Coevolution in Hybrid Genomes: Nuclear-Encoded Rubisco Small Subunits and Their Plastid-Targeting Translocons Accompanying Sequential Allopolyploidy Events in *Triticum*

Changping Li,[†,1] Xiaofei Wang,[†,1] Yaxian Xiao,[1] Xuhan Sun,[1] Jinbin Wang,[1] Xuan Yang,[1] Yuchen Sun,[1] Yan Sha,[1] Ruili Lv,[1] Yanan Yu,[1] Baoxu Ding,[1] Zhibin Zhang,[1] Ning Li,[1] Tianya Wang,[1] Jonathan F. Wendel,[2] Bao Liu,[1] and Lei Gong*[,1]

[1]Key Laboratory of Molecular Epigenetics of the Ministry of Education (MOE), Northeast Normal University, Changchun, China
[2]Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA
[†]These authors contributed equally to this work.

***Corresponding author:** E-mail: gongl100@nenu.edu.cn.
Associate editor: Stephen Wright

Abstract

The *Triticum*/*Aegilops* complex includes hybrid species resulting from homoploid hybrid speciation and allopolyploid speciation. Sequential allotetra- and allohexaploidy events presumably result in two challenges for the hybrids, which involve 1) cytonuclear stoichiometric disruptions caused by combining two diverged nuclear genomes with the maternal inheritance of the cytoplasmic organellar donor; and 2) incompatibility of chimeric protein complexes with diverged subunits from nuclear and cytoplasmic genomes. Here, we describe coevolution of nuclear *rbcS* genes encoding the small subunits of Rubisco (ribulose 1,5-bisphosphate carboxylase/oxygenase) and nuclear genes encoding plastid translocons, which mediate recognition and translocation of nuclear-encoded proteins into plastids, in allopolyploid wheat species. We demonstrate that intergenomic paternal-to-maternal gene conversion specifically occurred in the genic region of the homoeologous *rbcS*3 gene from the D-genome progenitor of wheat (abbreviated as *rbcS*3D) such that it encodes a maternal-like or B-subgenome-like SSU3D transit peptide in allohexaploid wheat but not in allotetraploid wheat. Divergent and limited interaction between SSU3D and the D-subgenomic TOC90D translocon subunit is implicated to underpin SSU3D targeting into the chloroplast of hexaploid wheat. This implicates early selection favoring individuals harboring optimal maternal-like organellar SSU3D targeting in hexaploid wheat. These data represent a novel dimension of cytonuclear evolution mediated by organellar targeting and transportation of nuclear proteins.

Key words: **hybrid species, *Triticum*/*Aegilops* complex, Rubisco, cytonuclear coordination, CECs, translocon.**

Introduction

Hybridization is an important process that can lead to speciation, both at the homoploid level and when accompanied by genome doubling via polyploidy (Soltis and Soltis 2009; Abbott et al. 2013; Soltis et al. 2014; Yakimowski and Rieseberg 2014; Yakimowski and Barrett 2016). Both homoploid hybrid speciation (HHS) and allopolyploidy speciation (AS) are expected to involve overcoming two potential challenges: 1) cytonuclear stoichiometric disruptions caused by the merger of nuclear genomes in a context involving, usually, only the maternal cytoplasmic organellar genomes; and 2) incompatibility of chimeric protein complexes with diverged subunits from nuclear and cytoplasmic genomes. Related phenomena have been extensively studied in HHS diploid plants at the early stages (e.g., the cytoplasmic male sterility and the nuclear restoration Rf alleles) (Coyne and Orr 2004; Abbott et al. 2010; Chase et al. 2010; Caruso et al. 2012) and during speciation and evolution (Arnold 1997; Gross and Rieseberg 2005; Abbott et al. 2009; Yakimowski and Rieseberg 2014; Folk et al. 2016, 2018; Nieto Feliner et al. 2017).

The prevalence and role of cytonuclear coordination during AS has also begun to be explored (Leitch et al. 2006; Gong et al. 2012, 2014; Sehrish et al. 2015; Sharbrough et al. 2017; Wang, Dong, et al. 2017; Ferreira de Carvalho et al. 2019; Zhai et al. 2019). The most frequently employed protein models in allopolyploids are enzyme complexes composed of both nuclear-encoded and cytoplasmically encoded subunits (hereafter abbreviated as CECs; Rand et al. 2004; Millar et al. 2005; Woodson and Chory 2008; van Wijk and Baginsky 2011; Sloan et al. 2018). Nuclear-encoded subunits of CECs targeted to plastids represent complex systems of organellar protein sorting, targeting, and translocation (Chotewutmontri et al. 2017), which entail posttranslational recognition and

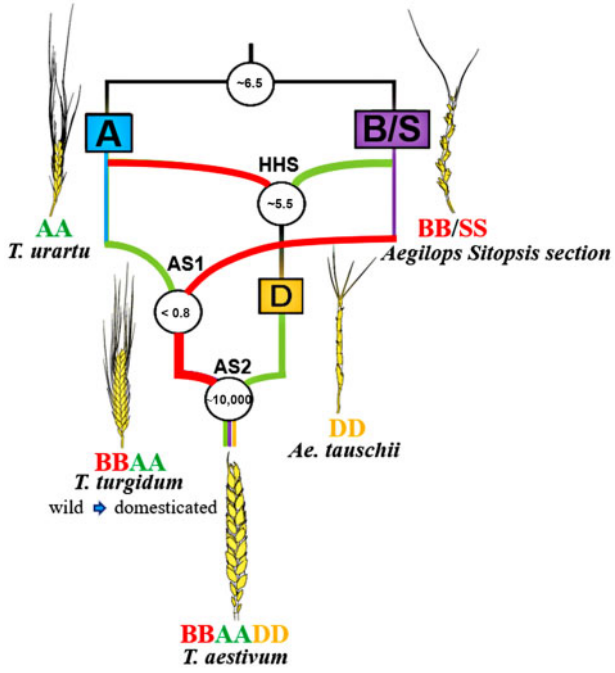

FIG. 1. Phylogenetic relationships of *Triticum*/*Aegilops* species with occurrences of diploid speciation (black lines), homoploid hybrid speciation (HHS; curved lines), and two rounds of allopolyploid speciation (AS1 and AS2). Maternal and paternal contributions are denoted by red and green curved lines. Estimated times are labeled within the circles (HHS and AS1 are estimated as million years ago; AS2 is estimated as ~10,000 years ago;This figure is adapted from "Ancient hybridizations among the ancestral genomes of bread wheat" by Marcussen et al. 2014. Copyright 2014 by " American Association for the Advancement of Science". Adapted with permission.).

translocation of nuclear-encoded proteins to plastids by oligomeric membrane complexes termed translocons (Walter and Lingappa 1986). Most translocons are nuclear-encoded whereas cytoplasmically targeted proteins (hereafter abbreviated as CTPs). Accordingly, CECs and CTP translocons offer us the opportunity to characterize and trace the evolutionary footprints of cytonuclear coordination as responses to the disruptions resulted by the aforementioned genome merger and/or genome doubling in HHS and AS (Bock et al. 2014; Sloan et al. 2014; Weng et al. 2016).

An initial whole-genome investigation of cytonuclear evolutionary responses to HHS was conducted using the diploid D-genome lineage of the *Triticum*/*Aegilops* complex (Li et al. 2019). This study revealed a biased maintenance of maternal ancestry in nuclear genes encoding subunits of CECs and CTPs, which is suggestive of cytonuclear selection for enhanced function in plant HHS (Li et al. 2019). The most-studied CEC model is Rubisco (1,5-bisphosphate carboxylase/oxygenase), which derives from nuclear *rbcS*-encoded small subunits (SSUs) and plastome *rbcL*-encoded large subunits (LSUs). These studies have established that many angiosperm allopolyploid lineages may have achieved cytonuclear compatibility at both genomic and transcriptional levels, which include biased paternal-to-maternal intergenomic gene conversion (paternal homoeologous *rbcS* genes are converted to be maternal-like) and preferential expression of maternal and maternal-like homoeologous nuclear genes, respectively (Gong et al. 2012, 2014; Sloan 2015). Additionally, given the targeting step of nuclear-encoded proteins into the organelles via the CTP translocon machinery, cytonuclear coordination in hybrid speciation might involve evolutionary adjustments in protein targeting and transportation into organelles.

Triticum/*Aegilops* is an ideal system to explore cytonuclear coordination in hybrid speciation because of its co-occurrences of species experiencing different histories of HHS and/or multiple rounds of ASs (fig. 1; Marcussen et al. 2014; Glémin et al. 2019). Briefly, within the *Triticum*/*Aegilops* complex, multiple ancient HHS(s) involving the A-genome *Triticum* species (represented by *Triticum urartu*) and B(S)-genome diploid species (represented by *Ae. speltoides* and *Ae. mutica*) have contributed to the origin of D-genome species (represented by *Ae. tauschii*) ~5.5 Ma (fig. 1; Gong et al. 2012, 2014; Marcussen et al. 2014; Glémin et al. 2019). In addition to this complex history at the diploid level, two well-known sequential AS events occurred within the *Triticum*/*Aegilops* complex: first, allotetraploidization occurred ~0.36–0.5 Ma between an A-genome *Triticum* species (represented by *T. urartu*) and an undefined or extinct goatgrass B/S-genome species closely related to the *Sitopsis* section of *Aegilops* (including *Ae. bicornis*, *Ae. longissima*, *Ae. searsii*, *Ae. sharonensis*, *Ae. speltoides*) as the maternal parent, leading to the origin of wild allotetraploid wheat (represented by *T. turgidum* ssp. *dicoccoides*). Second, after human domestication of tetraploid wheat for ~1,000–1,200 years, allohexaploidization (~10,000 years ago) between primitive domesticated *T. turgidum* (represented by *T. turgidum* ssp. *durum*; as the maternal parent) and a D-genome goatgrass (represented by *Ae. tauschii*) resulted in the formation of the hexaploid common or bread wheat (*T. aestivum* ssp. *aestivum*) (Feldman et al. 1995; Salamini et al. 2002; Zhang et al. 2014; Feldman and Levy 2015; fig. 1). These allotetraploid and allohexaploid wheat lineages offer excellent models to explore the cytonuclear coordination trajectory following sequential ASs.

Within this study, we characterized the *rbcS* gene family composition in representative diploid progenitors and reconstructed their evolutionary history. In addition, by analyzing the genomic composition of *rbcS* homoeologs in natural wild and (or) domesticated allopolyploid wheats and comparing them with corresponding orthologs from diploid progenitors, we demonstrate postallohexaploidy unidirectional gene conversion in the D-genome *rbcS*3 homoeolog, which results in maternal-like (B-like) paternal SSU3D signal peptides. Interestingly, there is no such genomic gene conversion detected in either natural wild or domesticated allotetraploid wheats. Further analyses of cytoplasm-to-chloroplast transport of SSUs implicate cytonuclear selection pressure involving interactions with a divergent D-subgenomic form of the TOC90 translocon. Here, we characterize this novel dimension of nuclear *coevolution* and its potential functional implications in cytonuclear coordination.

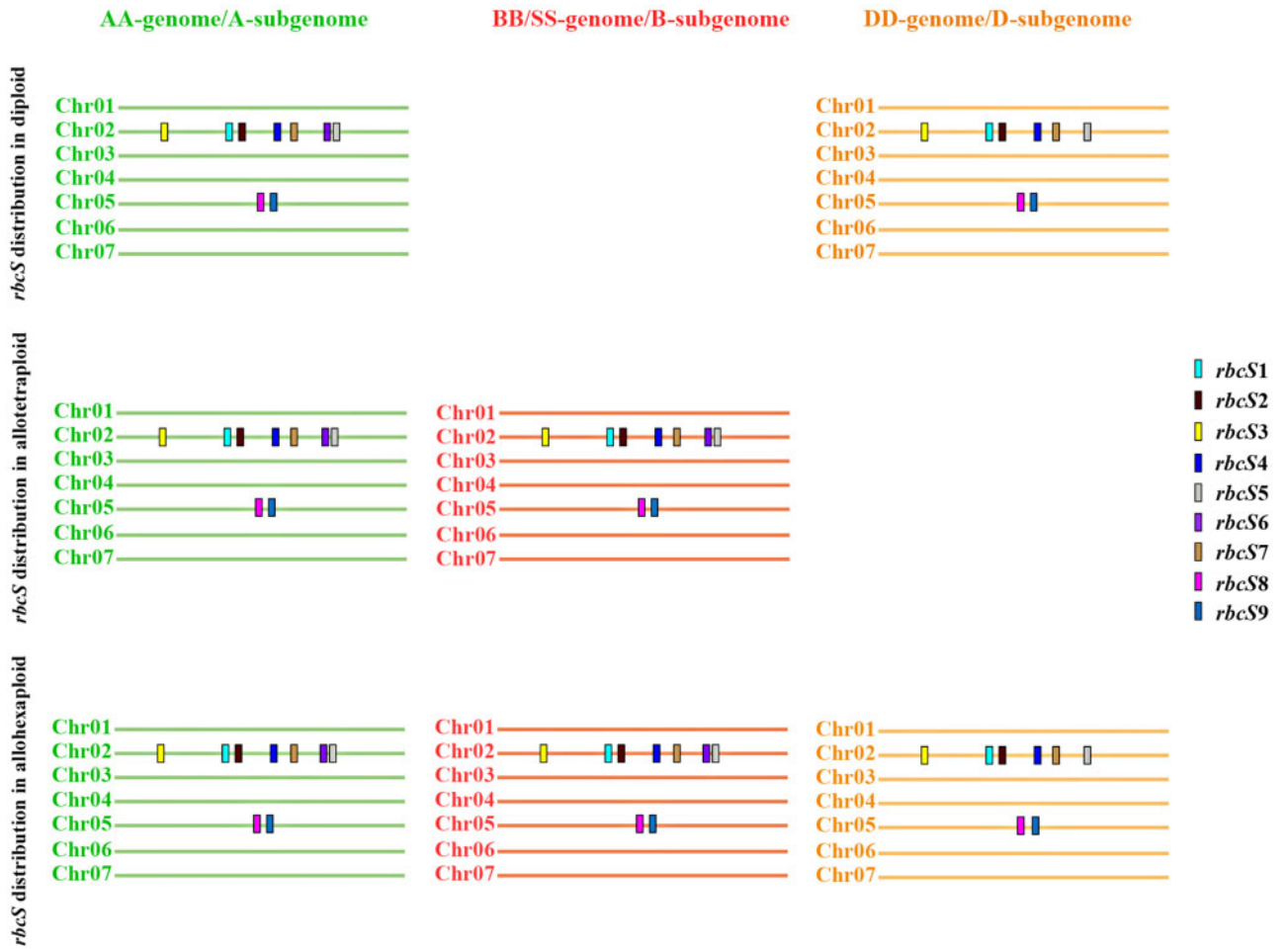

FIG. 2. Chromosomal distribution of *rbcS* gene homologs in representative diploid, allotetraploid, and allohexaploid *Triticum*/*Aegilops* species (no genome assembly with anchored chromosomes is available for B/S-genome diploids).

Results

Stable Inheritance of Most *rbcS* Homolog Lineages but Loss of the D-Genome *rbcS6* Lineage in Diploid and Allopolyploid Wheats

Prior to exploring cytonuclear evolution of *rbcS* homoeologs in the *Triticum*/*Aegilops* complex, we characterized *rbcS* family composition in representative A-genome (*T. urartu* TMU06 and TMU38), B/S-genome (*Ae. bicornis* TB01; *Ae. longissima* TL05; *Ae. searsii* TE10; *Ae. sharonensis* TH02; and *Ae. speltoides* TS01), D-genome (*Ae. tauschii* TQ18 and TQ27) diploid species, BBAA-genome allotetraploid species (*T. turgidum* ssp. *dicoccoides* TD265 and *T. turgidum* ssp. *durum* TTR13), and BBAADD-genome allohexaploids (*T. aestivum* ssp. *aestivum* TAA10, CS, and AK58) by cloning and sequencing as well as using available genome assemblies (Materials and Methods; supplementary table 1, Supplementary Material online). As introduced above, the evolutionary history of these species, which includes classical diploid speciation, HHS, and sequential AS events, are illustrated in figure 1.

To determine *rbcS* homolog lineages, we initially utilized the lineage-specific diagnostic SNPs and further discriminate the lineages using evidence from their collinear positions on homologous chromosomes (Materials and Methods). As shown (fig. 2), *rbcS*1, *rbcS*2, and *rbcS*3 comprise orthologous groups on homologous Chr02 separate from clusters of *rbcS*5–*rbcS*6, and *rbcS*4–*rbcS*7 on the same homologous chromosome. *rbcS*8 and *rbcS*9 are located on homologous Chr05. These conserved relative syntenic positions demonstrated the presence of nine homolog lineages (*rbcS*1–*rbcS*9) generated by ancient gene duplication(s) (fig. 2).

To reveal the evolutionary history of the *rbcS* gene homologs, a phylogenetic neighbor-joining gene tree was constructed based on aligned genomic *rbcS* homologs from diploid *Triticum*/*Aegilops* species and the outgroup barley (*Hordeum vulgare*) (fig. 3*a*; Materials and Methods). Notably, even though most internal nodes within the clades were not well supported (fig. 3*a*), most nodes linking to the gene homologs from the outgroup still received decent supportive bootstrap values (>70), which allows some degree of inference about the duplication history of the *rbcS* lineages. Specifically, given the close clustering of *rbcS*4-7 and *rbcS*8-9 in *Triticum*/*Aegilops* and the outgroup, gene duplication is inferred to have occurred prior to the divergence of *Hordeum* from *Triticum*/*Aegilops*. In contrast, other gene duplications generating paralogous *rbcS* lineages (i.e., *rbcS*1 and *rbcS*3) appear to have arisen near the base or root of the *Triticum*/*Aegilops* complex.

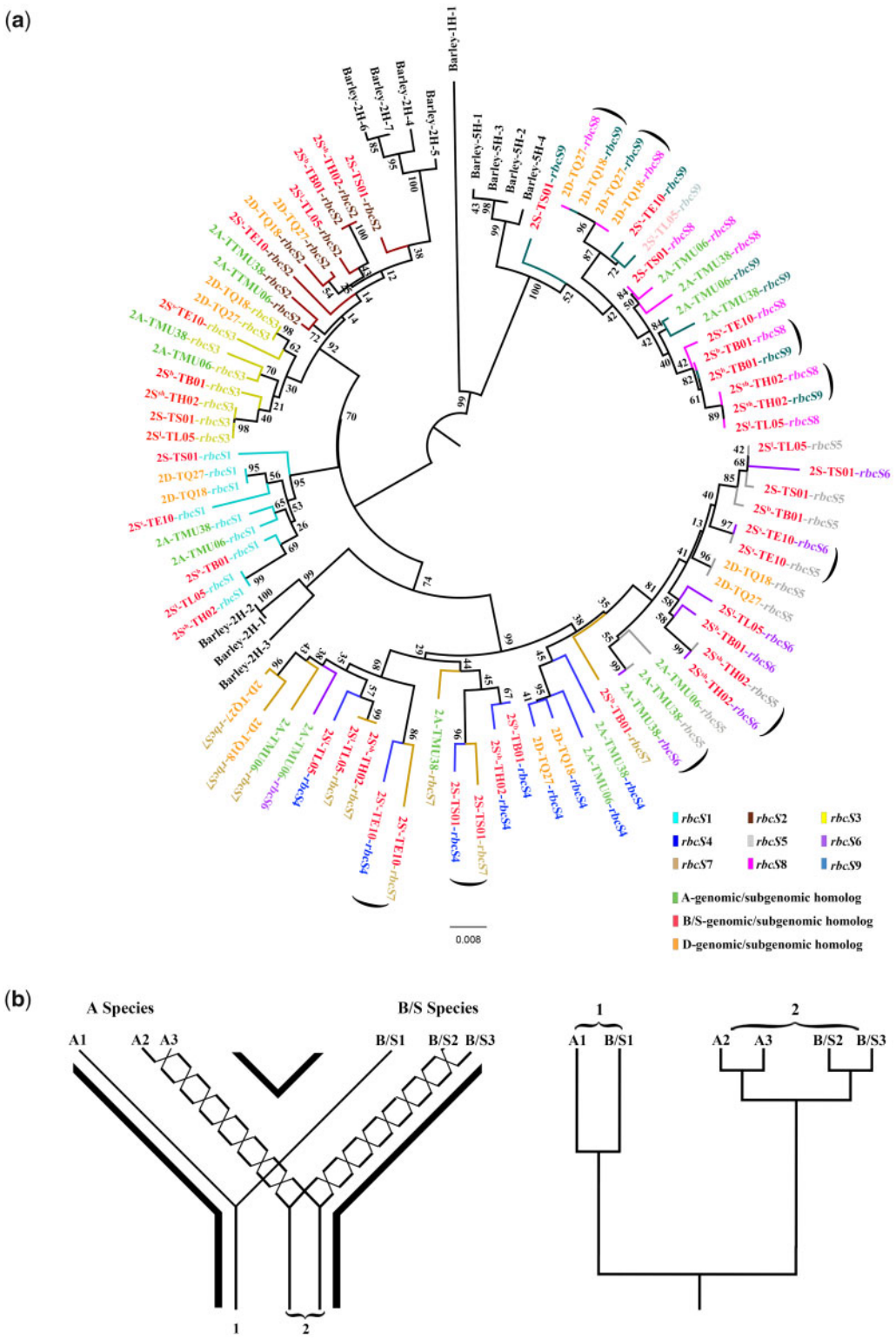

FIG. 3. (*a*) Simplified phylogenetic trees illustrating relationships among *rbcS* gene homologs in diploid *Triticum*/*Aegilops* species and the outgroup, barley. Phylogenetic neighbor-joining gene trees integrating *rbcS* gene homologs cloned and sequenced in representative diploids (2A, five *Sitopsis* species: *Ae. speltoides* [2S], *Ae. bicornis* [$2S^b$], *Ae. longissima* [$2S^l$], *Ae. searsii* [$2S^s$], and *Ae. sharonensis* [$2S^{sh}$], 2D, and *Hordeum vulgare* [barley], panel *a*). Each sequence name includes its accession name of the source species and the homolog group, denoted after the first and second dashes, respectively. Bootstrap values are denoted at nodes and the pairs of paralogous genes displaying concerted evolution are specified by the arcs on the periphery. The color codes of the homolog groups and the genomic/subgenomic origins are tabulated in the right bottom corner. (*b*) Simplified gene trees illustrate possible evolution patterns within species (left) and resulting phylogenetic gene trees (right) of *rbcS* gene homologs in two A and B/S diploid species. Intercrossing lines uniting genes denote concerted evolution in respective species (left). Gene 1 homologs have not experienced concerted evolution in either lineage and are designated as type 1 gene group; genes 2 and 3, within both the A and B/S species lineages, exhibit concerted evolution which partially homogenizes *rbcS* paralogs, and are designated as type 2 gene group (right).

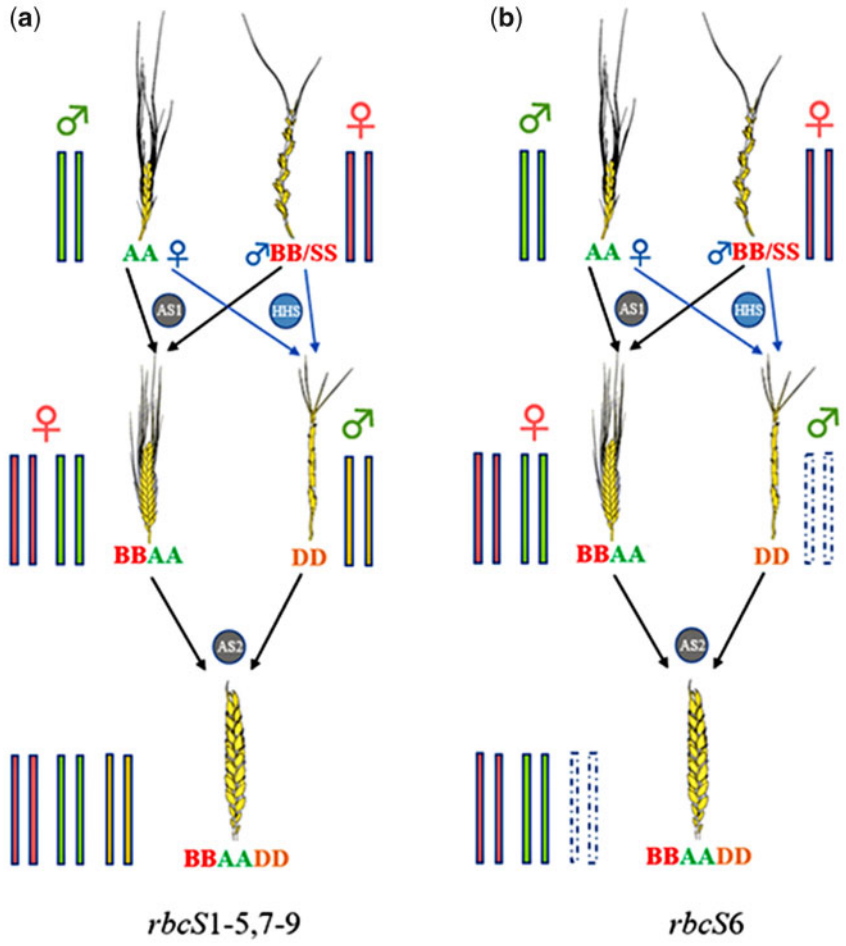

FIG. 4. Evolutionary history of *rbcS* gene homologs in D-genome diploid, allotetraploids, and allohexaploids. Illustrated are homoploid hybrid (HHS) and allopolyploid speciation (AS) events. Stable inheritance of *rbcS1-5* and *rbcS7-9* homologs (in filled colored columns) in both HHS and sequential AS events is summarized in panel (*a*); loss of *rbcS6* (dotted white column) in HHS and the subsequent allohexaploidy AS event is illustrated in panel (*b*).

Some *rbcS* homolog groups displayed greater within-species paralogous similarity than would be expected given their ancestry (e.g., *rbcS4* and *rbcS7* in accessions of TE10 and TS01; *rbcS5* and *rbcS6* in accessions of TE10, TH02, and TMU38; *rbcS8* and *rbcS9* in accessions of TB01, TH02, TQ18, and TQ27; fig. 3*a*). This within-species paralogous similarity implicates concerted evolution during the history of these genes, a characteristic feature of *rbcS* in many angiosperms (Meagher et al. 1989; Sanderson and Hufford 1996; Clegg et al. 1997; Gong et al. 2014). To determine the specific paralogs involved in concerted evolution, we combined the results from phylogenetic analysis with alternative expectations illustrated in simplified gene trees (fig. 3*b*). This analysis led us to infer that the paired lineages of *rbcS4* and *rbcS7*, *rbcS5* and *rbcS6*, and *rbcS8* and *rbcS9* each experienced concerted evolution (schematically illustrated by the arcs in fig. 3*a* and intercrossing lines in fig. 3*b*) leading to high paralogous similarity in many but not all species (fig. 3*b*). In contrast, in spite of *rbcS1*, *rbcS2*, and *rbcS3* are inferred to not have experienced concerted homogenization.

After integrating the *rbcS* homologs cloned and retrieved from allopolyploid wheats (fig. 2), most *rbcS* homologs were inferred to have been inherited in allotetraploid and allohexaploid species (fig. 2). Intriguingly, in spite of stable inheritance of *rbcS1*–*rbcS5*, *rbcS7*, *rbcS8*, and *rbcS9*, the *rbcS6* D-genome homolog lineage was not detected (by cloning or in the genome assemblies) in the D-genome diploid nor in the D-subgenome of BBAADD-genome allohexaploids (figs. 2 and 3*a*; Materials and Methods). We inferred that the D-genome *rbcS6* homolog lineage was completely lost during or subsequent to the HHS event that led to the D-genome diploid species (Marcussen et al. 2014; El Baidouri et al. 2017), and in the subsequent allopolyploidy that gave rise to hexaploid wheat, as illustrated in figure 4.

Unidirectional Nonsynonymous Paternal-to-Maternal (Maternal-Like) Gene Conversion in the D-Genome *rbcS3* Homoeolog Following Wheat Allohexaploidy

To characterize the evolution of biparental *rbcS* gene homoeologs in allotetraploid and allohexaploid wheats, we compared each *rbcS* homoeolog with its respective parental ortholog and inspected them for genome-diagnostic SNPs (Materials and Methods; Gong et al. 2012, 2014). This process permitted us to infer intra- and intergenomic gene conversions, which involved changes of diagnostic SNPs

Table 1. Intergenomic Conversions Mediating the Synonymous and Nonsynonymous Substitutions in Sampled Allotetra- and Allohexaploid Wheat Accessions.

Allotetraploid	TD265		TTR13		Svevo[a]		Zavitan[a]	
Intergenomic conversions	A(♂)-to-B(♀)	B(♀)-to-A(♂)	A(♂)-to-B(♀)	B(♀)-to-A(♂)	A(♂)-to-B(♀)	B(♀)-to-A(♂)	A(♂)-to-B(♀)	B(♀)-to-A(♂)
Synonymous substitutions	0	0	0	0	0	0	0	0
Nonsynonymous substitutions	0	0	0	0	0	0	0	0
Allohexaploid	**TAA10**							
Intergenomic conversions	A(♂)-to-B(♀)	B(♀)-to-A(♂)	A(♂)-to-D(♂)	D(♂)-to-A(♂)	D(♂)-to-B(♀)	B(♀)-to-D(♂)		
Synonymous substitutions	1	0	1	5	4	0		
Nonsynonymous substitutions	0	0	0	0	1[b]	0		
Allohexaploid	**AK58[a]**							
Intergenomic conversions	A(♂)-to-B(♀)	B(♀)-to-A(♂)	A(♂)-to-D(♂)	D(♂)-to-A(♂)	D(♂)-to-B(♀)	B(♀)-to-D(♂)		
Synonymous substitutions	0	0	1	1	5	0		
Nonsynonymous substitutions	0	0	0	0	1[b]	0		
Allohexaploid	**CS[a]**							
Intergenomic conversions	A(♂)-to-B(♀)	B(♀)-to-A(♂)	A(♂)-to-D(♂)	D(♂)-to-A(♂)	D(♂)-to-B(♀)	B(♀)-to-D(♂)		
Synonymous substitutions	0	0	0	0	3	0		
Nonsynonymous substitutions	0	0	0	0	1[b]	0		

[a]In those accessions, sequences were retrieved from genome.
[b]Denotes the nonsynonymous paternal-to-maternal gene conversion.

among homoeologs of the same and different parental origins, respectively. As described in previous studies (Gong et al. 2012, 2014), only intergenomic conversions were considered to be potentially involved in cytonuclear coordination. To facilitate this analysis, the nuclear subgenomic *rbcS* homologs consistent with the origin of cytoplasm were defined as "maternally contributed." Accordingly, B-genome *rbcS* homologs comprise the maternally contributed gene homoeologs in both allotetraploid and allohexaploid wheats, whereas A-genome *rbcS* homologs were paternally contributed in those two allopolyploids; D-genome *rbcS* homoeologs were paternally contributed in the allohexaploidy event giving rise to *T. aestivum* (Materials and Methods; Introduction). Because a gene conversion with a functional consequence is more likely to be one that causes an amino acid change, synonymous and nonsynonymous intergenomic gene conversions were studied separately (table 1). These data show: 1) that no intergenomic gene conversion was detected in wild (*T. turgidum* ssp. *dicoccoides* TD265 and Zavitan accession; table 1) or domesticated (*T. turgidum* ssp. *durum* TTR13 and Svevo accession; table 1) tetraploid wheat; 2) that in allohexaploid wheat (*T. aestivum* ssp. *aestivum* TAA10, AK58, and CS accession; table 1), from 4 to 11 and 0 to 1 SNPs were involved in the paternal-to-maternal (maternal-like) and maternal-to-paternal (paternal-like) gene conversions, respectively, mostly involving synonymous substitutions (table 1). One exception was a nonsynonymous SNP change (position 109 at the 5′-end CDS region) in the *rbcS*3 homoeolog group, which instead of having a paternal D-subgenomic adenine (A), exhibits a maternal-like B/S-subgenomic guanine (G) (fig. 5), which resulted in an amino acid change from a paternal serine (S) to a maternal arginine (R) (supplementary fig. 1, Supplementary Material online). This change makes the paternal D-subgenomic SSU3 (SSU3D; supplementary fig. 1, Supplementary Material online) harbor the same transit peptide (mediating the targeting of SSU to the chloroplast) as that of the maternal B/S-subgenomic SSU3 (SSU3B; fig. 5 and supplementary fig. 1, Supplementary Material online); and 3) given that there were many intronic and synonymous exonic maternal-like converted SNPs adjacent to the nonsynonymous change, we infer that the latter reflects part of a larger gene conversion event (supplementary figs. 1 and 2, Supplementary Material online), as opposed to a simple convergent mutation of a single SNP. Because of the many conserved nucleotide positions of all three parental homoeologs, it is not possible to demarcate the exact boundaries of the gene conversion event, so we conservatively focused on the smallest unit involved in gene conversion, which is the converted diagnostic SNP.

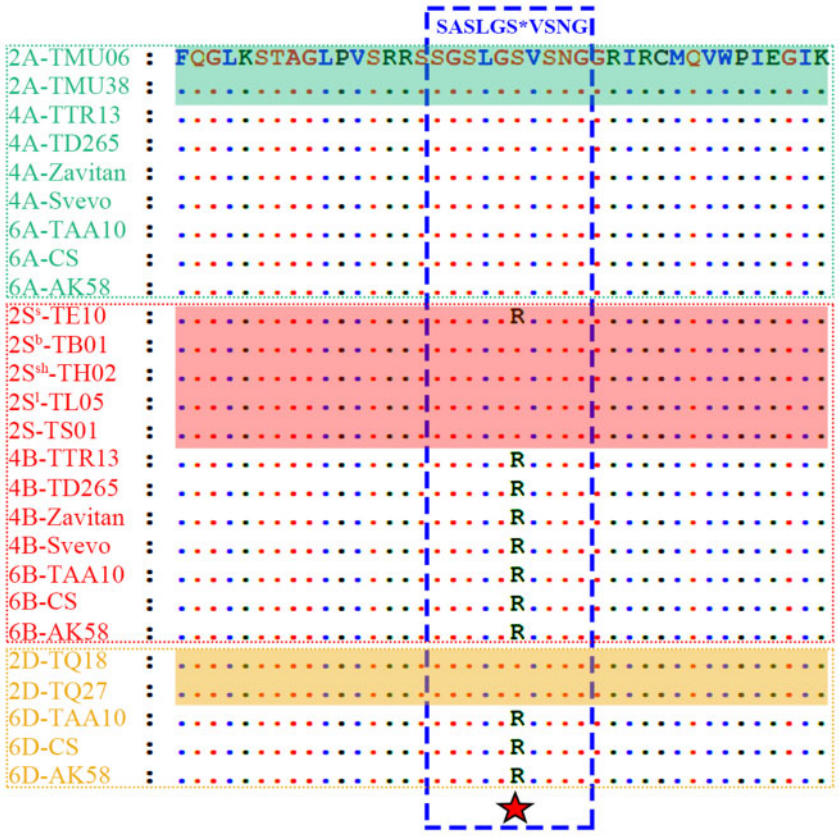

FIG. 5. Alignment of homologous SSU3 peptides around the SASLGS*VSNG motif cloned and/or retrieved from representative diploid (2A, five *Sitopsis* species: *Aegilops speltoides* [2S], *Ae. bicornis* [2S^{b}], *Ae. longissima* [2S^{l}], *Ae. searsii* [2S^{s}], and *Ae. sharonensis* [2S^{sh}], and 2D), allotetraploid (4A- and 4B-subgenomes), and allohexaploid (6A-, 6B-, and 6D-subgenomes) *Triticum*/*Aegilops* species. Dotted green, red, and yellow frames enclose respective A-genome/-subgenomic, B-/S-genome/-subgenomic, and D-genome/-subgenomic SSU3 homologs. The red star marks the amino acid changes from paternal "Serine (S)" to maternal "Arginine (R)" caused by the intergenomic gene conversion.

Uniform Paternal-to-Maternal Converted *SSU3* Transit Peptides in Allohexaploid Wheats

To assess the phylogenetic scope of the foregoing nonsynonymous change in the transit peptide of D-subgenome SSU3 homoeologs, the nucleotide composition at the corresponding position in other allohexaploid wheat accessions was tabulated from resequencing and transcriptomic RNA sequencing data, which included 29 Chinese local landraces and 34 varieties (Cheng et al. 2019), 100 European and 90 Chinese accessions (Miller et al. 2016; Wang, Yu, et al. 2017), and 65 Chinese Spring landraces/cultivars (Ramírez-González et al. 2018) (supplementary tables 2 and 3, Supplementary Material online). Without exception, all accessions shared this same SNP and amino acid change (supplementary table 2, Supplementary Material online). This observation implicates an early origin of this gene conversion event, possibly near the dawn of allohexaploid wheat evolution.

Divergent and Limited Interaction between SSU3D and the D-Subgenomic TOC90 Translocon Subunit

Given the nonsynonymous maternal-like conversion resulting in identical transit peptides in maternal and paternal subgenomic SSU3 homoeologs (SSU3B and SSU3D) in hexaploid wheat, the question about why only SSU3D has experienced this gene conversion deserves further exploration. Considering that the amino acid change is in the transit peptide region, which is essential for SSU targeting to the CTP translocon on the chloroplast (Walter and Lingappa 1986; VanderVere et al. 1995; Cline and Henry 1996; Schnell et al. 1997; Richter and Lamppa 2002; Chotewutmontri et al. 2017), we hypothesized: 1) that the interaction between SSU3D and certain CTP translocon subunit(s) could be under selection for optimal function involving efficient protein sorting and translocation of SSU3D into the chloroplast; and 2) that such selection was not exerted on the A-subgenomic SSU3A or other SSUs of different homoeolog groups.

To test these hypotheses, we studied the protein sequences of the interface domains involving both SSU transit peptides and their interacting CTP translocon subunits. SSU transit peptides in all parental diploids and respective allopolyploids (including allotetraploids and allohexaploids) were aligned and categorized into homolog groups (fig. 5). As noted, these peptide alignments revealed that the foregoing maternal-like amino acid change is within the conserved monocot "SASLGS*VSNG" sequence motif (Chotewutmontri et al. 2017), where * denotes the amino acid position corresponding to the amino acid change. Earlier experimental studies have confirmed that this motif is the region where the SSUs interact with the G-domains of

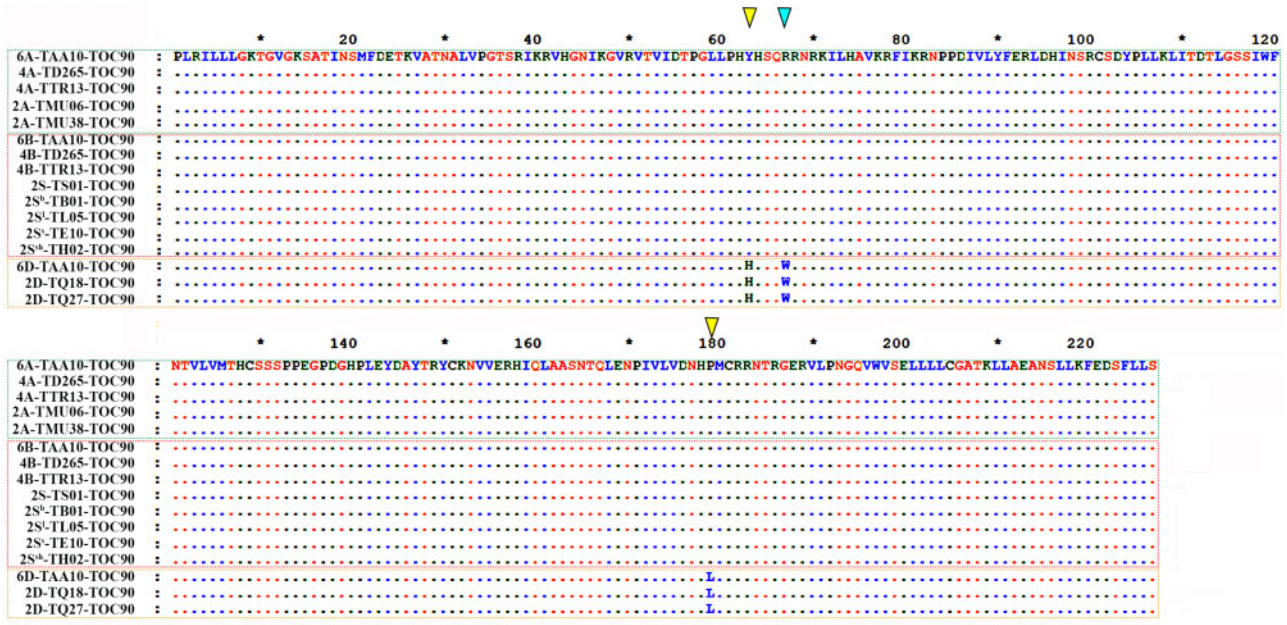

FIG. 6. Aligned G domains of TOC90 translocon homologs from representative diploid (2A, five *Sitopsis* species: *Aegilops speltoides* [2S], *Ae. bicornis* [2S^b], *Ae. longissima* [2S^l], *Ae. searsii* [2S^s], and *Ae. sharonensis* [2S^{sh}], and 2D), allotetraploid (4A- and 4B-subgenomes), and allohexaploid (6A-, 6B-, and 6D-subgenomes) *Triticum*/*Aegilops* species. Yellow and blue triangles mark amino acid substitutions of distinct chemical polarity and charge, respectively.

the TOC159 translocon complexes (Pinnaduwage and Bruce 1996; Kubis et al. 2004; Lee et al. 2009; Infanger et al. 2011; Chotewutmontri et al. 2017; Wiesemann et al. 2019). Interestingly, within most SSU homolog groups, their component diploid parental SSU homologs harbored identical amino acids in this "SASLGS*VSNG" motif (fig. 5); however, the unique polymorphic amino acid substitution, serine (S) in paternal A and D diploid species, versus arginine (R) in maternal B/S diploid species, corresponds to the star-marked gene-converted position in allohexaploid wheat (fig. 5). As B-subgenomic *rbcS*3s in both allotetra- and allohexaploid wheat harbor the same "R (arginine)" at the gene conversion position, we infer that the *rbcS*3 homolog in the B/S maternal diploid also likely had the same "R" amino acid at this position. It is possible that functional divergence of parental SSU3B/3S and SSU3D proteins could explain why only SSU3D was involved in homoeologous conversion after allohexaploidy.

To explore this further, we studied the specific CTP translocon subunit(s) that potentially interact with SSU homologs in protein targeting and importation. More specifically, we were interested in identifying the exact CTP translocon subunit interacting with the "SASLGS*VSNG" motif in the transit peptides of SSU homologs. Earlier experimental studies in *Arabidopsis thaliana* and *Pisum sativum* (pea) have demonstrated that the G-domains of the TOC159 and TOC90 directly interact with the SSUs homologs at their dicot "NDITSIASNG" motif which corresponds to the "SASLGS*VSNG" motif in monocots (Pinnaduwage and Bruce 1996; Becker et al. 2004; Lee et al. 2009; Chotewutmontri et al. 2017). Based on this, we hypothesized that the pairwise and/or cross interaction of homoeologous subgenomic SSUs with TOC159 and/or TOC90 translocons could have favored maternal-like gene conversion in SSU3D, which we view as a possible cytonuclear evolutionary response. To evaluate this, respective homoeologous genic regions encoding the G domains of TOC159 and TOC90 in allopolyploid wheats were cloned and/or retrieved from the available genomes (fig. 6 and supplementary fig. 3, Supplementary Material online; Materials and Methods). Sequence alignment of homoeologous TOC159 and TOC90 proteins revealed: 1) that all TOC159 homoeologs (A-, B-, and D-subgenomic) in allotetraploid and allohexaploid wheats had identical amino acids in their G domains (supplementary fig. 3, Supplementary Material online); 2) that A- and B-subgenomic homoeologous TOC90 G domains in allotetraploid wheat accessions displayed the same amino acid compositions (fig. 6), and that homoeologous G domains of TOC90 in allohexaploid wheats were identical with those in the allotetraploid; and 3) that in contrast, in the D-subgenomic TOC90 G-domains of allohexaploid wheats, amino acid substitutions occurred which were divergent from the (identical) A- and B-subgenomic TOC90 homoeologs (fig. 6). Moreover, these diverged amino acid substitutions involved changes in chemical polarity (D-subgenomic H[Histidine] versus B/A-subgenomic Y[Tyrosine] at the 63th position and D-subgenomic L[Lysine] versus B/A-subgenomic P[Proline] at the 179th position) and charge (D-subgenomic W[Tryptophan] versus B/A-subgenomic R[Arginine] at the 67th position) (fig. 6).

In addition to the sequence comparisons, based on the RNA sequencing data, we quantified and compared relative expression of A-, B-, and D-subgenomic TOC90-encoding homoeologs in leaves of allohexaploid wheats (Materials and Methods). The D-subgenomic TOC90 homoeolog was

Table 2. Relative Expression of Subgenomic TOC90 Homoeologs Quantified by the Number of Mapped Raw RNA Sequencing Reads in Two Allohexaploid Wheat Accessions (*Triticum aestivum* ssp. *aestivum* CS and TAA10).

CS-rep1		CS-rep2		CS-rep3	
Toc90-A	266	Toc90-A	746	Toc90-A	1188
Toc90-B	164	Toc90-B	612	Toc90-B	914
Toc90-D	112	Toc90-D	350	Toc90-D	458
TAA10-rep1		**TAA10-rep2**		**TAA10-rep3**	
Toc90-A	278	Toc90-A	232	Toc90-A	249
Toc90-B	118	Toc90-B	109	Toc90-B	95
Toc90-D	43	Toc90-D	60	Toc90-D	43

NOTE.—Three replicates are abbreviated as -rep1, -rep2, and -rep3, respectively.

consistently repressed compared with the A- and B-subgenomic homoeologs in two allohexaploid wheat accessions (relative to the hypothesized equal homoeologous expression; table 2).

With respect to SSU3D, the sequence difference and repressed expression of TOC90D relative to TOC90A and TOC90B suggest that interaction with its "native" (in the D-dipoid) TOC90D is significantly divergent from and limited relative to SSU3A/SSU3b with TOC90A/TOC90B. This interaction could provide an evolutionary explanation for coevolution between SSUs and their respective CTP translocons targeting into the chloroplast.

Discussion

Regional Concerted Evolution of *rbcS* Genes

It is well known that some multigene families, exemplified by *rbcS* and rDNA (ribosomal DNA), often displayed concerted evolution whereby paralogs within one species are more mutually similar than they are with corresponding orthologs in closely related species (Meagher et al. 1989; Sanderson and Hufford 1996; Clegg et al. 1997; Nei and Rooney 2005; Gong et al. 2014). Within the diploid species of the *Triticum*/*Aegilops* complex, concerted evolution of most *rbcS* paralogs generated by ancient duplication is also evident (fig. 3*a*). Representative examples involve paralogous *rbcS*4 and *rbcS*7, *rbcS*5 and *rbcS*6 (in all diploid species except the D-genome *Ae. tauschii*), and *rbcS*8 and *rbcS*9 groups within each diploid species (fig. 3*a*). It is notable that the sequence similarity of *rbcS* paralog groups associates with their genomic locations (fig. 2). For example, *rbcS*8 and *rbcS*9 paralogs cluster on homologous Chr05 chromosomes instead of with other *rbcS* paralog groups on homologous Chr02 chromosome (fig. 2). The *rbcS*8 and *rbcS*9 paralog groups harbor significant among-group sequence similarities, distinct from other *rbcS* paralog groups (figs. 2 and 3*a*). This observation is consistent with the hierarchical similarity of *rbcS* genes characterized in other species, which often display location-dependent regional homogenization via gene conversion (Meagher et al. 1989; Clegg et al. 1997; Gong et al. 2012). Our results are consistent with this pattern of dispersed *rbcS* paralog groups experiencing historical local gene conversion rather than across the entire genome.

Other common evolutionary outcomes for members in multigene families are also illustrated by the *rbcS* homologs characterized in the present study, namely, both divergent and birth-and-death evolutions (Nei and Rooney 2005). Specifically, as shown here for *rbcS*1, *rbcS*2, and *rbcS*3 homolog groups (fig. 3*a*), sequences and phylogenetic relationships demonstrate that mutation accumulation could outpace regional homogenization pressure (Gong et al. 2012, 2014) for some ortholog groups, leading to within-clustering among sequences from diploid species (fig. 3*a*). However, the weakly supported internal phylogenetic nodes suggest that regional homogenization mediated by gene conversion has been too low to produce significant paralogous homogenization. The *rbcS*6 homolog lineage deserves further consideration, as it was lost in the D-genome species *Ae. tauschii* (fig. 4). Considering the history of HHS in the background of *Ae. tauschii* generated by initial hybridization between A-genome *Triticum* species and S-/B-genome *Aegilops* species (Marcussen et al. 2014; El Baidouri et al. 2017), it is possible that pseudogenization and loss of both A- and S-/B-subgenomic *rbcS*6 alleles accompanied or followed this HHS process. Given the absence of homoeologous *rbcS*6D in allohexaploid wheat (fig. 4), such loss is inferred to have preceded the allohexaploidy event. Although *rbcS*6 was involved in concerted homogenization with *rbcS*5 in many *Triticum*/*Aegilops* species, as described above, nonfunctionalization of duplicated *rbcS*6 genes implicate birth-and-death evolution (Nei and Rooney 2005) in the D-genome *Ae. tauschii*.

Cytonuclear Coordination Involving *rbcS*3 Homologs

Coadaptation or coordination between the nuclear and cytoplasmic organellar genomes has been recognized as an essential component of evolutionarily successful hybrid speciation, either at the homoploid or allopolyploid level (Caruso et al. 2012; Burton et al. 2013; Yakimowski and Rieseberg 2014; Sloan 2015; Folk et al. 2016, 2018; Nieto Feliner et al. 2017). Based on cytonuclear epistasis and maternal–offspring coadaptation theory (Wolf and Hager 2006; Wolf 2009), cytonuclear intergenomic interactions may be interrupted when hybridization entails combining divergent nuclear genomes with only a single set of cytoplasmic organellar genomes. In the present case for Rubisco, diverged nuclear allelic or homoeologous *rbcS*s from parental species may have experienced selection pressure for optimal function in a novel cytoplasmic background, as implicated in a number of recent studies (Gong et al. 2012, 2014; Sehrish et al. 2015; Wang, Dong, et al. 2017; Ferreira de Carvalho et al. 2019; Li et al. 2019; Zhai et al. 2019). One novel aspect of the present study is that it involves two forms of hybridization (HHS and AS) at two ploidy levels, tetraploid and hexaploid.

Here, we show that early in the establishment of allohexaploid wheat, *rbcS3D* became fixed for a nonsynonymous, paternal-to-maternal (maternal-like) gene conversion (fig. 5 and supplementary fig. 2, Supplementary Material online), similar to reports from multiple other (but not all) allopolyploid lineages (Gong et al. 2012, 2014; Sehrish et al. 2015; Wang, Dong, et al. 2017; Ferreira de Carvalho et al. 2019; Zhai et al. 2019). Intriguingly, at the diploid level, that is,

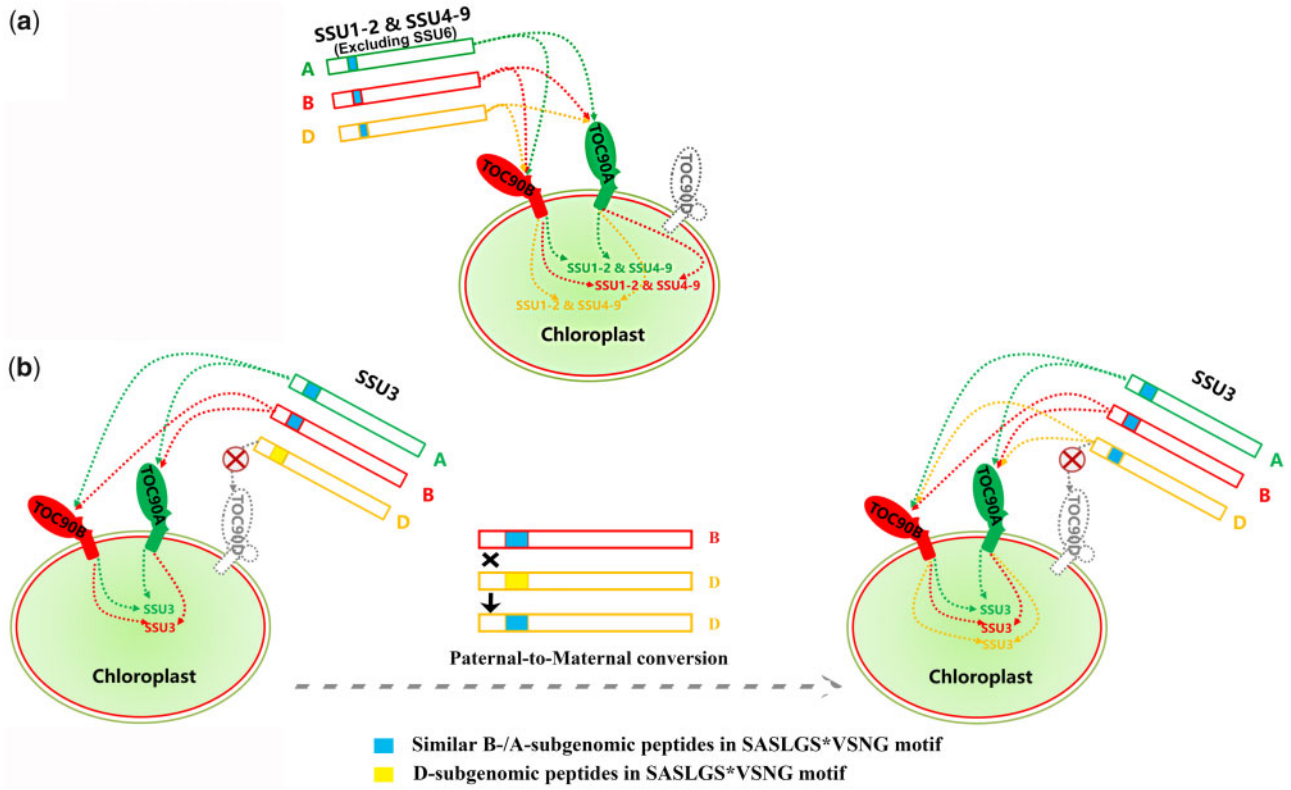

FIG. 7. Model of cytonuclear coordination at the level of SSU targeting and transportation in allohexaploid wheat. Relatively less transcriptionally active TOC90D and abundant TOC90A and TOC90B are represented by hollow dotted and colorful solid polygons on the chloroplast membrane, respectively. As for three subgenomic SSU1-2 and SSU4-9 homoeologs (excluding SSU6; panel *a*), they could be efficiently targeted and transported through TOC90A and TOC90B. As for three subgenomic SSU3 homoeologs (panel *b*), targeting and transportation with and through TOC90A and TOC90B are still allowed for SSU3A and SSU3B. As TOC90D is relative less transcriptionally active, the respective path via TOC90D is blocked, which is represented by the "No parking" sign. We hypothesize that under cytonuclear coordination pressure, SSU3D was converted to be SSU3A-/SSU3B-like via paternal-to-maternal (maternal-like) gene conversion (panel *b*), which achieves SSU3D targeting to TOC90A and/or TOC90B and transport into the chloroplast.

during the HHS event in the history of *Ae. tauschii*, the paternal B-allelic *rbcS3* was lost (Li et al. 2019). It is notable that the same set of *rbcS3* homologs are implicated in possible cytonuclear coordination in both AS and HHS, different and independent types of hybridization events. This result is suggestive of possible consistent cytonuclear selection for optimized SSU3 function, although we recognize that functional studies are still lacking to prove this. Future "mix and match" transgenic replacement experiments integrating native and alternative (converted and nonconverted) *rbcS3D* into those *Triticum*/*Aegilops* complex species and subsequent functional and phenotypic comparisons could prove informative with respect to this speculation.

A Novel Dimension of Cytonuclear Coordination Involving CTP Translocons

Cytonuclear enzyme complexes not only require compatible nuclear-encoded and cytoplasmically encoded protein partners but also compatible protein-targeting and transport mechanisms. Here, we focused on nuclear-encoded CTP translocons, involving the major carriers, channels, and chaperons that mediate organellar targeting and transportation of nuclear-encoded subunits of CEC proteins (Chotewutmontri et al. 2017; Li et al. 2019). Our concern here was whether or not evidence is consistent with the hypothesis that there has been coevolution of diverged *rbcS3D* genes and the TOC90 translocon genes at the level of SSU targeting and transport (figs. 5 and 6). This form of coevolution does not involve plastid genes per se, but instead represents selection for enhanced interactions between the products of two different nuclear gene systems, SSUs and their respective translocons (fig. 7). Based on sequence comparisons and homoeologous expression levels, our model suggests: 1) that A-, B-, and D-subgenomic SSU1-2 and SSU4-9 homoeologs (excluding SSU6; fig. 7*a*) are correctly targeted and transported using homoeologous TOC90A and TOC90B, noting their similar amino acid compositions in their transit peptide regions (SASLGS*VSNG motif); 2) that for A-, B-, and D-subgenomic SSU3 homoeologs (fig. 7*b*), targeting and transport using TOC90A and TOC90B remains functional for SSU3A and SSU3B; however, as the path via TOC90D is impaired (TOC90D is less transcriptionally active; fig. 7*b*), a specific cytonuclear response involving a maternal-like gene conversion among SSU3 homoeologs (SSU3D was converted to be SSU3A/B-like) resulted in a maternal-like SSU3D at its interface with Toc90 proteins, thereby facilitating interactions

with the more abundant Toc90B and Toc90A counterparts (figs. 6 and 7*b*); 3) as SSU3D was not converted to be paternal, A-genome like (paternal-to-paternal conversion; table 1), this suggests a possible evolutionary benefit of paternal-to-maternal conversion in SSU3D in promoting compatibility with the maternal B/S-genome LSU and/or other maternal plastid-encoded factors in the assembly of the rubisco holoenzyme. These results add to a growing literature about how gene conversion may be a common molecular mechanism of cytonuclear coordination at both the genomic and the organellar targeting levels (Gong et al. 2012, 2014; Sehrish et al. 2015; Wang, Dong, et al. 2017; Ferreira de Carvalho et al. 2019; Zhai et al. 2019).

This coevolutionary model (fig. 7) is, in principle, testable using at least two approaches. First, in vivo and/or in vitro experimental comparisons of the targeting and transport efficiencies and cross-talking interactions between Toc90B/90A and maternal-like SSU3D versus via interaction between Toc90D and nonconverted SSU3D (diploid paternal type) could be adopted. Second, one might take advantage of the existence of the D-genome diploid that has the B/S-genome organellar background, experimentally extracted from allohexaploid wheat (Jahier et al. 2017); this material should contain only Toc90D. Notably, as there is no suspected mismatch between Toc90B/90A with their respective SSU3B and SSU3A homoeologs in tetraploid wheat, we would expect little gene conversion pressure on B- and A-subgenomic *rbc*S3B (SSU3B) and *rbc*S3A (SSU3A), as modeled in figure 7.

From an evolutionary standpoint, the homogeneity of SSU3Ds detected in an extensive survey of allohexaploid wheat accessions suggests a scenario whereby natural and/or human-mediated directional selection acted early in hexaploid wheat establishment, favoring individuals harboring this type of gene-converted SSU3D such that it regained a more optimal organellar targeting and transport dynamics. Experiments to test this novel dimension of cytonuclear coordination are possible using resynthesized wheat allohexaploids, combined with functional assays.

Materials and Methods

Cloning, and Sequencing of *rbcS*, TOC159, and TOC90

Genome assemblies of all representative extant progenitors and offspring species make it possible to characterize the cytonuclear coevolution pattern of whole-genomic genes encoding nuclear subunits of CECs (such as the *rbcS* genes encoding SSUs of CEC Rubisco) and nuclear CTP translocons in our case (Li et al. 2019) (a draft *Ae. speltoides* genome is available in International Wheat Genome Sequencing Consortium, IWGSC). Accordingly, we searched for *rbcS*, TOC159, and TOC90 gene homologs in genome assemblies of A-genome diploid *T. urartu* (Ling et al. 2018), D-genome diploid *Ae. tauschii* (Luo et al. 2017), BBAA-genome allotetraploid *T. turgidum* ssp. *dicoccoides* (Avni et al. 2017) and *T. turgidum* ssp. *durum* (Maccaferri et al. 2019), and BBAADD-genome allohexaploid *T. aestivum* ssp. *aestivum* var. Chinese Spring (Appels et al. 2018) and var. AK58 (unpublished data), using NCBI local BlastN search (default settings) and previously characterized *Arabidopsis thaliana* homologs as respective query sequences (*rbc*S1A: At1g67090, *rbc*S1B: At5g38430, *rbc*S2B: At5g38420, *rbc*S3B: At5g38410, TOC159 gene: AT4G02510, and TOC90 gene: AT5G20300).

To obtain genomic sequences of *rbcS* and *rbcL* gene homologs, homolog-specific PCR primers were designed in conserved 5′- and 3′-UTRs of each homolog group (supplementary table 1, Supplementary Material online). For example, for the *rbc*S1A homolog group, the forward and reverse primers (5′-CCACATCTTGATTAAAGATAGGAGC-3′ and 5′-AGGTTTTGTATATGTGTAGCCGGTT-3′, respectively), were located in the conserved region of the *rbc*S1A homologs retrieved from *T. urartu*, *T. turgidum* ssp. *dicoccoides*, *T. turgidum* ssp. *durum*, and *T. aestivum* ssp. *aestivum*. PCR amplifications were carried out using template gDNA extracted from all accessions. Further cloning and sequencing was completed following the same methods as described earlier (Wang, Dong, et al. 2017). As for the gene homologs in possible B/S- parental diploid species, five representative *Sitopsis* species (including *Ae. bicornis*, *Ae. longissima*, *Ae. searsii*, *Ae. sharonensis*, and *Ae. speltoides*; accessions tabulated in Results) were included. As there is no genome assembly available for these species, primers amplifying their corresponding gene homologs were designed based on the B-subgenomic sequences assembled in BBAA (*T. turgidum* ssp. *dicoccoides* and *T. turgidum* ssp. *durum*) and BBAADD (*T. aestivum* ssp. *aestivum*) allopolyploid wheats.

For the TOC159 and TOC90 translocon gene homologs, only genic regions encoding their G-domains (reasons described in Results) were cloned and sequenced. The primer design strategy was similar as for *rbcS* gene homologs, using primers located in conserved regions flanking the G-domain encoding region (supplementary table 1, Supplementary Material online).

To ensure no false amplification artifacts, for each gene homolog, three independent experiments, including PCR amplification, cloning, and sequencing, were carried out in parallel.

Phylogenetic Tree Construction and Gene Conversion Analyses

To discriminate *rbcS* lineages, cloned and retrieved *rbcS* gene homologs from diploid species were initially aligned using local MAFFT (default settings; Katoh and Standley 2013). The possible lineage-diagnostic SNPs were specifically identified. Lineage determination was facilitated by using the genomic locations of respective *rbcS* homologs in genome assemblies of diploid (*T. urartu* and *Ae. tauschii*) and allopolyploid species (BBAA-genome *T. turgidum* ssp. *dicoccoides* and *T. turgidum* ssp. *durum*; BBAADD-genome *T. aestivum* ssp. *aestivum* var. Chinese Spring).

Based on the aligned sequences, a neighbor-joining phylogenetic tree was constructed using MEGA6.0 (Tamura et al. 2013) under the Jukes–Cantor substitution model with bootstrap evaluation for the cloned and retrieved gene homologs using the outgroup barley (*Hordeum vulgare*) sequences. The

tree was illustrated using Figtree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/; last accessed Feburary 2020).

Intra- and intergenomic gene conversions among subgenomic homoeologous *rbcS* copies were inferred separately following the same methods as described previously (Gong et al. 2012, 2014; Wang, Dong, et al. 2017). Briefly, within the allopolyploid species, the possible *rbcS* genomic conversion regions (in one direction from paternal to maternal homoeolog or vice versa) 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; Gong et al. 2014). Conversion copies were further visually confirmed. Each *rbcS* homoeolog in the polyploids was searched against both parental diploid orthologs as well as 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 (Gong et al. 2014).

Given the similar organellar genomic composition (all B/S-like) in BBAA and BBAADD allopolyploids and B/S-genome *Sitopsis* species, B/S-subgenomic *rbcS* were considered as maternal homoeologs, whereas both A- and D-genome *rbcS* homoeologs were paternally contributed. As it is not known which *Sitopsis* species is the closest extant diploid representing the maternal (B/S-genome) progenitor of allotetraploid wheat (Feldman and Levy 2015), we adopted a conservative approach in which only conversions involving changes of subgenome-diagnostic SNPs unambiguously supported by at least one *Sitopsis* species were considered to result from intergenomic conversion (examples shown in supplementary fig. 2, Supplementary Material online).

Analyses of NGS Resequencing and Transcriptomic Data

To characterize the prevalence of exonic gene conversion events in worldwide allohexaploid populations, genomic resequencing and RNA sequencing data from 318 allohexaploid varieties (supplementary tables 2 and 3, Supplementary Material online) were mapped to the reference genomic and exonic CDSs of B-, A-, and D-subgenomic *rbcS3* gene homoeologs (low-quality reads with the percentage of unqualified [phred quality $<$ 20] bases $>$ 20% were discarded before mapping; bwa mapping tool with requirements of perfect match and properly and unique mapping was adopted [mapping quality threshold $>$30]; Li and Durbin 2010). Samtools (samtools view option) was adopted to check the nucleotide composition of all *rbcS3* homoeologs in all samples.

To quantify the relative expression of subgenomic TOC90 translocon genes, transcriptomic RNA-seq data generated from the leaves of two exemplary allohexaploid wheat accessions (*T. aestivum* ssp. *aestivum* TAA10; unpublished data and *T. aestivum* ssp. *aestivum* CS; data available from NCBI SRP06335) were mapped to the exonic CDSs of cloned and sequenced or retrieved TOC90 homoeologs (bowtie2 mapping tools with requirements of perfect match and properly and unique mapping [mapping quality threshold $>$30]; Langmead and Salzberg 2012). The number of raw reads mapped to respective homoeologs was calculated to represent relative expression abundance.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 31970238), the National Key Research and Development Program of China (2016YFD0101004), and the National Science Foundation Plant Genome Program (to J.F.W.). We appreciate the knowledge and trainings given by the course of Evolution Biology in Northeast Normal University.

References

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, et al. 2013. Hybridization and speciation. *J Evol Biol*. 26(2):229–246.

Abbott RJ, Brennan AC, James JK, Forbes DG, Hegarty MJ, Hiscock SJ. 2009. Recent hybrid origin and invasion of the British Isles by a self-incompatible species, Oxford ragwort (*Senecio squalidus* *l*., Asteraceae). *Biol Invasions*. 11(5):1145–1158.

Abbott RJ, Hegarty MJ, Hiscock SJ, Brennan AC. 2010. Homoploid hybrid speciation in action. *Taxon* 59(5):1375–1386.

Appels R, Eversole K, Stein N, Feuillet C, Keller B, Rogers J, Pozniak CJ, Choulet F, Distelfeld A, Poland J, et al. 2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191.

Arnold ML. 1997. Natural hybridization and evolution. New York: Oxford University Press.

Avni R, Nave M, Barad O, Baruch K, Twardziok SO, Gundlach H, Hale I, Mascher M, Spannagl M, Wiebe K, et al. 2017. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science* 357(6346):93–97.

Becker T, Jelic M, Vojta A, Radunz A, Soll J, Schleiff E. 2004. Preprotein recognition by the Toc complex. *EMBO J*. 23(3):520–530.

Bock DG, Andrew RL, Rieseberg LH. 2014. On the adaptive value of cytoplasmic genomes in plants. *Mol Ecol*. 23(20):4899–4911.

Burton RS, Pereira RJ, Barreto FS. 2013. Cytonuclear genomic interactions and hybrid breakdown. *Annu Rev Ecol Evol Syst*. 44(1):281–302.

Caruso CM, Case AL, Bailey MF. 2012. The evolutionary ecology of cytonuclear interactions in Angiosperms. *Trends Plant Sci*. 17(11):638–643.

Chase CD, Ribarits A, Heberle-Bors E. 2010. Male sterility. In: Pua EC, Davey MR, editors. Plant developmental biology - biotechnological perspectives. Heidelberg, Germany: Springer. p. 437–457.

Cheng H, Liu J, Wen J, Nie X, Xu L, Chen N, Li Z, Wang Q, Zheng Z, Li M, et al. 2019. Frequent intra- and inter-species introgression shapes the landscape of genetic variation in bread wheat. *Genome Biol*. 20(1):136.

Chotewutmontri P, Holbrook K, Bruce BD. 2017. Plastid protein targeting: preprotein recognition and translocation. *Int Rev Cell Mol Biol*. 330:227–294.

Clegg MT, Cummings MP, Durbin ML. 1997. The evolution of plant nuclear genes. *Proc Natl Acad Sci U S A*. 94(15):7791–7798.

Cline K, Henry R. 1996. Import and routing of nucleus-encoded chloroplast proteins. *Annu Rev Cell Dev Biol*. 12(1):1–26.

Coyne JA, Orr HA. 2004. Speciation. Sunderland (MA): Sinauer Associates. p. 545.

El Baidouri M, Murat F, Veyssiere M, Molinier M, Flores R, Burlot L, Alaux M, Quesneville H, Pont C, Salse J. 2017. Reconciling the evolutionary origin of bread wheat (*Triticum aestivum*). *New Phytol*. 213(3):1477–1486.

Feldman M, Levy AA. 2015. Origin and evolution of wheat and related Triticeae species. In: Molnár-Láng M, Ceoloni C, Doležel J, editors. Alien introgression in wheat: cytogenetics, molecular biology, and genomics. Cham: Springer International Publishing. p. 21–76.

Feldman M, Lupton FGH, Miller TE. 1995. Wheats. In: Smartt, JSimmonds NW, editors. Evolution of crop plants. London: Longman Scientific. p. 184–192.

Ferreira de Carvalho J, Lucas J, Deniot G, Falentin C, Filangi O, Gilet M, Legeai F, Lode M, Morice J, Trotoux G, et al. 2019. Cytonuclear interactions remain stable during allopolyploid evolution despite repeated whole-genome duplications in Brassica. *Plant J*. 98(3):434–447.

Folk RA, Mandel JR, Freudenstein JV. 2016. Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of Angiosperms. *Syst Biol*. 66:320–337.

Folk RA, Soltis PS, Soltis DE, Guralnick R. 2018. New prospects in the detection and comparative analysis of hybridization in the tree of life. *Am J Bot*. 105(3):364–375.

Glémin S, Scornavacca C, Dainat J, Burgarella C, Viader V, Ardisson M, Sarah G, Santoni S, David J, Ranwez V. 2019. Pervasive hybridizations in the history of wheat relatives. *Sci Adv*. 5(5):eaav9188.

Gong L, Olson M, Wendel JF. 2014. Cytonuclear evolution of Rubisco in four allopolyploid lineages. *Mol Biol Evol*. 31(10):2624–2636.

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(10):3023–3036.

Gross BL, Rieseberg LH. 2005. The ecological genetics of homoploid hybrid speciation. *J Hered*. 96(3):241–252.

Infanger S, Bischof S, Hiltbrunner A, Agne B, Baginsky S, Kessler F. 2011. The chloroplast import receptor Toc90 partially restores the accumulation of Toc159 client proteins in the *Arabidopsis thaliana* ppi2 mutant. *Mol Plant*. 4(2):252–263.

Jahier J, Coriton O, Deffains D, Arnaud D, Chalhoub B. 2017. Revisiting meiotic pairing in wheat synthetics in relation to the evolution of the meiotic system in wheat. *Plant Syst Evol*. 303(9):1311–1316.

Katoh K, Standley DM. 2013. Mafft multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.

Kubis S, Patel R, Combe J, Bédard J, Kovacheva S, Lilley K, Biehl A, Leister D, Ríos G, Koncz C, et al. 2004. Functional specialization amongst the *Arabidopsis* Toc159 family of chloroplast protein import receptors. *Plant Cell*. 16(8):2059–2077.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with bowtie 2. *Nat Methods*. 9(4):357–359.

Lee DW, Lee S, Oh YJ, Hwang I. 2009. Multiple sequence motifs in the Rubisco small subunit transit peptide independently contribute to Toc159-dependent import of proteins into chloroplasts. *Plant Physiol*. 151(1):129–141.

Leitch AR, Lim KY, Skalicka K, Kovarik A. 2006. Nuclear cytoplasmic interaction hypothesis and the role of translocations in nicotiana allopolyploids. In: Cigna AA, Durante M, editors. Radiation risk estimates in normal and emergency situations. Dordrecht: Springer Netherlands. p. 319–326.

Li C, Sun X, Conover JL, Zhang Z, Wang J, Wang X, Deng X, Wang H, Liu B, Wendel JF, et al. 2019. Cytonuclear coevolution following homoploid hybrid speciation in *Aegilops tauschii*. *Mol Biol Evol*. 36(2):341–349.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26(5):589–595.

Ling H-Q, Ma B, Shi X, Liu H, Dong L, Sun H, Cao Y, Gao Q, Zheng S, Li Y, et al. 2018. Genome sequence of the progenitor of wheat a subgenome *Triticum urartu*. *Nature* 557(7705):424–428.

Luo MC, Gu YQ, Puiu D, Wang H, Twardziok SO, Deal KR, Huo N, Zhu T, Wang L, Wang Y, et al. 2017. Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature* 551(7681):498–502.

Maccaferri M, Harris NS, Twardziok SO, Pasam RK, Gundlach H, Spannagl M, Ormanbekova D, Lux T, Prade VM, Milner SG, et al. 2019. Durum wheat genome highlights past domestication signatures and future improvement targets. *Nat Genet*. 51(5):885–895.

Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M, Jakobsen KS, Wulff BBH, Steuernagel B, Mayer KFX, Olsen O-A. 2014. Ancient hybridizations among the ancestral genomes of bread wheat. *Science* 345(6194):1250092.

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(19):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(4):845–863.

Millar AH, Heazlewood JL, Kristensen BK, Braun H-P, Møller IM. 2005. The plant mitochondrial proteome. *Trends Plant Sci*. 10(1):36–43.

Miller CN, Harper AL, Trick M, Werner P, Waldron K, Bancroft I. 2016. Elucidation of the genetic basis of variation for stem strength characteristics in bread wheat by associative transcriptomics. *BMC Genomics* 17(1):500.

Nei M, Rooney AP. 2005. Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet*. 39(1):121–152.

Nieto Feliner G, Álvarez I, Fuertes-Aguilar J, Heuertz M, Marques I, Moharrek F, Piñeiro R, Riina R, Rosselló JA, Soltis PS, et al. 2017. Is homoploid hybrid speciation that rare? An empiricist's view. *Heredity* 118(6):513–516.

Pinnaduwage P, Bruce BD. 1996. In vitro interaction between a chloroplast transit peptide and chloroplast outer envelope lipids is sequence-specific and lipid class-dependent. *J Biol Chem*. 271(51):32907–32915.

Ramírez-González RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, Davey M, Jacobs J, van Ex F, Pasha A, et al. 2018. The transcriptional landscape of polyploid wheat. *Science* 361(6403):eaar6089.

Rand DM, Haney RA, Fry AJ. 2004. Cytonuclear coevolution: the genomics of cooperation. *Trends Ecol Evol*. 19(12):645–653.

Richter S, Lamppa GK. 2002. Determinants for removal and degradation of transit peptides of chloroplast precursor proteins. *J Biol Chem*. 277(46):43888–43894.

Salamini F, Özkan H, Brandolini A, Schäfer-Pregl R, Martin W. 2002. Genetics and geography of wild cereal domestication in the Near East. *Nat Rev Genet*. 3(6):429–441.

Sanderson MJ, Hufford L. 1996. Homoplasy and the evolutionary process: an afterword. In: Sanderson MJ, Hufford L, editors. Homoplasy. San Diego: Academic Press. p. 327–330.

Sawyer S. 1989. Statistical tests for detecting gene conversion. *Mol Biol Evol*. 6(5):526–538.

Schnell DJ, Blobel G, Keegstra K, Kessler F, Ko K, Soll J. 1997. A consensus nomenclature for the protein-import components of the chloroplast envelope. *Trends Cell Biol*. 7(8):303–304.

Sehrish T, Symonds VV, Soltis DE, Soltis PS, Tate JA. 2015. Cytonuclear coordination is not immediate upon allopolyploid formation in *Tragopogon miscellus* (Asteraceae) allopolyploids. *PLoS One* 10(12):e0144339.

Sharbrough J, Conover JL, Tate JA, Wendel JF, Sloan DB. 2017. Cytonuclear responses to genome doubling. *Am J Bot*. 104(9):1277–1280.

Sloan DB. 2015. Using plants to elucidate the mechanisms of cytonuclear co-evolution. *New Phytol*. 205(3):1040–1046.

Sloan DB, Triant DA, Wu M, Taylor DR. 2014. Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. *Mol Biol Evol*. 31(3):673–682.

Sloan DB, Warren JM, Williams AM, Wu Z, Abdel-Ghany SE, Chicco AJ, Havird JC. 2018. Cytonuclear integration and co-evolution. *Nat Rev Genet*. 19(10):635–648.

Soltis DE, Segovia-Salcedo MC, Jordon-Thaden I, Majure L, Miles NM, Mavrodiev EV, Mei W, Cortez MB, Soltis PS, Gitzendanner MA. 2014. Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al. (2011). *New Phytol*. 202(4):1105–1117.

Soltis PS, Soltis DE. 2009. The role of hybridization in plant speciation. *Annu Rev Plant Biol*. 60(1):561–588.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. Mega6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 30(12):2725–2729.

van Wijk KJ, Baginsky S. 2011. Plastid proteomics in higher plants: current state and future goals. *Plant Physiol*. 155(4):1578–1588.

VanderVere PS, Bennett TM, Oblong JE, Lamppa GK. 1995. A chloroplast processing enzyme involved in precursor maturation shares a zinc-binding motif with a recently recognized family of metalloendopeptidases. *Proc Natl Acad Sci U S A*. 92(16):7177–7181.

Walter P, Lingappa VR. 1986. Mechanism of protein translocation across the endoplasmic reticulum membrane. *Annu Rev Cell Biol*. 2(1):499–516.

Wang X, Dong Q, Li X, Yuliang A, Yu Y, Li N, Liu B, Gong L. 2017. Cytonuclear variation of Rubisco in synthesized rice hybrids and allotetraploids. *Plant Genome* 10(3):1–11.

Wang Y, Yu H, Tian C, Sajjad M, Gao C, Tong Y, Wang X, Jiao Y. 2017. Transcriptome association identifies regulators of wheat spike architecture. *Plant Physiol*. 175(2):746–757.

Weng M-L, Ruhlman TA, Jansen RK. 2016. Plastid–nuclear interaction and accelerated coevolution in plastid ribosomal genes in Geraniaceae. *Genome Biol Evol* . 8(6):1824–1838.

Wiesemann K, Simm S, Mirus O, Ladig R, Schleiff E. 2019. Regulation of two gtpases Toc159 and Toc34 in the translocon of the outer envelope of chloroplasts. *Biochim Biophys Acta Proteins Proteom*. 1867(6):627–636.

Wolf JB. 2009. Cytonuclear interactions can favor the evolution of genomic imprinting. *Evolution* 63(5):1364–1371.

Wolf JB, Hager R. 2006. A maternal-offspring coadaptation theory for the evolution of genomic imprinting. *PLoS Biol*. 4(12):e380.

Woodson JD, Chory J. 2008. Coordination of gene expression between organellar and nuclear genomes. *Nat Rev Genet*. 9(5):383–395.

Yakimowski SB, Barrett S. 2016. The role of hybridization in the evolution of sexual system diversity in a clonal, aquatic plant. *Evolution* 70(6):1200–1211.

Yakimowski SB, Rieseberg LH. 2014. The role of homoploid hybridization in evolution: a century of studies synthesizing genetics and ecology. *Am J Bot*. 101(8):1247–1258.

Zhai Y, Yu X, Zhu Z, Wang P, Meng Y, Zhao Q, Li J, Chen J. 2019. Nuclear-cytoplasmic coevolution analysis of rubisco in synthesized *Cucumis* allopolyploid. *Genes* 10(11):869.

Zhang H, Zhu B, Qi B, Gou X, Dong Y, Xu C, Zhang B, Huang W, Liu C, Wang X, et al. 2014. Evolution of the BBAA component of bread wheat during its history at the allohexaploid level. *Plant Cell*. 26(7):2761–2776.