













# Variation in cytonuclear expression accommodation among allopolyploid plants

Corrinne E. Grover <sup>1,\*</sup>, Evan S. Forsythe <sup>2</sup>, Joel Sharbrough <sup>3</sup>, Emma R. Miller <sup>1</sup>, Justin L. Conover <sup>1</sup>, Rachael A. DeTar <sup>2</sup>, Carolina Chavarro <sup>4</sup>, Mark A. Arick II <sup>5</sup>, Daniel G. Peterson <sup>5</sup>, Soraya C.M. Leal-Bertioli <sup>4,6</sup>, Daniel B. Sloan <sup>2</sup>, Jonathan F. Wendel <sup>1</sup>

<sup>1</sup>Ecology, Evolution, and Organismal Biology Department, Iowa State University, Ames, IA 50010, USA,

<sup>2</sup>Department of Biology, Colorado State University, Fort Collins, CO 80523, USA,

<sup>3</sup>Biology Department, New Mexico Institute of Mining and Technology, Socorro, NM 87801, USA,

<sup>4</sup>Institute of Plant Breeding, Genetics and Genomics, University of Georgia, Athens, GA 30602, USA,

<sup>5</sup>Institute for Genomics, Biocomputing & Biotechnology, Mississippi State University, Mississippi State, MS 39762, USA,

<sup>6</sup>Department of Plant Pathology, University of Georgia, Athens, GA 30602, USA

\*Corresponding author: Ecology, Evolution, and Organismal Biology Department, Iowa State University, Ames, IA 50010, USA. Email: [corrinne@iastate.edu](mailto:corrinne@iastate.edu)

## Abstract

Cytonuclear coevolution is a common feature among plants, which coordinates gene expression and protein products between the nucleus and organelles. Consequently, lineage-specific differences may result in incompatibilities between the nucleus and cytoplasm in hybrid taxa. Allopolyploidy is also a common phenomenon in plant evolution. The hybrid nature of allopolyploids may result in cytonuclear incompatibilities, but the massive nuclear redundancy created during polyploidy affords additional avenues for resolving cytonuclear conflict (i.e. cytonuclear accommodation). Here we evaluate expression changes in organelle-targeted nuclear genes for 6 allopolyploid lineages that represent 4 genera (i.e. *Arabidopsis*, *Arachis*, *Chenopodium*, and *Gossypium*) and encompass a range in polyploid ages. Because incompatibilities between the nucleus and cytoplasm could potentially result in biases toward the maternal homoeolog and/or maternal expression level, we evaluate patterns of homoeolog usage, expression bias, and expression-level dominance in cytonuclear genes relative to the background of noncytonuclear expression changes and to the diploid parents. Although we find subsets of cytonuclear genes in most lineages that match our expectations of maternal preference, these observations are not consistent among either allopolyploids or categories of organelle-targeted genes. Our results indicate that cytonuclear expression evolution may be subtle and variable among genera and genes, likely reflecting a diversity of mechanisms to resolve nuclear-cytoplasmic incompatibilities in allopolyploid species.

**Keywords:** polyploidy; cytonuclear incompatibility; expression bias; *Gossypium*; *Arachis*; *Chenopodium*; *Arabidopsis suecica*

## Introduction

Intergenomic coevolution between the nucleus and organelle(s) is a common feature among eukaryotes. Gene loss and transfers to the nucleus have greatly reduced the coding regions of modern mitochondrial and plastid genomes to a limited number of essential genes (Greiner and Bock 2013; Budar and Mireau 2018; Giannakis et al. 2021). Consequently, these organelles must coordinate transcripts and protein products from 2 or more different genomic compartments to carry out essential cellular functions. Over time, this functional interdependence results in coadaptation between the nucleus and each organelle; however, differences in mode of inheritance (i.e. biparental for the nucleus and cytoplasmic for the organelles) can lead to incompatibilities between nuclear and organellar alleles, particularly in hybrid lineages (Camus et al. 2022). These cytonuclear incompatibilities are widespread among species and can have dramatic consequences for fitness (Fishman and Willis 2006; Hill 2017; Fishman and Sweigart 2018; Postel and Touzet 2020), even leading to hybrid breakdown in some cases (Burke and Arnold 2001; Greiner et al.

2011; Burton and Barreto 2012; Burton et al. 2013; Budar and Mireau 2018).

Cytonuclear incompatibilities arising when evolutionarily distinct lineages merge to form allopolyploids may experience additional complex fates compared to incompatibilities in homoploid lineages (Sharbrough et al. 2017). The combined effects of genome merger and doubling have generally been associated with a diverse array of genomic and transcriptional changes, including nonrandom gene loss, intergenomic gene conversion, and epigenetic or regulatory changes leading to (sometimes biased) alterations in gene expression (Chen 2007; Doyle et al. 2008; Freeling 2009; Gaeta and Pires 2010; Jackson and Chen 2010; Salmon et al. 2010; Grover et al. 2012; Madlung and Wendel 2013; Yoo et al. 2014; Song and Chen 2015; Bao et al. 2019; Gallagher et al. 2020). While often evaluated on an individual gene basis, many genes are sensitive to the abundance of interacting partners, particularly those involved in multisubunit complexes (Birchler and Veitia 2010, 2014, 2021). In allopolyploid lineages, coordination of gene products becomes more complicated when interactions

Received: March 10, 2022. Accepted: July 28, 2022

© The Author(s) 2022. Published by Oxford University Press on behalf of Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

between previously isolated genomes occur and redundancy affords the possibility of gene loss or divergence (Adams and Wendel 2005; Conant and Wolfe 2008; Buggs et al. 2011; Conant et al. 2014; Gout and Lynch 2015; Panchy et al. 2016; Cheng et al. 2018; Nieto Feliner et al. 2020).

While cytonuclear incompatibilities arising in homoploid hybrid species and their roles in homoploid hybrid speciation have been described for many species (Levin 2003; Greiner et al. 2011; Burton and Barreto 2012; Burton et al. 2013; Sloan et al. 2017), the problem of maintaining coordinated expression after genome merger coupled with whole-genome duplication has only recently been considered and may be particularly acute for nuclear-encoded organelle-targeted proteins whose organelle-encoded interacting partners derive from only one of the 2 parents (Sharbrough et al. 2017). In addition to issues surrounding parental divergence and potential copy number variability in some organelle-interacting genes, allopolyploid species both face additional challenges relating to their massive duplication, including nucleotypic effects (Doyle and Coate 2019), and harbor additional mechanisms for resolving conflict, such as homoeologous exchange (Gaeta and Pires 2010; Bird et al. 2018; Mason and Wendel 2020). Consequently, a number of co-evolutionary processes might operate to balance the interaction between the nucleus and organelles, including copy number changes in organelle-interacting nuclear genes, increased organellar biogenesis, upregulation of maternal and/or organellar genes with concomitant paternal downregulation, selection for gene conversion or other mutations favoring maternal-like sequences, and pseudogenization of incompatible paternal copies (Sharbrough et al. 2017; Doyle and Coate 2019).

Recent research has begun to shed light on the extent and consequences of cytonuclear incompatibility in polyploid species. One of the first examples came from the genus *Gossypium*, in which the Rubisco complex exhibits maternally biased homoeolog retention, expression levels, and asymmetric gene conversion (Gong et al. 2012), and these observations were extended for Rubisco in phylogenetically disparate allopolyploids including *Arabidopsis*, *Arachis*, *Brassica*, and *Nicotiana* (Gong et al. 2014). Similar results were seen for the organelle-interacting gene *MS1* in allohexaploid wheat (ABD genomes in a B cytoplasm), where only B-homoeologs exhibited expression, and homoeologs from the nonmatching (AD) genomes were epigenetically silenced (Wang, Li, et al. 2017). The recently formed allotetraploid *Tragopogon miscellus* also exhibited maternal bias for cytonuclear related genes, but only for a subset of the naturally occurring *T. miscellus* individuals surveyed and none of the synthetic individuals (Sehrish et al. 2015; Shan et al. 2020). Similar observations were made for synthetic allopolyploids from *Cucumis* (Zhai et al. 2019), rice (Wang, Dong, et al. 2017), and in both the recent natural and newly synthesized forms of allopolyploid *Brassica* (Ferreira de Carvalho et al. 2019), suggesting that cytonuclear coordination may not occur immediately in nascent polyploid species. Notably, many of these studies considered only one or few genes, or were limited to patterns found within a single genus.

Here we examine the evolutionary consequences of genome merger and doubling on the expression of nuclear-encoded genes whose products are targeted to the mitochondria or plastids and interact with mitochondrial and/or plastid gene products (i.e. cytonuclear genes). Using 6 allopolyploid lineages representing 5 independent polyploid events in 4 genera (Fig. 1), which encompass a range of polyploid ages and diploid divergence times, we

quantify patterns of homoeolog usage in cytonuclear genes and patterns of total expression. We look for evidence of cytonuclear accommodation by testing the hypotheses that cytonuclear genes of allopolyploid taxa exhibit (1) maternally biased homoeolog expression and/or (2) maternal expression-level dominance (ELD) (i.e. expression patterns that more closely resemble maternal diploids than paternal diploids), reflecting a response to the historical coevolution between the maternal subgenome and the maternally inherited organelles.

## Methods

### Plant materials and sequencing

Five plants were grown for each diploid and polyploid representative from 4 genera: *Arabidopsis*, *Arachis*, *Chenopodium*, and *Gossypium* (Supplementary Table 1). Samples, here consisting of fully expanded leaves, were collected from mature plants at a uniform time of day (midday, here 11 AM–1 PM) and flash frozen in liquid nitrogen prior to RNA extraction. Growth conditions for each genus are listed below.

#### *Arabidopsis*

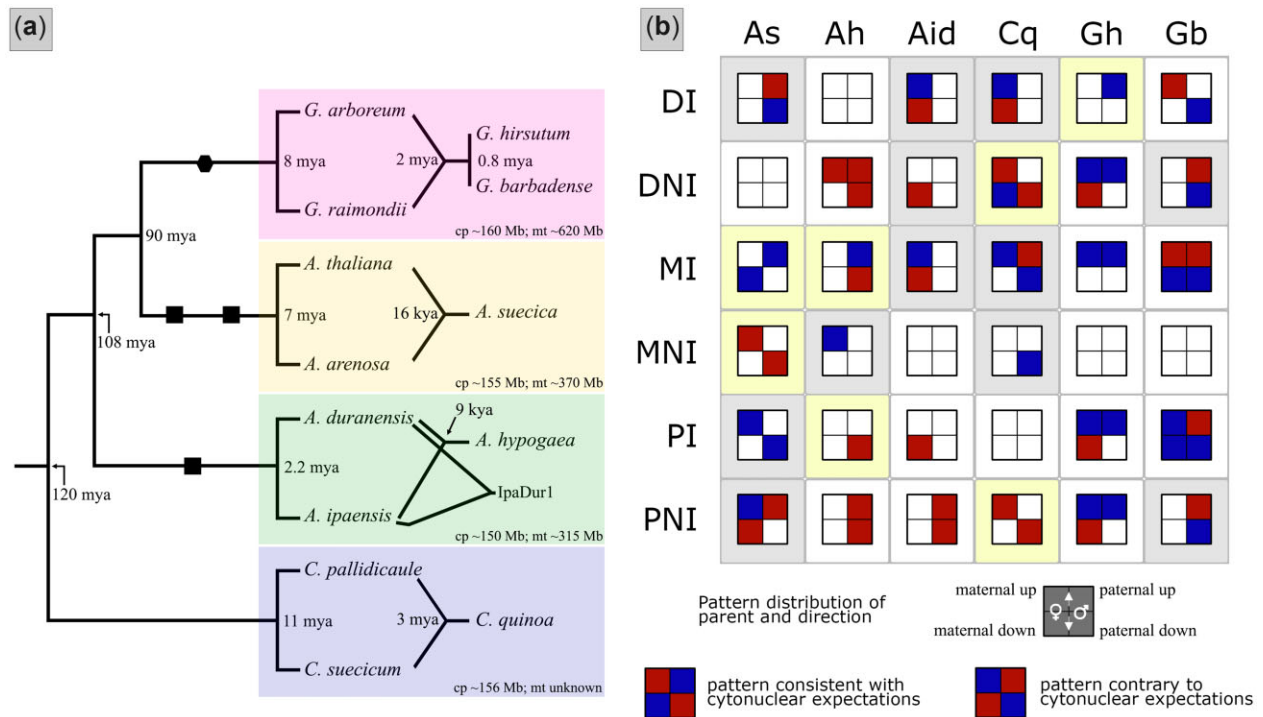
Allopolyploid *Arabidopsis suecica* (*Arabidopsis thaliana* x *Arabidopsis arenosa*) accession CS22505 seeds were acquired from Andreas Madlung (University of Puget Sound, Washington USA). These were grown in a common incubator with representatives of the parental species, *A. arenosa* (paternal, accession CS3901xKB3) and *A. thaliana* Landsberg *erecta* (maternal) whose seeds were provided by Roswitha Shmickl (Charles University, Prague) and Andreas Madlung, respectively. Seeds were surface sterilized using 70% v/v ethanol and placed on Morishige and Skoog (MS) plates for vernalization and germination. After the vernalization period (i.e. 2 weeks at 4°C), plates were moved to their growing conditions (20°C, 16/8 h light/dark). Once germinated, seeds were moved to 6-in diameter pots with potting soil (Sungro SUN52128CFLP). After several weeks of growth, plants were winterized (8°C, 10/14 h light/dark) to induce flowering (a marker of maturity). Once plants were mature, leaves were harvested from each plant at a uniform time of day (midday) and flash frozen for RNA extraction.

#### *Arachis*

*Arachis* was represented by 2 allopolyploid genotypes, i.e. *Arachis hypogaea* cv. Tifrunner (Holbrook and Culbreath 2007) and the synthetic (*Arachis ipaensis* x *Arachis duranensis*)<sup>4x</sup> known as IpaDur1 (Fávero et al. 2006; Leal-Bertioli et al. 2018; hereafter *Arachis* IpaDur1), as well as their 2 model diploid progenitors, *A. duranensis* (accession V14167) and *Arachis ipaensis* (accession K30076). Notably, these 2 allopolyploid species have opposite parentage; *A. duranensis* is maternal for *A. hypogaea* but paternal for *Arachis* IpaDur1. All species were grown in an environmentally controlled greenhouse at the University of Georgia. The first expanded leaves were collected from 8-week-old plants; these were flash frozen in liquid nitrogen and shipped on dry ice to Iowa State University for RNA extraction.

#### *Chenopodium*

The allopolyploid species *Chenopodium quinoa* accession QQ74 was grown to maturity along with the model progenitor species *Chenopodium pallidicaule* (maternal; PI 478407) and *Chenopodium suecicum* (paternal) by David Brenner in the United States



**Fig. 1.** Summary relationships among polyploids used and patterns of differential gene expression in cytonuclear categories for each polyploid species relative to each model diploid progenitor, partitioned as homoeologs. a) Cladogram depicting the relationships among the genera used, the divergence times for each set of model polyploid progenitors, and the age of the polyploidy event for each genus. In the case of *Gossypium*, 2 natural polyploids were sampled, so their age of divergence is also included. For *Arachis*, one natural (*A. hypogaea*, formed 9 KYA) and one synthetic polyploid (*IpaDur1*) was used. Ages are given in MYA (millions of years ago) and KYA (thousands of years ago). Approximate organellar sizes are given for each genus, although the mitochondrial size of *Arachis* is unavailable. Lineage-specific polyploidy events (Wang et al. 2019) for each genus are represented by squares (genome doubling) or a hexagon for *Gossypium*, which experienced a 5- to 6-fold ploidy increase (Paterson et al. 2012). b) This pictogram displays the statistically significant (Fisher's exact  $P < 0.05$ ) overrepresentation (red) or underrepresentation (blue) of the number of up- or downregulated genes for each category, relative to noncytonuclear genes. Each species/category is represented by a 4-square grid, where the rows specify regulation (up or down) and columns specify the homoeolog comparison (i.e. maternal homoeolog vs maternal progenitor and paternal homoeolog vs paternal progenitor, respectively). In each quadrant, red indicates that there were more genes statistically significant in that parent-category combination than was expected based on the NOT distribution, whereas blue indicates there were fewer statistically significant genes in that parent-category combination. Example color patterns consistent with and contrary to cytonuclear expectations are shown on the bottom. Species-category combinations highlighted in yellow are consistent with the hypothesis that cytonuclear accommodation in polyploid species favors expression from the "more compatible" maternal genome (via upregulation) and/or diminishes expression from the potentially "less favorable" paternal genome (via downregulation), whereas species-category highlighted in gray specifically contradict cytonuclear expectations. Species include *Arabidopsis suecica* (As), *Arachis hypogaea* (Ah), *Arachis IpaDur1* (Aid), *Chenopodium quinoa* (Cq), *Gossypium hirsutum* (Gh), and *Gossypium barbadense* (Gb). Categories include dual-targeted interacting (DI), dual-targeted noninteracting (DNI), mitochondria-targeted interacting (MI), mitochondria-targeted noninteracting (MNI), plastid-targeted interacting (PI), plastid-targeted noninteracting (PNI). All comparisons were performed relative to NOT genes.

Department of Agriculture (USDA, Ames, Iowa, USA) greenhouse at Iowa State University and provided as living material (3–4-month-old plants). Samples were harvested directly from the greenhouse at a uniform time of day and flash frozen in liquid nitrogen for RNA extraction.

### Gossypium

*Gossypium* was represented by 2 allopolyploid species, i.e. *Gossypium hirsutum* cultivar TM1 and *Gossypium barbadense* accession GB379, and their 2 model diploid progenitors, *Gossypium arboreum* (maternal) and *Gossypium raimondii* (paternal). Samples were grown from seed in a common environment in the Pohl Conservatory at Iowa State University. Seeds were planted in 2 gallon pots with a custom potting mixture of 4:2:2:1 Sungro soil: perlite: bark: chicken grit. *Gossypium* was grown to maturity (minimum of 6 months) under typical greenhouse conditions,

collected at a uniform time of day, and flash frozen in liquid nitrogen for RNA extraction.

### All plants

A minimum of 5 replicates (leaf tissue) were collected for each species. RNA was extracted from the *Arabidopsis*, *Arachis*, and *Chenopodium* samples using the Direct-zol RNA kit (Zymo Research), including 600  $\mu$ l of Trizol. For *Arachis*, an additional grind step in 600  $\mu$ l of Trizol using  $\frac{1}{8}$  in. diameter steel beads (1–2 min of vortexing) immediately followed the initial grind in liquid nitrogen, and 400  $\mu$ l of additional Trizol was added for extraction. All other steps follow the manufacturer protocol. *Gossypium* samples were extracted with the Spectrum Total Plant RNA kit (Sigma) following the manufacturer protocol. In total, 17 *Arabidopsis*, 20 *Arachis*, 15 *Chenopodium*, and 20 *Gossypium* samples were extracted for RNAseq (Supplementary Table 1). RNA was quantified using the Agilent 2100 BioAnalyzer and sent to the

Yale Center for Genome Analysis (YCGA) for library construction and sequencing. Illumina libraries were constructed using the TruSeq Stranded Total RNA kit with Ribo-Zero Plant and sequenced on a NovaSeq 6000 S4 flow cell. A minimum of 20 million read pairs ( $2 \times 150$  bp) was generated for each sample. Raw sequencing reads are available through the Short Read Archive under PRJNA726938.

## Reference preparation and RNA-seq processing

Reference sequences for each genus were prepared by concatenating primary transcripts for each polyploid species with transcripts for each organelle (Supplementary Table 2). Primary transcripts were derived from recent genome sequences published for *A. suецica* (Novikova et al. 2017), *A. hypogaea* (Bertioli et al. 2019), *C. quinoa* (Jarvis et al. 2017), and *G. hirsutum* (Chen et al. 2020). RepeatMasker (Smit et al. 2015) was used to mask each set of nuclear primary transcripts with both the organellar genomes and transcriptomes (Supplementary Table 2, and see below) for each species, and any transcript with fewer than 75 nucleotides of nonorganelle-derived sequence was discarded. Mitochondrial and plastid transcripts for each genus were derived from publicly available organelle genome annotations for a single representative species from each genus (Supplementary Table 2), with the exception of *Arachis* mitochondrial genes (see below). Each protein-coding gene set was manually curated to (1) add genes that were absent from the GenBank annotations (via BLAST identification; Camacho et al. 2009), (2) remove duplicate gene copies from the plastid inverted repeat, (3) remove nonconserved hypothetical genes, and (4) standardize gene naming conventions. Because there is no complete mitochondrial genome published for any *Arachis* species, we used available transcriptomic and genomic resources to extract protein-coding sequences for *Arachis* mitochondrial genes. Most genes were recovered by performing tBLASTN of *Arabidopsis* protein sequences against an unpublished dataset of *A. hypogaea* full-length cDNAs generated with PacBio Iso-Seq technology (NCBI Sequence Read Archive accession SRR14414925), and the remaining mitochondrial genes were extracted by searching against *A. hypogaea* genomic contigs in PeanutBase (Dash et al. 2016). Our curated mitochondrial and plastid protein-coding reference sequences for each taxon are available via <https://github.com/Wendellab/CytonuclearExpression>.

RNA-seq reads for each species were processed via Kallisto v0.46.1 (Bray et al. 2016) (i.e. *kallisto quant*) to assign orthologs and/or homoeologs to genes and quantify transcripts. Following Kallisto quantification, a principal component analysis (PCA) was generated for each genus using SNPRelate (Zheng et al. 2012) in R/4.0.2 (R Core Team 2020) to verify sample identity and generate an overview of the count data. PCA plots were visualized using ggplot2 (Wickham 2016) in R. Clustering heatmaps were generated using pheatmap (Kolde 2012). Code pertaining to this project can be accessed at <https://github.com/Wendellab/CytonuclearExpression>.

## Ortholog identification and targeting inference

We followed the methods of Sharbrough et al. (2022) to identify orthologous genes arising from allopolyploidy (i.e. “quartets” consisting of one homolog from each diploid parent and 2 homoeologs from the allopolyploid). Briefly, we used Orthofinder (v2.3.8) (Emms and Kelly 2019) to cluster protein coding genes into homologous gene families. We retained orthogroups containing 3 or more homologs, extracted coding sequences (CDS) for those proteins, and aligned each using the L-INS-i algorithm in MAFFT (v7.480) (Kato and Standley 2013). Model selection was done using jModelTest2 v2.1.10 (Darriba et al. 2012) and phylogenetic inference was performed in PhyML v3.3.20211021 (Guindon and Gascuel 2003), as

described previously (Sharbrough et al. 2022). Because these gene trees often contain multiple orthologous groups resulting from ancient duplications, we extracted subtrees containing potential quartets (i.e. subtrees with the expected number of genes from each species) using *subTreeIterator.py* (Sharbrough et al. 2022). We merged these phylogenetically based quartet predictions with independent synteny-based quartet predictions (generated via pSONIC; Conover et al. 2021) to identify high-confidence quartets. Quartets that were predicted by at least one method and were not in conflict with the second method were retained for analysis. Each quartet was analyzed for organelle targeting information using combined information from (1) CyMIRA (Forsythe et al. 2019); (2) de novo targeting software, including iPSORT v0.94 (Bannai et al. 2002), LOCALIZER v1.0.4 (Sperschneider et al. 2017), Predotar v1.03 (Small et al. 2004), and TargetP v1.1b (Emanuelsson et al. 2007); and (3) Orthofinder-based homology to the *A. thaliana* Araport 11 proteome. Full details can be found in Sharbrough et al. (2022), and relevant scripts can be found at [https://github.com/jsharbrough/CyMIRA\\_gene\\_classification](https://github.com/jsharbrough/CyMIRA_gene_classification) and <https://github.com/jsharbrough/allopolyploidCytonuclearEvolutionaryRate/blob/master/scripts/subTreeIterator.py>, as well as <https://github.com/Wendellab/CytonuclearExpression>.

## Differential gene expression

Differential gene expression analyses were conducted in R/4.0.2 (R Core Team 2020) using DESeq2 (Love et al. 2014) with the design “~species” and with the reference transcriptomes detailed above. Genes with a Benjamini–Hochberg (Benjamini and Hochberg 1995) adjusted *P*-value  $< 0.05$  ( $P_{adj}$ , as implemented in DESeq2) were considered differentially expressed. Expression PCA and heatmaps were made in R using the base R package and pheatmap v1.0.12. Differential expression (DE) was evaluated 3 ways: (1) DE between diploid progenitor and corresponding polyploid subgenome, (2) DE between each diploid progenitor and the total polyploid expression (i.e. summed homoeolog expression), and (3) DE between maternal and paternal homoeologs. Enrichment of DE genes in cytonuclear gene categories was conducted using Fisher’s exact test (*fisher.test*) relative to the nonorganelle-targeted (NOT) gene category.

We evaluated both homoeolog expression bias (HEB) and ELD using the above-generated lists of significantly differentially expressed genes (DESeq  $P_{adj} < 0.05$ ). HEB was inferred when the relative expression of the homoeologs was statistically different (based on #3 from the preceding paragraph). Genes were partitioned into ELD categories based on whether the summed homoeolog expression (i.e. polyploid expression) was statistically greater than or less than zero, one, or both parents (#3 from the preceding paragraph), and whether parental expression was statistically different. Classifications were made based on previously identified categories (Rapp et al. 2009; Yoo et al. 2013). Briefly, a gene in the polyploid was classified as “intermediate expression” if it exhibited the expression patterns  $M > P > D$  or  $M < P < D$ , where  $M$  = maternal progenitor,  $D$  = paternal progenitor, and  $P$  = polyploid. Paternal ELD was inferred when  $D = P < M$  or  $D = P > M$ , where “equals” in this case refers to expression that is not statistically different between the paternal progenitor and the polyploid. Likewise, maternal ELD was inferred when  $M = P < D$  or  $M = P > D$ . Transgressive expression was inferred when the polyploid exhibited statistically different expression (higher or lower) from both parents, regardless of whether the parents were statistically different from each other. No change in expression was inferred when there were no statistical differences in expression between the polyploid and either parent and no statistical

differences between the parents themselves (i.e. all expression values were statistically equivalent).

We employed a mixed-effects modeling approach to test whether differences in expression across homoeologs were related to cytonuclear targeting category (inferred from CyMIRA), legacy effects of diploid progenitors (estimated here as the difference in expression across diploid relatives), and the interaction between targeting category and legacy effects. DESeq2 rlog-normalized counts were used, where rlog represents a log<sub>2</sub>-based transformation that both normalizes the library size and minimizes differences between samples for genes with low counts (Love et al. 2014). Expression modeling was conducted in R/4.1.1 and considered the 2 models: (1)  $\Delta r \log \sim \text{Targeting}$ , and (2)  $\Delta r \log \sim \text{Targeting} + \Delta r \log_{\text{Diploid}} + \text{Targeting} \times \Delta r \log_{\text{Diploid}}$ , where  $\Delta r \log$  represents the difference in DESeq2-derived rlog normalized counts (maternal—paternal homoeolog),  $\Delta r \log_{\text{Diploid}}$  represents the difference in DESeq2-derived rlog normalized counts between the model diploid progenitors, Targeting represents the CyMIRA identified targeting category, and Targeting  $\times$   $\Delta r \log_{\text{Diploid}}$  represents the interaction between category and diploid expression levels. Fixed effects for each model were evaluated using emmeans v1.7.0 and the analysis of variance (ANOVA) was evaluated using car v3.0-11, with a type III computation of the sums of squares. Because model 1 is nested within model 2, we compared these 2 models for each species using lrttest from lmer v0.9-39 in R/4.1.1.

## Functional enrichment tests

CyMIRA-based results were verified for *A. suecica*, *G. hirsutum*, and *G. barbadense* using FUNC-E (<https://github.com/SystemsGenetics/FUNC-E>) in conjunction with existing functional annotations from INTERPRO (Jones et al. 2014), GO ontology (The Gene Ontology Consortium 2019), and Plant Ontology (available for *Arabidopsis* only; Avraham et al. 2008). *Arabidopsis* functional annotations were downloaded from TAIR (Cheng et al. 2017), and the *Gossypium* functional annotations were downloaded from CottonFGD (Zhu et al. 2017), both accessed in January 2022. These custom ontology lists were used to generate vocabulary terms for each FUNC-E analysis (one per species). Two sets of genes were used as queries in functional enrichment analyses, both of which are restricted to ortholog-homoeolog quartets with statistically significant DE

between homoeologs (DESeq2 P-value < 0.05) that was also greater than 4-fold. An additional criterion for the second query gene set required that the difference in fold-change (FC) between homoeologs and FC between parental orthologs also had to be greater than 4 (i.e.  $|\Delta FC| > 4$ ). In both cases, the reference (i.e. background) set was composed of quartets regardless of P-value and/or FC; these comprised 11,307 for *A. suecica*, 18,669 for *G. hirsutum*, and 18,099 for *G. barbadense*. Functional enrichment was determined in FUNC-E via a 1-sided Fisher's exact test for each comparison, and multiple tests were subjected to Benjamini-Hochberg correction; significance was determined as adjusted  $P < 0.05$ . By default, upregulated and downregulated genes were tested separately.

## Results

### Generation and categorization of reference sequences

Representative transcriptomes for each genus were downloaded along with both organellar genomes and transcriptomes (Supplementary Table 2). In the case of *Arachis*, only putative transcripts were available for the mitochondria (see Methods). Because reference genomes frequently have nuclear insertions of organellar genes that can be included in predicted transcripts, we first masked each nuclear transcriptome (primary transcripts only) with both the matching organellar genomes and transcriptomes, and we subsequently removed transcripts with fewer than 75 surviving nucleotides. Between 206 and 2,510 nuclear transcripts were filtered from each reference, leaving between 44,175 and 73,595 nonorganellar nuclear transcripts. These were combined with the curated organellar transcriptomes, consisting of 108–112 genes in total (see Methods), resulting in polyploid reference transcriptomes ranging from 44,283 to 73,707 genes (Table 1).

A curated set of high-confidence homoeologs was generated for each reference genome using a combination of phylogenetics and synteny (see Methods), which were subsequently characterized by their potential to interact with either/both organelles (Table 1). The number of homoeologous pairs in each genome ranged from 9,231 in *C. quinoa* to 20,124 in *G. hirsutum*, representing twice that number of genes (18,462 and 40,248 homoeologs, respectively). As expected, most genes (80–87%) were not

**Table 1.** Composition of the mapping reference for each genus.

	<i>Arabidopsis</i>	<i>Arachis</i>	<i>Chenopodium</i>	<i>Gossypium</i>
<b>Mitochondrial transcripts</b>	32	32	30	35
<b>Chloroplast transcripts</b>	78	76	78	77
<b>Nuclear transcripts</b>	44,625	67,150	44,770	74,902
Nuclear transcripts, excluding norgs	44,419	64,640	44,175	73,595
<b>Total transcripts</b>	44,735	67,258	44,878	75,014
Removed genes	206	2,510	595	1,307
<b>Total transcripts, excluding norgs</b>	44,529	64,748	44,283	73,707
<b>Homoeologous pairs (genome)</b>	12,254	11,671	9,231	20,124
Not targeted	9,830	10,121	7,575	17,606
Dual-targeted, interacting	45	52	62	76
Dual-targeted, noninteracting	185	746	771	1,103
Mitochondria-targeted, interacting	263	156	169	326
Mitochondria-targeted, noninteracting	467	94	84	135
Plastid-targeted, interacting	168	133	159	246
Plastid-targeted, noninteracting	1,296	369	411	632

Primary transcripts from each nuclear transcriptome were masked using the organellar transcriptomes and genomes, and nuclear transcripts matching organellar sequences (i.e. norgs) were removed. Gene quartets composed of a single gene for each diploid species and 2 paired homoeologs from the polyploid reference were identified. Each quartet was classified with respect to putative organellar targeting. “Dual-targeted” transcripts are those that have targeting information for both organelles. “Interacting” transcripts code for products that interact with organellar gene products, whereas “noninteracting” transcripts are those which function in one or both organelles but do not physically interact with an organellar gene product.

**Table 2.** The number of paternal or maternal homoeologs lost from *Arabidopsis* and *Gossypium* for each category, and the proportion that represent maternal losses.

	<i>Arabidopsis</i>			<i>Gossypium</i>		
	Paternal loss	Maternal loss	% maternal	Paternal loss	Maternal loss	% maternal
Not-organelle-targeted	673	439	39%	342	542	61%
All cytonuclear	106	81	46%	21	30	59%
Dual-targeted_interacting	2	3	60%	0	2	100%
Dual-targeted_noninteracting	2	8*	80%	9	13	59%
Mitochondria-targeted_interacting	19	10	34%	4	4	50%
Mitochondria-targeted_noninteracting	20	16	44%	1	2	67%
Plastid-targeted_interacting	12	7	37%	1	2	67%
Plastid-targeted_noninteracting	51	45	47%	6	7	54%

If broad cytonuclear incompatibilities exist, we expect that the number of maternal homoeologs lost should be fewer in cytonuclear gene categories than for the rest of the genome, represented by low numbers in the % maternal columns. Cytonuclear categories that are statistically different in distribution from NOT genes are marked with an \* in the column where loss is greater than expected by the NOT category (Fisher's exact  $P < 0.05$ ).

predicted to be targeted to either organelle, and, on average, 2–3% of genes were placed (per category) in the 6 organelle-related categories (i.e. mitochondria-/plastid-/dual-targeted, interacting/noninteracting genes; range = 0–11%; Table 1), as determined by CyMIRA (see Methods). Of those genes exhibiting signatures of organelle targeting, homoeolog pairs that function in the organelle but do not have direct interactions with organellar-encoded proteins were generally more abundant, with the exception of mitochondria-targeted interacting genes, which were 1.5–2 times more abundant in most species (except *A. thaliana*; Table 1) than the noninteracting mitochondrial genes. This observed difference in prediction for *Arabidopsis* may reflect the extensive prior knowledge and validation available for this model plant versus the *Arabidopsis*-based prediction in the other, nonmodel organisms. Nevertheless, these targeting predictions were subsequently applied to the reference transcriptome generated for each genus (Table 1 and see Methods).

We also evaluated the degree of homoeolog loss between the maternal and paternal genome for genes where orthologs were recovered from both model progenitors but only one polyploid subgenome (Table 2). If there is a general cytonuclear incompatibility between the diploid progenitors, then we would expect an excess in paternal homoeolog loss for genes involved in cytonuclear categories, i.e. dual-targeted interacting (DI), dual-targeted noninteracting (DNI), mitochondria-targeted interacting (MI), mitochondrial-targeted noninteracting (MNI), plastid-targeted interacting (PI), and plastid-targeted noninteracting (PNI). Because the *C. quinoa* genome has a large number of genes not assigned to maternal/paternal subgenome, and the *A. hypogaea* genome exhibits a high degree of homoeologous exchange (thereby reducing the number of reliable quartets; Chen et al. 2016; Bertoli et al. 2019; Zhuang et al. 2019), we restricted our analysis of putative homoeolog loss to *A. suecica* and *G. hirsutum* (Table 2). For most categories, there was no significant difference in paternal versus maternal homoeolog loss relative to background (i.e. genes whose products are not targeted to either organelle (NOT); Fisher's exact  $P > 0.05$ ). Only one cytonuclear category from the 2 genomes (i.e. DNI from *Arabidopsis*) exhibited biased homoeolog loss, and the distribution of loss was contrary to what is expected given maternal inheritance of organelles.

### Sequencing yields and general gene expression

Because the aim of this study was to characterize cytonuclear accommodation at the level of gene expression in polyploid species, total RNA was extracted for each accession and ribodepletion was used to remove ribosomal RNAs, circumventing the bias of

polyA-selection protocols that exclude some organellar transcripts (Slomovic et al. 2006, 2008; Smith 2013). As expected, transcripts from the organelles were abundant (Supplementary Table 3 and analyzed in Forsythe et al. 2022); however, sufficient nuclear transcriptome coverage was achieved, ranging from 4 to 26 M reads per sample (mean reads per sample were *Arabidopsis* = 13 M, *Arachis* = 11 M, *Chenopodium* = 7 M, *Gossypium* = 20 M). One replicate each for *A. hypogaea* and *Arachis* IpaDur1 was removed due to low overall mapping rates (i.e. <25% of reads mapped to nuclear or organellar genomes). Principal components analysis and hierarchical clustering of the gene expression data exhibit clustering of replicates for each species within a genus, with one exception. *C. suecicum* replicate #1 was placed intermediate among all *Chenopodium* species via PCA (Supplementary Fig. 1), and it clustered with *C. quinoa* via hierarchical clustering. Because this sample may represent a contaminated hybrid, it was excluded from subsequent analyses.

In general, polyploid species exhibited more upregulated genes than downregulated genes (expression relative to their diploid counterparts), both with respect to homoeolog-progenitor comparisons and total polyploid expression (Table 3). This pattern was most prominent in *Gossypium*, where all comparisons exhibited more upregulated than downregulated genes in polyploids ( $\chi^2$ ,  $P < 0.05$ ), followed by *A. suecica*, where all maternal comparisons exhibited more upregulation. Conversely, *C. quinoa* only exhibited more upregulation of the total polyploid expression (i.e. the summed expression of homoeologs), and the natural peanut polyploid, *A. hypogaea*, only exhibited more upregulation of maternal homoeologs relative to expression in the model maternal diploid progenitor, *A. duranensis* (Table 3). Interestingly, the synthetic allotetraploid, *Arachis* IpaDur1 also exhibits more upregulation of *A. duranensis* homoeologs, here functioning as the paternal diploid progenitor, with concomitant downregulation in expression of homoeologs from the maternal diploid parent, *Arachis ipaensis*, potentially indicating a general bias toward *A. duranensis* expression.

### Expression-level dominance in cytonuclear genes

ELD is a phenomenon whereby the combined expression of homoeologs in a polyploid is statistically similar to one diploid parent and statistically dissimilar from the other parent. In the context of cytonuclear compatibility, we might expect a bias toward the maternal diploid expression level (i.e. ELD) for the combined expression of both homoeologs in cytonuclear gene categories. When we consider ELD of nuclear genes within each species, irrespective of category (i.e. NOT or any cytonuclear

**Table 3.** The total number of genes passing filter (see *Methods*), and the number that are differentially expressed (parsed as up or downregulated).

		<i>Gossypium hirsutum</i>		<i>Gossypium barbadense</i>		<i>Chenopodium quinoa</i>		<i>Arabidopsis suecica</i>		<i>Arachis hypogaea</i>		<i>Arachis IpaDur1</i>	
Diploid divergence		5–10 MYA				11 MYA		6–8 MYA		2.2 MYA			
Polyploid origin		1–2 MYA				2.5–3 MYA		16 KYA		9,400 YA		Synthetic	
		Maternal	Paternal	Maternal	Paternal	Maternal	Paternal	Maternal	Paternal	Maternal	Paternal	Maternal	Paternal
Diploid-polyploid, parsed as homoeologs	Total genes	18,197	18,355	18,197	18,355	8,603	8,616	11,154	11,225	10,154	10,404	10,404	10,154
	Downregulated	2,803 (15%)	3,912 (21%)	2,617 (14%)	3,461 (19%)	2,390 (28%)	1,984 (23%)	1,358 (12%)	1,888 (17%)	328 (3%)	1,778 (17%)	<b>2,290 (22%)</b>	28 (0%)
	Upregulated	<b>3,166 (17%)</b>	<b>4,390 (24%)</b>	<b>2,908 (16%)</b>	<b>4,052 (22%)</b>	2,519 (29%)	1,965 (23%)	<b>1,478 (13%)</b>	1,987 (18%)	<b>593 (6%)</b>	1,716 (16%)	1,868 (18%)	<b>166 (2%)</b>
Diploid-polyploid, total expression in polyploids versus one or the other diploid	Total genes	18,792	18,792	18,792	18,792	8,813	8,813	11,610	11,610	10,645	10,645	10,645	10,645
	Downregulated	3,387 (18%)	4,012 (21%)	3,133 (17%)	3,529 (19%)	2,322 (26%)	2,435 (27%)	1,828 (16%)	2,040 (18%)	669 (6%)	1,722 (16%)	1,617 (15%)	90 (1%)
	Upregulated	<b>4,085 (22%)</b>	<b>4,691 (25%)</b>	<b>3,944 (21%)</b>	<b>4,351 (23%)</b>	<b>2,551 (29%)</b>	2,398 (27%)	<b>2,160 (19%)</b>	2,086 (18%)	609 (6%)	1,862 (18%)	<b>1,893 (18%)</b>	108 (1%)

Cells that are in bold are significantly different from equal (upregulation vs downregulation);  $\chi^2 < 0.05$ . Note that the different number of genes in the diploid-polyploid comparison (parsed as homoeologs) reflect differences in survivability in the DE analysis.

**Table 4.** Number of genes exhibiting ELD toward each parental expression level, parsed by cytonuclear category.

Category	<i>Arabidopsis suecica</i>		<i>Arachis hypogaea</i>		IpaDur1		<i>Chenopodium quinoa</i>		<i>Gossypium barbadense</i>		<i>Gossypium hirsutum</i>	
	Maternal ELD	Paternal ELD	Maternal ELD <sup>a</sup>	Paternal ELD	Maternal ELD	Paternal ELD <sup>a</sup>	Maternal ELD	Paternal ELD	Maternal ELD <sup>a</sup>	Paternal ELD	Maternal ELD <sup>a</sup>	Paternal ELD
Not-organelle-targeted	1,481	1,486	1,659	365	41	2,026	1,294	1,293	2,961	2,421	3,063	2,273
All cytonuclear	419	409	361 <sup>b</sup>	54	1	420	361 <sup>b</sup>	288	430	345	462	325
DI	8	8	11	1	0	11	15	14	9	12	12	15
DNI	31	30	184 <sup>b</sup>	25	0	203	184 <sup>b</sup>	126	194	156	213	156
MI	51	39	23	8	0	39	42 <sup>b</sup>	21	47	46	54	33
MNI	85	82	26	4	1	28	10	14	23	26	29	19
PI	22	37	32	2	0	38	31	29	48 <sup>b</sup>	21	46	30
PNI	222	213	85	14	0	101	79	84	109	84	108	72

Categories are dual-targeted interacting (DI), dual-targeted noninteracting (DNI), mitochondria-targeted interacting (MI), mitochondrial-targeted noninteracting (MNI), plastid-targeted interacting (PI), and plastid-targeted noninteracting (PNI).

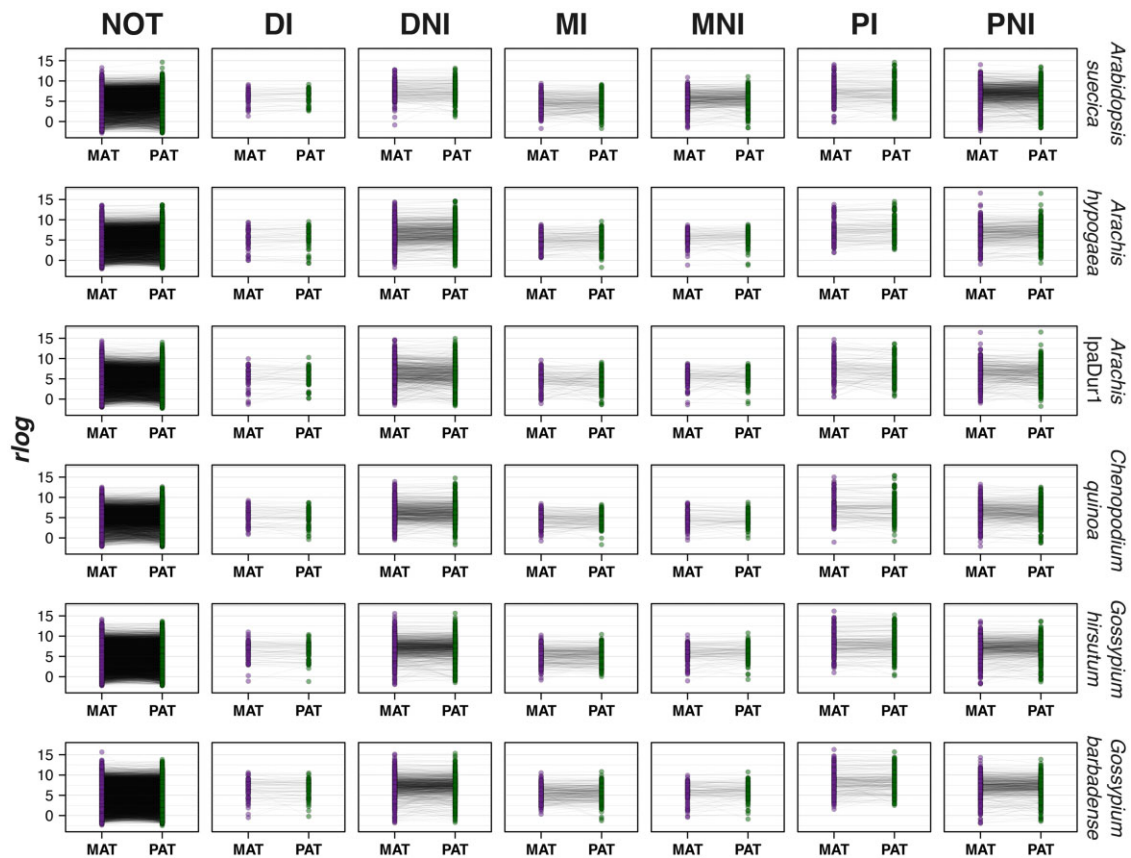
<sup>a</sup> Parental ELD is significantly different from 1:1 and biased toward the noted parent

<sup>b</sup> Cytonuclear category distribution (maternal versus paternal) is significantly different from the distribution in the NOT category and overrepresented by the noted parent

category), we see a general bias toward maternal ELD for *A. hypogaea* and both species of *Gossypium* (binomial test,  $P < 0.05$ ), but not for *A. suecica* or *C. quinoa* (binomial test,  $P > 0.05$ ; [Table 4](#) and [Supplementary Table 4](#)). These results are also reflected in the NOT category itself, where *A. hypogaea* and both *Gossypium* species exhibit bias toward maternal ELD. Interestingly, however, when we compare patterns of ELD for all organelle targeted genes versus those in the NOT category, we find that *A. hypogaea* and *C. quinoa* have significantly more genes (Fisher's exact,  $P < 0.05$ ) exhibiting ELD in maternally biased categories (i.e. categories IV and IX; [Supplementary Table 4](#)) than expected from the overall distribution of maternal and paternal ELD, whereas both species of *Gossypium* exhibited similar patterns of ELD for cytonuclear genes as NOT genes. Notably, *Arachis IpaDur1* exhibited an excess of paternal ELD, which is in contrast to the maternal ELD exhibited by *A. hypogaea* but biased toward the same diploid parent (i.e. biased toward *Arachis duranensis* in both cases). On the level of individual categories, 4 categories in 3 species exhibit an excess of ELD (Fisher's exact,  $P < 0.05$ ), all maternally biased: *A. hypogaea*, DNI; *C. quinoa*, DNI and MI; and *G. barbadense*, PI. All other

individual categories exhibited similar ELD bias as displayed by the NOT genes for that species ([Table 4](#)).

We also identified some genes in these polyploids with expression levels that fell outside the range of the 2 parental diploid models (i.e. transgressive expression), which may be associated with organelle copy number in a cytonuclear context. When considering all genes, regardless of targeting, *A. suecica* and both *Arachis* species have statistically similar numbers of genes that are transgressive downregulated (categories III, VII, and X in [Supplementary Table 4](#)) as transgressive upregulated (categories V, VI, and VIII), whereas *C. quinoa* and both species of *Gossypium* have ~20–35% more genes exhibiting transgressive upregulation (vs downregulation; [Supplementary Table 4](#)). Accounting for these global patterns, we find no species-category combinations exhibiting transgressive expression patterns in cytonuclear genes that are statistically different from NOT genes (after Benjamini-Hochberg P-value correction for multiple testing), although we note that many of these cytonuclear categories had very few genes ([Supplementary Table 4](#)) and are therefore difficult to statistically characterize.



**Fig. 2.** Mean normalized gene expression across homoeologs of 6 allotetraploids. Mean  $r$  log values (circles) from 4 to 5 biological replicates each are depicted for maternal (left, purple) and paternal (right, green) homoeologs, partitioned into 7 functional categories: NOT, dual-targeted noninteracting (DNI), mitochondria-targeted noninteracting (MNI), plastid-targeted noninteracting (PNI), dual-targeted interacting (DI), mitochondria-targeted interacting (MI), and plastid-targeted interacting (PI). Semitransparent lines connect maternal and paternal homoeologs.

### Homoeolog expression in cytonuclear genes

We evaluated homoeolog expression for each polyploid species in the context of the 6 cytonuclear categories with the biological expectation that maternal homoeologs should be preferentially upregulated relative to paternal homoeologs (Fig. 1 and Supplementary Tables 5–8). Figure 1 summarizes the results of the homoeolog comparisons for each homoeolog in  $2 \times 2$  grids for each species category, where maternal (left) and paternal (right) expression is measured relative to the model diploid progenitor and over-/underrepresentation is determined relative to the pattern observed in NOT genes (i.e. background; Fisher's exact test  $P < 0.05$ , see Methods). Because cytonuclear incompatibility predicts upregulation of the coevolved maternal cytonuclear homoeologs and downregulation of the evolutionarily more distant paternal homoeologs, we expect a combination of the following patterns (Fig. 1): (1) overrepresentation (depicted in red) for maternal homoeolog upregulation (upper left square), (2) overrepresentation (red) for paternal homoeolog downregulation (lower right square), (3) underrepresentation (depicted in blue) for maternal homoeolog downregulation (lower left square), and/or (4) underrepresentation (blue) for paternal homoeolog upregulation (upper right square). In general, fewer than half of the categories per polyploid species are consistent with cytonuclear incompatibility expectations, and, in both *Arachis IpaDur1* and *G. barbadense*, we do not observe any categories whose patterns are consistent with our biological expectations. None of the categories were consistent with cytonuclear expectations in more than 2 species, although each category was consistent in at least one.

Interestingly, the most frequently observed patterns were contrary to cytonuclear expectations (Fig. 1); that is, 12 species-category comparisons contradict cytonuclear expectations (vs 7 consistent species categories), although these contradictory patterns were also observed in no more than half of the categories per species.

We also directly compared expression between homoeologs to ascertain the extent (or lack) of maternal expression bias, both in general and with respect to cytonuclear categories (Fig. 2). HEB is distinct from ELD in that HEB reports statistically different expression levels *between* homoeologs, whereas ELD (see above) refers to instances where the *total* gene expression (of both homoeologs) is similar to only one parent. We find that most of the polyploids (except the synthetic *Arachis IpaDur1*) exhibit more genes with paternal HEB versus maternal, for all paired homoeologs regardless of category (Table 5). When these genes are partitioned into cytonuclear categories, however, we detect maternal bias for some individual categories, most notably *A. suecica* and *C. quinoa*, where 4 of the 6 cytonuclear categories have more genes with maternal bias than paternal. In most cases, this directional shift toward maternal bias is not statistically significant from the NOT distribution (Fisher's exact test,  $P > 0.05$ ) and may either represent a lack of biological relevance or a lack of statistical power due to the small numbers in many of these categories (Table 5). The only categories that did exhibit statistically significant higher numbers of maternally HEB were the PI and PNI categories from *A. suecica* and DI from *G. hirsutum*. The latter may be somewhat surprising not only because this is the sole



**Table 5.** HEBs for each polyploid, partitioned as maternal and paternal bias.

	<i>Arabidopsis suecica</i>		<i>Arachis hypogaea</i>		<i>Arachis IpaDur1</i>		<i>Chenopodium quinoa</i>		<i>Gossypium hirsutum</i>		<i>Gossypium barbadense</i>	
	Maternal bias	Paternal bias	Maternal bias	Paternal bias	Maternal bias	Paternal bias	Maternal bias	Paternal bias	Maternal bias	Paternal bias	Maternal bias	Paternal bias
<b>Total</b>	<b>1,634</b>	<b>1,887<sup>a</sup></b>	<b>757</b>	<b>836<sup>a</sup></b>	<b>2,282<sup>a</sup></b>	<b>1,678</b>	<b>1,376</b>	<b>1,613<sup>a</sup></b>	<b>3,282</b>	<b>3,690<sup>a</sup></b>	<b>2,536</b>	<b>2,734<sup>a</sup></b>
<b>NOT</b>	1,251	1,527	628	689	1,878	1,397	1,088	1,310	2,836	3,151	2,262	2,345
<b>DI</b>	5	4	3	3	11	12	10	9	22 <sup>b</sup>	13	7	8
<b>DNI</b>	33	29	64	74	198	136	125	142	184	231	127	172 <sup>b</sup>
<b>MI</b>	53	53	6	11	36	24	16	35 <sup>b</sup>	64	73	24	36
<b>MNI</b>	67	75	8	9	24	20	18	14	23	29	14	24
<b>PI</b>	28 <sup>b</sup>	17	13	18	46	26	37 <sup>b</sup>	22	41	55	27	47 <sup>b</sup>
<b>PNI</b>	197 <sup>b</sup>	182	35	32	89	63	82	81	112	138	75	102 <sup>b</sup>

Bias is considered when homoeolog expression is statistically significant (adjusted  $P < 0.05$ ), regardless of the magnitude of the change. The distribution of maternally paternally biased genes for each cytonuclear category was evaluated relative to the NOT category using a Fisher's exact test. Significant deviations ( $P < 0.05$ ) from the NOT distribution are noted by superscript letter b, and the column (maternal or paternal) designates the parental bias that is overrepresented for that category.

<sup>a</sup> Parental HEB is significantly different from 1:1 and biased toward the noted parent.

<sup>b</sup> Cytonuclear category distribution (maternal vs paternal) is significantly different from the distribution in the NOT category and overrepresented by the noted parent.

maternally biased category from either *Gossypium* species, but also because the closely related species *G. barbadense* exhibits 3 cytonuclear categories with bias in the opposite direction (more paternal HEB than is expected from the NOT distribution, i.e. DNI, PI, and PNI; Table 5).

We further evaluated the possible effects of cytonuclear category membership on homoeolog expression using linear modeling. We began with a model that asked if the difference in observed expression between maternal and paternal homoeologs was a function of where it was targeted ( $\Delta r \log \sim \text{Targeting}$ ) using the 6 aforementioned categories. For this model, we evaluated expression in each polyploid as a difference in  $r \log$  normalized counts (derived from DESeq2) between the maternal homoeolog and the paternal homoeolog (as  $\Delta r \log = r \log_{\text{Maternal}} - r \log_{\text{Paternal}}$ ). The results of this model (Fig. 3 and Supplementary Table 9) suggest that membership in a cytonuclear category (i.e. Targeting) does have an effect on the difference between homoeolog expression levels for *A. suecica*, *Arachis IpaDur1*, *C. quinoa*, *G. hirsutum*, and *G. barbadense*, but it is not significant for *A. hypogaea* (ANOVA,  $P > 0.05$ ). The number and identities of categories with fixed effects significantly different from NOT vary between polyploids (Fig. 3 and Supplementary Fig. 2 and Table 9), with the MNI category exhibiting significant fixed effects most frequently (3 of 5 significant polyploids) while MI is not significant for any polyploid. Contrasts among categories are even less suggestive of expression differences due to targeting for most species, although in *A. suecica* most categories (except DI and MI) exhibited significantly greater expression differences between homoeologs than the NOT category ( $P < 0.05$ ) and in the expected direction (i.e. expression differences between homoeologs in those cytonuclear categories are more maternally biased than NOT). In the remaining species, only PI in *C. quinoa* and DNI in *G. barbadense* were significantly different from the NOT category, the latter of which contradicted our expectations (i.e. NOT in *G. barbadense* is more maternally biased than is DNI; Fig. 3).

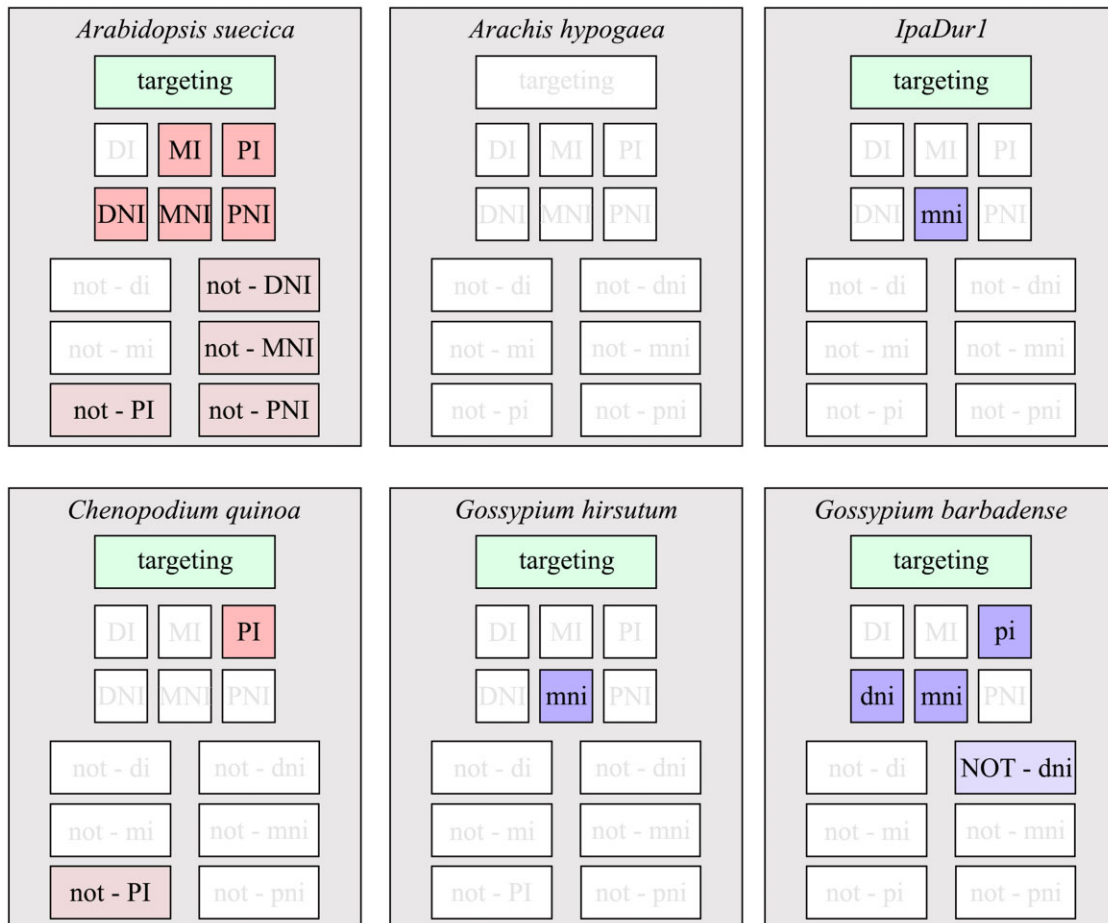
Importantly, this first model fails to account for the effects of parental legacy on expression levels in the polyploid and how deviations from parental expression levels may occur within the polyploid, the latter of which may be important depending on functional category (Supplementary Fig. 3). Therefore, we repeated the analysis with a second model that also considered the

difference in diploid expression as an explanatory term for the observed difference in homoeolog expression (i.e.  $\Delta r \log \sim \text{Targeting} + \Delta r \log_{\text{Diploid}} + \text{Targeting} \times \Delta r \log_{\text{Diploid}}$ ). We find that both targeting category and legacy expression differences ( $\Delta r \log_{\text{Diploid}}$ , representing the difference in the  $r \log$  values for the maternal and paternal diploid model species) affect homoeolog expression differences and strongly interact ( $\text{Targeting} \times \Delta r \log_{\text{Diploid}}$ ) in all comparisons (ANOVA,  $P < 0.05$ ; Fig. 4 and Supplementary Table 10).

Unlike the previous model, all of the targeting categories are significant predictors of  $\Delta r \log$  in at least 2 polyploid species (Fig. 4 and Supplementary Table 10). Additionally, contrasts in all species (except *A. hypogaea*) suggest that 2–4 targeting categories per species are significant predictors of differences in homoeolog expression beyond that predicted by NOT genes (Fig. 4 and Supplementary Table 10). Interestingly, however, the direction of these differences is not consistent and in some cases are contrary to the biological expectation that homoeolog expression differences will be more maternally biased in categories that interact with the maternally inherited organelles. Here we find few instances of greater expression divergence between homoeologs in targeting categories (vs NOT), which are limited to most categories for *A. suecica* and the DNI category in *C. quinoa* (Fig. 4 and Supplementary Table 10). Conversely, 3 categories each in *Arachis IpaDur1*, *G. hirsutum*, and *G. barbadense* and one in *C. quinoa* (MI) exhibit a greater difference between homoeologs for the NOT category, which contradicts the assumption that organelle-targeted homoeologs should preferentially upregulate maternal homoeologs and/or downregulate paternal homoeologs, both of which increase the difference in expression between homoeologs.

## Tests of functional enrichment

Functional enrichment analyses were conducted for *Arabidopsis* and *Gossypium* to further assess whether the lack of clear cytonuclear patterns were also observable through broad functional categories (vs the heretofore used CyMIRA categorizations) for those species where suitable information was available. Using a list of species-specific vocabulary terms from existing resources [i.e. TAIR (Cheng et al. 2017) and CottonFGD (Zhu et al. 2017)] to annotate our gene sets, we compared the suite of genes with greater than 4-fold differences in homoeolog expression (maternal vs

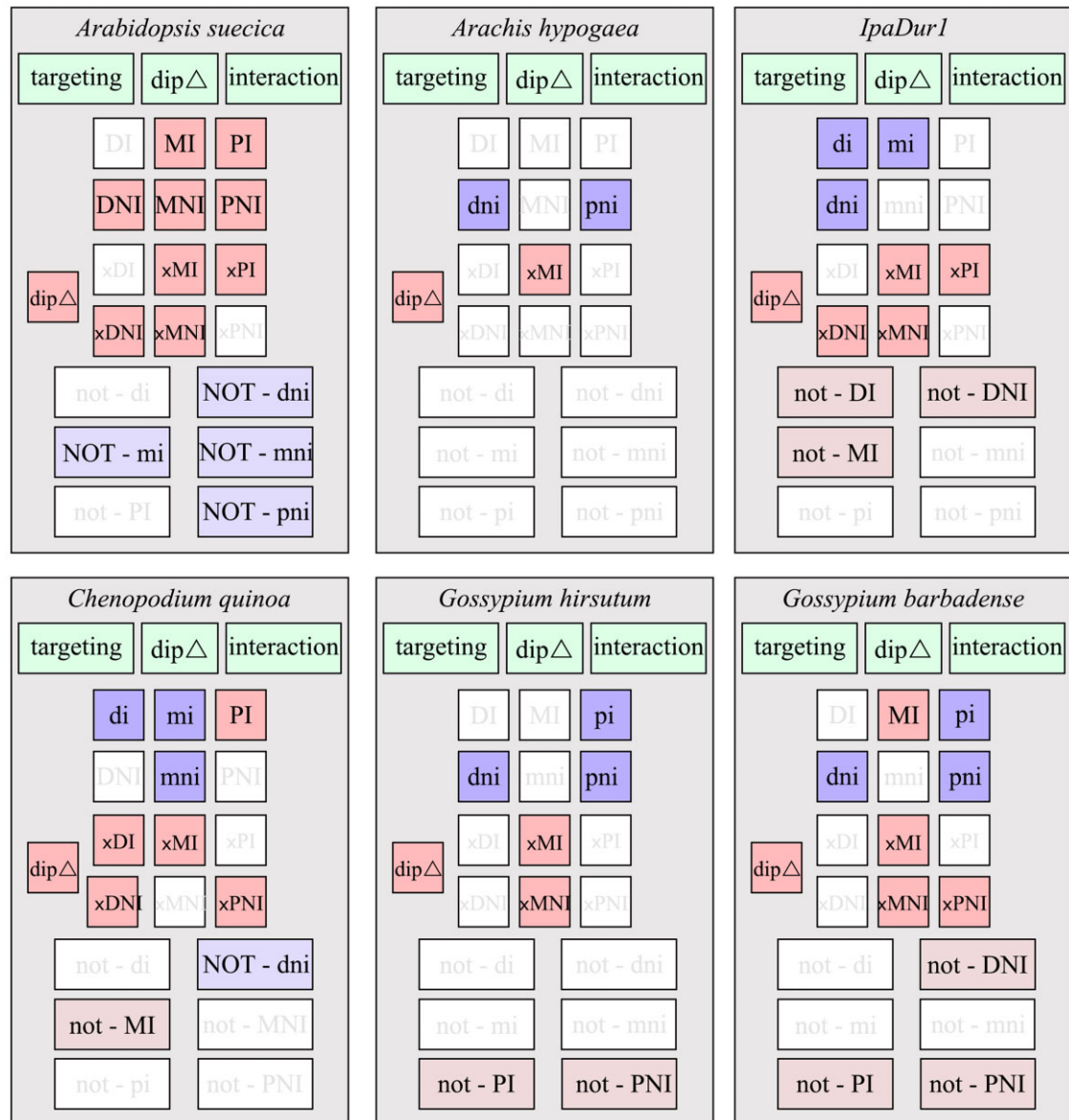


**Fig. 3.** Linear modeling results for the linear model  $expression \sim Targeting$ . Each panel represents the results from one polyploid species, with nonsignificant results displayed in gray and significant results displayed in black text. The top box in each panel indicates whether the ANOVA-based P-value for the Targeting category was significant in that species ( $P < 0.05$ ). Significant fixed effects are listed for each species (first sextet), where a red box (or uppercase letters) indicates a positive fixed effect and blue (or lowercase letters) indicates a negative fixed effect (all relative to the NOT category). Contrasts between the NOT category and each organellar category are indicated in the second sextet, where red indicates higher expression in the organellar category and blue indicates higher expression in the NOT category. Significance for all comparisons is computed at  $P < 0.05$ .

paternal) with those that exhibited any difference in homoeolog expression (regardless of FC or significance). More functional annotations are available in the model genus *Arabidopsis*, so it is unsurprising that a greater number of terms were enriched for *Arabidopsis* (126 terms; [Supplementary Table 11](#)) compared to *Gossypium* (75 and 52 terms for *G. hirsutum* and *G. barbadense*, respectively; [Supplementary Table 12](#)). Enriched terms in both *A. suecica* and *G. barbadense* were nearly evenly split with respect to parental bias, contrary to the general bias toward paternal homoeolog expression. In *A. suecica*, 65 (out of 126) terms exhibited paternal expression bias; likewise, 24 (out of 52) enriched terms exhibited paternal bias in *G. barbadense*. Conversely, *G. hirsutum* exhibited a clear maternal bias in enriched terms, i.e. 53 maternally biased terms versus 22 paternally biased ([Supplementary Table 12](#)). Of those terms exhibiting enrichment in DE genes ( $>4$ -FC, relative to background), only *G. barbadense* contained organelle relevant terms (i.e. GO : 0009523, photosystem II; GO : 0009654, photosystem II oxygen evolving complex; IPR002683, PsbP C-terminal; and GO : 0015979, photosynthesis) all of which exhibited a general bias toward maternal expression ([Supplementary Table 12](#)). Because a given gene can have multiple Gene Ontology (GO) and/or InterPro (IPR) terms associated with it, these 4 vocabulary terms represent only 5 genes in *G. barbadense* with an average 4.9-fold difference between homoeologs.

Notably, all 4 organelle relevant terms exhibited a general bias toward maternal expression ([Supplementary Table 12](#)), consistent with the biological expectation of preference for maternal cytonuclear genes.

Interestingly, although *G. barbadense* had the only organelle-relevant terms, none of these remain enriched when the analysis is restricted to only those genes exhibiting more than a 4-fold difference in expression between homoeologs and whose difference in FC between homoeologs compared to FC between diploid orthologs was at least 4.0 (i.e.  $\Delta FC_{Homoeolog} > 4$  and  $|\Delta FC_{Homoeolog} - \Delta FC_{Diploid}| > 4$ ; see [Methods](#); [Supplementary Table 12](#)), possibly indicating that some of the observed differences reflect expression patterns established in the diploid progenitors. Conversely, while no organelle-related terms were enriched in *A. suecica* when only homoeolog FC ( $\Delta FC_{Homoeolog} > 4$ ) was thresholded, a different functional term (i.e. GO : 0009941; “chloroplast envelope”) did show enrichment in the restricted set ( $\Delta FC_{Homoeolog} > 4$  and  $|\Delta FC_{Homoeolog} - \Delta FC_{Diploid}| > 4$ ). Chloroplast envelope is associated with 6 pairs of maternally biased, DE homoeologs whose average 10.6-fold difference in expression is substantially different from the average 0.6-fold difference in expression between parental orthologs. Interestingly, while chloroplast envelope alone is enriched here (and maternally biased) for *A. suecica*, the expression patterns in plastid-related CyMIRA categories (relative



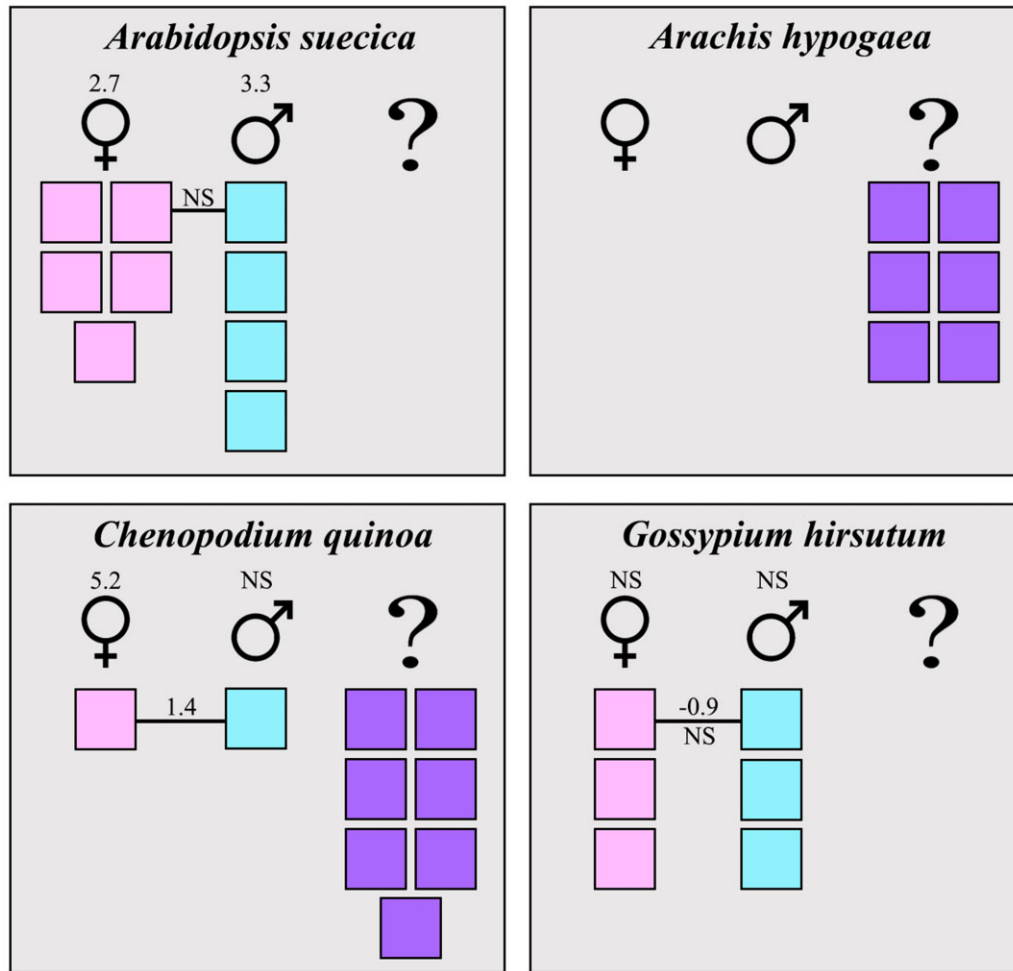
**Fig. 4.** Linear modeling results for the linear model  $\text{expression} \sim \text{Targeting} + \Delta r \log_{\text{Diploid}} + \text{Targeting} \times \Delta r \log_{\text{Diploid}}$ . Each panel represents the results from one polyloid species, with nonsignificant results displayed in gray and significant results displayed by black text. The top row of boxes in each panel indicates whether the ANOVA-based P-value for each term (or the interaction) was significant in that species ( $P < 0.05$ ). Significant fixed effects are listed for each species (first sextet), followed by the interaction between  $\Delta r \log_{\text{Diploid}}$  and each category (septet). In all cases,  $\Delta r \log_{\text{Diploid}}$  produced a significant, positive fixed effect (dip $\Delta$ ). For either set of terms, a red box (uppercase letters) indicates a positive fixed effect and blue (lowercase letters) indicates a negative fixed effect (all relative to the NOT category). Contrasts between the NOT category and each organellar category are indicated in the lower sextet, where red indicates higher expression in the organellar category and blue indicates higher expression in the NOT category. Significance for all comparisons is computed at  $P < 0.05$ .

to the diploid parents) generally contrast our expectation of maternal upregulation and/or paternal downregulation (Fig. 1).

### Expression accommodation in Rubisco

Previous analyses of the Rubisco small subunit (*rbcS*) cytonuclear gene family in multiple polyloid species reported patterns of maternally biased gene conversion and preferential expression of maternally converted paternal homoeologs (Gong et al. 2012, 2014); therefore, we specifically extracted expression patterns for *rbcS* from the current data. Consistent with previous results (Gong et al. 2012, 2014), we found that *rbcS* is composed of a small gene family in each polyloid species (Fig. 5 and Supplementary

Table 13). Because our analyses are based on available genomic and/or transcriptomic reference sequences, which are far less developed for *A. hypogaea*, we were unable to assign subgenomes (nor assess expression) for the 6 *rbcS* copies detected in either *Arachis* polyloid. For the remaining polyloids, the number of copies assigned to subgenome and/or paired as homoeologs varied depending on available information. In most cases where strict homoeologs could not be identified, it was due to copy number variation in the annotation. For example, the *A. suecica* genome is divided into *A. thaliana* (maternal) and *A. arenosa* (paternal) contigs; however, 7 of the 9 *rbcS* copies are in 2 tandem arrays (*AsAa\_g20535-37* and *AsAt\_g19714-17*), making orthology



**Fig. 5.** The number and distribution of *rbcS*-like homoeologs in each reference genome as maternal (♀; pink), paternal (♂; blue), or parentage-undetermined homoeologs (?; purple). For *Arachis* and *Chenopodium*, incomplete information prohibited assignment of individual *rbcS* copies to subgenome (i.e. all classified as "?"). Homoeolog pairs are indicated by a black line. Only one homoeologous pair was reliably determined from each genome (except *A. hypogaea*). The expression FC between homoeologs (as paternal vs maternal) is listed above the line, with the exception of the *G. hirsutum* genome reference, which lists the expression FC for *G. hirsutum* above the line and *G. barbadense* below the line. Nonsignificant FC is listed as "NS." The average expression difference (FC) between each polyploid subgenome and its model diploid progenitor is listed for each species, with the exception of *A. hypogaea*, where subgenome assignment was not possible.

difficult to determine; the unpaired copies of *rbcS* in *Chenopodium* and *Gossypium* were also a result of tandem duplications complicating orthology assignment. In general, comparisons of *rbcS* between subgenome and diploid progenitor suggest upregulation of *rbcS* in the *Arabidopsis* and *Chenopodium*, but not in either *Gossypium* species (Fig. 5 and Supplementary Table 13). Notably, the *Chenopodium rbcS* genes assigned to subgenome (i.e. paternal: AUR62042566 and maternal: AUR62018154) follow our biological expectations in that the maternal homoeolog exhibits upregulation relative to the diploid state; however, a comparison of expression between these homoeologs suggests that the paternal homoeolog is expressed 1.4-fold greater than the maternal homoeolog, contrary to the expectation that the maternal homoeolog would be preferentially expressed. We also note that 7 copies of *rbcS* were omitted from the *Chenopodium* analysis because they were not assignable to a subgenome, which may contribute to an overall bias that cannot be determined here. *G. hirsutum*, on the other hand, exhibits a slight, but statistically significant, maternal homoeolog bias (Fig. 5 and Supplementary Table 13), congruent with biological expectations; however, a

similar limitation resulting in the omission of 4 *rbcS* copies may also affect our inferences in the present analysis.

## Discussion

Allopolyploids face a complex array of challenges stemming both from whole-genome duplication and from hybridization of divergent genomes. These challenges include maintaining stoichiometric balance among interacting molecules (Birchler and Veitia 2010, 2012, 2014, 2021; Forsythe et al. 2022), which may be even more problematic for interactions between the biparentally inherited, organelle-targeted genes and those occurring in the maternally coevolved organelles (Wolf and Hager 2006; Sharbrough et al. 2017). These potential cytonuclear incompatibilities may underlie observations of rapid and repeated return to single copy for organelle-targeted genes in polyploid species (De Smet et al. 2013; Li et al. 2016) and the expectation that any paternal cytonuclear homoeologs that exhibit deleterious interactions should evolve rapidly when not immediately lost (Rand et al. 2004; Bock et al. 2014; Sloan et al. 2014). Evidence from homoploid

hybrids (Turelli and Moyle 2007; Greiner et al. 2011; Bock et al. 2014) suggests that stabilizing cytonuclear interactions is key to establishing a successful lineage, and surveys of Rubisco in diverse plant lineages (Gong et al. 2012, 2014) report differential homoeolog retention, biased expression, and asymmetric gene conversion favoring maternal homoeologs, although exceptions exist (Wang, Dong, et al. 2017; Zhai et al. 2019).

Emerging research into cytonuclear accommodation in allopolyploid species both supports and contradicts a priori cytonuclear expectations of maternal bias (Gong et al. 2012, 2014; Sehrish et al. 2015; Wang, Dong, et al. 2017; Wang, Li, et al. 2017; Ferreira de Carvalho et al. 2019; Zhai et al. 2019; Shan et al. 2020; Sharbrough et al. 2022), meaning that only some allopolyploids exhibited maternal bias in some cytonuclear genes whereas others did not. Against this backdrop of observations and expectations, we surveyed global gene expression for 5 allopolyploid species and one synthetic representing 4 different genera encompassing a wide range of divergence times to evaluate the extent to which gene expression patterns change in accordance with cytonuclear expectations. While we analyze these data here for the purpose of evaluating gene expression changes in allopolyploids, we also note that because these data include ncRNAs and organellar reads, they represent a valuable resource for the allopolyploid community and the plant community at large.

### Total gene expression exhibits limited evidence of cytonuclear maternal expression level dominance

Cytonuclear imbalance in polyploids could potentially arise due to the changes in dosage balance between organellar and nuclear genomes that accompany polyploidy. In response, the nascent polyploid might be expected to experience selection to mitigate any dosage-related detrimental effects by altering total gene expression in either the organelle or nucleus. Changes in organelle copy number and/or genome copy per organelle have been associated with polyploidy (Bingham 1968; Beversdorf 1979; Dean and Leech 1982; Butterfass 1987; Murti et al. 2012; Oberprieler et al. 2019; Coate et al. 2020; Fernandes Gyorfy et al. 2021; He et al. 2021), and these have been associated with cytonuclear compensation at the expression level (Doyle and Coate 2019; Coate et al. 2020; Forsythe et al. 2022). On the other hand, it is common for polyploids to undergo rapid changes in nuclear expression (Chen 2007; Doyle et al. 2008; Freeling 2009; Gaeta and Pires 2010; Jackson and Chen 2010; Salmon et al. 2010; Grover et al. 2012; Madlung and Wendel 2013; Yoo et al. 2014; Song and Chen 2015; Bao et al. 2019; Gallagher et al. 2020), which could include changes that compensate for deleterious cytonuclear stoichiometric imbalances.

In the present study, we characterized how total expression of nuclear-encoded cytonuclear genes changes relative to the rest of the transcriptome and whether those changes are biased toward the maternal parent. We evaluated each polyploid for evidence of maternally biased ELD in cytonuclear genes that is statistically different from any global, or background, bias exhibited by genes not involved in cytonuclear processes. Our expectation was that we would observe some degree of ELD for cytonuclear genes that might provide evidence of cytonuclear compensation to coordinate expression with the maternally coevolved organelles. While 3 of the 5 polyploids (i.e. *A. hypogaea*, *G. hirsutum*, and *G. barbadense*) exhibited a general bias toward maternal ELD, only *A. hypogaea* and *C. quinoa* exhibited an excess of maternal ELD in cytonuclear genes (in general) relative to the remaining transcriptome (i.e. NOT genes; Table 4), with only 1–2

categories exhibiting evidence of significant ELD (DNI in both species and MI in *C. quinoa*). Interestingly, however, *G. barbadense*, while not exhibiting a general parental bias in cytonuclear ELD, did exhibit maternally biased ELD in the PI cytonuclear category alone. While these results suggest that global ELD in cytonuclear genes is not a general consequence of cytonuclear accommodation, it is noteworthy that in many cases, and for all species, the number of genes exhibiting maternally biased ELD in cytonuclear categories does exceed the expected number (although not significantly so). This may be a function of the limited numbers of genes in each category, trends of partial yet nonubiquitous maternally biased ELD in cytonuclear categories, and/or both. While we also evaluated patterns of transgressive expression in cytonuclear categories relative to NOT genes, we did not find evidence of biased transgressive expression that would indicate a global up- or downregulation of cytonuclear genes to compensate for the number of organelles/organelle genomes; however, we again note that most categories were limited in membership, leading to low statistical power. In addition, evidence of cytonuclear incompatibility is best detected among rapidly evolving cytonuclear genomes (Burton and Barreto 2012; Osada and Akashi 2012; Sloan et al. 2014; Havird et al. 2015; Forsythe et al. 2021; Sharbrough et al. 2022), perhaps indicating that the protein differences between the allopolyploid parents in the organelles from these genera is insufficient to stimulate strong patterns of cytonuclear accommodation on this time scale.

### Variability in cytonuclear homoeolog expression patterns

Cytonuclear imbalance in allopolyploid species can also arise from incompatibilities between the organellar genomes and the more divergent paternal cytonuclear genes, and we expect these to become more common as the divergence time between progenitor genomes increases. Reconciliation of potentially maladaptive mutations is possible through a variety of mechanisms, as previously noted (Gong et al. 2012, 2014; Sharbrough et al. 2017). For example, at the genomic level, gene loss and maternally biased gene conversion could either remove or “correct” maladaptive mutations acquired by the paternal genome since its divergence from a common ancestor, minimizing their deleterious potential (Sharbrough et al. 2022).

With respect to expression, compensation for maladaptive paternal mutations could present as a combination of upregulated maternal homoeologs and/or downregulated paternal homoeologs. This, however, does not appear to be a global reaction to allopolyploidy in the species surveyed. When we compared homoeolog expression for each of the 6 allopolyploid species with their respective diploid progenitor genomes, we observed no clear and consistent pattern of homoeolog up-/downregulation within polyploids or for any of the cytonuclear categories. At most, any given polyploid displayed 2 cytonuclear categories consistent with our biological expectations of excess maternal upregulation and/or paternal downregulation (Fig. 1), and concomitantly have as many or more categories that directly contradict our cytonuclear predictions (i.e. enrichment of maternal downregulation and/or paternal upregulation). Individual cytonuclear categories were no more consistent, with the MI category being most frequently consistent (i.e. agreed with expectations in 2 species, *A. suecica* and *A. hypogaea*), while also being contradictory in the same number of species (*Arachis IpaDur1* and *C. quinoa*).

Maternal HEB (i.e. genes where maternal expression outweighs paternal, irrespective of diploid expression) was similarly intermittent in cytonuclear categories. When compared to any

global HEB exhibited by each species, few cytonuclear categories exhibited an excess of maternal HEB (i.e. *A. suecica* PI/PNI and *G. hirsutum* DI only; Table 5). Interestingly, a single category in *C. quinoa* (MI) and several in *G. barbadense* (DNI/PI/PNI) exhibited an excess of paternal HEB, which is contrary to cytonuclear expectations. We do note, however, that these relative expression biases are often parentally inherited, as noted by the previous analysis.

Importantly, our analytical methodology was designed to disentangle parental or progenitor legacy effects (i.e. differences at the diploid level vertically inherited in the polyploids at formation) from evolved cytonuclear responses subsequent to polyploid formation (Buggs et al. 2014). When we combined our assessment of homoeolog expression differences with legacy parental effects (Fig. 4 and Supplementary Table 10), we find that not only do targeting (i.e. cytonuclear category) and legacy diploid expression influence the difference in homoeolog expression, but there is also an interaction between targeting and legacy expression differences. Interestingly, however, many of the fixed effects are not congruent with our expectations under the cytonuclear hypotheses, i.e. that the difference in  $r$  log counts between maternal and paternal homoeologs should be greater in the cytonuclear categories (or positive relative to the intercept established by NOT genes). Contrasts between each cytonuclear category and NOT genes also exhibited sporadic significance and were frequently incongruent with expectations (i.e. that the cytonuclear categories would exhibit greater HEB when accounting for diploid legacy) for most species. Only *A. suecica* showed significant, congruent cytonuclear effects for most categories, suggesting that DNI, MI/MNI, and PNI were generally composed of genes whose maternal HEB was greater than expected by NOT and diploid legacy.

In light of previous research that both supports and contrasts the results presented here, we speculate that cytonuclear accommodation is variable among lineages, among cytonuclear categories, and among genes within categories themselves. It also may be that for most genes (and especially those in the organellar genomes, which experience low mutation rates), the rates of molecular evolution are too low to permit signals of cytonuclear selection to become evident on the divergence scales studied here. It is possible, for example, that cytonuclear selection is ongoing and even pervasive, but that for the most part it is subtle, involving expression level changes or genomic signatures that simply do not rise to the level of statistical significance given the timescales encompassed by the allopolyploids studied here. Some polyploids, such as *A. suecica*, provide a modest level of support for our a priori expectations for cytonuclear accommodation vis-à-vis gene expression, whereas others, such as *G. barbadense*, contradict expectations more frequently than not. The variability in our observations may suggest that species with fewer cytonuclear-congruent expression changes either have fewer detrimental cytonuclear incompatibilities and/or have other methods for resolving deleterious conflict between the coevolved maternal sub-genome, the potentially detrimental paternal homoeologs, and the cytoplasmically inherited organelles.

## Data availability

All sequence data used in the analysis are available from NCBI under PRJNA726938, and all scripts used to analyze the data are available from Github under <https://github.com/Wendellab/CytonuclearExpression> commit 0509853.

Supplemental material is available at GENETICS online.

## Acknowledgements

We thank Andreas Madlung and Roswitha Shmickl for providing *Arabidopsis* seeds, and David Brenner for providing *Chenopodium* plants. We thank Kenneth McCabe for supportive plant care in the Pohl Conservatory, and ResearchIT at Iowa State University for computational support. This work utilized the Summit supercomputer, which was supported by the National Science Foundation (awards ACI-1532235 and ACI-1532236), the University of Colorado Boulder, and Colorado State University. The Summit supercomputer is a joint effort of the University of Colorado Boulder and Colorado State University.

## Funding

This work was supported by the National Science Foundation (NSF) awards IOS-1829176 (DBS, JFW, JS, and CEG), IOS-2145811 (JS), and U.S. Department of Agriculture ARS 58-6066-0-066 (DGP, CEG, and JFW).

## Conflicts of interest

The authors declare no conflict of interest.

## Literature cited

- Adams KL, Wendel JF. Polyploidy and genome evolution in plants. *Curr Opin Plant Biol.* 2005;8(2):135–141.
- Avraham S, Tung C-W, Ilic K, Jaiswal P, Kellogg EA, McCouch S, Pujar A, Reiser L, Rhee SY, Sachs MM, et al. The Plant Ontology Database: a community resource for plant structure and developmental stages controlled vocabulary and annotations. *Nucleic Acids Res.* 2008;36(Database issue):D449–D454.
- Bannai H, Tamada Y, Maruyama O, Nakai K, Miyano S. Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics.* 2002;18(2):298–305.
- Bao Y, Hu G, Grover CE, Conover J, Yuan D, Wendel JF. Unraveling cis and trans regulatory evolution during cotton domestication. *Nat Commun.* 2019;10(1):5399.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 1995;57(1):289–300.
- Bertioli DJ, Jenkins J, Clevenger J, Dudchenko O, Gao D, Seijo G, Leal-Bertioli SCM, Ren L, Farmer AD, Pandey MK, et al. The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nat Genet.* 2019;51(5):877–884.
- Beversdorf WD. Influence of ploidy level on several plant characteristics in soybeans. *Can J Plant Sci.* 1979;59(4):945–948.
- Bingham ET. Stomatal chloroplasts in alfalfa at four ploidy levels. *Crop Sci.* 1968;8(4):509–510.
- Birchler JA, Veitia RA. The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. *New Phytol.* 2010;186(1):54–62.
- Birchler JA, Veitia RA. Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc Natl Acad Sci USA.* 2012;109(37):14746–14753.
- Birchler JA, Veitia RA. The gene balance hypothesis: dosage effects in plants. *Methods Mol Biol.* 2014;1112:25–32.
- Birchler JA, Veitia RA. One hundred years of gene balance: how stoichiometric issues affect gene expression, genome evolution, and quantitative traits. *Cytogenet Genome Res.* 2021;161:529–550.

- Bird KA, VanBuren R, Puzey JR, Edger PP. The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytol.* 2018;220(1):87–93.
- Bock DG, Andrew RL, Rieseberg LH. On the adaptive value of cytoplasmic genomes in plants. *Mol Ecol.* 2014;23(20):4899–4911.
- Bray NL, Pimentel H, Melsted P, Pachter L. Erratum: near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* 2016;34(8):888.
- Budar F, Mireau H. The cross-talk between genomes: how co-evolution shaped plant mitochondrial gene expression. In: *Annual Plant Reviews Online*. JA Roberts, editor. Chichester, UK: John Wiley & Sons, Ltd; 2018. p. 33–66.
- Buggs RJA, Wendel JF, Doyle JJ, Soltis DE, Soltis PS, Coate JE. The legacy of diploid progenitors in allopolyploid gene expression patterns. *Phil Trans R Soc B.* 2014;369(1648):20130354.
- Buggs RJA, Zhang L, Miles N, Tate JA, Gao L, Wei W, Schnable PS, Barbazuk WB, Soltis PS, Soltis DE, et al. Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. *Curr Biol.* 2011;21(7):551–556.
- Burke JM, Arnold ML. Genetics and the fitness of hybrids. *Annu Rev Genet.* 2001;35:31–52.
- Burton RS, Barreto FS. A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? *Mol Ecol.* 2012;21(20):4942–4957.
- Burton RS, Pereira RJ, Barreto FS. Cytonuclear genomic interactions and hybrid breakdown. *Annu Rev Ecol Evol Syst.* 2013;44(1):281–302.
- Butterfass T. Cell volume ratios of natural and of induced tetraploid and diploid flowering plants. *Cytologia.* 1987;52(2):309–316.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinform.* 2009;10:421.
- Camus MF, Alexander-Lawrie B, Sharbrough J, Hurst GDD. Inheritance through the cytoplasm. *Heredity.* 2022;129(1):31–43.
- Chen X, Li H, Pandey MK, Yang Q, Wang X, Garg V, Li H, Chi X, Doddamani D, Hong Y, et al. Draft genome of the peanut A-genome progenitor (*Arachis duranensis*) provides insights into geocarpy, oil biosynthesis, and allergens. *Proc Natl Acad Sci USA.* 2016;113(24):6785–6790.
- Chen ZJ. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu Rev Plant Biol.* 2007;58:377–406.
- Chen ZJ, Sreedasyam A, Ando A, Song Q, De Santiago LM, Hulse-Kemp AM, Ding M, Ye W, Kirkbride RC, Jenkins J, et al. Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. *Nat Genet.* 2020;52(5):525–533.
- Cheng C-Y, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD. Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* 2017;89(4):789–804.
- Cheng F, Wu J, Cai X, Liang J, Freeling M, Wang X. Gene retention, fractionation and subgenome differences in polyploid plants. *Nat Plants.* 2018;4(5):258–268.
- Coate JE, Schreyer WM, Kum D, Doyle JJ. Robust cytonuclear coordination of transcription in nascent *Arabidopsis thaliana* autopolyploids. *Genes.* 2020;11(2):134.
- Conant GC, Birchler JA, Pires JC. Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Curr Opin Plant Biol.* 2014;19:91–98.
- Conant GC, Wolfe KH. Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet.* 2008;9(12):938–950.
- Conover JL, Sharbrough J, Wendel JF. pSONIC: ploidy-aware syntenic orthologous networks identified via collinearity. *G3 (Bethesda).* 2021;11:jkab170.
- Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods.* 2012;9(8):772.
- Dash S, Cannon EKS, Kalberer SR, Farmer AD, Cannon SB. Chapter 8—PeanutBase and other bioinformatic resources for peanut. In: *Peanuts*. HT Stalker and RF Wilson, editor. Amsterdam, the Netherlands: AOCSS Press; 2016. p. 241–252.
- Dean C, Leech RM. Genome expression during normal leaf development: I. Cellular and chloroplast numbers and DNA, RNA, and protein levels in tissues of different ages within a seven-day-old wheat leaf. *Plant Physiol.* 1982;69(4):904–910.
- De Smet R, Adams KL, Vandepoele K, Van Montagu MCE, Maere S, Van de Peer Y. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proc Natl Acad Sci USA.* 2013;110(8):2898–2903.
- Doyle JJ, Coate JE. Polyploidy, the nucleotype, and novelty: the impact of genome doubling on the biology of the cell. *Int J Plant Sci.* 2019;180(1):1–52.
- Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF. Evolutionary genetics of genome merger and doubling in plants. *Annu Rev Genet.* 2008;42:443–461.
- Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc.* 2007;2(4):953–971.
- Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics, 2019;20:238. <https://doi.org/10.1186/s13059-019-1832-y>.
- Fávero AP, Simpson CE, Valls JFM, Vello NA. Study of the evolution of cultivated peanut through crossability studies among *Arachis ipaënsis*, *A. duranensis*, and *A. hypogaea*. *Crop Sci.* 2006;46(4):1546–1552.
- Fernandes Gyorfy M, Miller ER, Conover JL, Grover CE, Wendel JF, Sloan DB, Sharbrough J. Nuclear-cytoplasmic balance: whole genome duplications induce elevated organellar genome copy number. *Plant J.* 2021;108(1):219–230.
- Ferreira de Carvalho J, Lucas J, Deniot G, Falentin C, Filangi O, Gilet M, Legeai F, Lode M, Morice J, Trotoux G, et al. Cytonuclear interactions remain stable during allopolyploid evolution despite repeated whole-genome duplications in *Brassica*. *Plant J.* 2019;98(3):434–447.
- Fishman L, Sweigart AL. When two rights make a wrong: the evolutionary genetics of plant hybrid incompatibilities. *Annu Rev Plant Biol.* 2018;69:707–731.
- Fishman L, Willis JH. A cytonuclear incompatibility causes anther sterility in *Mimulus* hybrids. *Evolution.* 2006;60(7):1372–1381.
- Forsythe ES, Grover CE, Miller ER, Conover JL, Arick MA, Chavarro MCF, Leal-Bertioli SCM, Peterson DG, Sharbrough J, Wendel JF, et al. Organellar transcripts dominate the cellular mRNA pool across plants of varying ploidy levels. *Proc Natl Acad Sci USA.* 2022;119(30):e2204187119.
- Forsythe ES, Sharbrough J, Havird JC, Warren JM, Sloan DB. CyMIRA: the Cytonuclear Molecular Interactions Reference for *Arabidopsis*. *Genome Biol. Evol.* 2019;11(8):2194–2202. <https://doi.org/10.1093/gbe/evz144>.
- Forsythe ES, Williams AM, Sloan DB. Genome-wide signatures of plastid-nuclear coevolution point to repeated perturbations of plastid proteostasis systems across angiosperms. *Plant Cell.* 2021;33(4):980–997.
- Freeling M. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu Rev Plant Biol.* 2009;60:433–453.
- Gaeta RT, Pires JC. Homoeologous recombination in allopolyploids: the polyploid ratchet. *New Phytol.* 2010;186(1):18–28.

- Gallagher JP, Grover CE, Hu G, Jareczek JJ, Wendel JF. Conservation and divergence in duplicated fiber coexpression networks accompanying domestication of the polyploid *Gossypium hirsutum* L. G3 (Bethesda). 2020;10(8):2879–2892.
- Giannakis K, Arrowsmith SJ, Richards L, Gasparini S, Chustecki JM, et al. Universal features shaping organelle gene retention, 2021. bioRxiv 2021.10.27.465964. <https://doi.org/10.1101/2021.10.27.465964>
- Gong L, Olson M, Wendel JF. Cytonuclear evolution of rubisco in four allopolyploid lineages. *Mol Biol Evol.* 2014;31(10):2624–2636.
- Gong L, Salmon A, Yoo M-J, Grupp KK, Wang Z, Paterson AH, Wendel JF. The cytonuclear dimension of allopolyploid evolution: an example from cotton using rubisco. *Mol Biol Evol.* 2012;29(10):3023–3036.
- Gout J-F, Lynch M. Maintenance and loss of duplicated genes by dosage subfunctionalization. *Mol Biol Evol.* 2015;32(8):2141–2148.
- Greiner S, Bock R. Tuning a ménage à trois: co-evolution and co-adaptation of nuclear and organellar genomes in plants. *Bioessays.* 2013;35(4):354–365.
- Greiner S, Rauwolf U, Meurer J, Herrmann RG. The role of plastids in plant speciation. *Mol Ecol.* 2011;20(4):671–691.
- Grover CE, Gallagher JP, Szadkowski EP, Yoo MJ, Flagel LE, Wendel JF. Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytol.* 2012;196(4):966–971.
- Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 2003; 52(5):696–704.
- Havird JC, Whitehill NS, Snow CD, Sloan DB. Conservative and compensatory evolution in oxidative phosphorylation complexes of angiosperms with highly divergent rates of mitochondrial genome evolution. *Evolution.* 2015;69(12):3069–3081.
- He H, Xie W, Liang Z, Wu H, Bai M. The expansion of mesophyll cells is coordinated with the division of chloroplasts in diploid and tetraploid *Arabidopsis thaliana*. *Planta.* 2021;253(3):64.
- Hill GE. The mitonuclear compatibility species concept. *Auk.* 2017; 134(2):393–409.
- Holbrook CC, Culbreath AK. Registration of “Tifrunner” peanut. *J Plant Regist.* 2007;1:10–3198.
- Jackson S, Chen ZJ. Genomic and expression plasticity of polyploidy. *Curr Opin Plant Biol.* 2010;13(2):153–159.
- Jarvis DE, Ho YS, Lightfoot DJ, Schmöckel SM, Li B, Borm TJA, Ohyanagi H, Mineta K, Michell CT, Saber N, et al. The genome of *Chenopodium quinoa*. *Nature.* 2017;542(7641):307–312.
- Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014;30(9): 1236–1240.
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772–780.
- Kolde R. Pheatmap: pretty heatmaps. R package version, 2012.
- Leal-Bertioli SCM, Godoy IJ, Santos JF, Doyle JJ, Guimarães PM, Abernathy BL, Jackson SA, Moretzsohn MC, Bertioli DJ. Segmental allopolyploidy in action: increasing diversity through polyploid hybridization and homoeologous recombination. *Am J Bot.* 2018; 105(6):1053–1066.
- Levin DA. The cytoplasmic factor in plant speciation. *Syst Bot.* 2003; 28:5–11.
- Li Z, Defoort J, Tasdighian S, Maere S, Van de Peer Y, De Smet R. Gene duplicability of core genes is highly consistent across all angiosperms. *Plant Cell.* 2016;28(2):326–344.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
- Madlung A, Wendel JF. Genetic and epigenetic aspects of polyploid evolution in plants. *Cytogenet Genome Res.* 2013;140(2–4): 270–285.
- Mason AS, Wendel JF. Homoeologous exchanges, segmental allopolyploidy, and polyploid genome evolution. *Front Genet.* 2020; 11:1014.
- Murti RH, Kim HY, Yeoung YR. Morphological and anatomical characters of ploidy mutants of strawberry. *Int J Agric Biol.* 2012;14: 1560–8530.
- Nieto Feliner G, Casacuberta J, Wendel JF. Genomics of evolutionary novelty in hybrids and polyploids. *Front Genet.* 2020;11:792.
- Novikova PY, Tsuchimatsu T, Simon S, Nizhynska V, Voronin V, Burns R, Fedorenko OM, Holm S, Säll T, Prat E, et al. Genome sequencing reveals the origin of the allotetraploid *Arabidopsis suecica*. *Mol Biol Evol.* 2017;34(4):957–968.
- Oberprieler C, Talianova M, Griesenbeck J. Effects of polyploidy on the coordination of gene expression between organellar and nuclear genomes in *Leucanthemum* Mill. (Compositae, Anthemideae). *Ecol Evol.* 2019;357:1451.
- Osada N, Akashi H. Mitochondrial–nuclear interactions and accelerated compensatory evolution: evidence from the primate cytochrome c oxidase complex. *Mol Biol Evol.* 2012;29(1):337–346.
- Panchy N, Lehti-Shiu M, Shiu S-H. Evolution of gene duplication in plants. *Plant Physiol.* 2016;171(4):2294–2316.
- Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D, Showmaker KC, Shu S, Udall J, et al. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature.* 2012;492(7429):423–427.
- Postel Z, Touzet P. Cytonuclear genetic incompatibilities in plant speciation. *Plants.* 2020;9(4):487.
- Rand DM, Haney RA, Fry AJ. Cytonuclear coevolution: the genomics of cooperation. *Trends Ecol Evol.* 2004;19(12):645–653.
- Rapp RA, Udall JA, Wendel JF. Genomic expression dominance in allopolyploids. *BMC Biol.* 2009;7:18.
- R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020.
- Salmon A, Flagel L, Ying B, Udall JA, Wendel JF. Homoeologous non-reciprocal recombination in polyploid cotton. *New Phytol.* 2010; 186(1):123–134.
- Sehrish T, Symonds VV, Soltis DE, Soltis PS, Tate JA. Cytonuclear coordination is not immediate upon allopolyploid formation in *Tragopogon miscellus* (Asteraceae) allopolyploids. *PLoS One.* 2015; 10(12):e0144339.
- Shan S, Boatwright JL, Liu X, Chanderbali AS, Fu C, Soltis PS, Soltis DE. Transcriptome dynamics of the inflorescence in reciprocally formed allopolyploid *Tragopogon miscellus* (Asteraceae). *Front Genet.* 2020;11:888.
- Sharbrough J, Conover JL, Fernandes Gyorfy M, Grover CE, Miller ER, Wendel JF, Sloan DB. Global patterns of subgenome evolution in organelle-targeted genes of six allotetraploid angiosperms. *Mol Biol Evol.* 2022;39(4):msac074.
- Sharbrough J, Conover JL, Tate JA, Wendel JF, Sloan DB. Cytonuclear responses to genome doubling. *Am J Bot.* 2017;104(9):1277–1280.
- Sloan DB, Havird JC, Sharbrough J. The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Mol Ecol.* 2017;26(8):2212–2236.
- Sloan DB, Triant DA, Wu M, Taylor DR. Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. *Mol Biol Evol.* 2014;31(3):673–682.
- Slomovic S, Portnoy V, Liveanu V, Schuster G. RNA polyadenylation in prokaryotes and organelles; different tails tell different tales. *CRC Crit Rev Plant Sci.* 2006;25(1):65–77.



- Slomovic S, Portnoy V, Schuster G. Detection and characterization of polyadenylated RNA in Eukarya, Bacteria, Archaea, and organelles. *Methods Enzymol.* 2008;447:501–520.
- Small I, Peeters N, Legeai F, Lurin C. Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics.* 2004;4(6):1581–1590.
- Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2015; 2013–2015.
- Smith DR. RNA-Seq data: a goldmine for organelle research. *Brief Funct Genomics.* 2013;12(5):454–456.
- Song Q, Chen ZJ. Epigenetic and developmental regulation in plant polyploids. *Curr Opin Plant Biol.* 2015;24:101–109.
- Sperschneider J, Catanzariti A-M, DeBoer K, Petre B, Gardiner DM, Singh KB, Dodds PN, Taylor JM. LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. *Sci Rep.* 2017;7:44598.
- The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* 2019;47:D330–D338.
- Turelli M, Moyle LC. Asymmetric postmating isolation: Darwin's corollary to Haldane's rule. *Genetics.* 2007;176(2):1059–1088.
- Wang J, Qin J, Sun P, Ma X, Yu J, Li Y, Sun S, Lei T, Meng F, Wei C, et al. Polyploidy index and its implications for the evolution of polyploids. *Front Genet.* 2019;10:807.
- Wang X, Dong Q, Li X, Yuliang A, Yu Y, Li N, Liu B, Gong L. Cytonuclear variation of Rubisco in synthesized rice hybrids and allotetraploids. *Plant Genome.* 2017;10(3):lantgenome2017.05.0041.
- Wang Z, Li J, Chen S, Heng Y, Chen Z, Yang J, Zhou K, Pei J, He H, Deng XW, et al. Poaceae-specific MS1 encodes a phospholipid-binding protein for male fertility in bread wheat. *Proc Natl Acad Sci USA.* 2017;114(47):12614–12619.
- Wickham H. *ggplot2: Elegant Graphics for Data Analysis.* New York: Springer-Verlag; 2016.
- Wolf JB, Hager R. A maternal–offspring coadaptation theory for the evolution of genomic imprinting. *PLoS Biol.* 2006;4(12):e380.
- Yoo M-J, Liu X, Pires JC, Soltis PS, Soltis DE. Nonadditive gene expression in polyploids. *Annu Rev Genet.* 2014;48:485–517.
- Yoo M-J, Szadkowski E, Wendel JF. Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity.* 2013;110(2):171–180.
- Zhai Y, Yu X, Zhu Z, Wang P, Meng Y, Zhao Q, Li J, Chen J. Nuclear–cytoplasmic coevolution analysis of RuBisCO in synthesized *Cucumis* allopolyploid. *Genes.* 2019;10(11):869.
- Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics.* 2012;28(24):3326–3328.
- Zhuang W, Chen H, Yang M, Wang J, Pandey MK, Zhang C, Chang W-C, Zhang L, Zhang X, Tang R, et al. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat Genet.* 2019;51(5):865–876.
- Zhu T, Liang C, Meng Z, Sun G, Meng Z, Guo S, Zhang R. CottonFGD: an integrated functional genomics database for cotton. *BMC Plant Biol.* 2017;17(1):101.

Communicating editor: J. Birchler