Plant trait heterosis is quantitatively associated with expression heterosis of the plastid ribosomal proteins

Devon Birdseye[1], Laura A. de Boer[1], Hua Bai[1], Peng Zhou[2], Zhouxin Shen[1], Eric A. Schmelz[1], Nathan M. Springer[2], Steven P. Briggs[1]

[1] Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093, USA

[2] Department of Plant and Microbial Biology, University of Minnesota, Saint Paul, MN 55108, USA

Abstract

The use of hybrids is widespread in agriculture, yet the molecular basis for hybrid vigor (heterosis) remains obscure. To identify molecular components that may contribute to the known higher photosynthetic capacity of maize hybrids, we analyzed proteomic and transcriptomic data from seedling leaf tissues of the B73xMo17 hybrid and its inbred parents. Subunits of complexes required for chloroplast protein synthesis and for the light reactions of photosynthesis were expressed above mid-parent and high-parent levels, respectively. Nuclear and plastid-encoded subunits were expressed similarly and in a dominant pattern with B73 as the high-parent for most proteins. The reciprocal hybrid displayed the same pattern with B73 still the dominant parent. Protein expression patterns were mostly the same in the adult leaf blade. To identify potential seedling biomarkers for adult trait heterosis, we compared seedling leaf protein levels to adult plant heights of 15 hybrids with varied levels of heterosis. Expression heterosis (hybrid/mid-parent expression levels) was most positively correlated with plant height heterosis (hybrid/mid-parent plant height) for the plastid-encoded and nuclear-encoded chloroplast ribosomal proteins; the cytosolic ribosomal proteins were slightly less correlated. Ethylene biosynthetic enzymes were expressed below mid-parent levels in the hybrids, and the ethylene biosynthesis mutant, *acs2/acs6*, largely phenocopied the hybrid proteome, indicating that a reduction in ethylene biosynthesis may mediate the differences between inbreds and their hybrids. Levels of jasmonic acid biosynthetic enzymes were reduced in both *acs2/acs6* and the hybrid. Expression heterosis of these proteins was most negatively correlated with plant height heterosis, indicating a potential role in heterosis.

Significance Statement

Heterosis (hybrid vigor) boosts the growth and resilience of crops and livestock above the levels of both parents, yet its underlying mechanisms remain unknown. We analyzed expression patterns of all proteins in maize hybrids and their inbred parents. Differences in several molecular machines and biochemical pathways were found and quantitatively assessed using a panel of 15 hybrids to determine association with trait heterosis. Seedling leaf chloroplast ribosomal proteins were able to quantitatively infer levels of adult plant heterosis. Expression levels of biosynthetic enzymes for the stress hormone, ethylene, were reduced in hybrids and mutation of these genes in an inbred caused the proteome to resemble a hybrid. Repression of ethylene biosynthesis may play a key role in maize heterosis.

Introduction

Plants, animals, and humans display inbreeding depression associated with increased genetic uniformity and characterized by reduced vigor [1]. Mating between genetically distinct inbred parents can produce hybrid vigor, or heterosis, defined as the difference in vigor between a hybrid and (a) the average of its parents or (b) the better-performing parent. Hybrids may out-perform their parents in terms of size, vigor, yield, abiotic and biotic stress resistance, longevity, and reproductive advantage [2,3]. Maize is the most important crop in the United States, and it was the first hybrid crop to be made and sold. Heterosis in maize increases many traits far above the levels of the higher parent, including biomass and harvestable grain [4]. Despite its importance in agriculture, the changes in physiology that cause hybrid vigor remain obscure. Breeding hybrids requires expensive and labor-intensive field tests, of both the hybrids and their parents. Progress in maize breeding has greatly improved yield through increased tolerance of high-density planting stress and this primarily constitutes the non-heterosis portion of yield [5]. It is unclear whether substantial improvements in heterosis are possible and yet have lagged because breeders lack seedling biomarkers for adult plant heterosis.

QTL mapping has shown that minor effect loci for heterosis are distributed throughout the genome [1]. Transcriptome [6] and proteome [7] profiles indicate that gene expression levels in hybrids are generally the average of their parents. A minority of mRNAs and proteins are expressed

above or below mid-parent levels and most of these gene products are not obviously related to each other in function or to hybrid phenotypes, making their unusual levels difficult to interpret [8]. Despite these ambiguities, progress has been made in understanding hybrid vigor.

Recent work has implicated the plant hormone ethylene (ET) in heterosis. Transcript levels of ET biosynthesis genes are lower than expected in maize hybrids [9]. Knock-down of ET biosynthesis gene expression enhanced the agronomic performance of maize hybrids [10]. In Arabidopsis hybrids, ET is produced at reduced levels and application of exogenous ET reduced hybrid vigor [11]. Thus, ET potentially mediates at least some aspects of heterosis and it is unclear which genes are affected by reduced ET biosynthesis.

Maize hybrids have greater photosynthetic capacity than their inbred parents [9], which could contribute to their increased biomass and yield. In both maize and Arabidopsis, various genes for photosynthesis have been observed to be expressed above mid-parent levels [9,12,13]. Some metabolites and enzymes of the photosynthetic carbon reactions are elevated in maize hybrids while others in photorespiration are repressed [14]. We sought to identify protein biomarkers in seedling leaves that contribute new insights to the physiology of hybrid vigor and can be used to predict levels of adult trait heterosis.

Results

Dominant and Overdominant Expression Patterns of Chloroplast Protein Complexes

Multiplexed proteomic analyses were performed using TMT peptide tags and high-resolution mass spectrometry to quantify differences in the levels of proteins extracted from seedling leaves of hybrids and their inbred parents. Expression levels in the hybrid, B73xMo17, were compared to high-parent (HP) levels, low-parent (LP) levels, and calculated mid-parent (MP) levels to identify non-additive expression. Readers can use our website to visualize expression of their chosen genes in the hybrid and inbred parents (https://devonbirdseye.shinyapps.io/ExpressionViewer/).

Of 10,141 measured leaf proteins, 638 (6%) were expressed above HP. Functional enrichment performed using The Database for Annotation, Visualization and Integrated Discovery (DAVID)

[15,16] revealed that these were enriched for chloroplast-localization. Of the 1,068 chloroplast-localized proteins observed, 256 (24%) were expressed above HP levels. This included 65 of the Photosynthesis-Associated Nuclear Genes (PhANGs) and Photosynthesis-Associated Plastid Genes (PhAPGs). The PhAPG/PhANG proteins comprise photosystem I, photosystem II, cytochrome b6/f complex, NAD(P)H dehydrogenase-like complex, ATP synthase, and the carbon-fixation enzyme, Rubisco. PhANG and PhAPG proteins were expressed 19% above HP levels, on average (Figure 1A). The remaining 189 chloroplast-localized proteins were weakly enriched for carbon fixation.

Expression between MP and HP levels was observed for 956 proteins (9%). They were enriched for plastid ribosomal proteins, encoded by either the nuclear or plastid genome. The ribosome subunits were expressed 8% above MP, on average (Figure 1B). Less enrichment was observed for enzymes of carbon metabolism. While not enriched, carbon fixation-related proteins were differentially expressed in the hybrid, some above MP and some below (Figure S1).

Proteins that were expressed non-additively were frequently expressed at different levels between the parents, and expression in B73 was generally higher than in Mo17 (Figure 1C). However, some proteins such as photosynthetic electron transfer C, ribose-5-phosphate isomerase, and the oxygen evolving complex assembly proteins were expressed higher in Mo17 (Supplemental Table 1).

Expression below LP levels was observed for 517 proteins (5%). They were weakly enriched for amino acid biosynthesis, fatty acid degradation, chaperone proteins, and peroxisomal proteins. Expression between MP and LP was observed for 1,222 proteins (12%). These showed weak enrichment for chaperone proteins and protein processing in the ER.

The non-additive patterns of expression seen in the proteome of juvenile leaves were also observed in the blades of mature leaves (Figure S2 A). However, there was a reversal of relative expression levels between the parents for plastid ribosomal proteins (Figure S2 B). B73 was high parent in the seedling leaf, whereas Mo17 was high parent in the leaf blade.

Correlations between Trait Heterosis and Expression Heterosis

To quantitatively characterize the relationship between trait heterosis and protein expression heterosis across a set of hybrids, we examined six hybrids generated from diverse germplasm. B73xMo17 and its reciprocal hybrid, Mo17xB73 (the female parent is listed first), displayed the same expression patterns, indicating that the patterns are caused by combining the nuclear genomes, with little or no effect due to maternal inheritance of the plastid genome (Figure 2 C-D). Inbreds B73 and B84 are related members of the Stiff Stalk pool of germplasm. Inbreds Mo17 and A682 are part of the Non-Stiff Stalk pool. Hybrids made from crosses between these pools have strong heterosis in contrast to hybrids made from crosses within each pool [17]. The highest expression heterosis of the PhANG/PhAPG proteins was observed in hybrids with the least heterosis (Figure 2 E-F). In contrast, the plastid ribosomal proteins were expressed at or below HP levels in the hybrids with low levels of heterosis and substantially above HP levels in the hybrid with the greatest heterosis (Figure 2). High-heterosis hybrids expressed non-ribosomal proteins involved in protein biosynthesis above mid-parent levels whereas proteins involved in metabolic pathways were expressed below mid-parent levels.

We additionally analyzed nine hybrids made by back-crossing four IBM recombinant inbred lines (RILs) to both parents, B73 and Mo17 . The RIL hybrids were approximately 50% as heterozygous as the parent hybrid and were selected based on their varied levels of plant height heterosis (Supplemental Table 2 and Supplemental Table 3). We examined collectively these 15 hybrids and their parents, including the RIL hybrid set and the six hybrids described above. As a measure of trait heterosis, we utilized plant height, which is associated with grain yield [18]. We calculated the Pearson's correlations between expression heterosis and plant height heterosis (hybrid height/MP height). Readers can obtain Pearson correlation values for their genes of interest using our website (https://devonbirdseye.shinyapps.io/ExpressionViewer/). It is important to note that there was no correlation between expression levels and plant height *per se*; only the hybrid/mid-parent values for expression and height were correlated. This underscores the specificity of our biomarkers for the trait of heterosis. We refer to these calculated values as expression heterosis and plant height heterosis.

Expression heterosis levels of protein complexes and pathways were correlated with plant height heterosis levels (Supplemental Table 4) and 581 proteins had individual Pearson correlations greater than 0.5. This set of proteins was enriched for the terms ribosome (93 proteins), chloroplast (48 proteins), protein biosynthesis (20 proteins), photosynthesis (11 proteins), plastid chromosome (5 proteins), and tetratricopeptide repeat (12 proteins) (Figure 3). Of the top 1.5% of proteins whose expression heterosis was most correlated with plant height heterosis, 18 out of 54 were from the plastid ribosome. Pearson's correlations of 0.79-0.91 were observed for these proteins, half of which were nuclear-encoded and half were plastid-encoded. Correlations were most positive for the plastid ribosomal proteins, slightly less positive for the cytosolic ribosomal proteins, and moderately positive for the PhANG/PhAPG proteins (Figure 3). Correlations of hybrid/HP values between expression and plant height were less strong for the PhANG/PhAPG proteins, and approximately as strong for the plastid ribosomal proteins (Figure 3). There were 413 proteins with a Pearson's correlation less than -0.5; these were most strongly enriched for oxidoreductase (39 proteins), biosynthesis of secondary metabolites (63 proteins), protease (18 proteins), biosynthesis of antibiotics (32 proteins), and alpha-linoleic acid metabolism (7 proteins) terms (Figure 3).

Transcript Expression Heterosis Differs from Protein Expression Heterosis

The proteome differences between chloroplasts of the hybrid and its inbred parents were not mirrored in their nuclear transcriptomes (Figure S3). The nuclear transcripts for PhANGs were between MP and HP. Nuclear transcripts for plastid ribosomal subunits were substantially below MP in contrast to their proteins which were above MP. Most PhANG transcripts were expressed higher in B73, consistent with their protein expression. However, most plastid ribosomal subunit transcripts were higher in Mo17, strikingly discordant with parental expression of their proteins.

Expression of nuclear transcripts for the plastid ribosome relative to MP levels was less strongly correlated with plant height heterosis than that of their proteins, and expression relative to HP levels was even weaker (Figure S4). PhANG transcript expression relative to MP levels was substantially more positively correlated than that of their proteins, while expression relative to HP levels had a weakly positive correlation, similar to their proteins. Transcripts for the cytosolic

ribosome had no correlation with plant height heterosis, unlike their proteins. Plastid-encoded transcripts were not measured.

Overall, 640 transcripts were expressed above HP, with functional enrichment for circadian rhythm accounting for 7 transcripts; the remaining transcripts had no functional enrichment. With a few exceptions, most circadian-related transcripts were expressed above MP levels (Figure S5). There were 40 transcription factors expressed above HP, including 5 ethylene-responsive factors, 4 HOX-domain proteins, 1 auxin-responsive factor, and the circadian regulator, LHY. Conversely, 517 transcripts were expressed below LP, with functional enrichment for metabolic pathways, especially biosynthesis of amino acids. Only one transcription factor was expressed below LP, PHD16.

Expression Heterosis is Phenocopied by an *acs* Mutant

Mutations in two of the maize genes encoding ACC synthase have been functionally characterized (*Zmacs2* and *Zmacs6*) and were shown to have decreased endogenous ET levels [19–22]. While the ZmACS6 transcript was not detected, the ZmACS2 gene transcript, plus transcripts for ACO and SAM1/2 homologs were observed and found to be expressed substantially below MP levels (Figure S6). Several other isoforms of the enzymes predicted to be involved in ET biosynthesis were also expressed significantly below MP levels, although it remains unclear if reduced expression of these enzymes would cause a reduction in ET levels.

Proteomic analysis of the ET biosynthesis double mutant, *Zmacs2/6*, in the B73 genetic background [19] revealed that the mutant phenocopied the hybrid molecular phenotype. Proteins expressed above or below MP in A682xB73 (the hybrid with the greatest heterosis), were expressed similarly in the mutant relative to the inbred, B73. (Figure 4A). Protein levels of the majority of both the PhANGs/PhAPGs and plastid ribosome subunits were elevated in the mutant at equal proportions as in A682xB73 (Figure 4B). The exceptions were two ATP synthase subunits and two plastid ribosomal proteins, which were significantly above MP in the hybrid but unchanged in the mutant. Expression levels of jasmonic acid (JA) biosynthesis enzymes were reduced in the *Zmacs2/6* mutant relative to B73 and in hybrids relative to MP (Figure S7). Of the 538 proteins expressed above MP in A682xB73, only 18 were repressed in

Zmacs2/6; these were enriched for biosynthesis of secondary metabolites and also included three PLASTID TRANSCRIPTIONALLY ACTIVE CHROMSOME (PTAC) proteins. Of the 285 proteins expressed below MP in A682xB73, only 15 were elevated in *Zmacs2/6*; no functional enrichment was found for these.

Discussion

Non-additive expression patterns were observed for both protein and RNA in the hybrid, B73xMo17. Photosynthesis-related proteins and plastid ribosomal proteins represented the major groups expressed above HP and MP levels, respectively. For both groups, the nuclear-encoded and plastid-encoded subunits were expressed similarly. Most plastid-localized proteins were expressed at mid-parent levels. Elevated expression of photosynthetic proteins in hybrids is diurnally-regulated, with the highest levels in the morning [14], which corresponds to the sampling time in our experiments. We observed that circadian-related genes were enriched in the transcripts expressed above HP, consistent with previous reports of an altered circadian rhythm in hybrids [9,11,23]. Many transcription factors were also expressed above HP. Proteins and transcripts expressed below LP were enriched most strongly for amino acid biosynthesis, consistent with previous measurements of hybrid metabolite levels [14].

We hypothesize that coupling of photosynthesis to growth - specifically, the heterosis portions of each - may be through the plastid ribosome, which was positively correlated with plant height heterosis at both protein and transcript levels. The cytoplasmic ribosome is nearly as well correlated and is also likely to contribute to heterosis traits. Non-photosynthetic seminal roots of the hybrid, B73xMo17, express 16 cytoplasmic ribosomal proteins at or above high-parent levels [24]. In yeast, proteome reallocation to ribosomes enables faster growth in rich media [25]. Elevated photosynthetic capacity may be important for heterosis even though it does not appear to be quantitatively coupled to the levels of heterosis. Hybrids made from inbreds of the same heterotic pool expressed the PhANG and PhAPG proteins above high-parent levels whereas PhANG transcripts were higher in hybrids made from crosses between heterotic pools. Consequently, PhANG transcripts were quantitatively associated with plant height heterosis, while their protein levels were not. Nevertheless, hybrids have greater photosynthetic capacity than their parents and this seems necessary to produce the greater biomass and yield of hybrids [9].

Other groups of proteins with positive correlations to heterosis included PTAC proteins and protein biosynthesis-related proteins, consisting of elongation factors and initiation factors. Elevated expression of both of these groups of proteins in the hybrid may contribute to the hybrid-specific expression patterns of the chloroplast- and nuclear-encoded proteins. Tetratricopeptide repeat (TPR) proteins were enriched in the positively correlated group, and some of these are known to mediate protein-protein interactions and play roles in stress and hormone signaling [26].

Groups of proteins enriched amongst those with negative correlations included oxidoreductase, consisting of several dehydrogenases, lipoxygenases, and one of the ethylene biosynthetic enzymes, ACC oxidase. Proteases were negatively correlated. The two most negatively correlated groups were biosynthesis of secondary metabolites and alpha-linoleic acid metabolism, consistent with the proposed tradeoff between stress responses and plant growth. Carbon fixation-related proteins showed altered mixed expression pattern in the hybrid, and were neither positively nor negatively correlated with plant height heterosis. Others report that hybrids are associated with either elevated levels of photosynthetic proteins (hybrid ZD909) or stress proteins (hybrid ZD808) but not both [27].Our results suggest that the physiology of heterosis may be conserved between monocots and dicots. ET biosynthetic enzyme gene expression was repressed in maize hybrids as was previously reported for Arabidopsis hybrids [11]. Reduced expression of ET biosynthetic enzymes caused by the *Zmacs2/6* mutation in B73, phenocopied the heterosis expression levels of chloroplast proteins. Traits of the maize *acs2* and *acs6* mutants are known to resemble those of hybrids including transpiration, stomatal closure, carbon dioxide assimilation, and elevated levels of foliar chlorophyll, Rubisco, and soluble protein [19]. JA biosynthetic enzyme levels were repressed in both the hybrid and the *Zmacs2-6* mutant, indicating that JA biosynthesis may be regulated by ET. Interestingly, the JA biosynthetic pathway was enriched in the proteins that were negatively correlated with plant height heterosis, suggesting a potential role for reduced JA biosynthesis. The most notable differences between A682xB73 (the hybrid with the greatest heterosis) and the *Zmacs2-6* mutant were expression levels of three PTAC proteins which were elevated in the hybrid and repressed in the mutant.

Levels of these plastid RNA polymerase protein subunits were positively correlated with plant height heterosis.

The discrepancy between accumulation patterns of transcripts and their proteins argues that post-transcriptional mechanisms are selectively regulating protein levels in the inbreds and their hybrids. Developmental and circadian patterns of expression can uncouple transcripts from their proteins [28]. Expression of plastid ribosomal proteins was particularly interesting. They were expressed above MP in B73xMo17 while their transcripts were below MP. Across a panel of 15 hybrids, expression heterosis of these proteins was highly correlated with plant height heterosis. Most nuclear-encoded and plastid-encoded subunits were expressed higher in B73 than in Mo17. In contrast, transcripts for the nuclear-encoded subunits were mostly expressed higher in Mo17. It's unclear how B73 accumulates more of each ribosomal protein from substantially lower transcript levels than in Mo17. The levels of PhANG and PhAPG proteins were higher in B73 than in Mo17, in parallel with their nuclear transcript levels. Thus, for the digenomic protein complexes of the chloroplast, one parent is dominant for expression of nuclear and plastid-encoded subunits. Their levels are higher in B73 and above MP or HP in the hybrid, even when their transcripts follow an opposite pattern. This is true whether B73 is the female or male parent. Proteome dominance has been observed in the non-photosynthetic immature ear of the maize hybrid, ZD909 [7]. The authors found that many enzymes involved in carbon and nitrogen assimilation were expressed at high-parent levels. In Arabidopsis, transcripts for cell division are more highly expressed in one parent, and in the other parent transcripts for photosynthesis are higher. The hybrid expressed both sets above MP levels and the combination of the two pathways may be important for heterosis [29]. Further work is needed to understand the mechanisms of expression dominance and the roles it may play in heterosis.

Materials and Methods

Plant growth and sampling

Plant materials were all grown in the green house for collection of seedling tissues and grown in the field for collection of adult tissues. The seedling leaf and mature leaf blade tissues of B73, Mo17 and B73xMo17 were collected in May of 2016 in St Paul, Minnesota, while tissues from the six hybrids and RIL hybrids experiments were collected from July to November of 2019 in St

Paul, Minnesota. For seedling tissues, seeds were imbibed for 24 hours in distilled water and grown in growth chambers at 25°C under 12h light-dark cycles to V2/V3 stage (day 9) for collection of the V2 leaves. Tissue from at least four plants was pooled for each replicate and the three biological replicates were sampled at the same time (12PM) on different dates in order to sample the same growth stage for each genotype. For the *acs2/6* mutant and B73 seedling leaf samples, seeds were imbibed for 48 hours and manually examined to select uniformly germinating seedlings based on radical emergence and the lack of mold. They were then grown in growth chambers at 25°C under 12h light-dark cycles to V2/V3 stage (day 10). The V3 leaf was collected for 5 biological replicates of each genotype.

RNA-Seq analysis

Sequence libraries were prepared using the standard TruSeq Stranded mRNA library protocol and sequenced on NovaSeq 150bp paired end S4 flow cell to produce at least 20 million reads for each sample (Table S1). Both library construction and sequencing were done in the University of Minnesota Genomics Center. Sequencing reads were then processed through the nf-core RNA-Seq pipeline [30] for initial QC and raw read counting. In short, reads were trimmed by Trim Galore! (https://github.com/FelixKrueger/TrimGalore) and aligned to the B73 maize reference genome (Zm-B73-REFERENCE-GRAMENE-4.0) [31] using the variant-aware aligner Hisat2 [32] which also takes 90 million common variants to account for mapping bias. Uniquely aligned reads were then counted per feature by featureCounts [33]. Raw read counts were then normalized by library size and corrected for library composition bias using the TMM normalization approach [34] to give CPMs (Counts Per Million reads) for each gene in each sample, allowing direct comparison across samples. CPM values were then normalized by gene CDS lengths to give FPKM (Fragments Per Kilobase of exon per Million reads) values. Hierarchical clustering, principal component analysis and t-SNE clustering were used to explore sample cluster patterns and to remove or substitute bad replicates.

Proteomics Sample Preparation

Ground tissue powders were suspended in extraction buffer (8M Urea/100mM Tris/5mM Tris(2-carboxyethyl)phosphine (TCEP)/phosphatase inhibitors, pH 7). Proteins were precipitated by adding 4 volumes of cold acetone and incubated at 4oC for 2 hours. Samples were centrifuged at

4,000xg, 4oC for 5 minutes. Supernatant was removed and discarded. Proteins were re-suspended in Urea extraction buffer (8M Urea/100mM Tris/5mM TCEP/phosphatase inhibitors, pH 7) and precipitated by cold acetone one more time. Protein pellets were washed by cold methanol to further remove non-protein contaminants. Protein pellets were suspended in extraction buffer (8M Urea/100mM Tris/5mM TCEP/phosphatase inhibitors, pH 7). Proteins were first digested with Lys-C (Wako Chemicals, 125-05061) at 37oC for 15 minutes. Protein solution was diluted 8 times to 1M urea with 100mM Tris and digested with trypsin (Roche, 03 708 969 001) for 4 hours. Digested peptides were purified on a Waters Sep-Pak C18 cartridges, eluted with 60% acetonitrile. TMT-10 labeling was performed in 50% acetonitrile/150mM Tris, pH7. TMT labeling efficiency was checked by LC-MS/MS to be greater than 99%. Labeled peptides from different samples (Table 3) were pooled together for 2D-nanoLC-MS/MS analysis. An Agilent 1100 HPLC system was used to deliver a flow rate of 600 nL min-1 to a custom 3-phase capillary chromatography column through a splitter. Column phases were a 20 cm long reverse phase (RP1, 5 µm Zorbax SB-C18, Agilent), 6 cm long strong cation exchange (SCX, 3 µm PolySulfoethyl, PolyLC), and 20 cm long reverse phase 2 (RP2, 3.5 µm BEH C18, Waters), with the electrospray tip of the fused silica tubing pulled to a sharp tip (inner diameter <1 µm). Peptide mixtures were loaded onto RP1, and the 3 sections were joined and mounted on a custom electrospray adapter for on-line nested elutions. Peptides were eluted from RP1 section to SCX section using a 0 to 80% acetonitrile gradient for 60 minutes, and then were fractionated by the SCX column section using a series of 20 step salt gradients of ammonium acetate over 20 min, followed by high-resolution reverse phase separation on the RP2 section of the column using an acetonitrile gradient of 0 to 80% for 150 minutes.

Proteomics data acquisition

Spectra are acquired on a Q-exactive-HF mass spectrometer (Thermo Electron Corporation, San Jose, CA) operated in positive ion mode with a source temperature of 275 °C and spray voltage of 3kV. Automated data-dependent acquisition was employed of the top 20 ions with an isolation windows of 1.0 Da and collision energy of 30. The mass resolution is set at 60,000 for MS and 30,000 for MS/MS scans, respectively. Dynamic exclusion is used to improve the duty cycle.

Proteomics data analysis

The raw data was extracted and searched using Spectrum Mill vB.06 (Agilent Technologies). MS/MS spectra with a sequence tag length of 1 or less were considered to be poor spectra and were discarded. The remaining high quality MS/MS spectra were searched against a maize protein database including maize B73 (Zm-B73-REFERENCE-GRAMENE-4.0) and Mo17 (Zm-Mo17-REFERENCE-CAU-1.0) reference genomes [31,35]. A 1:1 concatenated forward-reverse database was constructed to calculate the false discovery rate (FDR). There were 435,212 protein sequences in the final protein database. Search parameters were set to Spectrum Mill's default settings with the enzyme parameter limited to full tryptic peptides with a maximum mis-cleavage of 1. Cutoff scores were dynamically assigned to each dataset to obtain the false discovery rates (FDR) of 0.1% for peptides, and 1% for proteins. Proteins that share common peptides were grouped using principles of parsimony to address protein database redundancy. Total TMT-10 reporter intensities were used for relative protein quantitation. Peptides shared among different protein groups were removed before TMT quantitation. Isotope impurities of TMT-10 reagents were corrected using correction factors provided by the manufacturer (Thermo). Median normalization was performed to normalize the protein TMT-10 reporter intensities in which the log ratios between different TMT-10 tags were adjusted globally such that the median log ratio was zero.

Proteomics data deposition

The raw spectra for the proteome data have been deposited in the Mass Spectrometry Interactive Virtual Environment (MassIVE) repository (massive.ucsd.edu/ProteoSAFe/static/massive.jsp) (accession ID MSV000085916). FTP download link before publication: ftp://MSV000085916@massive.ucsd.edu; FTP download link after publication: ftp://massive.ucsd.edu/MSV000085916/.

Expression analysis

For protein analyses, tandem-mass tag (TMT) abundances were normalized to the arithmetic mean of the B73 replicates in each run (Supplemental Table 5). For RNA analysis of the original seedling leaf and leaf blade, previously published counts per million (CPM) values were used [36] For six hybrid and RIL hybrid RNA analyses, CPM values were used (Supplemental Tables 6

and 7 for six hybrids and RIL hybrids, respectively). For each experiment, genes or proteins were removed if they were not detected in all biological replicates. The arithmetic mean of the biological replicates was then calculated and used to calculate hybrid/mid-parent and parent/parent ratios. The log base 2 was then calculated for each ratio. A T-test was used to calculate p-values. Data was manipulated using the R packages dplyr [37], plyr [38], tidyr [39], reshape2 [40], and data.table [41]. Scatterplots and volcano plots were generated using ggplot2 [42], density plots were generated using cowplot [43], and plots were assembled into paneled figures using patchwork [44]. For correlation analyses, the Pearson correlation was calculated from the hybrid/mid-parent or hybrid/BxM expression ratios to the hybrid/mid-parent plant height ratios (Supplemental Table 4) and plotted using ggridges [45]. KEGG maps were generated using pathview [46].

Protein group assignment

The plastid proteome was defined according to a previous report [47]. PhANGs were defined using KEGG maps zma00195, excluding the electron transport proteins, zma00196, and CornCyc reactions 1.97.1.12 and 1.10.3.9 [48,49], all of which were downloaded on February 26, 2020. The PhAPGs were defined as the plastid-encoded proteins in PSI, PSII, cytochrome b6f, and ATP synthase. Nuclear-encoded plastid ribosomal proteins were defined as those containing “30S” or “50S” in their NCBI annotation as well as “plastid” or “chloroplast” in their maize-GAMER GO annotations [50], with term descriptions retrieved using the R package AnnotationDbi [51]. Cytosolic ribosomal proteins were defined as those containing “40S” or “60S” in their NCBI annotation, except for those containing “mitochondrial” or “biogenesis”. Ethylene biosynthetic proteins were defined using CornCyc reactions 2.5.1.6, 4.4.1.14 and 1.14.17.4 [49], downloaded on May 5, 2020. All accession conversions were made using the MaizeGDB accession conversion table, downloaded on May 18, 2020 [49]. Proteins involved in carbon fixation, biosynthesis of secondary metabolites, and alpha-linoleic acid metabolism were defined using KEGG maps zma00710, zma01110, and zma00592, respectively. TPR and PTAC proteins were defined as those containing “TPR” and “plastid transcriptionally active” in their NCBI annotation, respectively. Oxidoreductase and protease proteins were defined as those with the term GO:0016491 and GO:0008233 in their maize-GAMER GO annotations [50], respectively. Protein biosynthesis proteins were defined as those containing “elongation factor” or “initiation factor” in their NCBI

annotation. A complete list of gene group assignments and their sources is provided in Supplemental Table 8.

Code availability

All source codes used for expression analyses are available on Github: https://github.com/devonbirdseye/HeterosisManuscript

Acknowledgements

This material is based upon work supported by the National Science Foundation Grant No. 1546899 to S.P.B and N.M.S., and the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1650112 to D.B. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Author Contributions

S.P.B., N.M.S., L.D., Z.S., and D.B. conceived the experiments. Z.S., D.B. and P.Z. performed the experiments. D.B. performed the data analysis and developed visualizations of data. L.D. provided feedback on the data analysis and visualizations. S.P.B and D.B. wrote the paper. E.A.S. and H.B. provided expertise, ideas, and feedback.

References

1. Schnable, P. S. & Springer, N. M. Progress toward understanding heterosis in crop plants. *Annu. Rev. Plant Biol.* **64**, 71–88 (2013).
2. Dobzhansky, T. Genetics of natural populations. XIX. Origin of heterosis through natural selection in populations of Drosophila pseudoobscura. *Genetics* **35**, 288–302 (1950).
3. Shull, G. H. What Is 'Heterosis'? *Genetics* **33**, 439–446 (1948).
4. Chen, K., Kumudini, S. V., Tollenaar, M. & Vyn, T. J. Plant biomass and nitrogen partitioning changes between silking and maturity in newer versus older maize hybrids. *F. Crop. Res.* **183**, 315–328 (2015).
5. Duvick, D. N. Genetic progress in yield of United States maize (Zea mays L.). *Maydica* vol. 50 193–202 (2005).

6. Stupar, R. M. *et al.* Gene expression analyses in maize inbreds and hybrids with varying levels of heterosis. *BMC Plant Biol.* **8**, 33 (2008).
7. Hu, X. *et al.* Genome-wide proteomic profiling reveals the role of dominance protein expression in heterosis in immature maize ears. *Sci. Rep.* **7**, 1–12 (2017).
8. Zhou, P. *et al.* Detecting small plant peptides using SPADA (Small Peptide Alignment Discovery Application). *BMC Bioinformatics* **14**, (2013).
9. Ko, D. K. *et al.* Temporal Shift of Circadian-Mediated Gene Expression and Carbon Fixation Contributes to Biomass Heterosis in Maize Hybrids. *PLoS Genet.* **12**, 1–31 (2016).
10. Habben, J. E. *et al.* Transgenic alteration of ethylene biosynthesis increases grain yield in maize under field drought-stress conditions. *Plant Biotechnol. J.* **12**, 685–693 (2014).
11. Song, Q. *et al.* Diurnal down-regulation of ethylene biosynthesis mediates biomass heterosis. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 5606–5611 (2018).
12. Fujimoto, R., Taylor, J. M., Shirasawa, S., Peacock, W. J. & Dennis, E. S. Heterosis of Arabidopsis hybrids between C24 and Col is associated with increased photosynthesis capacity. **109**, 7109–7114 (2012).
13. Song, Y. *et al.* Association of the molecular regulation of ear leaf senescence/stress response and photosynthesis/metabolism with heterosis at the reproductive stage in maize. *Sci. Rep.* **6**, 1–12 (2016).
14. Li, Z. *et al.* Temporal Regulation of the Metabolome and Proteome in Photosynthetic and Photorespiratory Pathways Contributes to Maize Heterosis. *Plant Cell* **32**, tpc.00320.2020 (2020).
15. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
16. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
17. Dudley, J. W., Maroof, M. A. S. & Rufener, G. K. Molecular Markers and Grouping of Parents in Maize Breeding Programs. *Crop Sci.* **31**, (1991).
18. Freeman, K. W. *et al.* By-plant prediction of corn forage biomass and nitrogen uptake at various growth stages using remote sensing and plant height. *Agron. J.* **99**, 530–536

(2007).

19. Young, T. E., Meeley, R. B. & Gallie, D. R. ACC synthase expression regulates leaf performance and drought tolerance in maize. *Plant J.* **40**, 813–825 (2004).
20. Park, C. H. *et al.* Arabidopsis ACC oxidase 1 coordinated by multiple signals mediates ethylene biosynthesis and is involved in root development. *Mol. Cells* **41**, 923–932 (2018).
21. Cheng, W. H., Chiang, M. H., Hwang, S. G. & Lin, P. C. Antagonism between abscisic acid and ethylene in Arabidopsis acts in parallel with the reciprocal regulation of their metabolism and signaling pathways. *Plant Mol. Biol.* **71**, 61–80 (2009).
22. Mao, D. *et al.* FERONIA receptor kinase interacts with S-adenosylmethionine synthetase and suppresses S-adenosylmethionine production and ethylene biosynthesis in Arabidopsis. *Plant Cell Environ.* **38**, 2566–2574 (2015).
23. Guo, B. *et al.* Comparative Proteomic Analysis of Embryos between a Maize Hybrid and Its Parental Lines during Early Stages of Seed Germination. *PLoS One* **8**, (2013).
24. Marcon, C. *et al.* Heterosis-associated proteome analyses of maize (Zea mays L.) seminal roots by quantitative label-free LC – MS ☆. *J. Proteomics* **93**, 295–302 (2013).
25. Björkeroth, J. *et al.* Proteome reallocation from amino acid biosynthesis to ribosomes enables yeast to grow faster in rich media. *Proc. Natl. Acad. Sci.* **117**, 201921890 (2020).
26. Sharma, M. & Pandey, G. K. Expansion and function of repeat domain proteins during stress and development in plants. *Front. Plant Sci.* **6**, 1–15 (2016).
27. Wang, D. *et al.* Comparative proteomic analysis reveals that the Heterosis of two maize hybrids is related to enhancement of stress response and photosynthesis respectively. *BMC Plant Biol.* **21**, 1–15 (2021).
28. Walley, J. W. *et al.* Reconstruction of protein networks from an atlas of maize seed proteotypes. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E4808-17 (2013).
29. Liu, W., He, G. & Deng, X. W. Biological pathway expression complementation contributes to biomass heterosis in Arabidopsis . *Proc. Natl. Acad. Sci.* **118**, e2023278118 (2021).
30. Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020).
31. Jiao, Y. *et al.* Improved maize reference genome with single-molecule technologies. *Nature* **546**, 524–527 (2017).

32. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
33. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
34. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, (2010).
35. Sun, S. *et al.* Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. *Nat. Genet.* **50**, 1289–1295 (2018).
36. Zhou, P., Hirsch, C. N., Briggs, S. P. & Springer, N. M. Dynamic Patterns of Gene Expression Additivity and Regulatory Variation throughout Maize Development. *Mol. Plant* **12**, 410–425 (2019).
37. Wickham, H., François, R., Henry, L. & Müller, K. dplyr: A Grammar of Data Manipulation. (2019).
38. Wickham, H. The Split-Apply-Combine strategy for data analysis. *J. Stat. Softw.* **40**, 1–29 (2011).
39. Wickham, H. & Henry, L. tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions. (2019).
40. Wickham, H. Reshaping Data with the {reshape} Package. *J. Stat. Softw.* **21**, 1–20 (2007).
41. Dowle, M. & Srinivasan, A. data.table: Extension of `data.frame`. (2019).
42. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).
43. Wilke, C. O. cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. (2019).
44. Pedersen, T. patchwork: The Composer of Plots. (2019).
45. Wilke, C. O. ggridges: Ridgeline Plots in 'ggplot2'. (2020).
46. Luo, W. & Brouwer, C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **29**, 1830–1831 (2013).
47. Friso, G., Majeran, W., Huang, M., Sun, Q. & van Wijk, K. J. Reconstruction of metabolic pathways, protein expression, and homeostasis machineries across maize bundle sheath and mesophyll chloroplasts: Large-scale quantitative proteomics using the first maize genome assembly. *Plant Physiol.* **152**, 1219–1250 (2010).
48. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms.

Protein Sci. **28**, 1947–1951 (2019).

49. Portwood, J. L. *et al.* Maizegdb 2018: The maize multi-genome genetics and genomics database. *Nucleic Acids Res.* **47**, D1146–D1154 (2019).

50. Wimalanathan, K., Friedberg, I., Andorf, C. M. & Lawrence-Dill, C. J. Maize GO Annotation—Methods, Evaluation, and Review (maize-GAMER). *Plant Direct* **2**, 1–15 (2018).

51. Pagès, H., Carlson, M., Falcon, S. & Li, N. AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor. (2019).

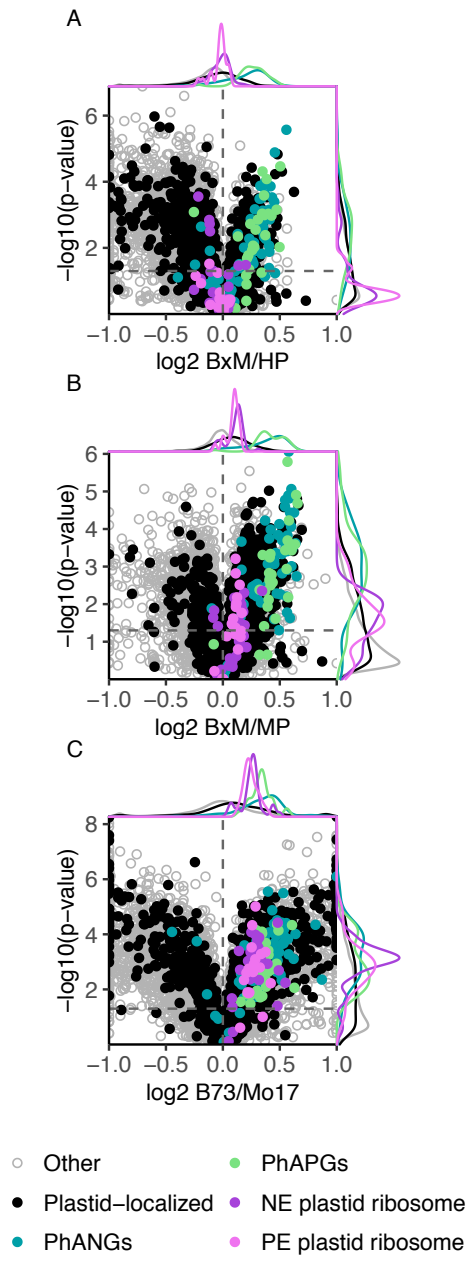

Figure 1. Volcano plots displaying expression patterns of the most significantly non-additive proteins, representing B73xMo17/high-parent (A), B73xMo17 /mid-parent (B), and B73/Mo17 (C). Photosynthesis-associated Nuclear Genes (PhANGs), Photosynthesis-associated Plastid Genes (PhAPGs), Nuclear-Encoded (NE) plastid ribosome, and Plastid-Encoded (PE) plastid ribosome proteins are color-coded.

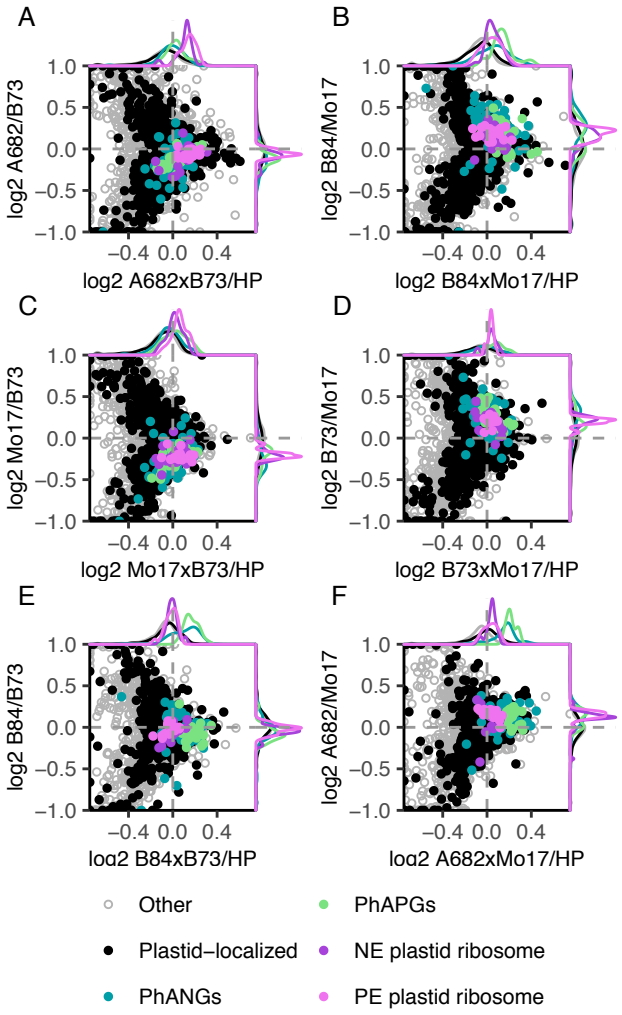

Figure 2. Volcano plots displaying plastid-localized proteins in seedling leaves of six hybrids relative to high-parent (HP) levels. Hybrids represented in A-F are ordered from greatest plant height heterosis (A) to least plant height heterosis (F). Each data point represents the mean of three biological replicates. Photosynthesis-associated Nuclear Genes (PhANGs), Photosynthesis-associated Plastid Genes (PhAPGs), Plastid-Encoded (PE) plastid ribosomes, and Nuclear-Encoded (NE) plastid ribosomes are color-coded.

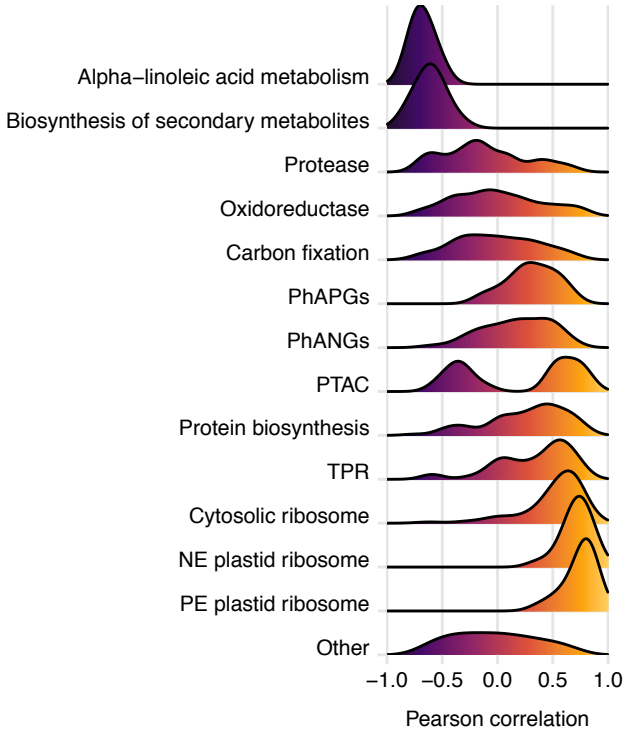

Figure 3. Density curves of Pearson correlations between protein expression heterosis and plant height heterosis in the combined RIL and six hybrids datasets.

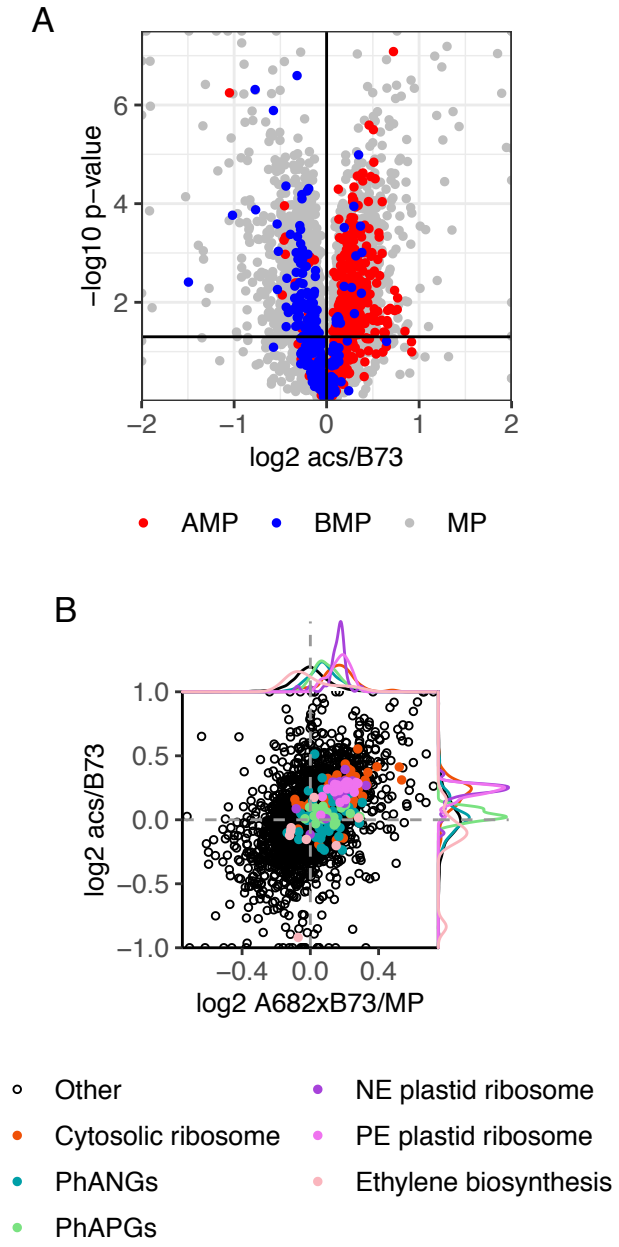

Figure 4. (A) Volcano plot of protein expression in the *acs2-6* double mutant in B73 background relative to B73. Each data point represents the mean of five biological replicates. Colors correspond to their expression in the hybrid with greatest plant height heterosis, A682xB73, designated as mid-parent (MP), above mid-parent (AMP), or below mid-parent (BMP) (B) Plastid-localized proteins in seedling leaves of the *acs2-6* double mutant in B73 background relative to B73, compared to expression in A682xB73 relative to MP. Each data point represents the mean of five biological replicates for *acs2-6*/B73 and three biological replicates for the hybrid and its inbred parents. Photosynthesis-Associated Nuclear Genes (PhANGs), Photosynthesis-Associated Plastid Genes (PhAPGs), Plastid-Encoded (PE) plastid ribosomes, Nuclear-Encoded (NE) plastid ribosomes, and ethylene biosynthesis proteins are color-coded.

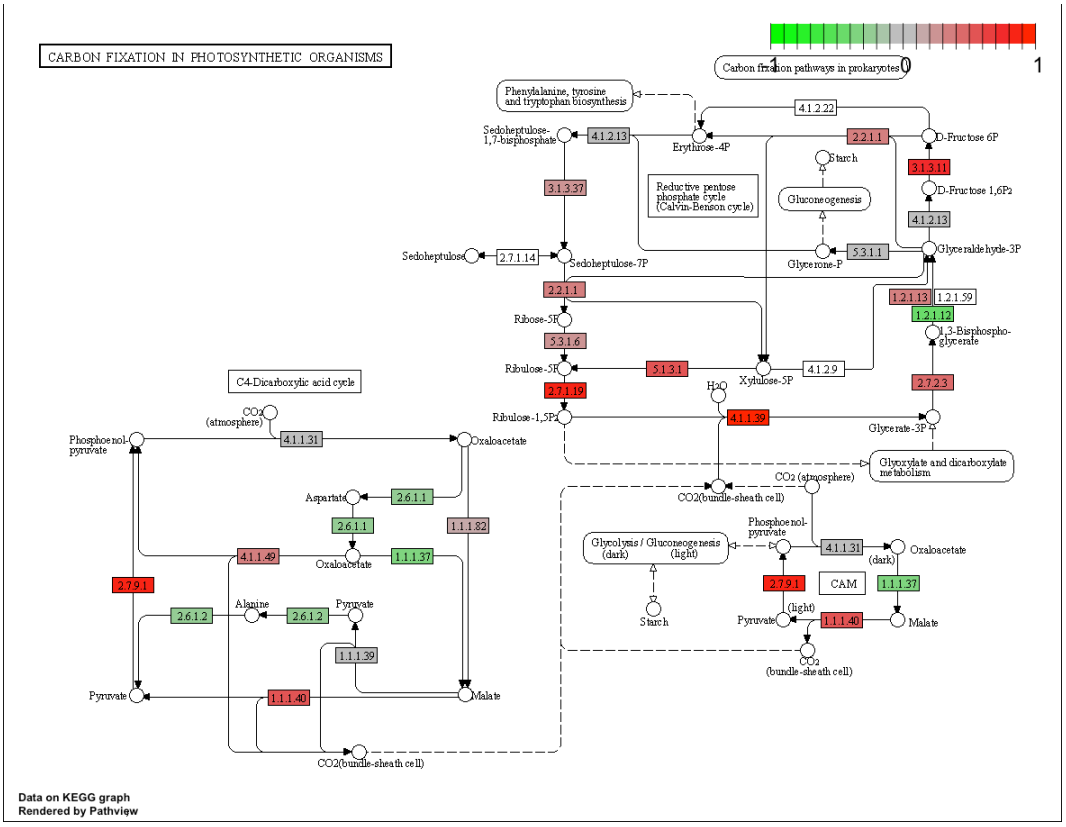

Figure S1. Expression of carbon fixation enzymes in the hybrid B73xMo17 relative to its mid-parent level

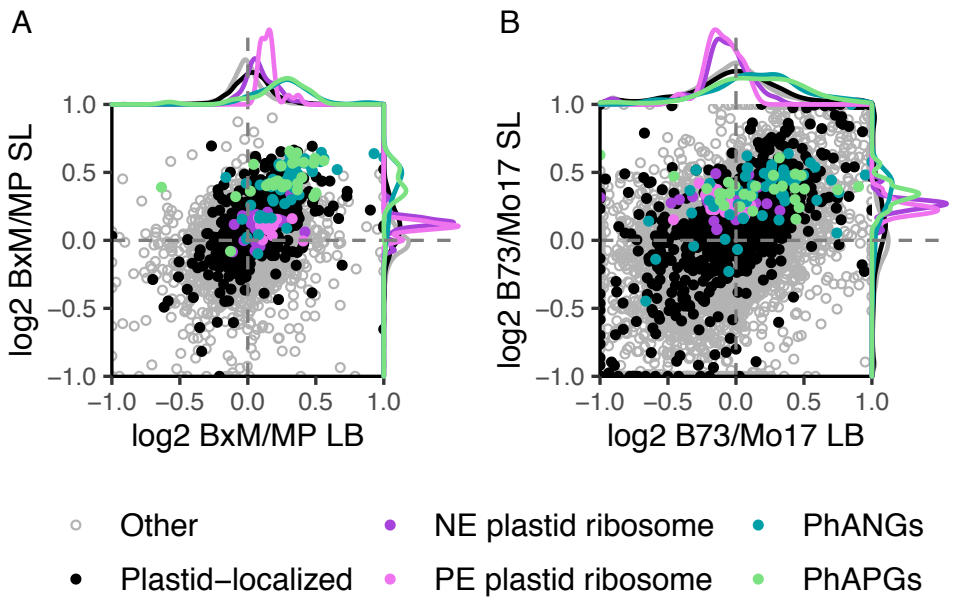

Figure S2. Comparison of expression of the significantly non-additive proteins in seedling leaf versus leaf blade tissue. A compares hybrid/mid-parent values; B compares parent/parent values. Photosynthesis-associated Nuclear Genes (PhANGs), Photosynthesis-associated Plastid Genes (PhAPGs), Plastid-Encoded (PE) plastid ribosomes, and Nuclear-Encoded (NE) plastid ribosomes are color-coded.

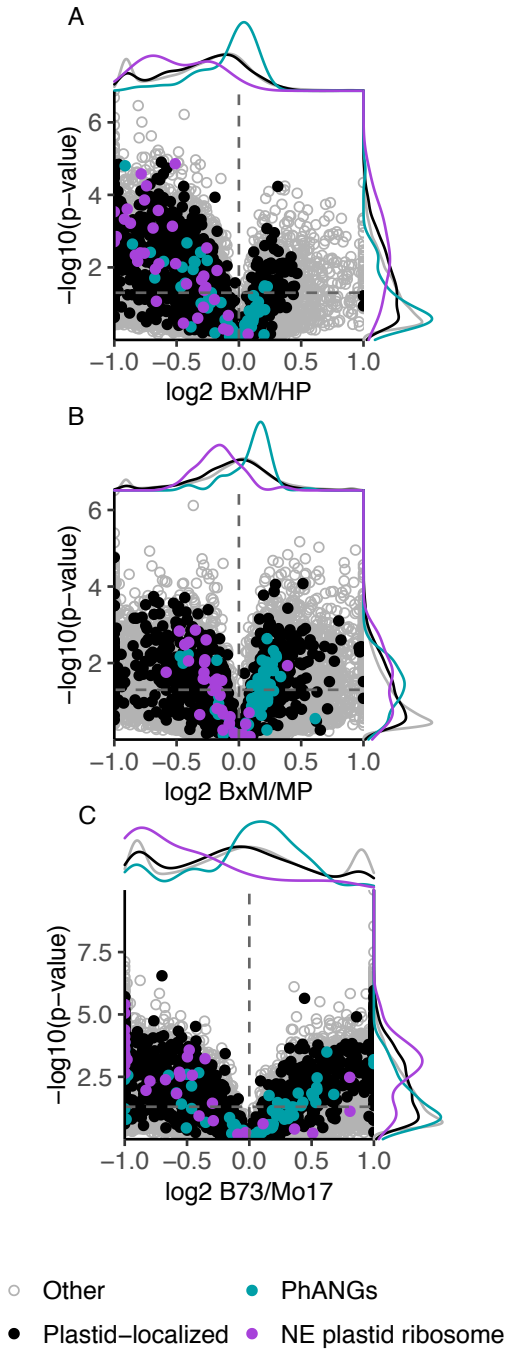

Figure S3. Volcano plots displaying RNA expression patterns of the most significantly non-additive proteins, representing B73xMo17/high-parent (A), B73xMo17 /mid-parent (B), and B73/Mo17 (C). Photosynthesis-associated Nuclear Genes (PhANGs), Nuclear-Encoded (NE) plastid ribosome, and are color-coded.

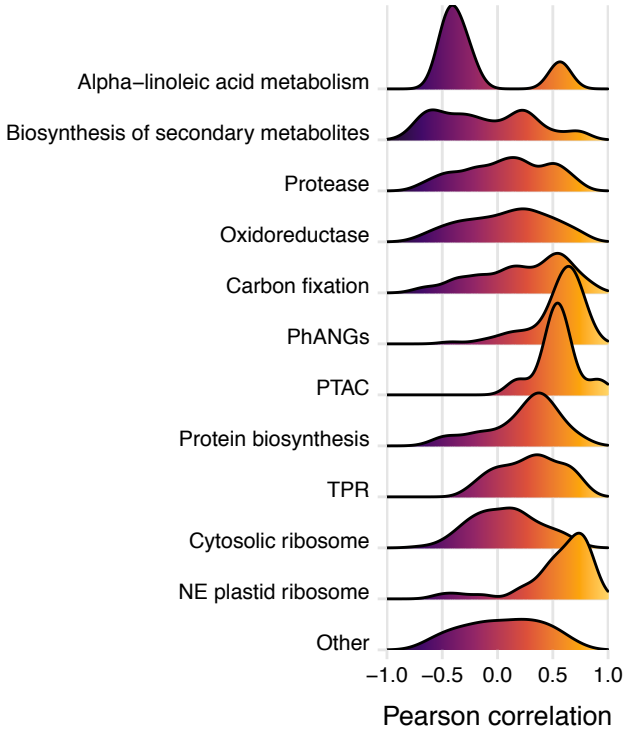

Figure S4. Density curves of Pearson correlations between RNA expression heterosis and plant height heterosis in the combined RIL and six hybrids datasets.

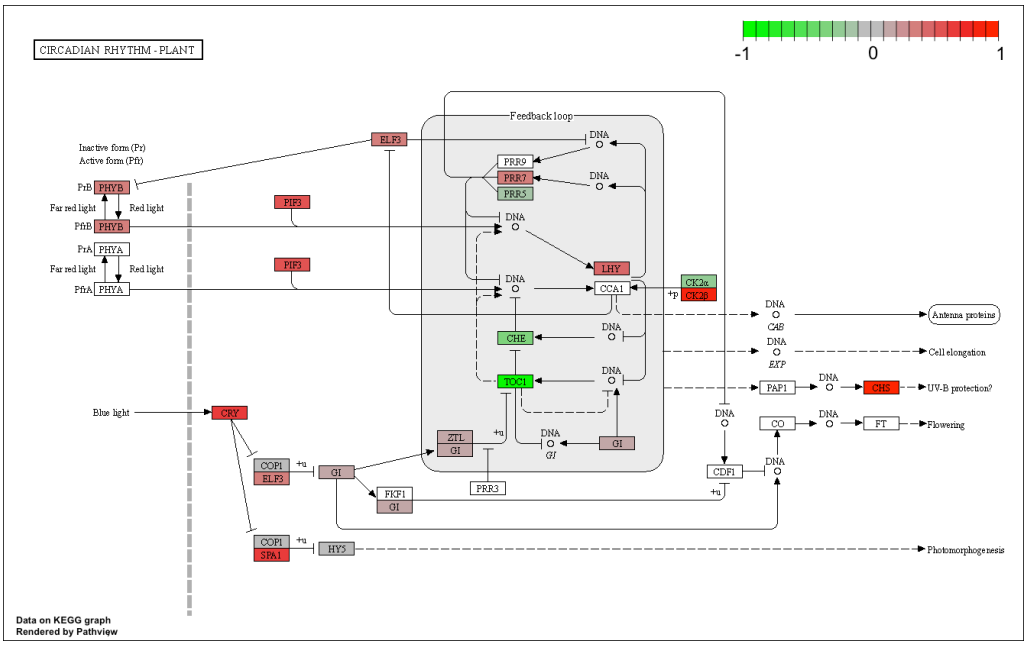

Figure S5. Expression of circadian rhythm-related transcripts in the hybrid, B73xMo17, relative to its mid-parent level.

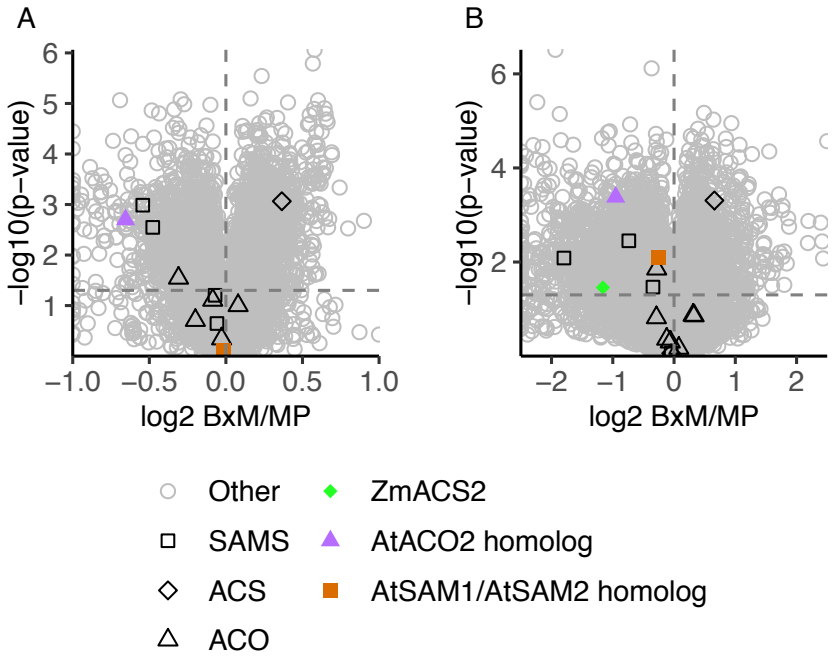

Figure S6. Volcano plots showing expression heterosis of ethylene biosynthesis enzymes in the proteome (A) and transcriptome (B) of the BxM hybrid. Of the six enzymes for which the mutant is known to produce reduced ethylene (ZmACS2, ZmACS6, AtACO1, AtACO2, AtSAMS1, and AtSAMS2), two were detected in the proteome (the AtACO2 homolog and the AtSAMS1/AtSAMS2 homolog) and three were detected in the transcriptome (ZmACS2, the AtACO2 homolog and the AtSAMS1/AtSAMS2 homolog).

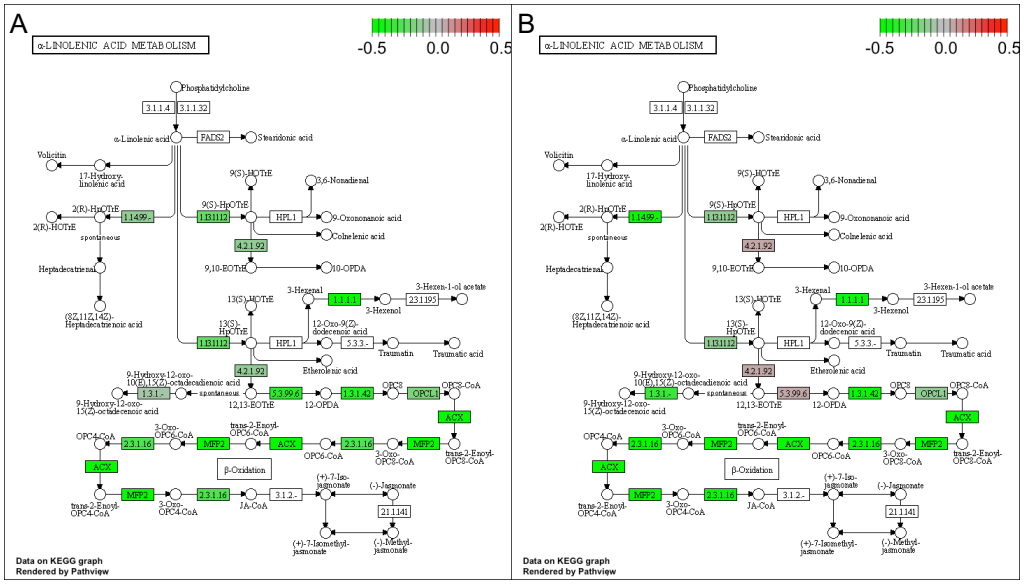

Figure S7. Expression of JA biosynthesis enzymes in the (A) hybrid B73xMo17 relative to its mid-parent level; and (B) in the ET biosynthesis mutant (*acs2-6*) compared to its inbred B73 background.