The metabolic pathways underlying dynamic heterosis for plant height are robust biomarkers in rice

Zhiwu Dan[‡], Yunping Chen[‡], Wenchao Huang[*]

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.

[*]Correspondence address: Wenchao Huang, College of Life Sciences, Wuhan University, Wuhan 430072, China. E-mail: wenchaoh@whu.edu.cn

[‡]These authors contributed equally to this study.

Abstract

Robust biomarkers for complex phenotypes enable accurate predictions not only across different populations and environments but also developmental stages because quantitative changes in specific molecules over time are directly connected to phenotypic transformations that determine outcomes of both crop agronomic traits and human diseases. In this study, we conducted a network analysis of untargeted metabolite profiles to investigate dynamic heterosis for rice plant height, a trait that varies during development and contributes to yield heterosis. We found that the levels of pyruvaldehyde were predictive of height heterosis at the seedling stage, while 4-hydroxycinnamic acid showed significant positive correlations with height heterosis across all four developmental stages examined. Our analysis identified metabolic pathways associated with height heterosis, revealing that metabolomic variations at the elongation stage had a more substantial impact than those at other stages. We developed 11 significantly enriched pathways into metabolomic biomarkers with robust predictability for mature hybrids from another population grown under different conditions. This work elucidates the metabolomic landscape of dynamic height heterosis in rice and provides a practical strategy for identifying biomarkers capable of predicting complex phenotypes, irrespective of discrepancies in population, environment, and developmental stage.

Keywords

dynamic heterosis, biomarker, metabolomics, plant height, rice

Introduction

Complex phenotypes are influenced by genotypes, environmental factors, and their interactions [1-3]. The impacts of genetic and environmental factors on phenotypic changes have been extensively studied in both crop agronomic traits and human diseases [4-6]. To accurately represent 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 [7-10]. However, the dynamic nature of the processes that form complex phenotypes is often underestimated [11,12]. Quantitative changes in specific molecules during development have been shown to directly connect with phenotypic transformations, including plant agronomic traits [13-16] and risks for human diseases such as diabetes or cancers [17,18]. Consequently, biomarkers identified from tissues at a single developmental stage (referred to as static biomarkers) may demonstrate low predictive efficiencies for new-stage phenotypes in genetically distant individuals, partially due to molecular heterogeneity across individuals and stages [17,19]. 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 of dynamic biomarkers, offers a viable approach for advancing precision breeding and precision medicine.

Plant height is a critical factor influencing crop grain yield [20]. Although plant height does not directly represent component traits of grain yield, it significantly

65 contributes to yield through its associations with grain number and flowering time.
66 The close relationship between plant height and grain yield in rice has been attributed
67 to multiple polygenic genes, such as *Ghd7* and *DTH7* [21,22]. In a partial correlation
68 analysis comparing the contributions of four yield components and plant height to
69 yield heterosis, heterosis for plant height (R = 0.22) had a higher correlation
70 coefficient than that of grain weight (R = 0.16) in rice [23]. Notably, *Ghd7* plays a
71 central role in the yield heterosis of the elite hybrid rice Shanyou63 [24]. However,
72 plant height is a dynamic trait that varies from young to mature stages, and most
73 current research primarily focuses on plant height at the maturation stage. Given the
74 substantial increase in crop grain yield achieved through heterosis, investigating the
75 molecular mechanisms underlying height heterosis has become essential for further
76 enhancing yield heterosis.
77 In this study, we recorded the height of rice plants from diallel crosses throughout
78 four successive developmental stages: seedling, elongation, flowering, and
79 maturation. We observed a significant change in height heterosis between the
80 elongation and flowering stages. Subsequently, we screened for heterosis-associated
81 analytes at each stage and examined the connections between these metabolites and
82 height heterosis. We identified metabolic pathways involved in dynamic height
83 heterosis and discovered overlapping pathways, primarily in amino acid and
84 carbohydrate metabolism, that revealed metabolomic connections among stages.
85 Finally, we compared the levels of metabolites in significantly enriched pathways,
86 observing dramatic metabolic changes at the elongation stage. Pathway biomarkers

were developed and subsequently validated with test crosses. Our findings reveal the metabolomic landscape of dynamic height heterosis and expand the application potential of pathway biomarkers in predicting complex phenotypes across developmental stages.

Results

Dynamic heterosis for plant height in rice diallel crosses

In this study, we combined untargeted metabolite profiles with better-parent heterosis (BPH) for plant height across four growth stages to explore dynamic biomarkers for the yield heterosis-related trait in rice (Fig. 1A and Supplementary Table S1). The plant height heterosis observed in a hybrid population designed as a complete diallel cross showed that the degree of heterosis was significantly greater at the later stages compared to the earlier stages (Fig. 1B). Additionally, the correlation coefficients of heterosis among stages decreased as the stage interval increased (Fig. 1C).

Metabolic analytes and pathways involved in dynamic height heterosis

To identify the metabolic pathways associated with changes in height heterosis, we firstly obtained metabolite profiles from young seedlings of the parents. Since additive effect is the predominant inheritance pattern for F_1 hybrids across developmental stages [16,25-27], we averaged the parental metabolite levels to represent the hybrid metabolome as previously described [28,29]. Subsequently, we divided the whole hybrid population into low- and high-BPH groups according to the

25th and 75th percentiles of height heterosis at each stage, as shown in Fig. 1B. We conducted dysregulated network analysis [30], which involves pathway enrichment analysis using a hypergeometric test on the differential metabolic peaks, to identify heterosis-associated metabolic pathways. Additionally, we employed a partial least square-based variable selection strategy to screen for heterosis-associated analytes relevant to plant height heterosis at the four developmental stages [23,31].

Similar correlation patterns were observed between the heatmaps of significantly enriched pathways and heterosis-associated analytes (Fig. 2A-H). Notably, more stage-specific heterosis-associated analytes were identified at the earlier stages compared to the later ones (Fig. 2I). Importantly, 4-hydroxycinnamic acid, which has been proven to be linked to plant height [23,32], was consistently identified as a heterosis-associated analyte across all four stages (Fig. 2B, D, F, and H). The levels of 4-hydroxycinnamic acid demonstrated 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). Additionally, the accurately annotated metabolite pyruvaldehyde (peak tag: M71T32_NEG; Fig. 2K), which is involved in pyruvate metabolism and propanoate metabolism (Fig. 2A-B), was identified as a heterosis-associated analyte for height heterosis at the seedling stage, showing significant negative correlations with plant height heterosis (Fig. 2L).

Metabolic pathways underlying dynamic height heterosis

The majority of the heterosis-associated metabolic pathways were found in amino

128 acid (48.28%) and carbohydrate (31.03%) metabolism (Fig. 3A). Specifically, three,
129 five, and six metabolic pathways were significantly enriched at two, three, and all four
130 stages, respectively. With respect to the metabolite levels of these significantly
131 enriched pathways, namely, the average levels of all metabolites in each pathway
132 [30], the pathways associated with carbohydrate metabolism (60-75%) consistently
133 exhibited high levels in high-BPH hybrids, while those related to amino acid
134 metabolism (50-90%) displayed lower levels (Fig 3B). This pattern complements the
135 changing trends observed for significantly enriched metabolic pathways related to
136 heterosis of yield and its components [23,28].

137 In alignment with the phenotypic differences and variations in the number of
138 heterosis-associated analytes (Fig. 1A-C and Fig. 2I), the number of metabolic
139 pathways showing stage-specific enrichment (8, 6, 1, and 0 pathways) indicated
140 substantial metabolomic changes between the earlier and later stages of height
141 heterosis (Fig. 3C). Thus, the metabolic pathways implicated in height heterosis
142 across the four investigated stages were identified, with amino acid and carbohydrate
143 metabolism serving as the primary pathways underlying these variations.

144 Metabolomic changes at the elongation stage play critical roles in height 145 heterosis transformation

146 According to our metabolomic landscape analysis, six pathways were significantly
147 enriched across all four stages. Three of these pathways (citrate cycle, pentose and
148 glucuronate interconversions, and pyruvate metabolism) are classified under

carbohydrate metabolism, while two pathways are from amino acid metabolism (phenylalanine metabolism and alanine, aspartate and glutamate metabolism). The remaining pathway pertains to monobactam biosynthesis. We compared the metabolite levels of these six pathways among stages to investigate variations in the metabolomic landscape related to dynamic height heterosis. The metabolite levels of the six pathways exhibited differences in either the high- or low-BPH hybrids across the four developmental stages. Nearly all pathways demonstrated significant differences in metabolite levels in at least one stage when compared to the others. The metabolite level of the citrate cycle initially increased at the elongation stage before decreasing in the later stages for high-BPH hybrids (Fig. 4A), while it showed the opposite trend in low-BPH hybrids. Similar changing patterns were observed in the other two pathways from carbohydrate metabolism (Fig. 4B-C). For the two pathways from amino acid metabolism and monobactam biosynthesis, fewer changes in metabolite levels were noted in low-BPH hybrids compared to high-BPH hybrids (Fig. 4D-F). Notably, significant changes in metabolite levels were frequently observed at the elongation stage for both high- and low-BPH hybrids, which demonstrated a critical role in the metabolomic changes during elongation to the transformation of height heterosis.

Pathway biomarkers for predicting dynamic height heterosis across stages

To identify biomarkers for the dynamic height heterosis, we conducted both univariate

170 and multivariate receiver operating characteristic (ROC) curve analyses on the 18
171 significantly enriched metabolic pathways for height heterosis at the seedling stage
172 (Supplementary Fig. S1). The area under the curve (AUC) was 0.885 with a single
173 predictive feature, increasing to 0.933 when utilizing one hundred features. The
174 analysis of the six pathways that were enriched at all four stages revealed no
175 substantial increase in AUC values or predictive accuracies (uni-ROC: AUC = 0.887,
176 multi-ROC: AUC = 0.931, predictive accuracy = 0.882 with 21 features;
177 Supplementary Fig. S1). We then explored the 11 metabolic pathways that were
178 significantly enriched at three or four stages, which were primarily pathways in amino
179 acid and carbohydrate metabolism, as predictive features (Fig. 5A). The AUC was
180 0.883 with a single predictor (Fig. 5B), while the best predictive model, which
181 included only 10 features, achieved an AUC of 0.929 and a predictive accuracy of
182 0.873 (Fig. 5C-D). Unexpectedly, despite being significantly enriched across all four
183 stages, the citrate cycle did not rank with high importance among the top ten features
184 (Fig. 5E). Furthermore, biomarker analysis of the 11 selected metabolic pathways
185 demonstrated high predictive accuracies for height heterosis at the three later
186 developmental stages: elongation, flowering, and maturation (Supplementary Fig. S2).
187 These results suggested that the development of biomarkers from the 11 metabolic
188 pathways significantly enriched across most developmental stages could be an
189 effective strategy for identifying robust biomarkers, taking into account predictive
190 accuracy, feature number, and feature ranking.

Validation of pathway biomarkers with test crosses

To further validate the predictive capability of the pathway biomarkers, we utilized a group of Honglian-type test crosses, characterized by distant genetic relationships and cultivated under different field conditions compared to the diallel crosses [28]. The height heterosis at the maturation stage of the two hybrid populations was comparable, although the test crosses exhibited a narrower range of height heterosis due to the limited genetic relatedness of their male parents, which were mainly recombinant inbred lines (Fig. 6A). After categorizing the test crosses into low- and high-BPH groups based on the quartiles of height heterosis (i.e., 25th and 75th percentiles), we obtained metabolite levels for four pathways through dysregulated network analysis (Supplementary Table S2), three of which belonged to the previously identified 11 pathways. Using the metabolite levels of nicotinate and nicotinamide metabolism as a single predictor yielded an AUC of 0.868 (Fig. 6B), while AUC values ranging from 0.8 to 0.834 were achieved with 2-6 variables (Supplementary Fig. S3). Negative correlations were observed between height heterosis and the metabolite levels of nicotinate and nicotinamide metabolism, as well as glyoxylate and dicarboxylate metabolism (Fig. 6C), whereas no correlation was found with phenylalanine metabolism. Additionally, the metabolite levels of nicotinate and nicotinamide metabolism were significantly lower in the high-BPH hybrids compared to low-BPH hybrids (Fig. 6D). Conversely, the levels of glyoxylate and dicarboxylate metabolism were also lower in the high-BPH hybrids (Fig. 6E),

which contrasted with the result shown in Fig. 3A. No significant change in phenylalanine metabolism was detected (Fig. 6F). Furthermore, higher AUC values were observed when grouping the low- and high-BPH hybrids at a more stringent level, while lower values were obtained when dividing the test crosses into approximately equal halves (Supplementary Figs S4-S5). Our results indicated that the 11 identified metabolic pathways, which were significantly enriched at most developmental stages in the diallel crosses, effectively predicted height heterosis at the maturation stage in the test crosses.

Discussion

In this study, we identified robust biomarkers for dynamic height heterosis based on metabolic pathways significantly enriched at most developmental stages in rice. In alignment with the phenotypic transformations observed in the field, metabolomic changes at the elongation stage were found to contribute more substantially to height heterosis than those at the other stages. Recognizing this critical stage through a time-series analysis may inspire the identification of biomarkers for predicting biomass or height in the breeding programs of perennial energy crops and trees [33,34].

Among the 11 selected metabolic pathways, all of which were significantly enriched at the elongation stage, pathways related to carbohydrate and amino acid metabolism comprised the majority and exhibited approximately opposite trends in metabolite levels. Given the tight connection between plant height and grain yield in rice [21,23,35], variations in these two types of pathways may determine the optimal balance between vegetative and reproductive traits in the formation of heterosis in

234 hybrid rice [23,36,37]. Furthermore, the strategy of incorporating pathways that are
235 significantly enriched at most developmental stages—rather than exclusively those
236 enriched at all stages—could enhance the robustness of pathway biomarkers for new
237 predictions across different populations, environments, and stages. We hypothesize
238 that this advantage arises from the comprehensive information provided by predictive
239 features.

240 While we have identified numerous metabolic pathways associated with dynamic
241 height heterosis, their specific contributions to height heterosis require in-depth
242 investigation. Simultaneous mutations in multiple key genes from relevant metabolic
243 pathways may help elucidate the detailed functions of each heterosis-associated
244 pathway. Notably, the number of heterosis-associated pathways and analytes at the
245 earlier stages was greater than at the later stages, suggesting that the metabolomic
246 basis underlying height heterosis during early developmental stages may be more
247 complex and significant than that at later stages. Moreover, the information on
248 enriched metabolic pathways in our study is largely dependent on the completeness of
249 the KEGG database. Other metabolic pathways not included in the database may also
250 contribute to heterosis for plant height. The addition of new heterosis-associated
251 pathways will further refine the metabolomic landscape of dynamic heterosis and
252 enhance the predictive accuracy.

253 Although some pathway variation patterns were consistent across hybrid
254 populations, we also observed discrepancies in metabolite levels of certain pathways.
255 The metabolomic heterogeneity between populations [23], similar to gene expression-

level variability among individual patients with the same phenotype [19,38], may be a primary factor driving these inconsistencies. Given the variations in predictive accuracies, further investigation is warranted for hybrids exhibiting intermediate heterosis. Comprehensive metabolomics data that include metabolite profiles obtained via both the positive and negative ionization modes, as well as through both GC and LC methods, are essential. The integration of multiple layers of omics information [39,40], including genomics, transcriptomics, proteomics, metabolomics, and microbiome data, may contribute to the optimization of pathway biomarkers for predicting complex phenotypes across populations (individuals), environments, and developmental stages.

Methods

Plant materials

The rice accessions selected for hybridization were previously described [28]. 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 days after

transplantation), flowering stage (time of flowering), and maturation stage (time of maturation). The plant height data for the maturation stage of the diallel cross population were reported in a previous study [23], and the complete plant height data are provided in Supplementary table 1. For comparative purpose, several analysis results related to better-parent heterosis for plant height at the maturation stage were retrieved from earlier research [23]. The plant height of the test cross population was only measured at the maturation stage. The two hybrid populations were grown under different growth conditions and planting densities at the same location (Ezhou, Hubei Province) in 2012 and 2015. Three biological replicates were designed for each genotype, with five plants from the middle region of each row selected for height measurement.

Metabolite profiling

Untargeted metabolite profiles in the negative ionization mode of 15-d-old parental seedlings were obtained using previously described methods [28]. In brief, a 1290 Infinity liquid chromatography system (Agilent Technologies), an Agilent quadrupole time-of-flight mass spectrometer (Agilent 6550 iFunnel QTOF), and a Triple TOF 6600 mass spectrometer (AB SCIEX) were employed to acquire the raw metabolomic profiles. The stability of the equipment and the accuracy of the metabolite profiles were assessed using intermittently injected quality control samples. A total of 3,746 analytes were detected, and 114 metabolites were identified through matching with in-house spectral libraries of standards by Shanghai Applied Protein Technology Co.,

298 Ltd. To prepare the metabolomic data for statistical analyses, sample normalization
299 (normalization by sum), data transformation (none), and data scaling (auto scaling)
300 were performed using MetaboAnalyst 4.0 [41]. The hybrid metabolome used for the
301 analyses was calculated by averaging the metabolite levels of the corresponding
302 parents according to previously reported methods [28,29].

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

304 The Metabolite identification and Dysregulated Network Analysis web server
305 (MetDNA version 1.2.1) [30], which annotates metabolites based on the rule that
306 neighboring metabolites tend to share similar MS2 spectra, was used to perform
307 pathway enrichment analysis for height heterosis. The hybrids were categorized into
308 high-better-parent heterosis (control group) and low-better-parent heterosis (case
309 group) groups according to quantiles (≥ 75th percentile and ≤ 25th percentile) of the
310 entire hybrid population, and the calculated hybrid metabolome was utilized for
311 dysregulated network analysis. Analysis parameters were set according to the
312 metabolite profiling experiments, with *Arabidopsis thaliana* (thale cress) selected as
313 the specie. Student's *t*-test was employed for univariate statistics, with a cutoff *P*-
314 value of 0.05 (with adjustment). Enriched metabolic pathways and quantitative
315 information on metabolic pathways, which were included in the analysis report, were
316 combined to identify significantly enriched metabolic pathways for height heterosis at
317 each stage. The metabolic pathways that were significantly enriched (*P* values less
318 than 0.05) and exhibited significant differences in metabolite levels between the high-

319 and low-better-parent heterosis hybrids were treated as height heterosis-associated
320 metabolic pathways.

321 For the identification of heterosis-associated analytes, we employed a partial least
322 square-based strategy [23,31]. Initially, the number of latent factors in the models was
323 determined based on the highest achieved *r* values. The performance of the predictive
324 models was assessed using 10-fold cross-validation and permutation tests conducted
325 with MetaboAnalyst 4.0 [41]. Subsequently, average values of variable importance in
326 the projection for all latent factors were calculated to evaluate the contributions of
327 specific analytes to plant height heterosis. Redundant analytes in the predictive
328 models were removed by reducing the numbers of predictive analytes. The final
329 number of predictive analytes for plant height heterosis at the seedling, elongation,
330 flowering, and maturation stages were 300, 300, 100, and 100, respectively.

331 **Statistical analyses**

332 Better-parent heterosis for plant height was calculated using the equation: BPH = (F_1-
333 P_H)/P_H, where F_1 represents the plant height of the hybrid and P_H denotes the height of
334 the taller parent. The hybrid metabolome was calculated as follows: M_H = (M_F +
335 M_M)/2, where M_H, M_F, and M_M indicate the metabolite levels of the F_1 hybrid, female
336 parent, and male parent, respectively. Pearson correlation analysis of height heterosis
337 across the four developmental stages was performed using the "Statistical Analysis"
338 module in MetaboAnalyst 4.0 [41]. Comparisons of height heterosis and metabolite
339 levels of significantly enriched pathways across stages were analyzed using one-way

340 ANOVA (with least significant difference for post hoc multiple comparisons, $P <$
341 0.05) via SPSS (IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM
342 Corp.). Analyses comparing plant height heterosis between the diallel and test crosses,
343 as well as metabolite levels between high- and low-BPH hybrids, were conducted
344 using independent samples *t*-test in SPSS (two-tailed). Pearson correlation analysis
345 between height heterosis and the metabolite levels of metabolic pathways was also
346 performed using SPSS. Both univariate and multivariate ROC curve analyses were
347 conducted using the "Biomarker Analysis" module in MetaboAnalyst 4.0 [41]. The
348 metabolic ratios of the top 100 were computed and included in the subsequent
349 analysis. Random forest analysis [42] was selected as the classification and feature
350 ranking method for the multivariate ROC curve analysis.

351 Additional Files

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

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

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

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

361 **Supplementary Fig. S3.** Multivariate ROC curve analysis of three metabolic

pathways for height heterosis of the test crosses.

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

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

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.

Fundings

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.

Data availability

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

[28].

Competing Interests

The authors declare no competing interest.

References

1. Hunter DJ. Gene-environment interactions in human diseases. Nature Reviews Genetics 2005;6(4):287-298. https://doi.org/10.1038/nrg1578.
2. Scheres B, & van der Putten WH. The plant perceptron connects environment to development. Nature 2017;543(7645):337-345. https://doi.org/10.1038/nature22010.
3. Crossa J, Perez-Rodriguez P, Cuevas J, et al. Genomic selection in plant breeding: methods, models, and perspectives. Trends in Plant Science 2017;22(11):961-975. https://doi.org/10.1016/j.tplants.2017.08.011.
4. Li X, Guo T, Mu Q, et al. Genomic and environmental determinants and their interplay underlying phenotypic plasticity. Proceedings of the National Academy of Sciences of the United States of America 2018;115(26):6679-6684. https://doi.org/10.1073/pnas.1718326115.
5. Fave MJ, Lamaze FC, Soave D, et al. Gene-by-environment interactions in urban populations modulate risk phenotypes. Nature Communications 2018;9(1):827. https://doi.org/10.1038/s41467-018-03202-2.
6. de Los Campos G, Perez-Rodriguez P, Bogard M, et al. A data-driven simulation platform to predict cultivars' performances under uncertain weather conditions. Nature Communications 2020;11(1):4876. https://doi.org/10.1038/s41467-020-18480-y.
7. Sarkizova S, Klaeger S, Le PM, et al. A large peptidome dataset improves HLA class I epitope prediction across most of the human population. Nature Biotechnology 2020;38(2):199-209. https://doi.org/10.1038/s41587-019-0322-9.
8. Tong H, Kuken A, & Nikoloski Z. Integrating molecular markers into metabolic models improves genomic selection for Arabidopsis growth. Nature Communications 2020;11(1):2410. https://doi.org/10.1038/s41467-020-16279-5.
9. Riedelsheimer C, Czedik-Eysenberg A, Grieder C, et al. Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nature Genetics 2012;44(2):217-220. https://doi.org/10.1038/ng.1033.
10. Zhao Y, Thorwarth P, Jiang Y, et al. Unlocking big data doubled the accuracy in predicting the

grain yield in hybrid wheat. Science Advances 2021;7(24):eabf9106. https://doi.org/10.1126/sciadv.abf9106.

11. Lesterhuis WJ, Bosco A, Millward MJ, et al. Dynamic versus static biomarkers in cancer immune checkpoint blockade: unravelling complexity. Nature Reviews Drug Discovery 2017;16(4):264-272. https://doi.org/10.1038/nrd.2016.233.

12. Miao C, Xu Y, Liu S, et al. Increased power and accuracy of causal locus identification in time series genome-wide association in sorghum. Plant Physiology 2020;183(4):1898-1909. https://doi.org/10.1104/pp.20.00277.

13. Varoquaux N, Cole B, Gao C, et al. 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 2019;116(52):27124-27132. https://doi.org/10.1073/pnas.1907500116.

14. Ko DK, Rohozinski D, Song Q, et al. Temporal shift of circadian-mediated gene expression and carbon fixation contributes to biomass heterosis in maize hybrids. PLoS Genetics 2016;12(7):e1006197. https://doi.org/10.1371/journal.pgen.1006197.

15. Obermeyer G, Fragner L, Lang V, et al. Dynamic adaption of metabolic pathways during germination and growth of lily pollen tubes after inhibition of the electron transport chain. Plant Physiology 2013;162(4):1822-1833. https://doi.org/10.1104/pp.113.219857.

16. Zhou P, Hirsch CN, Briggs SP, et al. Dynamic patterns of gene expression additivity and regulatory variation throughout maize development. Molecular Plant 2019;12(3):410-425. https://doi.org/10.1016/j.molp.2018.12.015.

17. Price ND, Magis AT, Earls JC, et al. A wellness study of 108 individuals using personal, dense, dynamic data clouds. Nature Biotechnology 2017;35(8):747-756. https://doi.org/10.1038/nbt.3870.

18. Taylor IW, Linding R, Warde-Farley D, et al. Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nature Biotechnology 2009;27(2):199-204. https://doi.org/10.1038/nbt.1522.

19. Menche J, Guney E, Sharma A, et al. Integrating personalized gene expression profiles into predictive disease-associated gene pools. npj Systems Biology and Applications 2017;3(10. https://doi.org/10.1038/s41540-017-0009-0.

20. Wang W, Guo W, Le L, et al. Integration of high-throughput phenotyping, GWAS, and predictive models reveals the genetic architecture of plant height in maize. Molecular Plant 2023;16(2):354-373. https://doi.org/10.1016/j.molp.2022.11.016.

21. Xue W, Xing Y, Weng X, et al. Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. Nature Genetics 2008;40(6):761-767. https://doi.org/10.1038/ng.143.

22. Gao H, Jin M, Zheng XM, et al. 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 2014;https://doi.org/10.1073/pnas.1418204111.

23. Dan Z, Chen Y, Li H, et al. The metabolomic landscape of rice heterosis highlights pathway biomarkers for predicting complex phenotypes. Plant Physiology 2021;187(2):1011-1025. https://doi.org/10.1093/plphys/kiab273.

24. Wang X, Zhou T, Li G, et al. A Ghd7-centered regulatory network provides a mechanistic approximation to optimal heterosis in an elite rice hybrid. Plant Journal 2022;112(1):68-83. https://doi.org/10.1111/tpj.15928.

25. Dan Z, Chen Y, Zhou W, et al. Additive and partially dominant effects from genomic variation contribute to rice heterosis. bioRxiv 2024;10.1101/2024.1107.1116.603817. https://doi.org/10.1101/2024.07.16.603817.

26. Liu W, He G, & Deng XW. Biological pathway expression complementation contributes to biomass heterosis in Arabidopsis. Proc Natl Acad Sci U S A 2021;118(16):https://doi.org/10.1073/pnas.2023278118.

27. Dan Z, Chen Y, & Huang W. Structural variations contribute to genetic diversity and heterosis in rice. bioRxiv 2024;10.1101/2024.1111.1103.621459. https://doi.org/10.1101/2024.11.03.621459.

28. Dan Z, Chen Y, Zhao W, et al. Metabolome-based prediction of yield heterosis contributes to the breeding of elite rice. Life Science Alliance 2020;3(1):e201900551. https://doi.org/10.26508/lsa.201900551.

29. Dan Z, Chen Y, Xu Y, et al. A metabolome-based core hybridisation strategy for the prediction of rice grain weight across environments. Plant Biotechnology Journal 2019;17(5):906-913. https://doi.org/10.1111/pbi.13024.

30. Shen X, Wang R, Xiong X, et al. Metabolic reaction network-based recursive metabolite annotation for untargeted metabolomics. Nature Communications 2019;10(1):1516. https://doi.org/10.1038/s41467-019-09550-x.

31. Wold H. Soft modelling by latent variables: the nonlinear iterative partial least squares approach. Journal of Applied Probability 1975;12(S1):117-142. https://doi.org/10.1017/s0021900200047604.

32. Li Y, Kim JI, Pysh L, et al. Four isoforms of arabidopsis 4-coumarate:CoA ligase have

overlapping yet distinct roles in phenylpropanoid metabolism. Plant Physiology 2015;169(4):2409-2421. https://doi.org/10.1104/pp.15.00838.

33. Maddison AL, Camargo-Rodriguez A, Scott IM, et al. Predicting future biomass yield in *Miscanthus* using the carbohydrate metabolic profile as a biomarker. GCB Bioenergy 2017;9(1264-1278. https://doi.org/10.1111/gcbb.12418.

34. Canas RA, Canales J, Munoz-Hernandez C, et al. Understanding developmental and adaptive cues in pine through metabolite profiling and co-expression network analysis. Journal of Experimental Botany 2015;66(11):3113-3127. https://doi.org/10.1093/jxb/erv118.

35. Chen Y, Dan Z, Gao F, et al. Rice *GROWTH-REGULATING FACTOR7* modulates plant architecture through regulating GA and indole-3-acetic acid metabolism. Plant Physiology 2020;184(1):393-406. https://doi.org/10.1104/pp.20.00302.

36. Birchler JA. Genomic balance plays out in evolution. Plant Cell 2019;31(6):1186-1187. https://doi.org/10.1105/tpc.19.00329.

37. Li Z, Zhu A, Song Q, et al. Temporal regulation of the metabolome and proteome in photosynthetic and photorespiratory pathways contributes to maize heterosis. Plant Cell 2020;32(12):3706-3722. https://doi.org/10.1105/tpc.20.00320.

38. Zhou W, Sailani MR, Contrepois K, et al. Longitudinal multi-omics of host-microbe dynamics in prediabetes. Nature 2019;569(7758):663-671. https://doi.org/10.1038/s41586-019-1236-x.

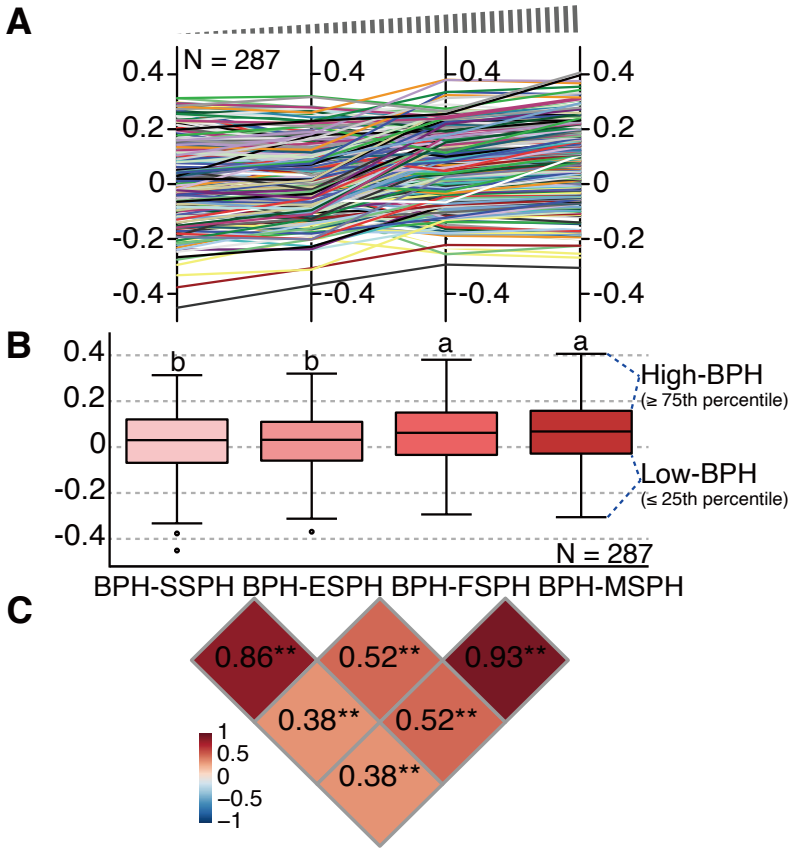
39. Weckwerth W, Ghatak A, Bellaire A, et al. PANOMICS meets germplasm. Plant Biotechnology Journal 2020;18(7):1507-1525. https://doi.org/10.1111/pbi.13372.

40. Xu Y, Zhao Y, Wang X, et al. Incorporation of parental phenotypic data into multi-omic models improves prediction of yield-related traits in hybrid rice. Plant Biotechnology Journal 2021;19(2):261-272. https://doi.org/10.1111/pbi.13458.

41. Chong J, Soufan O, Li C, et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids Research 2018;46(W1):W486-W494. https://doi.org/10.1093/nar/gky310.

42. Breiman L. Random forests. Machine Learning 2001;45(5-32. https://doi.org/10.1023/A:1010933404324.

508 ## Figure and figure legends



509

510 **Figure 1:** Dynamic heterosis for plant height of the diallel crosses. (A-B) Dynamic
511 better-parent heterosis for plant height of the hybrids from the diallel cross population
512 at the individual (A) and populational levels (B). The hybrids with better-parent
513 heterosis (BPH) ≤ 25th and ≥ 75th percentiles at four developmental stages were
514 grouped as low- and high-BPH hybrids, respectively. For box plots in (B), the middle
515 line is the median. The lower and upper boundaries of the boxes are the first and third
516 quantiles, respectively. The whiskers represent the 1.5 times interquartile range, and
517 the circles represent the outliers. Different letters above the boxes in (B) indicate
518 significant differences between groups (one-way ANOVA with the least significant
519 difference test for post hoc multiple comparisons, $P < 0.05$). (C) Pearson correlations
520 among height heterosis across the four developmental stages. Abbreviations: SSPH,
521 seedling stage plant height; ESPH, elongation stage plant height; FSPH, flowering
522 stage plant height; MSPH, maturation stage plant height. **, significant at the 0.01
523 level.

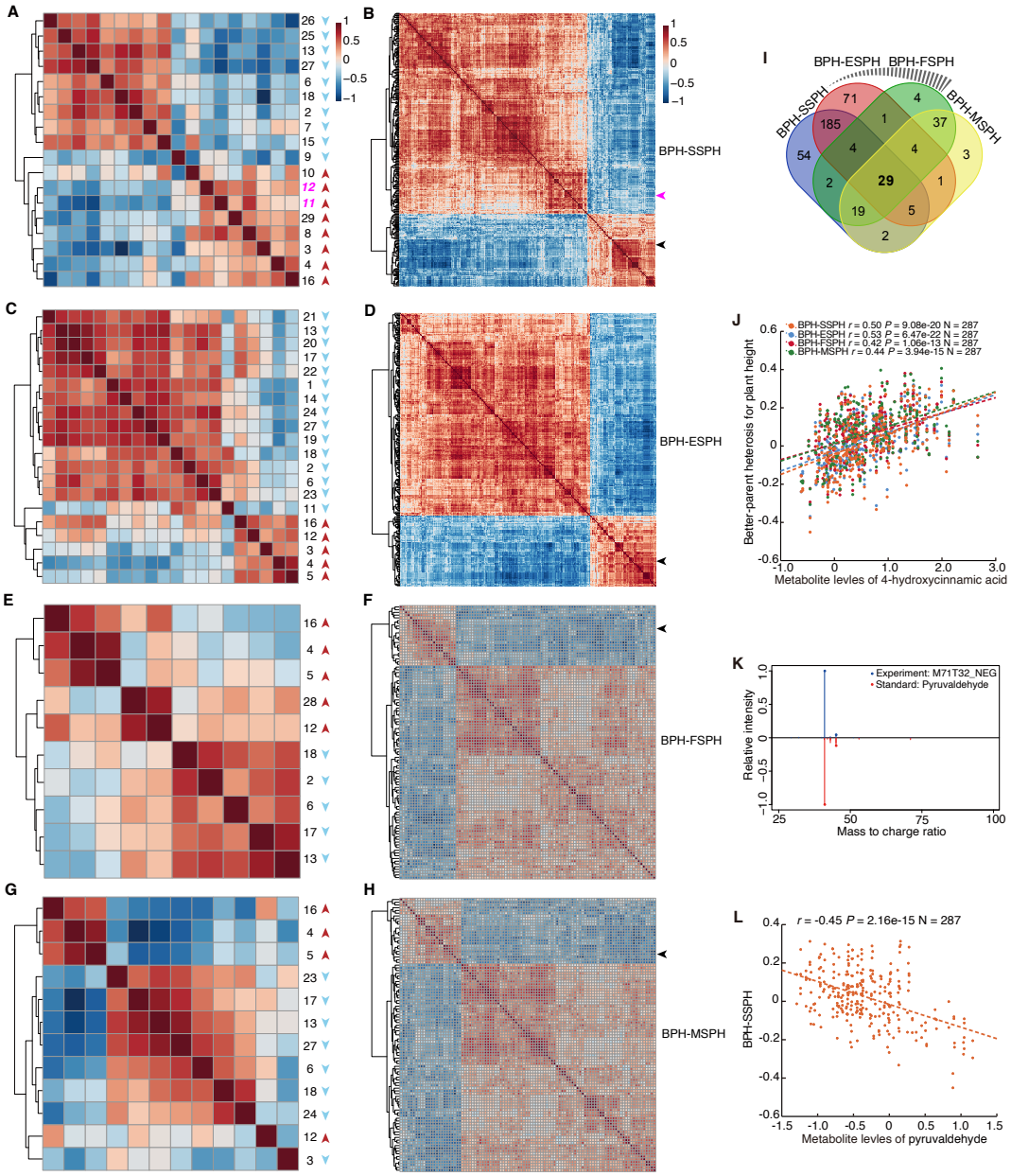


Figure 2: Heterosis-associated metabolic pathways and analytes for dynamic plant height. (A-H) Heatmaps of correlations among the significantly enriched metabolic pathways (A, C, E, and G) and heterosis-associated analytes (B, D, F, and H) for plant height heterosis at the seedling (A and B), elongation (C and D), flowering (E and F), and maturation (G and H) stages, respectively. The metabolite levels of the enriched metabolic pathways for height heterosis were obtained through dysregulated network analysis of the hybrid metabolome between the high- and low-BPH hybrids that was calculated by averaging the corresponding parental metabolite levels. Independent samples *t*-tests were conducted to compare the metabolite levels between the low- and high-BPH hybrids and thus identify the significantly enriched metabolic pathways

that were treated as height heterosis-associated pathways. The average values of metabolites per pathway were calculated as metabolite levels of the pathways. The heterosis-associated analytes were screened with the partial least square method which outputs predictive variables with high contributions in the models. Pearson correlation was performed among pathways or analytes. The red or light blue arrow of a pathway indicated that the high-BPH hybrids had high or low metabolite levels. 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. (I) Venn diagram of heterosis-associated analytes across four developmental stages. (J) Correlations between metabolite levels of 4-hydroxycinnamic acid and plant height heterosis at all four developmental stages. (K) MS/MS spectra of M71T32_NEG and the pyruvaldehyde standard. (L) Correlations between metabolite levels of pyruvaldehyde and plant height heterosis at the seedling stage. Abbreviations: BPH, better-parent heterosis; SSPH, seedling stage plant height; ESPH, elongation stage plant height; FSPH, flowering stage plant height; MSPH, maturation stage plant height.

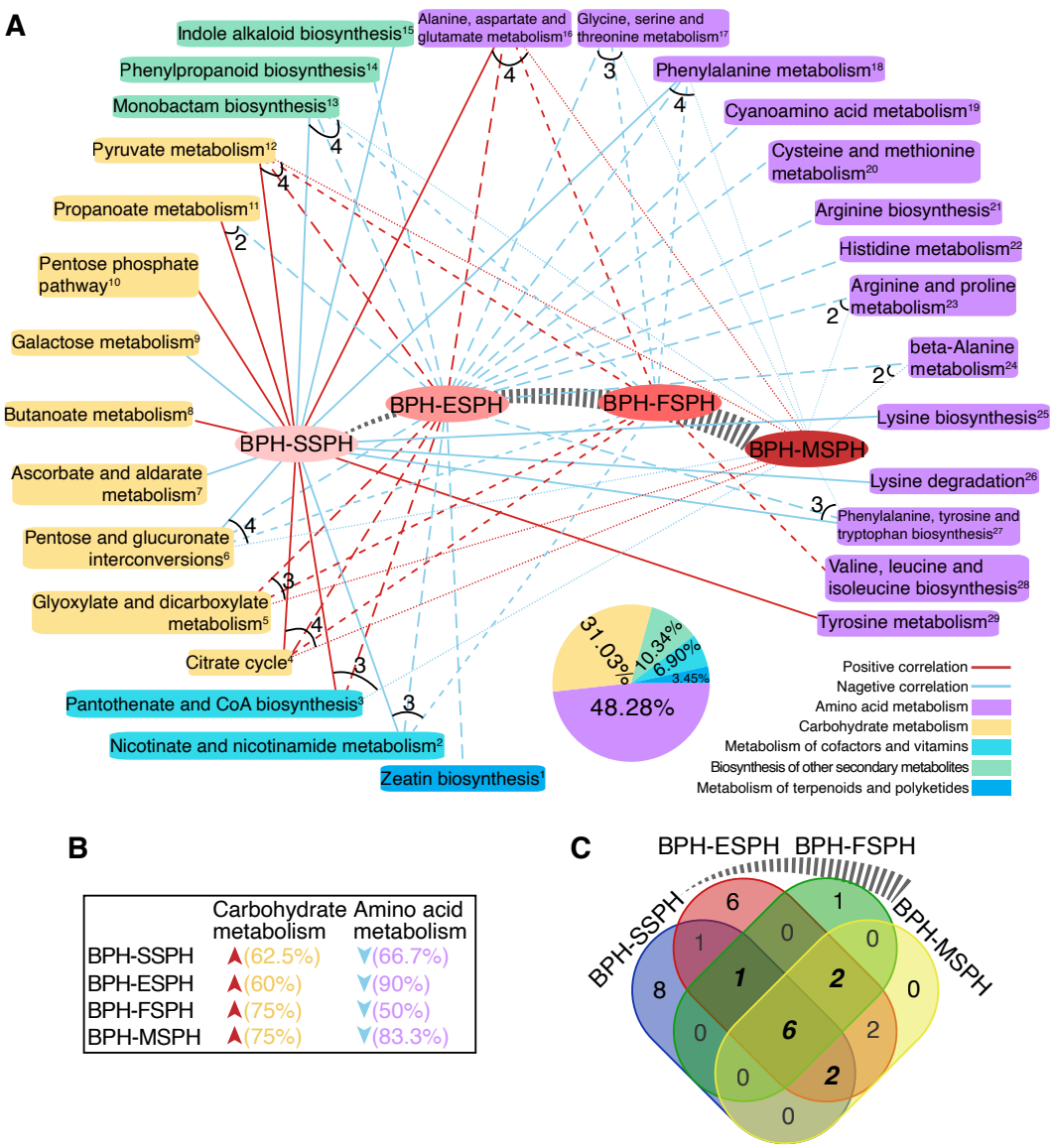


Figure 3: The metabolic pathways underlying dynamic height heterosis. (A) Connections between heterosis-associated metabolic pathways and height heterosis at four developmental stages. Pearson correlations were performed between height heterosis and metabolite levels of the significantly enriched pathways to investigate the correlation manners. Positive and negative correlations are shown with red and light blue lines, respectively. The pathways were divided into different types of metabolism (shown in the pie chart). The numbers on the lines indicate that the pathways were significantly enriched at multiple stages. The numbers next to the pathway names were used to represent the pathways shown in Fig. 2A, C, E, and G. (B) Percentages of the pathways from carbohydrate and amino acid metabolism with high and low metabolite levels in the high-BPH hybrids. The 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) are shown. (C) 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 significantly enriched at the elongation stage, and the numbers are marked in bold italic font. Abbreviations: BPH, better-parent heterosis; SSPH, seedling stage plant height; ESPH, elongation stage plant height; FSPH, flowering stage plant height; MSPH, maturation stage plant height.

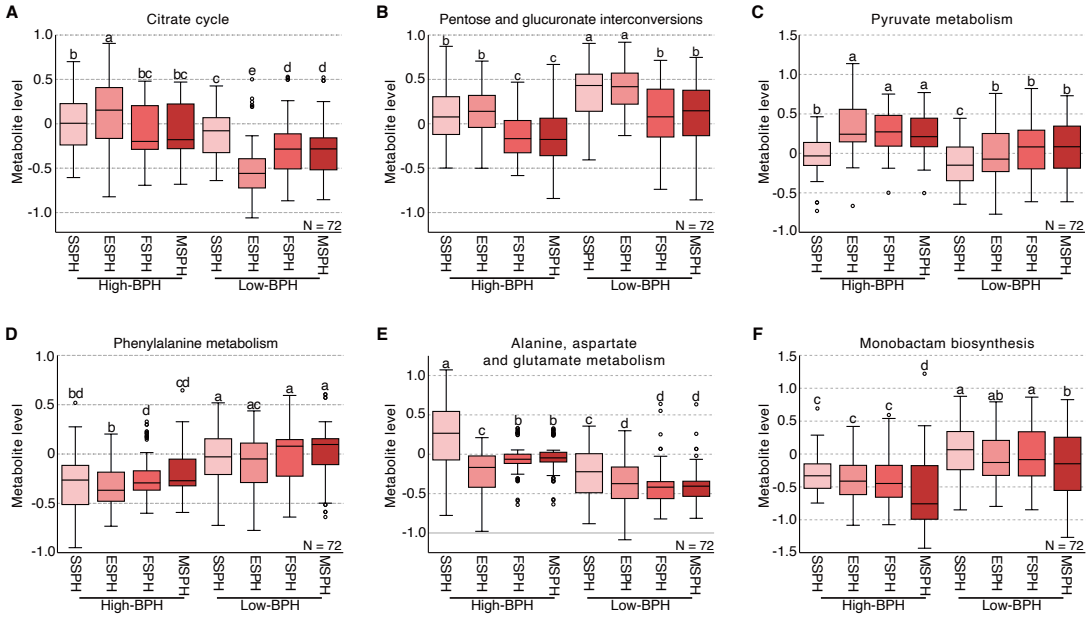


Figure 4: Metabolite levels of six metabolic pathways in the high- and low-BPH hybrids. Metabolite levels of six metabolic pathways, including the citrate cycle (A), pentose and glucuronate interconversions (B), pyruvate metabolism (C), phenylalanine metabolism (D), alanine, aspartate and glutamate metabolism (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, $P < 0.05$). Abbreviations: BPH, better-parent heterosis; SSPH, seedling stage plant height; ESPH, elongation stage plant height; FSPH, flowering stage plant height; MSPH, maturation stage plant height.

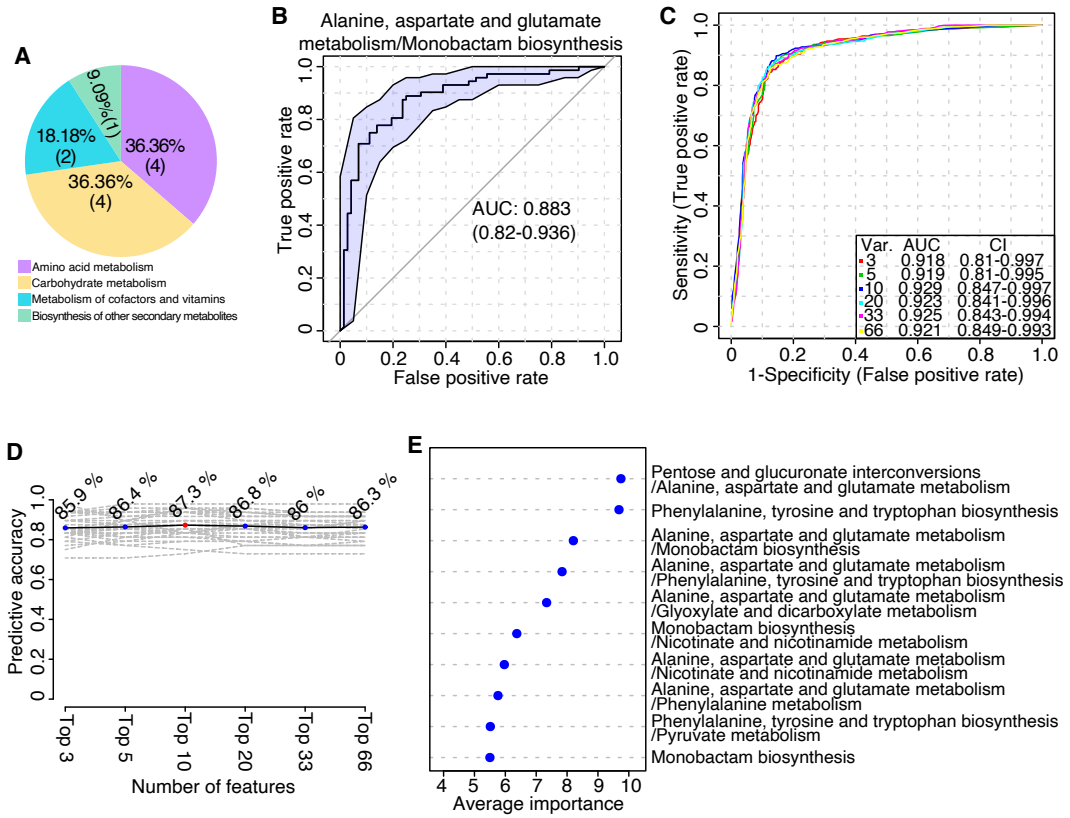


Figure 5: The development of pathway biomarkers for dynamic height heterosis. (A) Categories of the 11 metabolic pathways that were significantly enriched at most (three and four) developmental stages. (B) Univariate receiver operating characteristic (ROC) curve analysis of height heterosis at the seedling 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.

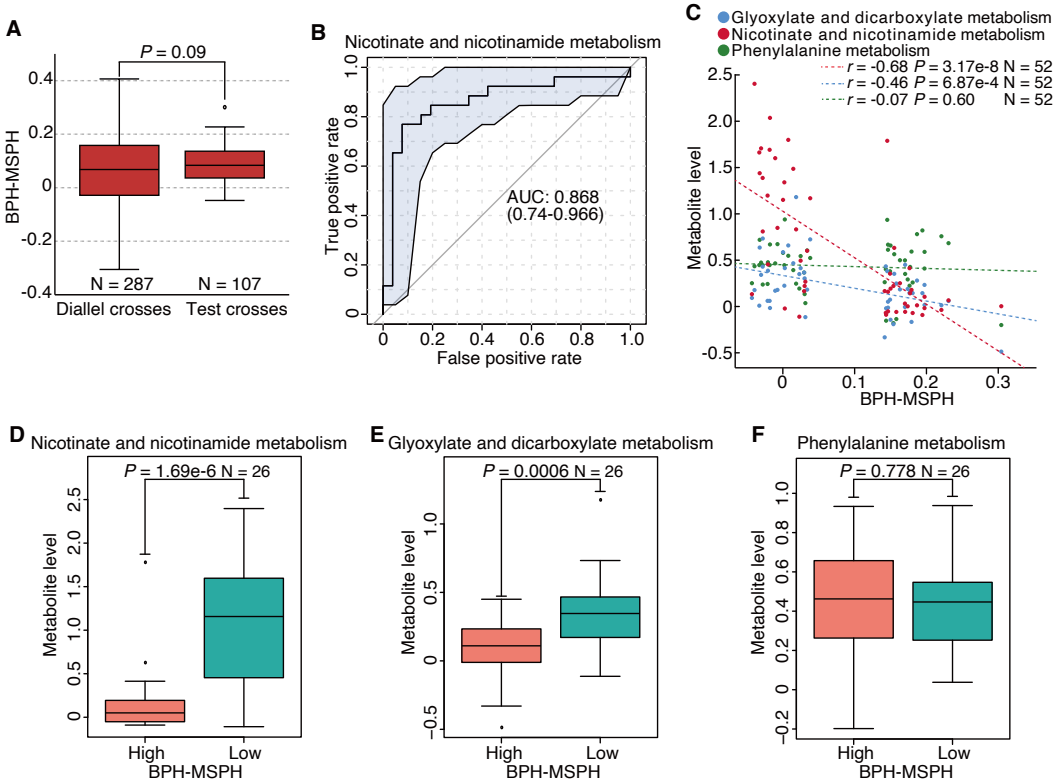


Figure 6: Validation of pathway biomarkers in test crosses. (A) Box plots of plant height heterosis at the maturation stage of the diallel and test crosses. (B) Univariate ROC curve analysis of height heterosis at the maturation stage of the test crosses. The test crosses were divided into low- and high-BPH 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. Abbreviations: BPH, better-parent heterosis; MSPH, maturation stage plant height.