varoquaux et al., 2019; Zhou et al., 2019a) and risks for human diseases such as diabetes or cancers (Taylor et al., 2009; Price et al., 2017). Consequently, static biomarkers—identified from tissues at a single developmental stage—often exhibit limited predictive power for new-stage phenotypes in genetically distant individuals, partly due to molecular heterogeneity across developmental stages and genotypes (Menche et al., 2017; Price et al., 2017). Furthermore, the question of at which stages quantitative changes in biomarkers contribute more to the phenotypic outcome than at others remains largely unexplored. Thus, investigating the molecular mechanisms of complex phenotypes in a time series

manner, followed by the exploration and validation of dynamic biomarkers, offers a viable approach for advancing precision breeding and precision medicine.

Plant height is a key determinant of crop grain yield (Wang et al., 2023). Although plant height is not a component trait of grain yield, it significantly contributes to yield through its associations with grain number and flowering time. In rice, the close relationship between plant height and yield has been attributed to multiple polygenic genes, such as *Ghd7* and *DTH* (Xue et al., 2008; Gao et al., 2014). In a partial correlation analysis comparing the contributions of four yield components and plant height to yield heterosis in rice, heterosis for plant height (R = 0.22) had a higher correlation coefficient than that of grain weight (R = 0.16) (Dan et al., 2021). Notably, *Ghd7* plays a central role in yield heterosis of the elite hybrid rice Shanyou63 (Wang et al., 2022). However, plant height is a dynamic trait that evolves from seedling to maturity, yet most studies primarily focus on the maturation stage. Given the substantial increase in crop grain yield achieved through heterosis, investigating the molecular mechanisms underlying height heterosis has become essential for further enhancing yield potential.

In this study, we measured plant height in rice diallel crosses across four developmental stages: seedling, elongation, flowering, and maturation. We observed a significant shift in height heterosis between the elongation and flowering stages. Subsequently, we identified heterosis-associated analytes at each stage and examined their relationships with height heterosis. We also detected metabolic pathways involved in dynamic height heterosis and discovered that the overlapping pathways—

primarily in amino acid and carbohydrate metabolism—connected different developmental stages. Notably, we observed pronounced metabolomic changes at the elongation stage. Using random forest, we developed pathway biomarkers and validated them with test crosses. Our findings provide insights into the metabolomic mechanisms of dynamic height heterosis and demonstrate the potential of pathway biomarkers in predicting complex phenotypes across developmental stages.

Results

Dynamic heterosis for plant height of diallel crosses in rice

To investigate the molecular mechanisms underlying dynamic heterosis in rice, we selected 18 representative rice inbred lines, which consist of both *indica* and *japonica* varieties with varying plant height (Dan et al., 2021; Dan et al., 2024a; Dan et al., 2024b), as parental lines. Using a complete diallel design, we generated a hybrid population and measured plant height of the population (including both inbred lines and F_1 hybrids) at four developmental stages: seedling stage (SS), elongation stage (ES), flowering stage (FS), and maturation stage (MS). Plant height data of 18 inbred lines and a total of 287 F_1 hybrids were successfully collected, and better-parent heterosis for plant height (BPH_PH) were calculated for the hybrid population.

Height heterosis varied among individuals across the four stages (Fig. 1A and Supplementary Table S1). When analyzed collectively, the F_1 hybrids exhibited significantly stronger heterosis at later stages compared to earlier stages (Fig. 1B). Additionally, we classified the F_1 hybrids into high- and low-BPH groups based on the 75th and 25th percentiles of heterosis at each stage for subsequently comparative

analysis. Despite stage-specific variations, heterosis was positively correlated across stages, with correlation coefficients decreasing as the interval between stages increased (Fig. 1C).

Metabolic analytes associated with dynamic height heterosis

To explore the metabolic underling of dynamic heterosis, we integrated untargeted metabolite profiles with height heterosis across the four stages. We first obtained metabolite profiles from young seedlings of the parental inbred lines to calculate metabolite levels for F_1 hybrids. Since additive effect, namely at middle parental levels, is among the predominant inheritance patterns in F_1 hybrids (Zhou et al., 2019a; Liu et al., 2021; Dan et al., 2024a; Dan et al., 2024b), we averaged the parental metabolite levels to represent the hybrid metabolome as previously described (Dan et al., 2019; Dan et al., 2020). Subsequently, we employed a partial least square-based strategy to connect the calculated hybrid metabolome and dynamic heterosis (Wold, 1975; Dan et al., 2021). Predictive models for height heterosis of each stage were built and these models displayed best performance with four latent factors (Fig. 2A). The top *r* values of these predictive models ranged from 0.68 to 0.77, indicating high accuracy of the additive effect for representing metabolic levels of F_1 hybrids.

Contributions of each analyte in predictive models were assesses by average values of variable importance in projection, and the top contributed analytes were identified for heterosis at each stage. The final numbers of heterosis-associated analytes for the four stages were 300, 300, 100, and 100, respectively, suggesting the involvement of more metabolic pathways with early-stage heterosis (Fig. 2B).

We performed correlation analysis on these heterosis-associated analytes to investigate their correlation patterns. Both positive and negative correlations were found and most of these analytes had positive correlations with other analytes across the four stages (Fig. 2C-F). Notably, an accurately annotated metabolite pyruvaldehyde (peak tag: M71T32_NEG; Fig. 2G), which is involved in pyruvate metabolism and propanoate metabolism, showed a significant negative correlation with height heterosis at the seedling stage (Fig. 2H). In contrast, 4-hydroxycinnamic acid, which has been proven to be linked to plant height (Li et al., 2015; Dan et al., 2021), was consistently identified as a heterosis-associated analyte across all four stages (Fig. 2C-F). The levels of 4-hydroxycinnamic acid showed significant positive correlations with plant height heterosis throughout the stages, with the later stages exhibiting lower correlation coefficients compared to the earlier ones (Fig. 2J).

Metabolic pathways underlying dynamic height heterosis

Due to limited annotation information of the heterosis-associated analytes, it was challenging to classify them into metabolic pathways. We thus adopted dysregulated network analysis (Shen et al., 2019), which involves pathway enrichment analysis using a hypergeometric test on the differential metabolic peaks, to identify heterosis-associated metabolic pathways. By treating the low- and high-BPH hybrids as control and case groups, respectively, we obtained significantly enriched pathways (heterosis-associated pathways) for each stage. Consistent with the trend in numbers of heterosis-associated analytes, more pathways were implicated in height heterosis at early stages (Fig. 3A-D). Through metabolic levels of these pathways, namely the average levels of

153 all metabolites in a pathway (Shen et al., 2019), we performed correlation analyses and
154 observed similar correlation patterns of these pathways to those of heterosis-associated
155 analytes (Fig. 3A-C).

156 Among these pathways, pyruvate metabolism and propanoate metabolism were
157 associated with height heterosis, while only pyruvate metabolism was also involved in
158 heterosis at the later stages. Pyruvate metabolism exhibited a positive correlation with
159 height across all four stages, whereas propanoate metabolism showed a positive
160 correlation at the seedling stage but a negative correlation at elongation. These findings
161 suggested that the relationships between metabolic pathways and heterosis might differ
162 from those of individual analytes, underscoring the importance of exploring height
163 heterosis at the pathway level.

164 By interconnecting overlapping pathway across stages, we obtained the
165 metabolomic landscape of dynamic height heterosis (Fig. 3E). A total of 29 metabolic
166 pathways were detected, with three, five, and six metabolic pathways involved in height
167 heterosis at two, three, and all four stages, respectively. The majority of these heterosis-
168 associated pathways belonged to amino acid (48.28%) and carbohydrate (31.03%)
169 metabolism (Fig. 3F). Notably, the pathways in carbohydrate metabolism (60-75%)
170 consistently exhibited high metabolite levels in high-BPH hybrids, while those in amino
171 acid metabolism (50-90%) displayed lower levels (Fig. 3G). This pattern underscored
172 the dominant roles of carbohydrate and amino acid metabolism in height heterosis,
173 which complements the changing trends observed for metabolic pathways related to
174 heterosis of yield and its components (Dan et al., 2020; Dan et al., 2021).

175 Furthermore, the numbers of heterosis-associated pathways were higher for the
176 earlier stages (8 and 6 pathways) than the later stages (1 and 0 pathway; Fig. 3H),
177 suggesting the changes in heterosis-associated pathways drove height heterosis
178 variation. However, the contributions of these heterosis-associated pathways at each
179 stage to height heterosis remained unclear.

180 **Metabolomic changes at the elongation stage play critical roles in height**
181 **heterosis**

182 According to the metabolomic landscape of height heterosis, six pathways were
183 significantly enriched across all four stages. Three of these pathways (citrate cycle,
184 pentose and glucuronate interconversions, and pyruvate metabolism) were classified
185 under carbohydrate metabolism, while two pathways were from amino acid
186 metabolism (phenylalanine metabolism and alanine, aspartate and glutamate
187 metabolism). The remaining pathway was monobactam biosynthesis. To assess
188 metabolomic variations linked to dynamic heterosis, we compared metabolite levels
189 of these pathways across stages (Fig. 4A-F). Significant differences in metabolite
190 levels of the six pathways were observed in either the high- or low-BPH hybrids
191 across the four stages. Nearly all pathways exhibited metabolomic differences in at
192 least one stage when compared to the others. In high-BPH hybrids, the citrate cycle
193 showed an initial increase at elongation followed by a decline in later stages (Fig.
194 4A), whereas the opposite trend occurred in low-BPH hybrids. Similar changing
195 patterns were observed for the other two pathways from carbohydrate metabolism
196 (Fig. 4B-C). In contrast, the two pathways from amino acid metabolism and

197 monobactam biosynthesis displayed fewer changes in metabolite levels in low-BPH
198 hybrids (Fig. 4D-F). Notably, metabolite levels frequently shifted significantly at the
199 elongation stage in both high- and low-BPH hybrids, underscoring its critical role in
200 shaping height heterosis.

201 **Pathway biomarkers for predicting dynamic height heterosis**

202 To explore biomarkers for the dynamic height heterosis, we first performed univariate
203 and multivariate receiver operating characteristic (ROC) curve analyses on the 18
204 heterosis-associated metabolic pathways for height heterosis at the seedling stage
205 (Supplementary Fig. S1). A single predictor achieved an area under the curve (AUC)
206 of 0.885, increasing to 0.933 with 100 features. Analysis of the six heterosis-
207 associated pathways across all stages revealed no substantial increase in AUC values
208 or predictive accuracies (uni-ROC: AUC = 0.887, multi-ROC: AUC = 0.931,
209 predictive accuracy = 0.882 with 21 features; Supplementary Fig. S1).

210 We then evaluated the 11 heterosis-associated pathways that were for three or
211 four stages, which were primarily pathways in amino acid and carbohydrate
212 metabolism, as predictive features (Fig. 5A). A single predictor yielded an AUC was
213 0.883 (Fig. 5B), while the optimal predictive model (10 features) achieved an AUC of
214 0.929 and predictive accuracy of 0.873 (Fig. 5C-D). Unexpectedly, despite being
215 significantly enriched across all four stages, the citrate cycle was not among the top-
216 ranked ten predictors (Fig. 5E). Further biomarker analysis demonstrated high
217 predictive accuracies for height heterosis at the three later stages: elongation,
218 flowering, and maturation (Supplementary Fig. S2). These results showed that the

219 development of biomarkers from the metabolic pathways associated with heterosis
220 across most developmental stages could be an effective strategy, taking into account
221 predictive accuracy, feature number, and feature ranking.

222 **Validation of pathway biomarkers with test crosses**

223 To validate predictive capability of these pathway biomarkers, we utilized height
224 heterosis of a group of Honglian-type test crosses, which exhibited distant genetic
225 relationships and were cultivated under different field conditions than the diallel
226 crosses (Dan et al., 2020). The test crosses exhibited a narrower range of height
227 heterosis due to limited genetic diversity of their male parents, which were mainly
228 recombinant inbred lines (Fig. 6A). Meanwhile, the height heterosis (at the maturation
229 stage) of the test crosses was comparable to those of the diallel ones. After
230 categorizing the test crosses into low- and high-BPH groups based on the 25th and
231 75th percentiles of height heterosis, we obtained metabolite levels for four pathways
232 through dysregulated network analysis (Supplementary Table S2), three of which
233 overlapped with the 11 previously identified pathways. Using the metabolite levels of
234 nicotinate and nicotinamide metabolism as a single predictor yielded an AUC of 0.868
235 (Fig. 6B), while AUC values ranging from 0.8 to 0.834 were achieved with 2-6
236 variables (Supplementary Fig. S3).

237 Height heterosis was negatively correlated with the metabolite levels of nicotinate
238 and nicotinamide metabolism, as well as glyoxylate and dicarboxylate metabolism
239 (Fig. 6C). But no correlation was found between height heterosis and phenylalanine
240 metabolism. Additionally, the metabolite levels of nicotinate and nicotinamide

241 metabolism were significantly lower in the high-BPH hybrids compared to low-BPH
242 hybrids (Fig. 6D). The levels of glyoxylate and dicarboxylate metabolism were also
243 lower in the high-BPH hybrids (Fig. 6E), which contrasted with the result shown in
244 Fig. 3E. Phenylalanine metabolism showed no significant changes (Fig. 6F).
245 Furthermore, higher AUC values were observed when grouping the low- and high-
246 BPH hybrids at a more stringent level, while lower values were obtained when
247 dividing the test crosses into approximately equal halves (Supplementary Figs S4-S5).
248 Our results indicated that the 11 identified heterosis-associated pathways, which were
249 significantly enriched at most developmental stages in the diallel crosses, effectively
250 predicted height heterosis in the test crosses.

251 Discussion

252 In this study, we identified robust biomarkers for dynamic height heterosis in rice,
253 based on metabolic pathways significantly enriched at most developmental stages. In
254 alignment with the phenotypic transformations observed in fields, metabolomic
255 changes at the elongation stage contribute more substantially to height heterosis than
256 those at the other stages. Recognizing these critical stages through time-series analysis
257 may facilitate the identification of biomarkers for predicting biomass or height in
258 breeding programs for perennial energy crops and trees (Canas et al., 2015; Maddison
259 et al., 2017).

260 Among the 11 selected metabolic pathways, all of which were significantly
261 enriched at the elongation stage, pathways related to carbohydrate and amino acid
262 metabolism comprised the majority and exhibited approximately opposite trends in

263 metabolite levels. Given the tight connection between plant height and grain yield in
264 rice (Xue et al., 2008; Chen et al., 2020; Dan et al., 2021), variations in these two
265 types of pathways may determine the optimal balance between vegetative and
266 reproductive traits in the formation of heterosis in hybrid rice (Birchler, 2019; Li et
267 al., 2020; Dan et al., 2021). Furthermore, the strategy of incorporating pathways that
268 are significantly enriched at most developmental stages—rather than exclusively
269 those enriched at all stages—could enhance the robustness of pathway biomarkers for
270 predictions across diverse populations, environments, and developmental stages. We
271 think that this advantage arises from the comprehensive information provided by the
272 predictive features.

273 While we have identified multiple metabolic pathways associated with dynamic
274 height heterosis, their mechanistic roles require in-depth investigation. Simultaneous
275 mutations of key genes within these metabolic pathways could clarify their functional
276 contributions. Notably, the numbers of heterosis-associated pathways and analytes
277 were greater at the earlier stages than at the later ones, suggesting a more complex and
278 potentially pivotal metabolomic basis for heterosis at early developmental stages.
279 Moreover, our pathway analysis is inherently limited by the coverage of the KEGG
280 database; unannotated pathways may also influence height heterosis. Expanding the
281 catalog of heterosis-associated pathways will refine the metabolomic landscape and
282 improve predictive accuracy.

283 While the variation patterns of some pathways were consistent across hybrid
284 populations, we observed discrepancies in metabolite levels of certain pathways. The

metabolomic heterogeneity between populations (Dan et al., 2021), similar to transcriptional variability among phenotypically similar patients (Menche et al., 2017; Zhou et al., 2019b), may underlie these inconsistencies. Given the observed variations in predictive accuracies, hybrids exhibiting intermediate heterosis warrant further investigation. Comprehensive metabolomics data that incorporate metabolite profiles obtained via both positive and negative ionization modes, as well as through both GC and LC methods, will be essential. The integration of multiple layers of omics information (Weckwerth et al., 2020; Xu et al., 2021), including genomics, transcriptomics, proteomics, metabolomics, and microbiome data, will also optimize pathway biomarkers for predicting complex phenotypes across populations (individuals), environments, and developmental stages.

Material and methods

Plant materials

The rice accessions selected for hybridization were previously described (Dan et al., 2020). In brief, two rice hybrid populations were utilized to develop biomarkers for plant height heterosis. The first population adopted a complete diallel cross design involving 18 *indica* and *japonica* accessions as parents, while the second population consisted of 107 test crosses with recombinant inbred lines crossed with the Honglian-type cytoplasmic male sterile line, Yuetai A. The plant height, defined as the distance from the plant base to the highest part of the leaf or panicle, of the diallel cross population was measured with tape measures at four growth stages from seedling to maturation: the seedling stage (28 days after transplantation), elongation stage (47

307 days after transplantation), flowering stage (time of flowering), and maturation stage
308 (time of maturation). The plant height data for the maturation stage of the diallel cross
309 population were reported in a previous study (Dan et al., 2021), and the complete
310 plant height data are provided in Supplementary table S1. For comparative purpose,
311 several analysis results related to better-parent heterosis for plant height at the
312 maturation stage were retrieved from earlier research (Dan et al., 2021). The plant
313 height of the test cross population was only measured at the maturation stage. The two
314 hybrid populations were grown under different growth conditions and planting
315 densities at the same location (Ezhou, Hubei Province) in 2012 and 2015. Three
316 biological replicates were designed for each genotype, with five plants from the
317 middle region of each row selected for height measurement.

318 **Metabolite profiling**

319 Untargeted metabolite profiles in the negative ionization mode of 15-d-old parental
320 seedlings were obtained using previously described methods (Dan et al., 2020). In
321 brief, a 1290 Infinity liquid chromatography system (Agilent Technologies), an
322 Agilent quadrupole time-of-flight mass spectrometer (Agilent 6550 iFunnel QTOF),
323 and a Triple TOF 6600 mass spectrometer (AB SCIEX) were employed to acquire the
324 raw metabolomic profiles. The stability of the equipment and the accuracy of the
325 metabolite profiles were assessed using intermittently injected quality control
326 samples. A total of 3,746 analytes were detected, and 114 metabolites were identified
327 through matching with in-house spectral libraries of standards by Shanghai Applied
328 Protein Technology Co., Ltd. To prepare the metabolomic data for statistical analyses,

329 sample normalization (normalization by sum), data transformation (none), and data
330 scaling (auto scaling) were performed using MetaboAnalyst 4.0 (Chong et al., 2018).
331 The hybrid metabolome used for the analyses was calculated by averaging the
332 metabolite levels of the corresponding parents according to previously reported
333 methods (Dan et al., 2019; Dan et al., 2020).

334 **Identification of heterosis-associated metabolic pathways and analytes**

335 The Metabolite identification and Dysregulated Network Analysis web server
336 (MetDNA version 1.2.1) (Shen et al., 2019), which annotates metabolites based on the
337 rule that neighboring metabolites tend to share similar MS2 spectra, was used to
338 perform pathway enrichment analysis for height heterosis. The hybrids were
339 categorized into high-better-parent heterosis (control group) and low-better-parent
340 heterosis (case group) groups according to quantiles (≥ 75th percentile and ≤ 25th
341 percentile) of the entire hybrid population, and the calculated hybrid metabolome was
342 utilized for dysregulated network analysis. Analysis parameters were set according to
343 the metabolite profiling experiments, with *Arabidopsis thaliana* (thale cress) selected
344 as the species. Student's *t*-test was employed for univariate statistics, with a cutoff *P*-
345 value of 0.05 (with adjustment). Enriched metabolic pathways and quantitative
346 information on metabolic pathways, which were included in the analysis report, were
347 combined to identify significantly enriched metabolic pathways for height heterosis at
348 each stage. The metabolic pathways that were significantly enriched (*P* values less
349 than 0.05) and exhibited significant differences in metabolite levels between the high-
350 and low-better-parent heterosis hybrids were treated as height heterosis-associated

351 metabolic pathways.

352 For the identification of heterosis-associated analytes, we employed a partial least
353 square-based strategy (Wold, 1975; Dan et al., 2021). Initially, the number of latent
354 factors in the models was determined based on the highest achieved *r* values. The
355 performance of the predictive models was assessed using 10-fold cross-validation and
356 permutation tests conducted with MetaboAnalyst 4.0 (Chong et al., 2018).
357 Subsequently, average values of variable importance in the projection for all latent
358 factors were calculated to evaluate the contributions of specific analytes to plant
359 height heterosis. Redundant analytes in the predictive models were removed by
360 reducing the numbers of predictive analytes. The final number of predictive analytes
361 for plant height heterosis at the seedling, elongation, flowering, and maturation stages
362 were 300, 300, 100, and 100, respectively.

363 **Statistical analyses**

364 Better-parent heterosis for plant height was calculated using the equation: BPH = (F_1-
365 P_H)/P_H, where F_1 represents the plant height of the hybrid and P_H denotes the height of
366 the taller parent. The hybrid metabolome was calculated as follows: M_H = (M_F +
367 M_M)/2, where M_H, M_F, and M_M indicate the metabolite levels of the F_1 hybrid, female
368 parent, and male parent, respectively. Pearson correlation analysis of height heterosis
369 across the four developmental stages was performed using the "Statistical Analysis"
370 module in MetaboAnalyst 4.0 (Chong et al., 2018). Comparisons of height heterosis
371 and metabolite levels of significantly enriched pathways across stages were analyzed
372 using one-way ANOVA (with least significant difference for post hoc multiple

comparisons, $P < 0.05$) via SPSS (IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.). Analyses comparing plant height heterosis between the diallel and test crosses, as well as metabolite levels between high- and low-BPH hybrids, were conducted using independent samples *t*-test in SPSS (two-tailed). Pearson correlation analysis between height heterosis and the metabolite levels of metabolic pathways was also performed using SPSS. Both univariate and multivariate ROC curve analyses were conducted using the "Biomarker Analysis" module in MetaboAnalyst 4.0 (Chong et al., 2018). The metabolic ratios of the top 100 were computed and included in the subsequent analysis. Random forest analysis (Breiman, 2001) was selected as the classification and feature ranking method for the multivariate ROC curve analysis.

Author contributions

Z.D. designed the research; Z.D. and W.H. recorded field phenotypic data; Z.D. and Y.C. performed most of the experiments; Z.D. performed the data analyses; W.H. supervised the experiments; Z.D., Y.C., and W.H. wrote the manuscript.

Funding

This research was supported by the National Natural Science Foundation of China (Grant No. 31801439, 32472185, and 32101667), the China Postdoctoral Science Foundation (2019M660186, and 2022T150500, and 2023T160497), the Key Research and Development Program of Hubei Province (2022BFE003), the National Key R&D Program of China (2017YFD0100400) and the Hubei Agriculture Science and Technology Innovation Center Program.

Conflict of interest statement

The authors declare no competing interest.

Data availability

Raw metabolite profiles of experimental and quality control samples and corresponding metabolomic data are available from the MetaboLights database under the accession number MTBLS742.

Supplementary data

Supplementary Table S1 Plant height of inbred lines and the F_1 hybrids.

Supplementary Table S2 Metabolite levels of four metabolic pathways for 26 high- and 26 low-BPH test crosses.

Supplementary Figure S1 Univariate and multivariate ROC curve analysis of different numbers of significantly enriched pathways for height heterosis at seedling stage of the diallel crosses.

Supplementary Figure S2 Univariate and multivariate ROC curve analysis of the 11 selected metabolic pathways for height heterosis at the three later developmental stages of the diallel crosses.

Supplementary Figure S3 Multivariate ROC curve analysis of three metabolic pathways for height heterosis of the test crosses.

Supplementary Figure S4 Validation of pathway biomarkers with the top 15 and bottom 15 test crosses.

Supplementary Figure S5 Validation of pathway biomarkers with the entire test crosses.

References

Birchler JA (2019) Genomic balance plays out in evolution. Plant Cell **31**(6)**:**1186-1187.

Breiman L (2001) Random forests. Machine Learning **45:**5-32.

Canas RA, Canales J, Munoz-Hernandez C, Granados JM, Avila C, Garcia-Martin ML, and Canovas FM (2015) Understanding developmental and adaptive cues in pine through metabolite profiling and co-expression network analysis. Journal of Experimental Botany **66**(11)**:**3113-3127.

Chen Y, Dan Z, Gao F, Chen P, Fan F, and Li S (2020) Rice *GROWTH-REGULATING FACTOR7* modulates plant architecture through regulating GA and indole-3-acetic acid metabolism. Plant Physiology **184**(1)**:**393-406.

Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, Wishart DS, and Xia J (2018) MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids Research **46**(W1)**:**W486-W494.

Crossa J, Perez-Rodriguez P, Cuevas J, Montesinos-Lopez O, Jarquin D, de Los Campos G, Burgueno J, Gonzalez-Camacho JM, Perez-Elizalde S, Beyene Y, Dreisigacker S, Singh R, Zhang X, Gowda M, Roorkiwal M, Rutkoski J, and Varshney RK (2017) Genomic selection in plant breeding: methods, models, and perspectives. Trends in Plant Science **22**(11)**:**961-975.

Dan Z, Chen Y, and Huang W (2024a) Structural variations contribute to genetic diversity and heterosis in rice. bioRxiv**:**10.1101/2024.1111.1103.621459.

Dan Z, Chen Y, Li H, Zeng Y, Xu W, Zhao W, He R, and Huang W (2021) The metabolomic landscape of rice heterosis highlights pathway biomarkers for predicting complex phenotypes. Plant Physiology **187**(2)**:**1011-1025.

Dan Z, Chen Y, Xu Y, Huang JR, Huang JS, Hu J, Yao G, Zhu Y, and Huang W (2019) A metabolome-based core hybridisation strategy for the prediction of rice grain weight across environments. Plant Biotechnology Journal **17**(5)**:**906-913.

Dan Z, Chen Y, Zhao W, Wang Q, and Huang W (2020) Metabolome-based prediction of yield heterosis contributes to the breeding of elite rice. Life Science Alliance **3**(1)**:**e201900551.

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

de Los Campos G, Perez-Rodriguez P, Bogard M, Gouache D, and Crossa J (2020) A data-driven simulation platform to predict cultivars' performances under uncertain weather conditions. Nature Communications **11**(1)**:**4876.

Fave MJ, Lamaze FC, Soave D, Hodgkinson A, Gauvin H, Bruat V, Grenier JC, Gbeha E, Skead K, Smargiassi A, Johnson M, Idaghdour Y, and Awadalla P (2018) Gene-by-environment interactions in urban populations modulate risk phenotypes. Nature Communications **9**(1)**:**827.

Gao H, Jin M, Zheng XM, Chen J, Yuan D, Xin Y, Wang M, Huang D, Zhang Z, Zhou K, Sheng P, Ma J, Ma W, Deng H, Jiang L, Liu S, Wang H, Wu C, Yuan L, and Wan J (2014) Days to heading 7, a major quantitative locus determining photoperiod sensitivity and regional adaptation in rice. Proceedings of the National Academy of Sciences of the United States of America.

Hunter DJ (2005) Gene-environment interactions in human diseases. Nature Reviews Genetics **6**(4)**:**287-298.

Ko DK, Rohozinski D, Song Q, Taylor SH, Juenger TE, Harmon FG, and Chen ZJ (2016) Temporal shift of circadian-mediated gene expression and carbon fixation contributes to biomass heterosis in maize hybrids. PLoS Genetics **12**(7)**:**e1006197.

Lesterhuis WJ, Bosco A, Millward MJ, Small M, Nowak AK, and Lake RA (2017) Dynamic versus static biomarkers in cancer immune checkpoint blockade: unravelling complexity. Nature Reviews Drug Discovery **16**(4)**:**264-272.

Li X, Guo T, Mu Q, Li X, and Yu J (2018) Genomic and environmental determinants and their interplay underlying phenotypic plasticity. Proceedings of the National Academy of Sciences of the United States of America **115**(26)**:**6679-6684.

Li Y, Kim JI, Pysh L, and Chapple C (2015) Four isoforms of arabidopsis 4-coumarate:CoA ligase have overlapping yet distinct roles in phenylpropanoid metabolism. Plant Physiology **169**(4)**:**2409-2421.

Li Z, Zhu A, Song Q, Chen HY, Harmon FG, and Chen ZJ (2020) Temporal regulation of the metabolome and proteome in photosynthetic and photorespiratory pathways contributes to maize heterosis. Plant Cell **32**(12)**:**3706-3722.

Liu W, He G, and Deng XW (2021) Biological pathway expression complementation contributes to biomass heterosis in Arabidopsis. Proc Natl Acad Sci U S A **118**(16).

Maddison AL, Camargo-Rodriguez A, Scott IM, Jones CM, Elias DMO, Hawkins S, Massey A, Clifton-Brown J, McNamara NP, Donnison IS, and Purdy SJ (2017) Predicting future biomass yield in *Miscanthus* using the carbohydrate metabolic profile as a biomarker. GCB Bioenergy **9:**1264-1278.

Menche J, Guney E, Sharma A, Branigan PJ, Loza MJ, Baribaud F, Dobrin R, and Barabasi AL (2017) Integrating personalized gene expression profiles into predictive disease-associated gene pools. npj Systems Biology and Applications **3:**10.

Miao C, Xu Y, Liu S, Schnable PS, and Schnable JC (2020) Increased power and accuracy of causal locus identification in time series genome-wide association in sorghum. Plant Physiology **183**(4)**:**1898-1909.

Obermeyer G, Fragner L, Lang V, and Weckwerth W (2013) Dynamic adaption of metabolic pathways during germination and growth of lily pollen tubes after inhibition of the electron transport chain. Plant Physiology **162**(4)**:**1822-1833.

Price ND, Magis AT, Earls JC, Glusman G, Levy R, Lausted C, McDonald DT, Kusebauch U, Moss CL, Zhou Y, Qin S, Moritz RL, Brogaard K, Omenn GS, Lovejoy JC, and Hood L (2017) A wellness study of 108 individuals using personal, dense, dynamic data clouds. Nature Biotechnology **35**(8)**:**747-756.

Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, Altmann T, Stitt M, Willmitzer L, and Melchinger AE (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nature Genetics **44**(2)**:**217-220.

Sarkizova S, Klaeger S, Le PM, Li LW, Oliveira G, Keshishian H, Hartigan CR, Zhang W, Braun DA, Ligon KL, Bachireddy P, Zervantonakis IK, Rosenbluth JM, Ouspenskaia T, Law T, Justesen S, Stevens J, Lane WJ, Eisenhaure T, Lan Zhang G, Clauser KR, Hacohen N, Carr SA, Wu CJ, and Keskin DB (2020) A large peptidome dataset improves HLA class I epitope prediction across most of the human population. Nature Biotechnology **38**(2)**:**199-209.

Scheres B, and van der Putten WH (2017) The plant perceptron connects environment to development. Nature **543**(7645)**:**337-345.

Shen X, Wang R, Xiong X, Yin Y, Cai Y, Ma Z, Liu N, and Zhu Z-J (2019) Metabolic reaction network-based recursive metabolite annotation for untargeted metabolomics. Nature Communications **10**(1)**:**1516.

Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, Bull S, Pawson T, Morris Q, and Wrana JL (2009) Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nature Biotechnology **27**(2)**:**199-204.

Tong H, Kuken A, and Nikoloski Z (2020) Integrating molecular markers into metabolic models improves genomic selection for Arabidopsis growth. Nature Communications **11**(1)**:**2410.

Varoquaux N, Cole B, Gao C, Pierroz G, Baker CR, Patel D, Madera M, Jeffers T, Hollingsworth J, Sievert J, Yoshinaga Y, Owiti JA, Singan VR, DeGraaf S, Xu L, Blow MJ, Harrison MJ, Visel A, Jansson C, Niyogi KK, Hutmacher R, Coleman-Derr D, O'Malley RC, Taylor JW, Dahlberg J, Vogel JP, Lemaux PG, and Purdom E (2019) Transcriptomic analysis of field-droughted sorghum from seedling to maturity reveals biotic and metabolic responses. Proceedings

of the National Academy of Sciences of the United States of America **116**(52)**:**27124-27132.

Wang W, Guo W, Le L, Yu J, Wu Y, Li D, Wang Y, Wang H, Lu X, Qiao H, Gu X, Tian J, Zhang C, and Pu L (2023) Integration of high-throughput phenotyping, GWAS, and predictive models reveals the genetic architecture of plant height in maize. Molecular Plant **16**(2)**:**354-373.

Wang X, Zhou T, Li G, Yao W, Hu W, Wei X, Che J, Yang H, Shao L, Hua J, Li X, Xiao J, Xing Y, Ouyang Y, and Zhang Q (2022) A Ghd7-centered regulatory network provides a mechanistic approximation to optimal heterosis in an elite rice hybrid. Plant Journal **112**(1)**:**68-83.

Weckwerth W, Ghatak A, Bellaire A, Chaturvedi P, and Varshney RK (2020) PANOMICS meets germplasm. Plant Biotechnology Journal **18**(7)**:**1507-1525.

Wold H (1975) Soft modelling by latent variables: the nonlinear iterative partial least squares approach. Journal of Applied Probability **12**(S1)**:**117-142.

Xu Y, Zhao Y, Wang X, Ma Y, Li P, Yang Z, Zhang X, Xu C, and Xu S (2021) Incorporation of parental phenotypic data into multi-omic models improves prediction of yield-related traits in hybrid rice. Plant Biotechnology Journal **19**(2)**:**261-272.

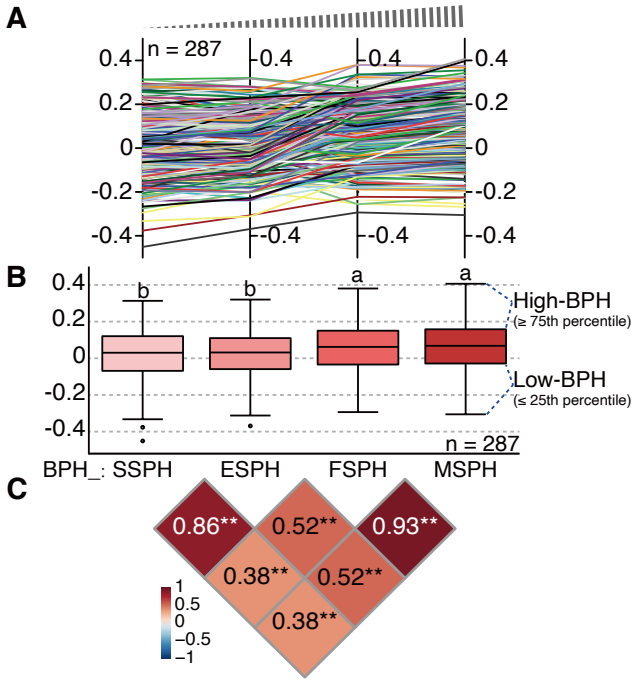
Xue W, Xing Y, Weng X, Zhao Y, Tang W, Wang L, Zhou H, Yu S, Xu C, Li X, and Zhang Q (2008) Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. Nature Genetics **40**(6)**:**761-767.

Zhao Y, Thorwarth P, Jiang Y, Philipp N, Schulthess AW, Gils M, Boeven PHG, Longin CFH, Schacht J, Ebmeyer E, Korzun V, Mirdita V, Dörnte J, Avenhaus U, Horbach R, Cöster H, Holzapfel J, Ramgraber L, Kühnle S, Varenne P, Starke A, Schürmann F, Beier S, Scholz U, Liu F, Schmidt RH, and Reif JC (2021) Unlocking big data doubled the accuracy in predicting the grain yield in hybrid wheat. Science Advances **7**(24)**:**eabf9106.

Zhou P, Hirsch CN, Briggs SP, and Springer NM (2019a) Dynamic patterns of gene expression additivity and regulatory variation throughout maize development. Molecular Plant **12**(3)**:**410-425.

Zhou W, Sailani MR, Contrepois K, Zhou Y, Ahadi S, Leopold SR, Zhang MJ, Rao V, Avina M, Mishra T, Johnson J, Lee-McMullen B, Chen S, Metwally AA, Tran TDB, Nguyen H, Zhou X, Albright B, Hong BY, Petersen L, Bautista E, Hanson B, Chen L, Spakowicz D, Bahmani A, Salins D, Leopold B, Ashland M, Dagan-Rosenfeld O, Rego S, Limcaoco P, Colbert E, Allister C, Perelman D, Craig C, Wei E, Chaib H, Hornburg D, Dunn J, Liang L, Rose SMS, Kukurba K, Piening B, Rost H, Tse D, McLaughlin T, Sodergren E, Weinstock GM, and Snyder M (2019b) Longitudinal multi-omics of host-microbe dynamics in prediabetes. Nature **569**(7758)**:**663-671.

548 ## Figure and figure legends



549

550 **Figure 1** Dynamic heterosis for plant height of the diallel crosses. **(A-B)** Better-parent
551 heterosis for plant height of the F_1 hybrids from the diallel cross population at the
552 individual (**A**) and populational levels (**B**). The F_1 hybrids with better-parent heterosis
553 (BPH) ≤ 25th and ≥ 75th percentiles at each stage were grouped as low- and high-
554 BPH hybrids, respectively. For box plots in (**B**), the middle line is the median. The
555 lower and upper boundaries of the boxes are the first and third quantiles, respectively.
556 The whiskers represent the 1.5 times interquartile range, and the circles represent the
557 outliers. Different letters above the boxes in (**B**) indicate significant differences
558 between groups (one-way ANOVA with the least significant difference test for post
559 hoc multiple comparisons, $P < 0.05$). **(C)** Pearson correlations among height heterosis
560 across the four developmental stages. Abbreviations: SSPH, seedling stage plant
561 height; ESPH, elongation stage plant height; FSPH, flowering stage plant height;
562 MSPH, maturation stage plant height. **, significant at the 0.01 level.

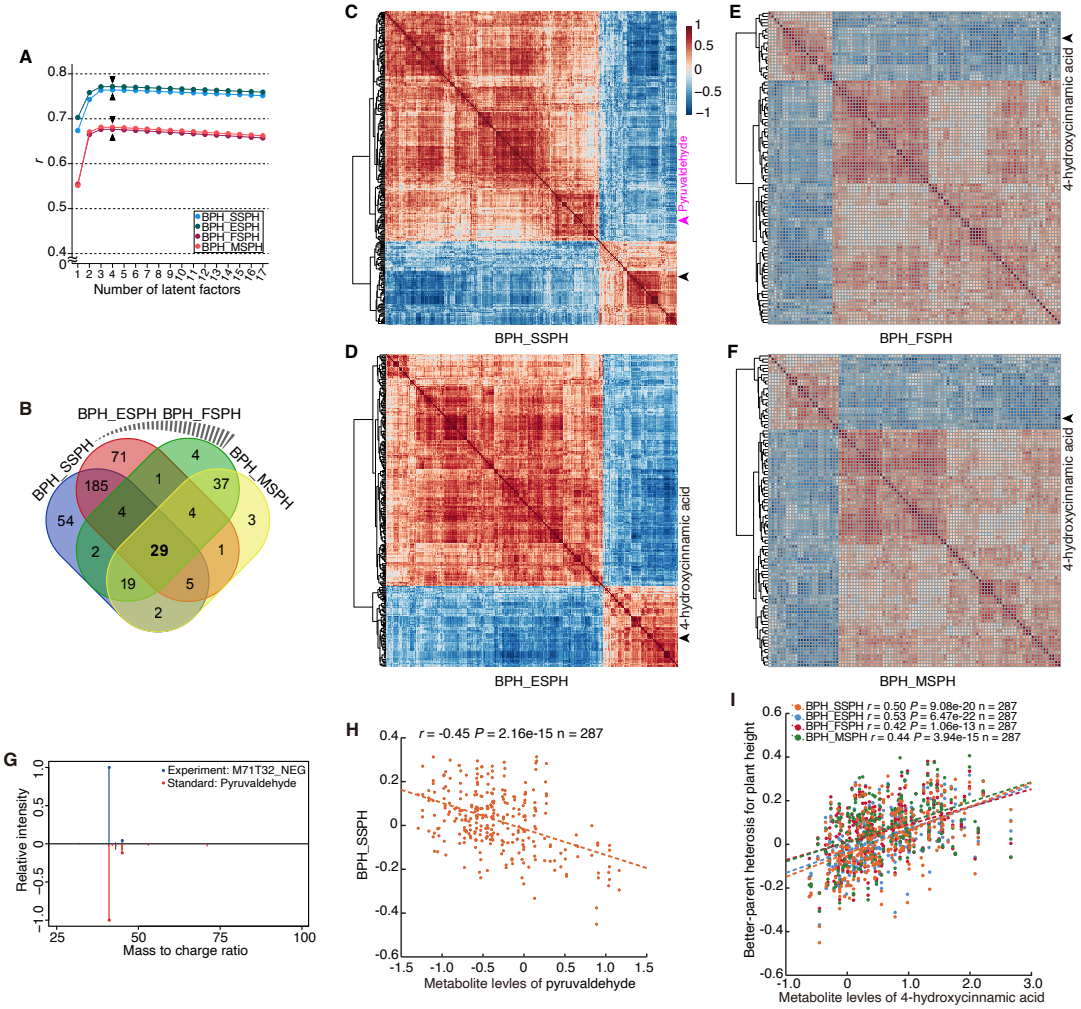


Figure 2 Heterosis-associated analytes for dynamic height heterosis. **(A)** Changes in *r* values with varied number of latent factors. Additive effect was adopted to calculate metabolite levels of F_1 hybrids. Predictive models were built for height heterosis at four stages with the partial least square method. **(B)** Venn diagram of heterosis-associated analytes across four developmental stages. **(C-F)** Heatmaps of correlations among the heterosis-associated analytes for plant height at the seedling (**C**), elongation (**D**), flowering (**E**), and maturation (**F**) stages, respectively. Pearson correlation was performed among analytes. The locations of analytes annotated as 4-hydroxycinnamic acid and pyruvaldehyde are marked with black (at four stages) and purple (only the seedling stage) arrows, respectively. **(G)** MS/MS spectra of M71T32_NEG and the pyruvaldehyde standard. **(H)** Correlations between metabolite levels of pyruvaldehyde and height heterosis at the seedling stage. **(L)** Correlations between metabolite levels of 4-hydroxycinnamic acid and height heterosis at four developmental stages.

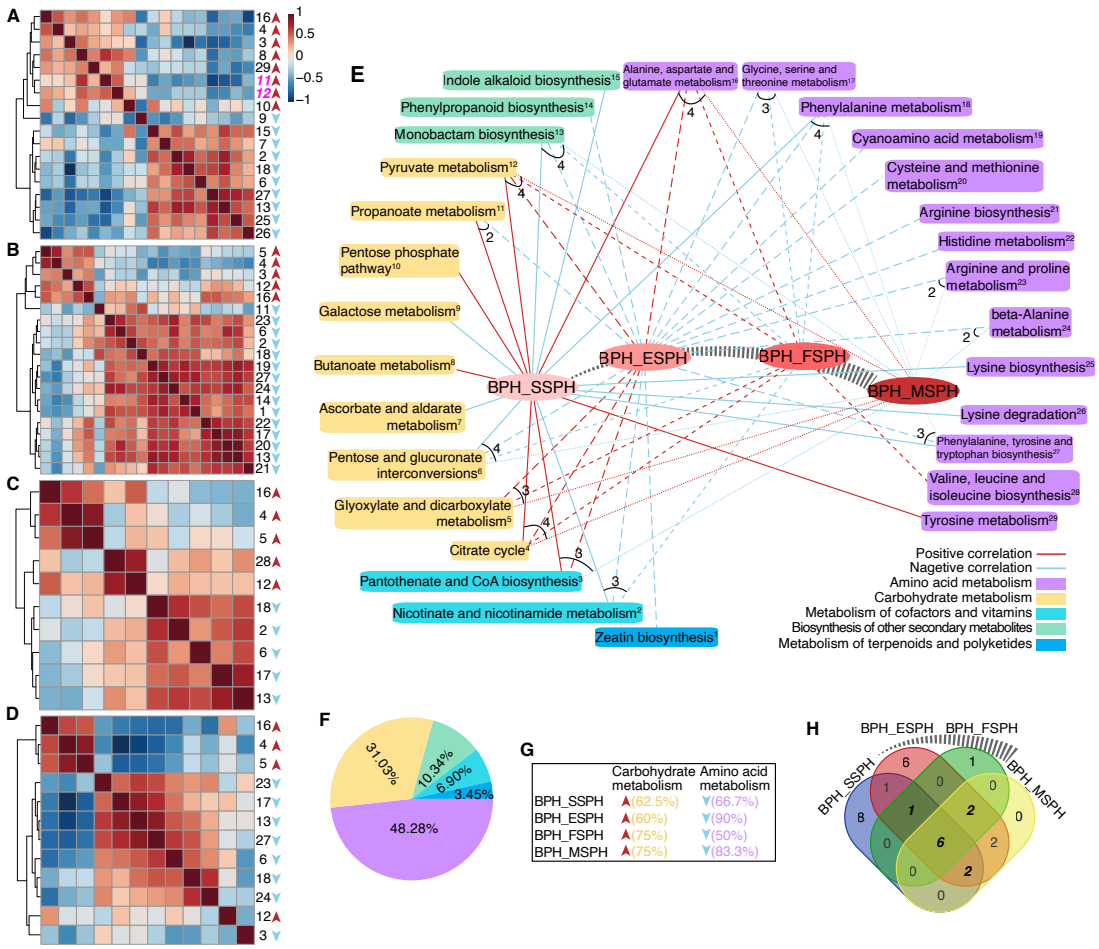


Figure 3 Heterosis-associated pathways underlying dynamic height heterosis. **(A-D)** Heatmaps of correlations among the heterosis-associated pathways for height heterosis at the seedling (**A**), elongation (**B**), flowering (**C**), and maturation (**D**) stages, respectively. The metabolite levels of the pathways for height heterosis were obtained through dysregulated network analysis on the high- and low-BPH hybrids. **(E)** The metabolomic landscape of dynamic height heterosis. Pearson correlations were performed between height heterosis and metabolite levels of the heterosis-associated pathways. **(F)** The pathways were from five groups with different percentages. **(G)** Percentages of the pathways in carbohydrate metabolism that had high metabolite levels in the high-BPH hybrids (upward red arrow) and those of the pathways in amino acid metabolism that had low metabolite levels in the high-BPH hybrids (downward light blue arrow). **(H)** Venn diagram of the heterosis-associated pathways for height heterosis across four developmental stages. The 11 metabolic pathways that were significantly enriched at most stages were among the heterosis-associated pathways for plant height at the elongation stage.

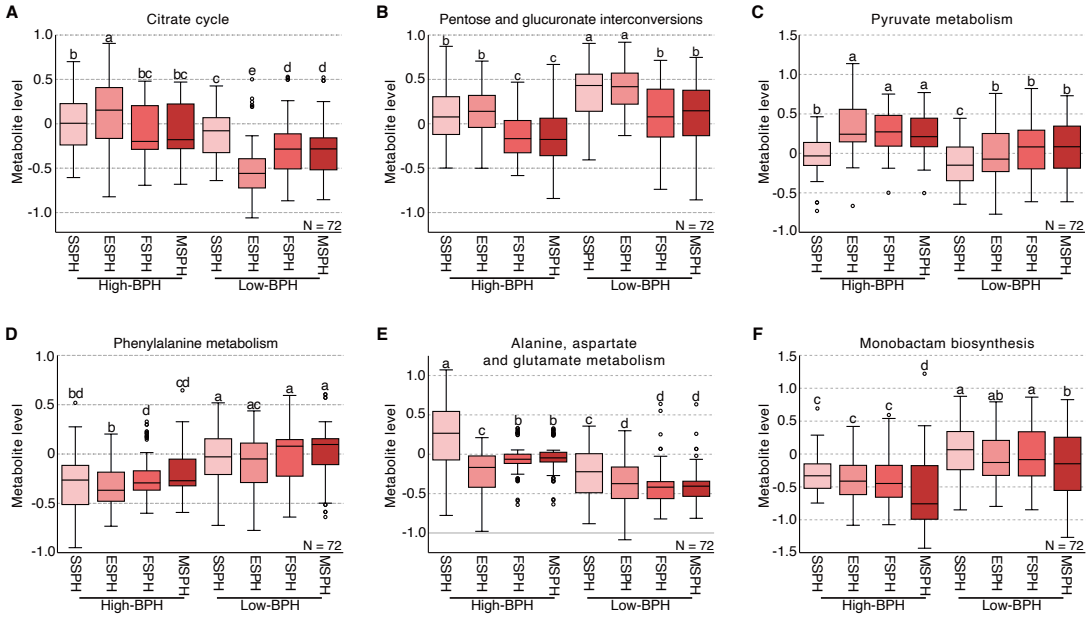


Figure 4 Metabolite levels of six metabolic pathways in the high- and low-BPH hybrids. Metabolite levels of six metabolic pathways, including the citrate cycle (**A**), pentose and glucuronate interconversions (**B**), pyruvate metabolism (**C**), phenylalanine metabolism (**D**), alanine, aspartate and glutamate metabolism (**E**), and monobactam biosynthesis (**F**), were compared to investigate changing patterns across stages. For box plots, the middle line is the median. The lower and upper boundaries of the boxes are the first and third quantiles, respectively. The whiskers represent the 1.5 times interquartile range and the circles represent the outliers. Different letters above the boxes indicate significant differences between groups (one-way ANOVA with the least significant difference test for post hoc multiple comparisons, $P < 0.05$).

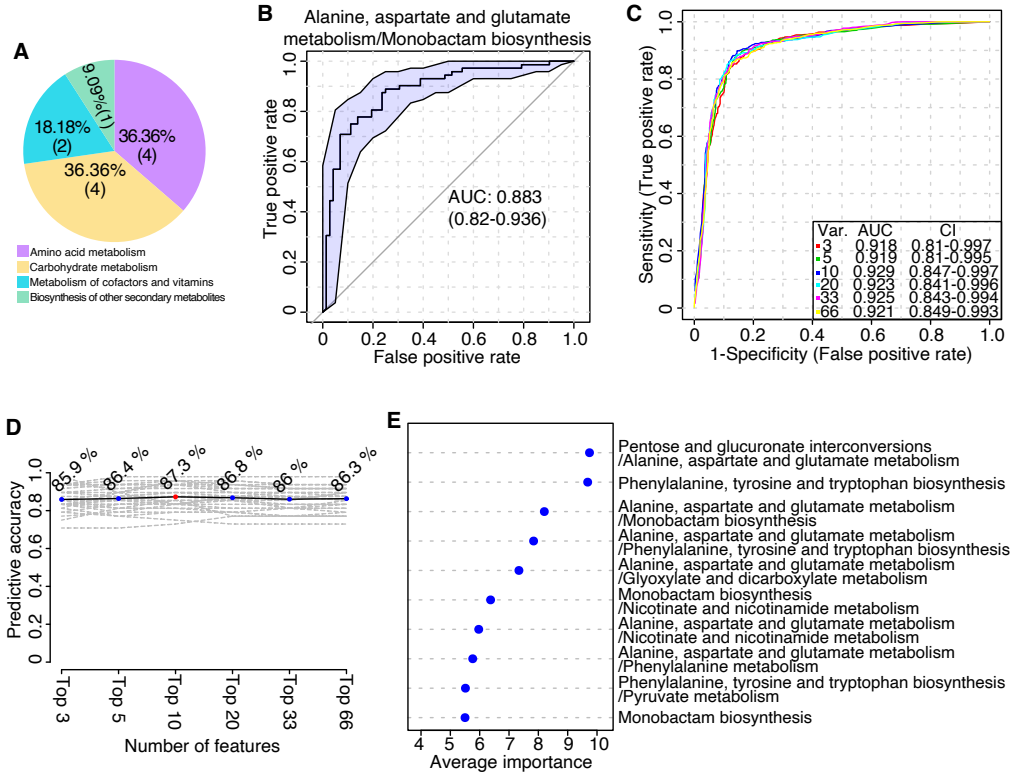


Figure 5 The development of pathway biomarkers for dynamic height heterosis. **(A)** Categories of the 11 metabolic pathways that were significantly enriched at most (three and four) developmental stages. **(B)** Univariate receiver operating characteristic (ROC) curve analysis of height heterosis at the seedling stage of the diallel crosses. The metabolic ratios of alanine, aspartate and glutamate metabolism and monobactam biosynthesis were calculated and used as predictors. AUC, area under the curve. **(C-E)** Multivariate ROC curve analysis of height heterosis at the seedling stage of the diallel crosses. The changes in the AUC and predictive accuracies with different numbers of variables (VAR.) or features are shown in (**C**) and (**D**), respectively. The average importance of the top ten features is listed in (**E**). CI, confidence interval.

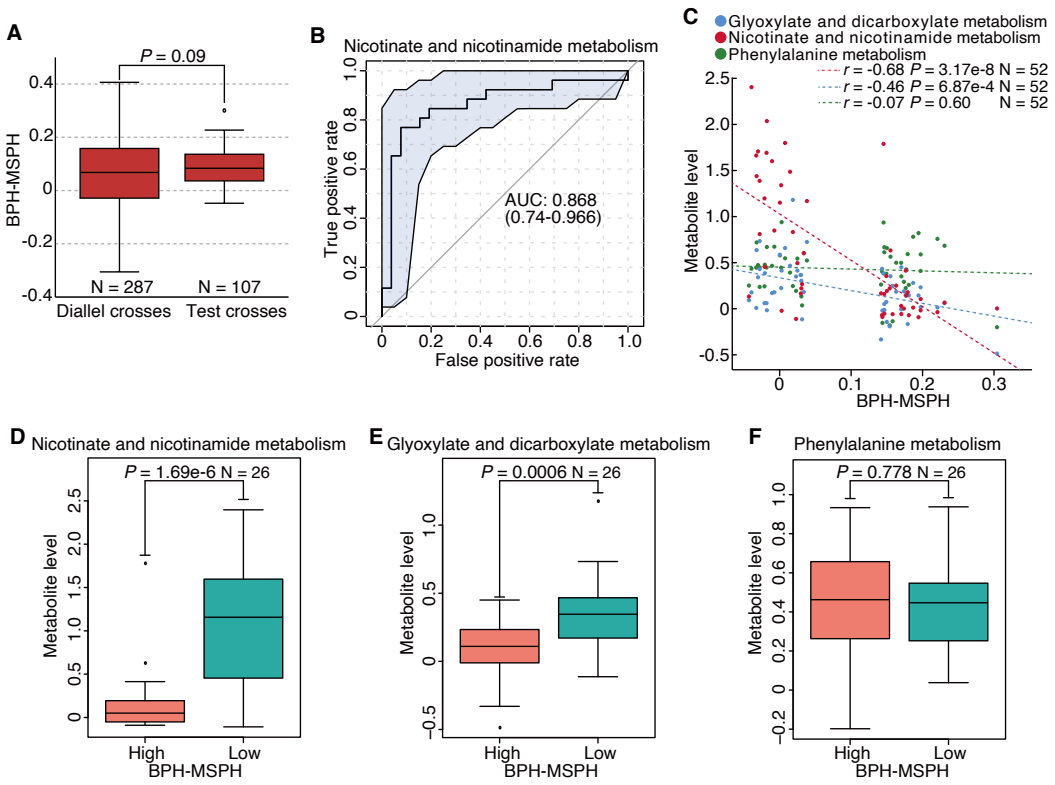


Figure 6 Validation of pathway biomarkers in test crosses. **(A)** Box plots of plant height heterosis at the maturation stage of the diallel and test crosses. **(B)** Univariate ROC curve analysis of height heterosis at the maturation stage of the test crosses. The test crosses were divided into low- and high-BPH hybrids based on the 25th and 75th percentiles, respectively. The metabolite levels of nicotinate and nicotinamide metabolism were identified as predictors. **(C)** Pearson correlations between height heterosis at the maturation stage of the test crosses and metabolite levels of three metabolic pathways. **(D-F)** Comparisons of the metabolite levels of three pathways between the high- and low-BPH hybrids from the test crosses. The metabolite levels of nicotinate and nicotinamide metabolism (**D**), glyoxylate and dicarboxylate metabolism (**E**), and phenylalanine metabolism (**F**) were compared. In (**A**) and (**D-F**), the data were analyzed by two-tailed independent samples *t*-test. For the box plots in (**A**) and (**D-F**), the middle line is the median. The lower and upper boundaries of the boxes are the first and third quantiles, respectively. The whiskers represent the 1.5 times interquartile range, and the circles represent the outliers.