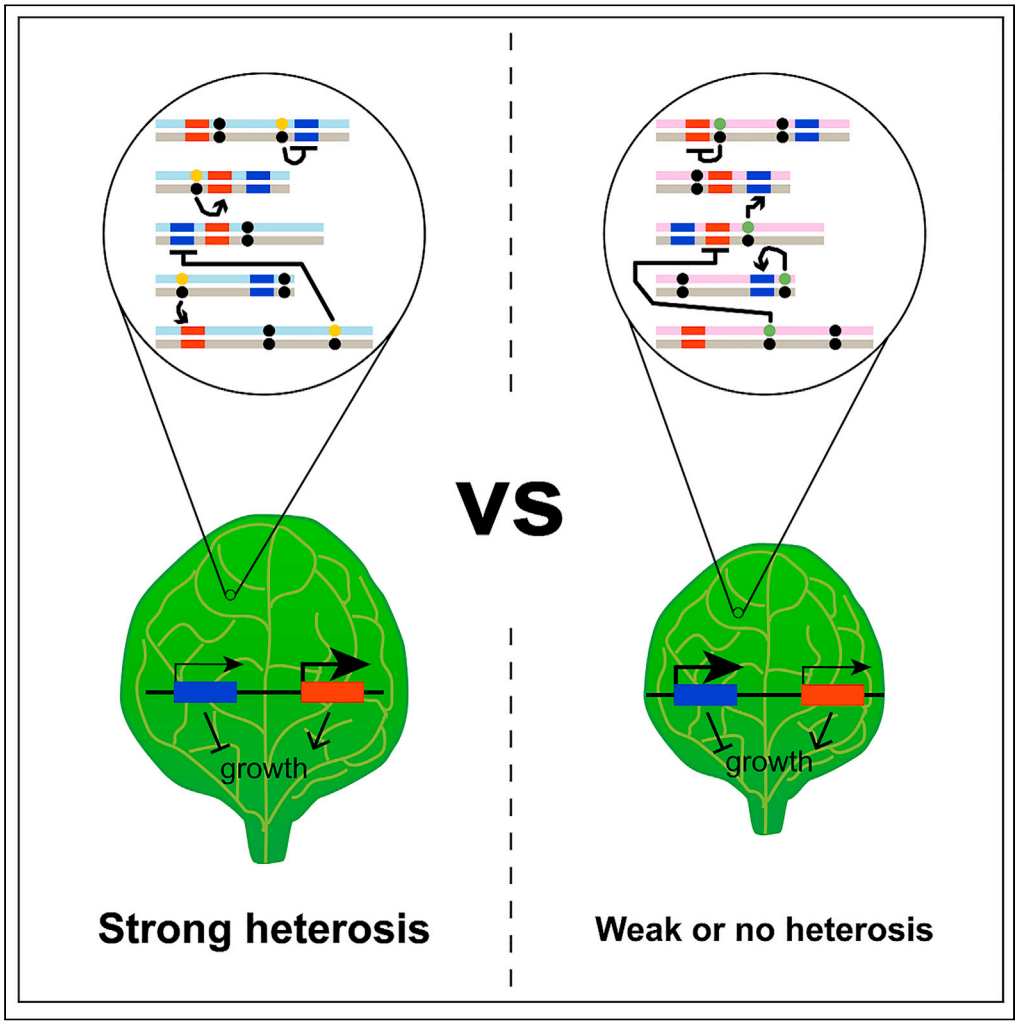
iScience

CellPress
OPEN ACCESS

Article

Genetic and molecular regulation of increased photosynthetic cell number contributes to leaf size heterosis in *Arabidopsis*



Wenwen Liu, Diqiu Ren, Wenyi Yang, ..., Xingwei Wang, Guangming He, Xing Wang Deng

heguangming@pku.edu.cn (G.H.)
deng@pku.edu.cn (X.W.D.)

Highlights

Up-regulated degree of cell cycle and photosynthesis genes relates to leaf heterosis

Increased photosynthetic cell number is a major driver for leaf size heterosis

Superior heterozygous eQTL accumulation optimizes F_1 gene expression and phenotype

Liu et al., iScience *26*, 107366
August 18, 2023 © 2023
https://doi.org/10.1016/j.isci.2023.107366

iScience

CellPress
OPEN ACCESS

Article

Genetic and molecular regulation of increased photosynthetic cell number contributes to leaf size heterosis in *Arabidopsis*

Wenwen Liu,[1] Diqiu Ren,[1] Wenyi Yang,[1] Miqi Xu,[1] Yi Zhang,[1] Xingwei Wang,[1] Guangming He,[1,*] and Xing Wang Deng[1,2,3,*]

SUMMARY

Heterosis is an important genetic phenomenon that has been observed and widely utilized in agriculture. However, the genetic and molecular bases of heterosis are unclear. Through transcriptome-wide association studies (TWAS) and expression quantitative trait locus (eQTL) analysis to integrate genome, transcriptome, and heterotic phenotype of a half-sibling *Arabidopsis* hybrid population, we report that the genetic and molecular bases of variations in leaf growth heterosis can be explained by the varied expression levels of growth-regulating genes resulting from distinct sets of heterozygous eQTLs carried by the half-sibling hybrids. In F_1 versus parent, the degree of up-regulated gene expression in the cell cycle pathway in the shoot apex and the photosynthesis pathway in true leaf positively correlates with true leaf area heterosis level, and this is affected by the accumulation of superior heterozygous eQTLs. This was further corroborated by the major contribution of increased photosynthetic cell number to leaf area heterosis.

INTRODUCTION

Heterosis, or hybrid vigor, is a widely observed genetic phenomenon in plants and animals that describes the F_1 hybrid's phenotype (i.e., biomass, yield, or environmental fitness) as being superior to that of its parents.[1–3] Despite its great influence in agriculture, the mechanism underlying heterosis remains a complicated conundrum of biology with no consensus as to how it works. Also unknown is why replacing half a parental genome produces superior trait performance in the resulting F_1 hybrid compared to the parents and why heterosis levels vary in different intraspecific hybrids. Differences in the genome-wide gene expression in F_1 hybrids relative to their parents are important links between genomic variation and heterosis phenotypes.[4–6] However, how gene expression changes in F_1 hybrids influence the heterosis phenotype is unclear, and the genetic causes of differential gene expression in hybrids relative to parents remain uncharacterized.

Heterosis in *Arabidopsis* is most commonly and obviously manifested in the greater biomass of F_1 hybrids versus their parents.[2] A previous population-level study revealed that *Arabidopsis* leaf area heterosis is highly, positively correlated with biomass heterosis.[7] Here, we systematically investigated the relationships among heterozygous genome, gene expression changes, and leaf area heterotic phenotypes in a half-sibling *Arabidopsis* F_1 hybrid population using combined eQTL and transcriptome-wide association studies (TWAS) analyses. Our results dissect the molecular network of how allelic divergence positively or negatively influences growth-related gene expression, and how the varied gene expression change degrees affect heterosis level variations in different half-sibling hybrids because of the distinct classes of heterozygous loci accumulated in hybrid genomes.

RESULTS

Specific gene expression changes significantly influence leaf area heterosis

Using the homozygous *Arabidopsis* ecotype Col-0 as a common maternal parent and 96 other homozygous ecotypes as paternal parents, we created a half-sibling *Arabidopsis* hybrid population and used it to explore how leaf area heterosis manifests among the 96 hybrids that differ from each other by half their

[1]School of Advanced Agricultural Sciences and School of Life Sciences, State Key Laboratory of Protein and Plant Gene Research, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China

[2]Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences in Weifang, Weifang, Shandong 261325, China

[3]Lead contact

*Correspondence: heguangming@pku.edu.cn (G.H.), deng@pku.edu.cn (X.W.D.)

https://doi.org/10.1016/j.isci.2023.107366

genomes (Figure S1). Because the first true leaf area of *Arabidopsis* 15 days after sowing (DAS) accounts for >70% of the leaf's eventual size and is well into cell expansion,[8] we determined the first true leaf area heterosis in those hybrids at this stage (Figure 1A; Table S1). To examine the transcriptomic regulation of heterosis during leaf development which undergo initiation from the shoot apical meristem and growth controlled by coordinated cell division and expansion,[9] we performed RNA-seq of the shoot apex 6 DAS and the first true leaf 15 DAS in all 96 half-sibling hybrids and their parents (Figure S2; Table S2). Then, we performed TWAS for leaf area heterosis using the corresponding transcriptome-wide gene expression ratios in hybrids versus the fixed parent Col-0. It showed that the significantly associated genes ($p < 0.01$) were specific to the tissue and developmental stage (Figure 1B), suggesting that the molecular regulation network of leaf growth heterosis is spatially and temporally dynamic.

In detail, for shoot apex, the changes in the expression of 355 genes in F_1 versus Col-0 significantly correlated (200 positively and 155 negatively) with leaf area heterosis, while for true leaf, the changes in the expression of 232 genes in F_1 versus Col-0 significantly correlated (115 positively and 117 negatively) with leaf area heterosis, and very few genes are found to be correlated simultaneously with the heterosis for both tissues (Figure 1C). The correlation coefficients varied from −0.41 to 0.43 (Table S3), indicating a cumulative effect of those genes' expression changes on the heterotic phenotype of leaf growth. Meanwhile, differential gene expression analysis of TWAS candidate genes also validated the contribution of their expression changes in hybrid to the heterosis levels (Figure S3).

Different biological pathways cooperatively contribute to leaf area heterosis in a temporally dynamic context

To gain a deeper understanding of how those genes function in leaf area heterosis, Gene Ontology (GO) enrichment analysis was separately conducted on each of the aforementioned 4 sets of significantly associated genes. The positively correlated genes for shoot apex were overrepresented in cell cycle regulation, whereas the negatively correlated genes were overrepresented in cell maturation (Figure 1D). This suggests that enhanced cell cycle and weakened cell maturation in the shoot apex might contribute to leaf area heterosis, which leads to increased leaf cell number in hybrids relative to the parents, thus resulting in a larger leaf area in a later period of leaf development. With the leaf well into the expansion phase 15 DAS, the photosynthesis pathway was significantly enhanced in hybrid relative to parents and thus could maintain the leaf area heterosis by more energy producing, fixing, and efficiently allocating. However, the function of negatively correlated genes was overrepresented in mitochondrion organization (Figure 1D), suggesting a potential negative role of energy decomposition in heterosis. Remarkably, at the population level, the leaf area heterosis levels rose along with the expression ratio of those genes in hybrids versus parents (Figure 1E). In particular, significant difference in the degree for the expression change of those genes displayed between the two groups of hybrids showing the highest and lowest levels of leaf area heterosis (Figure 1F). These results indicate that the transcriptional regulation of leaf growth heterosis is quite dynamic, involving distinct sets of genes during different developmental periods. Moreover, in hybrids, both up-regulated and down-regulated pathways cooperatively contribute to leaf area heterosis.

Increased photosynthetic cell number is the prominent cellular source for leaf area heterosis

Palisade mesophyll cells, which are located near the upper surface of the leaf, contain a large number of chloroplasts for both harnessing light and photosynthesis. To verify the contribution of enhanced gene transcription in cell cycle and photosynthesis pathways to leaf area heterosis, we investigated the palisade mesophyll cell area and cell number in the first true leaf area 15 DAS for the 96 hybrid combinations (Figures 2A and S1). Significantly increased photosynthetic cell number was displayed in F_1 hybrids relative to parents, but no significant difference was observed in regard to photosynthetic cell area (Figure 2B). Accordingly, mid-parent and best-parent heterosis (MPH and BPH) for leaf area and cell number, but not cell area, was prevalent in the hybrid population (Figure 2C). These results further prove the main contribution of increased cell number and photosynthesis to leaf area heterosis at the population level. Likewise, TWAS performed for the cellular-level phenotypes also identified the gene expression change related to their heterosis (Figure S4). Consistently, many shared significant genes were uncovered by the TWAS for leaf area and leaf cell number (Figure S5A). However, the TWAS significant genes for cell number heterosis showed no clear function enrichment, and that for cell area heterosis indicates the positive role of glucan biosynthesis and negative role of jasmonic acid (JA) biosynthesis (Figure S5B). Notably, JA-induced glucosinolate accumulation and its negative effects on cell size have been uncovered.[10,11] Furthermore, different from the comparisons between hybrids and the fixed maternal parent, for hybrids versus paternal

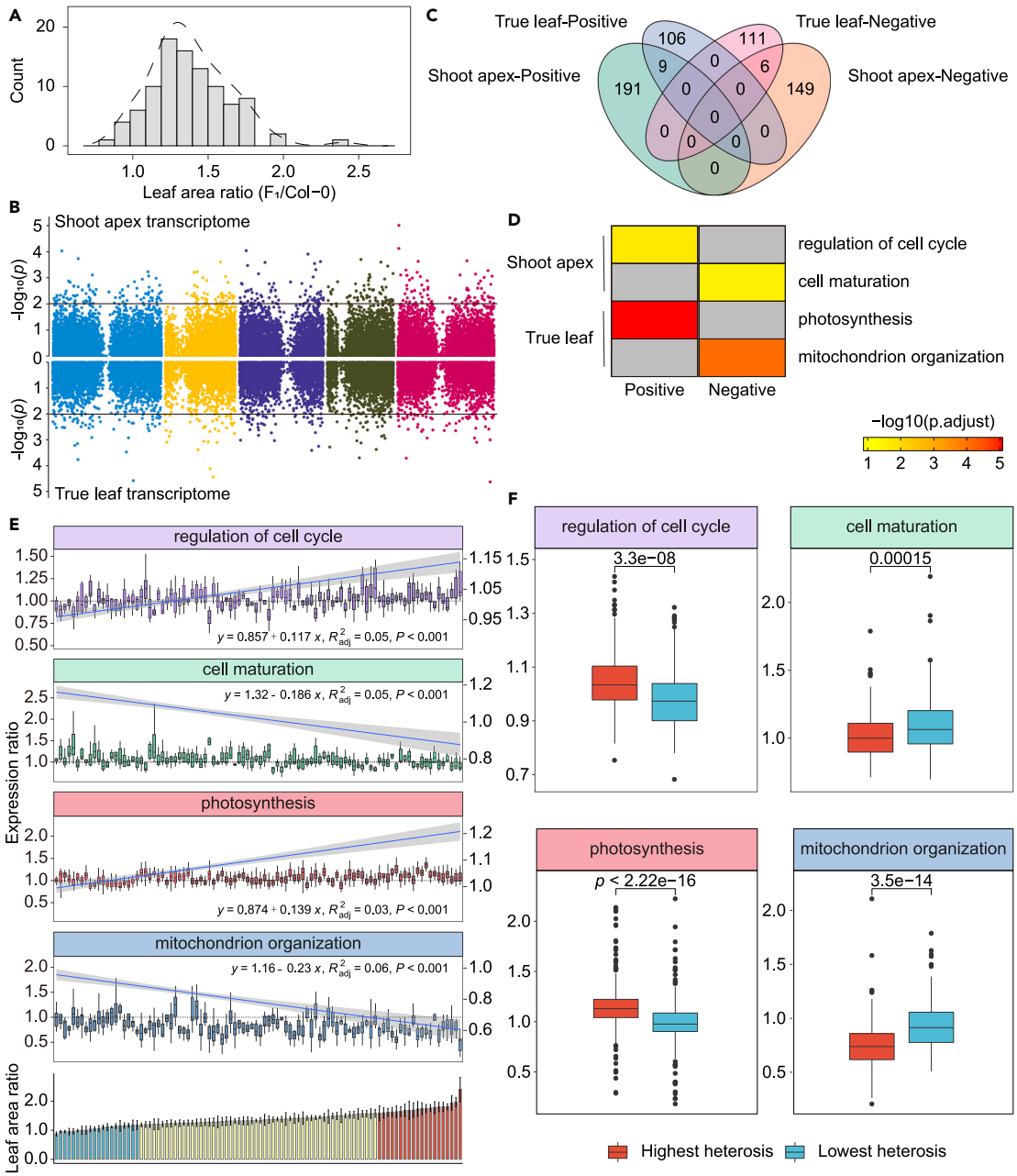


Figure 1. Mechanism of how the transcriptomic alterations in the specific pathways function in leaf area heterosis variation

(A) Normal distribution of leaf area heterosis (F_1/Col-0) in the 96 half-sibling hybrids.

(B) Manhattan plots of transcriptome-wide association study (TWAS) performed between leaf area heterosis and gene expression alterations in shoot apex and true leaf in hybrids compared to the maternal parent (Col-0). The black horizontal lines indicate the threshold levels for $p < 0.01$.

(C) Overlap among TWAS candidate genes positively or negatively correlated to leaf area heterosis.

(D) Functional analysis of the 4 sets of TWAS candidate genes in (C). Gray means no significant enrichment.

(E) Relationships between leaf area heterosis levels (leaf area ratio) and the degrees of gene expression change (expression ratio) in hybrids versus Col-0. Curvilinear regression lines were generated using the lm method (n = 2–5; Error bar of bar plot indicates mean ± sd).

(F) Gene expression ratios to Col-0 of the hybrids that had the highest and lowest levels of leaf area heterosis. Above each plot are p values obtained by Student's *t* test.

ecotypes, few biological processes were significantly enriched for the TWAS candidate genes (Figure S5), suggesting the diverse biological functions of the genes from fixed maternal parent (Col-0) to complement the allelic expression from different paternal parents in half-sibling hybrids, which is likely because of the

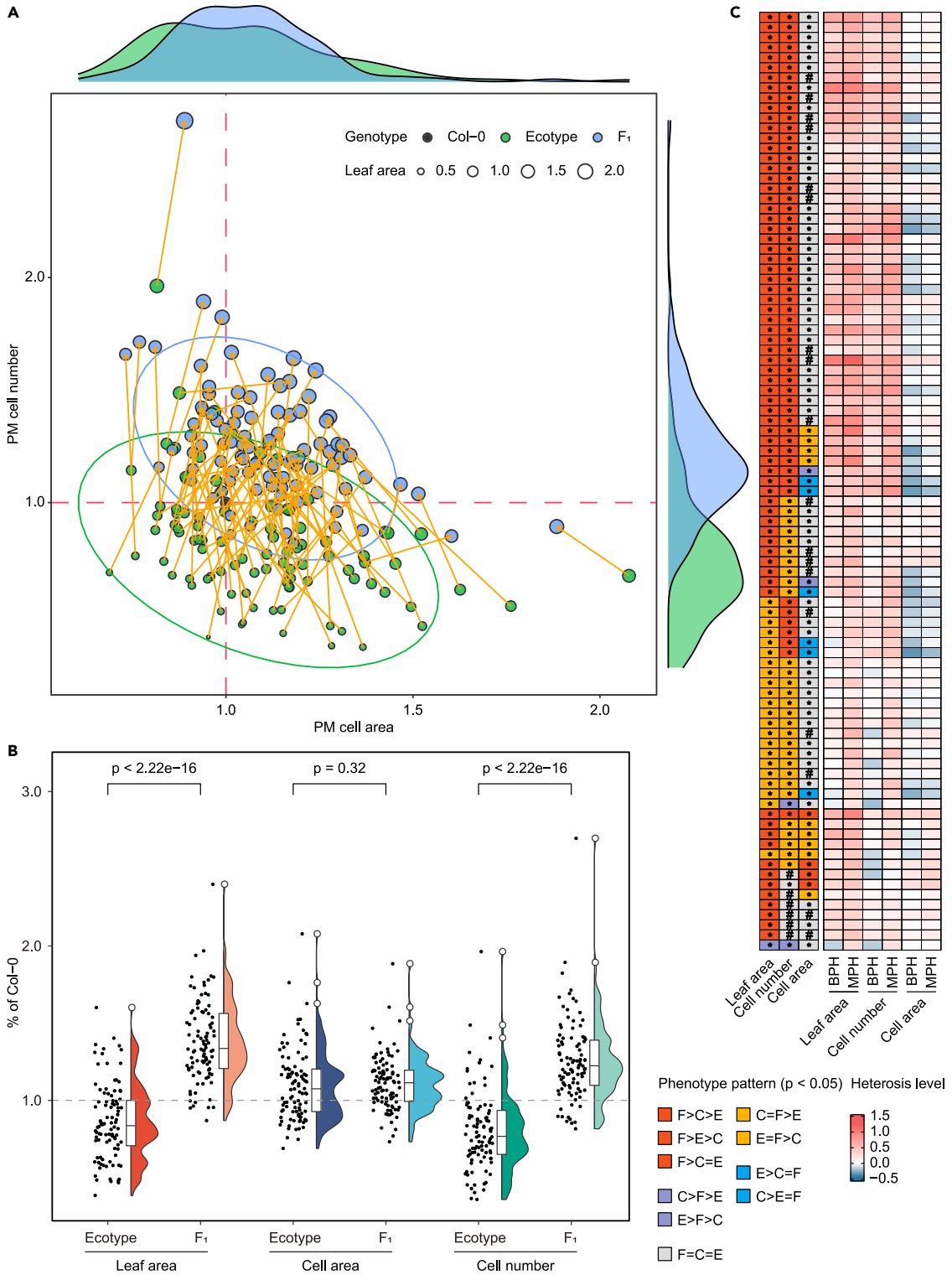


Figure 2. The prominent contribution of increased photosynthetic cell number to leaf area heterosis

(A) Leaf area, cell number, and cell area changes in hybrids compared to the parents (maternal Col-0 and paternal ecotypes). The data were normalized by Col-0 trait values. Red dashed lines indicate the normalized value of Col-0. Density plots outside the top and right of the scatterplot visualize the cell number and cell area distribution of the hybrids and paternal parents. The colors in the density plots were mapped according to the colors of the points in the scatterplot.

Figure 2. *Continued*

(B) Raincloud plots show the distribution of trait value in hybrids and parents. Above each plot are p values obtained by Student's *t* test.

(C) The roles of cell number or cell area increasing in leaf area heterosis for each F_1 hybrid (n = 3–5, $p < 0.05$, Student's *t* test). * Significance supported by all the 3 pairwise comparisons; # Significance supported by 2 of the 3 pairwise comparisons. F, F_1 hybrid; C, Col-0; E, ecotype.

natural genomic variation in different paternal ecotypes. Next, in order to understand how increased cell number in hybrid compared to parent happened at the tissue level, we used the mitotic cell reporter line of Col-0×Per-1 F_1 hybrid that gains leaf area heterosis by increased cell number to visualize the mitotic cell regions of the first true leaves (Table S1). It showed that a significantly larger mitotic cell region was presented in F_1 versus Col-0 (Figure S6), suggesting that the ongoing division of more cells resulted in the increased cell number in hybrid.

Genomic architecture affects gene expression alteration in hybrids to shape heterosis performance

To discover how replacing half the genome of a parent could lead to the obviously divergent degrees of expression fluctuation of the heterosis-related genes in different half-sibling hybrids and thus affect their leaf area heterosis performance, we studied the genetic regulation of those genes' expression through eQTL mapping. Both *cis*-eQTLs (physical distance to the eGene <5 Mb) and *trans*-eQTLs (physical distance to the eGene >5 Mb) were significantly recognized ($p < 10^{-6}$) for those eGenes.[12] Significant differences in the degree of the changes in each eGene expression in hybrids versus parent (Col-0) were observed between the half-sibling hybrids containing homozygous and heterozygous forms of the eQTLs. And the effects of the heterozygous eQTL on the changes in eGene expression in hybrids versus parent differed from each other, which could be either up-regulation or down-regulation (Figures 3A and 3B). In shoot apex, the change in expression in F_1 versus parent in the cell cycle regulation core gene *CYCA1* was associated with the genotype of the eQTL at Chr1 24129636, and the expression ratio of *CSLD5*, which is involved in cell plate formation during cytokinesis, was associated with the genotype of the eQTL at Chr1 29745973 (Figure 3A).[13,14] Similar regulation modes were also found for genes related to cell maturation, such as *GH9A1*, which is involved in cellulose synthesis and cell elongation (with the eQTL at Chr1 25320108), and *EXPANSIN3*, which functions in cell wall organization and cell growth (with the eQTL at Chr3 5519925).[15–17] Moreover, we found that the change in gene expression in hybrids versus the parent could be simultaneity associated with several genetic loci. For example, in true leaf, the change in expression of *CP12-1*, which participates in the formation of an enzyme complex embedded in the Calvin cycle,[18] was associated with two eQTLs at Chr2 15354570 and Chr2 19457218 (Figure 3B). In addition, the genotype of eQTL at Chr1 22134246 was found associated with the change in the expression of *COX17*, a gene that mediates the assembly of a functional cytochrome oxidase complex in the mitochondria.[19]

To validate the effect of the aforementioned genetic and molecular mechanisms on the level of corresponding heterotic phenotypes, we further analyzed the relationships between the eQTL genotypes and the leaf area heterosis in the 96 F_1 hybrids. Obvious divergences displayed in the lead SNP genotypes (heterozygous or homozygous) between the two sets of hybrids with the highest and lowest levels of leaf area heterosis (Figures 3C and 4A–4E). Besides, some eQTLs' effects on heterosis for leaf area were validated in other hybrid populations even when the plants were grown in different environments and phenotyped at different developmental stages (Figures 4F–4H). Thus, although altering half of the Col-0 genome led to heterozygosity of numerous alleles in hybrids, not all of the heterozygous loci functioned in the performance of leaf area heterosis. Instead, only some of the heterozygous loci in the genome could affect spatiotemporal-specific changes in the expression of genes related to leaf growth in hybrids versus parent. And variation in the leaf area heterosis phenotype in different hybrids was coordinately conferred by the changes in the growth-related gene expression that dynamically regulated during leaf development.

DISCUSSION

Plant heterosis has contributed greatly to improving the yield of many crops and vegetables, yet the genetic and molecular bases of heterosis remain unclear, particularly for these complicated quantitative traits such as biomass accumulation vigor in hybrids.[3] Currently, both modern molecular biology and genomics ask why substitution of half the parental genome often causes improved trait performance in the resulting hybrids and how introducing different allogeneic genomes results in different levels of hybrid

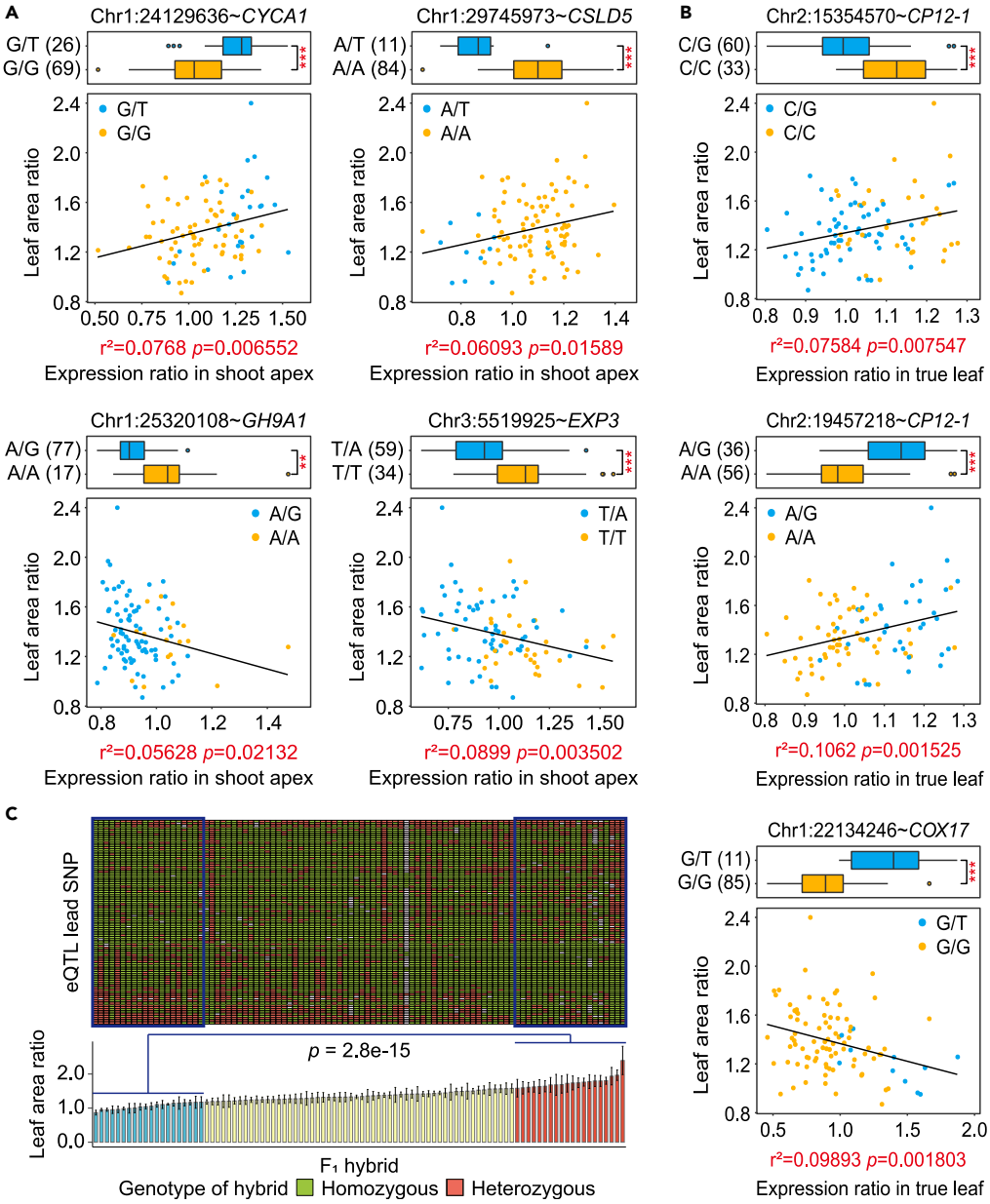


Figure 3. The superior heterozygous loci gathering in hybrid aids its gene expression changes to benefits leaf growth

(A and B) Boxplots and scatterplots show the relationships among the eQTL genotype, eGene expression change, and leaf area heterosis. The horizontal axes shared by the boxplots and scatterplots show the gene expression ratio in (A) shoot apex and (B) true leaf. The vertical axes of boxplots show the genotypes of each eQTL in the hybrids represented by the lead SNP, followed by the corresponding number of hybrids. Curvilinear regression lines were generated using the lm method. *p < 0.05, **p < 0.01, ***p < 0.001 (Student's *t*-test).

(C) The relationship between the eQTL genotype and the level of leaf area heterosis. Each row in the heatmap indicates a lead SNP and each column indicates a hybrid. The leaf area ratio is displayed under the heatmap (n = 2–5; Error bar indicates mean ± sd; p value was calculated by Student's *t*-test).

vigor. Our study explored the genetic and molecular mechanisms as well as the cellular basis underlying leaf growth heterosis using a half-sibling hybrid *Arabidopsis* population (Figure 5). During leaf development, half-sibling hybrids with more heterozygous loci that up-regulate positive growth-regulating genes and down-regulate negative growth-regulating genes benefit from the comprehensive effects of changes

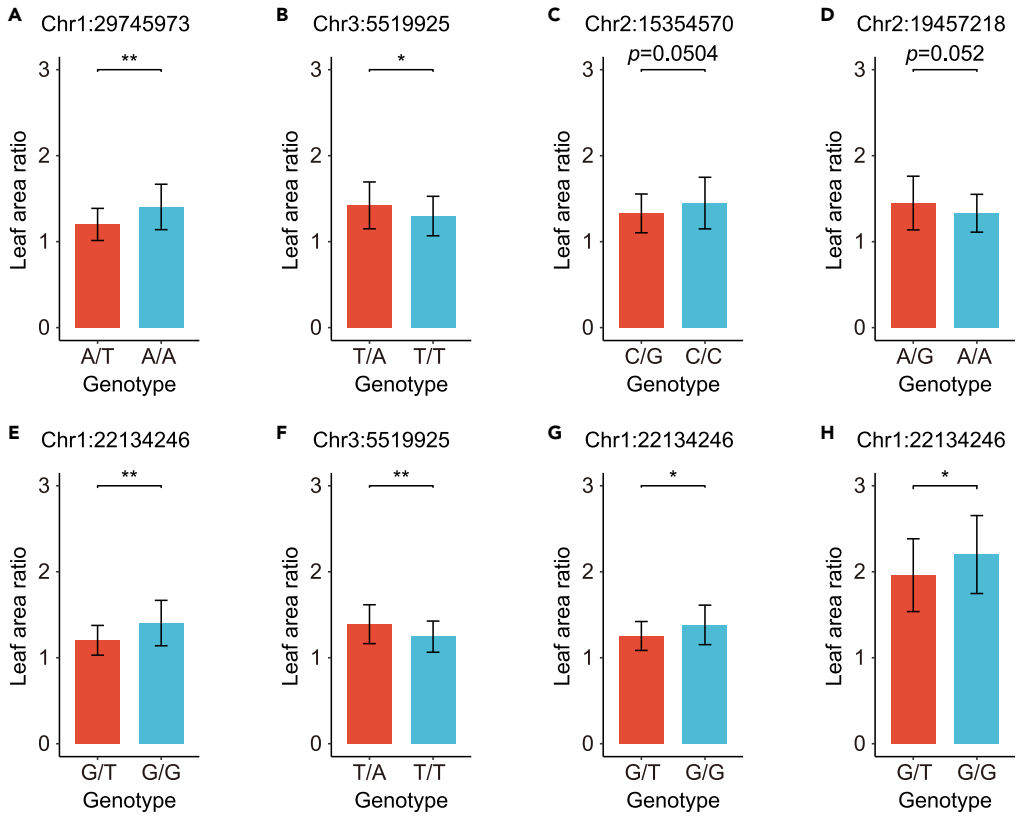


Figure 4. Validation of the leaf area heterosis-related eQTLs in different hybrid populations

(A–E) Different eQTL genotypes can distinguish the leaf area heterosis levels in the 96 half-sibling hybrids.

(F and G) Different eQTL genotypes can distinguish the leaf area heterosis levels in another population containing 94 half-sibling hybrids grown at the same condition and at the same developmental stage (15 DAS cultured at plates) as for the 96 hybrids.

(H) Different eQTL genotypes can distinguish the leaf area heterosis levels in a population containing 171 half-sibling hybrids grown in soil for 7 days after 7-day culturing at plate (Original data are from Yang et al., 2017). *$p < 0.05$, **$p < 0.01$, Student's *t*-test.

in growth-regulating gene expression, resulting in high levels of leaf area heterosis through increased photosynthetic cell number. In contrast, half-sibling hybrids with more heterozygous loci that down-regulate positive growth-regulating genes and up-regulate negative growth-regulating genes are hindered by the comprehensive effects of changes in growth-regulating gene expression, resulting in low levels of leaf area heterosis. Finally, the mature leaf area heterosis level would be determined by the whole effect of dynamic profile of gene expression changes in the hybrid versus parents during leaf development.

Besides this general principle, a few exceptions were observed in some hybrids (Figures 1E, 3A, and 3B), likely because of other ecotype-specific genomic variations or the genome-wide epigenomic divergence. Given the divergent heterozygous loci carried in the different half-sibling hybrids' genomes and the genome-wide interactions among them, the size of the effect of an eQTL on the corresponding changes in gene expression in different hybrids could be distinct. Thus, a more detailed study will help to fully elucidate the heterosis mechanism in a specific trait of a specific hybrid. However, varying divergence in the epigenomes of different half-sibling hybrids may also contribute to changes in their gene expression compared to their parents. Consistently, we identified *cis*- or *trans*-eQTLs for only part of the significant genes identified by TWAS. This implies that, in addition to the genomic regulators acting in a *cis* or *trans* manner, epigenetic factors (e.g., DNA methylation, histone modification, and small RNAs) could also influence the degree of change in gene expression in hybrids versus their parents, and this can be further characterized through epigenome-wide association studies.

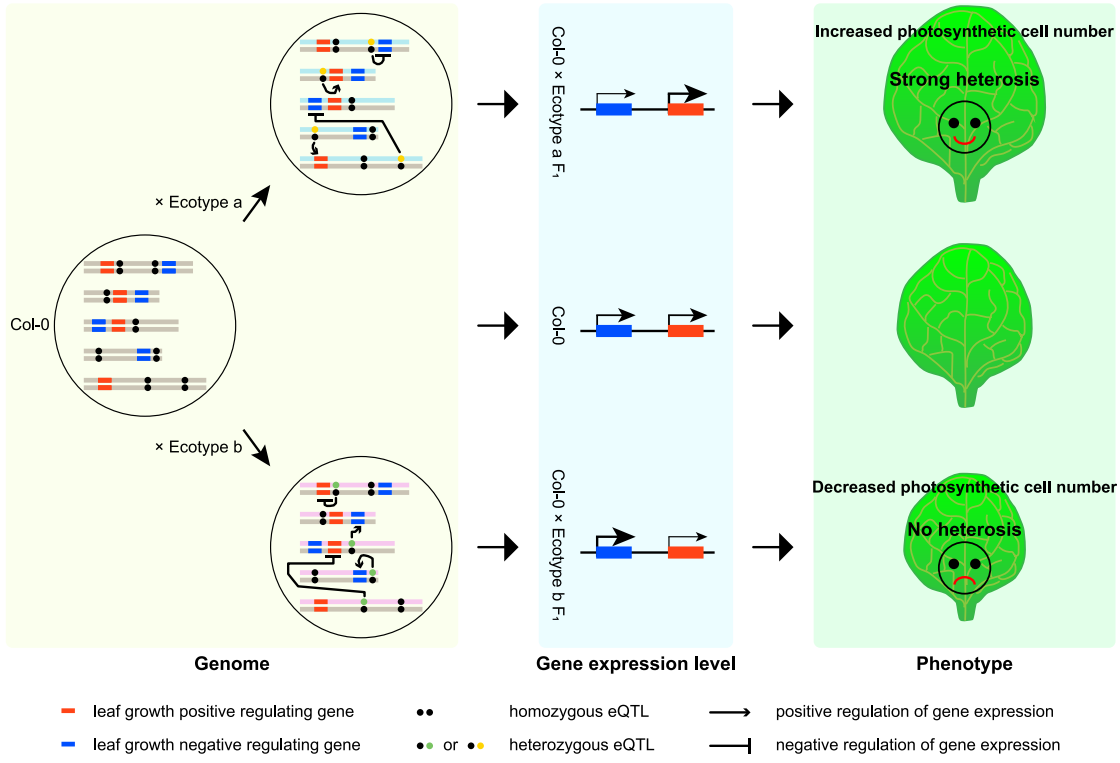


Figure 5. Genetic and molecular mechanism underlying the contribution of increased photosynthetic cell number to leaf area heterosis

In half-sibling hybrids (Col-0 as the maternal parent and different ecotypes as paternal parents), distinct regions of their genome became heterozygous relative to the homozygous Col-0 genome. The heterozygous loci could spatiotemporally affect (either in *cis* or *trans*) the expression levels of leaf growth-regulating genes in the F_1 hybrids displayed as 4 modes: (1) The heterozygous locus acts upon the genes that positively regulate leaf growth, leading to up-regulation of their expression in hybrids versus Col-0; (2) the heterozygous locus acts upon the genes that negatively regulate leaf growth, leading to down-regulation of their expression in hybrids versus Col-0; (3) the heterozygous locus acts upon the genes that positively regulate leaf growth, leading to down-regulation of their expression in hybrids versus Col-0; or (4) the heterozygous locus acts upon the genes that negatively regulate leaf growth, leading to up-regulation of their expression in hybrids versus Col-0. For the hybrids that substantially accumulated type (1) and (2) heterozygous loci because of genomic variation between parents, the expression of the genes that positively regulate leaf growth was up-regulated while the expression of the genes negatively regulating leaf growth was down-regulated relative to Col-0. Together, this resulted in a larger leaf area in the hybrids versus Col-0. In contrast, for the hybrids that substantially accumulated type (3) and (4) heterozygous loci because of genomic variation between parents, the expression of the specific genes positively regulating leaf growth was down-regulated while the expression of the specific genes negatively regulating leaf growth was up-regulated relative to Col-0. The combined results were a smaller leaf area in hybrids versus Col-0.

The genes involved in the cell cycle and photosynthesis were revealed as critical contributors to biomass or leaf growth heterosis in a few *Arabidopsis* hybrids.[7,8,20–24] After examining this from a population perspective, our findings emphasize the general roles of these two pathways in leaf growth heterosis in transcriptomic level, and provide a clue to understand it by studying the mechanism that how photosynthetic cell number improved in hybrids.

Overall, we revealed a general mechanism of how divergent hybrid genomes influence variation in gene expression changes in hybrid relative to parent and lead to heterosis level variation. Our study proved that TWAS and eQTL are feasible and effective approaches that help us understand the genetic and molecular mechanisms of heterosis. Moreover, the methods used in this study provide bioinformatic clues for hybrid breeding programs in crops, including mining of the superior alleles of target traits, predicting heterosis performance, and selecting optimal parental combinations for breeding ideal hybrids.

Limitations of the study

By TWAS for leaf area heterosis, we have identified a set of genes positively or negatively affect heterosis level through gene expression change in hybrid relative to parent. However, no putative main-effect candidate gene was recognized, which might be due to the limited population size in this study. We have

performed eQTL analysis and uncovered the relationships between eQTL genotypes and leaf area heterosis phenotypes in different hybrid populations, as well as the relationships between eQTL genotypes and TWAS candidate gene expression changes. However, the biological function of these eQTLs or genes remained to be verified through genetic modification.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Trait phenotyping
  - Mitotic cell reporter line generation and GUS staining
  - DNA preparation and whole-genome resequencing
  - RNA preparation and sequencing
  - Genome resequencing data analysis and variant calling
  - RNA-seq data processing
  - TWAS
  - eQTL identification
  - GO enrichment analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.107366.

ACKNOWLEDGMENTS

This work was supported by grants from the National Natural Science Foundation of China (31871221 and 31621001), the State Key Laboratory of Protein and Plant Gene Research, and the Peking-Tsinghua Center for Life Sciences (to X.W.D.).

AUTHOR CONTRIBUTIONS

G.H., X.W.D., and W.L. conceived the study. G.H., W.L., and D.R. designed the experiments. W.L., D.R., W.Y., M.X., Y.Z., and X.W. conducted the experiments. W.L. and G.H. analyzed the data. W.L. developed the programs and analytic codes. W.L. created the figures. W.L., G.H., and X.W.D wrote the manuscript. All authors approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 20, 2022
Revised: May 31, 2023
Accepted: July 10, 2023
Published: July 13, 2023

REFERENCES

1. Birchler, J.A., Yao, H., Chudalayandi, S., Vaiman, D., and Veitia, R.A. (2010). Plant Cell *22*, 2105–2112. https://doi.org/10.1105/tpc.110.076133.

2. Chen, Z.J. (2013). Genomic and epigenetic insights into the molecular bases of heterosis. Nat. Rev. Genet. *14*, 471–482. https://doi.org/10.1038/nrg3503.

3. Fujimoto, R., Uezono, K., Ishikura, S., Osabe, K., Peacock, W.J., and Dennis, E.S. (2018). Recent research on the mechanism of heterosis is important for crop and vegetable breeding systems. Breed Sci. *68*, 145–158. https://doi.org/10.1270/jsbbs.17155.

4. Liu, W., Zhang, Y., He, H., He, G., and Deng, X.W. (2022). From hybrid genomes to heterotic trait output: Challenges and opportunities. Curr. Opin. Plant Biol. *66*, 102193. https://doi.org/10.1016/j.pbi.2022.102193.

5. Baranwal, V.K., Mikkilineni, V., Zehr, U.B., Tyagi, A.K., and Kapoor, S. (2012). Heterosis: emerging ideas about hybrid vigour. J. Exp. Bot. *63*, 6309–6314. https://doi.org/10.1093/jxb/ers291.

6. Hochholdinger, F., and Baldauf, J.A. (2018). Heterosis in plants. Curr. Biol. *28*, R1089–R1092. https://doi.org/10.1016/j.cub.2018.06.041.

7. Yang, M., Wang, X., Ren, D., Huang, H., Xu, M., He, G., and Deng, X.W. (2017). Genomic architecture of biomass heterosis in Arabidopsis. Proc. Natl. Acad. Sci. USA *114*, 8101–8106. https://doi.org/10.1073/pnas.1705423114.

8. Groszmann, M., Gonzalez-Bayon, R., Greaves, I.K., Wang, L., Huen, A.K., Peacock, W.J., and Dennis, E.S. (2014). Intraspecific Arabidopsis hybrids show different patterns of heterosis despite the close relatedness of the parental genomes. Plant Physiol. *166*, 265–280. https://doi.org/10.1104/pp.114.243998.

9. Gonzalez, N., Vanhaeren, H., and Inzé, D. (2012). Leaf size control: complex coordination of cell division and expansion. Trends Plant Sci. *17*, 332–340. https://doi.org/10.1016/j.tplants.2012.02.003.

10. Skirycz, A., Reichelt, M., Burow, M., Birkemeyer, C., Rolcik, J., Kopka, J., Zanor, M.I., Gershenzon, J., Strnad, M., Szopa, J., et al. (2006). DOF transcription factor AtDof1.1 (OBP2) is part of a regulatory network controlling glucosinolate biosynthesis in Arabidopsis. Plant J. *47*, 10–24. https://doi.org/10.1111/j.1365-313X.2006.02767.x.

11. Mikkelsen, M.D., Petersen, B.L., Glawischnig, E., Jensen, A.B., Andreasson, E., and Halkier, B.A. (2003). Modulation of CYP79 genes and glucosinolate profiles in Arabidopsis by defense signaling pathways. Plant Physiol. *131*, 298–308. https://doi.org/10.1104/pp.011015.

12. Sul, J.H., Raj, T., de Jong, S., de Bakker, P.I.W., Raychaudhuri, S., Ophoff, R.A., Stranger, B.E., Eskin, E., and Han, B. (2015). Accurate and fast multiple-testing correction in eQTL studies. Am. J. Hum. Genet. *96*, 857–868. https://doi.org/10.1016/j.ajhg.2015.04.012.

13. Gu, F., Bringmann, M., Combs, J.R., Yang, J., Bergmann, D.C., and Nielsen, E. (2016). Arabidopsis CSLD5 Functions in Cell Plate Formation in a Cell Cycle-Dependent Manner. Plant Cell *28*, 1722–1737. https://doi.org/10.1105/tpc.16.00203.

14. Vandepoele, K., Raes, J., De Veylder, L., Rouzé, P., Rombauts, S., and Inzé, D. (2002). Genome-wide analysis of core cell cycle genes in Arabidopsis. Plant Cell *14*, 903–916. https://doi.org/10.1105/tpc.010445.

15. Cosgrove, D.J. (2000). Loosening of plant cell walls by expansins. Nature *407*, 321–326. https://doi.org/10.1038/35030000.

16. Kwon, Y.R., Lee, H.J., Kim, K.H., Hong, S.W., Lee, S.J., and Lee, H. (2008). Ectopic expression of Expansin3 or Expansin beta 1 causes enhanced hormone and salt stress sensitivity in Arabidopsis. Biotechnol. Lett. *30*, 1281–1288. https://doi.org/10.1007/s10529-008-9678-5.

17. Nicol, F., His, I., Jauneau, A., Vernhettes, S., Canut, H., and Höfte, H. (1998). A plasma membrane-bound putative endo-1,4-beta-D-glucanase is required for normal wall assembly and cell elongation in Arabidopsis. EMBO J. *17*, 5563–5576. https://doi.org/10.1093/emboj/17.19.5563.

18. Elena López-Calcagno, P., Omar Abuzaid, A., Lawson, T., and Anne Raines, C. (2017). Arabidopsis CP12 mutants have reduced levels of phosphoribulokinase and impaired function of the Calvin-Benson cycle. J. Exp. Bot. *68*, 2285–2298. https://doi.org/10.1093/jxb/erx084.

19. Garcia, L., Welchen, E., Gey, U., Arce, A.L., Steinebrunner, I., and Gonzalez, D.H. (2016). The cytochrome c oxidase biogenesis factor AtCOX17 modulates stress responses in Arabidopsis. Plant Cell Environ. *39*, 628–644. https://doi.org/10.1111/pce.12647.

20. Fujimoto, R., Taylor, J.M., Shirasawa, S., Peacock, W.J., and Dennis, E.S. (2012). Heterosis of Arabidopsis hybrids between C24 and Col is associated with increased photosynthesis capacity. Proc. Natl. Acad. Sci. USA *109*, 7109–7114. https://doi.org/10.1073/pnas.1204464109.

21. Zhu, A., Greaves, I.K., Liu, P.C., Wu, L., Dennis, E.S., and Peacock, W.J. (2016). Early changes of gene activity in developing seedlings of Arabidopsis hybrids relative to parents may contribute to hybrid vigour. Plant J. *88*, 597–607. https://doi.org/10.1111/tpj.13285.

22. Liu, P.C., Peacock, W.J., Wang, L., Furbank, R., Larkum, A., and Dennis, E.S. (2020). Leaf growth in early development is key to biomass heterosis in Arabidopsis. J. Exp. Bot. *71*, 2439–2450. https://doi.org/10.1093/jxb/eraa006.

23. Liu, W., He, G., and Deng, X.W. (2021). Biological pathway expression complementation contributes to biomass heterosis in Arabidopsis. Proc. Natl. Acad. Sci. USA *118*, e2023278118. https://doi.org/10.1073/pnas.2023278118.

24. Ni, Z., Kim, E.D., Ha, M., Lackey, E., Liu, J., Zhang, Y., Sun, Q., and Chen, Z.J. (2009). Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. Nature *457*, 327–331. https://doi.org/10.1038/nature07523.

25. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

26. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv. https://doi.org/10.48550/arXiv.1303.3997.

27. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575. https://doi.org/10.1086/519795.

28. Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360. https://doi.org/10.1038/nmeth.3317.

29. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics *28*, 882–883. https://doi.org/10.1093/bioinformatics/bts034.

30. Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., Yuan, X., Zhu, M., Zhao, S., Li, X., and Liu, X. (2021). rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated tool for Genome-Wide Association Study. Dev. Reprod. Biol. *19*, 619–628. https://doi.org/10.1016/j.gpb.2020.10.007.

31. Endelman, J.B. (2011). Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. Plant Genome *4*, 250–255. https://doi.org/10.3835/plantgenome2011.08.0024.

32. Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics *28*, 1353–1358. https://doi.org/10.1093/bioinformatics/bts163.

33. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287. https://doi.org/10.1089/omi.2011.0118.

34. Rymen, B., Coppens, F., Dhondt, S., Fiorani, F., and Beemster, G.T.S. (2010). Kinematic analysis of cell division and expansion. Methods Mol. Biol. *655*, 203–227. https://doi.org/10.1007/978-1-60761-765-5_14.

35. Colon-Carmona, A., You, R., Haimovitch-Gal, T., and Doerner, P. (1999). Spatio-temporal analysis of mitotic activity with a labile cyclin-GUS fusion protein. Plant J. *20*, 503–508. https://doi.org/10.1046/j.1365-313x.1999.00620.x.

36. 1001 Genomes Consortium. Electronic address: magnus.nordborg@gmi.oeaw.ac.at; 1001 Genomes Consortium; The-1001-Genomes-Consortium (2016) (2016). 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell *166*, 481–491. https://doi.org/10.1016/j.cell.2016.05.063.

37. VanRaden, P.M. (2008). Efficient methods to compute genomic predictions. J. Dairy Sci. *91*, 4414–4423. https://doi.org/10.3168/jds.2007-0980.

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Seeds of *Arabidopsis thaliana* Col-0 and other ecotypes	*Arabidopsis* Biological Resource Center (Ohio State University, Columbus, Ohio, USA)	https://abrc.osu.edu/
Chemicals, peptides, and recombinant proteins		
MURASHIGE & SKOOG BAS SALT MIX	PhytoTechnology Laboratories, USA	M524
Chloral hydrate	Sigma-Aldrich	C8383; CAS: 302-17-0
Critical commercial assays		
GUS staining kit	HUAYUEYANG biotechnology, Beijing, China	GT0391
Plant Genomic DNA Extraction Kit	Tiangen Biotech, Beijing, China	DP342
Universal DNA Fragmentase Kit	Annoroad, Beijing, China	AN200102
Universal DNA Library Prep Kit	Annoroad, Beijing, China	AN200101
RNeasy Plant Mini Kit	Qiagen, Hilden, Germany	Cat#74904
RNase-Free DNase Set	Qiagen, Hilden, Germany	Cat#79254
NEBNext® Ultra™ RNA Library Prep Kit for Illumina	New England Biolabs, Ipswich, Massachusetts, USA	Cat#E7530
Deposited data		
Raw and analyzed data	This paper	GEO: GSE227500
Code	This paper	https://github.com/WenwenLiu54
Experimental models: Organisms/strains		
$Promoter_{Col\text{-}0\ CYCB1;1}$::*DB*-EGFP-*GUS*	This paper	N/A
$Promoter_{Per\text{-}1\ CYCB1;1}$::*DB*-EGFP-*GUS*	This paper	N/A
$Promoter_{Col\text{-}0\ CYCB1;1}$::*DB*-EGFP-*GUS* ×$Promoter_{Per\text{-}1\ CYCB1;1}$::*DB*-EGFP-*GUS* F_1	This paper	N/A
Recombinant DNA		
Plasmid: pKGWFS7.0	BioVector NTCC Inc.	N/A
Software and algorithms		
fastp v0.20.0	Chen et al.[25]	https://github.com/OpenGene/fastp
bwa-mem v0.7.17	Li[26]	https://github.com/lh3/bwa
Picard v2.23.4	N/A	http://broadinstitute.github.io/picard/
GATK4	N/A	https://gatk.broadinstitute.org
PLINK v1.9	Purcell et al.[27]	https://zzz.bwh.harvard.edu/plink/
HISAT2 v2.1.0	Kim et al.[28]	https://daehwankimlab.github.io/hisat2/
Sva	Leek et al.[29]	https://www.rdocumentation.org/packages/sva/
edgeR v3.40.2	N/A	https://bioconductor.org/packages/release/bioc/html/edgeR.html
rMVP v1.0.6	Yin et al.[30]	https://github.com/xiaolei-lab/rMVP
rrBLUP v4.6.1	Endelman[31]	https://www.rdocumentation.org/packages/rrBLUP/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Matrix eQTL v2.3	Shabalin[32]	https://github.com/andreyshabalin/MatrixEQTL
clusterProfiler v3.10	Yu et al.[33]	https://www.rdocumentation.org/packages/clusterProfiler/

STAR★METHODS

KEY RESOURCES TABLE

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Xing Wang Deng (deng@pku.edu.cn).

Materials availability

Materials generated in this study will be made available on request.

Data and code availability

All original sequences, gene expression, and SNP variant data generated in this study have been deposited in the Gene Expression Omnibus database (https://www.ncbi.nlm.nih.gov/geo) under accession number GSE227500. All code for the bioinformatics analysis is publicly available on github at https://github.com/WenwenLiu54. Additional data, such as raw image files, that support this study are available from the corresponding authors upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Seeds of *Arabidopsis thaliana* Col-0 and other ecotypes were obtained from the *Arabidopsis* Biological Resource Center (Ohio State University, Columbus, Ohio, USA). Using hand pollination, we generated half-sibling hybrids with Col-0 as the maternal parent and each of other ecotypes as the paternal parent.

METHOD DETAILS

Trait phenotyping

To measure leaf area heterosis levels, we measured the PM cell area and number of the first or second true leaves of several representative 15-DAS seedlings from each half-sibling hybrid and its two parents (Col-0 and the paternal ecotype), all of which had been planted on the same Murashige and Skoog (MS) plate. Seeds were surface-sterilized and stratified in the dark at 4°C to break dormancy before sowing. Sampled true leaves with each genotype were treated with ethyl alcohol overnight to remove chlorophyll, and then were transferred to a chloral hydrate solution to better clear the leaves.[34] We used chloral hydrate solution as a mounting medium, mounted each leaf on a microscope slide, and then used microscopes equipped with a camera to acquire images of the leaves and PM cells. To measure each genotype's leaf area, five representative leaves were selected for each genotype and viewed under a binocular microscope (Leica M165 C) with calibration performed by photographing a ruler at the same magnification for each leaf, and then the leaf area was obtained using ImageJ (https://imagej.nih.gov/ij/). Next, we used a differential interference contrast microscope (Leica DM6000) at 40× magnification and ImageJ to measure the PM cell area of the same five leaves for each genotype. Images were acquired while visualizing the PM cells approximately midway between the tip and bottom of each leaf and halfway between the leaf margin and the midvein. PM cell number per leaf was calculated for each genotype by dividing the average leaf area by the average PM cell area. Finally, to remove a potential batch effect among the half-sibling crosses planted on different plates, the leaf area and PM cell area and number for each genotype were normalized to that of the Col-0 planted on its shared plate. For each trait, MPH was calculated as (F_1-MP)/MP, where F_1 is the hybrid phenotype value and MP is the mid-parent phenotype value. BPH was calculated as (F_1-BP)/BP, where BP is the better parent phenotype value.

Mitotic cell reporter line generation and GUS staining

To identify mitotic cells in the first true leaf, a construct (p*CYCB1;1*::DB-EGFP-GUS) using plasmid pKGWFS7.0 to fuse the ecotype-specific promoter and destruction box of the CYCB1;1 to GUS, which is expressed only in cells during G2M and is destroyed in mitosis, was generated for Col-0 and Per-1 to obtain their transgenic mitotic cell reporter lines, respectively. The F_1 hybrid reporter line was obtained by hand pollination with the Col-0 reporter line as maternal parent. The construct assay was the same as a published protocol.[35] Histochemical GUS staining was performed on the seedlings of Col-0 and Col-0×Per-1 F_1 hybrid 7 days after sowing using a GUS staining kit (GT0391, HUAYUEYANG biotechonology, Beijing, China). And the GUS-staining intensity was examined with a Leica microscope equipped with a digital camera and software imageJ.

DNA preparation and whole-genome resequencing

Surface-sterilized seeds from five *Arabidopsis* ecotypes (Er-0, Bch-1, Me-0, Old-1, Le-0) were stratified in the dark at 4°C and then sown on MS plates containing 1% sucrose. After 10 days, the seedlings were planted in pots with soil and grown under long-day conditions for a few weeks. A few representative young leaves of each ecotype were sampled and quick-frozen in liquid nitrogen; DNA was extracted using a Plant Genomic DNA Extraction Kit (DP342, Tiangen Biotech, Beijing, China). Subsequent DNA sequencing libraries were constructed using a Universal DNA Fragmentase Kit (AN200102, Annoroad, Beijing, China) and the Universal DNA Library Prep Kit (AN200101, Annoroad). Each library was sequenced on a NovaSeq 6000 platform (Illumina, San Diego, California, USA) to generate 150-nucleotide paired-end reads (~5 Gb of clean data). Whole genome resequence data for the other ecotypes were downloaded from the NCBI Sequence Read Archive, accession number SRP056687.[36]

RNA preparation and sequencing

RNA-seq samples from the SA 6 DAS or the first TL 15 DAS were prepared for the 96 half-sibling cross population in 11 batches, and each batch included one shared sample of Col-0 as a reference for removing the batch effect in RNA-seq data analysis. For each batch, samples were prepared as follows. Seeds of Col-0, a few ecotypes, and their half-sibling hybrids were cultured and grown as described for trait phenotyping above. Either the SA 6 DAS or the first two TLs 15 DAS were collected from several representative seedlings of each genotype and then quick frozen in liquid nitrogen. For each organ of each genotype, we sampled one biological replicate, ultimately obtaining 406 samples in 11 batches, 203 RNA-seq samples for each organ, including 96 from hybrids, 96 from ecotypes, and 11 from Col-0 (Table S2). Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) and the DNA was digested using the RNase-Free DNase Set (Qiagen). Next, we constructed libraries with the NEBNext® Ultra™ RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, Massachusetts, USA) and sequenced all of the RNA-seq samples in only one batch on a NovaSeq platform (Illumina), ultimately generating 150-nucleotide paired-end reads (~5 Gb of clean data for each sample).

Genome resequencing data analysis and variant calling

Quality control was performed for the DNA read data for each of the 96 ecotypes using fastp v0.20.0.[25] Using BWA-MEM v0.7.17,[26] the high-quality reads that remained after filtering were mapped to the *Arabidopsis* Col-0 reference genome obtained from TAIR10. PCR duplicate reads were removed by Picard v2.23.4, and only unique mapped reads with mapping quality >20 were retained for variant calling using GATK4 tools as follows. First, the read base quality was recalibrated using BaseRecalibrator and then ApplyBQSR with the parameter "–known-sites" was set to the published, high-quality variation in the 1001 Genomes Project.[36] Next, variants were detected using HaplotypeCaller, CombineGVCFs, and GenotypeGVCFs, and then high-quality SNP variants were selected using SelectVariant and a threshold of "QD < 2.0 || MQ < 40.0 || FS > 60.0 || SOR > 3.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0". The SNPs were filtered using PLINK v1.9[27] using the parameters "–geno 0.05" and "–maf 0.05", and only biallelic nuclear genome SNPs were retained. Consequently, we obtained 1,067,509 SNPs for the 96 ecotypes.

RNA-seq data processing

Once again, we used fastp (v0.20.0)[25] to obtain high-quality reads, and these reads for each sample were mapped to the TAIR10 *Arabidopsis* Col-0 reference genome using the aligner HISAT2 v2.1.0[28] with the parameter "–score-min L, 0, -0.4", which reports only up to a 1.5-base-mismatched alignment. In addition,

the multiple mapped reads were filtered to retain only unique mapped reads for quantification of gene expression. Next, the read counts mapped to each gene were quantified using featureCounts according to the gene structure annotation information in the *Arabidopsis* gtf from TAIR10. Potential batch effects for each tissue's (SA 6 DAS or TL 15 DAS) 203 samples were corrected using the ComBat-seq function in the sva software package.[29] Finally, transcript per million (TPM) gene expression was calculated by normalizing gene length and library size among all 406 samples; the resulting data were normalized using the trimmed mean of m-values (TMM) method to obtain final gene expression values. In our analysis, genes with expression levels < 1 (TPM < 1) in more than 98% of samples were discarded, as were organelle genes. This left 21,014 genes in the reference nuclear genome, and these genes' expression profiles were used for sample Pearson correlation analyses. Using the prcomp function in R, we performed principal components analysis (PCA) on the expression profiles of the 1,000 genes that showed the largest TPM variation (standard deviation) among the 406 samples. Both sample correlation calculation and PCA showed that batch effect interference in gene expression was removed successfully (Figure S2). Differential gene expression analysis was performed using edgeR v3.40.2 with a cutoff of FDR < 0.05.

TWAS

Genotype imputation was conducted based on the 1,067,509 SNPs in the 96 ecotypes, and the kinship among them was calculated using the VanRaden method[37] in rMVP v1.0.6.[30] TWASs were then performed between the change in gene expression and the corresponding trait changes in leaf area, cell number, and cell area of the hybrids versus each of their parents (numbered as TWAS I–XII). For each TWAS, the TPM ratio for gene expression in a hybrid versus its parent was calculated for each gene, after which the genes were filtered without the TPM ratio varying among the 96 hybrids. Genes with a TPM ratio of 0 in at least 5% of the 96 hybrids were also removed. Next, we conducted TWAS through the GWAS function in rrBLUP v4.6.1[31] with default settings using the quantile normalized TPM ratio (–1∼1) based on a mixed linear model that considers kinship between different ecotypes. The cutoff for determining significantly associated genes was $p < 0.01$. Correlation levels between significantly associated gene expression ratios and trait ratios of hybrids versus parents were calculated using the Pearson coefficient.

eQTL identification

The significantly associated genes identified by TWAS I and VII underwent eQTL analysis using Matrix eQTL v2.3[32] based on a linear model, with the TPM gene ratio of the hybrids versus that of the common maternal parent (Col-0) in the SA 6 DAS or TL 15 DAS as the expression trait and the 1,067,509 SNPs of the 96 ecotypes as the genotype. The cutoff for determining significant SNPs was $p < 1 \times 10^{-6}$. The significant SNPs for each gene were grouped into clusters with a maximum distance of 10 kb between two consecutive SNPs, and only the cluster with more than three significant SNPs was retained as a putative eQTL. For each eQTL, the significant SNP showing the lowest *p*-value was identified as the lead SNP. Local (*cis*) and distant (*trans*) eQTLs were distinguished according to the distance of each SNP-eGene pair. A *cis*-eQTL had a distance between the eGene transcription start site and the lead SNP of < 5 Mb, whereas that distance for a *trans*-eQTL was >5 Mb.

GO enrichment analysis

GO enrichment analysis of the associated genes identified by the TWASs was performed with GO Biological Process Complete using clusterProfiler v3.10[33] with all the associated genes from TWAS I-VI or TWAS VII-XII as a reference and a hypergeometric test with FDR correction (adjusted p < 0.05) applied. Gene description information and GO terms were assigned according to TAIR (https://www.arabidopsis.org/) and the corresponding R package, org.At.tair.db.

QUANTIFICATION AND STATISTICAL ANALYSIS

The statistical analysis details have been described in methods and results.