Global patterns of subgenome evolution in organelle-targeted genes of six allotetraploid angiosperms

Joel Sharbrough[1, 2, *, ¶], Justin L. Conover[3, ¶], Matheus Fernandes Gyorfy[1], Corrinne E. Grover[3], Emma R. Miller[3], Jonathan F. Wendel[3], Daniel B. Sloan[1]

[1] – Biology Department, Colorado State University, Fort Collins, CO, USA
[2] – Biology Department, New Mexico Institute of Mining and Technology, Socorro, NM, USA
[3] – Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA, USA

*Corresponding author
Email: joel.sharbrough@nmt.edu

[¶] – These authors contributed equally to this work.

JS: 0000-0002-3642-1662
JLC: 0000-0002-3558-6000
MFG: 0000-0002-2027-8128
CEG: 0000-0003-3878-5459
ERM: 0000-0001-9009-5303
JFW: 0000-0003-2258-5081
DBS: 0000-0002-3618-0897

ABSTRACT

Whole-genome duplications (WGDs), in which the number of nuclear genome copies is elevated as a result of autopolyploidy or allopolyploidy, are a prominent process of diversification in eukaryotes. The genetic and evolutionary forces that WGD imposes upon cytoplasmic genomes are not well understood, despite the central role that cytonuclear interactions play in eukaryotic function and fitness. Cellular respiration and photosynthesis depend upon successful interaction between the 3000+ nuclear-encoded proteins destined for the mitochondria or plastids and the gene products of cytoplasmic genomes in multi-subunit complexes such as OXPHOS, organellar ribosomes, Photosystems I and II, and Rubisco. Allopolyploids are thus faced with the critical task of coordinating interactions between nuclear and cytoplasmic genes that were inherited from different species. Because cytoplasmic genomes share a more recent history of common descent with the maternal nuclear subgenome than the paternal subgenome, evolutionary "mismatches" between the paternal subgenome and the cytoplasmic genomes in allopolyploids might lead to accelerated rates of evolution in the paternal homoeologs of allopolyploids, either through relaxed purifying selection or strong directional selection to rectify these mismatches. We tested this hypothesis in maternal vs. paternal copies of organelle-targeted genes in six allotetraploids: *Brachypodium hybridum*, *Chenopodium quinoa*, *Coffea arabica*, *Gossypium hirsutum*, *Nicotiana tabacum*, and *Triticum dicoccoides*. We report evidence that allopolyploid subgenomes exhibit unequal rates of protein-sequence evolution, but we did not observe global effects of cytonuclear incompatibilities on paternal homoeologs of organelle-targeted genes. Analyses of gene content revealed mixed evidence for whether organelle-targeted genes re-diploidize more rapidly than non-organelle-targeted genes. Together, these global analyses provide insights into the complex evolutionary dynamics of allopolyploids, showing that allopolyploid subgenomes have separate evolutionary trajectories despite sharing the same nucleus, generation time, and ecological context.

Keywords:
Allopolyploidy, *Brachypodium hybridum*, *Chenopodium quinoa*, chloroplast, *Coffea arabica,* CyMIRA, cytonuclear incompatibility, *Gossypium hirsutum*, mitochondrion, *Nicotiana tabacum, Triticum dicoccoides*

AUTHOR SUMMARY

Whole genome duplication, in which the size and content of the nuclear genome is instantly doubled, represents one of the most profound forms of mutational change. The consequences of duplication events are equally monumental, especially considering that almost all eukaryotes have undergone whole genome duplications during their

evolutionary history. While myriad genetic, cellular, organismal, and ecological effects of whole genome duplications have been extensively documented, relatively little attention has been paid to the diminutive but essential “other” genomes present inside the cell, those of chloroplasts and mitochondria. In this study, we compared the evolutionary patterns of >340,000 genes from 23 species to test whether whole genome duplications are associated with genetic mismatches between the nuclear, mitochondrial, and chloroplast genomes. We discovered tremendous differences between duplicated copies of nuclear genomes; however, mitochondria-nuclear and chloroplast-nuclear mismatches do not appear to be common following whole genome duplications. Together these genomic data represent the most extensive analysis yet performed on how polyploids maintain the delicate and finely tuned balance between the nuclear, mitochondrial, and chloroplast genomes.

INTRODUCTION

Whole genome duplication (WGD) events, in which the nuclear genome is doubled via polyploidization, are among the most profound mutational changes observed in nature. The high frequency of WGDs, especially among flowering plants [1–4], makes them a major force in genome evolution. Accordingly, evolutionary biologists have had a great deal of interest in exploring the consequences of and responses to WGD. The ensuing studies have shown that the effects of WGDs are far-ranging, including silencing and loss of duplicated genes [5–11], mobilization of previously dormant transposable elements [12–17], inter-genomic gene conversion and homoeologous chromosome exchanges [18–25], alterations of epigenetic marks [26–33], massive, genome-wide transcriptional rewiring [6,34–41], and a host of other associated physiological, ecological, and life-history changes [42–54]. Whole genome duplications are also expected to produce novel interactions between the nuclear genome and the mitochondrial and plastid genomes [55], but this dimension of allopolyploid evolution has received relatively little attention (but see [56–62]).

Cytonuclear interactions are themselves the result of gene transfers from the cytoplasmic genomes (mitochondrial and plastid) to the nuclear genome or the recruitment of existing nuclear-encoded proteins to function in these organelles [63,64]. As a result, the vast majority of the ~2000 proteins that comprise the mitochondrial proteome [65] and ~3000 proteins that comprise the plastid proteome [66] are nuclear-encoded [67]. Many of these nuclear-encoded proteins directly interact with gene products from the cytoplasmic genomes to form heteromeric complexes (e.g., Rubisco, Photosystems I and II, organellar ribosomes and the enzymes that comprise the mitochondrial electron transport chain). Additionally, the replication, expression, and post-transcriptional modifications of cytoplasmic genomes are dependent on nuclear-encoded proteins [68–71], as are the many biosynthetic and signaling functions of mitochondria and plastids [72–78]. Taken together, the cellular and metabolic functions that result from cytonuclear interactions, especially aerobic respiration and photosynthesis, are critically important to eukaryotic health and fitness [79–83]. Perturbations to one genomic compartment can therefore have dramatic consequences for the other genomic compartments [84–90], so much so that incompatibilities between the nuclear and cytoplasmic genomes may be a potent force in generating and reinforcing species boundaries [91–95].

Allopolyploidization, a WGD event resulting from a genome merger of two differentiated species [96–98], is expected to perturb cytonuclear interactions because the cytoplasmic genomes have a more recent history of shared descent with one nuclear subgenome than the other [55]. Researchers have hypothesized a number of immediate and evolutionary responses that may mitigate any resulting deleterious consequences. First, maternally biased nuclear gene expression in recently formed

allopolyploid lineages could alleviate the deleterious consequences of incompatibilities between the paternal nuclear subgenome and the cytoplasmic genomes [56]. Over time, evolutionary rates may vary across nuclear subgenomes, with paternal copies of organelle-targeted genes evolving faster than maternal copies, either as a reflection of relaxed selection [99] or positive selection to rectify mismatches with the cytoplasmic genomes [90]. In the long run, paternal copies of organelle-targeted genes may be altered more frequently than maternal copies as a result of maternally biased gene conversion [57,62] and homoeologous exchange [25], or completely excised from the genome via pseudogenization and gene loss [58].

Investigations into the predicted outcomes of cytonuclear incompatibilities in allopolyploids have so far had mixed results. Rubisco in particular has been a primary focus, as the nuclear-encoded small subunit *rbcS* appears to have undergone maternally biased gene conversion and maternally biased gene expression in some allopolyploids, such as cotton, tobacco, *Arabidopsis suecica,* peanut, and wheat [56,57,62]. Synthetic and recently formed allopolyploids show more inconsistent support. For example, *Tragopogon miscellus* exhibits maternally biased expression of *rbcS*, while its reciprocally formed congener *Tragopogon mirus* does not [58]. Synthetic allotetraploid rice showed little evidence of maternally biased expression of *rbcS [59]*, and synthetic allopolyploid *Cucumis x hytivus* displayed paternally biased expression of *rbcS* [61]. Generalizing rules of cytonuclear biology from these handful of somewhat contradictory studies is made even more difficult because they have all only considered a single cytonuclear complex.

A more extensive survey of 110 nuclear genes encoding subunits involved in plastid protein complexes in allopolyploid *Brassica napus* did not find evidence for maternally biased expression or retention of organelle-targeted genes [60]. What remains to be evaluated is whether there are systematic rules that might explain the discrepancies among these earlier studies and more generally what the principles are that govern cytonuclear evolution in plant allopolyploids. There are as yet no genome-wide investigations of the signatures of cytonuclear incompatibilities in a set of independently formed allopolyploids that differ both in terms of the amount of divergence between diploid progenitors (and therefore the probability of cytonuclear incompatibilities [100]), or time since allopolyploidization (and therefore the probability of an evolutionary response to cytonuclear incompatibilities [101]). The rapidly increasing availability of genome sequences for a number of allopolyploid genomes and their diploid relatives (e.g., *Brassica* [8,19,102,103], cotton [104–107], wheat [108–112], peanut [23,113], coffee [114–117], tobacco [118,119], quinoa [21,120], and *Brachypodium* [121,122]) makes it possible to better understand the rules of cytonuclear biology in allopolyploid lineages.

Here we evaluate genome-wide patterns of molecular evolution in organelle-targeted gene sets for six separate allotetraploid species: *Brachypodium*

hybridum, Chenopodium quinoa (quinoa), *Coffea arabica* (coffee), *Gossypium hirsutum* (cotton), *Nicotiana tabacum* (tobacco), and *Triticum dicoccoides* (wild emmer wheat). We document strong effects of subgenome on overall rates and patterns of evolution but find little evidence for global signatures of cytonuclear incompatibilities across polyploid systems. We also find that organelle-targeted gene content is generally less biased across subgenomes than the rest of the genome. Together, these genome-wide analyses of six independently formed allotetraploid species provide insights into the rules of polyploidy, a prominent process in eukaryotic diversification.

RESULTS

Orthologous genes in six allopolyploid species and their diploid relatives

To compare rates and patterns of molecular evolution across subgenomes of six allotetraploid angiosperms (Figure 1a), we inferred orthologous gene groups from the two polyploid subgenomes, the closest available diploid species for each subgenome, and an outgroup (Figure 2) using a combination of phylogenetic and syntenic methods. The resulting orthologous gene groups are summarized in Table 1, and additional details regarding their inference are provided in the Materials & Methods section as well as in Figure S1.

Table 1. Orthologous gene groups in six allotetraploid angiosperms.

Species	Phylogenetic orthologous gene groups	Syntenic orthologous gene groups	Merged, single-copy quintets[1] (Union/Intersect)	Filtered, merged, single-copy quintets (Union/Intersect)
Quinoa	10511	17896	(9088 / 3284)	(8201 / 3121)
Wheat	25454	24212	(10180 / 3603)	(4204 / 1759)
Cotton	29504	31841	(18715 / 10222)	(17133 / 10023)
Coffee	19399	20926	(6672 / 3869)	(4032 / 2379)
Tobacco	24797	32088	(9059 / 167)	(8699 / 163)
Brachypodium	24854	34440	(14449 / 8084)	(14000 / 7912)

[1] – Single-copy quintets include orthologous gene groups with one and only one sequence from an outgroup, two closely related diploids, and two sequences from the allopolyploid.

The goal of our orthology inference methods was to produce orthologous “quintets”, containing one gene sequence each from the outgroup species and the two diploid model species and two gene sequences from the polyploid species, while also requiring that gene trees be consistent with the overall species tree. Both syntenic and phylogenetic methods produced sizable numbers of identical quintets; however, there were many quintets only detectable using one method or the other. Tobacco was especially challenging for syntenic inference, as the highly fragmented assemblies of all three *Nicotiana* reference genomes made identifying syntenic blocks difficult. The largest syntenic block between any two of the genomes in this clade was only 57 genes long (*N. tabacum* and *Solanum lycopersicum*), and no syntenic block including *N. tomentosiformis* or *N. sylvestris* was longer than 22 genes. Quinoa highlighted a different issue that represents a common feature of polyploid genome assemblies in that many genes were located on contigs that are not anchored to chromosomes. Genes present in this fraction of the assembly can only be included in orthologous groups by phylogenetics, and they are often replete with repetitive elements, making it a likely spot for genome misassemblies (and subsequent errors in analyses that depend upon them). Moreover, the quinoa genome contains cases of apparent homoeologous

exchange in which genes were located on chromosomes from opposing subgenomes (see also [21]).

Variation in assembly and annotation quality also represented a significant challenge in identifying orthologous genes across genome assemblies produced by different groups with different underlying data. The most extreme example of this issue was the maternal diploid model for polyploid wheat, *Aegilops speltoides*, which was represented only by a transcriptome assembly. Despite these and other hurdles, we were able to identify orthologous gene groups as well as the more strict group of single-copy quintets for each of these polyploid systems, which should present a useful resource for polyploid genomics moving forward. The *Aegilops speltoides* transcriptome, all OrthoFinder results, phylogenetic gene trees with branch lengths, multi-species synteny networks, merged orthologous gene groups, CDS alignments, and analyses of molecular evolution have been made available at https://doi.org/10.6084/m9.figshare.13473207. For the remainder of the manuscript, we report only on the results from the “Union” group of quintets that were identified by either phylogenetic or syntenic inference, but we have performed all the same analyses on the “Intersection” group, comprised only of those quintets that were identified by both methods, and have provided the results from those analyses in Supplementary File 1. Results obtained using the Intersection dataset did not substantively differ from those obtained using the Union dataset.

Subgenomic distributions of organelle-targeted genes

To evaluate whether cytonuclear interactions affect subgenomic evolution in allopolyploid species, we first partitioned genes by predicted subcellular targeting localization and cytonuclear interaction activity in each allopolyploid system. The Cytonuclear Molecular Interactions Reference for *Arabidopsis* (CyMIRA) database indicates that the *Arabidopsis thaliana* nuclear genome has 1773 genes that encode mitochondria-targeted products and 2931 genes that encode plastid-targeted products [67]. By propagating this classification across the six allotetraploids studied here, we found means of 3880 (SD = 730) genes with mitochondria-targeted products and 4464 (SD = 731) genes with plastid-targeted products (Table 2), which varies ~60-70% among allotetraploid taxa. At least some of the observed variation among polyploids appears to be due to phylogeny, as the number of mitochondria-targeted genes and plastid-targeted genes varies extensively among diploids (25-30%, Figure S2). Diploid relatives of our focal allotetraploids ranged from 17% fewer (*Chenopodium* diploids) to 108% more (*Nicotiana* diploids) mitochondria-targeted genes and from 37% fewer (*Triticum*, *Chenopodium* diploids) to 33% more (*Nicotiana* diploids) plastid-targeted genes than documented in *Arabidopsis* (Figure S2).

Table 2. Functional gene partitioning in six allotetraploid angiosperms.

Species	Mitochondria-targeted	Mitochondria-targeted Interacting[1]	Mitochondria Enzyme Complexes[2]	Plastid-targeted	Plastid-targeted Interacting[1]	Plastid Enzyme Complexes[2]
Quinoa	2830	894	279	3528	686	215
Wheat	4077	1048	378	4419	693	245
Cotton	4728	1232	458	5670	800	307
Coffee	3221	921	285	3889	621	193
Tobacco	3851	1092	402	4567	740	297
Brachypodium	4540	981	339	4684	674	238
Mean (SD)	**3880 (730)**	**1031 (121)**	**358 (68)**	**4464 (731)**	**704 (61)**	**250 (45)**
***Arabidopsis thaliana* (diploid)**	**1773**	**617**	**180**	**2931**	**375**	**128**

[1] – Mitochondria- and plastid-targeted interacting genes are a subset of the total number of mitochondria- and plastid-targeted genes

[2] – Mitochondria and plastid enzyme complex genes are a subset of the total number of mitochondria- and plastid-targeted interacting genes

Among the genes with mitochondria-targeted products, CyMIRA lists 617 *A. thaliana* genes that have interactions with mitochondrial genes or gene products and 180 genes with products that are directly involved in enzyme complexes with mitochondrially encoded subunits (i.e., mitoribosome, OXPHOS complexes, TAT complex). We expected to find roughly twice as many genes in each functional category for tetraploids as are present in *Arabidopsis*. In the six focal allotetraploids, we found that functional categories were increased 40-250% (per category/species) relative to *A. thaliana*, with means of 1031 (SD = 121) genes having interactions with mitochondrial genes or gene products and 358 (SD = 68) genes with products that are directly involved in those three mitochondrial enzyme complexes. A similar pattern was observed for genes with plastid-targeted products. Where CyMIRA lists 375 *A. thaliana* genes that have interactions with plastid genes or gene products and 128 genes with products that are directly involved in enzyme complexes with plastid-encoded subunits (i.e., chlororibosome, Photosystems I and II, NDH, ATP synthase, Cytochrome b6f, Rubisco, Clp protease, ACCase), we found means of 704 (SD = 61) and 250 (SD = 45) genes in the allotetraploids for those categories, respectively. Gene numbers for all 55 functional gene categories are described in Table S1, gene IDs for each category and *de novo* targeting predictions are available at https://github.com/jsharbrough/CyMIRA_gene_classification/tree/master/Species_CyMIRA, and the physical distribution of organelle-targeted genes along polyploid chromosomes are shown in Figure S3.

Polyploidization events are expected to perturb cytonuclear interactions in part because cytoplasmic genomes suddenly exist inside a cell in which all of their nuclear-encoded interacting partners have been doubled. One possible evolutionary

response to altered cytonuclear stoichiometry in the wake of whole genome duplication is that nuclear-encoded organelle-targeted genes experience selection to rapidly return to a diploid-like state [123,124]. We tested this hypothesis for both mitochondria- and plastid-targeted nuclear genes in six independently formed allopolyploids using the combined diploid relatives as models for the ancestral allopolyploid state. We found that quinoa ($\chi^2 = 54.40$, $p < 0.0001$), wheat ($\chi^2 = 660.23$, $p < 0.0001$), tobacco ($\chi^2 = 243.85$, $p < 0.0001$), and *Brachypodium* ($\chi^2 = 50.15$, $p < 0.0001$) retain a significantly smaller proportion of organelle-targeted genes in duplicate than non-organelle-targeted genes, whereas, cotton ($\chi^2 = 134.12$, $p < 0.0001$) and coffee ($\chi^2 = 13.40$, $p = 0.00025$) exhibit the opposite pattern by retaining a significantly larger proportion of organelle-targeted genes than non-organelle-targeted genes (Table S2). Notably, the excess retention of organelle-targeted genes in cotton was also evident when we restricted our analysis to only include the subset of genes directly involved in mitochondrial ($\chi^2 = 7.90$, $p = 0.0049$) or plastid enzyme complexes ($\chi^2 = 5.58$, $p = 0.018$). Although levels of retention within each category varied among species, we did not find a difference in retention levels between mitochondria-targeted versus plastid targeted genes in any of the six species ($p > 0.05$ for all species). Wheat ($\chi^2 = 18.35$, $p < 0.0001$) and cotton ($\chi^2 = 11.05$, $p = 0.00089$) both exhibited significantly more pentatricopeptide repeat (PPR) genes (relative to non-organelle-targeted genes) compared to the combined diploids, while the tobacco genome encoded significantly fewer PPR genes than expected (relative to non-organelle-targeted genes) compared to the combined diploids ($\chi^2 = 68.09$, $p < 0.0001$). Together, these results provide mixed support for rapid loss of organelle-targeted genes compared to the rest of the genome in allopolyploids, but do indicate that similar forces may equally affect mitochondria- and plastid-targeted genes.

A second possible consequence of polyploidy is incompatibility between the paternally derived nuclear subgenome and the maternally derived cytoplasmic genomes, potentially resulting in preferential loss of paternally-derived organelle-targeted genes in hybrid (allo)polyploid species. This effect could exaggerate a general subgenome bias for paternal loss or partially compensate for a general bias towards maternal loss. For five of the allotetraploid genomes, it was possible to assign genes to subgenomes based on their chromosomal position, thereby permitting a relative assessment of parental gene loss; the sole exception, *N. tabacum*, has experienced extensive genomic rearrangement since polyploidization that precludes subgenome assignment based on physical location. In general, we found significant differences in non-organelle-targeted gene abundance across subgenomes for all five allotetraploid species (Table 3), with quinoa, wheat, and coffee exhibiting more paternal homoeolog loss, whereas cotton and *Brachypodium* exhibit a deficit in maternal homoeologs (Figure 3, left panel). Interestingly, however, when considering biases in organelle-targeted genes after correcting for genome-wide levels, these biases flip for quinoa, wheat, and *Brachypodium*. That is, while both quinoa and wheat exhibit biased

loss of *paternal* homoeologs for non-organellar targeting genes, those that are targeted to the organelles exhibit biased *maternal* loss (again, relative to background; Figure 3 right panels, Table S3). Similarly, *Brachypodium* organelle-targeted genes exhibit biased paternal loss (relative to background), whereas the genome-wide pattern shows biased maternal loss. These patterns were also found using the diploid relatives to correct for different gene abundances at the time of allopolyploidization (Figure S4). While the maternal homoeolog deficit for organelle-targeted genes found in wheat and quinoa is contrary to predictions based on cytonuclear incompatibilities, we note that this reflects homoeolog retention relative to the genome-wide rate and suggests that these species exhibit a lower degree of subgenomic bias in their organelle-targeted genes than the genome-wide rate.

Table 3. Biased gene content of non-organelle-targeted genes across subgenomes of five allotetraploid angiosperms.

Species	Paternal subgenome	Maternal subgenome	r_{PAT} - r_{MAT} (95% CI)	Binomial test *p* value
Quinoa	9786	11053	-0.061 (-0.074 – -0.047)	< 0.0001
Wheat	48786	52571	-0.037 (-0.044 – -0.031)	< 0.0001
Cotton	29762	28871	0.015 (0.007 – 0.023)	0.00024
Coffee	19008	19773	-0.020 (-0.030 – -0.010)	0.00011
Brachypodium	34860	29605	0.082 (0.074 – 0.089)	< 0.0001

Evolutionary rate differences across subgenomes and gene functional categories

We used the CyMIRA gene classifications from the maternal diploid models of each allotetraploid to classify single-copy orthologous quintets into functional gene categories, except in the case of wheat. For wheat, the paternal diploid model, *Triticum urartu*, was used because the maternal diploid model (i.e., *Aegilops speltoides*) is only represented by a transcriptome. These functional categories served as the basis for our concatenated and gene-level analyses of evolutionary rate. Summary statistics describing the number of orthologous quintets in each functional category are presented for each allopolyploid system in Table 4 and Figure S5, along with the rates of synonymous (d_S) and nonsynonymous (d_N) evolution in concatenated alignments.

Table 4. Single-copy orthologous quintets partitioned by functional category in six allotetraploid species.

Species	Functional Category	Number of Quintets	d_S	d_N	d_N/d_S
Quinoa	Not-organelle-targeted	6885	0.499	0.096	0.193
	Mitochondria-targeted Non-interacting	615	0.444	0.079	0.179
	Mitochondria-targeted Interacting	213	0.477	0.105	0.220
	Mitochondria Enzyme Complexes	69	0.465	0.084	0.180
	Plastid-targeted Non-interacting	900	0.449	0.081	0.180
	Plastid-targeted Interacting	212	0.463	0.091	0.197
	Plastid Enzyme Complexes	74	0.483	0.081	0.168
Wheat	Not-organelle-targeted	3507	0.1882	0.035	0.187
	Mitochondria-targeted Non-interacting	476	0.179	0.030	0.169
	Mitochondria-targeted Interacting	67	0.162	0.033	0.206
	Mitochondria Enzyme Complexes	38	0.191	0.039	0.206
	Plastid-targeted Non-interacting	561	0.179	0.031	0.171
	Plastid-targeted Interacting	86	0.171	0.030	0.175
	Plastid Enzyme Complexes	38	0.228	0.030	0.131
Cotton	Not-organelle-targeted	14957	0.108	0.038	0.348
	Mitochondria-targeted Non-interacting	1076	0.106	0.033	0.309
	Mitochondria-targeted Interacting	375	0.103	0.034	0.332
	Mitochondria Enzyme Complexes	100	0.119	0.037	0.310
	Plastid-targeted Non-interacting	1502	0.106	0.033	0.309
	Plastid-targeted Interacting	270	0.102	0.031	0.303
	Plastid Enzyme Complexes	94	0.100	0.029	0.289
Coffee	Not-organelle-targeted	3397	0.181	0.055	0.306
	Mitochondria-targeted Non-interacting	306	0.181	0.051	0.281
	Mitochondria-targeted Interacting	121	0.170	0.052	0.306
	Mitochondria Enzyme Complexes	31	0.187	0.057	0.307
	Plastid-targeted Non-interacting	420	0.180	0.051	0.285
	Plastid-targeted Interacting	88	0.163	0.049	0.300
	Plastid Enzyme Complexes	25	0.159	0.043	0.273
Tobacco	Not-organelle-targeted	7323	0.438	0.090	0.205
	Mitochondria-targeted Non-interacting	675	0.375	0.071	0.190
	Mitochondria-targeted Interacting	209	0.374	0.082	0.220
	Mitochondria Enzyme Complexes	59	0.392	0.070	0.178
	Plastid-targeted Non-interacting	952	0.380	0.072	0.191
	Plastid-targeted Interacting	183	0.370	0.074	0.200
	Plastid Enzyme Complexes	72	0.406	0.070	0.173
Brachypodium	Not-organelle-targeted	11886	0.449	0.1052	0.234
	Mitochondria-targeted Non-interacting	1310	0.388	0.0759	0.196
	Mitochondria-targeted Interacting	367	0.398	0.086	0.216
	Mitochondria Enzyme Complexes	116	0.399	0.0645	0.162
	Plastid-targeted Non-interacting	1497	0.389	0.0763	0.196
	Plastid-targeted Interacting	256	0.396	0.0829	0.209
	Plastid Enzyme Complexes	83	0.485	0.0626	0.129

Rates of protein-sequence evolution vary substantially across CyMIRA functional categories, likely indicative of variation in functional constraint (Figure S5a). In particular, protein sequences of mitochondrial OXPHOS complexes, several of the

plastid photosynthesis complexes (but not all, see e.g., the NADH dehydrogenase-like [NDH] complex), as well as the mitochondrial and plastid RNA polymerases appear to evolve especially slowly, indicating that they have experienced relatively stringent purifying selection in these angiosperms. In addition to complex-level effects, we also observed differences in protein-sequence evolution across our focal angiosperm systems, with coffee and cotton genomes exhibiting higher quintet-wide d_N/d_S values compared to quinoa, wheat, tobacco, and *Brachypodium* (Figure S5b).

Cytonuclear incompatibilities between maternally derived cytoplasmic genomes and the paternal subgenomes of allopolyploids are expected to result in accelerated rates of protein-sequence evolution in the paternal homoeologs of organelle-targeted genes. We tested for signatures of these cytonuclear incompatibilities first by estimating differences in rates of protein-sequence evolution (i.e., $d_N/d_S = \omega$) in concatenated and individual gene alignments of paternal (ω_{PAT}) vs. maternal (ω_{MAT}) subgenomes in non-organelle-targeted (NOT) genes to assess whether genome-wide biases exist in our six focal allopolyploids. In concatenated analyses, quinoa, wheat, cotton, and tobacco all showed significant departures (i.e., < 2.5% overlap of bootstrap distributions between ω_{PAT} and ω_{MAT}) from equal rates of evolution across subgenomes. In particular, quinoa, cotton, and tobacco exhibited higher ω values in maternally derived homoeologs of NOT genes than paternal homoeologs (i.e., ω_{PAT}: ω_{MAT} ratio < 1), while coffee and wheat showed the opposite pattern in which paternally derived homoeologs exhibit faster rates of protein-sequence evolution than maternal homoeologs (i.e., ω_{PAT}: ω_{MAT} ratio > 1; Figure 4a). We observed similar patterns in gene-level analyses as compared to concatenated analyses in the three older polyploids (Figure 4b): a significantly higher proportion of maternal homoeologs (p_{MAT}) exhibited faster rates of evolution than paternal homoeologs (p_{PAT}) in quinoa (binomial test, $p = 0.0022$) and cotton (binomial test, $p < 0.0001$), while p_{PAT} was significantly greater than p_{MAT} in wheat (binomial test, $p < 0.0001$). Although p_{MAT} was greater than p_{PAT} in the concatenated analysis of tobacco subgenomes, the difference was not significant at the gene level (binomial test, $p = 0.183$). A similar result was obtained in coffee, with the concatenated analysis showing a significant paternal bias, but gene-level patterns did not appear to be paternally biased (binomial test, $p = 0.375$). Bootstrap distributions of ω_{MAT} in *Brachypodium* estimated from concatenated alignments were higher than bootstrap distributions of ω_{PAT} but were not significantly different (i.e., > 2.5% overlap), while p_{MAT} was significantly greater than p_{PAT} at the individual gene level (binomial test, $p = 0.00026$). The higher ω values in the maternal subgenomes of quinoa, cotton, and *Brachypodium* and the higher ω values in the paternal subgenome of coffee were primarily driven by differences in d_N as opposed to d_S (Figure 2), indicating that these subgenomes experience different rates of protein-sequence evolution. By contrast, the elevated ω values in the maternal subgenome of tobacco and the paternal subgenome of wheat were primarily driven by d_S (Figure 2), potentially indicating that different

subgenomes experience different mutation rates or that the diploids used here represent highly asymmetric models of the diploid progenitors. Taken together, these analyses of NOT genes indicate that allopolyploids experience significant biases in rates of evolution across subgenomes present inside the same cell.

We next performed concatenated and gene-level analyses of ω_{PAT} and ω_{MAT} in organelle-targeted genes (normalized by NOT genes) to test whether paternal homoeologs exhibited faster rates of protein-sequence evolution than maternal homoeologs, as predicted if paternal subgenomes harbor incompatibilities with the cytoplasmic genomes. We found evidence that concatenations of wheat genes involved in mitochondrial enzyme complexes exhibited significantly higher ω_{PAT} values (median = 0.661, 95% CI = 0.268 – 0.807) compared to ω_{MAT} values (median = 0.0771, 95% CI = 0.0460 – 0.125), relative to NOT genes (ω_{PAT} = 0.444, 95% CI = 0.414 – 0.476; ω_{MAT} = 0.201, 95% CI = 0.189 – 0.215); however, no other species or functional classes exhibited the predicted pattern (Figure 5). To further investigate the patterns of molecular evolution the wheat mitochondrial enzyme complex genes, we manually inspected and trimmed concatenated alignments from NOT genes, mitochondrial enzyme complex genes, and plastid enzyme complex genes and re-inferred ω_{PAT} and ω_{MAT} in all three gene categories. Importantly, we found two small regions from two genes in the mitochondrial enzyme complexes that were poorly aligned only in the paternal subgenome, contributing to elevated ω_{PAT} but not ω_{MAT}. The poorly aligned regions appeared to be caused by a combination of an apparent frameshift in the paternal homoeologs of one gene encoding a protein involved in the NADH Dehydrogenase (OXPHOS Complex I – TRIDC1AG048530) and another gene encoding a protein that functions in the large subunit of the mitoribosome (TRIDC4AG029590) had an exon on the 3’ end of the gene with no apparent homology to the other sequences in the quintet (likely due to misannotation or misassembly, as the new *T. turgidum* assembly, GCA_900231445.1, does not have this same issue). Both genes exhibited substantially different d_S and d_N values compared to other genes in the same functional gene category (Table S4). Trimming the poorly aligned regions resulted in a substantially lower d_N value for concatenated alignments of mitochondrial enzyme complex genes, which in turn caused a lower ω_{PAT} value that was not significantly different from the ω_{MAT} value (Figure S6). All trimmed alignments and analyses are available at https://github.com/jsharbrough/allopolyploidCytonuclearEvolutionaryRate. For gene-level analyses, we did not find any functional categories in any species that exhibited significantly different normalized proportions of genes with higher ω_{PAT} or ω_{MAT} (Figure S7), a pattern which did not change when d_N was used in place of ω. Thus, there do not appear to be global accelerations in protein-sequence evolutionary rate of paternal homoeologs of organelle-targeted genes in the wake of allopolyploidization.

We next evaluated ω values at the level of specific cytonuclear interactions (Table S5) and found scattered patterns of both paternal and maternal bias across various cytonuclear interactions in the three older polyploids (i.e., quinoa, wheat, and cotton). In particular, paternal homoeologs of quinoa exhibited significantly higher ω values (i.e., ω values from concatenated alignments +/- 1 SE were outside bootstrap-constructed 95% confidence intervals of NOT genes) than maternal homoeologs in mitochondrial tRNA base modification, plastid NDH, and plastid tRNA base modification, and maternal homoeologs exhibited significantly higher ω values than paternal homoeologs in both subunits of the chlororibosome and Photosystem I (PSI). As seen at higher levels of organization, wheat mitochondrial enzyme complexes generally exhibited higher ω values in paternal vs. maternal homoeologs (see below for detailed discussion) compared to NOT genes. However, the reverse was true in plastid enzyme complexes, with plastid PSII exhibiting significantly higher ω values in maternal vs. paternal homoeologs, relative to NOT genes. Wheat organellar tRNA aminoacyl synthetases, which are largely dual-targeted [125], also exhibited significant maternal bias compared to NOT genes. Cotton had fewer CyMIRA categories that showed evidence of bias over-and-above genome-wide levels, with just the mitochondria- and plastid-targeted recombination, replication, and repair (RRR) genes (also commonly dual-targeted [67]) exhibiting elevated ω values in paternal vs. maternal homoeologs and the large subunit of the mitoribosome and mitochondria-targeted PPR genes exhibiting higher ω values in maternal vs. paternal homoeologs compared to NOT genes. Coffee, tobacco, and *Brachypodium* all appear to be too young for this analysis, as only a single functional category (plastid transcription and transcript maturation) in coffee showed significant (maternal) bias compared to NOT genes, despite genome-wide bias in ω values of coffee and tobacco. There were no CyMIRA categories that exhibited consistent patterns across even the older three allopolyploids, highlighting the highly context-specific nature of evolutionary dynamics of cytonuclear interactions in allopolyploids.

Because incompatibilities are only likely to arise in genes that are divergent at the time of allopolyploidization, we also performed the analyses described above on high and low divergence gene bins. To do so we split single-copy orthologous quintets into two groups: those with high amino acid sequence divergence between diploid models (measured by d_N) and those with low amino acid sequence divergence. We used a similar approach as before to normalize ω_{PAT} and ω_{MAT} using the NOT genes. There were only two cases in which high and low divergence classes differed by more than one standard error: mitochondrial and plastid enzyme complexes of wheat (Figure S8). In particular, the low divergence class of wheat mitochondrial enzyme complexes (MTEC) exhibited more extreme paternal bias than the high divergence class, while the low-divergence class of wheat plastid enzyme complexes (PTEC) exhibited a more extreme maternal bias compared to the high-divergence class. This somewhat

surprising result notwithstanding, the lack of signal in the high divergence classes across the other functional categories and species indicates that cytonuclear incompatibilities of allopolyploids are not resolved by faster rates of protein-sequence evolution in paternal homoeologs.

We compared patterns of autapomorphic amino acid changing mutations at sites that were conserved throughout the rest of the quintet in genes encoding subunits of mitochondrial enzyme complexes. For each species, we observed several gene functional categories that exhibited an excess number of autapomorphic amino acid changes compared to genes not targeted to the mitochondria or plastids in one subgenome compared to the other. However, the direction of excess was not consistent across species or even across functional gene categories (Table S6).

Because derived amino acids with substantially different biochemical properties compared to ancestral residues (i.e., radical amino acid changes) are especially likely to alter protein structure and function [126–131], we next restricted these analyses of derived amino acid changes in the tetraploids to only include radical amino acid changes (as defined by the Conservative/Radical Index CRI [132]). As was the case with total derived amino acid changes, there existed several functional categories in each species that exhibited significant biases in the accumulation of radical autapomorphies across subgenomes, but the direction of bias and the functional categories identified were not consistent across species. Several notable functional categories did exhibit bias across multiple species though (e.g., DNA replication, recombination, and repair genes [quinoa, cotton, *Brachypodium*], tRNA base modification genes [quinoa, cotton, coffee, *Brachypodium*], and tRNA aminoacyl synthetases [wheat, tobacco]), potentially indicating they are hotbeds for cytonuclear incompatibilities and/or diploidization. Together these results indicate that cytonuclear enzymes exhibit complex- and species-specific patterns of accumulation of derived amino acids at conserved sites.

In sum, our concatenated, gene-level, and site level analyses provide evidence that protein sequences of different allopolyploid subgenomes exhibit different ω values, potentially as a result of different rates of protein-sequence evolution, but cytonuclear incompatibilities resulting from the allopolyploidization event do not leave global signatures of accelerated protein sequence evolution in paternal homoeologs of organelle-targeted genes. Moreover, while organelle-targeted genes are often lost at higher rates than genome-wide rates of diploidization, this is not always the case, especially in cotton, and biased gene content of allopolyploid subgenomes does not appear to be related to cytonuclear incompatibilities. Rather, only species- and complex-specific cytonuclear dynamics appear to alter rates of evolution in organelle-targeted genes, and in directions not uniformly consistent with allopolyploidy induced cytonuclear incompatibilities.

DISCUSSION

We inferred orthologous gene sets, partitioned genes by subcellular targeting localization and cytonuclear interaction, and evaluated genome-wide patterns of gene content and natural selection across subgenomes of six allotetraploid angiosperms. We report significant genome-wide biases across maternal vs. paternal subgenomes in overall gene content in all five allopolyploids tested and in mutation-rate-corrected rates of protein-sequence evolution (i.e., ω) in all six allopolyploid genomes tested. The directions of bias in both gene content and higher ω were not consistent across independent allopolyploidization events, and the patterns observed in gene content did not appear to be similar in direction as bias in ω.

The analyses reported here support three primary conclusions: (1) allopolyploid subgenomes exhibit substantially different rates of protein-sequence evolution from one another despite existing inside the same cell for thousands to millions of years; (2) cytonuclear incompatibilities between the cytoplasmic genomes and the paternal subgenome are complex and taxon-specific and do not result in global increases in rates of protein-sequence evolution in paternal homoeologs of organelle-targeted genes; and (3) gene content is not equally distributed across subgenomes, with both species and cytonuclear functional classes contributing to variation in the rate at which genomes fractionate following WGDs. The foregoing conclusions suggest a number of questions that have implications for our understanding of polyploid biology.

Differential rates of protein-sequence evolution across allopolyploid subgenomes

Most prominent among our data are the remarkable differences in evolutionary patterns across subgenomes, raising the question of what evolutionary forces underlie these subgenomic biases? That is, allopolyploid subgenomes that have been (co-)evolving inside the same nucleus for thousands to millions of years [133], remain on separate evolutionary trajectories with respect to evolutionary rates in protein-coding genes. Here, we consider several phenomena that could play a role in establishing and maintaining subgenomic biases.

If ω is adequately inferring patterns of natural selection across subgenomes (but see below for alternative explanations), then the patterns of subgenomic biases in rates of protein-sequence evolution reported here could arise from differences in the efficacy of selection or effective population size (N_e) across subgenomes. In particular, genes that are more highly expressed [134,135], have higher local recombination rates [136–139], or lower local TE densities [140–142] (but see [143]), are expected to experience increased efficacies of natural selection and thus exhibit reduced rates of protein-sequence evolution [144]. That is, genome-wide differences between the progenitors at the time of allopolyploid formation (e.g., transcriptome size,

recombination rate, TE load) would not only be expected to give rise to subgenomic differences in the immediate aftermath of polyploidization [101,145–148], but also contribute to evolved differences across subgenomes [10,11,149–153].

Mutation rate varies tremendously across species, populations, individuals, and even within genomes [154–157], making it a potential candidate for generating subgenome biases in ω, if elevated mutation rate results in increased rates of background selection, thereby reducing N_e [144]. Such mutational biases could reflect ancestral differences in parental species (e.g., differences in DNA methylation [157]), or could potentially arise after polyploidization in association with other biased phenomena such as recombination [158], gene expression [8,36,38,159–166], epigenetic marks [26–33], or transposable element activity [14–17], which are all thought to themselves be mutagenic [167–170].

Subgenomes might also differ in N_e as a result of backcrossing, in which one polyploid subgenome experiences higher rates of introgression than the other [171–173]. Repeated allopolyploid formation or gene flow from diploids (e.g., *Brachypodium hybridum* – [122], *Arabidopsis suecica* – [174]) can cause N_e to differ across subgenomes. Finally, recombination could also act to bias inferences of ω artifactually because genetic material exchanged across subgenomes via homoeologous exchange [18–25,175–183], gene conversion [19,20,57,62,133,153,184–189], and other recombinational mechanisms (e.g., [190]) would be expected to bias ω inferred across a topologically constrained tree. However, we took steps to prevent this type of artifact from influencing our data by only including genes that exhibited gene-tree topologies that were consistent with the species tree topology.

The relative contributions of these various evolutionary dynamics are of central importance to the understanding of polyploid genomes, but testing each hypothesis in turn is made difficult by the fact that the sampled diploids are, to varying degrees, imprecise models of the ancestral progenitors. Therefore, an unknown fraction of each terminal “polyploid” branch in our quintet trees actually represents evolution as a diploid prior to hybridization. Wheat in particular is susceptible to artifactual inflation of ω because *Aegilops speltoides* is so much more distantly related to the B subgenome of the polyploid than *Triticum urartu* is to the A subgenome (Figure 2). The persistence of slightly deleterious changes since the divergence of the A subgenome and the diploid A genome may result in overestimates of ω in the A subgenome compared to the B subgenome. The same logic applies to all of our polyploid taxa to varying extents; however, it is worth noting that while differences in d_S across subgenomes were the primary drivers of differences in ω in wheat and tobacco, d_N had a proportionally larger effect than d_S on differences in ω in quinoa, cotton, coffee, and *Brachypodium*. This latter finding is consistent with selection being the driving factor in variation in evolutionary rates across subgenomes (but see prior caveat regarding quality of diploid

models and evolution prior to polyploidization), rather than mutation rate variation or artifactual inflation of ω in the more closely related diploid-subgenome pair. In the same vein, coffee and cotton, which are both thought to have extremely small effective population sizes [191], exhibited the highest overall ω values (Figure S5b). All told, investigating site frequency spectra, gene expression profiling, and recombination rates within populations and their relationships to the biased ω values reported here will help resolve these outstanding questions.

No global signature of mitonuclear incompatibilities in paternal homoeologs of allopolyploid genomes

To test the hypothesis that incompatibilities stemming from evolutionary mismatches between the maternally derived cytoplasmic genomes and the paternally derived nuclear subgenome result in preferential loss and accelerated rates of protein-sequence evolution in paternal homoeologs of organelle-targeted genes, we applied the same analyses described above to sets of CyMIRA-partitioned genes, after accounting for genome-wide effects. We did not discover evidence that cytonuclear incompatibilities shape either gene content or protein-sequence evolution in paternal homoeologs of organelle-targeted genes, despite multiple distinct tests of this hypothesis. In particular, patterns of gene content on organelle-targeted genes exhibited the opposite pattern as that observed in NOT genes in three of five allopolyploid taxa (the remaining two were not significantly different from genome-wide patterns), indicating that organelle-targeted genes tend to exhibit greater balance across subgenomes than the rest of the genome. While the proportion of organelle-targeted genes per subgenome did not appear to be especially maternally biased, four of six allotetraploids had reduced overall proportions of organelle-targeted genes compared to NOT genes. Overall, rates of protein-sequence evolution in organelle-targeted and interacting genes generally reflected the genome-wide patterns of bias observed in NOT genes, rather than rate accelerations peculiar to paternal but not maternal homoeologs.

One outstanding question stemming from our analyses of protein-sequence evolution in paternal vs. maternal homoeologs of organelle-targeted genes is why hybrid polyploid genomes appear to generally lack genome-wide signatures of cytonuclear incompatibilities, despite their apparent importance in homoploid hybridization [95] and introgression events [92]? It is possible that cytonuclear incompatibilities do leave signatures on genomes, but not in terms of accelerated rates of protein-sequence evolution in paternal homoeologs. For example, pseudogenization may be a rapid and common mechanism for adaptation in plant genomes (e.g., [192]), which would be missed by our quintet analyses. While we did not observe maternally biased gene content in CyMIRA datasets, direct analysis of bias in homoeologous pairs in which one copy is pseudogenized is necessary to rule out gene loss as a mechanism by which

cytonuclear incompatibilities are resolved. The seemingly stochastic patterns of homoeolog bias in accumulation of autapomorphic amino acid changes indicates that there are often cases in which homoeologs of cytonuclear interacting genes evolve very differently, perhaps reflecting cytonuclear incompatibilities or the precursor to gene loss and diploidization, but these biases do not appear to coincide with the allopolyploidization events in any systematic way. The presence of biased accumulation of autapomorphies in *Brachypodium* may indicate that cytonuclear incompatibilities are resolved rapidly. Cytonuclear incompatibilities may also be resolved via biased homoeolog expression [162], gene conversion [57,62] or homoeologous exchange [25], subfunctionalization of subcellular localization by differential isoform usage across homoeologs [193], or other potential mechanisms that would not generate global signatures of paternal acceleration in coding sequences of organelle-targeted quintets.

Biased homoeolog expression represents a potential mechanism by which allopolyploids could resolve cytonuclear incompatibilities, but has found mixed support in the studies that have so far attempted it. In particular, cotton, tobacco, *Arabidopsis*, peanut, and the extremely young allotetraploid *Tragopogon miscellus* exhibit biased maternal expression of the nuclear-encoded subunit of Rubisco [56–58], but others have not found similar patterns in rice [59] or *Brassica napus* [60]. Moving forward, large-scale genome-wide homoeolog expression bias could be evaluated across all the CyMIRA gene sets (not just Rubisco) to test this hypothesis. Additionally, the topological and alignment filtering steps we imposed on quintets here had the intended side effect of filtering out genes exhibiting gene conversion or homoeologous exchange. Notable among them was *rbcS*, which encodes the small subunit of Rubisco and was missing from filtered, single-copy quintets in five of six species complexes (present only in *Brachypodium*, the youngest allopolyploid). It is likely that because of *rbcS*' propensity for gene conversion [57], this apparent "hotbed" for cytonuclear incompatibilities might provide additional evidence that was missed here. Certainly, a careful analysis of maternal vs. paternal bias in gene conversion tracts and homoeologous exchanges among organelle-targeted genes may be a fruitful future approach.

An additional and perhaps likely possibility is that cytoplasmic genomes of these allopolyploids may evolve too slowly in protein-coding sequence to generate widespread incompatibilities in hybrid polyploids [194]. The relatively young allopolyploid *Brassica napus* may be a relevant example. The plastid genomes of *B. oleracea* and *B. rapa* have very few differences, and a recent analysis did not detect extensive incompatibilities with nuclear subgenomes [60]. By contrast, elevated rates of protein-sequence evolution and ω values in organelle-interacting genes have been detected repeatedly in lineages with rapidly evolving cytoplasmic genomes [86,87,89,195–202]. Therefore, genome-wide analyses of evolutionary rates appear to be sensitive enough to detect cytonuclear incompatibilities when their effects are strong.

Because cytonuclear interactions are critical for hybrid lineage success in many cases [203–205], allopolyploids with cytonuclear incompatibilities may also be evolutionarily short-lived, such that the relatively successful allopolyploids assayed here may be unlikely to exhibit cytonuclear incompatibilities. Along these lines, allopolyploid unisexual salamanders do not appear to exhibit maternally biased expression of nuclear-encoded OXPHOS genes [206], despite high rates of mitochondrial DNA sequence evolution and ancient divergence of the mitochondrial lineage from the paternal lineages [173]. The high incidence of asexuality and selfing species among polyploid lineages may speak to this possibility [44]. Overall, the data presented here and elsewhere appear most consistent with a scenario in which cytonuclear incompatibilities have minimal effects on rates of protein sequence evolution in allopolyploid plants.

Cytonuclear gene content evolution in allopolyploids

Polyploids often have both larger cells [52,207–210] and more chloroplasts per cell in leaf tissue ([211–216][53]). Together, these phenomena suggest that stoichiometry between nuclear and cytoplasmic genomes is important for cellular and organismal function [55]. Previous work investigating single-copy genes in plants indicated that organelle-targeted genes are among the first to return to diploidy following whole genome duplication events [123,124]. The gene content analyses presented here generally agree with those analyses, although cotton and coffee offer important exceptions that muddy the waters. By contrast, Ferreira de Carvalho and colleagues [60] reported higher levels of maintained duplicates in organelle-targeted genes in the allopolyploid *Brassica napus,* compared to genome-wide levels. The discrepancies between the former two studies (performed in diploids) and the latter two (performed in polyploids) indicates that cytonuclear stoichiometry may be highly responsive to nuclear gene content. In support of that hypothesis, diverse polyploids appear to compensate for elevated nuclear ploidy with increased organelle genome copy number [40,214,217–220]. Additional work investigating the immediate and evolved consequences of cytonuclear stoichiometry at the genomic, transcriptomic, proteomic, and organellar levels, especially by homoeologous pair analysis, will provide valuable insights into the unresolved question of how genome doubling can affect cellular energy production and homeostasis.

Summary

The genome-wide analyses of maternal vs. paternal evolutionary rates presented here represent the most extensive investigation of cytonuclear incompatibilities in allopolyploids performed to date, representing six distinct allopolyploidization events of

varying ages and divergences. We find clear evidence of differential evolution across subgenomes, but little evidence of paternal-homoeolog-specific accelerations of evolutionary rates in organelle-targeted genes. Additionally, we found that organelle-targeted gene content tends to be less biased than the rest of the genome, with mixed evidence of whether organelle-targeted genes tend to be lost more often than the rest of the genome. Further work investigating the forces underlying these observations and the consequences for organismal energy metabolism and homeostasis will be critical for understanding the cytonuclear dimension of allopolyploidy.

MATERIALS AND METHODS

Genomic datasets

The proliferation of genome assemblies for polyploid plants and their diploid relatives has enabled powerful phylogenomic analyses. We identified six allotetraploids that share hybrid origins (Figure 1a), have publicly available chromosome-scale genome assemblies for both the polyploid and the diploids that are most closely related to each subgenome (with the exception of the wild emmer wheat [*Triticum dicoccoides*] B subgenome, whose diploid relative [*Aegilops speltoides*] only has a transcriptome available), and varying degrees of divergence between their diploid progenitors and the amount of time since allopolyploidization (Figure 1b). We also included the closest available chromosome-scale assembly for an outgroup species to polarize substitutions. Accession numbers and references are provided for assemblies and annotations used from each species complex in Table 5.

Table 5. Genomic resources for six allotetraploid species complexes.

Species complex	Species	Ploidy	Version/Accession	Reference
Brachypodium	*Hordeum vulgare*[O]	2x	GCA_901482405.1	[110]
	Brachypodium distachyon	2x	GCA_000005505.4	[221]
	Brachypodium stacei[1]	2x	B_stacei_v1_1	[122]
	Brachypodium hybridum	4x	B_hybridum_v1_1	[122]
Coffee	*Gardenia jasminoides*[O]	2x	GCA_013103745.1	[117]
	Coffea canephora	2x	GCA_900059795.1	[116]
	Coffea eugenoides[2]	2x	GCA_003713205.1	[222]
	Coffea arabica	4x	GCA_003713225.1	[114,222]
Cotton	*Gossypioides kirkii*[O]	2x	Gossypioides_kirkii_ISU-v3.0	[106]
	Gossypium raimondii	2x	G.raimondii_JGI_221_v2.0	[104]
	Gossypium arboreum[3]	2x	G.arboreum_CRI-A2_assembly_v1.0	[223]
	Gossypium hirsutum	4x	Ghirsutum_458_v1.0	[224]
Quinoa	*Spinacia oleracea*[O]	2x	GCA_002007265.1	[225]
	Chenopodium suecicum	2x	Csuecicum_DT_PBjellyM2	
	Chenopodium pallidicaule[4]	2x	PGA_assembly_final_assembly_Cpallidicaule	[120]
	Chenopodium quinoa	4x	quinoa_pb_chicago-2-final_PBJELLY_pilon	[21]
Tobacco	*Solanum lycopersicum*[O]	2x	ITAG4.0	[226]
	Nicotiana tomentosiformis	2x	GCA_000390325.2	[119]
	Nicotiana sylvestris[5]	2x	GCA_000393655.1	[119]
	Nicotiana tabacum	4x	GCA_002210045.1	[118]
Wheat	*Hordeum vulgare*[O]	2x	GCA_901482405.1	[110]
	Triticum urartu	2x	GCA_003073215.1	[109]
	Aegilops speltoides[6]	2x	SRR949822	
	Triticum dicoccoides	4x	GCA_002162155.2	[111]

[O]– Species used as outgroup sequence

[1] – Closest extant relative to maternal progenitor inferred from plastome data [122]

[2] – Closest extant relative to maternal progenitor inferred from plastome data [227]

[3] – Closest extant relative to maternal progenitor inferred from mitochondrial and plastome data [107,228]

[4] – Closest extant relative to maternal progenitor inferred from mitochondrial and plastome data [229]

[5] – Closest extant relative to maternal progenitor inferred from mitochondrial and plastome data [230,231]

[6] – Closest extant relative to maternal progenitor inferred from plastome data [232]

Orthologous quintet inference

Each of the six allopolyploids have subgenomes that are more closely related to those of the sampled diploids than they are to each other. Combined with an outgroup lineage, the resulting tree topology characteristics of allopolyploids (Figure 2) allow for robust inference of lineage-specific rates of evolution in orthologous quintets. We used a combination of phylogenetic and syntenic methods to construct orthologous quintets (Figure S1).

To infer orthologous quintets using phylogenetic methods, we used Orthofinder v. 2.2.7 to infer orthologous groups of sequences, termed “orthogroups”, from the whole proteomes (primary transcripts only) of all four species [233]. For each orthogroup, we aligned CDS sequences in a codon-aware manner using the align_fasta_with_mafft_codon subroutine in the sloan.pm perl module (available at https://github.com/dbsloan/perl_modules) which translates CDS sequences into amino acid sequences, aligns those amino acid sequences with MAFFT v7.407 [234], and reverse translates the aligned amino acid positions into the CDS sequences to produce the final alignment. We selected models of molecular evolution for each alignment using jModelTest2 v2.1.10 to identify the model with the highest AICc score [235,236], and inferred phylogenetic trees with the MPI-compatible distribution of PhyML v3.3.20180214 [235]. Five random tree starts were performed, and the treespace was further searched using a combination of nearest-neighbor-interchange subtree pruning and re-grafting. Support for trees was assessed using 100 bootstrap replicates, and splits with ≤ 50 bootstrap support were collapsed into polytomies using collapeLowSupportBranches.py (unless otherwise stated all scripts are available at https://github.com/jsharbrough/allopolyploidCytonuclearEvolutionaryRate/tree/master/scripts).

All monophyletic, minimally inclusive, species-complete subtrees were pruned out of orthogroup trees using subTreeIterator.py. We next trimmed lineage-specific gene duplicates from subtrees using trimBranches.py, which keeps only the longest sequence or a random sequence in cases where sequence length is equal across copies. The resulting trimmed subtrees that contained exactly one sequence from each diploid and two sequences from the polyploid represented our set of phylogenetic orthologous quintets. All scripts developed for reading, writing, and manipulating trees are based on the DendroPy package (https://dendropy.org/) [237].

We used the pSONIC [238]program to create a genome-wide set of syntenic orthologs. In short, pSONIC employs MCScanX [239] to create a list of pairwise syntenic blocks between all possible pairs of species in each clade, combined with orthogroups identified from OrthoFinder [233] to choose which syntenic blocks contained the highest confidence orthologs that were direct descendants of the most recent common ancestor of all species in the clade. Notably, the filtering criteria of collinear groups from our run of pSONIC differed from its formal presentation in that we did not remove collinear groups in which more genes received a “not pass” than “pass” score, and the ends of each collinear block were not trimmed as described in the manuscript describing pSONIC. These developments were made after our analyses were performed with this tool, but before the tool was submitted and reviewed for publication.

To take advantage of both inference methods, we merged phylogenetic and syntenic orthologous quintets using mergeQuintets.py to produce a high quality set of quintets that were identical across both methods (i.e. “Intersection”) and a second set of quintets that included all identical quintets plus all the phylogenetic quintets whose members were not present in the syntenic quintets and vice versa (i.e., “Union”). Results from the Intersection dataset (Supplementary File 1, Figure S10, Figure S11) did not differ in any meaningful way from the Union, so only Union results are described in the main text. Phylogenetic quintets that overlapped with but were not identical to syntenic quintets were excluded. Likewise, syntenic quintets that overlapped with but were not identical to phylogenetic quintets were also removed from our final analysis. These conflicting quintets represent a small minority of total quintets and are likely a result of the different methods by which lineage-specific duplicates are handled in the phylogenetic vs. syntenic pipelines.

For all non-conflicting orthologous quintets, we re-aligned CDS sequences as before, trimmed alignments with Gblocks v0.91b using the codon setting with the parameter -p set to n [240]. We estimated new models of molecular evolution using jModelTest2 [235,236], and inferred phylogenetic trees as described above. We tested whether the resulting gene tree topologies agreed with the overall species tree using the quintetTopology.py script and excluded all genes with discordant tree topologies from subsequent evolutionary rate analyses.

CyMIRA-based gene classification

To evaluate the effect of cytonuclear interactions on subgenome-specific evolutionary dynamics, we used a combination of *de novo* targeting predictions and CyMIRA [67] to partition genes into distinct functional and interaction categories. *De novo* targeting predictions were obtained from four separate targeting prediction programs: iPSORT v0.94 [241], LOCALIZER v1.0.4 [242], Predotar 1.03 [243], and TargetP v1.1b [244]. In parallel, we used Orthofinder v2.2.7 to obtain orthology information with the *Arabidopsis*

thaliana Araport 11 proteome [245]. We combined the *de novo* targeting predictions with the *Arabidopsis*-inclusive orthogroups using the geneClassification.py script. Genes were classified as cytonuclear-interacting genes if they shared the same orthogroup as *Arabidopsis* genes whose products interact with mitochondrial/plastid genomes or gene products according to the CyMIRA classifications scheme [67]. Genes present in orthogroups lacking an *Arabidopsis* cytonuclear interacting gene were classified as organelle-targeted if at least one *de novo* prediction tool indicated a mitochondrial or plastid subcellular localization for the gene product and ≥50% of *Arabidopsis* genes present in the orthogroup encode products targeted to the mitochondria or plastids according to CyMIRA. Genes with evidence of dual targeting were included in both mitochondria-targeted and plastid-targeted data partitions. The resulting genome-wide targeting predictions and CyMIRA-guided classifications are available at https://github.com/jsharbrough/allopolyploidCytonuclearEvolutionaryRate/tree/master/geneClassification and the pipeline for performing this classification is available at https://github.com/jsharbrough/CyMIRA_gene_classification. The breakdowns of gene functional categories for each genome are provided in Table 2 and Table S2.

We next evaluated whether retention of genes targeted to the organelles differs across subgenomes by comparing CyMIRA gene counts across subgenomes for five out of six polyploid genomes (*N. tabacum* was excluded from this analysis owing to the difficulty in positively identifying subgenomic ancestry for genes lacking a corresponding homoeolog). We performed binomial tests of NOT genes against expectations of equal retention, and then used χ^2 tests of organelle-targeted gene groups against the genome-wide patterns observed among genes not targeted to the organelles.

Evolutionary rate comparisons

We evaluated genome-wide signatures of cytonuclear incompatibilities in organelle-targeted genes using a combination of single gene and concatenated analyses. For all single-copy quintets whose evolutionary history was consistent with the overall species tree, we removed poorly aligned quintets by estimating the total length of the tree in terms of synonymous substitutions per site (d_S) using model 1 in codeml within PAML v4.9j [246]. Maximum cutoff values for d_S were determined for each species complex separately and are depicted by red lines in Figure S8.

After quality filtering, we estimated d_N, d_S, and ω for individual quintets using model 1 in codeml as above, and the RateAncestor parameter set to 1. Other PAML parameters included the getSE parameter set to 1, the gamma shape parameter set to a fixed alpha of 0 (i.e., no rate variation among codons), initial omega set to 0.4, and initial kappa set to 2. For each quintet in each functional gene category, we evaluated whether the maternal vs. paternal subgenome had a higher ω value and a higher d_N. We used χ^2 tests to evaluate whether individual categories differed from the pattern

observed in the group of genes not targeted to the organelles. Using the inferred mutational changes from the RateAncestor output, we also evaluated whether maternal vs. paternal subgenomes had higher numbers of radical amino acid changes (i.e., substitutions between amino acids with substantially different biochemical properties) at sites that were otherwise conserved across the quintet. Substitutions were identified as radical if their score in the CRI matrix [132] was >0.5. Accumulation of derived conservative and radical amino acid changes was analyzed in a similar manner to ω and d_N results, using a Fisher's Exact Test to test whether there was a difference compared to genes not targeted to the organelles.

Next, we concatenated quintets according to gene functional category and estimated ω in maternal vs. paternal subgenomes using similar PAML parameters as before. For each PAML run, we repeated the analysis 1000 times to adequately sample the maximum likelihood plane and found median ω values from the replicates for each branch. We then calculated the ratio of paternal to maternal subgenome ω values ($\omega_{PAT}/\omega_{MAT}$), with a ratio >1.0 indicating faster rates of amino acid sequence evolution in the paternal subgenome and a ratio <1.0 indicating a faster rate of amino acid sequence evolution in the maternal subgenome. We assessed the statistical significance of the degree to which subgenomes exhibited different rates of amino acid sequence evolution by bootstrapping concatenated alignments at the gene level. For each bootstrap replicate we randomly sampled genes with replacement from the original concatenation and ran each bootstrapped alignment through five replicate runs of PAML. The median ω values of these five replicates were used as the bootstrap replicate values. We then found the ratio of paternal to maternal ω values for each bootstrap replicate and for each gene functional category to evaluate whether bootstrapped distributions departed from 1.0. To account for evolutionary forces that are not a result of cytonuclear interactions, we normalized these ratios by dividing by the paternal to maternal ω ratio of genes not targeted to either organelle. We inferred two-tailed p values directly from bootstrap distributions. For specific cytonuclear interaction categories, which are composed of only a few dozen genes or less, we manually inspected concatenated alignments, trimmed poorly aligned regions, bootstrapped alignments at the codon level using the python script bootstrapCodons.py, and performed PAML analyses with a similar approach as before.

Because cytonuclear incompatibilities are only expected when there exists divergence between the two progenitor genomes, we also binned our quintets based on high vs. low divergence between diploids for each species and repeated the gene-level bootstrap procedure described above. First, we estimated d_N between diploid relatives for each quintet individually from the gene-specific PAML runs described above and placed genes according to d_N into two equally sized bins. We then tested whether genes with high levels of amino acid divergence exhibit greater accelerations in ω in paternal copies than in genes with lower levels of amino acid sequence divergence. We

evaluated statistical significance by bootstrapping alignments at the gene level and comparing paternal to maternal ω ratio distributions from the same gene categories to one another.

FIGURES AND FIGURE LEGENDS

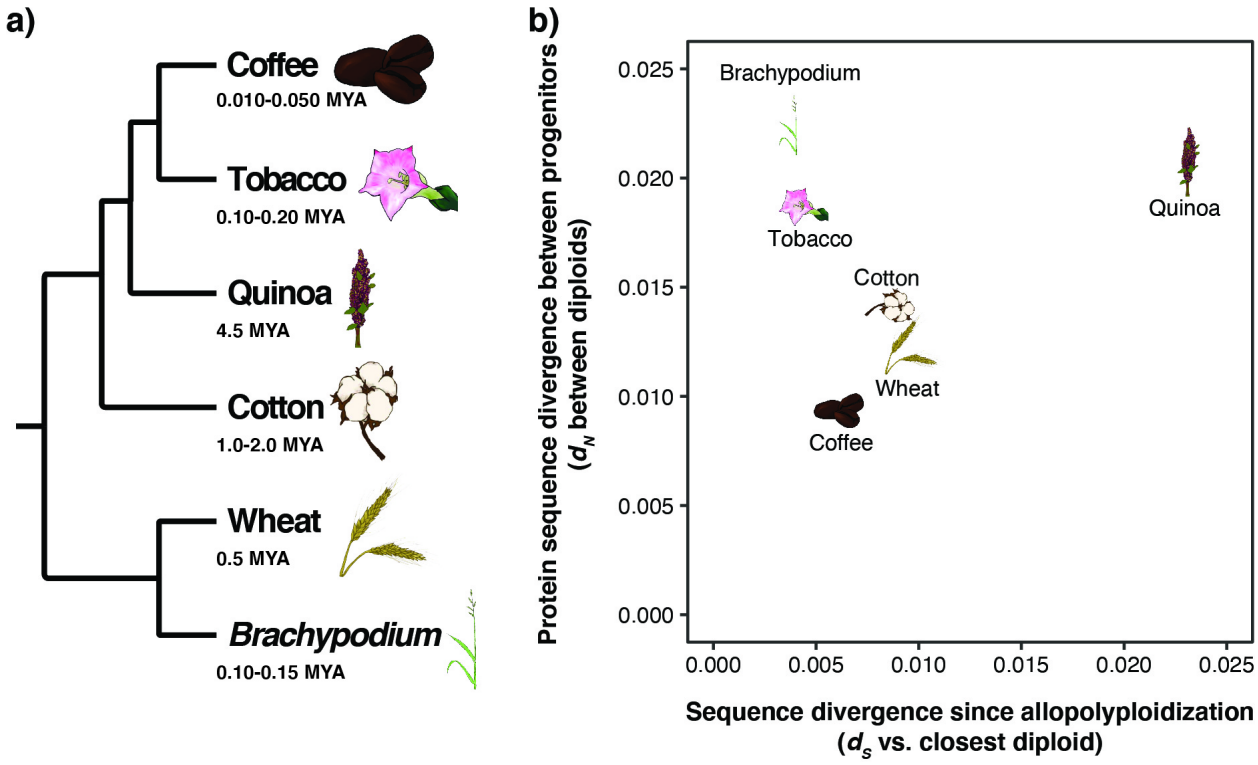

Figure 1. Evolutionary relationships and origins of six allotetraploid angiosperms. a) Cladogram depicting evolutionary relationships among six independently derived allotetraploid angiosperms. b) The scatter plot depicts the synonymous substitutions per synonymous site (d_S) between the polyploid subgenome-diploid pair with the lowest amount of divergence on the x-axis as a proxy for the amount of time since allopolyploidization. Amino acid sequence divergence between subgenomes, measured as nonsynonymous substitutions per nonsynonymous site (d_N) between the two diploid relatives, is shown on the y-axis. Higher levels of amino acid sequence divergence between subgenomes increase the probability of a genetic incompatibility in the polyploid, whereas longer periods of time since allopolyploidization increases the probability that evolutionary responses to incompatibilities are detectable in the polyploid.

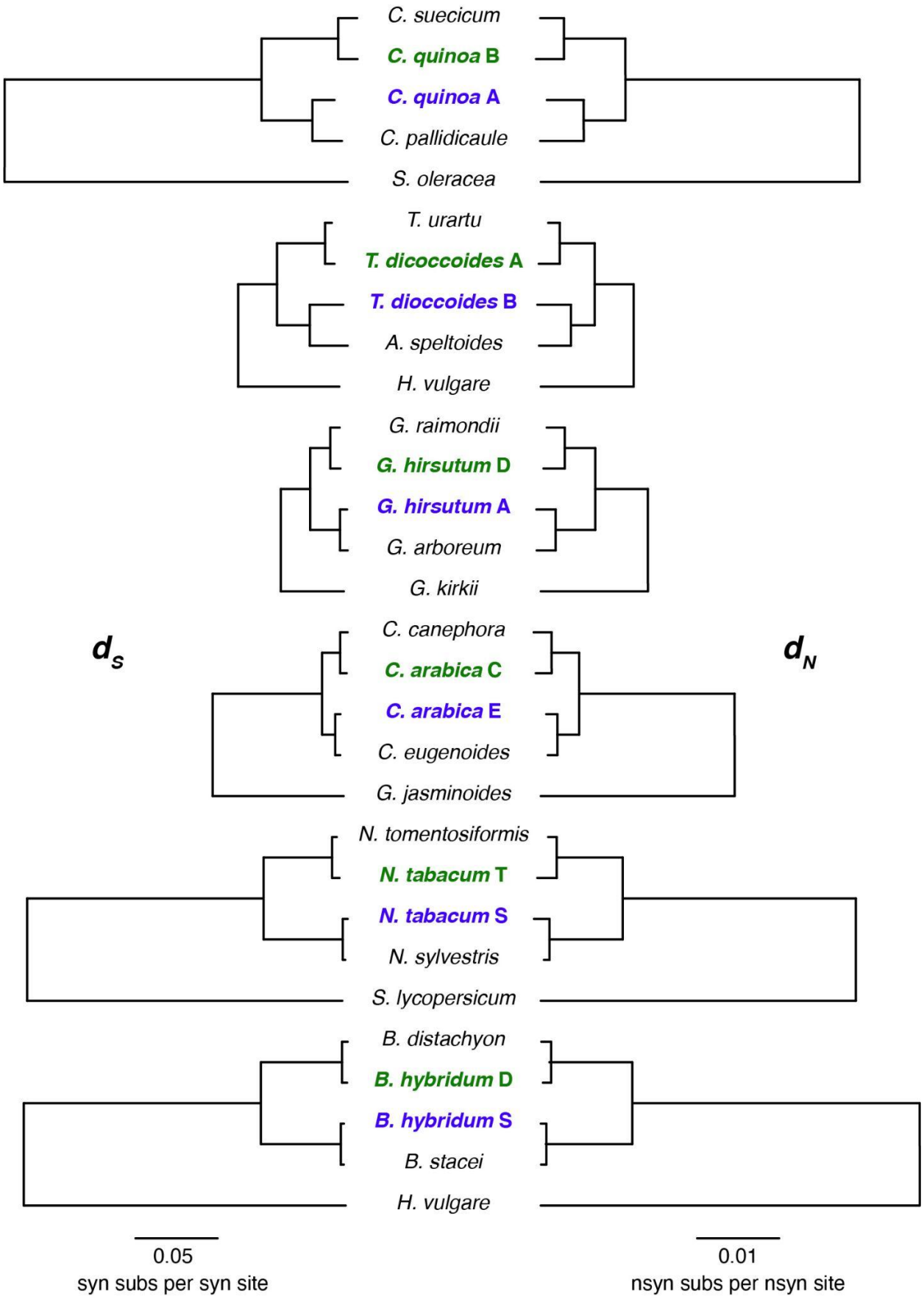

Figure 2. Synonymous and nonsynonymous rates of evolution in genomes (and subgenomes) of focal allopolyploid systems. Substitution rates per site for synonymous (d_S – left) and nonsynonymous (d_N – right) sites from concatenated analyses of non-organelle-targeted genes are represented by branch lengths for each genome (and subgenome). Allopolyploid systems are arranged from oldest (top) to youngest (bottom), as described on the x-axis of Figure 1. Paternal subgenomes of allotetraploids are bolded in green, and maternal subgenomes are bolded in purple.

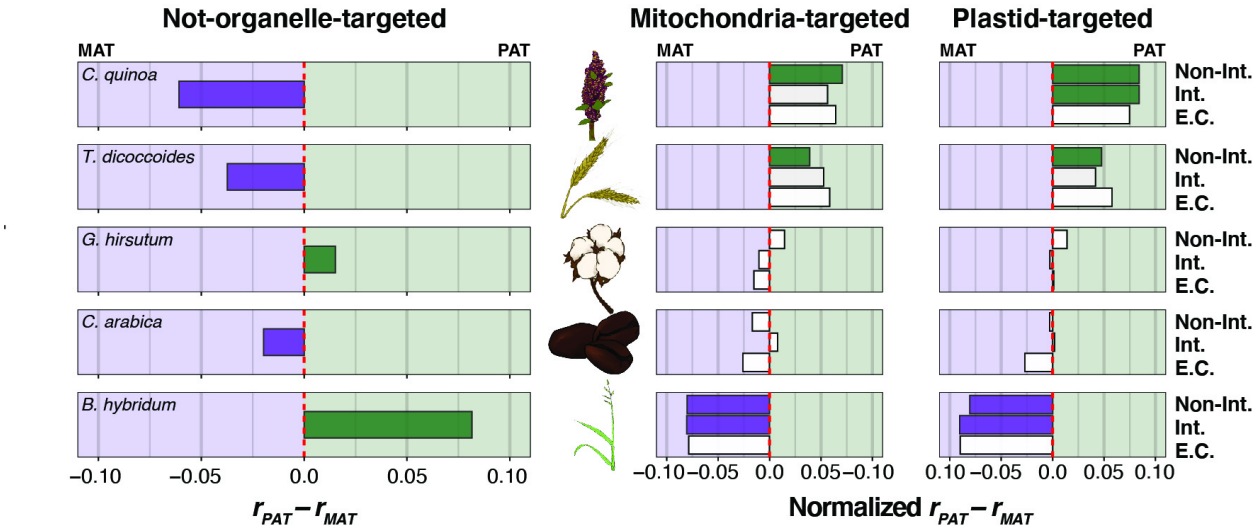

Figure 3. Gene content bias across allotetraploid subgenomes. The proportion of genes present in paternal (r_{PAT}) vs. maternal (r_{MAT}) subgenomes is depicted for each of five allotetraploid species arranged vertically from oldest (top) to youngest (bottom). Tobacco was excluded from this analysis because the massive rearrangement it has experienced makes subgenomic identification based on chromosomal position intractable. The left panel includes only non-organelle-targeted genes, the middle panel includes only mitochondria-targeted genes, and the right panel includes only plastid-targeted genes. In the left panel, the red-dashed line represents equal content across subgenomes. In the right two panels, the r_{PAT} and r_{MAT} are normalized by the proportions estimated from genes not targeted to the organelles, such that the red-dashed line reflects genome-wide patterns, rather than equality across subgenomes. Proportion deltas that depart significantly from the red line are filled in solid according to the direction of subgenomic bias (i.e., green: $r_{PAT} > r_{MAT}$; purple: $r_{PAT} < r_{MAT}$; no fill: $r_{PAT} \approx r_{MAT}$). The intimacy of interactions are depicted on the y-axis for each of the right two panels from low- or no- interaction with organelle gene products (top), to interacting genes (middle), to genes involved in mitochondrial or plastid enzyme complexes (bottom).

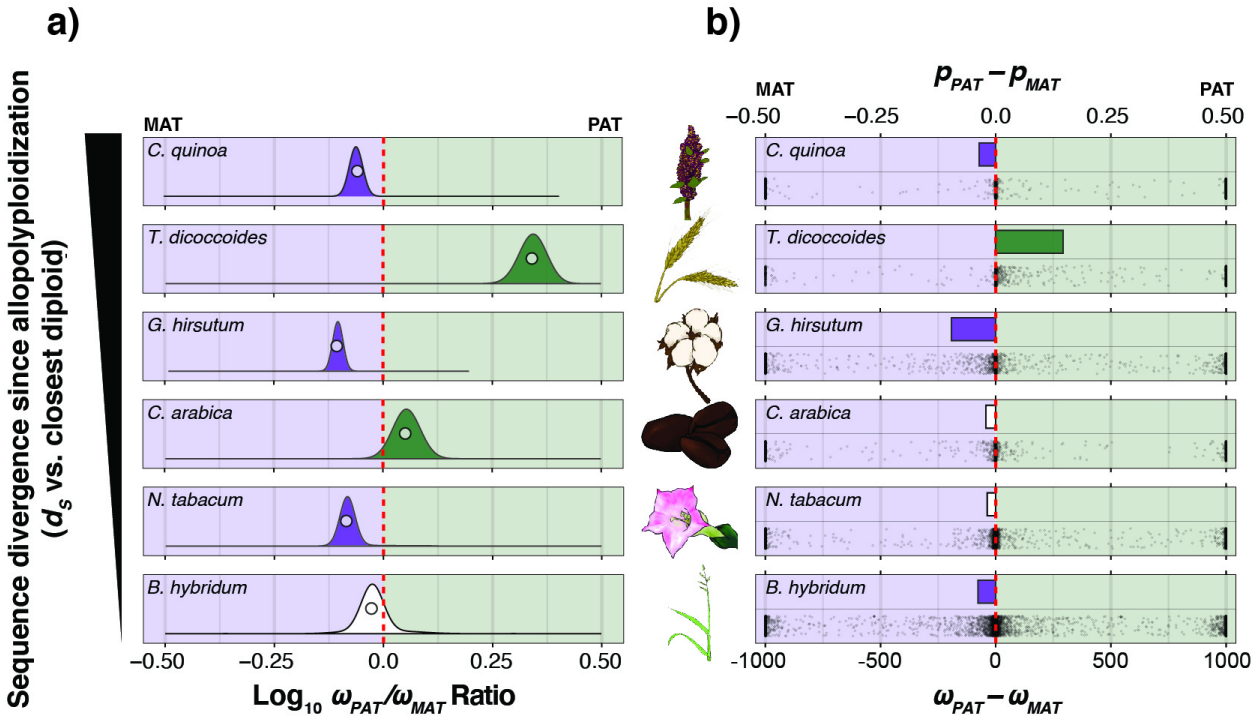

Figure 4. Genome-wide bias in ω (d_N/d_S) across maternal and paternal subgenomes. a) Log-transformed ratios of ω values in paternal (ω_{PAT}) vs. maternal (ω_{MAT}) subgenomes from concatenations (circles), and underlying bootstrap distributions (density curves) of genes encoding proteins that are not targeted to either the plastids or mitochondria. Species panels are arranged vertically from oldest (top) to youngest (bottom). The red-dashed line indicates equal ω values across subgenomes, left of the red line indicates higher ω values in the maternal subgenomes, and right of the red line indicates higher ω values in the paternal subgenome. Bootstrap distributions of ω ratios that depart significantly ($p < 0.05$) from the red line are filled in solid according to the direction of subgenomic bias (i.e., green: $\omega_{PAT}/\omega_{MAT} > 1.0$; purple: $\omega_{PAT}/\omega_{MAT} < 1.0$; no fill: $\omega_{PAT}/\omega_{MAT} \approx 1.0$). b) Estimates of $\omega_{PAT} - \omega_{MAT}$ for each individual gene is depicted on the bottom half of each species' panel and the proportion of genes with higher ω values in the paternal subgenome (p_{PAT}) minus the proportion of genes with higher ω values in the maternal subgenome (p_{MAT}) is depicted on the top half of each species' panel for all genes not-targeted to either the mitochondria or plastids. The red-dashed line represents equal proportions of genes with higher ω values across subgenomes, and bars are filled in when proportion deltas are significantly different from zero (i.e., green: $p_{PAT} > p_{MAT}$; purple: $p_{PAT} < p_{MAT}$; no fill: $r_{PAT} \approx r_{MAT}$).

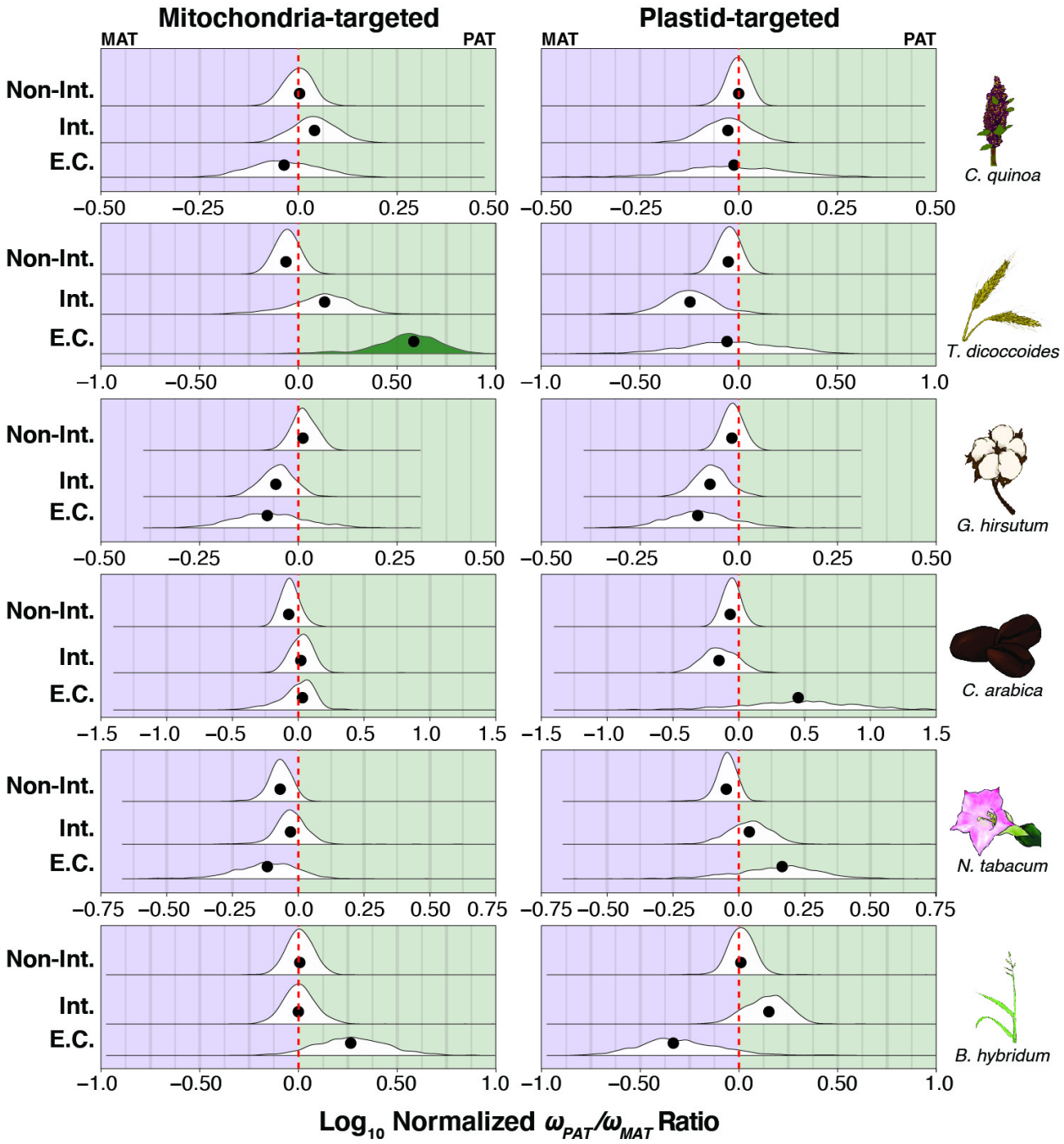

Figure 5. Ratios of maternal vs. paternal ω values in organelle-targeted genes. Log-transformed ratios of maternal vs. paternal ω values for concatenations (black circles) and underlying bootstrap distributions (density curves) of mitochondria- (left) and plastid-targeted (right) genes in six focal allotetraploid species. Species panels are arranged vertically from oldest (top) to youngest (bottom). The red-dashed line indicates the $\omega_{PAT}/\omega_{MAT}$ ratio for a concatenation of genes not-targeted to the organelles (Figure 4a). Ratios left of the red line indicate higher ω values in the maternal subgenome, and ratios right of the red line indicate higher ω values in the paternal subgenome, after

accounting for genome-wide patterns. Bootstrap distributions of ω ratios that depart significantly ($p < 0.05$) from the red line are filled in solid according to the direction of subgenomic bias (i.e., green: normalized $\omega_{PAT}/\omega_{MAT} > 1.0$; purple: normalized $\omega_{PAT}/\omega_{MAT} < 1.0$; no fill: normalized $\omega_{PAT}/\omega_{MAT} \approx 1.0$). The intimacy of interactions are indicated on the y-axis from low or no interaction with organelle gene products (top), to interacting genes (middle), to genes involved in mitochondrial or plastid enzyme complexes (bottom).

SUPPLEMENTAL FIGURES AND FIGURE LEGENDS

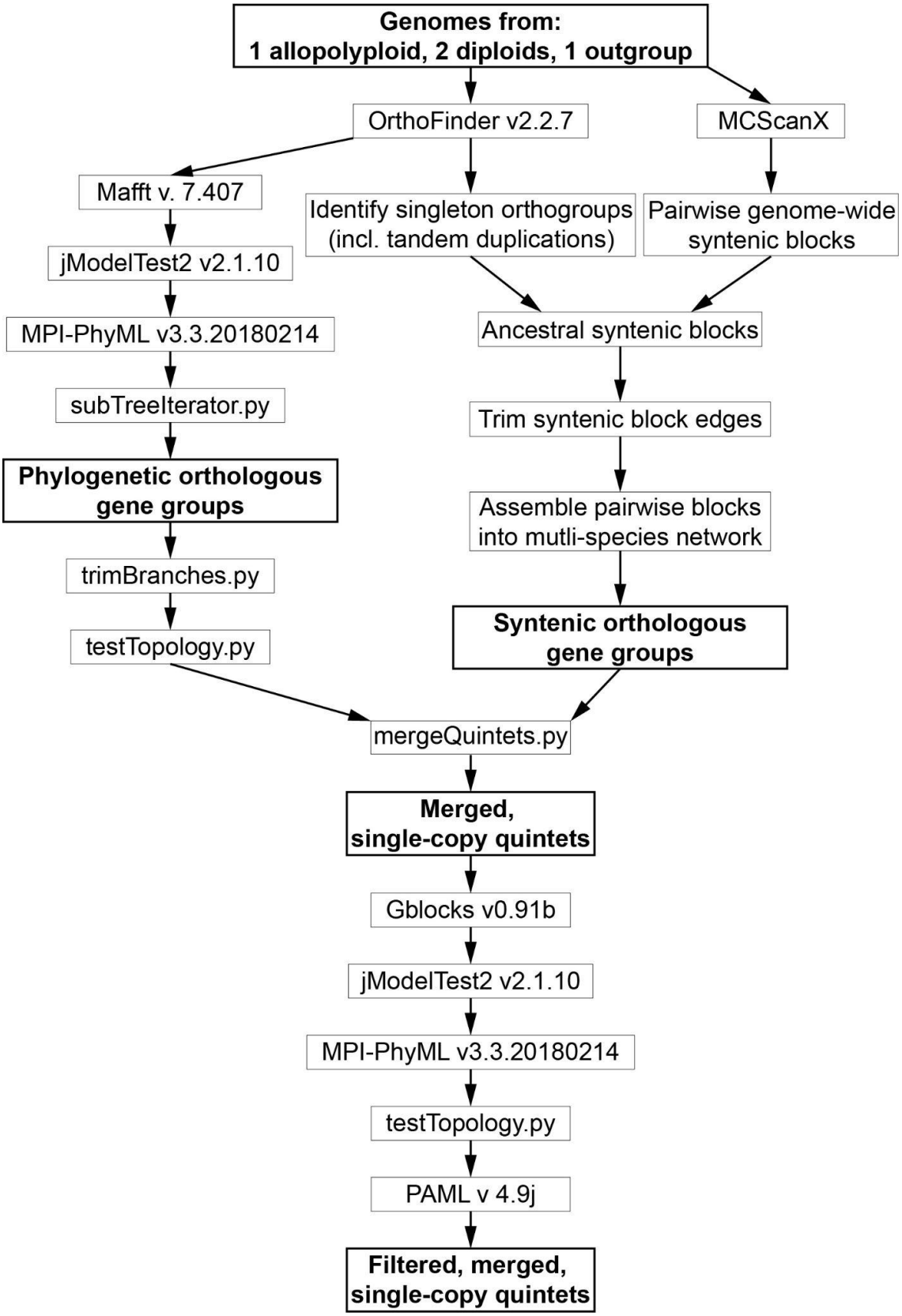

Figure S1. Schematic representation of phylogenetic and syntenic pipelines for inferring orthologous quintets in allopolyploid genomes. The final output of the pipeline was a set of filtered, merged, single-copy quintets, which was used in downstream analyses of rates of protein-sequence evolution.

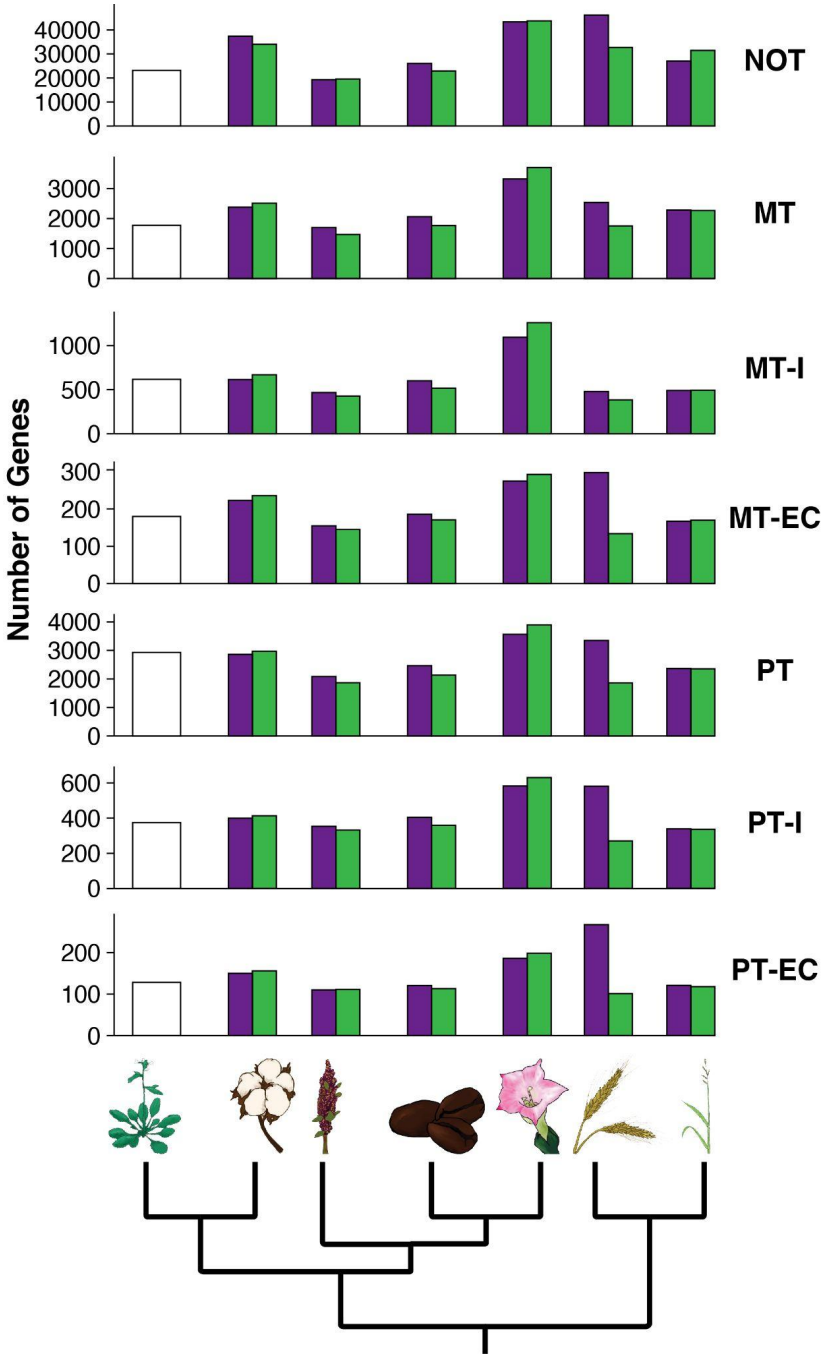

Figure S2. CyMIRA gene counts in diploid models compared to *Arabidopsis*. The number of genes per category is depicted on the y-axis for both maternal (purple) and paternal (green) diploid models compared to *Arabidopsis* (white). Functional gene categories are listed to the right of each plot: NOT – genes that are not-organelle-targeted, MTNI – mitochondria-targeted non-interacting genes, MTI – mitochondria-targeted interacting genes, MTEC – genes involved in mitochondrial enzyme complexes (subset of MTI), PTNI – plastid-targeted non-interacting genes, PTI

– plastid-targeted interacting genes, PTEC – genes involved in plastid enzyme complexes (subset of PTI).

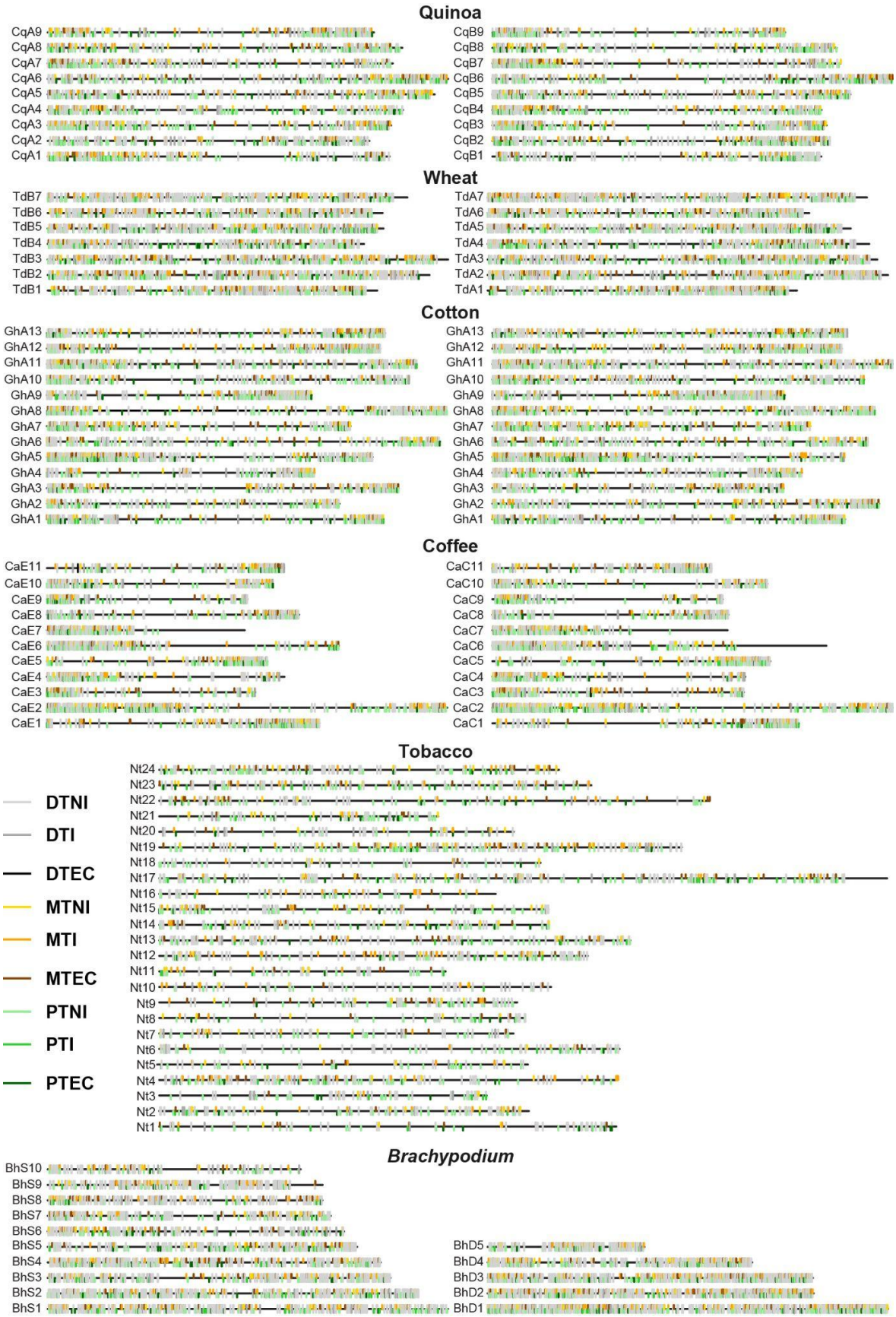

Figure S3. Physical distribution of organelle-targeted genes on chromosomes of focal allopolyploid genomes. Mitochondria-targeted (orange), plastid-targeted (green), and dual-targeted (grey) genes mapped onto chromosomes (black lines) of the six focal allotetraploid genomes. Taxa are arranged from oldest (top) to youngest (bottom), with maternally derived subgenomes on the left and paternally derived subgenomes on the right (excepting tobacco). Chromosome numbers are listed to the left of each chromosome (quinoa chromosomes are numbered according to similarity with *Chenopodium pallidicaule* chromosomes).

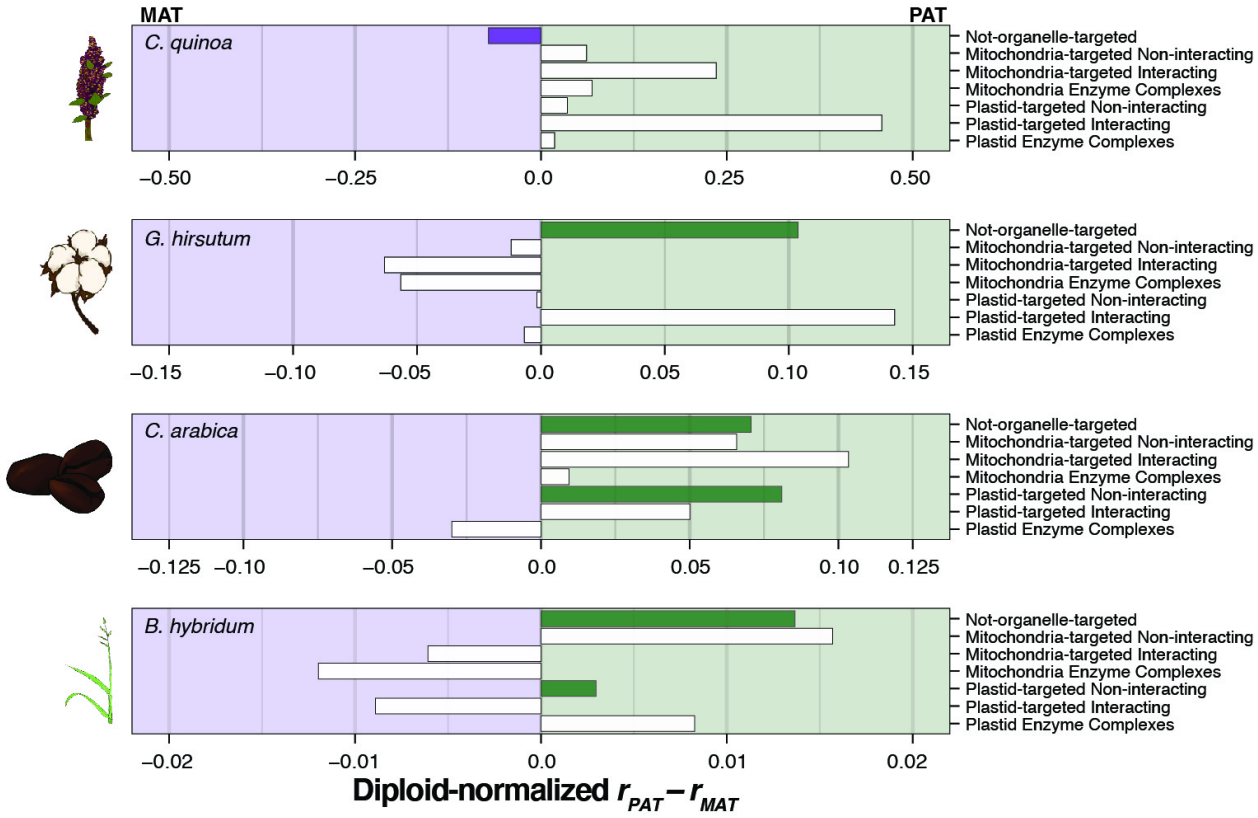

Figure S4. CyMIRA gene counts in maternal and paternal subgenomes of allotetraploids relative to diploid models. Bar graph depicting the number of genes present in the maternal subgenome as a proportion of the number of genes present in the maternal diploid model's genome (r_{MAT}) subtracted from the number of genes present in the paternal subgenome as a proportion of the number of genes present in the paternal diploid model's genome (r_{PAT}) for seven functional categories of genes: not-organelle-targeted, mitochondria-targeted non-interacting, mitochondria-targeted-interacting, mitochondria enzyme complexes, plastid-targeted non-interacting, plastid-targeted interacting, plasti enzyme complexes. Polyploid taxa are arranged vertically from oldest (top) to youngest (bottom) for quinoa, cotton, coffee, and *Brachypodium*. Wheat was excluded because the maternal diploid transcriptome from *Aegilops speltoides* was not a good indicator of gene counts.

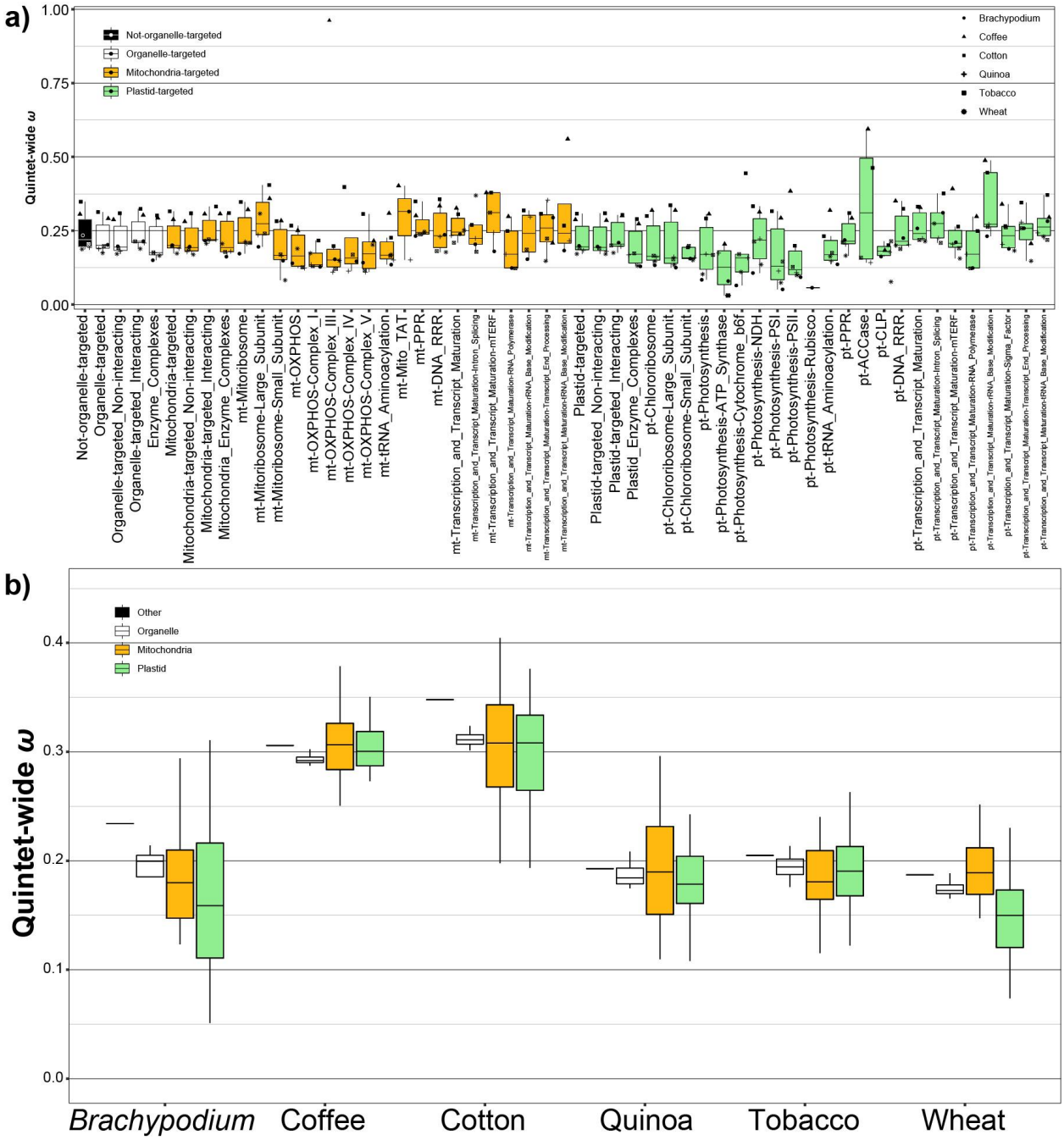

Figure S5. Rates of protein-sequence evolution in CyMIRA gene categories across the six focal allopolyploids. a) Quintet-wide rates of protein-sequence evolution across CyMIRA functional categories are depicted in box-and-whisker plots for the combined set of allopolyploids. Data points from each species complex are also shown, the legend for which is provided in the upper right corner of the plot. Boxes are filled according to subcellular compartment, with genes not targeted to the organelles represented by black boxes, genes targeted to either organelle by white boxes, mitochondria-targeted genes are represented by orange boxes, and plastid targeted

genes are represented by green boxes. b) Boxplot depicting rates of protein-sequence evolution across allopolyploid species complexes separated by subcellular compartment of localization. Fill color is as described in panel (a).

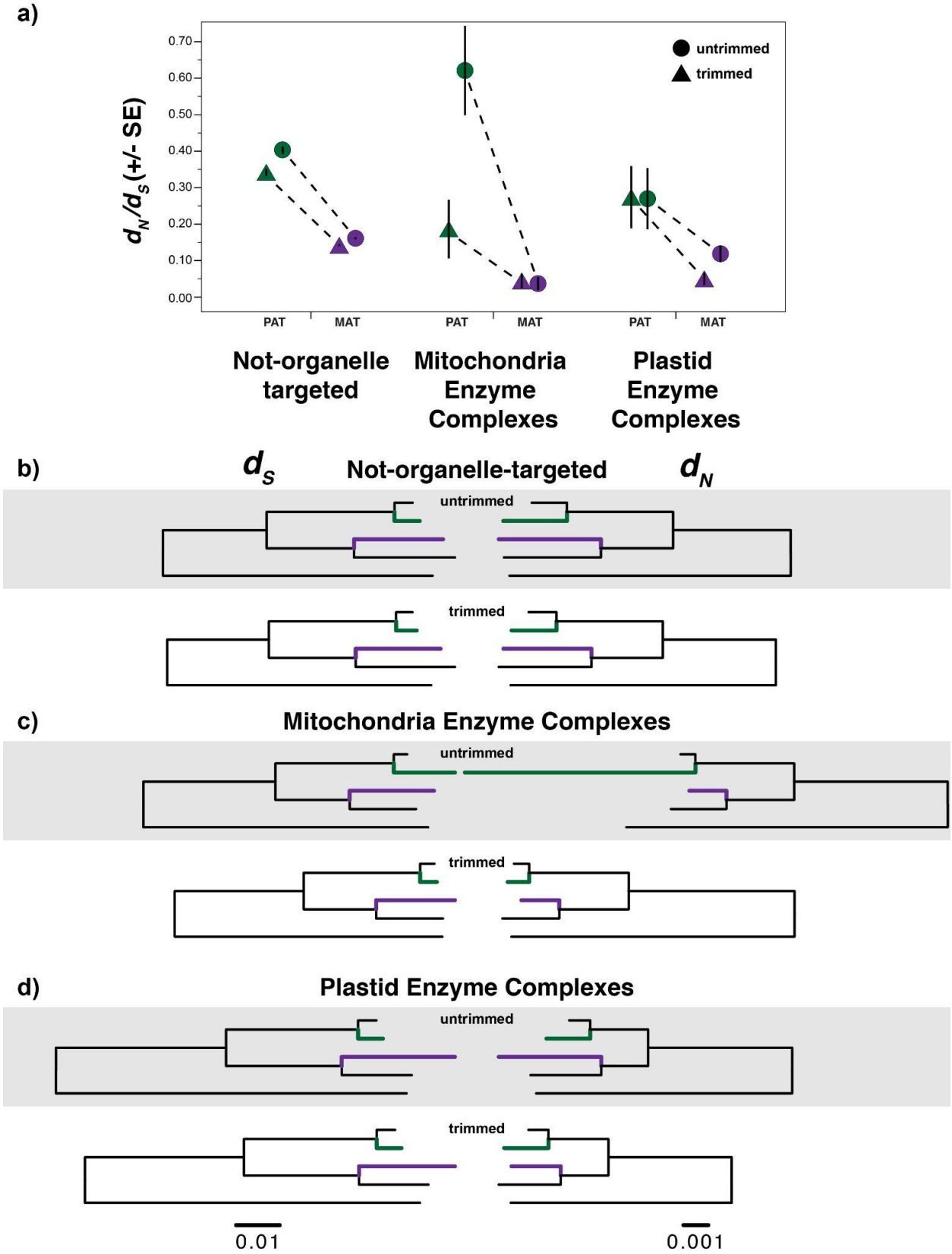

Figure S6. Poorly aligned regions largely explain elevated *ω* values in paternal homoeologs of wheat mitochondrial enzyme complexes. a) *ω* values from maternal (purple) vs. paternal (green) branches estimated from concatenations of genes that are not-organelle-targeted (left), involved in the mitochondrial enzyme complexes (middle), or involved in plastid enzyme complexes (right) in untrimmed (circles) vs. trimmed (triangles) alignments. The removal of two regions totalling ~240bp accounts for the apparently elevated *ω* values in paternal homoeologs of wheat mitochondrial enzyme complex genes. b-d) Deconstructed *ω* values from concatenated PAML runs for genes (b) not-targeted to the organelles, (c) genes involved in the mitochondrial enzyme complexes, and (d) genes involved in plastid enzyme complexes in untrimmed (top) vs. trimmed (bottom) alignments. Rates of evolution for synonymous (d_S - left) and nonsynonymous (d_N - right) sites are represented by branch lengths, and branches are scaled similarly across functional categories. Green branches represent rates in the paternal subgenome and purple branches represent evolutionary rates in the maternal subgenome of *T. dicoccoides*.

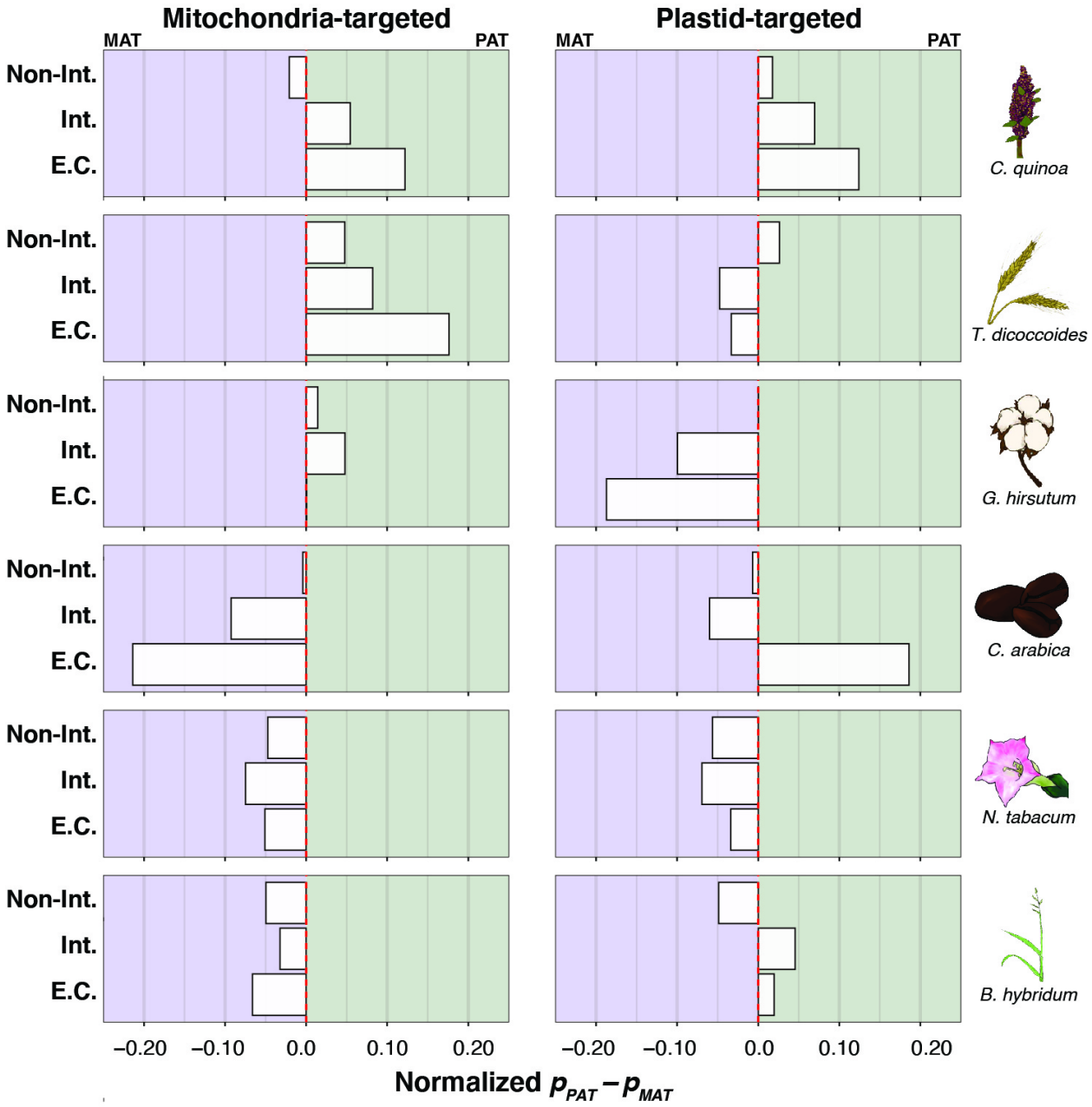

Figure S7. Proportions of genes that have higher *ω* values in paternal vs. maternal copies of organelle-targeted genes. The proportion of genes with higher *ω* values in the paternal homoeolog than in the maternal homoeolog (p_{PAT}) minus the proportion of genes with higher *ω* values in the maternal homoeolog than in the paternal homoeolog (p_{MAT}) is depicted along the x-axis. Mitochondria- (left) and plastid-targeted (right) genes are separated by the degree of interaction: non-interacting genes (top), interacting genes (middle), and genes involved in cytonuclear enzyme complexes (bottom, subset of interacting genes). Proportions are normalized by those found in non-organelle-targeted genes, and genomic bias is denoted by color with maternal bias (i.e., $p_{PAT} - p_{MAT} < 0$) colored purple and paternal bias (i.e., $p_{PAT} - p_{MAT} > 0$) colored

green. None of the values exhibited biased proportions according to χ^2 tests, relative to genes not targeted to the organelles. Allopolyploids are arranged from oldest (top) to youngest (bottom) as in Figure 2.

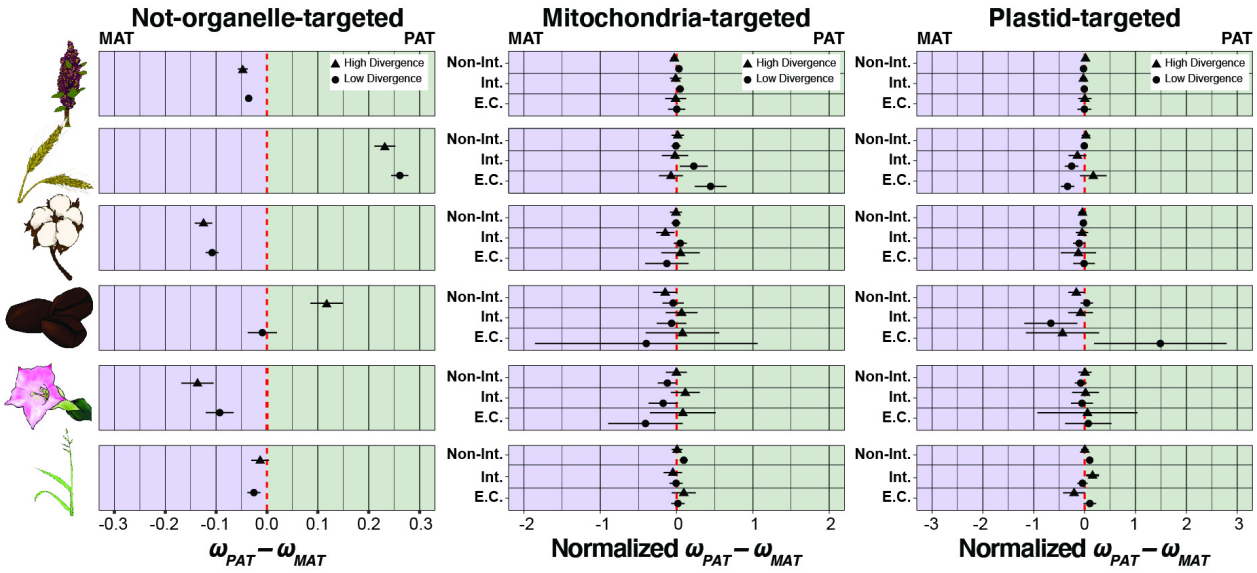

Figure S8. Divergence-binned analysis of subgenomic bias in rates of protein-sequence evolution in six focal allopolyploids. Rates of protein-sequence evolution paternal subgenomes (ω_{PAT}) minus those found in maternal subgenomes (ω_{MAT}) are depicted on the x-axis for non-organelle-targeted genes (left), mitochondria-targeted genes (middle), and plastid-targeted genes (right). Genes were concatenated by functional category and binned according to divergence with high-divergence bins (top) depicted by triangles and low-divergence bins (bottom) depicted by circles. Error bars represent standard errors inferred by PAML. Allopolyploids are arranged from oldest (top) to youngest (bottom). The right two panels are further divided by the degree of intimacy of interaction, with non-interacting genes on top, interacting genes in the middle, and genes that are part of enzyme complexes on bottom. The red-dashed line represents equal rates of protein-sequence evolution in the left panel, but on the right two panels the red-dashed line represents the genome-wide pattern taken from the left panel (i.e., organelle-targeted rates were normalized by non-organelle targeted rates). Maternal bias (i.e.,$\omega_{MAT} > \omega_{PAT}$) occurs left of the red-dashed lines and paternal bias (i.e., $\omega_{PAT} > \omega_{MAT}$) occurs to the right of the red-dashed line.

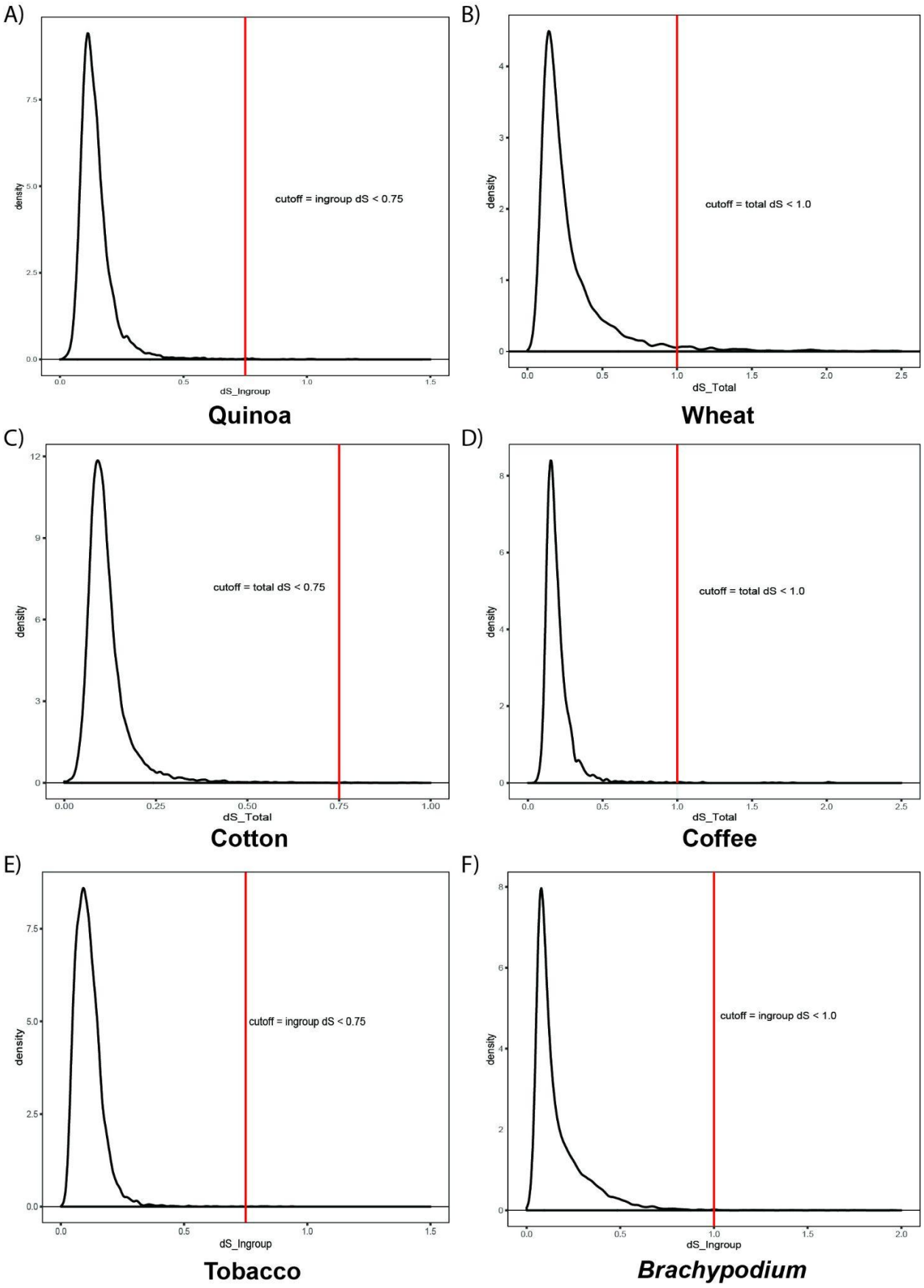

Figure S9. Alignment filtering based on d_S in orthologous quintets. Individual genes with either total d_S values (wheat, cotton, coffee) or ingroup-only d_S levels (quinoa, tobacco, *Brachypodium*) greater than the cutoff point (indicated by the red line) were excluded from analyses of rates of protein-sequence evolution, as we could not exclude the possibility that those quintets were poorly aligned vs. truly divergent.

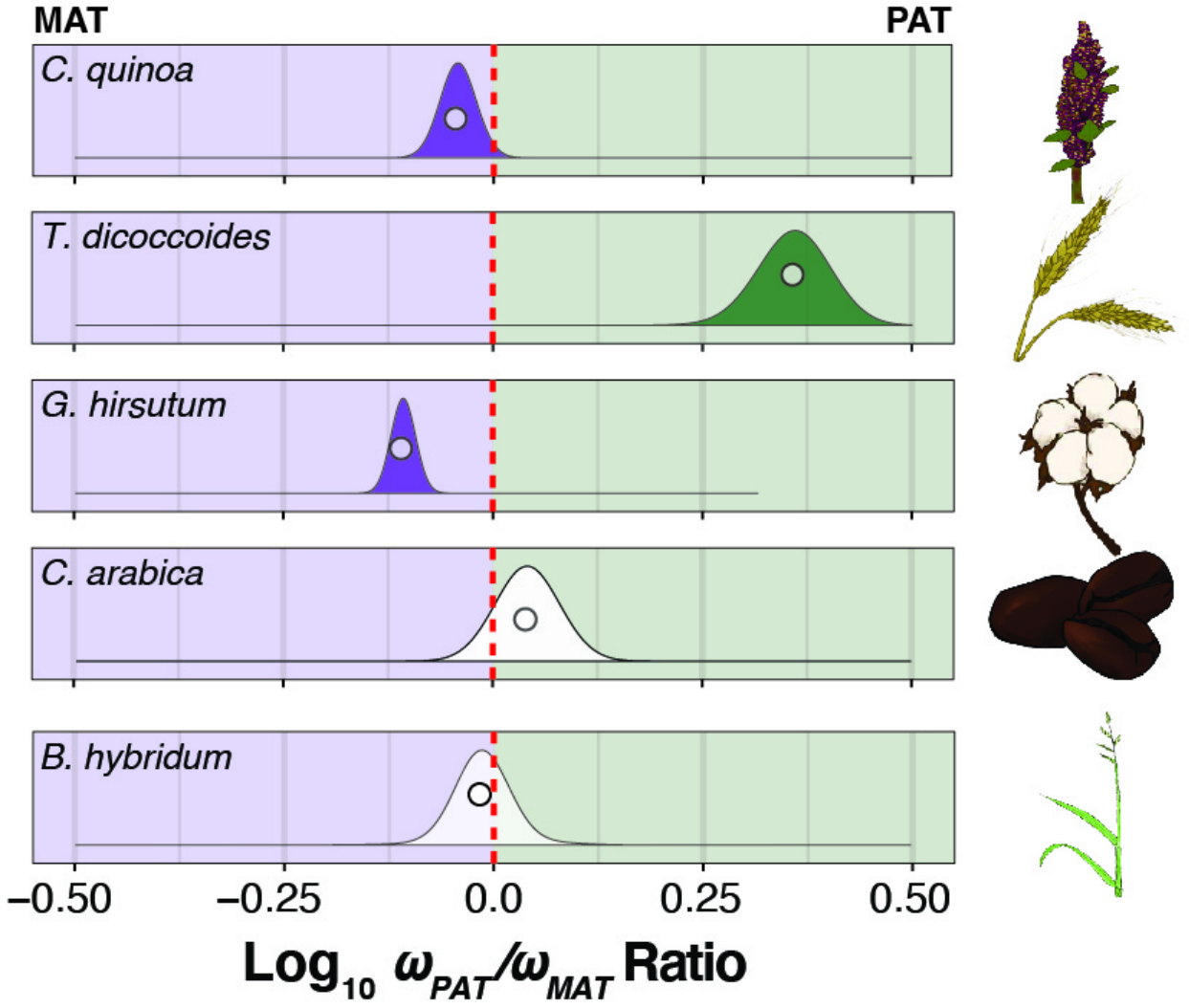

Figure S10. Genome-wide bias in ω (d_N/d_S) across maternal and paternal subgenomes, identical quintets only. Log-transformed ratios of ω values in paternal (ω_{PAT}) vs. maternal (ω_{MAT}) subgenomes from concatenations (circles), and underlying bootstrap distributions (density curves) of genes encoding proteins that are not targeted to either the plastids or mitochondria using only quintets that were identical across phylogenetic and syntenic methods. Species panels are arranged vertically from oldest (top) to youngest (bottom). Tobacco was excluded from this analysis because it produced so few syntenic quintets. The red-dashed line indicates equal ω values across subgenomes, left of the red line indicates higher ω values in the maternal subgenomes, and right of the red line indicates higher ω values in the paternal subgenome. Bootstrap distributions of ω ratios that depart significantly ($p < 0.05$) from the red line are filled in solid according to the direction of subgenomic bias (i.e., green: $\omega_{PAT}/\omega_{MAT} > 1.0$; purple: $\omega_{PAT}/\omega_{MAT} < 1.0$; no fill: $\omega_{PAT}/\omega_{MAT} \approx 1.0$).

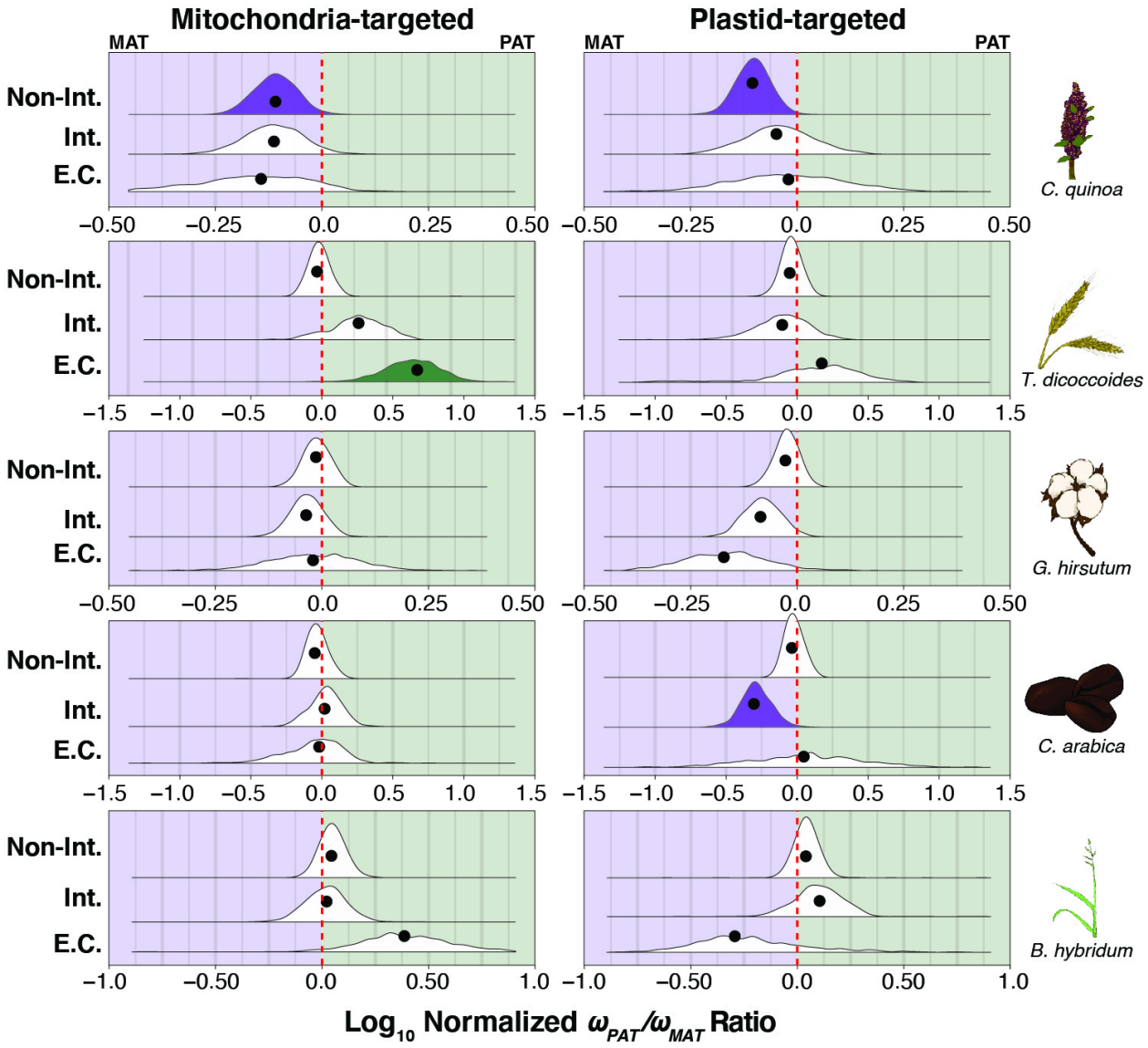

Figure S11. Ratios of maternal vs. paternal *ω* values in organelle-targeted genes, identical quintets only. Log-transformed ratios of maternal vs. paternal *ω* values for concatenations (black circles) and underlying bootstrap distributions (density curves) of mitochondria- (left) and plastid-targeted (right) genes, including only quintets that were identical across phylogenetic and syntenic methods. Species panels are arranged vertically from oldest (top) to youngest (bottom). Tobacco was excluded from this analysis because it produced so few syntenic quintets. The red-dashed line indicates the $\omega_{PAT}/\omega_{MAT}$ ratio for a concatenation of genes not-targeted to the organelles (Figure S9). Ratios left of the red line indicate higher *ω* values in the maternal subgenome, and ratios right of the red line indicate higher *ω* values in the paternal subgenome, after accounting for genome-wide patterns. Bootstrap distributions of *ω* ratios that depart significantly ($p < 0.05$) from the red line are filled in solid according to the direction of

subgenomic bias (i.e., green: normalized $\omega_{PAT}/\omega_{MAT} > 1.0$; purple: normalized $\omega_{PAT}/\omega_{MAT} < 1.0$; no fill: normalized $\omega_{PAT}/\omega_{MAT} \approx 1.0$). The intimacy of interactions are indicated on the y-axis from low or no interaction with organelle gene products (top), to interacting genes (middle), to genes involved in mitochondrial or plastid enzyme complexes (bottom).

ACKNOWLEDGEMENTS

This work was funded by the National Science Foundations Plant Genome Resources Program (IOS-1829176). We made extensive use of resources from the University of Colorado Boulder Research Computing Group, which is supported by the National Science Foundation (awards ACI-1532235 and ACI-1532236), the University of Colorado Boulder, and Colorado State University. We also thank the Iowa State University ResearchIT Unit for computational support. We thank Jeff Maughan, David Jarvis, Rick Jellen, and Mark Tester for access to the quinoa genomes, discussions of appropriate diploid models, and discussions of orthology inference in the face of homoeologous exchange. We thank Aaron Davis for discussions about polyploidy in *Coffea*, selection of an appropriate outgroup, and a great cup of coffee. We thank John Lovell, Jeremy Schmutz, Pilar Catalan, Sergio Gálvez Rojas, and Robert Hasterok for making the *Brachypodium* genome assemblies available in advance of their recent publication and for general discussions relating to *Brachypodium* genomics. We thank Evan Forsythe and the rest of the Sloan lab for helpful discussion of methods for orthology inference, especially relating to subtree extraction and branch trimming.

REFERENCES

1. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, et al. Ancestral polyploidy in seed plants and angiosperms. Nature. 2011;473: 97–100. doi:10.1038/nature09916

2. Wendel JF. The wondrous cycles of polyploidy in plants. Am J Bot. 2015;102: 1753–1756. doi:10.3732/ajb.1500320

3. Ruprecht C, Lohaus R, Vanneste K, Mutwil M, Nikoloski Z, Van de Peer Y, et al. Revisiting ancestral polyploidy in plants. Sci Adv. 2017;3: e1603195. doi:10.1126/sciadv.1603195

4. One Thousand Plant Transcriptomes Initiative. One thousand plant transcriptomes and the phylogenomics of green plants. Nature. 2019;574: 679–685. doi:10.1038/s41586-019-1693-2

5. Anssour S, Krügel T, Sharbel TF, Saluz HP, Bonaventure G, Baldwin IT. Phenotypic, genetic and genomic consequences of natural and synthetic polyploidization of *Nicotiana attenuata* and *Nicotiana obtusifolia*. Ann Bot. 2009;103: 1207–1217. doi:10.1093/aob/mcp058

6. Schnable JC, Pedersen BS, Subramaniam S, Freeling M. Dose–sensitivity, conserved non-coding sequences, and duplicate gene retention through multiple tetraploidies in the grasses. Front Plant Sci. 2011;2. doi:10.3389/fpls.2011.00002

7. Buggs RJA, Chamala S, Wu W, Tate JA, Schnable PS, Soltis DE, et al. Rapid, repeated, and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. Curr Biol. 2012;22: 248–252. doi:10.1016/j.cub.2011.12.027

8. Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IAP, et al. The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nat Commun. 2014;5: 3930. doi:10.1038/ncomms4930

9. Mirzaghaderi G, Mason AS. Revisiting pivotal-differential genome evolution in wheat. Trends Plant Sci. 2017;22: 674–684. doi:10.1016/j.tplants.2017.06.003

10. 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: 258–268. doi:10.1038/s41477-018-0136-7

11. Wendel JF, Lisch D, Hu G, Mason AS. The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. Curr Opin Genet Dev. 2018;49: 1–7. doi:10.1016/j.gde.2018.01.004

12. Petit M, Guidat C, Daniel J, Denis E, Montoriol E, Bui QT, et al. Mobilization of retrotransposons in synthetic allotetraploid tobacco. New Phytol. 2010;186: 135–147. doi:10.1111/j.1469-8137.2009.03140.x

13. Gao C, Zhou G, Ma C, Zhai W, Zhang T, Liu Z, et al. Helitron-like transposons contributed to the mating system transition from out-crossing to self-fertilizing in polyploid *Brassica napus* L. Sci Rep. 2016;6: 33785. doi:10.1038/srep33785

14. Senerchia N, Felber F, North B, Sarr A, Guadagnuolo R, Parisod C. Differential introgression and reorganization of retrotransposons in hybrid zones between wild wheats. Mol Ecol. 2016;25: 2518–2528. doi:10.1111/mec.13515

15. Springer NM, Lisch D, Li Q. Creating order from chaos: epigenome dynamics in plants with complex genomes. Plant Cell. 2016;28: 314–325. doi:10.1105/tpc.15.00911

16. Vicient CM, Casacuberta JM. Impact of transposable elements on polyploid plant genomes. Ann Bot. 2017;120: 195–207. doi:10.1093/aob/mcx078

17. Nieto Feliner G, Casacuberta J, Wendel JF. Genomics of evolutionary novelty in hybrids and polyploids. Front Genet. 2020;11: 792. doi:10.3389/fgene.2020.00792

18. Chester M, Gallagher JP, Symonds VV, Cruz da Silva AV, Mavrodiev EV, Leitch AR, et al. Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (*Asteraceae*). Proc Natl Acad Sci U S A. 2012;109: 1176–1181. doi:10.1073/pnas.1112041109

19. Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, et al. Plant genetics. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science. 2014;345: 950–953. doi:10.1126/science.1253435

20. Guo H, Wang X, Gundlach H, Mayer KFX, Peterson DG, Scheffler BE, et al. Extensive and biased intergenomic nonreciprocal DNA exchanges shaped a nascent polyploid genome, *Gossypium* (cotton). Genetics. 2014;197: 1153–1163. doi:10.1534/genetics.114.166124

21. Jarvis DE, Ho YS, Lightfoot DJ, Schmöckel SM, Li B, Borm TJA, et al. The genome of *Chenopodium quinoa*. Nature. 2017;542: 307–312. doi:10.1038/nature21370

22. Chen S, Ren F, Zhang L, Liu Y, Chen X, Li Y, et al. Unstable allotetraploid tobacco genome due to frequent homeologous recombination, segmental deletion, and chromosome loss. Mol Plant. 2018;11: 914–927. doi:10.1016/j.molp.2018.04.009

23. Bertioli DJ, Jenkins J, Clevenger J, Dudchenko O, Gao D, Seijo G, et al. The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. Nat Genet. 2019;51: 877–884. doi:10.1038/s41588-019-0405-z

24. Li N, Xu C, Zhang A, Lv R, Meng X, Lin X, et al. DNA methylation repatterning accompanying hybridization, whole genome doubling and homoeolog exchange in nascent segmental rice allotetraploids. New Phytol. 2019;223: 979–992. doi:10.1111/nph.15820

25. Mason AS, Wendel JF. Homoeologous exchanges, segmental allopolyploidy, and polyploid genome evolution. Front Genet. 2020;11: 1014. doi:10.3389/fgene.2020.01014

26. Madlung A, Masuelli RW, Watson B, Reynolds SH, Davison J, Comai L. Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids. Plant Physiol. 2002;129: 733–746. doi:10.1104/pp.003095

27. Salmon A, Ainouche ML, Wendel JF. Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (*Poaceae*). Mol Ecol. 2005;14: 1163–1175. doi:10.1111/j.1365-294X.2005.02488.x

28. Shcherban AB, Badaeva ED, Amosova AV, Adonina IG, Salina EA. Genetic and epigenetic changes of rDNA in a synthetic allotetraploid, *Aegilops sharonensis* x *Ae. umbellulata*. Genome. 2008;51: 261–271. doi:10.1139/G08-006

29. Fulneček J, Matyášek R, Kovařík A. Faithful inheritance of cytosine methylation patterns in repeated sequences of the allotetraploid tobacco correlates with the expression of DNA methyltransferase gene families from both parental genomes. Mol Genet Genomics. 2009;281: 407–420. Available: https://idp.springer.com/authorize/casa?redirect_uri=https://link.springer.com/article/10.1007/s00438-008-0420-8&casa_token=qYU2GIr7O6MAAAAA:84hwr5hunnznDK_s8vJpPhog1l4Pu7Th2GDdjMLoGsX9vIanRl5GGvulgB51xCH1E69iNtkCOEyUdAo10Q

30. Akagi T, Henry IM, Kawai T, Comai L, Tao R. Epigenetic regulation of the sex determination gene MeGI in polyploid persimmon. Plant Cell. 2016;28: 2905–2915. doi:10.1105/tpc.16.00532

31. Chen J, Li E, Zhang X, Dong X, Lei L, Song W, et al. Genome-wide nucleosome occupancy and organization modulates the plasticity of gene transcriptional status in maize. Mol Plant. 2017;10: 962–974. doi:10.1016/j.molp.2017.05.001

32. Song Q, Zhang T, Stelly DM, Chen ZJ. Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. Genome Biol. 2017;18: 99. doi:10.1186/s13059-017-1229-8

33. Ding M, Chen ZJ. Epigenetic perspectives on the evolution and domestication of polyploid plant and crops. Curr Opin Plant Biol. 2018;42: 37–48. doi:10.1016/j.pbi.2018.02.003

34. Combes M-C, Dereeper A, Severac D, Bertrand B, Lashermes P. Contribution of subgenomes to the transcriptome and their intertwined regulation in the allopolyploid *Coffea arabica* grown at contrasted temperatures. New Phytol. 2013;200: 251–260. doi:10.1111/nph.12371

35. Akama S, Shimizu-Inatsugi R, Shimizu KK, Sese J. Genome-wide quantification of

homeolog expression ratio revealed nonstochastic gene regulation in synthetic allopolyploid *Arabidopsis*. Nucleic Acids Res. 2014;42: e46. doi:10.1093/nar/gkt1376

36. Hu G, Hovav R, Grover CE, Faigenboim-Doron A, Kadmon N, Page JT, et al. Evolutionary conservation and divergence of gene coexpression networks in *Gossypium* (Cotton) seeds. Genome Biol Evol. 2016;8: 3765–3783. doi:10.1093/gbe/evw280

37. Yang J, Liu D, Wang X, Ji C, Cheng F, Liu B, et al. The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. Nat Genet. 2016;48: 1225–1232. doi:10.1038/ng.3657

38. Edger PP, Smith R, McKain MR, Cooley AM, Vallejo-Marin M, Yuan Y, et al. Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. Plant Cell. 2017;29: 2150–2167. doi:10.1105/tpc.17.00010

39. Ramírez-González RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, et al. The transcriptional landscape of polyploid wheat. Science. 2018;361. doi:10.1126/science.aar6089

40. 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;9: 9100–9110. doi:10.1002/ece3.5455

41. Landis JB, Kurti A, Lawhorn AJ, Litt A, McCarthy EW. Differential gene expression with an emphasis on floral organ size differences in natural and synthetic polyploids of *Nicotiana tabacum* (*Solanaceae*). Genes. 2020;11. doi:10.3390/genes11091097

42. Stebbins GL. The significance of polyploidy in plant evolution. Am Nat. 1940;74: 54–66. doi:10.1086/280872

43. Levin DA. Polyploidy and novelty in flowering plants. Am Nat. 1983;122: 1–25. doi:10.1086/284115

44. Otto SP, Whitton J. Polyploid incidence and evolution. Annu Rev Genet. 2000;34: 401–437. doi:10.1146/annurev.genet.34.1.401

45. Ramsey J, Schemske DW. Neopolyploidy in flowering plants. 2003 [cited 29 Nov 2020]. doi:10.1146/annurev.ecolsys.33.010802.150437

46. Otto SP. The evolutionary consequences of polyploidy. Cell. 2007;131: 452–462. doi:10.1016/j.cell.2007.10.022

47. Leitch AR, Leitch IJ. Genomic plasticity and the diversity of polyploid plants.

Science. 2008;320: 481–483. doi:10.1126/science.1153585

48. Van de Peer Y, Fawcett JA, Proost S, Sterck L, Vandepoele K. The flowering world: a tale of duplications. Trends Plant Sci. 2009;14: 680–688. doi:10.1016/j.tplants.2009.09.001

49. Madlung A. Polyploidy and its effect on evolutionary success: old questions revisited with new tools. Heredity . 2013;110: 99–104. doi:10.1038/hdy.2012.79

50. Soltis DE, Visger CJ, Soltis PS. The polyploidy revolution then… and now: Stebbins revisited. Am J Bot. 2014;101: 1057–1078. Available: https://bsapubs.onlinelibrary.wiley.com/doi/abs/10.3732/ajb.1400178

51. Yang Y, Moore MJ, Brockington SF, Mikenas J, Olivieri J, Walker JF, et al. Improved transcriptome sampling pinpoints 26 ancient and more recent polyploidy events in *Caryophyllales*, including two allopolyploidy events. New Phytol. 2018;217: 855–870. doi:10.1111/nph.14812

52. 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: 1–52. doi:10.1086/700636

53. Bomblies K. When everything changes at once: finding a new normal after genome duplication. Proc Biol Sci. 2020;287: 20202154. doi:10.1098/rspb.2020.2154

54. Fox DT, Soltis DE, Soltis PS, Ashman T-L, Van de Peer Y. Polyploidy: a biological force from cells to ecosystems. Trends Cell Biol. 2020;30: 688–694. doi:10.1016/j.tcb.2020.06.006

55. Sharbrough J, Conover JL, Tate JA, Wendel JF, Sloan DB. Cytonuclear responses to genome doubling. Am J Bot. 2017;104: 1277–1280. doi:10.3732/ajb.1700293

56. Gong L, Salmon A, Yoo M-J, Grupp KK, Wang Z, Paterson AH, et al. The cytonuclear dimension of allopolyploid evolution: an example from cotton using rubisco. Mol Biol Evol. 2012;29: 3023–3036. doi:10.1093/molbev/mss110

57. Gong L, Olson M, Wendel JF. Cytonuclear evolution of rubisco in four allopolyploid lineages. Mol Biol Evol. 2014;31: 2624–2636. doi:10.1093/molbev/msu207

58. 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: e0144339. doi:10.1371/journal.pone.0144339

59. Wang X, Dong Q, Li X, Yuliang A, Yu Y, Li N, et al. Cytonuclear variation of Rubisco in synthesized rice hybrids and allotetraploids. Plant Genome. 2017;10. doi:10.3835/plantgenome2017.05.0041

60. Ferreira de Carvalho J, Lucas J, Deniot G, Falentin C, Filangi O, Gilet M, et al.

Cytonuclear interactions remain stable during allopolyploid evolution despite repeated whole-genome duplications in *Brassica*. Plant J. 2019;98: 434–447. doi:10.1111/tpj.14228

61. Zhai Y, Yu X, Zhu Z, Wang P, Meng Y, Zhao Q, et al. Nuclear-cytoplasmic coevolution analysis of RuBisCO in synthesized *Cucumis* allopolyploid. Genes. 2019;10. doi:10.3390/genes10110869

62. Li C, Wang X, Xiao Y, Sun X, Wang J, Yang X, et al. Co-evolution in hybrid genomes: nuclear-encoded rubisco small subunits and their plastid-targeting translocons accompanying sequential allopolyploidy events in *Triticum*. Mol Biol Evol. 2020. doi:10.1093/molbev/msaa158

63. Kleine T, Maier UG, Leister D. DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. Annu Rev Plant Biol. 2009;60: 115–138. doi:10.1146/annurev.arplant.043008.092119

64. Sloan DB, Warren JM, Williams AM, Wu Z, Abdel-Ghany SE, Chicco AJ, et al. Cytonuclear integration and co-evolution. Nat Rev Genet. 2018;19: 635–648. doi:10.1038/s41576-018-0035-9

65. Millar AH. The plant mitochondrial proteome. In: Šamaj J, Thelen JJ, editors. Plant Proteomics. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007. pp. 226–246. doi:10.1007/978-3-540-72617-3_15

66. van Wijk KJ, Baginsky S. Plastid proteomics in higher plants: current state and future goals. Plant Physiol. 2011;155: 1578–1588. doi:10.1104/pp.111.172932

67. Forsythe ES, Sharbrough J, Havird JC, Warren JM, Sloan DB. CyMIRA: The Cytonuclear Molecular Interactions Reference for Arabidopsis. Genome Biol Evol. 2019;11: 2194–2202. doi:10.1093/gbe/evz144

68. Day A, Madesis P. DNA replication, recombination, and repair in plastids. In: Bock R, editor. Cell and Molecular Biology of Plastids. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007. pp. 65–119. doi:10.1007/4735_2007_0231

69. Cupp JD, Nielsen BL. Minireview: DNA replication in plant mitochondria. Mitochondrion. 2014;19 Pt B: 231–237. doi:10.1016/j.mito.2014.03.008

70. Gualberto JM, Newton KJ. Plant mitochondrial genomes: dynamics and mechanisms of mutation. Annu Rev Plant Biol. 2017;68: 225–252. doi:10.1146/annurev-arplant-043015-112232

71. Morley SA, Ahmad N, Nielsen BL. Plant organelle genome replication. Plants. 2019;8. doi:10.3390/plants8100358

72. Woodson JD, Chory J. Coordination of gene expression between organellar and nuclear genomes. Nat Rev Genet. 2008;9: 383–395. doi:10.1038/nrg2348

73. Liere K, Weihe A, Börner T. The transcription machineries of plant mitochondria and chloroplasts: Composition, function, and regulation. J Plant Physiol. 2011;168: 1345–1360. doi:10.1016/j.jplph.2011.01.005

74. Weihe A, Liere K, Börner T. Transcription and transcription regulation in chloroplasts and mitochondria of higher plants. In: Bullerwell CE, editor. Organelle Genetics: Evolution of Organelle Genomes and Gene Expression. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. pp. 297–325. doi:10.1007/978-3-642-22380-8_12

75. Chan KX, Phua SY, Crisp P, McQuinn R, Pogson BJ. Learning the languages of the chloroplast: retrograde signaling and beyond. Annu Rev Plant Biol. 2016;67: 25–53. doi:10.1146/annurev-arplant-043015-111854

76. Huang S, Van Aken O, Schwarzländer M, Belt K, Millar AH. The roles of mitochondrial reactive oxygen species in cellular signaling and stress response in plants. Plant Physiol. 2016;171: 1551–1559. doi:10.1104/pp.16.00166

77. Richardson LGL, Singhal R, Schnell DJ. The integration of chloroplast protein targeting with plant developmental and stress responses. BMC Biol. 2017;15: 118. doi:10.1186/s12915-017-0458-3

78. Krupinska K, Blanco NE, Oetke S, Zottini M. Genome communication in plants mediated by organelle–nucleus-located proteins. Philosophical Transactions of the Royal Society B: Biological Sciences. 2020. p. 20190397. doi:10.1098/rstb.2019.0397

79. Pike TW, Blount JD, Bjerkeng B, Lindström J, Metcalfe NB. Carotenoids, oxidative stress and female mating preference for longer lived males. Proc Biol Sci. 2007;274: 1591–1596. doi:10.1098/rspb.2007.0317

80. Barreto FS, Burton RS. Elevated oxidative damage is correlated with reduced fitness in interpopulation hybrids of a marine copepod. Proc Biol Sci. 2013;280: 20131521. doi:10.1098/rspb.2013.1521

81. Dowling DK. Evolutionary perspectives on the links between mitochondrial genotype and disease phenotype. Biochim Biophys Acta. 2014;1840: 1393–1403. doi:10.1016/j.bbagen.2013.11.013

82. Kremnev D, Strand A. Plastid encoded RNA polymerase activity and expression of photosynthesis genes required for embryo and seed development in *Arabidopsis*. Front Plant Sci. 2014;5: 385. doi:10.3389/fpls.2014.00385

83. Hill GE, Havird JC, Sloan DB, Burton RS, Greening C, Dowling DK. Assessing the fitness consequences of mitonuclear interactions in natural populations. Biol Rev Camb Philos Soc. 2019;94: 1089–1104. doi:10.1111/brv.12493

84. Rand DM, Haney RA, Fry AJ. Cytonuclear coevolution: the genomics of

cooperation. Trends Ecol Evol. 2004;19: 645–653. doi:10.1016/j.tree.2004.10.003

85. Weng M-L, Ruhlman TA, Jansen RK. Plastid-nuclear interaction and accelerated coevolution in plastid ribosomal genes in *Geraniaceae*. Genome Biol Evol. 2016;8: 1824–1838. Available: https://academic.oup.com/gbe/article-abstract/8/6/1824/2574029

86. Havird JC, Trapp P, Miller CM, Bazos I, Sloan DB. Causes and consequences of rapidly evolving mtDNA in a plant lineage. Genome Biol Evol. 2017;9: 323–336. doi:10.1093/gbe/evx010

87. Barreto FS, Watson ET, Lima TG, Willett CS, Edmands S, Li W, et al. Genomic signatures of mitonuclear coevolution across populations of *Tigriopus californicus*. Nat Ecol Evol. 2018;2: 1250–1257. doi:10.1038/s41559-018-0588-1

88. Li C, Sun X, Conover JL, Zhang Z, Wang J, Wang X, et al. Cytonuclear coevolution following homoploid hybrid speciation in *Aegilops tauschii*. Mol Biol Evol. 2019;36: 341–349. doi:10.1093/molbev/msy215

89. Yan Z, Ye G, Werren JH. Evolutionary rate correlation between mitochondrial-encoded and mitochondria-associated nuclear-encoded proteins in insects. Mol Biol Evol. 2019;36: 1022–1036. doi:10.1093/molbev/msz036

90. Hill GE. Mitonuclear compensatory coevolution. Trends Genet. 2020;36: 403–414. doi:10.1016/j.tig.2020.03.002

91. Greiner S, Rauwolf U, Meurer J, Herrmann RG. The role of plastids in plant speciation. Mol Ecol. 2011;20: 671–691. doi:10.1111/j.1365-294X.2010.04984.x

92. Burton RS, Barreto FS. A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? Mol Ecol. 2012;21: 4942–4957. doi:10.1111/mec.12006

93. Hill GE. Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap. Ecol Evol. 2016;6: 5831–5842. doi:10.1002/ece3.2338

94. Sloan DB, Havird JC, Sharbrough J. The on-again, off-again relationship between mitochondrial genomes and species boundaries. Mol Ecol. 2017. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.13959

95. Postel Z, Touzet P. Cytonuclear genetic incompatibilities in plant speciation. Plants. 2020;9. doi:10.3390/plants9040487

96. Grant V. Plant speciation. Columbia University Press; 1981. Available: https://www.degruyter.com/abstract/title/548067

97. Wendel J, Doyle J. Polyploidy and evolution in plants. Plant diversity and evolution: genotypic and phenotypic variation in higher plants. pp. 97–117. doi:10.1079/9780851999043.0097

98. Doyle JJ, Sherman-Broyles S. Double trouble: taxonomy and definitions of polyploidy. The New phytologist. 2017. pp. 487–493. doi:10.1111/nph.14276

99. Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. RELAX: detecting relaxed selection in a phylogenetic framework. Mol Biol Evol. 2015;32: 820–832. doi:10.1093/molbev/msu400

100. Maheshwari S, Barbash DA. The genetics of hybrid incompatibilities. Annu Rev Genet. 2011;45: 331–355. doi:10.1146/annurev-genet-110410-132514

101. Song K, Lu P, Tang K, Osborn TC. Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. Proc Natl Acad Sci U S A. 1995;92: 7719–7723. doi:10.1073/pnas.92.17.7719

102. Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, et al. The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet. 2011;43: 1035–1039. doi:10.1038/ng.919

103. Kioukis A, Michalopoulou VA, Briers L, Pirintsos S, Studholme DJ, Pavlidis P, et al. Intraspecific diversification of the crop wild relative *Brassica cretica Lam.* using demographic model selection. BMC Genomics. 2020;21: 48. doi:10.1186/s12864-019-6439-x

104. Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, et al. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. Nature. 2012;492: 423–427. doi:10.1038/nature11798

105. Udall JA, Long E, Hanson C, Yuan D, Ramaraj T, Conover JL, et al. *De novo* genome sequence assemblies of *Gossypium raimondii* and *Gossypium turneri*. G3. 2019;9: 3079–3085. doi:10.1534/g3.119.400392

106. Udall JA, Long E, Ramaraj T, Conover JL, Yuan D, Grover CE, et al. The genome sequence of *Gossypioides kirkii* illustrates a descending dysploidy in plants. Front Plant Sci. 2019;10: 1541. doi:10.3389/fpls.2019.01541

107. Chen Z, Nie H, Grover CE, Wang Y, Li P, Wang M, et al. Entire nucleotide sequences of *Gossypium raimondii* and *G. arboreum* mitochondrial genomes revealed A-genome species as cytoplasmic donor of the allotetraploid species. Plant Biol. 2017;19: 484–493. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/plb.12536?casa_token=4DCjrKoIWJQAAAAA:nO5tYJg51kNWiWLseGCthuGrXUAUZkj5bwI_1hTZaaL3Hj1Wu9fkqj64oWqJeaoWvS6BgsOD9PtU2SzJ

108. Avni R, Nave M, Barad O, Baruch K, Twardziok SO, Gundlach H, et al. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science. 2017;357: 93–97. doi:10.1126/science.aan0032

109. Ling H-Q, Ma B, Shi X, Liu H, Dong L, Sun H, et al. Genome sequence of the

progenitor of wheat A subgenome *Triticum urartu*. Nature. 2018;557: 424–428. doi:10.1038/s41586-018-0108-0

110. Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, et al. A chromosome conformation capture ordered sequence of the barley genome. Nature. 2017;544: 427–433. doi:10.1038/nature22043

111. Zhu T, Wang L, Rodriguez JC, Deal KR, Avni R, Distelfeld A, et al. Improved genome sequence of wild emmer wheat Zavitan with the aid of optical maps. G3. 2019;9: 619–624. doi:10.1534/g3.118.200902

112. Luo M-C, Gu YQ, Puiu D, Wang H, Twardziok SO, Deal KR, et al. Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. Nature. 2017;551: 498–502. doi:10.1038/nature24486

113. Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EKS, et al. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat Genet. 2016;48: 438–446. doi:10.1038/ng.3517

114. Tran HTM, Ramaraj T, Furtado A, Lee LS, Henry RJ. Use of a draft genome of coffee (*Coffea arabica*) to identify SNPs associated with caffeine content. Plant Biotechnol J. 2018;16: 1756–1766. doi:10.1111/pbi.12912

115. Dereeper A, Bocs S, Rouard M, Guignon V, Ravel S, Tranchant-Dubreuil C, et al. The coffee genome hub: a resource for coffee genomes. Nucleic Acids Res. 2015;43: D1028–35. doi:10.1093/nar/gku1108

116. Denoeud F, Carretero-Paulet L, Dereeper A, Droc G, Guyot R, Pietrella M, et al. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. Science. 2014;345: 1181–1184. doi:10.1126/science.1255274

117. Xu Z, Pu X, Gao R, Demurtas OC, Fleck SJ, Richter M, et al. Tandem gene duplications drive divergent evolution of caffeine and crocin biosynthetic pathways in plants. BMC Biol. 2020;18: 63. doi:10.1186/s12915-020-00795-3

118. Edwards KD, Fernandez-Pozo N, Drake-Stowe K, Humphry M, Evans AD, Bombarely A, et al. A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. BMC Genomics. 2017;18: 448. doi:10.1186/s12864-017-3791-6

119. Sierro N, Battey JND, Ouadi S, Bovet L, Goepfert S, Bakaher N, et al. Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. Genome Biol. 2013;14: R60. doi:10.1186/gb-2013-14-6-r60

120. Mangelson H, Jarvis DE, Mollinedo P, Rollano-Penaloza OM, Palma-Encinas VD, Gomez-Pando LR, et al. The genome of *Chenopodium pallidicaule*: an emerging Andean super grain. Appl Plant Sci. 2019;7: e11300. doi:10.1002/aps3.11300

121. International Brachypodium Initiative. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature. 2010;463: 763–768. doi:10.1038/nature08747

122. Gordon SP, Contreras-Moreira B, Levy JJ, Djamei A, Czedik-Eysenberg A, Tartaglio VS, et al. Gradual polyploid genome evolution revealed by pan-genomic analysis of *Brachypodium hybridum* and its diploid progenitors. Nat Commun. 2020;11: 3670. doi:10.1038/s41467-020-17302-5

123. 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. Proceedings of the National Academy of Sciences. 2013. pp. 2898–2903. doi:10.1073/pnas.1300127110

124. 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. The Plant Cell. 2016. pp. 326–344. doi:10.1105/tpc.15.00877

125. Duchêne A-M, Giritch A, Hoffmann B, Cognat V, Lancelin D, Peeters NM, et al. Dual targeting is the rule for organellar aminoacyl-tRNA synthetases in *Arabidopsis thaliana*. Proc Natl Acad Sci U S A. 2005;102: 16484–16489. doi:10.1073/pnas.0504682102

126. Perutz MF, Kendrew JC, Watson HC. Structure and function of haemoglobin: II. Some relations between polypeptide chain configuration and amino acid sequence. J Mol Biol. 1965;13: 669–678. doi:10.1016/S0022-2836(65)80134-6

127. Grantham R. Amino acid difference formula to help explain protein evolution. Science. 1974;185: 862–864. doi:10.1126/science.185.4154.862

128. Lesk AM, Chothia C. How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. J Mol Biol. 1980;136: 225–270. doi:10.1016/0022-2836(80)90373-3

129. Nakashima H, Nishikawa K, Ooi T. The folding type of a protein is relevant to the amino acid composition. J Biochem. 1986;99: 153–162. doi:10.1093/oxfordjournals.jbchem.a135454

130. Rumbley J, Hoang L, Mayne L, Englander SW. An amino acid code for protein folding. Proc Natl Acad Sci U S A. 2001;98: 105–112. doi:10.1073/pnas.98.1.105

131. Boyko AR, Williamson SH, Indap AR, Degenhardt JD, Hernandez RD, Lohmueller KE, et al. Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet. 2008;4: e1000083. doi:10.1371/journal.pgen.1000083

132. Sharbrough J, Luse M, Boore JL, Logsdon JM Jr, Neiman M. Radical amino acid mutations persist longer in the absence of sex. Evolution. 2018;72: 808–824.

doi:10.1111/evo.13465

133. Gaeta RT, Chris Pires J. Homoeologous recombination in allopolyploids: the polyploid ratchet. New Phytol. 2010;186: 18–28. doi:10.1111/j.1469-8137.2009.03089.x

134. Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. Why highly expressed proteins evolve slowly. Proc Natl Acad Sci U S A. 2005;102: 14338–14343. doi:10.1073/pnas.0504070102

135. Yang J-R, Liao B-Y, Zhuang S-M, Zhang J. Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. Proc Natl Acad Sci U S A. 2012;109: E831–40. doi:10.1073/pnas.1117408109

136. Hill WG, Robertson A. The effect of linkage on limits to artificial selection. Genet Res. 1966;8: 269–294. Available: https://www.ncbi.nlm.nih.gov/pubmed/5980116

137. Felsenstein J. The evolutionary advantage of recombination. Genetics. 1974;78: 737–756. Available: https://www.ncbi.nlm.nih.gov/pubmed/4448362

138. Zhou Y, Massonnet M, Sanjak JS, Cantu D, Gaut BS. Evolutionary genomics of grape (*Vitis vinifera ssp. vinifera*) domestication. Proc Natl Acad Sci U S A. 2017;114: 11715–11720. doi:10.1073/pnas.1709257114

139. Liu Q, Zhou Y, Morrell PL, Gaut BS. Deleterious variants in asian rice and the potential cost of domestication. Mol Biol Evol. 2017;34: 908–924. doi:10.1093/molbev/msw296

140. Hollister JD, Gaut BS. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res. 2009;19: 1419–1428. doi:10.1101/gr.091678.109

141. Freeling M, Woodhouse MR, Subramaniam S, Turco G, Lisch D, Schnable JC. Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. Curr Opin Plant Biol. 2012;15: 131–139. doi:10.1016/j.pbi.2012.01.015

142. Bird KA, VanBuren R, Puzey JR, Edger PP. The causes and consequences of subgenome dominance in hybrids and recent polyploids. New Phytol. 2018;220: 87–93. doi:10.1111/nph.15256

143. Wyler M, Stritt C, Walser J-C, Baroux C, Roulin AC. Impact of transposable elements on methylation and gene expression across natural accessions of Brachypodium distachyon. doi:10.1101/2020.06.16.154047

144. Charlesworth B. Effective population size and patterns of molecular evolution and variation. Nat Rev Genet. 2009;10: 195–205. doi:10.1038/nrg2526

145. Koh J, Soltis PS, Soltis DE. Homeolog loss and expression changes in natural populations of the recently and repeatedly formed allotetraploid *Tragopogon mirus* (*Asteraceae*). BMC Genomics. 2010;11: 97. doi:10.1186/1471-2164-11-97

146. Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C, et al. Impact of transposable elements on the organization and function of allopolyploid genomes. New Phytol. 2010;186: 37–45. doi:10.1111/j.1469-8137.2009.03096.x

147. Szadkowski E, Eber F, Huteau V, Lodé M, Huneau C, Belcram H, et al. The first meiosis of resynthesized *Brassica napus*, a genome blender. New Phytol. 2010;186: 102–112. doi:10.1111/j.1469-8137.2010.03182.x

148. Song MJ, Potter BI, Doyle JJ, Coate JE. Gene balance predicts transcriptional responses immediately following ploidy change in *Arabidopsis thaliana*. Plant Cell. 2020;32: 1434–1448. doi:10.1105/tpc.19.00832

149. Adams KL, Cronn R, Percifield R, Wendel JF. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc Natl Acad Sci U S A. 2003;100: 4649–4654. doi:10.1073/pnas.0630618100

150. Mutti JS, Bhullar RK, Gill KS. Evolution of gene expression balance among homeologs of natural polyploids. G3. 2017;7: 1225–1237. doi:10.1534/g3.116.038711

151. Emery M, Willis MMS, Hao Y, Barry K, Oakgrove K, Peng Y, et al. Preferential retention of genes from one parental genome after polyploidy illustrates the nature and scope of the genomic conflicts induced by hybridization. PLoS Genet. 2018;14: e1007267. doi:10.1371/journal.pgen.1007267

152. Wicker T, Gundlach H, Spannagl M, Uauy C, Borrill P, Ramírez-González RH, et al. Impact of transposable elements on genome structure and evolution in bread wheat. Genome Biol. 2018;19: 103. doi:10.1186/s13059-018-1479-0

153. Liu C, Wang J, Sun P, Yu J, Meng F, Zhang Z, et al. Illegitimate recombination between homeologous genes in wheat genome. Front Plant Sci. 2020;11: 1076. doi:10.3389/fpls.2020.01076

154. Drake JW, Charlesworth B, Charlesworth D, Crow JF. Rates of spontaneous mutation. Genetics. 1998;148: 1667–1686. Available: https://www.ncbi.nlm.nih.gov/pubmed/9560386

155. Baer CF, Miyamoto MM, Denver DR. Mutation rate variation in multicellular eukaryotes: causes and consequences. Nat Rev Genet. 2007;8: 619–631. doi:10.1038/nrg2158

156. Lynch M. Evolution of the mutation rate. Trends Genet. 2010;26: 345–352. doi:10.1016/j.tig.2010.05.003

157. Weng M-L, Becker C, Hildebrandt J, Neumann M, Rutter MT, Shaw RG, et al. Fine-grained analysis of spontaneous mutation spectrum and frequency in *Arabidopsis thaliana*. Genetics. 2019;211: 703–714. doi:10.1534/genetics.118.301721

158. Pelé A, Rousseau-Gueutin M, Chèvre A-M. Speciation success of polyploid plants closely relates to the regulation of meiotic recombination. Front Plant Sci. 2018;9: 907. doi:10.3389/fpls.2018.00907

159. Chen ZJ. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Annu Rev Plant Biol. 2007;58: 377–406. doi:10.1146/annurev.arplant.58.032806.103835

160. Akhunova AR, Matniyazov RT, Liang H, Akhunov ED. Homoeolog-specific transcriptional bias in allopolyploid wheat. BMC Genomics. 2010;11: 505. doi:10.1186/1471-2164-11-505

161. Flagel LE, Wendel JF. Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. New Phytol. 2010;186: 184–193. doi:10.1111/j.1469-8137.2009.03107.x

162. 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: 966–971. doi:10.1111/j.1469-8137.2012.04365.x

163. Yoo M-J, Szadkowski E, Wendel JF. Homoeolog expression bias and expression level dominance in allopolyploid cotton. Heredity . 2013;110: 171–180. doi:10.1038/hdy.2012.94

164. Li A, Liu D, Wu J, Zhao X, Hao M, Geng S, et al. mRNA and Small RNA transcriptomes reveal insights into dynamic homoeolog regulation of allopolyploid heterosis in nascent hexaploid wheat. Plant Cell. 2014;26: 1878–1900. doi:10.1105/tpc.114.124388

165. Wang X, Zhang H, Li Y, Zhang Z, Li L, Liu B. Transcriptome asymmetry in synthetic and natural allotetraploid wheats, revealed by RNA-sequencing. New Phytol. 2016;209: 1264–1277. Available: https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/nph.13678

166. Nomaguchi T, Maeda Y, Yoshino T, Asahi T, Tirichine L, Bowler C, et al. Homoeolog expression bias in allopolyploid oleaginous marine diatom *Fistulifera solaris*. BMC Genomics. 2018;19: 330. doi:10.1186/s12864-018-4691-0

167. Gonzalgo ML, Jones PA. Mutagenic and epigenetic effects of DNA methylation. Mutat Res. 1997;386: 107–118. doi:10.1016/s1383-5742(96)00047-6

168. Bennetzen JL, Wang H. The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annu Rev Plant Biol. 2014;65:

505–530. doi:10.1146/annurev-arplant-050213-035811

169. Jinks-Robertson S, Bhagwat AS. Transcription-associated mutagenesis. Annu Rev Genet. 2014;48: 341–359. doi:10.1146/annurev-genet-120213-092015

170. Halldorsson BV, Palsson G, Stefansson OA, Jonsson H, Hardarson MT, Eggertsson HP, et al. Characterizing mutagenic effects of recombination through a sequence-level genetic map. Science. 2019;363. doi:10.1126/science.aau1043

171. Slotte T, Huang H, Lascoux M, Ceplitis A. Polyploid speciation did not confer instant reproductive isolation in *Capsella* (*Brassicaceae*). Mol Biol Evol. 2008;25: 1472–1481. doi:10.1093/molbev/msn092

172. Zohren J, Wang N, Kardailsky I, Borrell JS, Joecker A, Nichols RA, et al. Unidirectional diploid-tetraploid introgression among British birch trees with shifting ranges shown by restriction site-associated markers. Mol Ecol. 2016;25: 2413–2426. doi:10.1111/mec.13644

173. Denton RD, Morales AE, Gibbs HL. Genome-specific histories of divergence and introgression between an allopolyploid unisexual salamander lineage and two ancestral sexual species. Evolution. 2018. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/evo.13528

174. Novikova PY, Tsuchimatsu T, Simon S, Nizhynska V, Voronin V, Burns R, et al. Genome sequencing reveals the origin of the allotetraploid *Arabidopsis suecica*. Mol Biol Evol. 2017;34: 957–968. doi:10.1093/molbev/msw299

175. Xiong Z, Gaeta RT, Pires JC. Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. Proc Natl Acad Sci U S A. 2011;108: 7908–7913. doi:10.1073/pnas.1014138108

176. Albertin W, Marullo P. Polyploidy in fungi: evolution after whole-genome duplication. Proc Biol Sci. 2012;279: 2497–2509. doi:10.1098/rspb.2012.0434

177. Cenci A, Combes M-C, Lashermes P. Genome evolution in diploid and tetraploid *Coffea* species as revealed by comparative analysis of orthologous genome segments. Plant Mol Biol. 2012;78: 135–145. doi:10.1007/s11103-011-9852-3

178. Allendorf FW, Bassham S, Cresko WA, Limborg MT, Seeb LW, Seeb JE. Effects of crossovers between homeologs on inheritance and population genomics in polyploid-derived salmonid fishes. J Hered. 2015;106: 217–227. doi:10.1093/jhered/esv015

179. He Z, Wang L, Harper AL, Havlickova L, Pradhan AK, Parkin IAP, et al. Extensive homoeologous genome exchanges in allopolyploid crops revealed by mRNAseq-based visualization. Plant Biotechnology Journal. 2017. pp. 594–604. doi:10.1111/pbi.12657

180. Lloyd A, Blary A, Charif D, Charpentier C, Tran J, Balzergue S, et al. Homoeologous exchanges cause extensive dosage-dependent gene expression changes in an allopolyploid crop. New Phytol. 2018;217: 367–377. Available: https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/nph.14836

181. Edger PP, Poorten TJ, VanBuren R, Hardigan MA, Colle M, McKain MR, et al. Origin and evolution of the octoploid strawberry genome. Nat Genet. 2019;51: 541–547. doi:10.1038/s41588-019-0356-4

182. Wu Y, Lin F, Zhou Y, Wang J, Sun S, Wang B, et al. Genomic mosaicism due to homoeologous exchange generates extensive phenotypic diversity in nascent allopolyploids. Natl Sci Rev. 2020 [cited 28 Dec 2020]. doi:10.1093/nsr/nwaa277

183. Zhang Z, Gou X, Xun H, Bian Y, Ma X, Li J, et al. Homoeologous exchanges occur through intragenic recombination generating novel transcripts and proteins in wheat and other polyploids. Proc Natl Acad Sci U S A. 2020;117: 14561–14571. doi:10.1073/pnas.2003505117

184. Salmon A, Flagel L, Ying B, Udall JA, Wendel JF. Homoeologous nonreciprocal recombination in polyploid cotton. New Phytol. 2010;186: 123–134. doi:10.1111/j.1469-8137.2009.03093.x

185. Wendel JF, Schnabel A, Seelanan T. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proc Natl Acad Sci U S A. 1995;92: 280–284. doi:10.1073/pnas.92.1.280

186. Kovarik A, Matyasek R, Lim KY, Skalická K, Koukalová B, Knapp S, et al. Concerted evolution of 18–5.8–26S rDNA repeats in *Nicotiana* allotetraploids. Biol J Linn Soc Lond. 2004;82: 615–625. doi:10.1111/j.1095-8312.2004.00345.x

187. Kovarik A, Pires JC, Leitch AR, Lim KY, Sherwood AM, Matyasek R, et al. Rapid concerted evolution of nuclear ribosomal DNA in two *Tragopogon* allopolyploids of recent and recurrent origin. Genetics. 2005;169: 931–944. doi:10.1534/genetics.104.032839

188. Page JT, Huynh MD, Liechty ZS, Grupp K, Stelly D, Hulse AM, et al. Insights into the evolution of cotton diploids and polyploids from whole-genome re-sequencing. G3 . 2013;3: 1809–1818. doi:10.1534/g3.113.007229

189. Leal-Bertioli SCM, Godoy IJ, Santos JF, Doyle JJ, Guimarães PM, Abernathy BL, et al. Segmental allopolyploidy in action: Increasing diversity through polyploid hybridization and homoeologous recombination. Am J Bot. 2018;105: 1053–1066. doi:10.1002/ajb2.1112

190. Mandáková T, Pouch M, Brock JR, Al-Shehbaz IA, Lysak MA. Origin and evolution of diploid and allopolyploid *Camelina* genomes were aaccompanied by chromosome shattering. Plant Cell. 2019;31: 2596–2612. doi:10.1105/tpc.19.00366

191. Simone S, Lucile T, Gabriele DG, Davide S, Gabriele M, Vidotto M, et al. A single polyploidization event at the origin of the tetraploid genome of *Coffea arabica* is responsible for the extremely low genetic variation in wild and cultivated germplasm. Scientific Reports (Nature Publisher Group); London. 2020;10: s41598–020. doi:10.1038/s41598-020-61216-7

192. Monroe JG, Powell T, Price N, Mullen JL, Howard A, Evans K, et al. Drought adaptation in *Arabidopsis thaliana* by extensive genetic loss-of-function. Elife. 2018;7. doi:10.7554/eLife.41038

193. Qiu Y, Van Tay Y, Ruan Y, Adams KL. Divergence of duplicated genes by repeated partitioning of splice forms and subcellular localization. New Phytol. 2020;225: 1011–1022. doi:10.1111/nph.16148

194. Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci U S A. 1987;84: 9054–9058. doi:10.1073/pnas.84.24.9054

195. Osada N, Akashi H. Mitochondrial–nuclear interactions and accelerated compensatory evolution: evidence from the Primate Cytochrome c Oxidase Complex. Mol Biol Evol. 2011;29: 337–346. doi:10.1093/molbev/msr211

196. Barreto FS, Burton RS. Evidence for compensatory evolution of ribosomal proteins in response to rapid divergence of mitochondrial rRNA. Mol Biol Evol. 2013;30: 310–314. doi:10.1093/molbev/mss228

197. Sloan DB, Triant DA, Wu M, Taylor DR. Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. Mol Biol Evol. 2014;31: 673–682. doi:10.1093/molbev/mst259

198. 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: 3069–3081. doi:10.1111/evo.12808

199. Zhang J, Ruhlman TA, Sabir J, Blazier JC, Jansen RK. Coordinated rates of evolution between interacting plastid and nuclear genes in *Geraniaceae*. Plant Cell. 2015;27: 563–573. doi:10.1105/tpc.114.134353

200. Rockenbach K, Havird JC, Monroe JG, Triant DA, Taylor DR, Sloan DB. Positive selection in rapidly evolving plastid-nuclear enzyme complexes. Genetics. 2016;204: 1507–1522. doi:10.1534/genetics.116.188268

201. Weng M-L, Ruhlman TA, Jansen RK. Plastid–nuclear interaction and accelerated coevolution in plastid ribosomal genes in *Geraniaceae*. Genome Biol Evol. 2016;8: 1824–1838. doi:10.1093/gbe/evw115

202. Forsythe ES, Williams AM, Sloan DB. Genome-wide signatures of plastid-nuclear

coevolution point to repeated perturbations of plastid proteostasis systems across angiosperms. bioRxiv. 2020. Available: https://www.biorxiv.org/content/10.1101/2020.08.28.272872v1.abstract

203. Burton RS, Pereira RJ, Barreto FS. Cytonuclear genomic interactions and hybrid breakdown. 2013 [cited 29 Nov 2020]. doi:10.1146/annurev-ecolsys-110512-135758

204. Bock DG, Andrew RL, Rieseberg LH. On the adaptive value of cytoplasmic genomes in plants. Mol Ecol. 2014;23: 4899–4911. doi:10.1111/mec.12920

205. Dai B, Guo H, Huang C, Zhang X, Lin Z. Genomic heterozygosity and hybrid breakdown in cotton (*Gossypium*): different traits, different effects. BMC Genet. 2016;17: 58. doi:10.1186/s12863-016-0366-5

206. McElroy KE, Denton RD, Sharbrough J, Bankers L, Neiman M, Gibbs HL. Genome expression balance in a triploid trihybrid vertebrate. Genome Biol Evol. 2017;9: 968–980. doi:10.1093/gbe/evx059

207. Butterfass T. Cell volume ratios of natural and of induced tetraploid and diploid flowering plants. Cytologia. 1987;52: 309–316. doi:10.1508/cytologia.52.309

208. Beaulieu JM, Leitch IJ, Patel S, Pendharkar A, Knight CA. Genome size is a strong predictor of cell size and stomatal density in angiosperms. New Phytol. 2008;179: 975–986. doi:10.1111/j.1469-8137.2008.02528.x

209. Marshall WF, Young KD, Swaffer M, Wood E, Nurse P, Kimura A, et al. What determines cell size? BMC Biol. 2012;10: 101. doi:10.1186/1741-7007-10-101

210. Roddy AB, Théroux-Rancourt G, Abbo T, Benedetti JW, Brodersen CR, Castro M, et al. The scaling of genome size and cell size limits maximum rates of photosynthesis with implications for ecological strategies. doi:10.1101/619585

211. Rhoades MM, Dempsey E. Induction of chromosome doubling at meiosis by the elongate gene in maize. Genetics. 1966;54: 505–522. Available: https://www.ncbi.nlm.nih.gov/pubmed/17248322

212. Bingham ET. Stomatal chloroplasts in *Alfalfa* at four ploidy levels. Crop Sci. 1968;8: cropsci1968.0011183X000800040036x. doi:10.2135/cropsci1968.0011183X000800040036x

213. Krishnaswami R, Andal R. Stomatal chloroplast number in diploids and polyploids of *Gossypium*. Proceedings of the Indian Academy of Sciences-Section B. Springer; 1978. pp. 109–112. Available: https://idp.springer.com/authorize/casa?redirect_uri=https://link.springer.com/content/pdf/10.1007/BF03046960.pdf&casa_token=INURrB2DwtYAAAAA:3BJM2LHbGub_oB0bXcpk4tcnjBnPcwIxL31e3oV-1Icr5QiQ0yg4X2jaFaTpeggfz4Cxb7JNDqqbTcyl-g

214. Bowman CM. Copy numbers of chloroplast and nuclear genomes are proportional in mature mesophyll cells of *Triticum* and *Aegilops* species. Planta. 1986;167: 264–274. doi:10.1007/BF00391425

215. Warner DA, Edwards GE. Effects of polyploidy on photosynthesis. Photosynth Res. 1993;35: 135–147. doi:10.1007/BF00014744

216. Kawade K, Horiguchi G, Ishikawa N, Hirai MY, Tsukaya H. Promotion of chloroplast proliferation upon enhanced post-mitotic cell expansion in leaves. BMC Plant Biol. 2013;13: 143. doi:10.1186/1471-2229-13-143

217. Whiteway MS, Lee RW. Chloroplast DNA content increases with nuclear ploidy in *Chlamydomonas*. Mol Gen Genet. 1977. Available: https://idp.springer.com/authorize/casa?redirect_uri=https://link.springer.com/article/10.1007/BF00268681&casa_token=CF220yDsOCYAAAAA:4Ve7mwZZX4_52Q-p9I88rvCelQuxkns6LYhhdvKiDaFS7HCJFG5BSG0YoslRjtDMJ2oi1G7GkRRlzNmoaw

218. 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: 904–910. doi:10.1104/pp.69.4.904

219. Coate JE, Schreyer WM, Kum D, Doyle JJ. Robust cytonuclear coordination of transcription in nascent *Arabidopsis thaliana* autopolyploids. Genes. 2020;11. doi:10.3390/genes11020134

220. Gyorfy MF, Miller ER, Conover JL, Grover CE, Wendel JF, Sloan DB, et al. Nuclear-cytoplasmic balance: whole genome duplications induce elevated organellar genome copy number. bioRxiv. 2021. p. 2021.06.08.447629. doi:10.1101/2021.06.08.447629

221. Gordon SP, Contreras-Moreira B, Woods DP, Des Marais DL, Burgess D, Shu S, et al. Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. Nat Commun. 2017;8: 2184. doi:10.1038/s41467-017-02292-8

222. Gaitán A, Yepes M, Zimin A, Maldonado CE, Navarro L, Flórez C, et al. Release of the *Coffea arabica* variety Caturra genome and that of its maternal diploid ancestor *C. eugenioides* to provide a strong foundation for breeding and functional genomics studies in coffee. Plant and Animal Genome XXVII Conference (January 12-16, 2019). PAG; 2019. Available: https://pag.confex.com/pag/xxvii/meetingapp.cgi/Paper/37040

223. Jia Y, Yinhua JIA, Zhaoe PAN, Shoupu HE, Gong W, Geng X, et al. Genetic diversity and population structure of *Gossypium arboreum L.* collected in China. Journal of Cotton Research. 2018. doi:10.1186/s42397-018-0011-0

224. Saski CA, Scheffler BE, Hulse-Kemp AM, Liu B, Song Q, Ando A, et al. Sub

genome anchored physical frameworks of the allotetraploid Upland cotton (*Gossypium hirsutum L.*) genome, and an approach toward reference-grade assemblies of polyploids. Sci Rep. 2017;7: 15274. doi:10.1038/s41598-017-14885-w

225. Xu C, Jiao C, Sun H, Cai X, Wang X, Ge C, et al. Draft genome of spinach and transcriptome diversity of 120 *Spinacia* accessions. Nat Commun. 2017;8: 15275. doi:10.1038/ncomms15275

226. Hosmani PS, Flores-Gonzalez M, van de Geest H, Maumus F, Bakker LV, Schijlen E, et al. An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. Cold Spring Harbor Laboratory. 2019. p. 767764. doi:10.1101/767764

227. Cros J, Combes MC, Trouslot P, Anthony F, Hamon S, Charrier A, et al. Phylogenetic analysis of chloroplast DNA variation in *Coffea*. Mol Phylogenet Evol. 1998;9: 109–117. doi:10.1006/mpev.1997.0453

228. Wendel JF. New World tetraploid cottons contain Old World cytoplasm. Proc Natl Acad Sci U S A. 1989;86: 4132–4136. doi:10.1073/pnas.86.11.4132

229. Kolano B, McCann J, Orzechowska M, Siwinska D, Temsch E, Weiss-Schneeweiss H. Molecular and cytogenetic evidence for an allotetraploid origin of *Chenopodium quinoa* and *C. berlandieri* (*Amaranthacea*e). Mol Phylogenet Evol. 2016;100: 109–123. doi:10.1016/j.ympev.2016.04.009

230. Bland MM, Matzinger DF, Levings CS 3rd. Comparison of the mitochondrial genome of *Nicotiana tabacum* with its progenitor species. Theor Appl Genet. 1985;69: 535–541. doi:10.1007/BF00251100

231. Sasaki T, Yukawa Y, Miyamoto T. Identification of RNA editing sites in chloroplast transcripts from the maternal and paternal progenitors of tobacco (*Nicotiana tabacum*): comparative analysis shows the involvement of distinct trans-factors for *ndhB* editing. Mol Biol. 2003. Available: https://academic.oup.com/mbe/article-abstract/20/7/1028/1046678

232. Gornicki P, Zhu H, Wang J, Challa GS, Zhang Z, Gill BS, et al. The chloroplast view of the evolution of polyploid wheat. New Phytol. 2014;204: 704–714. doi:10.1111/nph.12931

233. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20: 238. doi:10.1186/s13059-019-1832-y

234. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30: 772–780. doi:10.1093/molbev/mst010

235. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003;52: 696–704. doi:10.1080/10635150390235520

236. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. 2012;9: 772. doi:10.1038/nmeth.2109

237. Sukumaran J, Holder MT. DendroPy: a Python library for phylogenetic computing. Bioinformatics. 2010;26: 1569–1571. doi:10.1093/bioinformatics/btq228

238. Conover JL, Sharbrough J, Wendel JF. pSONIC: Ploidy-aware Syntenic Orthologous Networks Identified via Collinearity. G3 Genes|Genomes|Genetics. 2021 [cited 6 Jul 2021]. doi:10.1093/g3journal/jkab170

239. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40: e49. doi:10.1093/nar/gkr1293

240. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000;17: 540–552. doi:10.1093/oxfordjournals.molbev.a026334

241. Bannai H, Tamada Y, Maruyama O, Nakai K, Miyano S. Extensive feature detection of N-terminal protein sorting signals. Bioinformatics. 2002;18: 298–305. doi:10.1093/bioinformatics/18.2.298

242. Sperschneider J, Catanzariti A-M, DeBoer K, Petre B, Gardiner DM, Singh KB, et al. LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. Sci Rep. 2017;7: 44598. doi:10.1038/srep44598

243. Small I, Peeters N, Legeai F, Lurin C. Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics. 2004;4: 1581–1590. doi:10.1002/pmic.200300776

244. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc. 2007;2: 953–971. doi:10.1038/nprot.2007.131

245. 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: 789–804. doi:10.1111/tpj.13415

246. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24: 1586–1591. doi:10.1093/molbev/msm088