Cytonuclear Evolution of Rubisco in Four Allopolyploid Lineages

Lei Gong,[1] Mischa Olson,[1,2] and Jonathan F. Wendel*,[1]

[1]Department of Ecology, Evolution and Organismal Biology, Iowa State University
[2]Department of Plant Biology, Cornell University
*Corresponding author: E-mail: jfw@iastate.edu.
Associate editor: Naoki Takebayashi

Abstract

Allopolyploidization in plants entails the merger of two divergent nuclear genomes, typically with only one set (usually maternal) of parental plastidial and mitochondrial genomes and with an altered cytonuclear stoichiometry. Thus, we might expect cytonuclear coevolution to be an important dimension of allopolyploid evolution. Here, we investigate cytonuclear coordination for the key chloroplast protein rubisco (ribulose 1,5-bisphosphate carboxylase/oxygenase), which is composed of nuclear-encoded, small subunits (SSUs) and plastid-encoded, large subunits. By studying gene composition and diversity as well as gene expression in four model allopolyploid lineages, _Arabidopsis_, _Arachis_, _Brassica_, and _Nicotiana_, we demonstrate that paralogous nuclear-encoded _rbcS_ genes within diploids are subject to homogenization via gene conversion and that such concerted evolution via gene conversion characterizes duplicated genes (homoeologs) at the polyploid level. Many gene conversions in the polyploids are intergenomic with respect to the diploid progenitor genomes, occur in functional domains of the homoeologous SSUs, and are directionally biased, such that the maternal amino acid states are favored. This consistent preferential maternal-to-paternal gene conversion is mirrored at the transcriptional level, with a uniform transcriptional bias of the maternal-like _rbcS_ homoeologs. These data, repeated among multiple diverse angiosperm genera for an important photosynthetic enzyme, suggest that cytonuclear coevolution may be mediated by intergenomic gene conversion and altered transcription of duplicated, now homoeologous nuclear genes.

Key words: polyploid, cytonuclear coordination, rubisco, homoeologous recombination, biased expression.

Introduction

Polyploidy is a prominent evolutionary process in plants, in which two or more parental genomes are combined into the same nucleus. Through multiplying a single genome or via combining divergent genomes, autopolyploids and allopolyploids are formed, respectively (Soltis and Soltis 2000; Wendel 2000; Wendel and Doyle 2005). Although ancient polyploidy characterizes all flowering plant lineages (Jiao et al. 2011), recent allopolyploidy is observed in many plant lineages, including such well-known examples as _Arabidopsis_, _Arachis_ (peanut), _Brassica_ (cabbage), _Nicotiana_ (tobacco), and _Gossypium_ (cotton). In each of these genera, cytogenetic and molecular evidence have revealed extant diploid species that most closely resemble the diploid parents of the allopolyploids (Koch et al. 2000; Inaba and Nishio 2002; Chase et al. 2003; Jakobsson et al. 2006; Seijo et al. 2007; Leitch et al. 2008; Higgins et al. 2012; Bertioli et al. 2013). Comparative analyses of different allopolyploid species and their extant diploid relatives reveal that polyploidization results in complex and fascinating changes at different biological levels, including genomic alterations (loss of genes and nongenic elements and homoeologous genomic exchanges) (Lim et al. 2007; Salmon et al. 2010; Buggs et al. 2012), nonadditive

gene expression including expression dominance and biased homoeolog expression (Hegarty et al. 2008; Rapp et al. 2009; Flagel and Wendel 2010; Grover et al. 2012; Buggs 2013; Yoo et al. 2013), and changes in epigenetic modifications (Wang et al. 2004; Madlung and Wendel 2013).

In addition to these dynamic responses to polyploidization, there are potential stoichiometric disruptions caused by the combination of two nuclear genomes but inheritance of only one set of progenitor organellar genomes (usually maternal), suggesting a cytonuclear dimension to allopolyploid evolution. Many aspects of cytonuclear coevolution have been considered for diploid plants and animals (Rand et al. 2004; Wolf 2009; Caruso et al. 2012; Burton et al. 2013), addressing a number of key topics such as the effects of cytonuclear interaction on population fitness (Caruso et al. 2012; Burton et al. 2013), the occurrence of compensatory coadaptive cytonuclear mutations (Rand et al. 2004), participation of cytonuclear coordination in hybrid breakdown (Burton et al. 2013), and cytonuclear-epistasis-controlled nuclear genome imprinting (Wolf 2009). To date, though, the special circumstances surrounding cytonuclear evolution in polyploids remains largely unexplored. Previously, we investigated how homoeologous nuclear genes of _Gossypium_ allopolyploids

Open Access

encoding subunits of one protein complex evolved in a new context where they need to interact with a subunit encoded by a gene from the plastome, inherited (in cotton) from only one of the two progenitor diploids (Gong et al. 2012). The model protein complex we utilized is Rubisco (Ribulose 1,5-bisphosphate carboxylase/oxygenase), an essential enzyme in carbon fixation during photosynthesis, which functions as octamer holoenzymes of small subunits (SSUs) encoded by a nuclear *rbcS* multigene family and large subunits (LSUs) encoded by a single plastid *rbcL* gene (Rodermel et al. 1996). After characterizing *rbcS* and *rbcL* genic compositions in *Gossypium*, we explored their cytonuclear coordination at the genomic level, showing postpolyploidy, intergenomic, maternal-to-paternal gene conversion between nuclear homoeologs (Gong et al. 2012), in the direction opposite to that exhibited overall in *Gossypium* polyploids (Salmon et al. 2010; Flagel et al. 2012; Paterson et al. 2012; Guo et al. 2014). At the transcriptional level, biased maternal *rbcS* homoeolog expression was also demonstrated.

Intrigued by these findings for *Gossypium*, we asked whether similar cytonuclear coordination would be observed as a general phenomenon for rubisco evolution in other polyploids. Toward that end, we selected four exemplary angiosperm polyploid lineages, *Arabidopsis*, *Arachis*, *Brassica*, and *Nicotiana*, each of which has a well-understood phylogeny with extant model diploids and stabilized descendant allopolyploids. The rubisco *rbcS* and *rbcL* genes in each lineage were characterized. Within each lineage, phylogenies were constructed for *rbcS* gene paralogs and orthologs in the diploid species and placed in the context of their species divergence. By analyzing the *rbcS* gene sequences in representative parental diploids and allopolyploids, we demonstrate a consistent pattern of postpolyploidy gene conversion among *rbcS* homoeologs. In addition, biased homoeolog expression of paternal homoeologs carrying maternal conversions was also confirmed in most polyploid species. These results have

general significance with respect to cytonuclear evolution in plant allopolyploids.

Results

Maternal Inheritance and Divergence among *rbcL* Genes

rbcL genes from diploid and polyploid species of all four polyploid lineages were cloned and sequenced (table 1). Except in *Brassica*, there are from 0.43% to 0.65% nonsynonymous substitutions between the LSU proteins of the parental diploid species in each lineage. As expected, each polyploid has the copy inherited from the maternal parents. In *Brassica*, no amino acid differences exist between the parental diploid species (table 1). Similar to observations for *rbcL* genes in diploid cottons (Gong et al. 2012), diverged amino acid residues cluster in the C-terminal α/β- barrel domain and/or N-terminal domains of LSU subunits (table 1), which together form the active sites for rubisco (Spreitzer and Salvucci 2002). Notably, amino acid substitutions are also observed in the middle regions following the C-terminal domains, where the LSUs interact with the SSUs (Spreitzer and Salvucci 2002; Spreitzer et al. 2005). These raise the possibility of coevolutionary pressures in allopolyploids that might inherit divergent parental SSUs.

rbcS Composition in Diploids

Prior to cloning *rbcS* homoeologs in the polyploids, we cloned parental *rbcS* genes and aligned these into orthologs for inferences of homoeology in the polyploids. As shown in the exemplary *rbcS* sequence alignment for *Arabidopsis* (fig. 1), gene structure (introns/exons) was ascertained using cloned cDNAs. *rbcS* genes in most genera have three exons separated by two introns, the latter accumulating most of the substitutions and indels (fig. 1 and supplementary figs. S1–S3, Supplementary Material online). In *Nicotiana*, however,

Table 1. Nonsynonymous Substitutions of *rbcL* Sequences in Species[a] of Four Polyploid Lineages.

	Arabidopsis		
Amino Acid Position	♀*Arabidopsis thaliana* (Columbia-0)	*A. suecica* (Sue16)	♂*A. arenosa* (Strecno)
318	I	I	V
458	T	T	R
464	I	I	V
	Arachis		
Amino Acid Position	♀*Arachis duranensis* (PI 219823)	*Arac. hypogaea* (PI 161303)	♂*Arac. ipaensis* (PI 468322)
2	M	M	I
3	L	L	S
260	G	G	E
	Brassica		
Amino acid position	♀*B. rapa* (PI649186)	*B. napus* (PI633141)	♂*B. oleracea* (PI385959)
		No nonsynonymous substitution	
	Nicotiana		
Amino acid position	♀*Nicotiana sylvestris* (A403750326)	*N. tabacum* (095-55)	♂*N. tomentosiformis* (NIC 479/84)
124	R	R	C
422	K	K	Q

[a]Accession listed beside each species names. Allopolyploids are shown in the central column, with maternal and paternal parents on the left and right, respectively.
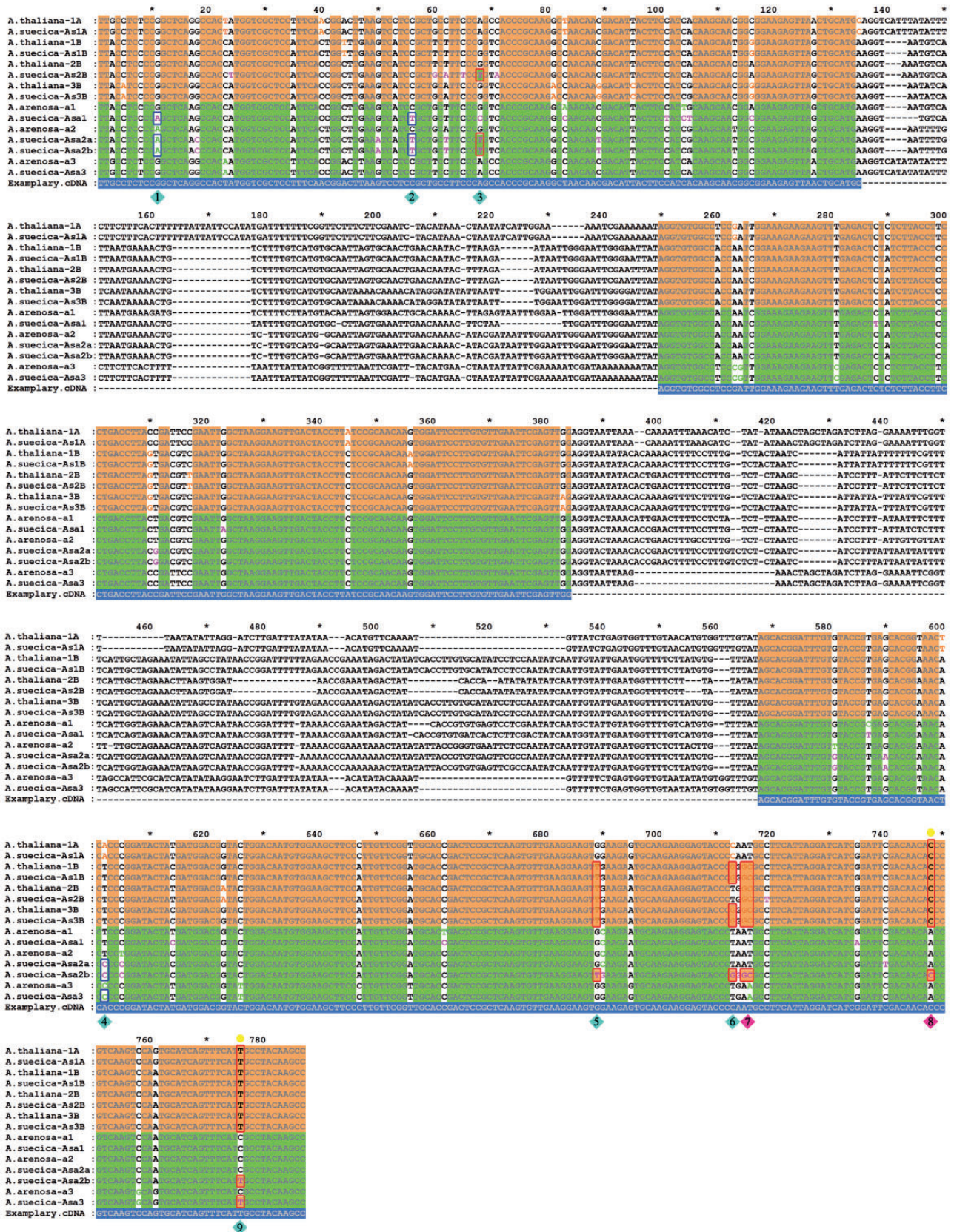
Fig. 1. Alignment of *Arabidopsis rbcS* orthologs and homoeologs with featured SNPs and gene conversion events highlighted in the exons. An exemplary cloned cDNA at the bottom (in light blue) is aligned with genomic *rbcS* homologs to ascertain *rbcS* exons/introns structure. Only featured SNPs and gene conversions in exonic regions are illustrated here. Conserved nucleotides in all orthologs and homoeologs are shown in gray. Homologs of maternal and paternal origins are highlighted in orange and green, respectively. Species-specific SNP positions (748 and 776) are marked by yellow ovals above the alignment blocks. Multiple genome-unique SNPs in diploid parental copies are shown in orange (maternal) and green (paternal) text.

there are three introns separating the coding region into four exons (supplementary fig. S3, Supplementary Material online). In exons of *rbcS* paralogs in each parental diploid, there are species-specific (consistent polymorphic substitution shared by all paralogs in the same species) and genome-unique (existing in a unique genome) single-nucleotide polymorphisms (SNPs), denoted in the exons of the alignment (fig. 1, supplementary figs. S1–S3, Supplementary Material online). Two groups of genome-unique SNPs are further recognized: Category I includes genome-unique SNPs present in at least two paralogs of a specific species; category II SNPs are carried by only one *rbcS* paralog (fig. 1, supplementary figs. S1–S3, Supplementary Material online; table 2). Species-specific SNPs shared by all paralogs of the same species were detected only in *Arabidopsis*, where the two species-specific SNPs have "C (Cytosine)" and "A (Adenine)" at the 748th position and "T (Thymine)" and "C (Cytosine)" at the 776th position in *Arabidopsis thaliana* and *A. arenosa*, respectively (fig. 1, table 2).

To compare the fixation rates of exonic, genome-unique SNPs, we tabulated their numbers in diploids of each lineage and included data generated previously for *Gossypium* species (supplementary fig. S4, Supplementary Material online, table 2). Because genome-unique SNPs in category I exist in multiple paralogs of the same diploid species, these SNPs are treated as nucleotide mutations that are fixed and spread by local gene conversions. As shown, the proportions of fixed genome-unique SNPs in category I are variable among lineages, ranging from 1.04% in *Arachis* to 4.47% in *Brassica* (table 2). This divergence is related to organismal divergence time (supplementary fig. S4, Supplementary Material online, table 2), with the notable exception of *Brassica*. For this genus, in which the progenitor diploids are thought to have diverged approximately 3.5 Ma (Higgins et al. 2012), a much higher proportion of genome-unique SNPs (4.47%) in category I is observed. This is significantly higher than in similarly aged *Arachis* (diverged 3.5 Ma, Seijo et al. 2007), older *Arabidopsis* (diverged 5 Ma, Jakobsson et al. 2006) and *Gossypium* (Wendel et al. 2010), or even the more ancient *Nicotiana* lineage (diverged 15 Ma, Leitch et al. 2008) (supplementary fig. S4, Supplementary Material online, and table 2). Possible explanations for this exceptional divergence in *Brassica* are discussed below. Accordingly, *Brassica* was not included in the correlation calculation but still is shown in the regression plot (supplementary fig. S4, Supplementary Material online). Apart from *Brassica*, a significant correlation was observed in fixation rate of exonic category I genome-unique SNPs ($R^2 = 0.53585$, P value < 0.05) (supplementary fig. S4, Supplementary Material online).

To understand the evolutionary history of the diploid *rbcS* orthologs, phylogenetic trees were constructed in the context of diploid species divergence within each lineage (fig. 2). In all cases, gene copy numbers are based on published genome sequences in conjunction with the cloning and sequence data. Unusually divergent *rbcS* paralogs are shown in blue, which includes orthologous groups 1A and a3 in *Arabidopsis*, A1 and B1 in *Arachis*, A1 and C1–C3 in *Brassica*, and S5 and T5 in *Nicotiana*. Because gene conversion at the diploid level has homogenized sequence pairs in many cases, the number of different gene copies is lower than the number of actual gene copies. In figure 2, homogenized copies are shown by interacting double helices. Among the species studied, the number of *rbcS* orthologs ranges from 4 to 12 (fig. 2). In some cases, autapomorphic substitutions arose following polyploidy, confirming the presence of gene converted and hence homogenized duplicates at the diploid level. There was no loss of any homoeolog in any of the four allopolyploids studied.

Gene Conversion Events Following Allopolyploidy

Comparison of each *rbcS* homoeolog with their parental orthologous copies revealed a number of autapomorphic nucleotide substitutions that have accumulated after formation of each polyploid (table 2). At the low end, in *Nicotiana tabacum*, 11 autapomorphic SNPs were detected, representing 2.40 % of the exonic nucleotide positions. The higher levels were for *A. suecica* and *Gossypium hirsutum* with the proportions 7.55% and 8.01% (table 2). As shown in supplementary figure S5, Supplementary Material online, the level of autapomorphic SNP presence is dependent on polyploid age; more recent polyploids have fewer SNPs. For example, in *Arachis hypogaea* and *Brassica napus*, polyploids of similar age ($> 5,000$ and $< 10,000$ years ago), almost equivalent proportions of autapomorphic SNPs are detected (3.73% and 3.35%, respectively). *Nicotiana tabacum*, a polyploid estimated as less than 200,000 years old, has an exceptionally small proportion of SNPs, whereas for *A. suecica* and *G. hirsutum*, the ancient polyploid species in our analysis (formed 12,000–300,000 years ago and 1–2 Ma, respectively) has the higher proportions of exonic autapomorphic SNPs (supplementary fig. S5, Supplementary Material online).

We inferred the parental origin of each homoeolog in the polyploids through comparisons with their diploid orthologs. We then inspected each homoeolog for genome-diagnostic SNPs from a different *rbcS* gene, mindful of the possibility (Gong et al. 2012) of intergenomic gene conversions. Alternatively, intragenomic gene conversions are implicated when they exclusively involve diagnostic SNPs among

Fɪɢ. 1. Continued

Autapomorphic substitutions in polyploid homoeologs are shown in pink. Inferred intra- and intersubgenomic gene conversion events are in blue and red boxes, respectively. For the positions involved in intersubgenomic gene conversion, the parental origin of each intersubgenomic converted nucleotide is illustrated by color (maternal origin: orange; paternal origin: green). Polyploid homoeologs with mosaic filled color boxes are the copies having intergenomic conversions. At the bottom of each alignment block, numbered gene conversion events resulting in synonymous/nonsynonymous substitutions are marked in blue and purple diamonds, respectively.

homoeologous copies of the same parental origin. A summary of these inferences of the intra- and intergenomic gene conversions is illustrated for each lineage (figs. 1 and 3, supplementary figs. S1–S3 and S6–S8, Supplementary Material online). Together with previous findings for *Gossypium* polyploids, we note several features of the inter/intragenomic gene conversions: 1) most conversion events were intergenomic (figs. 1 and 4 and supplementary figs. S1–S3, Supplementary Material online). Specifically, except for three intragenomic conversions in *A. suecica* (1st, 2nd, and 4th events among nine gene conversion events; fig. 1), there were no intragenomic conversions detected in other studied

polyploids, including *Gossypium* (Gong et al. 2012); 2) similar to the short *rbcS* genes in *Gossypium* (Gong et al. 2012), intergenomic events altered the originally identical *rbcS* duplicates (those linked by anastomosing lines in fig. 2) so they became distinguishable (different) at the polyploid level—for example, two identical paralogs in *A. arenosa* became two different homoeologs, *A. suecica*—Asa2a and *A. suecica*—Asa2b, when the latter copy obtained maternal diagnostic SNPs via 5th–9th intergenomic conversion events (fig. 1); and 3) most of the intergenomic events occurred in the paternal homoeologs, using templates from the maternal homoeologs (figs. 1, 4, and supplementary fig. S1–S3,

Table 2. Summary of Exonic Genome-Unique, Species-Specific, and Autapomorphic SNPs in Species of Five Polyploid Lineages.[a]

Lineage	Diploids		Polyploids
	Genome-Unique SNPs	Species-Specific SNPs	Autapomorphic SNPs
Arabidopsis	35 (7.14%) = 11 (2.24%) + 24 (4.90%)	748th and 776th	37 (7.55%)
Arachis	16 (3.32%) = 5 (1.04%) + 11 (2.28%)	None	18 (3.73%)
Brassica	54 (10.06%) = 24 (4.47%) + 30 (5.59%)	None	18 (3.35%)
Nicotiana	54 (11.79%) = 14 (3.06%) + 40 (8.73%)	None	11 (2.40%)
Gossypium	=26 (4.73%) = 24 (4.37%) + 2 (0.36%)	546th and 629th	44 (8.01%)

[a]Shown are the numbers and proportions of each SNP category across all sequenced exonic nucleotide positions.
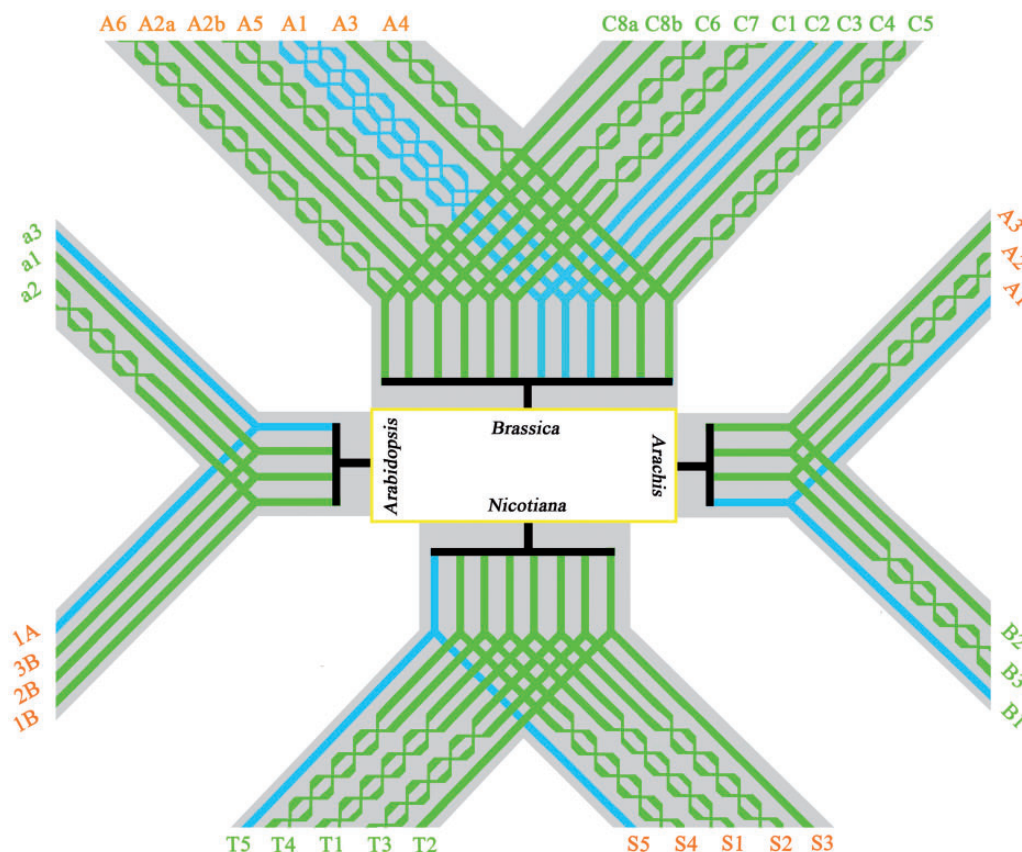

FIG. 2. Evolutionary history of *rbcS* genes in diploid species in four genera. Gene names in maternal and paternal diploid species are denoted in orange and green, respectively. Unusually divergent *rbcS* paralogs are shown in blue, which includes orthologous groups 1A and a3 in *Arabidopsis*, A1 and B1 in *Arachis*, A1 and C1–C3 in *Brassica*, and S5 and T5 in *Nicotiana*. Because gene conversion at the diploid level has homogenized sequence pairs in many cases, the number of different gene copies is lower than the number of actual gene copies; homogenized copies are shown by anastomosing double helices.
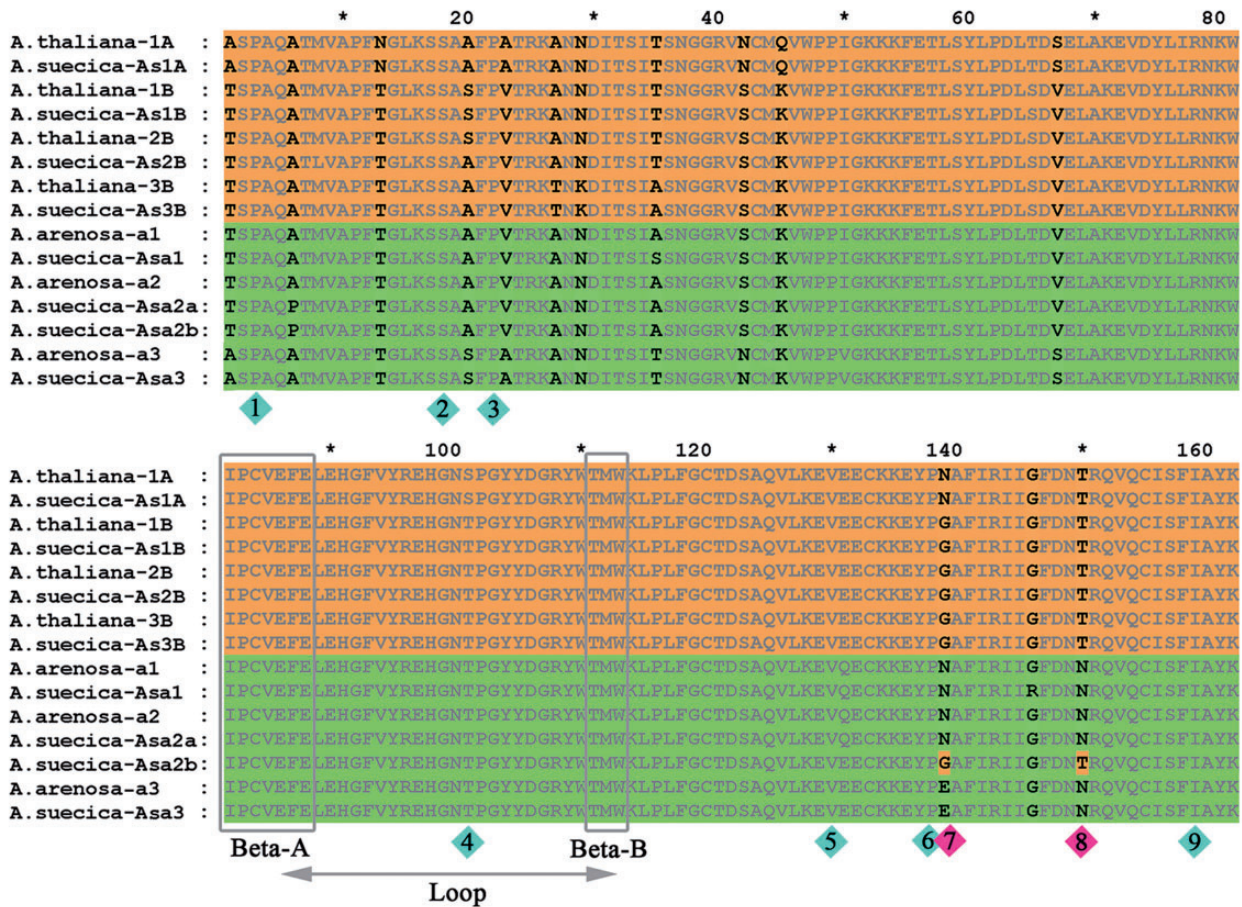
Fig. 3. Alignment of SSU proteins encoded by *rbcS* orthologs and homoeologs in *Arabidopsis* lineage. Maternal and paternal origin of each *rbcS* homolog is highlighted in orange and green color, respectively. Conserved amino acids are shown in gray, whereas polymorphic amino acid substitutions are in black. The synonymous/nonsynonymous substitutions caused by gene conversions are marked using different diamonds as in figure 1. Essential interface regions in SSUs, the predicted βA/βB loops where SSUs contacts with LSUs, are shown by open gray boxes.
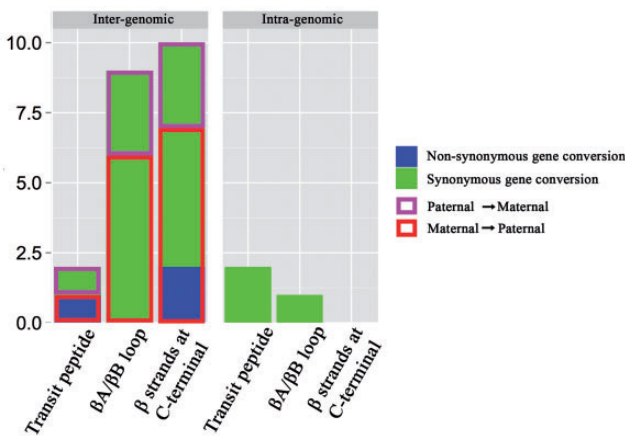

Fig. 4. Summary of gene conversions in multiple SSU domains. Gene conversion events among homoeologs from the same and different genomic origins, defined as intra- and intergenomic conversion events, are shown in the right and left panels, respectively. Within each functional SSU domain (on the x axis), the total numbers of conversion events introducing synonymous and nonsynonymous amino acid substitutions are denoted by green and blue bars, respectively. The pink and red frames around each green and blue bar highlight conversion directions, paternal to maternal (paternal state introduced into maternal homoeologs) and maternal-to-paternal (maternal state introduced into paternal homoeologs), respectively.

Supplementary Material online)—in other words, gene conversions occurred preferentially in the direction of introducing maternal-diagnostic SNPs into paternal homoeologs (simplified as "maternal-to-paternal" conversions). This is also the case in *Gossypium* polyploid species (Gong et al. 2012). Here, this is exemplified in *A. suecica*, where five of six intergenomic conversions entailed maternal-diagnostic SNPs detected in paternal homoeologs (fig. 1).

Protein sequences of all *rbcS* orthologs and homoeologs were predicted. Within the protein alignment, the aforementioned gene conversions were discovered to generate nonsynonymous amino acid substitutions only in *A. suecica* and *Arac. hypogaea*; most gene conversions did not result in amino acid changes (figs. 3, 4, and supplementary figs. S6–S8, Supplementary Material online). In *A. suecica*, the 7th and 8th conversion events brought maternal-specific "G (Glycine)" and "T (Threonine)" residues into the paternal homoeolog "*A. suecica*-Aa2b," in the process replacing the paternal amino acids "N (Asparagine)" at those two positions (fig. 3). Similarly, in *Arac. hypogaea*, the first conversion event caused nonsynonymous amino acid substitution in "*Arac. hypogaea*-AhB3b" homoeolog (supplementary fig. S6, Supplementary Material online).

Table 3. Comparisons of Homoeolog Expression in Five Polyploids.

Species	Homoeolog Pairs in Comparison[a]	Expression Differences[b]	Z Value = Difference/(Variance)$^{1/2}$	Significance (one side)
Arabidopsis Suecica	Asa2a vs. Asa2b	−327	−17.166	$P < 0.001$
Arachis hypogaea	AhB3a vs. AhB3b	−2,908[c]	−9.09	$P < 0.001$
Brassica napus	BnC6a vs. BnC6b	2,202	47.057	$P < 0.001$
Nicotiana tabacum	NtT3a vs. NtT3b	−4,559	−69.69	$P < 0.001$
	NtT4a vs. NtT4b	−4,166	−33.49	$P < 0.001$
Gossypium hirsutum	GhD-short1 vs. GhD-short2	−1,870[c]	−30.45	$P < 0.001$

[a]The homoeolog without maternal-to-paternal conversions is listed first.

[b]Negative expression differences are interpreted as biased expression of homoeolog copies with maternal-to-paternal gene conversions relative to the homoeologs without such conversions.

[c]Those two RNA sequencing experiment involved three biological replicates generated from mature leaves (Peggy Ozias-Akins, unpublished data and SRA056385 in Yoo et al. 2013). Expression difference shown is from one replicate of each experiment. Significant expression differences are consistently identified at the same P value level for all other replicates (not shown).

We summarized the distribution of types of gene conversion across the different SSU functional domains (fig. 4). SSU proteins were partitioned into four domains: Transit peptide (signaling peptide for pre-SSU targeting plastid and transportation into plastid); transit-loop interval region (mainly composed by α-helix A between signal peptide and βA/βB loop); βA/βB loop region (interface of SSU with LSU, which includes the β-strands and their enclosed loop); and all other β strands at the C-terminal end (Spreitzer and Salvucci 2002; Genkov and Spreitzer 2009; Kim et al. 2010). No gene conversion was detected in the transit-loop interval in any polyploid. Consequently, this region was excluded from the summary bar chart (fig. 4). In addition, the major intergenomic conversions preferentially occurred in the transit peptides and the C-terminal β strands rather than in the βA/βB loop region where SSUs interact with LSUs in the rubisco holoenzyme. Finally, in terms of the intergenomic conversion directions, the preferred "maternal-to-paternal" conversion events were detected in each SSU domain. All three nonsynonymous, intergenomic conversions introduced maternal amino acids into the paternal homoeologous SSUs (fig. 4).

Biased Expression of Paternal *rbcS* Homoeologs with Maternal-Converted Regions

To address whether there is biased homoeolog expression-related genomic origin of *rbcS* genes and if this is correlated with intergenomic gene conversions, we compared transcript levels for all polyploids (table 3). Homoeolog expressions were determined by multiplying the read coverage proportion of their specific SNPs by the total mapped *rbcS* reads (table 3 and supplementary table S4, Supplementary Material online). Within all polyploid species except *B. napus*, the paternal homoeologs with converted maternal segments were always significantly more highly expressed than their homoeologous counterparts without such intergenomic conversions (table 3). In contrast, in *B. napus*, the paternal homoeolog without gene conversion (BnC6a) had significantly higher expression than its counterpart paternal homoeolog (BnC6b) with maternal-to-paternal conversions (table 3).

Discussion

Here, we extend our results on cytonuclear coevolution of rubisco genes in *Gossypium* allopolyploids (Gong et al. 2012) to four other model allopolyploids, *Arabidopsis*, *Arachis*, *Brassica*, and *Nicotiana*. Our goal was to explore the extent to which the genic and transcriptional biases observed in cotton are mirrored in other allopolyploids and thereby gain insight into the generality of our indications of cytonuclear coevolution. Specifically, our aims were to discern the genic copy numbers and structures of nuclear *rbcS* genes in different genera, their propensity for "gene conversion" at both the diploid and allopolyploid levels, and the possible interplay between these dynamics and those of the plastid-encoded *rbcL* gene. We further wished to assess whether there is biased expression of homoeologs in other genera, how this relates to gene conversion, and the degree of similarities among multiple, phylogenetically dispersed angiosperm allopolyploids.

Potential Selection Pressure for Cytonuclear Coordination among *rbcS* Genes in Polyploids

rbcL is widely utilized as a slowly evolving plastid gene for purposes of phylogenetic reconstruction of angiosperm families and orders. Accordingly, we expected little sequence evolution among con-generic species and such is indeed the case for the data presented here (table 1). Yet several nonsynonymous differences are observed between *rbcL* genes from different diploid parents (except in *Brassica*), documenting maternal inheritance of the plastome in the allopolyploids, and indicating possible functional regions of LSUs that could conceivably apply selective pressure for optimization of biparentally inherited *rbcS*-derived SSU proteins. Specifically, during diploid divergence, the LSU in three of the four genera studied here accumulated several amino acid substitutions at both the C and N termini. Considering the C- and N- terminal domains are the catalytic centers and where the subunit interfaces with SSUs (Spreitzer et al. 2005; Genkov and Spreitzer 2009), the possibility exists that selection has operated on *rbcS* genes in the allopolyploid to optimize rubisco holoenzyme activity. As discussed below, the *rbcS* data are suggestive of this mechanism of compensation, for

most genera studied. Notably, *Brassica* is exceptional, with no amino acid divergence between parental LSUs, yet it too exhibits signatures of cytonuclear coevolution (see below), thereby implicating selection operating on other aspects of cytonuclear regulation.

Concerted Evolution of *rbcS* Genes in Diploid Species

In angiosperms studied to date, similarities of *rbcS* genes within species are often observed (Gong et al. 2012), especially among tandom *rbcS* paralogs, with lower similarities among physically dispersed *rbcS* paralogs. These observations, combined with phylogenetic evidence showing even lower similarities of *rbcS* orthologs in different species, have been taken as evidence that *rbcS* genes frequently are subjected to "concerted evolution" or sequence homogenization via gene conversion (Meagher et al. 1989; Clegg et al. 1997).

Concerted evolution is also evident in most species studied here (fig. 2). These inferences are based on two sources of information, that is, cloning and sequencing data, which provide diagnostic SNPs for *rbcS* paralogs, and genome sequence data, which provides gene number counts. The former includes both species-specific and genome-unique SNPs of the same genus. Species-specific SNPs reflect homogenization among paralogs within species, presumably from a gene conversion process that is evolutionarily sporadic. Interestingly, this process appears to be insufficiently frequent to completely homogenize paralogs but sufficiently common that its footprints are visible in the current suite of *rbcS* genes in each species. Similar results were previously reported for *Gossypium* (Gong et al. 2012). Genome-unique SNPs in each species, as described previously, can be further sorted into two categories, which have experienced distinct evolutionary histories. Category I includes genome-unique SNPs present in at least two paralogs of a specific species (table 2), which likely are derived from relatively recent homogenization via local/minor conversions among several rather than all paralogs. Category II includes most genome-unique SNPs (table 2), existing in single *rbcS* paralogs. We infer that these SNPs are the most recent substitutions generated in specific *rbcS* paralogs, such that they have not been homogenized across any other paralog. Possible mechanistic hypotheses for this failure to homogenize include recency of these SNPs relative to the pace of gene conversion, and/or spatial dispersal of these paralogs from other paralogs, so that the opportunities for gene conversion are lower.

Another interesting dimension of our data is the relatively consistent fixation rate of genome-unique SNPs in different lineages. Given a significant positive linear correlation of the proportion of category I genome-unique SNPs with divergence time in most genera (supplementary fig. S4, Supplementary Material online), the balance between nucleotide mutations in *rbcS* genes and their erasure via homogenization may generally be similar among plant lineages. This suggestion clearly will benefit from additional study using other plant genera. It may be, for example, that life history features such as mating system, population level dynamics, and effective population size create variation in this mutation

fixation balance. The higher fixation rates observed in the obligately outcrossing *Brassica*, for example, might reflect these factors (Wright et al. 2008; Ivanov and Gaude 2009).

One somewhat ironic observation is that in some cases, more *rbcS* genes are detectable at the allopolyploid than the diploid level. This reflects both the absence of gene loss following allopolyploidy and the evolution of novel SNPs postpolyploidy, which render previously identical paralogs (at the diploid level) nonidentical. For instance, in *Arabidopsis*, the similar but different *A.suecica*-Asa2a and *A.suecica*-Asa2b (corresponding to two identical *A. arenosa*-a2 paralogous copies) and in *Gossypium* (Gong et al. 2012), one more short-type *rbcS* homoeolog, are examples where different genes are observed at the polyploid level, caused by mutation being ahead of homogenization. A second example involves multiple distinct paralogs in one diploid species and a single group of identical paralogs in another diploid species, such as in *Brassica*, where there is orthology between *B. oleracea*-C8a and *B. oleracea*-C8b and two identical *B. rapa*-A6 genes, and between *B. rapa*-A2a and *B. rapa*-A2b and two identical *B. oleracea*-C6 genes (fig. 2).

The more extreme cases of escape from homogenization involve the near-independent *rbcS* copies in each diploid species of each lineage studied (blue lines in fig. 2). As proposed for *Gossypium*, this relative independence may be related to their distinct chromosomal locations (Gong et al. 2012). For example, in *A. thaliana*, three paralogs (1B, 2B, and 3B), with relative higher sequence similarities, are all located on chromosome 5, whereas the 1A paralog with the least similarity is on chromosome 1. In *Gossypium*, relatively independent long and short paralog groups are also clustered on chromosomes 11 and 1, respectively. In *Brassica*, three identical *B. rapa*-A1 copies and its three *B. oleracea*-C orthologs (-C1 to -C3) have the lowest sequence similarity to the other paralogs in each diploid species (fig. 2). A parsimonious explanation for this observation is that after the originally clustered gene copies translocated to new genomic regions in the common ancestor of *B. rapa* and *B. oleracea*, the three *rbcS* paralogs in *B. rapa* began to evolve independently from other *rbcS* paralogs, while still being subject to local gene conversion homogenization pressures; the other three gene copies in *B. oleracea*, however, came to be distinguishable via novel mutations. In brief, physical dispersal could protect independent copies from global homogenization.

Concerted Evolution of *rbcS* Homoeologs in Allopolyploids

Because allopolyploidy entails the merger of two sets of *rbcS* genes, gene conversion can, in principle, homogenize not only paralogs but also homoeologs. Notably, there are many autapomorphic SNPs in the allopolyploids, some identified in genomic conversion regions (shown as pink SNPs in each alignment file). Thus, these autapomorphic SNPs are new SNPs introduced by homogenization via gene conversion across homoeologs. These mutations appear to be related to the time since polyploidization, as the relatively older *A. suecica* and *G. hirsutum* have more of these SNPs than are

observed in the relatively younger *Arac. hypogaea* and *B. napus* sequences (supplementary fig. S5, Supplementary Material online).

Genomic Cytonuclear Coordination of *rbcS* in Allopolyploids

Intergenomic coadaptation or coordination between the nuclear and cytoplasmic organellar genomes is an essential component of evolutionarily successful hybridization events (Burton et al. 2013). Intergenomic interactions may be interrupted when hybridization occurs between genetically divergent populations, which combine divergent nuclear genomes with only a single set of cytoplasmic genomes. With respect to the rubisco complex, diverged nuclear *rbcS* homoeologs inherited from both parental species may be posited to be targets of selection following genome merger and doubling at the time of polyploid formation, in response to their new cellular milieu containing only the maternal cytoplasm.

As shown in *Gossypium*, one path toward reducing potential cytonuclear conflict is "maternal-to-paternal," intergenomic homogenization of *rbcS* homoeologs, presumably to stabilize or optimize rubisco holoenzyme activity. Specifically, in the N-terminal transit peptide region, which possesses the necessary information for SSU targeting and transport into the chloroplast (Bruce 2000; Lee et al. 2002), the potential relief from inefficient recognition and transport of paternal SSUs into the maternal chloroplast could conceivably be achieved by intergenomic, nonsynonymous gene conversions of paternally inherited *rbcS* copies. This possibility is exemplified by the 1st conversion event in *Arac. hypogaea*-AhB3b (supplementary figs. S1 and S7, Supplementary Material online). Similarly, at the C-terminal β-strands domain, which maintains holoenzyme structural stability and also potentially regulates LSU/SSU interactions (Esquível et al. 2002; Spreitzer and Salvucci 2002), paternal SSU homoeologs obtained maternal-like, C-terminal β-strands via intergenomic, nonsynonymous conversions both in *Gossypium* and the currently studied genera (fig. 4). This group of converted, paternal SSUs could also be favored during or after the assembly process with the maternal LSUs in the holoenzyme. However, in the βA/βB loop region where SSU proteins contact LSUs (Spreitzer et al. 2005; Genkov and Spreitzer 2009), there were no amino changes in the currently studied allopolyploids that were introduced by intergenomic conversions, so all paternal SSUs maintained their original protein sequences. Two scenarios can explain this observation: 1) paternal SSUs have sufficient compatibility with the cytoplasmic LSU at this interface region, such that fitness is not compromised and 2) insufficient time has elapsed for "more fit" genomic conversions to arise. In *G. hirsutum*, the loop regions of all divergent paternal SSUs have been replaced by the maternal loops via nonsynonymous, maternal-to-paternal gene conversions (Gong et al. 2012), suggesting an evolutionary future for these "caught in the act" younger allopolyploids. Targeting mutation experiments with artificial maternal-to-paternal conversions in the

βA/βB loop regions of paternal *rbcS* homoeologs would be interesting experiments to evaluate these scenarios.

The evidence presented here is consistent with, but does not prove, preferential selection for the products of intergenomic, maternal-to-paternal gene conversions (among intra, maternal-to-paternal or paternal-to-maternal events) across paternal nuclear homoeologs, followed by homogenization of the selected conversions across other copies originating from the paternal genome. Specially, for the polyploid species, in addition to the maternal-to-paternal intergenomic conversions, both intragenomic conversions and paternal-to-maternal, intergenomic conversions have probably occurred in paternal *rbcS* homoeologs following polyploidization, detected across different SSU domains (fig. 4). Cytonuclear co-evolutionary pressure may thus have preferentially selected intergenomic, maternal-to-paternal conversions. Given the relatively recent formation (<0.5 Ma) of all allopolyploids analyzed here, intragenomic and paternal-to-maternal conversions remain evident, perhaps having had insufficient time to homogenize the putatively beneficial maternal-to-paternal conversions across all *rbcS* copies, some of which retain their original parental diagnostic SNPs. In *Gossypium*, where polyploidy originated 1–2 Ma, the maternal, genome-specific SNPs have been homogenized across all paternal homoeologs. Additional evidence from other genera will further inform this possible evolutionary scenario.

A special case exists in *B. napus*, which inherited the maternal diploid LSUs with no amino acid divergence from the paternal LSUs. Given this observation, and the assumption that this would eliminate the possibility of selection at the level of SSU/LSU interaction, one might expect random interchanges among homoeologs irrespective of parental origin. Yet even in *Brassica* only intergenomic, maternal-to-paternal conversions were detected (in the paternal homoeolog, BnC6b). Relevant to this observation is the fact that SSU proteins need to be recognized by multiple cytoplasmic factors and transported to the surface membrane of the maternally derived plastid, where they are subjected to transmembrane transport into the plastids. It is possible that the gene conversion observed here reflects selection at this level, during some stage or process involved with maternal trans-membrane transport (Bruce 2000; Lee et al. 2002). Testing this idea is experimentally feasible, for example, through targeting mutations in the maternal-to-paternal conversion region in the BnC6b homoeolog and comparing its accumulated SSU proteins in the plastid stroma with SSU proteins from control *B. napus* individuals. At present, we are reporting an intriguing phenomenon that is suggestive of a newly described dimension of cytonuclear evolution.

Transcriptional Cytonuclear Coordination of *rbcS* Homoeologs

In addition to the gene sequence data and gene conversion evidence for cytonuclear accommodation to the polyploid state, we also explored gene expression levels to test whether there is biased expression of maternally derived *rbcS* genes. In three of the four allopolyploids (all but *B. napus*), relative to

the paternal homoeologs with no intergenomic conversion, paternal homoeologs with maternal-to-paternal conversions uniformly displayed preferential expression, consistent with our previous observations in *Gossypium* (Gong et al. 2012). This repeatedly observed, biased homoeolog expression among diverse allopolyploids is suggestive of selection at the level of transcript accumulation with a fitness advantage for SSU-encoding transcripts that carry maternal-like sequences. We note that biased expression of the paternal homoeolog with no maternal conversions was observed in *Brassica*. It could be explained by two possible scenarios: 1) relative weak selection of maternal LSU in plastid, which is identical with paternal LSU; and 2) insufficient time for transcriptional selection to arise.

Here, we have explored two dimensions of possible coordination and regulation of rubisco component subunits following allopolyploidization in plant species. We have confirmed that concerted evolution among divergent ancestral, duplicated copies of *rbcS* genes is a consistent feature of allopolyploid plants. We have shown that interparalog gene conversion is common at the diploid level and that it continues among homoeologs at the allopolyploid level, with a preferential occurrence of maternal-to-paternal, intergenomic conversions in signaling and regulatory domain of SSU genes. In most allopolyploids, this is accompanied by biased expression of paternal homoeologs carrying maternal-like gene conversions. Taken together, these data are consistent with cytonuclear selection following the reunion of two diverged genomes in a single cytoplasm as a consequence of allopolyploid speciation. Importantly, our analysis focuses only on cytonuclear coevolution of rubisco genes at the DNA and RNA levels; clearly much work remains for other potentially relevant dimensions of the problem, including studies of incorporation efficiency of divergent homoeologous SSUs into the rubisco holoenzyme, similar explorations in other cytonuclear coencoded complexes assembled in cytoplasmic organelles, stoichiometric changes in organelle and organellar genome abundances in each polyploid cell compared with the cells of their diploid parents, and many other dimensions of protein trafficking into organelles.

Materials and Methods

DNA and RNA Extraction and cDNA Synthesis

Four angiosperm polyploid lineages were selected, each of which included model progenitor diploids and derived allopolyploids (table 1). Fully expanded leaves of each species in each genus were sampled at the same developmental stages. After washing with Diethylpyrocarbonate (DEPC)-treated water, leaves were divided into two parts, which were used for DNA and RNA extraction, respectively. DNA extraction, RNA extraction, and cDNA synthesis were carried out following methods described previously (Gong et al. 2012).

Primer Design, Cloning, and Sequencing

rbcL is highly conserved among closed related species (Gielly and Taberlet 1994). We downloaded from National Center for Biotechnology Information (NCBI) all available *rbcL* genes in the genera studied for primer design (supplementary table S1, Supplementary Material online). For lineages not represented in the NCBI collection, the *rbcL* gene sequence from a closely related genus was used as the query sequence to BLASTn against the expressed sequence tags (ESTs) in PlantGDB (http://www.plantgdb.org/, last accessed April 28, 2014; supplementary table S1, Supplementary Material online). Manually aligned sequences of ESTs on the 5′- and 3′-end of the original BLAST query sequence (covering the start and stop codon, respectively) were used for primer design. Degenerate primers used to amplify full-length *rbcL* genes in each species are tabulated in supplementary table S1, Supplementary Material online.

Available genome assemblies of sequenced species and their ESTs deposited in PlantGDB were collected for *rbcS* primer design (supplementary table S2, Supplementary Material online). The *rbcS* genomic sequences of that species, or a related species in the same genus, or in some cases a related genus, were used as query sequences with BLASTn against genome assemblies or EST sequences. Significant homologous copies in each genome assembly or the manually aligned sequences of ESTs on the 5′- and 3′-end of the original BLAST query sequence were used for primer design. Primers specifically amplifying orthologs or homoeologs of *rbcS* in each species are tabulated in supplementary table S2, Supplementary Material online.

Polymerase chain reaction (PCRs) and PCR programs amplifying *rbcL* and *rbcS* genes, PCR product cloning, and sequencing followed the methods described earlier (Gong et al. 2012). Only the annealing temperature (at the initial step and in stabilized loops) was adjusted for each primer (supplementary table S2, Supplementary Material online). The same sets of primers designed above for amplifying genomic *rbcS* genes were also used for amplification of *rbcS* cDNAs. Final sequenced *rbcL* and *rbcS* genes were deposited into GenBank with accession numbers as KM025240–KM025251 and KM025252–KM025337, respectively.

When amplifying the *rbcS* genomic or transcript copies in each species, to avoid possible false *rbcS* PCR-recombination artifacts, three parallel independent PCRs were carried out for each primer sample. Only the *rbcS* copies, having at least 25% supportive clones sequenced in each independent PCR, were accepted as bona fide copies. Cloning and sequencing of the PCR products were also carried out as described (Gong et al. 2012).

Sequence Alignment and Phylogenetic Reconstruction

Sequences were aligned within each genus using the online MAFFT tool v7.122 (Katoh and Standley 2013). After manual adjustment, synonymous and nonsynonymous substitutions in the exons were noted. Phylogenetic histories of the *rbcS* multigene family in all diploids of each lineage were inferred based on parsimony analysis.

Detection of Homoeologous SNPs and Gene Conversion Events (Nonreciprocal Homoeologous Recombination)

For each genus, genome-diagnostic SNPs (including the species-specific and genome-unique SNPs) and autapomorphic SNPs were inferred in *rbcS* orthologs and homoeologs, respectively. Genome-diagnostic SNPs were used to determine the parental genomic origin of each homoeolog. Autapomorphic SNPs are defined as novel nucleotides arising at the polyploid level (in either homoeolog). Within the allopolyploid species, the possible exonic *rbcS* genomic conversion regions or points of "non-reciprocal recombination" (in only one direction from paternal to maternal homoeolog or vice versa; Salmon et al. 2010) were initially inferred using the GENECONV tool (automated recombination detection in triplet sequences), which is incorporated in RDP4 Beta 4.27 software (Sawyer 1989; Martin et al. 2010). Specifically, each *rbcS* homoeolog in the polyploids was searched against both reference diploid orthologs and other homoeologs: Any recombinations identified between homoeologs of the same genomic origin were inferred as intragenomic conversions, whereas those involving homoeologs of different genomic origin were accepted as products of intergenomic conversion events. Recombination detection program (RDP)-identified conversion copies were further processed by homemade Perl scripts, which tabulated the SNP information within converted homoeologs, as previously described (listing the coordinates in the alignment and nucleotide changes before and after the conversions; Gong et al. 2012). As noted previously, to avoid possible artificial PCR recombinants, only recombinants occurring in at least 25% of the total cloned sequences from each replicated PCR were accepted as true "gene conversion" copies.

Statistical Comparison of *rbcS* Homoeolog Transcript Level Based on RNAseq

Next-generation RNA sequencing data of all polyploids in four lineages were collected from SRA databases in NCBI and other resources (supplementary table S3, Supplementary Material online). Quality-filtered reads were mapped to all cloned *rbcS* homoeologs via Bowtie 1.0.0 with stringent perfect match control (Langmead et al. 2009). The final *rbcS* homoeolog-specific expression proportions were obtained by dividing the mapped reads covering all diagnostic homoeolog-specific SNPs of each homoeolog copy by the total reads mapped to those SNP positions in all expressed *rbcS* homoeologs. The coverage of each SNP in each homoeolog was obtained by running the mpileup module in the samtools package (Heng et al. 2009). The final observed *rbcS* homoeolog-specific expressions were obtained by multiplying their individual estimated expression proportion by the total mapped reads.

Given the high expression levels of *rbcS* genes in plant species, under the Central Limit Theorem (Rice 2006), a Z statistic evaluating the expression difference between homoeologs with intergenomic gene conversion and paralogous homoeologs without intergenomic gene conversions (here

abbreviated as "$H_{converted}$ vs. $H_{nonconversion}$"), was calculated as follows:

(1) The null hypothesis assumed no homoeolog expression difference in $H_{converted}$ versus $H_{nonconversion}$. Hence, the expectation of the expression difference was zero.

$$E(H_{converted} - H_{nonconversion}) = 0$$

(2) The variance of the homoeolog expression difference in $H_{converted}$ versus $H_{nonconversion}$ was derived:

$$Var(H_{converted} - H_{nonconversion}) = Var(H_{converted})$$
$$+ Var(H_{nonconversion})$$
$$- 2 \times Cov(H_{converted}, H_{nonconversion})$$

Under the assumption of the summed proportions of all *rbcS* homoeologs being 1, the probability of obtaining the observed combination of *rbcS* homoeolog expression should follow the multinomial distribution. According to the known variance of one variable and covariance of two component variables in the multinomial distribution,

$$Var(H_{converted}) = N_{total} \times p_{Hconverted} \times (1 - p_{Hconverted})$$

$$Var(H_{nonconversion}) = N_{total} \times p_{Hnonconversion}$$
$$\times (1 - p_{Hnonconversion})$$

$$Cov(H_{converted}, H_{nonconversion}) = -N_{total} \times p_{Hconverted}$$
$$\times p_{Hnonconversion,}$$

in which N_{total} was the total expression of all *rbcS* homoeologs in polyploid species, and $p_{Hconverted}$ and $p_{Hnonconversion}$ were expression proportions of homoeologs with and without intergenomic conversions.

(3) A final Z statistic was calculated with all terms replaced by their values calculated as above:

$$Z \text{ statistic} = \frac{\left\{ \begin{array}{c} Observed(H_{converted} - H_{nonconversion}) \\ - E(H_{converted} - H_{nonconversion}) \end{array} \right\}}{[Var(H_{nonconvertion} - H_{nonconversion})]^{1/2}}$$

P value of each estimated Z statistic was estimated based on the standard normal distribution.

Supplementary Material

Supplementary figures S1–S8 and tables S1–S4 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

References

Bertioli DJ, Vidigal B, Nielen S, Ratnaparkhe MB, Lee T-H, Leal-Bertioli SCM, Kim C, Guimaraes PM, Seijo G, Schwarzacher T, et al. 2013. The repetitive component of the A genome of peanut (*Arachis hypogaea*) and its role in remodelling intergenic sequence space since its evolutionary divergence from the B genome. *Ann Bot.* 112:545–559.

Bruce BD. 2000. Chloroplast transit peptides: structure, function and evolution. *Trends Cell Biol.* 10:440–447.

Buggs RJA. 2013. The consequences of polyploidy and hybridisation for transcriptome dynamics Unravelling gene expression of complex crop genomes. *Heredity (Edinb)* 110:97–98.

Buggs RJA, Chamala S, Wu W, Tate JA, Schnable PS, Soltis DE, Soltis PS, Barbazuk WB. 2012. Rapid, repeated, and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. *Curr Biol.* 22:248–252.

Burton RS, Pereira RJ, Barreto FS. 2013. Cytonuclear genomic interactions and hybrid breakdown. *Annu Rev Ecol Evol Syst.* 44:281–302.

Caruso CM, Case AL, Bailey MF. 2012. The evolutionary ecology of cytonuclear interactions in angiosperms. *Trends Plant Sci.* 17: 638–643.

Chase MW, Knapp S, Cox AV, Clarkson JJ, Butsko Y, Joseph J, Savolainen V, Parokonny AS. 2003. Molecular systematics, GISH and the origin of hybrid taxa in *Nicotiana* (Solanaceae). *Ann Bot.* 92:107–127.

Clegg MT, Cummings MP, Durbin ML. 1997. The evolution of plant nuclear genes. *Proc Natl Acad Sci U S A.* 94:7791–7798.

Esquível MG, Anwaruzzaman M, Spreitzer RJ. 2002. Deletion of nine carboxy-terminal residues of the rubisco small subunit decreases thermal stability but does not eliminate function. *FEBS Lett.* 520: 73–76.

Flagel LE, Wendel JF. 2010. Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. *New Phytol.* 186:184–193.

Flagel LE, Wendel JF, Udall JA. 2012. Duplicate gene evolution, homoeologous recombination, and transcriptome characterization in allopolyploid cotton. *BMC Genomics* 13:302–314.

Genkov T, Spreitzer RJ. 2009. Highly conserved small subunit residues influence rubisco large subunit catalysis. *J Biol Chem.* 284: 30105–30112.

Gielly L, Taberlet P. 1994. The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbcL* sequences. *Mol Biol Evol.* 11: 769–777.

Gong L, Salmon A, Yoo M-J, Grupp KK, Wang Z, Paterson AH, Wendel JF. 2012. The cytonuclear dimension of allopolyploid evolution: an example from cotton using rubisco. *Mol Biol Evol.* 29: 3023–3036.

Grover CE, Gallagher JP, Szadkowski EP, Yoo M-J, Flagel LE, Wendel JF. 2012. Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytol.* 196:966–971.

Guo H, Wang X, Gundlach H, Mayer KFX, Peterson DG, Scheffler BE, Chee PW, Paterson AH. 2014. Extensive and biased inter-genomic non-reciprocal DNA exchanges shaped a nascent polyploid genome, *Gossypium* (cotton). *Genetics* 197(4):1153–1163.

Hegarty MJ, Barker GL, Brennan AC, Edwards KJ, Abbott RJ, Hiscock SJ. 2008. Changes to gene expression associated with hybrid speciation in plants: further insights from transcriptomic studies in *Senecio*. *Philos Trans R Soc Lond B Biol Sci.* 363:3055–3069.

Heng L, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.

Higgins J, Magusin A, Trick M, Fraser F, Bancroft I. 2012. Use of mRNA-seq to discriminate contributions to the transcriptome from the constituent genomes of the polyploid crop species *Brassica napus*. *BMC Genomics* 13:247.

Inaba R, Nishio T. 2002. Phylogenetic analysis of Brassiceae based on the nucleotide sequences of the S-locus related gene, SLR1. *Theor Appl Genet.* 105:1159–1165.

Ivanov R, Gaude T. 2009. *Brassica* self-incompatibility: a glimpse below the surface. *Plant Signal Behav.* 4:996–998.

Jakobsson M, Hagenblad J, Tavare S, Sall T, Hallden C, Lind-Hallden C, Nordborg M. 2006. A unique recent origin of the allotetraploid species *Arabidopsis suecica*: evidence from nuclear DNA markers. *Mol Biol Evol.* 23:1217–1231.

Jiao Y, Wickett NJ, Chanderbali SAAS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473:97–100.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30:772–780.

Kim S, Lee D-S, Choi IS, Ahn S-J, Kim Y-H, Bae H-J. 2010. *Arabidopsis thaliana* Rubisco small subunit transit peptide increases the accumulation of *Thermotoga maritima* endoglucanase Cel5A in chloroplasts of transgenic tobacco plants. *Transgenic Res.* 19:489–497.

Koch MA, Haubold B, Mitchell-Olds T. 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis, Arabis*, and related genera (Brassicaceae). *Mol Biol Evol.* 17:1483–1498.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25.

Lee KH, Kim DH, Lee SW, Kim ZH, Hwang I. 2002. *In vivo* import experiments in protoplasts reveal the importance of the overall context but not specific amino acid residues of the transit peptide during import into chloroplasts. *Mol Cells.* 14:388–397.

Leitch IJ, Hanson L, Lim KY, Kovarik A, Chase MW, Clarkson JJ, Leitch AR. 2008. The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). *Ann Bot.* 101:805–814.

Lim KY, Kovarik A, Matyasek R, Chase MW, Clarkson JJ, Grandbastien MA, Leitch AR. 2007. Sequence of events leading to near-complete genome turnover in allopolyploid *Nicotiana* within five million years. *New Phytol.* 175:756–763.

Madlung A, Wendel JF. 2013. Genetic and epigenetic aspects of polyploid evolution in plants. *Cytogenet Genome Res.* 140:270–285.

Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P. 2010. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26:2462–2463.

Meagher RB, Berry-Lowe S, Rice K. 1989. Molecular evolution of the small subunit of ribulose bisphosphate carboxylase: nucleotide substitution and gene conversion. *Genetics* 123:845–863.

Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D, Showmaker KC, Shu S, Udall J, et al. 2012. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* 492:423–427.

Rand DM, Haney RA, Fry AJ. 2004. Cytonuclear coevolution: the genomics of cooperation. *Trends Ecol Evol.* 19:645–653.

Rapp RA, Udall JA, Wendel JF. 2009. Genomic expression dominance in allopolyploids. *BMC Biol.* 7:18–27.

Rice JA. 2006. Mathematical statistics and data analysis, 3rd ed. Belmont (CA): Duxbury Press.

Rodermel S, Haley J, Jiang C-Z, Tsai C-H, Bogorad L. 1996. A mechanism for intergenomic integration: abundance of ribulose bisphosphate carboxylase small-subunit protein influences the translation of the large-subunit mRNA. *Proc Natl Acad Sci U S A.* 93:3881–3885.

Salmon A, Flagel LE, Ying B, Udall JA, Wendel JF. 2010. Homoeologous nonreciprocal recombination in polyploid cotton. *New Phytol.* 186: 123–134.

Sawyer S. 1989. Statistical tests for detecting gene conversion. *Mol Biol Evol.* 6:526–538.

Seijo G, Lavia GI, Fernandez A, Krapovickas A, Ducasse DA, Bertioli DJ, Moscone EA. 2007. Genomic relationships between the cultivated peanut (*Arachis hypogaea*, Leguminosae) and its close relatives revealed by double GISH. *Am J Bot.* 94:1963–1971.

Soltis PS, Soltis DE. 2000. The role of genetic and genomic attributes in the success of polyploids. *Proc Natl Acad Sci U S A.* 97:7051–7057.

Spreitzer RJ, Peddi SR, Satagopan S. 2005. Phylogenetic engineering at an interface between large and small subunits imparts land-plant kinetic properties to algal Rubisco. *Proc Natl Acad Sci U S A.* 102: 17225–17230.

Spreitzer RJ, Salvucci ME. 2002. Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. *Annu Rev Plant Biol.* 53:449–475.

Wang J, Tian L, Madlung A, Lee H-S, Chen M, Lee JJ, Watson B, Kagochi T, Comai L, Chen ZJ. 2004. Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids. *Genetics* 167:1961–1973.

Wendel JF. 2000. Genome evolution in polyploids. *Plant Mol Biol.* 42: 225–249.

Wendel JF, Brubaker CL, Seelanan T. 2010. The origin and evolution of *Gossypium*. New York: Springer.

Wendel JF, Doyle JJ. 2005. Polyploidy and evolution in plants. In: Henry RJ, editor. Plant diversity and evolution. Wallingford (United Kingdom): CABI publishing. p. 97–117.

Wolf JB. 2009. Cytonuclear interaction can favor the evolution of genomic imprinting. *Evolution* 63:1364–1371.

Wright SI, Ness RW, Foxe JP, Barrett SCH. 2008. Genomic consequences of outcrossing and selfing in plants. *Int J Plant Sci.* 169: 105–118.

Yoo M-J, Szadkowski EP, Wendel JF. 2013. Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity (Edinb)* 110:171–180.