

# Enhanced photosynthesis and carbohydrate metabolic capability contributes to heterosis of cotton (*Gossypium hirsutum*) hybrid 'Huaza Mian H318' as revealed by genome-wide gene expression analysis

**Yuanhao Ding**

Hainan University

**Rui Zhang**

Huazhong Agriculture University

**Longfu Zhu**

Huazhong Agriculture University

**Maojun Wang**

Huazhong Agriculture University

**Yizan Ma**

Huazhong Agriculture University

**Daojun Yuan**

Huazhong Agriculture University

**Nian Liu**

Huazhong Agriculture University

**Haiyan Hu**

Hainan University

**Ling Min** (✉ [lingmin@mail.hzau.edu.cn](mailto:lingmin@mail.hzau.edu.cn))

Huazhong Agriculture University <https://orcid.org/0000-0003-4278-0626>

**Xianlong Zhang**

Huazhong Agriculture University

---

## Research article

**Keywords:** Cotton, Heterosis, RNA-Seq, Photosynthesis, Carbohydrate

**Posted Date:** June 17th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-33557/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at BMC Genomics on April 17th, 2021. See the published version at <https://doi.org/10.1186/s12864-021-07580-8>.

# Abstract

## Background

Heterosis has been exploited for decades in different crops for the dramatic increase in yield but less documented molecular evidences reported in cotton.

## Results

An elite cotton hybrid variety 'Huaza Mian H318' (H318) and its parental lines was used to explore the source of its yield heterosis. A four years' investigation of yield related traits showed that the boll number of the H318 was stable higher compare with its two parents whether in the climate suitable or unsuitable years. Besides, the hybrid H318 grows faster and showed higher fresh and dry weight than its parental lines at the seedling stage. Transcriptome analysis of seedlings identified 17,308 differential expressed genes (DEGs) between H318 and its parental lines, and 945 extremely changed DEGs were screened out for later analysis. In which, most DEGs (931/945) were gathered between H318 and its paternal line (4-5), but only 78 DEGs were found between H318 and its maternal line (B0011), which implied that H318 display more similar gene action with its maternal parent. GO and KEGG analysis showed that these DEGs highly enriched in photosynthesis and carbohydrate metabolic processes, the expression level of these DEGs was significantly higher in the H318 relative to its parental lines, which implies that sucrose and starch biosynthesis were enhanced in H318.

## Conclusion

The enhanced photosynthesis and carbohydrate metabolic capability contributes to the heterosis of H318 at seedling stage, and establish the material foundation for later higher boll setting rates in the complex field environments.

## Background

Heterosis, a mysterious and widespread biological phenomenon, has been scientifically investigated but few explicit descriptions of its molecular basis have been documented [1]. Heterosis is extremely prevalent in plants and animals but exhibits differently at different developmental stages in different species and across different traits [2]. Moreover, heterosis has benefited production for over a century, causing dramatic improvements in crop yield; some major crops exhibit heterosis in production such as rice, maize and rapeseeds [3].

Despite lack of a unified theory, several valuable hypotheses have been presented and widely accepted by biologists [4]. Dominance and overdominance are two classical genetic explanations, especially for single gene or single trait heterosis [5, 6]. Additionally, studies in rice over the last 20 years have developed two

different genetic models for heterosis, epistasis and pseudo-dominance [7, 8]. Common opinion about heterosis is that the genetic diversity of the corresponding inbreds of F1 hybrids is a key factor for heterosis performance. In *Arabidopsis*, the genetic distance of the inbred parents is proportional to the hybrid vigor [9]. However, the polyploid hybrids always exhibit higher heterosis than the diploid one, although crosses between genetically similar rice can also produce hybrids with significant heterosis [10, 11]. Thus, the hypotheses are insufficient in explaining all of the evidence regarding heterosis [2]. Presentation of a perfect unified theory to explain heterosis is difficult because of the genetic complexity and genomic diversity of different species.

In past years, many transcriptome studies have provided insights into the molecular basis of heterosis in different species like rice [8, 12], *Arabidopsis* [13, 14] and maize [15, 16]. In these studies, a vast number of DEGs associated with different biochemical pathways have been found related to heterosis. However, few specific biological pathways have been demonstrated to play key roles in heterosis, and these expression-changed genes have had a random distribution among biochemical pathways. In addition, genome-wide changes in gene expression have displayed different gene actions (additive and non-additive) in different studies and crops [9], which means that the differing gene actions in hybrids are related to the genetic distance of the parents [17]. Moreover, genome-wide studies of transcriptomes identify which biological pathways are changed in hybrids, including energy, metabolism and biomass, phytohormone signaling and stress responses [18], which aid our understanding of heterosis.

Cotton (*Gossypium spp.*) is an important cultivated crop for the economic value of its fibre. The acreage of cotton in China is keeping decrease in recent years. The challenge in China is using less land to produce more cotton to save acreage for producing urgently demanded grains. Hybrid cotton is chosen to maintain total cotton production capability with less land. As early as 1894, it was reported that hybrids between *Gossypium hirsutum* L. and *Gossypium barbadense* L. exhibited great heterosis in vegetative growth [19]. Subsequently, several attempts were made to apply similar principles in intraspecific cottons [20, 21]. Previous studies have shown that heterosis in cotton may be correlated with vegetative heterosis at the seedling and squaring stages [22]. During flowering time, environmental influences were correlated with the final yield of cotton in a hybrid [23]. Although increasing cotton heterosis research data has accumulated, it is far from a systematic understanding of heterosis.

H318, a released hybrid cotton variety derived from a cross between B0011 and 4–5, was largely adopted in cotton production in Southern China due to its high yield and wide range of adaptations. The hybrid and its parents were employed to investigate heterosis performance through comparing the genome-wide gene expression profiles by RNA-Seq technology and physiological analysis. A large number of DEGs were found to be involved in photosynthesis and carbohydrate metabolic processes. Moreover, the detection of physiological rate, sugar content and gene expression related to photosynthesis and carbohydrate supported to the findings from RNA-Seq. Thus, the enhanced photosynthesis and carbohydrate metabolic capability were considered contributes to the heterosis performance of H318 at seedling stages, which might also support the material foundation for the hybrid's high adaption to the complex field environments and finally benefit its higher boll setting rates in different years.

# Results

## **Increased boll number contributes to the yield heterosis of H318 in different years.**

H318, as an elite cotton cultivar, showed an obvious yield increase over the control variety, with an average 3,664.5 kg/ha seed cotton and 1,521 kg/ha lint yield production (9.2% higher than control) in two years (2007–2008) of regional testing.

To explore the effects of heterosis on yield, we conducted a four-year field investigation about the traits like the number of bolls, the number of fruit site, abscission rate and the number of fruit branches (Fig. 1). At the same time, considering the climate influences to plant growth and production, the climatic changes from May to October (the whole growth period of cotton) in Wuhan from 2010 to 2013 were also obtained from Statistics Bureau of Hubei Province (Figure S1).

In 2010, a long-term lasting rainfall and large volume of precipitation in July (Figure S1 C, D) extremely affected the fruit setting efficiency (July and August are the flourishing flowering stage), which causes high abscission rate of 4–5 (paternal line) and B0011 (maternal line) (Fig. 1C). However, relatively lower abscission rate and more fruit sites was found in the H318 (Fig. 1B-C), which might be the reason for its final higher boll numbers (Fig. 1A).

In later relatively ideal climate conditions in 2011, more water in growing stage (June in Figure S1 C, D); suitable sunshine hours, raining and warming temperature during flourishing flowering stage (July and August in Figure S1 A-D), little difference could be found in the abscission rates between the hybrid and both parents (Fig. 1B), but much more fruit sites were produced in hybrid H318, which directly causing the increased boll number of H318 (Fig. 1A).

In 2012 and 2013, long-term high temperature stress occurred in July and August (Figure S1 A) caused higher abscission rate than the other years, but H318 still showed much lower abscission rate than its parents and more bolls were produced at last (Fig. 1A, C). No obviously differences were found about fruit site in 2012 and 2013, and fruit branches have no difference in all years (Fig. 1B, D). Thus, we speculate that the higher yield production of H318 directly derives from the stable higher number of bolls, and which is the result of the increased fruit sites or decreased abscission rate in different years.

## **Hybrid H318 shows heterosis on biomass and growth speed at seedling stage**

Besides the yield heterosis, H318 also showed growth vigor than its parents at seedling stage (Fig. 2A). The fresh and dry weight (the whole plant) of H318 and its parents were measured at two-leaf stage, and results showed that H318 had obviously higher fresh weight than its parents and dry weight than 4–5 (Fig. 2B). Moreover, the fresh weight of B0011 was higher than 4–5, but without obviously changes about dry weight. To evaluate the plant growth status in detail, the cotyledon area was calculated every 2 days after cotyledons spreading until 14 DAS (Fig. 2C). At 6 DAS, the hybrid H318 had just a little larger area of cotyledon than both parents. After the 8 DAS, the cotyledon area was significantly increasing and reached its maximum area until the 14 DAS. The cotyledon area of H318 remarkably kept largest than its parents

after 8 DAS. These results suggested that H318 have obviously heterosis about biomass production and growth speed at seedling stage.

### Global analysis of differential gene expression of H318 and its parents

To analyse the global gene expression patterns of these three genotypes, a total of 9 RNA sequencing libraries (containing 3 independent biological repeats for each genotype) were constructed for Illumina sequencing using the whole seedlings of 8 DAS. In total, more than 43,000,000 clean reads (occupy 95% of raw reads) were generated from each sample (Table S1). All clean reads then were mapped to cotton genome (*G. hirsutum* TM-1 (AD)1) to obtain the genome location and annotation information (Table S2). Nearly 89% reads were mapped to genome and more than 80% reads were uniquely mapped. Most of mapped reads (> 85%) were found located in exon; reads in intergenic region occupied 11%-12%; and the rest (2.5%) located in intron (Figure S2).

The reads mapped to genome then were used for transcript assembling and gene expression level calculation. The FPKM of all genome annotated genes (70,478) and novel genes (6,564) were listed in Table S3. The correlation analysis showed perfect consistency between biological repeats (Figure S3). The differential expressed genes (DEGs) then were screened out with condition of P values < 0.05. Finally, a total of 17,308 DEGs containing 16,407 annotated genes and 1,261 novel genes were found between three genotypes (Table S4), and much less DEGs (306) between H318 and B0011 were found relatively (Figure S4).

To remove the low expressed or differences DEGs, a more stringent criterion (FPKM > 50 and  $\log_2(\text{fold changes}) > 1$  in at least one sample) was performed and left 945 DEGs (Fig. 3A). Among these, 931 DEGs were found between H318 and its paternal line (4–5), most DEGs (532) were up-regulated in H318; only 78 DEGs were found between H318 and its maternal line (B0011, Fig. 3B). The DEGs distribution in three genotypes showed that most DEGs (844/945) is different expressed between H318 and 4–5 or B0011 and 4–5; 77 DEGs were shared in three genotypes (Fig. 3C). These results indicated that the gene expression pattern of H318 is more similar to its maternal line and difference to its paternal line.

To validate the reliability of RNA-Seq, 26 DEGs was selected randomly for qRT-PCR confirmation. A correlation analysis was made using data from RNA-Seq and qRT-PCR (Fig. 3C). And the result shows that the  $R^2$  value reaches 0.9791, indicating the high quality for RNA-Seq.

## GO And KEGG Analysis

To gain a deeper understanding of the potential functions of these DEGs (945) in the three genotypes, GO and KEGG analyses were carry out for gene functional classification (Figure 4, Table S5 and Table S6). For GO analysis, the most significant enriched biological processes were Photosynthesis related processes (red, 67 DEGs); Energy production process (blue, 43 DEGs); Oxidation-reduction process (green, 84 DEGs); Carbohydrate metabolic related processed (yellow, 54 DEGs) and Metabolic process (purple,

341 DEGs) (Figure 4, Table S5). At the same time, Oxidoreductase activity (green) and Photosystem related components (red) mostly enriched in Molecular function and Cellular component respectively. For KEGG analysis, the significant enriched pathways were various metabolic related pathways (purple) with 211 DEGs (Figure 4, Table S6). And Photosynthesis (red,) and Carbon metabolism related pathways (green) were also found enriched with 45 DEGs and 62 DEGs respectively.

### **Photosynthesis, carbohydrate metabolic and oxidation-reduction pathways are enriched in H318.**

From above GO and KEGG analyses, photosynthesis, oxidation-reduction and carbohydrate metabolic were speculated significantly changed between H318 and its parental lines. Of the 67 DEGs involved in photosynthesis, all of them were up-regulated in H318 and B0011 relative to 4-5 (Table S7). However, most of these DEGs were slightly down-regulated (32) or no change (31) in H318 compared to maternal line (B0011, Table S7). In oxidation-reduction process (84 DEGs), 68 DEGs were highly up-regulated in H318 relative to 4-5, and 55 DEGs slightly up-regulated in H318 relative to B0011 (Table S7). For carbohydrate metabolic process (54 DEGs), 30 and 37 DEGs were up-regulated in H318 relative to 4-5 and B0011 respectively. Results indicated that the photosynthesis pathway were both enhanced in H318 and B0011 related to 4-5, oxidation-reduction process and carbohydrate metabolic process were enhanced in H318. These results indicated that hybrid H318 might have stronger capability on photosynthesis, carbohydrate metabolic and oxidation-reduction than its parents especially to the paternal line. We speculated that enhanced capability of these functions might contribute to higher fresh and dry weight and faster growth speed of H318 at seedling stage (Figure 2).

### **Higher photosynthesis rate and sugar accumulation were found in H318 at seedling stage.**

To validate the authenticity of this inference, the photosynthesis rate and the content of sucrose and starch at two leaf stage were assessed in both H318 and its parental lines (Figure 5). Results showed that the photosynthesis rate was significantly enhanced in H318 relative to its parental lines especially to the paternal line 4-5 (Figure 5 A). Correspondingly, the content of sucrose and starch were also found higher in H318 (Figure 5 B). Moreover, the relative expression level of six genes related to starch biosynthesis and photosynthesis process was also detected and showed in Figure 5 C including *GRANULE BOUND STARCH SYNTHASE 1 (GBSS1)*, *RUBISCO ACTIVASE (RCA)*, *PROTON GRADIENT REGULATION 5 (PGR5)*, *PHOTOSYSTEM II SUBUNIT X (PSBX)*, *PHOTOSYSTEM I SUBUNIT L (PSAL)* and *PHOSPHORIBULOKINASE (PRK)*. The expression trends of these genes were highly consistent to the photosynthesis rate and the content of sucrose and starch, which indicated an enhanced photosynthesis rate and sugar biosynthesis in H318 relative to its parental lines. All these results implied that stronger photosynthesis and metabolic processes contributes to the heterosis in H318 at seedling stage.

## **Discussion**

To produce more cotton using less land, heterosis application seems imperative for improving yield output per unit in China. Differential gene expression studies between hybrids and their parents have confirmed heterosis related effects in other crops, but little is known about cotton [11, 15]. In this study,

we used RNA-Seq technology to analyse the gene expression differences between a high yield cotton hybrid variety 'Huazamian H318' and its parents for illustrating the effects of heterosis in cotton.

### **Comprehensive stress tolerance and boll setting rate contribute to the yield heterosis of H318**

From our results about yield investigation, higher boll numbers is thought as the source of the yield heterosis performance of the hybrid H318 (Fig. 1). What's interesting is that H318 stably produced more bolls than its parental lines in the continuous four years whether under optimal or unsuitable climate conditions. So, we speculated that the stable higher boll setting rate of H318 might be derived from its potential capability of balancing the tradeoff between stress responses and growth. On the other hand, in the hot summer of 2013, the paternal line (4–5) of H318 was found more sensitive to high temperature and showed serious male sterility than H318 and B0011 (Figure S5), which implied the stress tolerance of H318 might inherit from its maternal line (B0011).

In past years, transcriptome researches about heterosis found that many stress response biological processes were enriched in F1 hybrids in different species [24–26]. In maize, a drought-stress tolerance gene *ZAR1* was found benefit to yield heterosis performance [27]. And some TCA cycle intermediates was reported contribute to the heterotic phenotype under freezing stress [28], and soluble sugar and flavonol content also have been shown to have a strong relationship with heterosis performance [29]. A viewpoint about balancing the tradeoff between a rapid requirement for stress responses and long-term maintenance of growth vigor was proposed [30]. Thus, we think that the comprehensive stress tolerance and growth vigor both benefit to the yield heterosis performance, and changing with the circumstance.

### **Enhanced photosynthesis and carbohydrate metabolism resulted in heterosis of H318 at seedling stages.**

In *Arabidopsis*, various processes or pathways have been reported related to the biomass heterosis performance of F1 hybrid like photosynthesis, PIF4 controlled auxin pathway, stimulus-responsive pathways and *et al.* [31–33]. And GA metabolism was found related to biomass heterosis in rice seedlings [34]. Our results showed that photosynthesis and carbohydrate metabolic process contributed to the hybrid vigor of H318 (Fig. 4, 5). For the samples for sequencing were obtained under a relative ideal growth condition, and none stress response related pathways were accumulated that seems natural although H318 showed high temperature tolerance at flowering stage. We speculated that the increased photosynthesis and energy production was the basis of biomass heterosis performance of hybrid, and many researches supporting this point [18, 33]. It is widely accepted that gene-gene and gene-environment interactions together determine the final phenotype of a plant. In the whole life of a plant, it can be seemed as a process conflicted with various abiotic and biotic stressors from the environment, thus stress-responsive pathways were found related to heterosis in many researches [31, 35]. However, in a relative suitable growth condition, higher expression of stress-responsive genes will waster a lot of energy, thus balancing the tradeoff between a rapid requirement for stress responses and long-term maintenance of growth vigor seems more economical for plants. So enhanced photosynthesis and carbohydrate metabolism not only directly result in the biomass heterosis of H318 at seedling stages but also established material foundation for the higher boll setting rates of H318 at flowering stages.

## Conclusions

In this study, an elite cotton hybrid variety H318 and its parental lines was used to explore the source of its yield heterosis. Four years' investigations of yield related traits and transcriptome analysis of seedlings between H318 and its parental lines revealed that photosynthesis and carbohydrate metabolic processes were enhanced in H318, which contributes to the heterosis performance of H318 not only at seedling stages but establish the material foundation for its higher boll setting rates in the complex field environments for many years.

## Methods

### Plant material, growth conditions and trait investigation

All plant materials used in this study are developed by the Group of Cotton Genetic Improvement (GCGI), Huazhong Agricultural University. Cotton cultivars 4–5 and B0011 were grown under normal field conditions in Wuhan, and hybrid H318 seeds were developed by hand-pollinating the pollen from 4–5 to the emasculated flowers of B0011 before 10 o'clock in the summer. Meanwhile, the two parents were self-pollinated to maintain the genetically pure lines.

Seeds of H318 and its parents were sowed in mid-April every year to investigate the performance of the traits related to heterosis with 12 plants for each line and 4 lines per plot in a randomised design, repeated in triplicate. Plant spacing was set at 30 cm, and the line spacing was 1 m. The traits like boll number, plant height, fruit branches and fruit sites were investigated at three time points: July, August and September. In October, 20 bolls from the mid nodes were harvested from the material in each plot to calculate the single boll weight, lint percentage and lint index.

To evaluate heterosis at the seedling stage, cotton seeds were germinated in an incubator at 28 °C for 2 days. They were then transplanted into sand by choosing 30 seedlings with similar radicle lengths to avoid the effects of seed vigor. The plants were watered every two days and grown in a controlled growth chamber (16/8 h photoperiod, day and night, at 28 °C, with light intensity of  $200 \mu\text{mol m}^{-2}\text{s}^{-1}$ ) until 14 days after sowing (DAS). Cotyledon areas were measured from 6 DAS to 14 DAS according to previously reported methods [36]. Fresh and dry weights were calculated at 14 DAS; 5 random plants were harvested as one replicate for each type, and 3 replicates were prepared for each material. And another 5 whole plants (aerial and underground parts) for each genotype were harvested and kept in liquid nitrogen for RNA extraction, 3 replicates were set. Photosynthesis was measured with the first expanded true leaf by LI-6400 XT (LI-COR, USA).

### RNA Isolation And Sequencing Library Construction

Samples were ground in liquid nitrogen, and total RNA was extracted by a modified guanidine thiocyanate method [37], and quantitated by a Nanodrop 2000 instrument (Thermo Scientific, USA). Three

total RNA samples (20 µg for each sample) were sent to the Novogene (Beijing, China) for construction of libraries and were then sequenced using the Illumina Genome Analyzer (HiSeq-PE150).

In briefly, the high-quality RNA samples (tested by Agilent 2100 Bioanalyzer system) were firstly enriched to obtain mRNA with poly (A<sup>+</sup>) by using magnetic beads contained OligdT. Then enriched mRNA were broken into short fragments by fragmentation buffer. The first-strand cDNAs were synthesised using random hexamers and reverse transcriptase. The second-strand cDNAs were generated by *Escherichia Coli* polymerase I. And the final cDNA libraries were ready after a round of purification, terminal repair, A-tailing, ligation of sequencing adapters, size selection and PCR enrichment. Completed libraries then were used for sequencing after a series of quality testing.

## Bioinformatics Analysis And Statistic Screening

The original raw data from Illumina HiSeq™ PE150 are first filtered to generate clean data by removing reads containing adapters or low quality. The clean data then were mapped to cotton genome (*G. hirsutum* TM-1 (AD)1) [38] to obtain the genome information of each reads. And the reads mapped to exon or genome were used to calculate the expression level of genes showing with FPKM (expected number of Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced). Differential expressed genes (DEGs) between samples were obtained by screening conditions of padj (P-value adjusted) < 0.05. Relatively critical screening conditions (FPKM > 50, |log<sub>2</sub>Ratio| > 1) were taken to ensure reliability for in-depth analysis.

### GO and KEGG analysis

Gene ontology (GO) term analysis ([www.geneontology.org](http://www.geneontology.org)) was applied to predict the functional category distribution frequency of screened DEGs using Blast2GO software. The functional categories with FDR < 0.05 were kept for later analysis. To understand the differences in pathways, KEGG analysis was also performed with KOBAS 3.0 (<http://kobas.cbi.pku.edu.cn/>). All pathways were screened with conditions of FDR < 0.005.

## qRT-PCR Analysis

A sub set of DEGs (20) were selected to validate the RNA-Seq result through qRT-PCR. Gene-specific primers were designed though Primer Premier 5.0 and synthesized by company of Genscript Bioscience (Nanjing, China). The cDNA was generated from 3 µg RNA samples by Superscript III RT (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. 20 µl qRT-PCR reactions were performed with 9.6 µl of 100 × diluted cDNA, 0.2 µl forward and reverse gene-specific primers, and 10 µl SYBR Green PCR Master Mix (Applied Biosystems) and then were run in four duplicates on an ABI Prism 7500 Sequence Detection System (Applied Biosystems). Thermal cycling was set as 30 s at 95 °C, followed by 40 cycles of 5 s at 95 °C and 35 s at 60 °C. Relative quantitation of gene expression was calculated and normalized

through using GhUBQ7 (GenBank accession number: DQ116441) as an internal standard according to the previous study [39].

## Sucrose And Starch Measurement

To validate the changes of photosynthesis and carbohydrate metabolites in hybrid, approximately 50 mg plant tissues (the entire plants were fully ground in liquid nitrogen) were obtained from each genotype and incubated in sterile ddH<sub>2</sub>O at room temperature for 24 h for extraction dissolved. The subsequent procedures were performed essentially according to previous reports [40]. The supernatant was used for the measurements of sucrose contents. The remaining pellets were used to assay starch content using an improved colorimetric method [41].

## Declarations

## Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Availability of data and material

The RNA sequencing data used in this study can be found in the National Center for Biotechnology Information (NCBI) SRA database under following accession number: PRJNA393079.

### Competing interests

The authors declare that they have no competing interests.

## Funding

This work was supported by the National Key Research and Development Program of China (2016YFD0101402; grant recipient: Ling Min).

## Authors' contributions

YD, LM and XZ initiated the research. LZ, LM and XZ designed the experiments. YD performed the four years' investigations of yield-related traits. YD, RZ, NL and HH performed molecular experiments. YD, MW,

YM and DY performed bioinformatics analysis. YD drafted the manuscript. YD, HH, LM and XZ finalized the manuscript. All authors contributed in the interpretation of results and approved the final manuscript.

## Acknowledgements

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

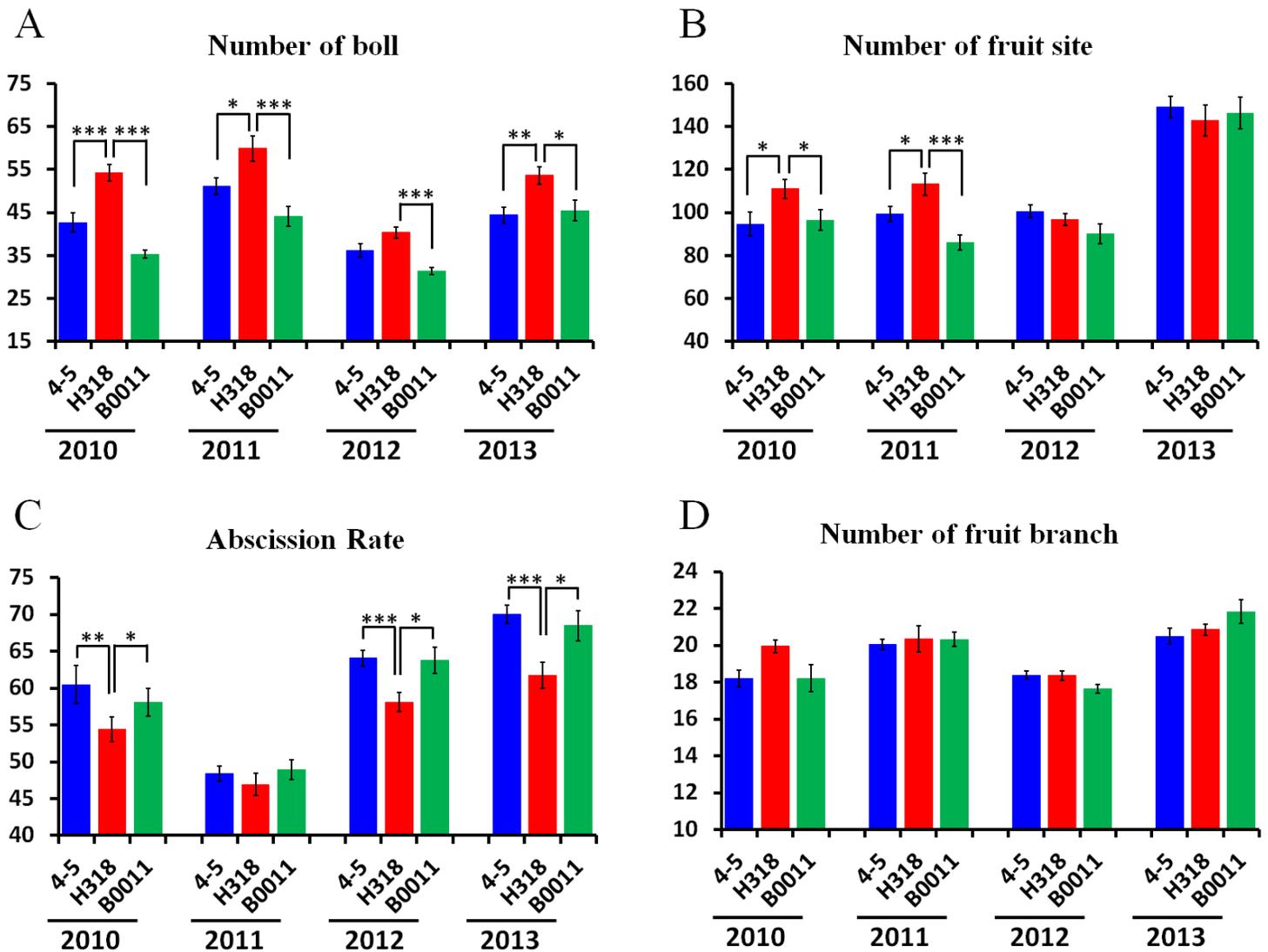
## References

1. Darwin CR. **The effects of cross- and self-fertilization in the vegetable kingdom.** *John Murray* 1876.
2. Chen ZJ. Genomic and epigenetic insights into the molecular bases of heterosis. *Nat Rev Genet.* 2013;14(7):471–82.
3. Fu DH, Xiao ML, Hayward A, Fu Y, Liu G, Jiang GJ, Zhang HH. Utilization of crop heterosis: a review. *Euphytica.* 2014;197(2):161–73.
4. Stupar RM, Springer NM. Cis-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F1 hybrid. *Genetics.* 2006;173(4):2199–210.
5. Shull GH. The composition of a field of maize. *J Hered.* 1908;4(1):296–301.
6. Bruce AB. **The Mendelian theory of heredity and the augmentation of vigor.** *Science* 1910(32):627–628.
7. Yu SB, Li JX, Xu CG, Tan YF, Gao YJ, Li XH, Zhang Q, Maroof MAS. Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proc Natl Acad Sci USA.* 1997;94(17):9226–31.
8. Zhou G, Chen Y, Yao W, Zhang C, Xie W, Hua J, Xing Y, Xiao J, Zhang Q. Genetic composition of yield heterosis in an elite rice hybrid. *Proc Natl Acad Sci USA.* 2012;109(39):15847–52.
9. Chen ZJ. Molecular mechanisms of polyploidy and hybrid vigor. *Trends Plant Sci.* 2010;15(2):57–71.
10. Riddle NC, Birchler JA. Comparative analysis of inbred and hybrid maize at the diploid and tetraploid levels. *Theor Appl Genet.* 2008;116(4):563–76.
11. He GM, Zhu XP, Elling AA, Chen LB, Wang XF, Guo L, Liang MZ, He H, Zhang HY, Chen FF, et al. Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. *Plant Cell.* 2010;22(1):17–33.
12. Katara JL, Verma RL, Parida M, Ngangkham U, Molla KA, Barbadikar KM, Mukherjee M, C P, Samantaray S, Ravi NR, et al. Differential expression of genes at panicle initiation and grain filling stages implied in heterosis of rice hybrids. *Int J Mol Sci.* 2020;21(3):1080.
13. Meyer RC, Witucka-Wall H, Becher M, Blacha A, Boudichevskaia A, Dormann P, Fiehn O, Friedel S, von Korff M, Lisek J, et al. **Heterosis manifestation during early *Arabidopsis* seedling development is**

- characterized by intermediate gene expression and enhanced metabolic activity in the hybrids. *Plant J.* 2012;71(4):669–83.
14. Liu P-C, Peacock WJ, Wang L, Furbank R, Larkum A, Dennis ES. **Leaf growth in early development is key to biomass heterosis in *Arabidopsis*.** *J Exp Bot* 2020, eraa006.
  15. Paschold A, Jia Y, Marcon C, Lund S, Larson NB, Yeh CT, Ossowski S, Lanz C, Nettleton D, Schnable PS, et al. **Complementation contributes to transcriptome complexity in maize (*Zea mays* L.) hybrids relative to their inbred parents.** *Genome Res.* 2012;22(12):2445–54.
  16. Zhao Y, Hu F, Zhang X, Wei Q, Dong J, Bo C, Cheng B, Ma Q. Comparative transcriptome analysis reveals important roles of nonadditive genes in maize hybrid An’ong 591 under heat stress. *Bmc Plant Biol.* 2019;19(1):273.
  17. Jackson S, Chen ZJ. Genomic and expression plasticity of polyploidy. *Curr Opin Plant Biol.* 2010;13(2):153–9.
  18. Fujimoto R, Taylor JM, Shirasawa S, Peacock WJ, Dennis ES. **Heterosis of *Arabidopsis* hybrids between C24 and Col is associated with increased photosynthesis capacity.** *Proc Natl Acad Sci USA.* 2012;109(18):7109–14.
  19. Mell PH. **Experiments in crossing for the purpose of improving the cotton fiber.** *Ala Agric Exp Stn Bull* 1894:56.
  20. Loden HDRTR. Hybrid vigor in cotton-cytogenetic aspects and practical applications. *Econ Bot.* 1951;5:387–408.
  21. D. DD: **Hybrid cotton: specific problems and potentials.** *Adv Agron* 1978, 303:129–157.
  22. Zhu XX, Ainijiang, Zhang YM, Guo WZ, Zhang TZ. Relationships between differential gene expression and heterosis in cotton hybrids developed from the foundation parent CRI-12 and its pedigree-derived lines. *Plant Sci.* 2011;180(2):221–7.
  23. Tyagi P, Bowman D, Bourland F, Edmisten K, Campbell BT, Fraser D, Wallace T, Kuraparthy V. **Components of hybrid vigor in upland cotton (*Gossypium hirsutum* L.) and their relationship with environment.** *Euphytica.* 2014;195(1):117–27.
  24. Wang L, Greaves IK, Groszmann M, Wu LM, Dennis ES, Peacock WJ. **Hybrid mimics and hybrid vigor in *Arabidopsis*.** *Proc Natl Acad Sci USA.* 2015;112(35):4959–67.
  25. Ding H, Cheng Q, Luo X, Li L, Chen Z, Liu H, Gao J, Lin H, Shen Y, Zhao M. Heterosis in early maize ear inflorescence development: a genome-wide transcription analysis for two maize inbred lines and their Hybrid. *Int J Mol Sci.* 2014;15(8):13892.
  26. Guo H, Mendrikahy JN, Xie L, Deng J, Lu Z, Wu J, Li X, Shahid MQ, Liu X. Transcriptome analysis of neo-tetraploid rice reveals specific differential gene expressions associated with fertility and heterosis. *Sci Rep.* 2017;7:40139.
  27. Guo M, Rupe MA, Wei J, Winkler C, Goncalves-Butruille M, Weers BP, Cerwick SF, Dieter JA, Duncan KE, Howard RJ, et al. Maize ARGOS1 (ZAR1) transgenic alleles increase hybrid maize yield. *J Exp Bot.* 2014;65(1):249–60.

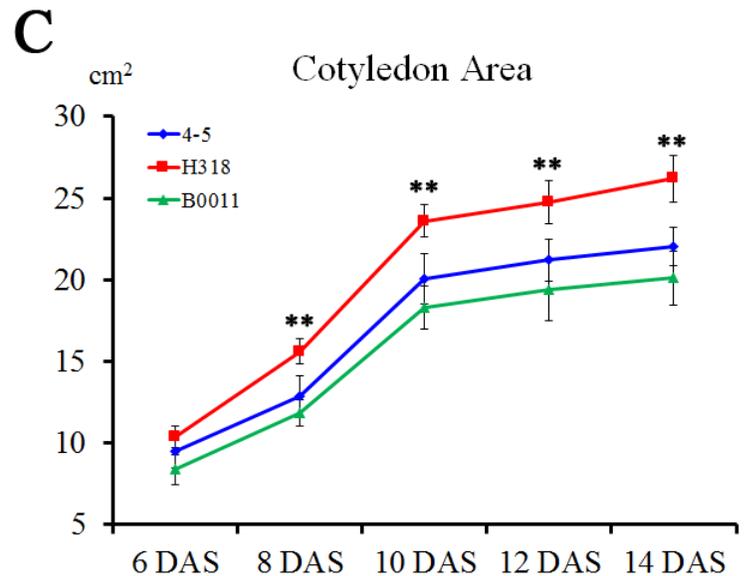
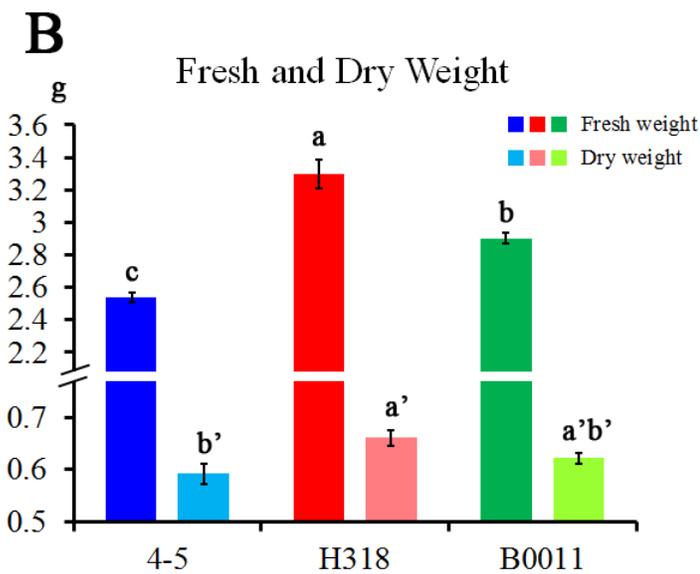
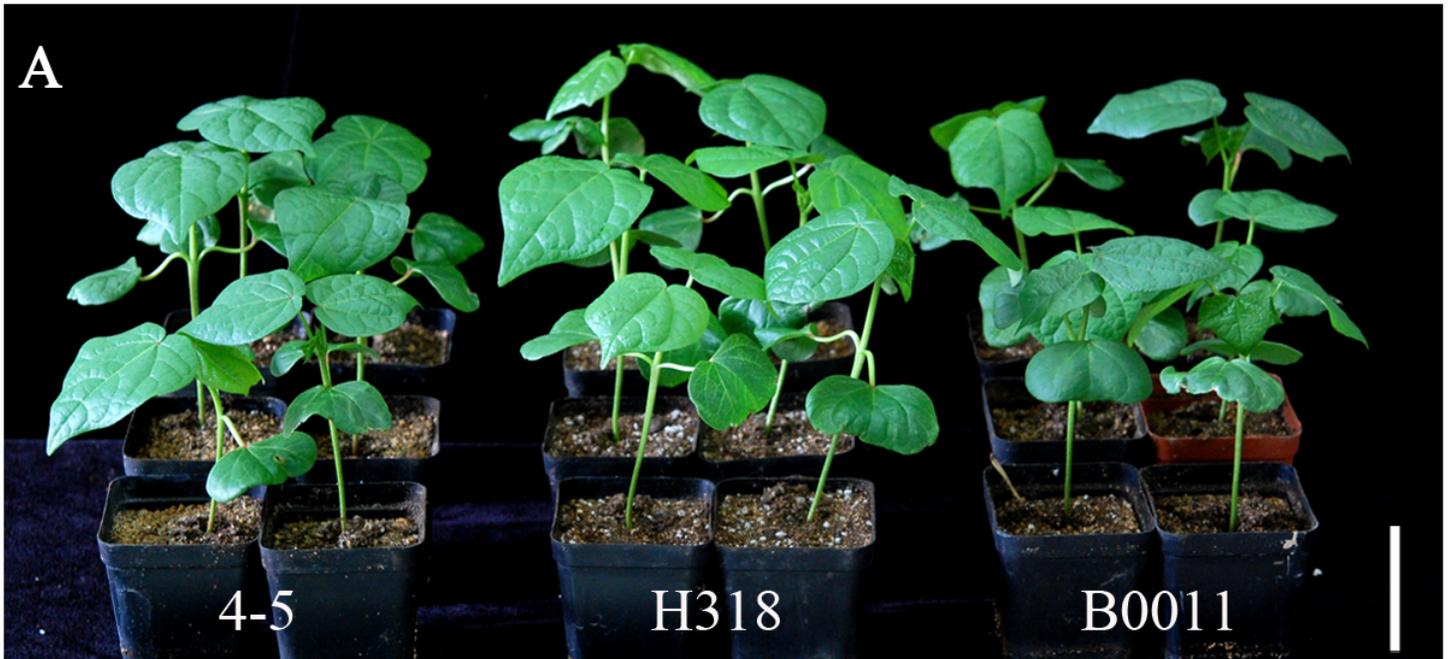
28. Korn M, Gärtner T, Erban A, Kopka J, Selbig J, Hincha DK. **Predicting *Arabidopsis* freezing tolerance and heterosis in freezing tolerance from metabolite composition.** *Mol Plant*. 2010;3(1):224–35.
29. Korn M, Peterek S, Mock HP, Heyer AG, Hincha DK. Heterosis in the freezing tolerance, and sugar and flavonoid contents of crosses between *Arabidopsis thaliana* accessions of widely varying freezing tolerance. *Plant Cell Environ*. 2008;31(6):813–27.
30. Miller M, Song Q, Shi X, Juenger TE, Chen ZJ. **Natural variation in timing of stress-responsive gene expression predicts heterosis in intraspecific hybrids of *Arabidopsis*.** *Nat Commun*. 2015;6:7453.
31. Yang M, Wang X, Ren D, Huang H, Xu M, He G, Deng XW. **Genomic architecture of biomass heterosis in *Arabidopsis*.** *Proc Natl Acad Sci USA*. 2017;114(30):8101–6.
32. Wang L, Wu LM, Greaves IK, Zhu A, Dennis ES, Peacock WJ. **PIF4-controlled auxin pathway contributes to hybrid vigor in *Arabidopsis thaliana*.** *Proc Natl Acad Sci USA*. 2017;114(17):E3555–62.
33. Zhu A, Greaves IK, Liu PC, Wu L, Dennis ES, Peacock WJ. **Early changes of gene activity in developing seedlings of *Arabidopsis* hybrids relative to parents may contribute to hybrid vigour.** *Plant J*. 2016;88(4):597–607.
34. Ma QM, Qian<sup>1,2</sup>; Hedden, P (Hedden, Peter)<sup>3</sup>; Zhang, QF (Zhang, Qifa)<sup>1,2</sup> **Heterosis in rice seedlings: its relationship to gibberellin content and expression of gibberellin metabolism and signaling genes.** *Plant Physiol Bioch* 2011.
35. Solhaug EM, Ihinger J, Jost M, Gamboa V, Marchant B, Bradford D, Doerge RW, Tyagi A, Replogle A, Madlung A. **Environmental regulation of heterosis in the allopolyploid *Arabidopsis suecica*.** *Plant Physiol*. 2016;170(4):2251–63.
36. Zuo X, Han B, Cheng J-L. A measurement approach of leaf area based on digital image processing. *Computer Engineering Applications*. 2006;42(27):194–6.
37. Zhu LTL, Zeng F, Liu D, Zhang X. **An improved simple protocol for isolation of high quality RNA from *Gossypium* spp. suitable for cDNA library construction.** *Acta Agron Sin*. 2005;31:1657–9.
38. Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J, Saski CA, Scheffler BE, Stelly DM, et al. **Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement.** *Nat Biotech*. 2015;33:531–7.
39. Tu L, Zhang X, Liu D, Jin S, Cao J, Zhu L, Deng F, Tan J, Zhang C. Suitable internal control genes for qRT-PCR normalization in cotton fiber development and somatic embryogenesis. *Chinese Sci Bull*. 2007;52(22):3110–7.
40. Min L, Zhu L, Tu L, Deng F, Yuan D, Zhang X. Cotton GhCKI disrupts normal male reproduction by delaying tapetum programmed cell death via inactivating starch synthase. *Plant J*. 2013;75(5):823–35.
41. Buysse J, Merckx R. An improved colorimetric method to quantify sugar content of plant tissue. *J Exp Bot*. 1993;44(267):1627–9.

## Figures



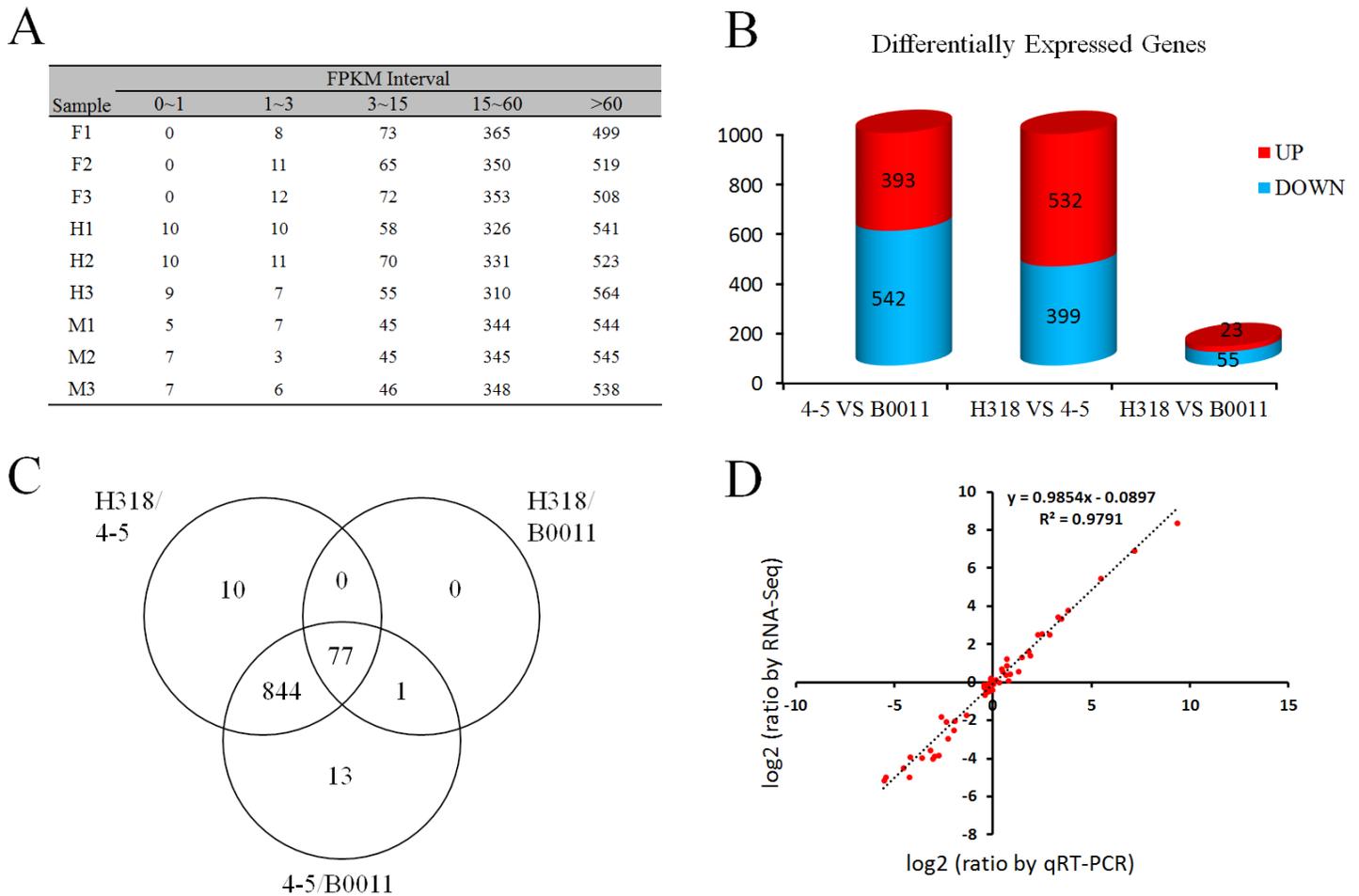
**Figure 1**

Four years' investigation of yield-related traits in H318 and its parental lines. A: The number of bolls was stably higher in the hybrid over 4 years. B: Boll-node number increased in the first two years. C: Abscission rate decreased in 3 out of the 4 years, corresponding to the number of bolls. D: The average number of fruit branches has no difference between hybrid and its parents. Asterisks indicate statistically significant differences (\*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ ) by Student's t test.



**Figure 2**

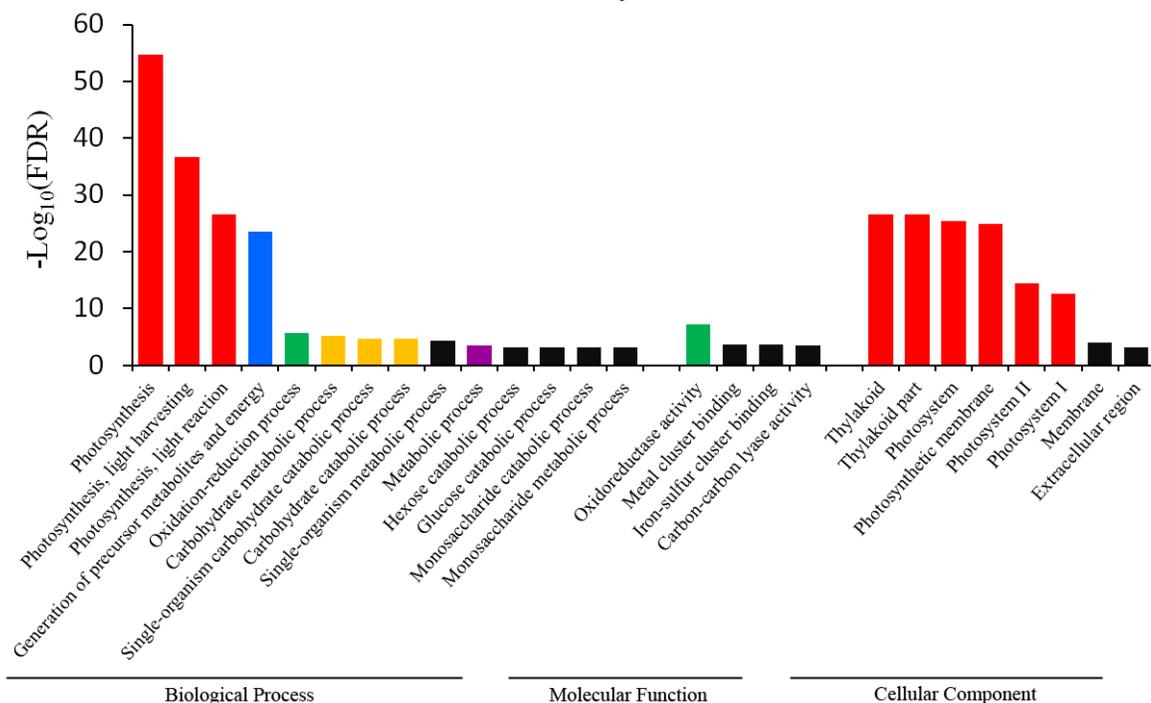
Heterosis performance of H318 at seedling stages. A: Phenotype of H318 and its parental lines at seedling stages. H318 showed obviously growth heterosis at seedling stages relative to its parental lines. Bar = 5 cm. B: Fresh and dry weight of H318 and its parental lines. Values with different letter are considered statistically significant (shortest significant range;  $P < 0.05$ ). C: Cotyledon area of H318 and its parental lines from the 6th to 14th day after sowing (DAS). Asterisks indicate statistically significant differences (\*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ ) by Student's t test.



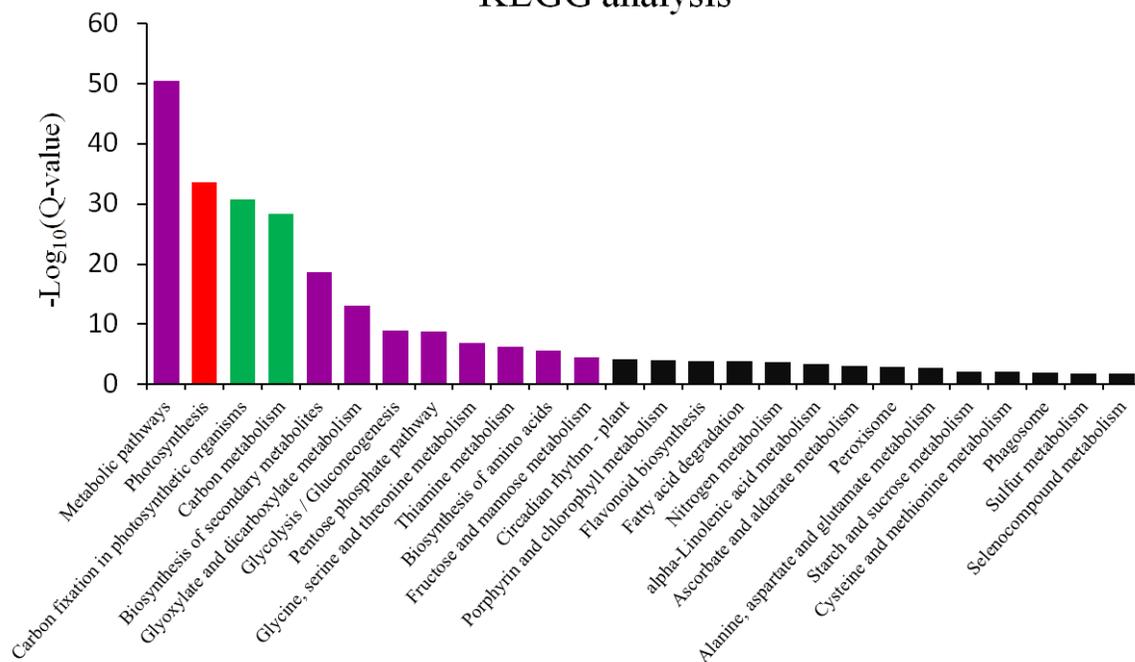
**Figure 3**

Statistic analysis and qRT-PCR validation of differentially expressed genes (DEGs) from RNA sequencing data. A: The expression distribution of filtered DEGs (945) from each sample according to the FPKM values. B: Up- and down-regulated DEGs between each sample. C: The distribution of all DEGs in three genotype. D: Correlation analysis between qRT-PCR results and data from RNA sequencing.

## GO analysis

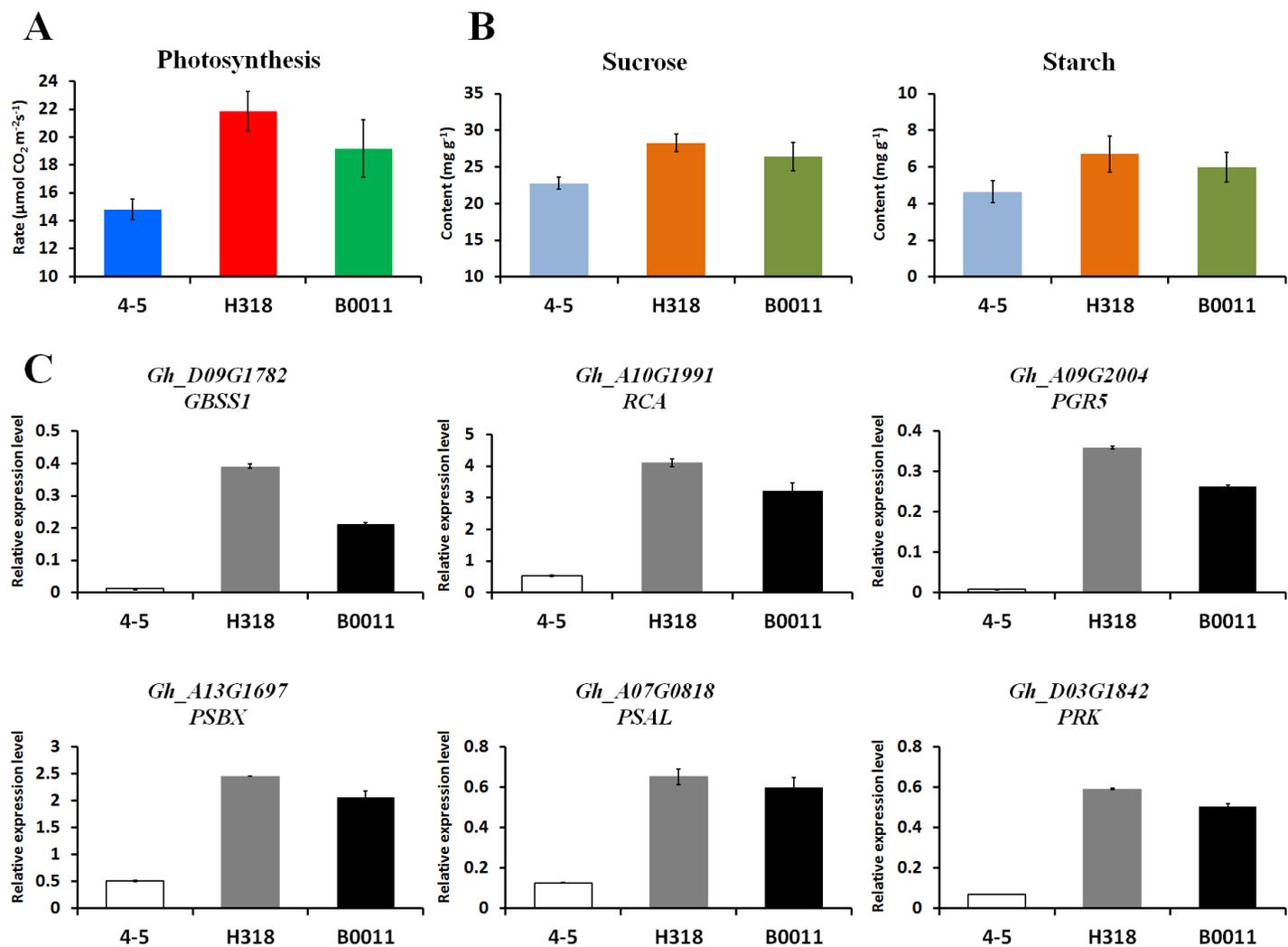


## KEGG analysis



**Figure 4**

GO and KEGG analysis of DEGs. All processes and pathways were ranked according to the value of  $-\log_{10}(\text{FDR})$ .



**Figure 5**

The detection of photosynthesis rate, sucrose, starch and the gene expression changes related to photosynthesis and carbohydrate metabolic processes. A: Photosynthesis rate of H318 and its parental lines at seedling stage. B: The content of sucrose and starch in H318 and its parental lines. C: Gene expression changes in photosynthesis and carbohydrate metabolic processes. GhUB7 was used as control.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplement1.xlsx](#)
- [supplement2.tif](#)
- [supplement3.tif](#)
- [supplement4.tif](#)

- [supplement5.tif](#)
- [supplement6.tif](#)