



## OPEN ACCESS

## EDITED BY

Yidan Ouyang,  
Huazhong Agricultural University, China

## REVIEWED BY

Zheng Yuan,  
Shanghai Jiao Tong University, China  
Xiao-Meng Wu,  
Huazhong Agricultural University, China

## \*CORRESPONDENCE

Z. Jeffrey Chen  
✉ zjchen@austin.utexas.edu

RECEIVED 04 July 2023

ACCEPTED 28 August 2023

PUBLISHED 14 September 2023

## CITATION

June V, Xu D, Papoulas O, Boutz D, Marcotte EM and Chen ZJ (2023) Protein nonadditive expression and solubility contribute to heterosis in *Arabidopsis* hybrids and allotetraploids. *Front. Plant Sci.* 14:1252564. doi: 10.3389/fpls.2023.1252564

## COPYRIGHT

© 2023 June, Xu, Papoulas, Boutz, Marcotte and Chen. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Protein nonadditive expression and solubility contribute to heterosis in *Arabidopsis* hybrids and allotetraploids

Viviana June<sup>1</sup>, Dongqing Xu<sup>1,2</sup>, Ophelia Papoulas<sup>1</sup>, Daniel Boutz<sup>1</sup>, Edward M. Marcotte<sup>1</sup> and Z. Jeffrey Chen<sup>1\*</sup>

<sup>1</sup>Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX, United States,

<sup>2</sup>State Key Laboratory of Crop Genetics and Germplasm Enhancement, College of Agriculture, Nanjing Agricultural University, Nanjing, China

Hybrid vigor or heterosis has been widely applied in agriculture and extensively studied using genetic and gene expression approaches. However, the biochemical mechanism underlying heterosis remains elusive. One theory suggests that a decrease in protein aggregation may occur in hybrids due to the presence of protein variants between parental alleles, but it has not been experimentally tested. Here, we report comparative analysis of soluble and insoluble proteomes in *Arabidopsis* intraspecific and interspecific hybrids or allotetraploids formed between *A. thaliana* and *A. arenosa*. Both allotetraploids and intraspecific hybrids displayed nonadditive expression (unequal to the sum of the two parents) of the proteins, most of which were involved in biotic and abiotic stress responses. In the allotetraploids, homoeolog-expression bias was not observed among all proteins examined but accounted for 17-20% of the nonadditively expressed proteins, consistent with the transcriptome results. Among expression-biased homoeologs, there were more *A. thaliana*-biased than *A. arenosa*-biased homoeologs. Analysis of the insoluble and soluble proteomes revealed more soluble proteins in the hybrids than their parents but not in the allotetraploids. Most proteins in ribosomal biosynthesis and in the thylakoid lumen, membrane, and stroma were in the soluble fractions, indicating a role of protein stability in photosynthetic activities for promoting growth. Thus, nonadditive expression of stress-responsive proteins and increased solubility of photosynthetic proteins may contribute to heterosis in *Arabidopsis* hybrids and allotetraploids and possibly hybrid crops.

## KEYWORDS

heterosis, proteome, protein solubility, hybrids, allopolyploids, genetics, genomics

## Introduction

Interspecific hybridization in plants often leads to allopolyploids including most important crops such as wheat, cotton, and canola, while many other crops such as corn and sorghum are grown as hybrids. Both allopolyploids and hybrids show hybrid vigor or heterosis. Heterosis or hybrid vigor refers to the observation that hybrid offspring show

greater growth and fitness than either parent and occurs across plant and animal kingdoms. phenomenon was systematically described by Charles Darwin in 1876 (Darwin, 1876), and rediscovered by Shull and East during maize breeding (Shull, 1908; East, 1936). Several genetic models are available to explain heterosis. The dominance model suggests complementation of deleterious alleles by the dominant ones in the heterozygous loci (Bruce, 1910; Jones, 1917). The overdominance model indicates that heterozygous loci in hybrids are expressed at a higher level than or advantageous over homozygous loci (East, 1936; Crow, 1948). Another model is related to epistasis, in which interactions between nonallelic genes contribute to the growth vigor in hybrids (Schnell and Cockerham, 1992; Yu et al., 1997). However, no single model can fully explain the basis of heterosis.

A notion in the field is to jump outside theoretical dogmas because these genetic models cannot address epistasis or complex regulatory network interactions in various biological pathways (Birchler et al., 2010). Indeed, transcriptomic analyses have revealed genome-wide nonadditive gene expression changes in *Arabidopsis* allotetraploids or interspecific hybrids (Wang et al., 2006b), which led to the discovery of linking enhanced circadian rhythms with biomass heterosis in plant hybrids (Ni et al., 2009). Expression peaks of circadian clock genes are epigenetically altered in the hybrids to enhance expression of the circadian output genes in photosynthesis and starch biosynthesis. The more starch is synthesized during the day, the more it can be degraded at night to promote growth (Chen, 2013). The role of altered circadian rhythms in heterosis has been consistently demonstrated in *Arabidopsis* (Shen et al., 2012; Miller et al., 2015; Yang L. et al., 2021), rice (Shen et al., 2015), and maize (Ko et al., 2016; Li et al., 2020; Birdseye et al., 2021), suggesting a conserved role of enhanced circadian rhythms in hybrid vigor.

Studies of proteomic changes in hybrids are very limited. Using protein two-dimensional gel electrophoresis analysis of the proteins extracted from mitochondria, Dahal et al. found a correlation between expression of specific alleles and/or post-translational modification of specific proteins and higher levels of heterosis in different maize hybrids (Dahal et al., 2012). Using isobaric tags for relative and absolute quantitation (iTRAQ) coupled with mass spectrometry, Ng et al. found that expression of ~8% of the proteins in *Arabidopsis* allotetraploids are nonadditive relative to the parents (mid-parent level) (Ng et al., 2012). Although the overall trend of nonadditive expression is consistent between transcript and protein levels, the percentage of differentially accumulated proteins that matched differentially expressed genes is relatively low. In natural allopolyploid *Tragopogon mirus*, hybridization generates more effects on proteomes than polyploidy (Koh et al., 2012).

In maize hybrids, metabolic changes correspond to nonadditive protein abundance and enzyme activities of key enzymes in the respective pathways, suggesting that concerted changes in metabolomes and proteomes contribute to maize heterosis (Li et al., 2020). Another study indicates increased expression of nuclear- and plastid-encoded subunits of protein complexes required for protein synthesis in chloroplasts and for photosynthetic activities in hybrid seedling leaves, and hybrid/mid-parent expression ratios of chloroplast ribosomal proteins

are correlated with plant height heterosis (Birdseye et al., 2021). These results suggest that post-transcriptional regulation and protein synthesis play a role in regulating the nonadditive expression of proteins in hybrids (Ng et al., 2012; Yang X. et al., 2021).

Metabolic and proteomic studies in maize further demonstrate that a large fraction of maize metabolites and proteins is diurnally regulated, and many show nonadditive abundance in the hybrids (Li et al., 2020). Metabolic heterosis is relatively mild, and metabolites in the photosynthetic pathway show positive mid-parent heterosis (MPH), whereas metabolites in the photorespiratory pathway show negative MPH. Hybrids may more effectively remove toxic metabolites generated during photorespiration, and thus maintain higher photosynthetic efficiency for heterosis. The cause of these changes remains elusive. One possibility is that the presence of multiple different alleles of a single gene in hybrids allows for selective expression of the more stable alleles (Goff, 2011). Fewer misfolded and aggregated proteins would increase metabolic efficiency in hybrids, as less energy would be required to refold or degrade misfolded proteins, and less protein synthesis would be required (Kristensen et al., 2002; Pedersen et al., 2005). This is because the presence of multiple different alleles of a single gene in hybrids allows for selective expression of the more stable alleles (Goff, 2011). Alternatively, the presence of these alternate alleles leads to a general increase in solubility through disrupting the homotypic aggregation of proteins (Ginn, 2010; Ginn, 2017). The increased metabolic efficiency caused by decreased protein aggregation would present a unified model for heterosis, but protein solubility has not been studied in plant hybrids.

Here, we investigated both changes in protein abundance and solubility in two sets of hybrids: reciprocal intraspecific hybrids between *Arabidopsis thaliana* ecotypes C24 and Col-0 (Miller et al., 2015), and allotetraploids *A. suecica* and Allo738 and their progenitors *A. arenosa* and *A. thaliana* (Wang et al., 2006b; Jiang et al., 2021). *A. thaliana* intraspecific hybrids have been extensively used as a model to study heterosis, as several hybrids (including Col/C24 hybrids) display high levels of growth vigor (Chen, 2013; Groszmann et al., 2014). However, the parental ecotypes have similar genomes with fewer non-synonymous mutations compared to interspecific hybrids or allotetraploids, which also display increased levels of heterosis (Chen, 2010; Chen, 2013). A comparison of proteome changes between allotetraploids and intraspecific hybrids would allow for testing the effect of genetic distance on protein changes.

We applied a protein fractionation approach coupled with label-free liquid chromatography-mass spectrometry (LC-MS) to investigate proteomic changes in the intraspecific hybrids and allotetraploids. We found nonadditive expression of proteins in stress response, photosynthesis, and protein biosynthesis in the hybrids and allotetraploids, which are consistent with transcriptome results related to heterosis. There were more soluble proteins in the intraspecific hybrids relative to the parents, but not in the allotetraploids. Most ribosomal proteins and proteins in the thylakoid lumen, membrane, and stroma, were in the soluble fractions. These results may suggest a role of nonadditive regulation

of stress-responsive and photosynthetic proteins in heterosis. Alternatively, reduced levels of protein synthesis may contribute to growth vigor in the hybrids and allotetraploids.

## Methods

### Plant materials

Two *A. thaliana* ecotypes Columbia (Col-0) and C24 were used as parents to generate reciprocal intraspecific hybrids by manually crossing as previously described (Miller et al., 2012). Each parent was also manually crossed as a control. Seeds were collected from these crosses once siliques had matured. Allotetraploid Allo738 was derived from an induced autotetraploid *A. thaliana* Ler ecotype (Ath4; ABRC CS3900) and *A. arenosa* (Aar, Care-1; ABRC; CS3901), an outcrossing tetraploid species (Comai et al., 2000; Wang et al., 2006b). Natural allotetraploid *A. suecica* strain As9502 (As; ABRC CS22509) and all other parental strains (Ath4 and Aar) were maintained in the lab.

### Plant growth conditions

Seeds were sterilized in 20% bleach for 10 minutes, followed by five rinses with 1 mL sterile ddH<sub>2</sub>O. Seeds were then plated onto 0.5 Murashige and Skoog media supplemented with 1% sucrose and stratified at 4°C in the dark for 48 hours. After stratification, seeds were transferred to a 22°C growth room with 16 hours of light and 8 hours of dark per day. Seven days after germination, seedlings were transplanted onto soil. A 3:1 mixture of Pro-Mix Biofungicide to Field and Fairway was used, and at first watering, plants were treated with 4g Miracle Gro Plant Food and 1 tsp Gnatrol Biological Larvicide (Valent Biosciences LLC, Libertyville, IL) per gallon of water. Plants were sprayed with Bonide copper soap fungicide weekly to prevent powdery mildew infection and with pesticide weekly to prevent thrips infestation.

### Protein extraction and fractionation

At 21 days after sowing, rosettes were harvested at zeitgeber time (ZT) 0 (dawn) to minimize circadian effects with 3 biological replicates for each genotype and flash-frozen in liquid nitrogen. A pool of 10 rosettes from 10 individual plants grown in a similar developmental stage was ground to a fine powder in a chilled mortar and pestle. An equivalent volume of lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 5 mM EGTA, 10% glycerol, 1% NP40) with plant protease inhibitor cocktail (Sigma-Aldrich, St. Louis, MO) and phosphatase inhibitor (PhosSTOP Easy, Roche, Basel, Switzerland) was added to each sample. Samples were then lysed at 4°C on a rotator for 30 minutes. Debris was pelleted via centrifugation at 1,000 g for 10 minutes. The supernatant was retained as the whole cell extract. The whole cell extract was then fractionated into the soluble and insoluble fractions through centrifugation at 10,000 g for 10 minutes. The supernatant was retained as the soluble fraction, and the pellet was resuspended in

lysis buffer to form the insoluble fraction. Fractions were then denatured in 50% trifluoroethanol (TFE) and 5 mM tris (2-carboxyethyl phosphine) (TCEP) at 55°C for 45 minutes. Samples were cooled to room temperature and alkylated in 15 mM iodoacetamide (IAM) at room temperature in the dark for 30 minutes. After the alkylation reaction was quenched with 7 mM dithiothreitol, the samples were diluted in trypsin digestion buffer (50 mM Tris, 2mM CaCl<sub>2</sub>, pH 8.0) to reduce the final TFE concentration to 5%. After adding 2 µg MS grade trypsin in the intraspecific hybrids (Pierce Biotechnology, Waltham, MA) and polyploids (Promega Corporation, Madison, WI) to each sample, the samples were digested at 37°C for 5 hours. Formic acid was added to a final concentration of 1% to quench the digestion. Sample volumes were reduced in a SpeedVac to 250 µL. Samples were then filtered using Amicon Ultra 10kD (Millipore Sigma, Burlington, MA) spin-caps to remove undigested protein and eluted in buffer C [95% H<sub>2</sub>O, 5% acetonitrile (ACN), 0.1% formic acid]. Samples were desalted using a 5-7 µL C18 Filter Plate (Glygen Corp.) and a vacuum manifold, eluted in 60% ACN, and reduced in volume to <10 µL in a SpeedVac. The final samples were resuspended in buffer C for mass spectrometry.

### Mass spectrometry

Mass spectra from each of three biological replicates were acquired on a Thermo Orbitrap Fusion Lumos. Peptides were separated using reverse phase chromatography on a Dionex Ultimate 3000 RSLCnano UHPLC system (Thermo Fisher Scientific, Waltham, MA) with a C18 trap to Acclaim C18 PepMap RSLC column (Dionex; Thermo Fisher Scientific) configuration. Peptides were eluted using a 5-40% acetonitrile gradient in 0.1% formic acid over 120 min for all samples. Peptides were injected directly into the mass spectrometer using nano-electrospray for data-dependent tandem mass spectrometry. The data acquisition used for the mass spectrometer was as follows: full precursor ion scans (MS1) collected at 120,000 m/z resolution. Monoisotopic precursor selection and charge-state screening were enabled using Advanced Peak Determination (APD), with ions of charge > +1 selected for high energy collision dissociation (HCD) with collision energy 30% stepped ± 3%. Dynamic exclusion was active with 20-second exclusion for ions selected twice within a 20 s window for intraspecific hybrid samples, and with 60 s exclusion for ions selected twice within a 60-second window for polyploid samples. All MS2 scans were centroid and done in rapid mode.

### Peptide assignment

For *A. thaliana*, the proteome was downloaded from Uniprot in July 2018 (UniProt, 2021). For the allotetraploids, the proteome was generated from the recent long read resequencing of the Allo738 genome (Jiang et al., 2021). We then created an orthogroup collapsed proteome by concatenating the sequences of all proteins within orthogroups with triple lysines between each protein, as described in a published paper (McWhite et al., 2020). Orthogroups used to create the proteome were those identified in a previously

published paper (Jiang et al., 2021). Peptide assignment was performed using Proteome Discoverer (v2.3 for the allotetraploid, and v2.2 for the *A. thaliana* hybrids). The MS spectra were searched against these proteomes as well as a database of common contaminants from MaxQuant using the SEQUEST HT node. For the search, a maximum of two missed trypsin cleavage sites was allowed. For MS1, a mass tolerance of 10 ppm was allowed, and for MS2, a mass tolerance of 0.6 Da was allowed. A maximum of 3 equal modifications were allowed per peptide, and 4 maximum dynamic modifications were allowed per peptide. For dynamic modifications, oxidation (+15.995 Da) was allowed, and for static modifications, carbamidomethyl (+57.021 Da) was allowed. We used the Percolator node to assign peptide spectral matches (PSMs) and for the decoy database search using a strict FDR of 1%. The Minora Feature Detector node was used to calculate extracted-ion chromatogram (XIC) peak area for quantitation, with a minimum trace length of 5, a minimum number of 2 peaks, and a max  $\Delta RT$  of 0.2 for isotope pattern multiplets.

## Protein quantification

The MSstats package (v. 3.22.1) was used to calculate protein level quantitation from peptide data, as well as to perform differential abundance analysis between fractions and samples (Choi et al., 2014). Peptides with only one or two counts across runs were removed, as were proteins with only one peptide. Only unique peptides were used for protein quantitation. Median normalization was performed to normalize extracted ion chromatogram (XIC) peak area across biological replicates and fractions. Protein quantification from peptides was performed using the TOP3 method, and missing values were imputed using an accelerated failure model.

## Differential abundance analysis

The MSstats package (v. 3.22.1) was used to perform differential abundance analysis to identify non-additively expressed proteins. A linear mixed model was used to calculate fold changes and p-values. The mean protein abundance of the hybrid was contrasted against the mean of both parental protein abundance means. Only proteins with measurements in at least two biological replicates per genotype in the soluble fractions were considered. Proteins with p-value  $\leq 0.05$  and  $\log_2 FC \geq |0.5|$  were considered differentially expressed. We used uncorrected p-values as using Benjamini-Hochberg adjusted p-values resulted in the identification of no differentially expressed proteins due to the relatively high variability among the samples. The use of multiple testing correction, although reducing the incidence of Type I errors (false positives), may increase Type II errors (false negatives), as observed in other proteomics studies (Pascovici et al., 2016). As a possible remedy, we used a fold-change threshold that may reduce the number of false positives. PCA analysis was performed using the `prcomp` function in R and drawn using `ggbiplot`.

## Solubility shift analysis

For this analysis, proteins that were not quantified in all three biological replicates or all fractions were discarded. In base R, a two-way ANOVA, with fraction (insoluble/soluble) and genotype (progenitors/hybrid), and the corresponding interaction term was performed. Proteins with a significant ( $P \leq 0.05$ ) interaction term displayed a significant shift in solubility between the parents and the hybrid. To quantify the degree to which solubility shifts between the parents and the hybrids, as well as the direction of this shift, a solubility score was calculated. For each biological replicate, the ratio of protein in the soluble fraction to the insoluble fraction was calculated. The median ratio for each progenitor and hybrid was then used for further analysis. Median ratios were used due to the high variability between biological replicates in the insoluble fraction. To get the mid-parent value, the mean was taken of the median ratios for each parent. The following formula was then used to calculate the overall solubility shift:

$$\text{Solubility score} = \frac{(\text{Soluble}_{MPV})}{(\text{Insoluble}_{MPV})} - \frac{(\text{Soluble}_{Hybrid})}{(\text{Insoluble}_{Hybrid})}$$

Proteins with p-value  $\leq 0.05$  and solubility score  $\geq 0.5$  were classified as being significantly more soluble in the parents than in the hybrids, and proteins with p-value  $\leq 0.05$  and solubility score  $\leq -0.5$  were classified as being significantly more soluble in the hybrids than in the parents.

## Homoeolog-specific protein expression

We assigned peptides to individual homoeologs using the `assigned_peptides` script from PIVO (<https://github.com/marcottelab/pivo>) (Drew et al., 2020). These peptide matches were then intersected with peptides that were uniquely assigned to an orthogroup. They were filtered to identify peptides that match proteins belonging to either the *A. thaliana* or *A. arenosa* sub-genome for each orthogroup. *A. thaliana* and *A. arenosa* specific peptides were summed separately by orthogroup for each sample. Orthogroups where peptides in the At4 or Aa samples matched to the incorrect parental proteome were discarded. Samples where >75% of peptides (>3:1 ratio) matched either the *A. arenosa* or *A. thaliana* proteome were classed as being biased towards that proteome.

## Gene ontology (GO) analysis

GO analysis was performed using the TopGO package (v. 2.42.0) using the `elim` algorithm (<https://bioconductor.org/packages/release/bioc/html/topGO.html>). GO annotations for *A. thaliana* from `org.At.tair.db` were downloaded for enrichment analysis. For the polyploids, orthogroups were annotated by lifting GO annotations from the *A. thaliana* proteins in each orthogroup, using GO annotations downloaded from Ensembl BioMart (Kinsella et al., 2011). Orthogroups without an *A. thaliana* member were annotated using InterProScan annotations for the orthogroups assigned in a previous paper (Jiang et al., 2021).

## Aggregation propensity and instability predictions:

Aggregation propensity was calculated using the TANGO algorithm (Fernandez-Escamilla et al., 2004). Instability scores were calculated using the ProtParam tool from ExPasy (<https://web.expasy.org/protparam>).

## RNA-seq analysis

Previously collected RNA-seq data from our lab was used to investigate homoeolog-specific RNA expression in Allo738 and *A. suecica* (NCBI's Gene Expression Omnibus accession numbers GSE29687 and GSE50715) (Shi et al., 2015). Reads were trimmed using trimmomatic (Bolger et al., 2014). Reads were then mapped to the Allo738 genome from Jiang et al., 2021, using STAR (Dobin et al., 2013) using the following settings—outFilterMismatchNoverLmax 0.04 —outFilterMultimapNmax 20 —alignIntronMin 25 —alignIntronMax 3000. Reads were then filtered to identify uniquely mapped reads using samtools (using the -q 60 setting) (Barnett et al., 2011). For differential expression analysis, reads overlapping each gene were counted using HTseq using the union and reverse stranded settings. EdgeR was used to calculate CPM values for each locus (Robinson et al., 2010). Log<sub>2</sub>-fold change (LFC) was calculated between homoeologs within orthogroups. Samples where there was a LFC >2 between homoeologs from either the *A. arenosa* or *A. thaliana* subgenome were classed as “biased.”

## Results

### Proteome in *Arabidopsis* intraspecific hybrids and allotetraploids

We investigated the proteomes of reciprocal hybrids between the *A. thaliana* accessions Col and C24 (Miller et al., 2015), a natural allotetraploid *A. suecica* (As), and a resynthesized allotetraploid Allo738, and their progenitors (Wang et al., 2006b). *A. thaliana* Col and C24 diverged after the last glacial period (Figure 1A), about 10,000 years ago (Consortium, 2016), while *A. thaliana* (At4) and *A. arenosa* (Aa) diverged around ~6 million years ago and hybridized to form *A. suecica* 16,000–300,000 years ago (Novikova et al., 2017; Jiang et al., 2021). Both the intraspecific hybrids and allotetraploids display high levels of growth vigor (Figure 1B), and the level of biomass vigor is higher in the allotetraploids than in intraspecific hybrids, indicating a role of genetic distance in heterosis (Chen, 2013; Miller et al., 2015).

We separated native protein extracts into soluble and insoluble fractions using a native and non-denaturing protein extraction method (see Methods), with NP-40 (1%), a non-ionic detergent, at 10,000g centrifugation, and analyzed with label-free liquid chromatography mass spectrometry (LC-MS/MS). These proteins were in normal distributions (Supplementary Figure 1) and highly reproducible among three biological replicates in intraspecific hybrids and their parents (Supplementary Figure 2) and

allotetraploids and their progenitors (Supplementary Figure 3). We identified a total of 5,144 protein groups, out of 12,769 (Castellana et al., 2008), across all fractions in the intraspecific hybrids, which were filtered down to 2,600 protein groups after removal of non-unique peptides and proteins with few supporting peptides (Supplementary Dataset 1). The recent genome assembly of Allo738 (Jiang et al., 2021), comprising the At and Aa subgenomes, was used to generate a proteome for allotetraploids. To increase protein identifications in the polyploid species, we used an orthogroup-collapsed approach (McWhite et al., 2020) for peptide assignment to preserve peptides that were mapped onto both subgenomes in the allotetraploids. The use of this approach had two primary benefits. Firstly, we identified 200 more protein groups, and 87,476 more peptide spectrum matches when orthogroups were collapsed than when only the *A. thaliana* proteome was used for peptide assignment. Secondly, it allowed us to evaluate nonadditive expression of proteins in Allo738 and *A. suecica* relative to At4 and Aa. In the allotetraploids, we identified 4,927 protein orthogroups, which were reduced to 2,519 protein orthogroups after removal of lower quality ones (Supplementary Dataset 2). Principal component analysis (PCA) of protein abundance in both *A. thaliana* hybrids (Figure 1C) and allotetraploids (Figure 1D) showed clear separation by fractions (PC1) and by genotypes (PC2). In both allotetraploids and intraspecific hybrids, the largest separation was between the two parents Col and C24 for the hybrids (Figure 1C) and Aa and At4 for the allotetraploids (Figure 1D), with the hybrids and allotetraploids falling between their respective parents. There was a greater spread along PC2 between allotetraploid progenitors, At4 and Aa (Figure 1D), than between *A. thaliana* hybrid parents (Col and C24) (Figure 1C), which could reflect the increased genetic diversity between At4 and Aa compared to Col and C24.

### Proteins are nonadditively expressed in *Arabidopsis* hybrids and allotetraploids

We evaluated protein abundance levels in both allotetraploids and intraspecific hybrids compared to the mid-parent value (MPV) (Supplementary Dataset 3), and differentially expressed proteins between the respective parents (Supplementary Dataset 4). In the intraspecific hybrids, numbers of nonadditively expressed proteins ( $\log_2FC > |0.5|$ ;  $p < 0.05$ ) were 109 and 73 in F<sub>1</sub> (ColxC24, by convention the maternal parent is listed first in a genetic cross) and the reciprocal F<sub>1</sub> (C24xCol), respectively (Figures 2A, C), and 279 and 228 proteins were nonadditively expressed in As and Allo738, respectively (Figures 2B, D). In the intraspecific hybrids, twice as many proteins that were down-regulated than up-regulated, consistent with more down-regulated genes than up-regulated genes in the transcriptome study (Miller et al., 2015). However, numbers of up-regulated and down-regulated proteins were relatively equal in both allotetraploids, which were consistent with previous proteomic data (Ng et al., 2012) but inconsistent with microarray data (Wang et al., 2006b). This may suggest a discordance between protein and transcript abundance (Ng et al., 2012) and/or different stages of plant materials assayed between two studies.

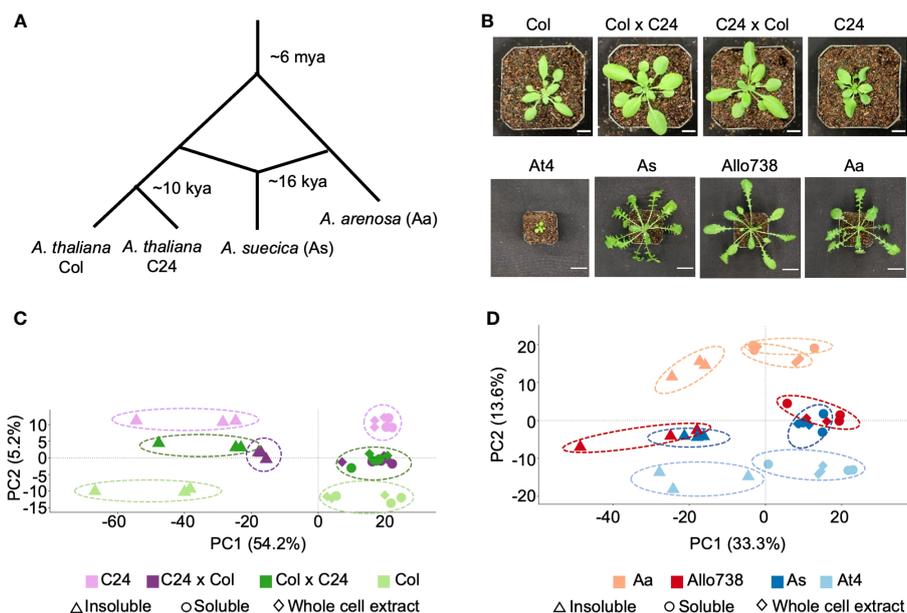


FIGURE 1

Proteome diversity in *Arabidopsis* intraspecific hybrids and allotetraploids. (A) Diagram of genetic divergence between *Arabidopsis* species and accessions. Mya: million years ago; kya: thousand years ago. (B) Photographs of *Arabidopsis* intraspecific hybrids and their parents, and of allotetraploids and their (extant) progenitors. Scale bars = 10 mm (intraspecific hybrids) and 30 mm (allotetraploids). (C) PCA plot showing protein abundance identified in the whole cell extract (diamond), soluble fraction (circle), and insoluble (triangle) fractions in the intraspecific hybrids (C24xCol, dark purple and ColxC24, dark green) between C24 (light purple) and Col (light green). There is separation of the samples along PC2, and separation of the samples by fraction along PC1 with percentage of variation explained (%). (D) PCA plot showing protein abundance in the whole cell extract (diamond), soluble (circle), and insoluble (triangle) fractions in *Arabidopsis* allotetraploids (Allo738, dark red and As, dark blue) and *A. thaliana* (light blue) and *A. arenosa* (light red). As with the intraspecific hybrids, there is separation of the samples by genotype along PC2, and separation of the samples by fraction along PC1.

The allotetraploids exhibit an increased genetic diversity between their progenitors, as well as an increased level of growth vigor compared to the intraspecific hybrids. This is reflected in the number of nonadditively expressed proteins identified. There was a large degree of overlap in proteins that showed nonadditive expression in the allotetraploids; 98 proteins (43.0%) and 89 proteins (31.9%) were nonadditively expressed in Allo738 and *A. suecica*, respectively, and were also differentially expressed between *A. thaliana* and *A. arenosa* (Supplementary Figures 4A, B). In the F<sub>1</sub> hybrids, 37 proteins (33.9%) in C24xCol and 27 proteins (37.0%) in ColxC24 were nonadditively expressed and showed differential expression between the parents Col and C24 (Supplementary Figures 4C, D). This high-level overlap suggests that protein differences between the parents need to be modified or reconciled in the intraspecific hybrids and allotetraploids, a notion supported by the transcriptome studies (Wang et al., 2006b; Miller et al., 2015).

## Gene ontology (GO) enrichment of nonadditively expressed proteins

GO analysis identified several functional terms as significantly enriched in the nonadditively expressed proteins in both the intraspecific hybrids and allotetraploids (Figures 2E, F). The GO enrichment terms of the nonadditively expressed proteins were

much more similar between the two reciprocal intraspecific hybrids than between the allotetraploids. In the intraspecific hybrids, a number of GO enrichment terms were related to stress response, which is consistent with overrepresentation of nonadditively expressed stress-responsive genes in both *Arabidopsis* intraspecific hybrids (Miller et al., 2015) and allotetraploids (Wang et al., 2006b). Interestingly, the GO enrichment of upregulated proteins was related to the abiotic stress response, such as response to cold (GO:0009409), toxin catabolic process (GO:0009407), and response to acid-containing chemical (GO:1901700) (Figure 2E), while GO terms of down-regulated proteins were related to the biotic stress response, such as defense response to bacterium (GO:0042742) and defense response to other organism (GO:0098542) (Figure 2F).

The GO enrichment categories showed little overlap between the nonadditively expressed proteins in As and Allo738, probably because of the large difference between the resynthesized (Allo738) and natural (As) allotetraploids. In Allo738, upregulation of the proteins involved in response to cytokinin (GO:0009735) and cold (GO:0009409) (Figure 2E) may suggest that natural *A. suecica*, with its origin in northern Europe (O'Kane et al., 1995; Lind-Hallden et al., 2002), has adapted to cold response. GO term enrichment of the down-regulated proteins was related to RNA and protein metabolism (Figure 2F), including tRNA aminoacylation for protein translation (GO:0016070) and proteolysis (GO:0006508). These protein expression changes agree with previous findings. For

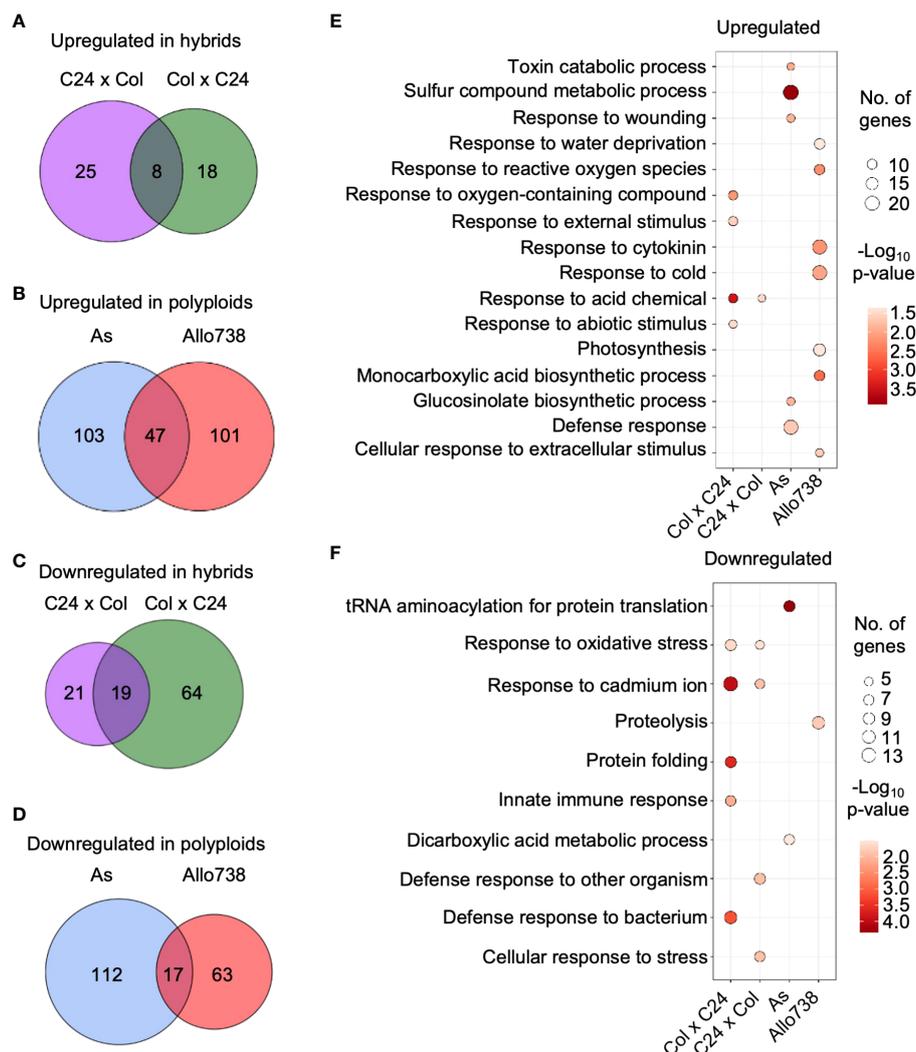


FIGURE 2

Nonadditive expression of proteins in intraspecific hybrids and allotetraploids. (A) Nonadditively expressed proteins (upregulated) in the hybrids relative to the mid-parent value (MPV). Venn diagrams indicate overlap between Col x C24 and C24 x Col. (B) Nonadditively expressed proteins (upregulated) in the allotetraploids relative to the MPV. Venn diagrams indicate overlap between As and Allo738. (C) Nonadditively expressed proteins (down-regulated) in the hybrids relative to the MPV. Venn diagrams indicate overlap between ColxC24 and C24xCol. (D) Nonadditively expressed proteins (down-regulated) in allotetraploids relative to the MPV. Venn diagrams indicate overlap between Allo738 and As. (E) Gene ontology analysis of biological process enrichment for the upregulated proteins in the hybrids and allotetraploids relative to the MPV. (F) GO analysis of biological process enrichment for the downregulated proteins in the hybrids and allotetraploids relative to the MPV.

example, in maize downregulation of proteins is related to proteasome formation and amino acid biosynthesis (Li et al., 2020), and in *Drosophila*, increased inbreeding is associated with an increase in *HSP70* expression (Kristensen et al., 2002).

## Analysis of soluble and insoluble proteomes in *Arabidopsis* hybrids and allotetraploids

Theoretical analyses suggest that misfolded proteins can form protein aggregates, leading to proteasomal degradation (Ginn, 2010; Ginn, 2017) (Figure 3A). Alternatively, coding sequence variants between alleles in hybrids could lead to a reduction in the rate of self-association during protein folding, leading to a decrease in

protein misfolding and aggregation in hybrids (Figure 3B). To investigate this, we employed a fractionation scheme previously used to investigate protein solubility shifts in *S. cerevisiae* in response to heat shock – the proteins enriched in the insoluble fraction were found to form foci after heat shock (O'Connell et al., 2014). This method uses a 10,000g centrifugation step to separate the soluble and insoluble fractions from whole cell extract (Figure 3C).

A similar method was previously used to separate insoluble and soluble proteins in *Arabidopsis* on the basis of aggregation propensity as calculated using the TANGO algorithm (Fernandez-Escamilla et al., 2004). We therefore evaluated whether there was a significant difference in the TANGO scores of proteins enriched in the soluble and insoluble fractions of the proteome in our samples. The insoluble fraction had proteins with a

significantly higher mean TANGO score than the proteins in the soluble fraction ( $P = 1.36 \times 10^{-7}$ , Wilcoxon Rank Sum test), indicating that it is enriched in aggregating proteins (Figure 3D).

GO analysis found that proteins more abundant in the soluble fraction were represented many cellular components and most cell regions, whereas proteins more abundant in the insoluble fraction only showed enrichment in a few cellular components primarily membrane-bound organelles such as the chloroplast envelope (GO:0009941) and the thylakoid membrane (GO:0009535) (Supplementary Figure 5).

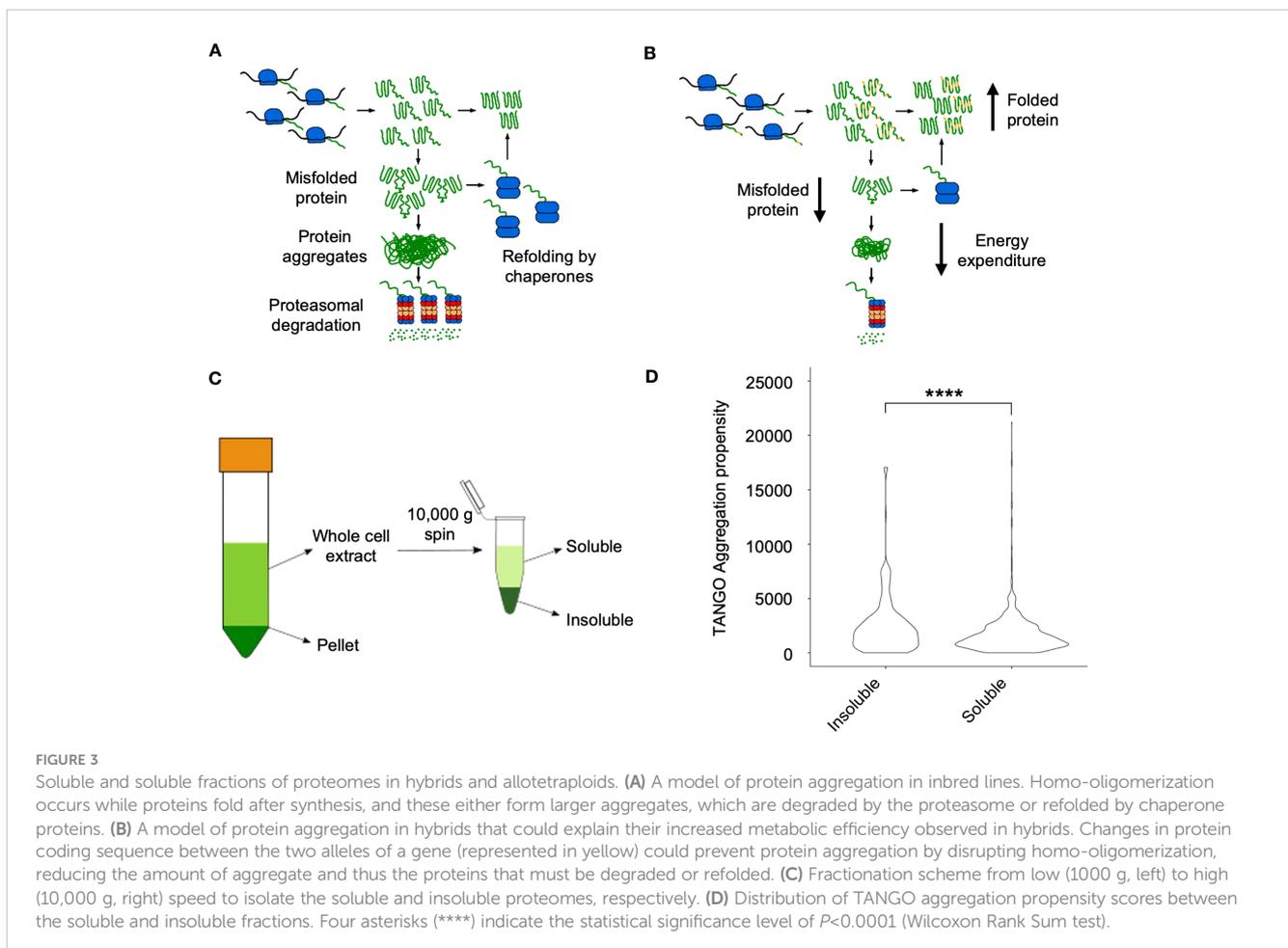
Cytosolic proteins such as ribosomal proteins were generally enriched in the soluble fraction instead of the insoluble fraction (Supplemental Dataset 5). As with the insoluble fraction, chloroplast localized proteins were enriched in the soluble fraction; however, unlike the insoluble fraction, proteins from the thylakoid lumen and stroma in addition to the thylakoid membrane were also enriched in the soluble fraction. This included both subunits of RuBisCO, which were significantly enriched in the soluble fraction of all samples (Figures 4B, C and Supplemental Dataset 5). This result argues that the abundance of thylakoid proteins in the chloroplast in the insoluble fraction is not due to intact chloroplasts accumulating in the insoluble fraction, but rather a reflection of the solubility of these proteins. This finding may also suggest a role for protein solubility in maintaining high photosynthetic activities, as they contribute to heterosis in

*Arabidopsis* (Ni et al., 2009; Miller et al., 2015; Yang L. et al., 2021), rice (Shen et al., 2015), and maize (Ko et al., 2016; Li et al., 2020).

## Changes in protein solubility between hybrids and their parents

To test a potential role of protein solubility changes in hybrid vigor, we examined whether there was a general shift in protein solubility of proteins between hybrids and their parents. Using ANOVA ( $P < 0.05$ ), we calculated pairwise ratios of the protein abundance between soluble and insoluble fractions. When the median ratio was greater than 0.5 between the MPV ratio and hybrid ratio in addition to a  $P$ -value of less than 0.05, the proteins were considered having a solubility shift between the hybrids and the parents (Figure 4A) (Supplementary Dataset 5).

In the intraspecific hybrids relative to the parents, there were more proteins in the soluble than in the insoluble fractions (Figures 4B, D). Among those soluble proteins that were localized in the chloroplast stroma, 14 out of 34 proteins were more soluble in both reciprocal hybrids than in their parents. More soluble proteins were identified in C24 x Col hybrids than in the reciprocal Col x C24 hybrids, probably because of the parent-of-origin effect (Ng et al., 2014). This effect on the transcriptome difference is related to RNA-directed DNA methylation,



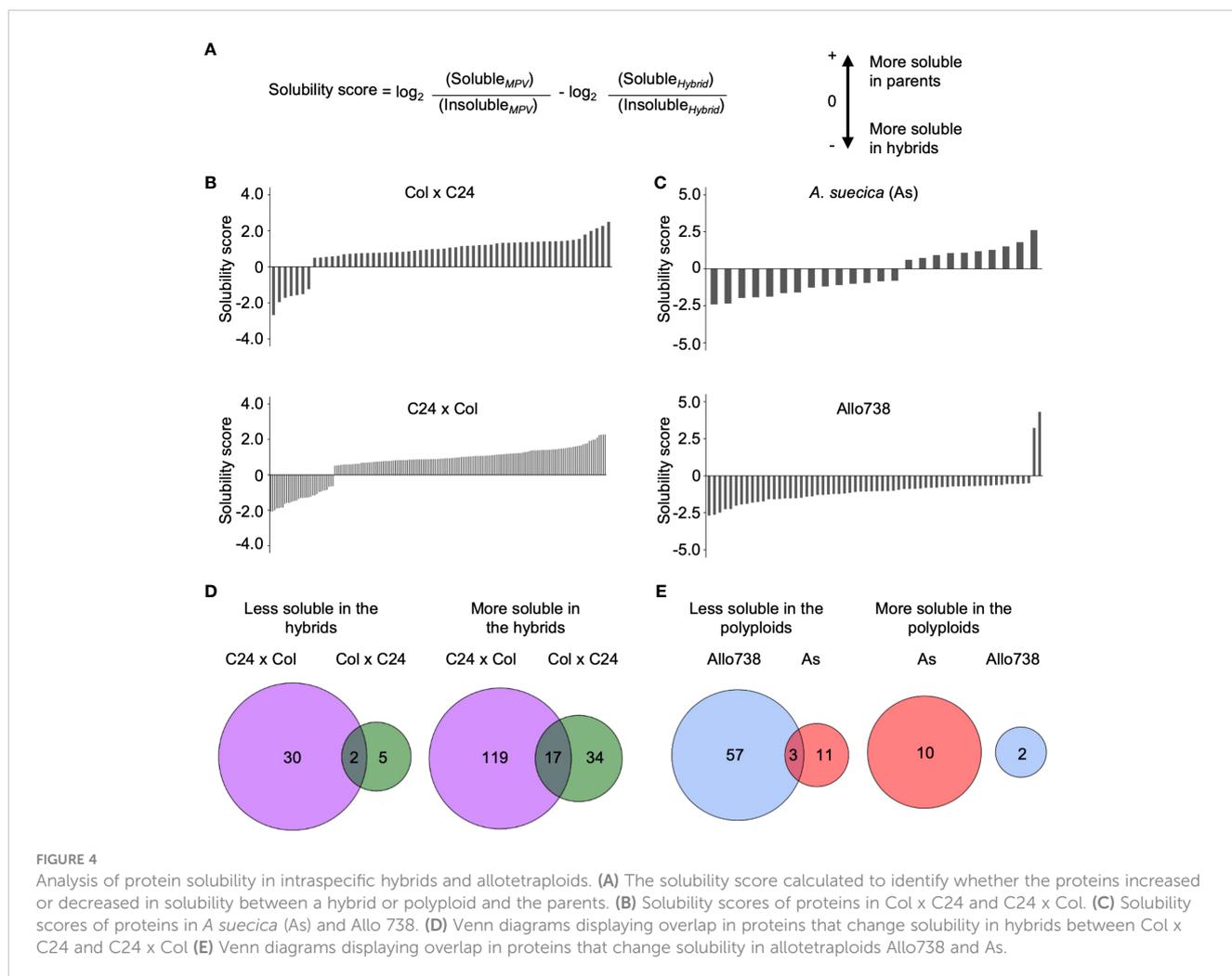


FIGURE 4

Analysis of protein solubility in intraspecific hybrids and allotetraploids. (A) The solubility score calculated to identify whether the proteins increased or decreased in solubility between a hybrid or polyploid and the parents. (B) Solubility scores of proteins in Col x C24 and C24 x Col. (C) Solubility scores of proteins in *A. suecica* (As) and Allo738. (D) Venn diagrams displaying overlap in proteins that change solubility in hybrids between Col x C24 and C24 x Col. (E) Venn diagrams displaying overlap in proteins that change solubility in allotetraploids Allo738 and As.

as previously reported (Ng et al., 2014). C24 x Col hybrids accumulate more starch and sugars than Col x C24 hybrids, which coincide with the increase in protein solubility. Very few proteins showed lower solubility in the hybrids than in their parents. Only two proteins were less soluble in both hybrids: *RBP31*, a chloroplast ribonucleoprotein, and *OEP16*, a chloroplast outer envelope pore protein.

Unexpectedly, fewer proteins displayed a solubility shift in the allotetraploids than in the intraspecific hybrids (Figures 4B, C). Ten and two proteins were more soluble in Allo738 and As, respectively (Figure 4E), compared to their progenitors, including a heat shock factor binding protein that is involved in acquired thermotolerance (Hsu et al., 2010). Three protein orthogroups showed a reduced solubility relative to both progenitors: an outer envelope membrane protein, a hydroxymethylglutaryl-CoA synthase involved in glucosinolate biosynthesis, and an orthogroup containing kinesin-like protein involved in cell division. In addition, there were more proteins that showed a decrease in protein solubility in the allotetraploids relative to their progenitors. These data may suggest protein solubility may not be directly related to genetic distance. Alternatively, protein solubility may change during different stages of development, as these allotetraploids grow slower and flower later than the diploids (Wang et al., 2006a).

## Expression of homoeolog-specific proteins in allotetraploids

The recent assembly of Allo738 genome (Jiang et al., 2021) allowed us to use the Allo738 proteome for peptide assignments. This improved reference proteome, along with the increased divergence between At4 and Aa helped us identify peptides that were unique to individual homoeologs in the allotetraploids (Supplementary Dataset 6). This would allow us to test if allelic-specific expression of proteins in hybrids and polyploids contributes to the metabolic efficiency in hybrids (Goff, 2011).

We used allele-specific peptides to calculate protein abundance in allotetraploids, and the abundance of proteins from the Aa and At sub-genomes were compared within each orthogroup. We found that similar numbers of proteins that displayed a bias towards either the *A. thaliana* or *A. arenosa* subgenome in Allo738 and natural *A. suecica* (Figures 5A, C); nearly 50% of these proteins in At-biased (Figure 5A) or Aa-biased (Figure 5C) group were shared between Allo738 and natural *A. suecica*. This finding is consistent with transcriptome data that no obvious expression dominance was found among multiple natural *A. suecica* accessions (Burns et al., 2021). Although expression dominance of specific homoeologs can

occur in the allotetraploids *Arabidopsis* (Wang et al., 2006a; Wang et al., 2006b), cotton (Adams et al., 2003; Zhang et al., 2015), and *Tragopogon* (Tate et al., 2006), our data support the notion of genomic and expression stability accompanied by epigenetic changes in many genetically stable allopolyploids like *Arabidopsis* (Jiang et al., 2021) and *Gossypium* (cotton) (Chen et al., 2020).

Among orthogroup proteins that display biased-homoeolog expression, *A. thaliana*-biased proteins showed more GO enrichment groups than *A. arenosa*-biased proteins (Figures 5B, D). *At*-biased proteins had more GO enrichment terms in Allo738 than natural *A. suecica* (As), many of which belonged to biosynthetic and metabolic processes, including RNA metabolic process (GO:0016070), organelle organization (GO:0006996), regulation of biosynthetic process (GO:0009889), gene expression (GO:0010467), and cellular nitrogen and aromatic compound metabolic processes. Many of these proteins are localized in chloroplasts; this may reflect inheritance of chloroplasts from the maternal *A. thaliana* ancestor of Allo738 (Wang et al., 2006b) and *A. suecica* (Sall et al., 2003). Alternatively, proteins, like transcripts, of *A. thaliana* origin, may be subject to biased expression (Wang et al., 2006b). The *A. arenosa*-biased proteins had fewer GO

enrichment terms, including response to light stimulus in both Allo738 and As, and the cellular response to stress in Allo738.

These homoeolog-biased proteins accounted for 20% of nonadditively expressed proteins in *A. suecica* and 17% in Allo738 (Supplementary Dataset 6), indicating a role of homoeolog-biased expression in the nonadditive protein accumulation in allotetraploids. Interestingly, about twice as many nonadditively expressed proteins that displayed homoeolog-expression bias were expressed above MPV than below MPV in As, despite the percentage of nonadditively expressed proteins displaying homoeolog-expression bias was similar in both allotetraploids. This contrasted with Allo738, in which equal numbers of nonadditively expressed proteins were expressed both above and below MPV. This may reflect changes in protein abundance (or silencing) between neo-allopolyploid Allo738 and old (natural) *A. suecica*.

To determine whether homoeolog-expression bias contributes to changes in protein solubility in hybrids, we estimated the mean TANGO score and instability score (Supplementary Figure 3) and compared them for both homoeologous proteins that displayed biased expression. If there was a trend towards expressing the more

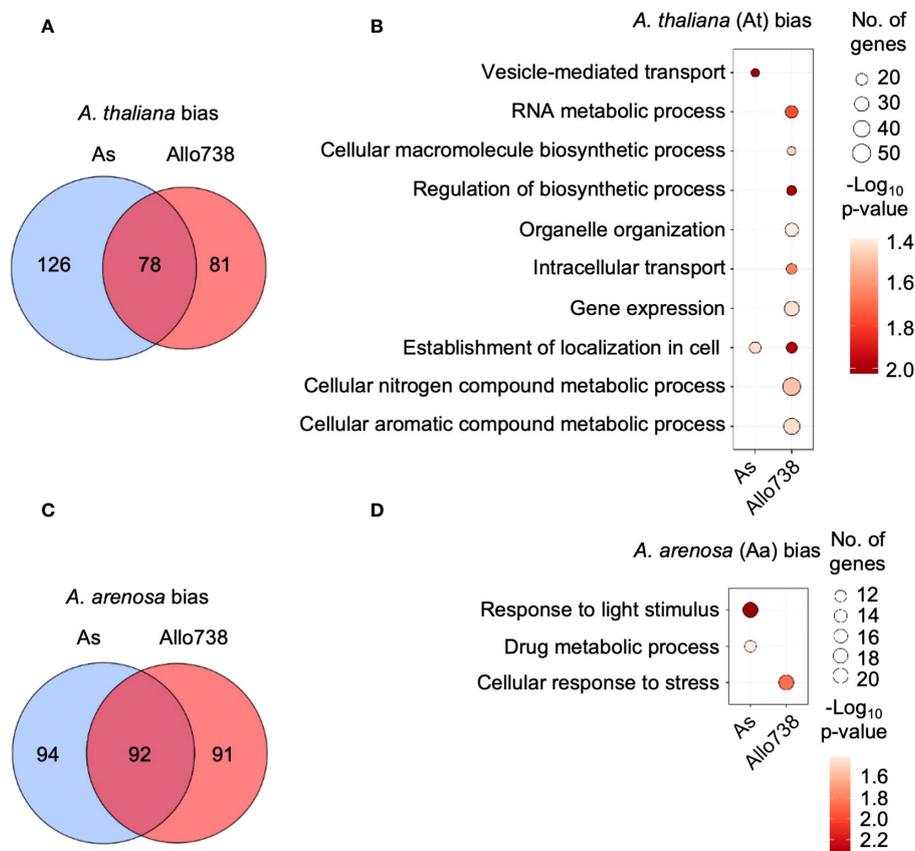


FIGURE 5

Homoeolog-specific expression of proteins in allotetraploids. (A) Venn diagram indicating the overlap between proteins that display biased expression towards the *A. thaliana* homoeologs in Allo738 and natural *A. suecica* (As). (B) Enrichment of GO biological process for proteins that display *A. thaliana* homoeolog-biased expression in Allo738 and As. (C) Venn diagram indicating the overlap between proteins that display biased expression of the *A. arenosa* homoeologs in Allo738 and As. (D) Enrichment of GO biological process for proteins that display *A. arenosa* homoeolog-biased expression in Allo738 and As.

stable homoeologs, we would expect to see an increase in the average solubility of At or Aa homoeologs. However, no obvious difference was observed in the aggregation propensity of the proteins that displayed biased expression in either Allo738 or *A. suecica* at the proteomic (Supplementary Figure 6) and transcriptomic (Supplementary Figure 7) levels. Our current data suggest that homoeolog-expression bias may not alter protein solubility in *Arabidopsis* allotetraploids.

## Discussion

### Nonadditive accumulation of stress-response proteins in intraspecific hybrids and allotetraploids

Our investigation of fractionated proteomes uncovered the role of non-additively expressed proteins in stress responses in *Arabidopsis* intraspecific hybrids and allotetraploids. Down-regulation of abiotic and biotic stress-responsive genes in normal conditions can save the energy to promote growth vigor (Miller et al., 2015). Results from the proteomic analysis largely support the findings of nonadditively expressed transcripts in *Arabidopsis* intraspecific hybrids (Miller et al., 2015) and allotetraploids (Wang et al., 2006b). For example, the biotic stress-responsive proteins that are downregulated include two proteins in the *PATHOGENESIS-RELATED GENES* family, *PR2* and *PR5*. Members of this gene family are also downregulated at mRNA levels in Col x C24 hybrids (Miller et al., 2015). Several genes encoding glutathione-S-transferase (*GST*), *GSTF6* and *GSFT2*, are downregulated in both reciprocal hybrids, and *GSTF7* and *GSTF8* are downregulated in one F<sub>1</sub> (Col x C24). *GST* genes are involved in response to bacterial or fungal infections by removing toxins associated with pathogen infection as well as in mediating a systemic immune response (Gullner et al., 2018). It is notable that in the diurnal transcriptome study (Miller et al., 2015), many abiotic stress-responsive genes were repressed in the afternoon, while biotic stress responsive genes were repressed in the morning as a trade-off mechanism for heterosis. In this study, the samples were collected at one time point (dawn), which may explain GO term enrichment of abiotic and biotic responsive proteins in the opposite directions. Alternatively, protein accumulation levels could be different from transcript abundance.

A reduction in oxidative stress can potentially downregulate protein metabolic machinery. In maize hybrids, catalase protein abundance is greater than mid-parent value at ZT21, leading to an overnight reduction in H<sub>2</sub>O<sub>2</sub> abundance (Li et al., 2020). In the allotetraploids, the orthogroup containing catalase gene orthologs (OG0000577) is significantly upregulated in Allo738 relative to the mid-parent value ( $P = 0.013$ ) and slightly in *A. suecica* ( $P = 0.053$ ). This increased expression of catalase in the allotetraploids may contribute to low levels of protein damage due to oxidative stress, thus leading to a reduction in the requirement of protein biosynthesis machinery in hybrids and polyploids.

Among the proteins upregulated relative to the MPV in both Allo738 and As, many are related to photosynthesis, consistent with

upregulation of these genes in resynthesized allotetraploids (Wang et al., 2006b; Ni et al., 2009). For example, *AMY3*, an alpha amylase protein involved in starch degradation, is upregulated in both allotetraploids, and its transcripts are also upregulated in resynthesized allotetraploids (Wang et al., 2006b; Ni et al., 2009). In addition, *PORC*, a protochlorophyllide oxidoreductase that is involved in the biosynthesis of chlorophyll, is upregulated in both allotetraploids. Other *POR* loci, such as *PORA* and *PORB* are also found to be consistently upregulated in allotetraploids (Wang et al., 2006b; Ni et al., 2009).

Upregulated proteins of *A. thaliana* homoeologs in the allotetraploids include  $\beta$ -glucosidases and jacalin-related lectin *JAL35*, which is involved in glucosinolate biosynthesis and ER body formation (Nagano et al., 2008). ER bodies are responsible for the formation of isothiocyanates, which are toxic to many herbivores (Wittstock et al., 2003). Upregulation of these proteins in the allotetraploids may contribute to glucosinolate turnover pathway. Furthermore, the *A. thaliana* homoeolog of the *GRP7* protein orthogroup is upregulated in both Allo738 and *A. suecica*, consistent with microarray results (Wang et al., 2006b). *GRP7* is an RNA binding protein that is involved in regulating circadian oscillation (Heintzen et al., 1997), as well as both biotic and abiotic stress-responsive genes (Meyer et al., 2017). Upregulation of this protein could mediate expression of stress-responsive genes by altering the circadian clock in the intraspecific hybrids and allotetraploids (Ni et al., 2009; Miller et al., 2015).

### Cytokinin responsive proteins were non-additively expressed in the allotetraploids

Phytohormones, including ethylene, salicylic acid, and auxin, have been shown to play roles in mediating growth vigor in hybrids (Wang et al., 2006b; Shen et al., 2012; Groszmann et al., 2015; Saeki et al., 2016; Song et al., 2018). In this study, we found upregulation of genes involved in response to cytokinin in the allotetraploids. Upregulation of the proteasome subunit *RPN12a* in Allo738 may be involved in promoting the degradation of inhibitors to the cytokinin response. For example, the *RPN12a* mutant shows slow leaf formation, reduced root elongation, and altered growth in response to exogenous cytokinins (Smalle et al., 2002). Cytokinins are generally involved in promoting cell division and plant growth – mutants that overexpress cytokinin biosynthesis genes are associated with increased shoot growth (Kieber and Schaller, 2014). This role of cytokinins in mediating heterosis in the allotetraploids remains to be tested.

### Changes in protein solubility in *Arabidopsis* intraspecific hybrids and allotetraploids

Theoretical studies of protein folding in yeast hybrids suggest hybrids have lower levels of protein aggregation, and thus more soluble proteins (Ginn, 2010; Ginn, 2017). Consistent with this, downregulation of genes involved in protein metabolism is observed in intraspecific hybrids in *Drosophila*

(Kristensen et al., 2002). Here we found a shift of protein solubility in the intraspecific hybrids relative to the mid-parent value, but not in the allotetraploids. It is possible that factors other than genetic distance affect the observed changes in solubility. Alternatively, a computational estimate of a protein's solubility in yeast hybrids may not reflect its stability *in vivo*. We also note that although the method of separating soluble and insoluble proteins has been successfully used in yeast studies (O'Connell et al., 2014), it should be refined for working with plants that have rigid cell walls and more debris than the yeast cells. Moreover, appropriate statistical methods and additional validation are needed to properly interpret these data.

Our results also confirm the enrichment of several components of the TIC-TOC complex (translocon on the inner chloroplast membrane - translocon on the inner chloroplast membrane), as well as many members of the photosystem II reaction center in the insoluble fraction in all samples. These proteins are largely located in the chloroplast's membranes and less soluble than cytosolic proteins. TIC214, a component of the TIC-TOC complex having amyloidogenic properties due to the QN-rich region of the protein's C-terminus (Antonets and Nizhnikov, 2017), is significantly enriched in the insoluble fraction of all samples where it was detected. Many of these proteins are highly abundant with about 80% of protein molecules in a mesophyll cell localized to the chloroplast (Heinemann et al., 2021).

In analysis of 400 and 350 proteins with homoeolog-specific expression in *A. suecica* and Allo738, respectively, we did not observe any significant differences in aggregation propensity of the homoeologs. This result is consistent with overall balanced expression among subgenomes (Jiang et al., 2021), despite expression bias can occur to rRNA genes and other protein-coding genes due to epigenetic changes (Chen et al., 1998; Lee and Chen, 2001).

What might cause the increase in metabolic efficiency and downregulation of protein biosynthesis? One possibility is novel functionality of protein complexes emerging from protein-protein interactions between different protein alleles in the hybrids (Herbst et al., 2017) or in the allotetraploids. In *S. cerevisiae* x *S. uvarum* hybrids, there is an overrepresentation of proteins involved in protein metabolism that displayed protein-protein interactions between diverged alleles (Berger and Landry, 2021; Dandage et al., 2021). Several complexes involved in protein metabolism, including the prefoldin complex and proteasome, consist of members from both parental copies. This is reminiscent of the abundance of metabolites and proteins in maize hybrids, where most amino acids show abundance peaks during the day and decrease at night (Li et al., 2020). In addition, investigation of the circadian control of protein synthesis in both *Arabidopsis* and dinoflagellates has found that ribosome loading and translation primarily occurs overnight (Cornelius et al., 1985; Missra et al., 2015). This may lead to the decreased expression in amino acid biosynthesis and tRNA synthesis as observed in this study and in maize hybrids (Li et al., 2020). Whether novel protein-protein interactions have altered function and impacted proteostasis in hybrids and hybrid vigor remains to be investigated.

## Data availability statement

All raw and interpreted mass spectrometry data were deposited to the ProteomeXchange <https://massive.ucsd.edu> with the MassIVE repository number MSV000089682 and ProteomeXChange number PXD034635. The datasets presented in this study can also be found in the article/Supplementary Material.

## Author contributions

VJ and ZJC conceived the research, analyzed the data, and wrote the paper. VJ, DX, OP, and DB performed the experiments. EM provided supervision, revision, technical, and intellectual support. All authors contributed to the article and approved the submitted version.

## Funding

The financial support for this work was partly provided by the National Institutes of Health (GM109076) to ZJC and the Welch Foundation (F1515) and Army Research Office (W911NF-12-1-0390) to EM.

## Acknowledgments

We thank Dr. Alan Lloyd at The University of Texas at Austin for supervision in the latter part of this project and Texas Advanced Computing Center for providing computing support for data analysis.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2023.1252564/full#supplementary-material>

## References

- Adams, K. L., Cronn, R., Percifield, R., and Wendel, J. F. (2003). Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc. Natl. Acad. Sci. U.S.A.* 100, 4649–4654. doi: 10.1073/pnas.0630618100
- Antonets, K. S., and Nizhnikov, A. A. (2017). Predicting amyloidogenic proteins in the proteomes of plants. *Int. J. Mol. Sci.* 18, 2155. doi: 10.3390/ijms18102155
- Barnett, D. W., Garrison, E. K., Quinlan, A. R., Stromberg, M. P., and Marth, G. T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* 27, 1691–1692. doi: 10.1093/bioinformatics/btr174
- Berger, C. M., and Landry, C. R. (2021). Yeast proteins do not practice social distancing as species hybridize. *Curr. Genet.* 67, 755–759. doi: 10.1007/s00294-021-01188-x
- Birchler, J. A., Yao, H., Chudalayandi, S., Vaiman, D., and Veitia, R. A. (2010). Heterosis. *Plant Cell* 22, 2105–2112. doi: 10.1105/tpc.110.076133
- Birdseye, D., de Boer, L. A., Bai, H., Zhou, P., Shen, Z., Schmelz, E. A., et al. (2021). Plant height heterosis is quantitatively associated with expression levels of plastid ribosomal proteins. *Proc. Natl. Acad. Sci. U.S.A.* 118, e2109332118. doi: 10.1073/pnas.2109332118
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1038/bioinformatics/bt170
- Bruce, A. B. (1910). The Mendelian theory of heredity and the augmentation of vigor. *Science* 32, 627–628. doi: 10.1126/science.32.827.627.b
- Burns, R., Mandakova, T., Gunis, J., Soto-Jimenez, L. M., Liu, C., Lysak, M. A., et al. (2021). Gradual evolution of allopolyploidy in *Arabidopsis suecica*. *Nat. Ecol. Evol.* 5, 1367–1381. doi: 10.1038/s41559-021-01525-w
- Castellana, N. E., Payne, S. H., Shen, Z., Stanke, M., Bafna, V., and Briggs, S. P. (2008). Discovery and revision of *Arabidopsis* genes by proteogenomics. *Proc. Natl. Acad. Sci. U.S.A.* 105, 21034–21038. doi: 10.1073/pnas.0811066106
- Chen, Z. J. (2010). Molecular mechanisms of polyploidy and hybrid vigor. *Trends Plant Sci.* 15, 57–71. doi: 10.1016/j.tplants.2009.12.003
- Chen, Z. J. (2013). Genomic and epigenetic insights into the molecular bases of heterosis. *Nat. Rev. Genet.* 14, 471–482. doi: 10.1038/nrg3503
- Chen, Z. J., Comai, L., and Pikaard, C. S. (1998). Gene dosage and stochastic effects determine the severity and direction of uniparental ribosomal RNA gene silencing (nuclear dominance) in *Arabidopsis* allopolyploids. *Proc. Natl. Acad. Sci. U.S.A.* 95, 14891–14896. doi: 10.1073/pnas.95.25.14891
- Chen, Z. J., Sreedasyam, A., Ando, A., Song, Q., De Santiago, L. M., Hulse-Kemp, A. M., et al. (2020). Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. *Nat. Genet.* 52, 525–533. doi: 10.1038/s41588-020-0614-5
- Choi, M., Chang, C. Y., Clough, T., Broudy, D., Killeen, T., MacLean, B., et al. (2014). MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* 30, 2524–2526. doi: 10.1093/bioinformatics/btu305
- Comai, L., Tyagi, A. P., Winter, K., Holmes-Davis, R., Reynolds, S. H., Stevens, Y., et al. (2000). Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. *Plant Cell* 12, 1551–1568. doi: 10.1105/tpc.12.9.1551
- Consortium, G. (2016). 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* 166, 481–491. doi: 10.1016/j.cell.2016.05.063
- Cornelius, G., Schroeder-Lorenz, A., and Rensing, L. (1985). Circadian-clock control of protein synthesis and degradation in *Gonyaulax polyedra*. *Planta* 166, 365–370. doi: 10.1007/BF00401174
- Crow, J. F. (1948). Alternative hypothesis of hybrid vigor. *Genetics* 33, 477–487. doi: 10.1093/genetics/33.5.477
- Dahal, D., Mooney, B. P., and Newton, K. J. (2012). Specific changes in total and mitochondrial proteomes are associated with higher levels of heterosis in maize hybrids. *Plant J.* 72, 70–83. doi: 10.1111/j.1365-3113.2012.05056.x
- Dandage, R., Berger, C. M., Gagnon-Arsenault, I., Moon, K. M., Stacey, R. G., Foster, L. J., et al. (2021). Frequent assembly of chimeric complexes in the protein interaction network of an interspecies yeast hybrid. *Mol. Biol. Evol.* 38, 1384–1401. doi: 10.1093/molbev/msaa298
- Darwin, C. R. (1876). *The Effects of Cross- and Self-fertilization in the Vegetable Kingdom* (London: John Murray).
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635
- Drew, K., Lee, C., Cox, R. M., Dang, V., Devitt, C. C., McWhite, C. D., et al. (2020). A systematic, label-free method for identifying RNA-associated proteins in vivo provides insights into vertebrate ciliary beating machinery. *Dev. Biol.* 467, 108–117. doi: 10.1016/j.ydbio.2020.08.008
- East, E. M. (1936). Heterosis. *Genetics* 21, 375–397. doi: 10.1093/genetics/21.4.375
- Fernandez-Escamilla, A. M., Rousseau, F., Schymkowitz, J., and Serrano, L. (2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* 22, 1302–1306. doi: 10.1038/nbt1012
- Ginn, B. R. (2010). The implications of gene heterozygosity for protein folding and protein turnover. *J. Theor. Biol.* 265, 554–564. doi: 10.1016/j.jtbi.2010.05.023
- Ginn, B. R. (2017). The thermodynamics of protein aggregation reactions may underpin the enhanced metabolic efficiency associated with heterosis, some balancing selection, and the evolution of ploidy levels. *Prog. Biophys. Mol. Biol.* 126, 1–21. doi: 10.1016/j.pbiomolbio.2017.01.005
- Goff, S. A. (2011). A unifying theory for general multigenic heterosis: energy efficiency, protein metabolism, and implications for molecular breeding. *New Phytol.* 189, 923–937. doi: 10.1111/j.1469-8137.2010.03574.x
- Groszmann, M., Gonzalez-Bayon, R., Greaves, I. K., Wang, L., Huen, A. K., Peacock, W. J., et al. (2014). Intraspecific *Arabidopsis* hybrids show different patterns of heterosis despite the close relatedness of the parental genomes. *Plant Physiol.* 166, 265–280. doi: 10.1104/pp.114.243998
- Groszmann, M., Gonzalez-Bayon, R., Lyons, R. L., Greaves, I. K., Kazan, K., Peacock, W. J., et al. (2015). Hormone-regulated defense and stress response networks contribute to heterosis in *Arabidopsis* F1 hybrids. *Proc. Natl. Acad. Sci. U.S.A.* 112, E6397–E6406. doi: 10.1073/pnas.1519926112
- Gullner, G., Komives, T., Kiraly, L., and Schroder, P. (2018). Glutathione S-transferase enzymes in plant-pathogen interactions. *Front. Plant Sci.* 9, 1836. doi: 10.3389/fpls.2018.01836
- Heinemann, B., Kunzler, P., Eubel, H., Braun, H. P., and Hildebrandt, T. M. (2021). Estimating the number of protein molecules in a plant cell: protein and amino acid homeostasis during drought. *Plant Physiol.* 185, 385–404. doi: 10.1093/plphys/kiaa050
- Heintzen, C., Nater, M., Apel, K., and Staiger, D. (1997). AtGRP7, a nuclear RNA-binding protein as a component of a circadian-regulated negative feedback loop in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 94, 8515–8520. doi: 10.1073/pnas.94.16.8515
- Herbst, R. H., Bar-Zvi, D., Reikhav, S., Soifer, I., Breker, M., Jona, G., et al. (2017). Heterosis as a consequence of regulatory incompatibility. *BMC Biol.* 15, 38. doi: 10.1186/s12915-017-0373-7
- Hsu, S. F., Lai, H. C., and Jinn, T. L. (2010). Cytosol-localized heat shock factor-binding protein, ATHSBP, functions as a negative regulator of heat shock response by translocation to the nucleus and is required for seed development in *Arabidopsis*. *Plant Physiol.* 153, 773–784. doi: 10.1104/pp.109.151225
- Jiang, X., Song, Q., Ye, W., and Chen, Z. J. (2021). Concerted genomic and epigenomic changes accompany stabilization of *Arabidopsis* allopolyploids. *Nat. Ecol. Evol.* 5, 1382–1393. doi: 10.1038/s41559-021-01523-y
- Jones, D. F. (1917). Dominance of linked factors as a means of accounting for heterosis. *Genetics* 2, 466–479. doi: 10.1093/genetics/2.5.466
- Kieber, J. J., and Schaller, G. E. (2014). Cytokinins. *Arabidopsis Book* 12, e0168. doi: 10.1199/tab.0168
- Kinsella, R. J., Kahari, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., et al. (2011). Ensembl BioMart: a hub for data retrieval across taxonomic space. *Database (Oxford)* 2011, bar030. doi: 10.1093/database/bar030
- Ko, D. K., Rohozinski, D., Song, Q., Taylor, S. H., Juenger, T. E., Harmon, F. G., et al. (2016). Temporal shift of circadian-mediated gene expression and carbon fixation contributes to biomass heterosis in maize hybrids. *PLoS Genet.* 12, e1006197. doi: 10.1371/journal.pgen.1006197
- Koh, J., Chen, S., Zhu, N., Yu, F., Soltis, P. S., and Soltis, D. E. (2012). Comparative proteomics of the recently and recurrently formed natural allopolyploid *Tragopogon mirus* (Asteraceae) and its parents. *New Phytol.* 196, 292–305. doi: 10.1111/j.1469-8137.2012.04251.x
- Kristensen, T. N., Dahlgaard, J., and Loeschke, V. (2002). Inbreeding affects Hsp70 expression in two species of *Drosophila* even at benign temperatures. *Evol. Ecol. Res.* 4, 1209–1216.
- Lee, H. S., and Chen, Z. J. (2001). Protein-coding genes are epigenetically regulated in *Arabidopsis* polyploids. *Proc. Natl. Acad. Sci. U.S.A.* 98, 6753–6758. doi: 10.1073/pnas.121064698
- Li, Z., Zhu, A., Song, Q., Chen, H. Y., Harmon, F. G., and Chen, Z. J. (2020). Temporal regulation of the metabolome and proteome in photosynthetic and photorespiratory pathways contributes to maize heterosis. *Plant Cell* 32, 3706–3722. doi: 10.1105/tpc.20.00320
- Lind-Halden, C., Halden, C., and Sall, T. (2002). Genetic variation in *Arabidopsis suecica* and its parental species *A. arenosa* and *A. thaliana*. *Hereditas* 136, 45–50. doi: 10.1034/j.1601-5223.2002.1360107.x
- McWhite, C. D., Papoulas, O., Drew, K., Cox, R. M., June, V., Dong, O. X., et al. (2020). A pan-plant protein complex map reveals deep conservation and novel assemblies. *Cell* 181, 460–474 e414. doi: 10.1016/j.cell.2020.02.049
- Meyer, K., Koster, T., Nolte, C., Weinholdt, C., Lewinski, M., Grosse, I., et al. (2017). Adaptation of iCLIP to plants determines the binding landscape of the clock-

- regulated RNA-binding protein AtGRP7. *Genome Biol.* 18, 204. doi: 10.1186/s13059-017-1332-x
- Miller, M., Song, Q., Shi, X., Juenger, T. E., and Chen, Z. J. (2015). Natural variation in timing of stress-responsive gene expression predicts heterosis in intraspecific hybrids of *Arabidopsis*. *Nat. Commun.* 6, 7453. doi: 10.1038/ncomms8453
- Miller, M., Zhang, C., and Chen, Z. J. (2012). Ploidy and hybridity effects on growth vigor and gene expression in *Arabidopsis thaliana* hybrids and their parents. *G3 (Bethesda)* 2, 505–513. doi: 10.1534/g3.112.002162
- Missra, A., Ernest, B., Lohoff, T., Jia, Q., Satterlee, J., Ke, K., et al. (2015). The circadian clock modulates global daily cycles of mRNA ribosome loading. *Plant Cell* 27, 2582–2599. doi: 10.1105/tpc.15.00546
- Nagano, T., Mitchell, J. A., Sanz, L. A., Pauler, F. M., Ferguson-Smith, A. C., Feil, R., et al. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 322, 1717–1720. doi: 10.1126/science.1163802
- Ng, D. W., Miller, M., Yu, H. H., Huang, T. Y., Kim, E. D., Lu, J., et al. (2014). A role for CHH methylation in the parent-of-origin effect on altered circadian rhythms and biomass heterosis in *Arabidopsis* intraspecific hybrids. *Plant Cell* 26, 2430–2440. doi: 10.1105/tpc.113.115980
- Ng, D. W., Zhang, C., Miller, M., Shen, Z., Briggs, S. P., and Chen, Z. J. (2012). Proteomic divergence in *Arabidopsis* autopolyploids and allopolyploids and their progenitors. *Heredity* 108, 419–430. doi: 10.1038/hdy.2011.92
- Ni, Z., Kim, E. D., Ha, M., Lackey, E., Liu, J., Zhang, Y., et al. (2009). Altered circadian rhythms regulate growth vigor in hybrids and allopolyploids. *Nature* 457, 327–331. doi: 10.1038/nature07523
- Novikova, P. Y., Tsuchimatsu, T., Simon, S., Nizhynska, V., Voronin, V., Burns, R., et al. (2017). Genome sequencing reveals the origin of the allotetraploid *Arabidopsis suecica*. *Mol. Biol. Evol.* 34, 957–968. doi: 10.1093/molbev/msw299
- O'Connell, J. D., Tschansky, M., Royal, A., Boutz, D. R., Ellington, A. D., and Marcotte, E. M. (2014). A proteomic survey of widespread protein aggregation in yeast. *Mol. Biosyst.* 10, 851–861. doi: 10.1039/c3mb70508k
- O'Kane, S., Schaal, B., and Al-Shehbaz, I. (1995). The origins of *Arabidopsis suecica* (Brassicaceae), as indicated by nuclear rDNA sequences, and implications for rDNA evolution. *Systematic Bot.* 21, 559–566. doi: 10.2307/2419615
- Pascovici, D., Handler, D. C., Wu, J. X., and Haynes, P. A. (2016). Multiple testing corrections in quantitative proteomics: A useful but blunt tool. *Proteomics* 16, 2448–2453. doi: 10.1002/pmic.201600044
- Pedersen, K. S., Kristensen, T. N., and Loeschcke, V. (2005). Effects of inbreeding and rate of inbreeding in *Drosophila melanogaster*-Hsp70 expression and fitness. *J. Evol. Biol.* 18, 756–762. doi: 10.1111/j.1420-9101.2005.00884.x
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinf.* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Saeki, N., Kawanabe, T., Ying, H., Shimizu, M., Kojima, M., Abe, H., et al. (2016). Molecular and cellular characteristics of hybrid vigor in a commercial hybrid of Chinese cabbage. *BMC Plant Biol.* 16, 45. doi: 10.1186/s12870-016-0734-3
- Sall, T., Jakobsson, M., Lind-Hallden, C., and Hallden, C. (2003). Chloroplast DNA indicates a single origin of the allotetraploid *Arabidopsis suecica*. *J. Evol. Biol.* 16, 1019–1029. doi: 10.1046/j.1420-9101.2003.00554.x
- Schnell, F. W., and Cockerham, C. C. (1992). Multiplicative vs. arbitrary gene action in heterosis. *Genetics* 131, 461–469. doi: 10.1093/genetics/131.2.461
- Shen, G., Hu, W., Zhang, B., and Xing, Y. (2015). The regulatory network mediated by circadian clock genes is related to heterosis in rice. *J. Integr. Plant Biol.* 57, 300–312. doi: 10.1111/jipb.12240
- Shen, H., He, H., Li, J., Chen, W., Wang, X., Guo, L., et al. (2012). Genome-wide analysis of DNA methylation and gene expression changes in two *Arabidopsis* ecotypes and their reciprocal hybrids. *Plant Cell* 24, 875–892. doi: 10.1105/tpc.111.094870
- Shi, X., Zhang, C., Ko, D. K., and Chen, Z. J. (2015). Genome-wide dosage-dependent and -independent regulation contributes to gene expression and evolutionary novelty in plant polyploids. *Mol. Biol. Evol.* 32, 2351–2366. doi: 10.1093/molbev/msv116
- Shull, G. H. (1908). The composition of a field of maize. *Amer. Breeders Assoc. Rep.* 4, 296–301. doi: 10.1093/jhered/os-4.1.296
- Smalle, J., Kurepa, J., Yang, P., Babiychuk, E., Kushnir, S., Durski, A., et al. (2002). Cytokinin growth responses in *Arabidopsis* involve the 26S proteasome subunit RPN12. *Plant Cell* 14, 17–32. doi: 10.1105/tpc.010381
- Song, Q., Ando, A., Xu, D., Fang, L., Zhang, T., Huq, E., et al. (2018). Diurnal down-regulation of ethylene biosynthesis mediates biomass heterosis. *Proc. Natl. Acad. Sci. U.S.A.* 115, 5606–5611. doi: 10.1073/pnas.1722068115
- Tate, J. A., Ni, Z., Scheen, A. C., Koh, J., Gilbert, C. A., Lefkowitz, D., et al. (2006). Evolution and expression of homeologous loci in *Tragopogon miscellus* (Asteraceae), a recent and reciprocally formed allopolyploid. *Genetics* 173, 1599–1611. doi: 10.1534/genetics.106.057646
- UniProt, C. (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49, D480–D489. doi: 10.1093/nar/gkaa1100
- Wang, J., Tian, L., Lee, H. S., and Chen, Z. J. (2006a). Nonadditive regulation of *FRI* and *FLC* loci mediates flowering-time variation in *Arabidopsis* allopolyploids. *Genetics* 173, 965–974. doi: 10.1534/genetics.106.056580
- Wang, J., Tian, L., Lee, H. S., Wei, N. E., Jiang, H., Watson, B., et al. (2006b). Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics* 172, 507–517. doi: 10.1534/genetics.105.047894
- Wittstock, U., Kliebenstein, D. J., Lambrix, V., Reichelt, M., and Gershenzon, J. (2003). Glucosinolate hydrolysis and its impact on generalist and specialist insect herbivores. *Recent Adv. Phytochem.* 37, 101–125. doi: 10.1016/S0079-9920(03)80020-5
- Yang, L., Liu, P., Wang, X., Jia, A., Ren, D., Tang, Y., et al. (2021). A central circadian oscillator confers defense heterosis in hybrids without growth vigor costs. *Nature Commun.* 12, 2317. doi: 10.1038/s41467-021-22268-z
- Yang, X., Yu, H., Sun, W., Ding, L., Li, J., Cheema, J., et al. (2021). Wheat in vivo RNA structure landscape reveals a prevalent role of RNA structure in modulating translational subgenome expression asymmetry. *Genome Biol.* 22, 326. doi: 10.1186/s13059-021-02549-y
- Yu, S. B., Li, J. X., Xu, C. G., Tan, Y. F., Gao, Y. J., Li, X. H., et al. (1997). Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proc. Natl. Acad. Sci. U.S.A.* 94, 9226–9231. doi: 10.1073/pnas.94.17.9226
- Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., et al. (2015). Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* 33, 531–537. doi: 10.1038/nbt.3207