PNAS RESEARCH ARTICLE | PLANT BIOLOGY OPEN ACCESS

A temporal gradient of cytonuclear coordination of chaperonins and chaperones during RuBisCo biogenesis in allopolyploid plants

Changping Li[a,1], Baoxu Ding[a,1], Xintong Ma[a,1], Xuan Yang[a], Hongyan Wang[b], Yuefan Dong[a], Zhibin Zhang[a], Jinbin Wang[a], Xiaochong Li[a], Yanan Yu[a], Yiyang Yu[a], Bao Liu[a], Jonathan F. Wendel[c], Yidan Li[d,2], Tianya Wang[a,2], and Lei Gong[a,2]

Edited by Douglas Soltis, University of Florida, Gainesville, FL; received January 21, 2022; accepted April 22, 2022

Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo) has long been studied from many perspectives. As a multisubunit (large subunits [LSUs] and small subunits[SSUs]) protein encoded by genes residing in the chloroplast (*rbcL*) and nuclear (*rbcS*) genomes, RuBisCo also is a model for cytonuclear coevolution following allopolyploid speciation in plants. Here, we studied the genomic and transcriptional cytonuclear coordination of auxiliary chaperonin and chaperones that facilitate RuBisCo biogenesis across multiple natural and artificially synthesized plant allopolyploids. We found similar genomic and transcriptional cytonuclear responses, including respective paternal-to-maternal conversions and maternal homeologous biased expression, in chaperonin/chaperon-assisted folding and assembly of RuBisCo in different allopolyploids. One observation is about the temporally attenuated genomic and transcriptional cytonuclear evolutionary responses during early folding and later assembly process of RuBisCo biogenesis, which were established by long-term evolution and immediate onset of allopolyploidy, respectively. Our study not only points to the potential widespread and hitherto unrecognized features of cytonuclear evolution but also bears implications for the structural interaction interface between LSU and Cpn60 chaperonin and the functioning stage of the Raf2 chaperone.

RuBisCo | cytonuclear coevolution/coordination | chaperonin | chaperone | allopolyploids

As the most abundant photosynthetic enzyme on earth (1, 2), ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo) has long attracted study, including that of its biogenesis (3), metabolic maintenance (4), engineering for increased catalytic specificity for CO_2 versus O_2 (5), and recombinant expression (3, 6–9). RuBisCo has also been widely employed in evolutionary biology, mostly for molecular phylogenetic inference using sequences of the large subunit (LSU), encoded by the relatively slow-evolving chloroplast *rbcL* genes (10–12). In contrast to the plastid-encoded LSU, the small subunit (SSU) is encoded by *rbcS*, which typically occurs in small multigene families that evolve nonindependently (13–15). Given that the RuBisCo enzyme complex is composed of both nuclear-encoded SSUs and cytoplasmically encoded LSUs, RuBisCo also has been an important model for the study of cytonuclear coevolution following allopolyploid speciation in plants (16–23).

Allopolyploid speciation entails the combination of two or more diploid parents with divergent nuclear genomes but in only one progenitor cytoplasm, which carries the mitochondrial and plastid genomes (24). Allopolyploid speciation is prevalent in the current plant kingdom (25–27), including in many well-studied models such as *Arabidopsis*, *Gossypium*, *Nicotiana*, *Tragopogon*, and *Triticum*. Of note, allopolyploid speciation is followed by evolutionary diploidization, which involves diverse genomic processes and mechanisms (24). Accompanying allopolyploid speciation, stoichiometric disruption and incompatibility between nuclear and cytoplasmic genomes, caused by the union of two nuclear genomes in only one set of parental organellar genomes (usually maternal), necessitates cytonuclear coevolution or coordination (16–23). Previous studies using RuBisCo have demonstrated nuclear, paternal-to-maternal gene conversion among subgenomic *rbcS* homologs leading to maternal-like paternal SSUs (carrying maternal conversions) and transcriptionally biased expression of maternal and maternal-like *rbcS* homologs (16–23). However, it is still underexplored about whether foregoing cytonuclear coordinative responses still hold for other nuclear-encoded proteins targeting to cytoplasmic organelles in the allopolyploid speciation process.

RuBisCo maturation involves folding and assembly stages, with a number of intermediate cytonuclear coencoded complexes formed from plastid-encoded LSUs and nuclear-encoded chaperonin/chaperones (Fig. 1). In the initial folding stage, nuclear-encoded chaperonin 60α, chaperonin 60β, cochaperonin 20 ,and cochaperonin 10 (hereafter abbreviated as Cpn60α, Cpn60β, and Cpn20/Cpn10, respectively) proteins

Significance

Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo), consisting of subunits encoded by nuclear and cytoplasmic genes, is a model for cytonuclear evolution in plant allopolyploids. To date, coordinated cytonuclear evolutionary responses of auxiliary cofactors involved in RuBisCo biogenesis remain unexplored. This study characterized and compared genomic and transcriptional cytonuclear coevolutionary responses of chaperonin/chaperones in RuBisCo folding and assembly processes across different allopolyploids. We discovered significant cytonuclear evolutionary responses in folding cofactors, with diminishing or attenuated responses later during assembly. Our results have general significance for understanding the unrecognized cytonuclear evolution of chaperonin/chaperone genes, structural and functional features of intermediate complexes, and the functioning stage of the Raf2 cofactor. Generally, the results reveal a hitherto unexplored dimension of allopolyploidy in plants.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]C.L., B.D., and X.M. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: liyidan@foxmail.com, wangty309@nenu.edu.cn, or gongl100@nenu.edu.cn.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2200106119/-/DCSupplemental.

Published August 15, 2022.

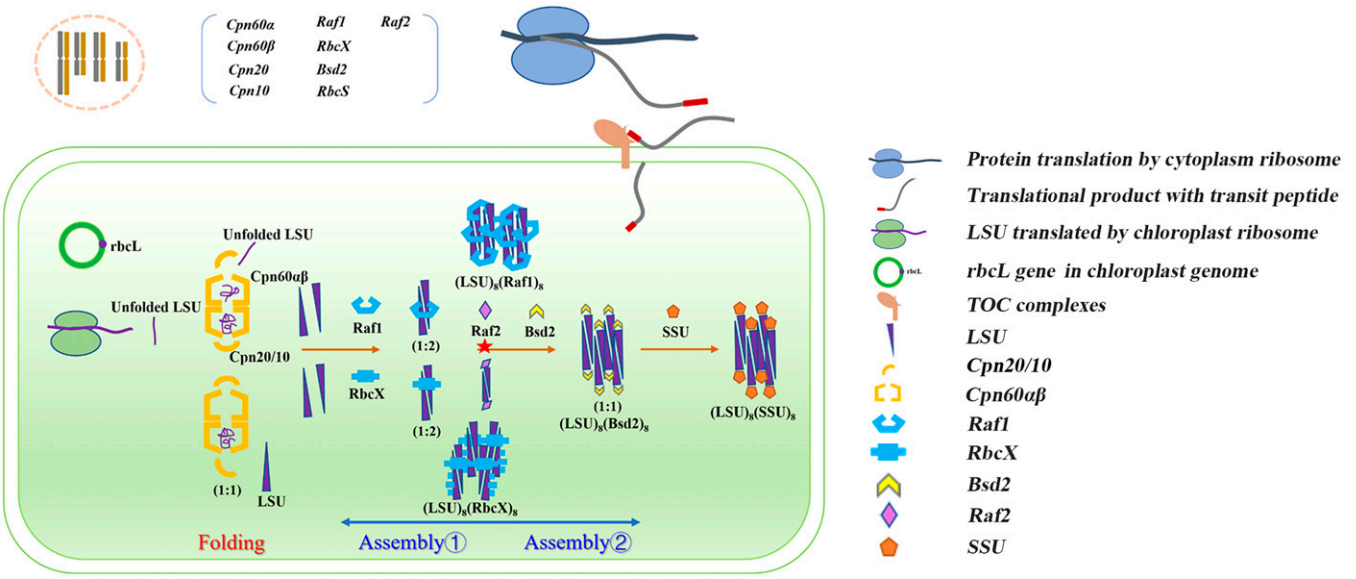

Fig. 1. Overview of folding and assembly stages/processes of RuBisCo biogenesis. Nuclear-encoded premature chaperonins, chaperones, and RuBisCo SSUs in parenthesis are targeted to chloroplasts by their N-terminal transit peptides and transported through TOC (outer membrane translocon) complexes into chloroplasts. Within chloroplasts, LSUs produced by translation from chloroplast-genome *rbcL* transcripts are input into subsequent folding and assembly steps (subdivided into assembly-1 and -2 stages). Intermediate complexes are arranged according to their functional position in the respective stage (8). Component combinations and ratios of LSU vs. respective auxiliary chaperonin/chaperon proteins or SSUs in each intermediate complex are labeled beneath each complex. The functioning stage of the Raf2 cofactor has not been ascertained; based on results from the present study, its most likely positioning in assembly is denoted by the red star.

are assembled into a complex (in a 1:1:1:1:1 ratio of subunits relative to LSU; Fig. 1), which folds LSU into a configuration appropriate for subsequent polymerization (28). Subsequently, nuclear-encoded RuBisCo accumulation factor 1 (hereafter abbreviated as Raf1) and/or RbcX assemble LSUs into dimers and even higher oligomers (in 1:2 or 2:2 ratio relative to LSU; Fig. 1), for assembly with other chaperones (28). RuBisCo accumulation factor 2 (hereafter abbreviated as Raf2) chaperones are implicated as binding with this complex (in an undetermined ratio relative to LSU) in a subsequent second assembly step, downstream of LSURaf1 and LSURbcX complex formation, or perhaps at a folding stage upstream of the Cpn60αβ/Cpn20 chaperonin complex. It is thought that the bundle sheath defective 2 (hereafter abbreviated as Bsd2) chaperone replaces the previous assembly chaperones to bind with LSU octamers (in a 1:1 ratio relative to LSU; Fig. 1), facilitating the binding of SSU with LSU to form the RuBisCo holoenzyme (28). To the best of our knowledge, none of the aforementioned proteins involved in RuBisCo biogenesis have been explored with respect to cytonuclear coevolutionary responses let alone comparing their cytonuclear coordinative responses in terms of their functional position in RuBisCo folding and assembly steps.

Here, we explore their cytonuclear evolution in multiple angiosperm allopolyploid models, including *Arabidopsis*, *Gossypium*, *Nicotiana*, and *Triticum*, each of which has a well-characterized phylogeny and living models of progenitor or parental diploids. Our focus was on nuclear-encoded chaperonin and chaperone genes, for which we explored copy number, genomic variation, and transcriptional bias in the context of folding and assembly steps potentially subjected to phenomena that might be reflective of cytonuclear interaction and selection. By quantifying and comparing genomic (intergenomic homeologous gene conversion) and transcriptional (maternal homeologous biased expression) cytonuclear coordination of chaperonins and chaperones across multiple natural and artificially synthesized allopolyploids, we demonstrate similar genomic and transcriptional cytonuclear responses of genes involved in chaperonin/chaperon-assisted folding and assembly of RuBisCo in different allopolyploids. We also show that these responses are temporally attenuated during the folding and assembly processes. Our results have general significance for understanding the cytonuclear evolution of chaperonin/chaperone encoding genes, as well as for elucidating structural and functional features of intermediate protein complexes during RuBisCo biogenesis.

Results

Reduction in Copy Numbers of Chaperonin and Chaperone Genes Downstream Relative to Upstream. Gene families encoding chaperonin and chaperone proteins during RuBisCo folding and assembly (6) have previously been characterized in some diploid angiosperms (29). Here, we tabulated chaperonin and chaperone gene families (Fig. 1) and placed them in a phylogenic context (Fig. 2). Notably, chaperonin proteins involved in the upstream folding process (including Cpn60α, Cpn60β, and Cpn20/Cpn10) were encoded by multiple copies, with the numbers of gene copies encoding *cpn60α*, *cpn60β*, and *cpn20/cpn10* being approximately the same (Fig. 2). Given their integration into the same Cpn60αβ/Cpn20/Cpn10 chaperonin complex, such a multiple-copy existence implicates gene balance as a constraint on copy number (30, 31). In contrast, the chaperone factors involved in RuBisCo assembly were encoded by fewer genes than those that function in the folding process (Figs. 1 and 2). Notably, there was no apparent association between gene copy number of each gene family with the number of whole-genome duplications (WGDs) in each lineage. In addition, a unique case involves *raf2*, which encodes an essential Raf2 protein for RuBisCo biogenesis. Without exception across the angiosperms studied, this protein was encoded by a single-copy gene (Fig. 2).

Attenuated Cytonuclear Genomic Responses to Polyploidy. Previous studies in allopolyploids have demonstrated coordinated cytonuclear evolution between the maternally derived

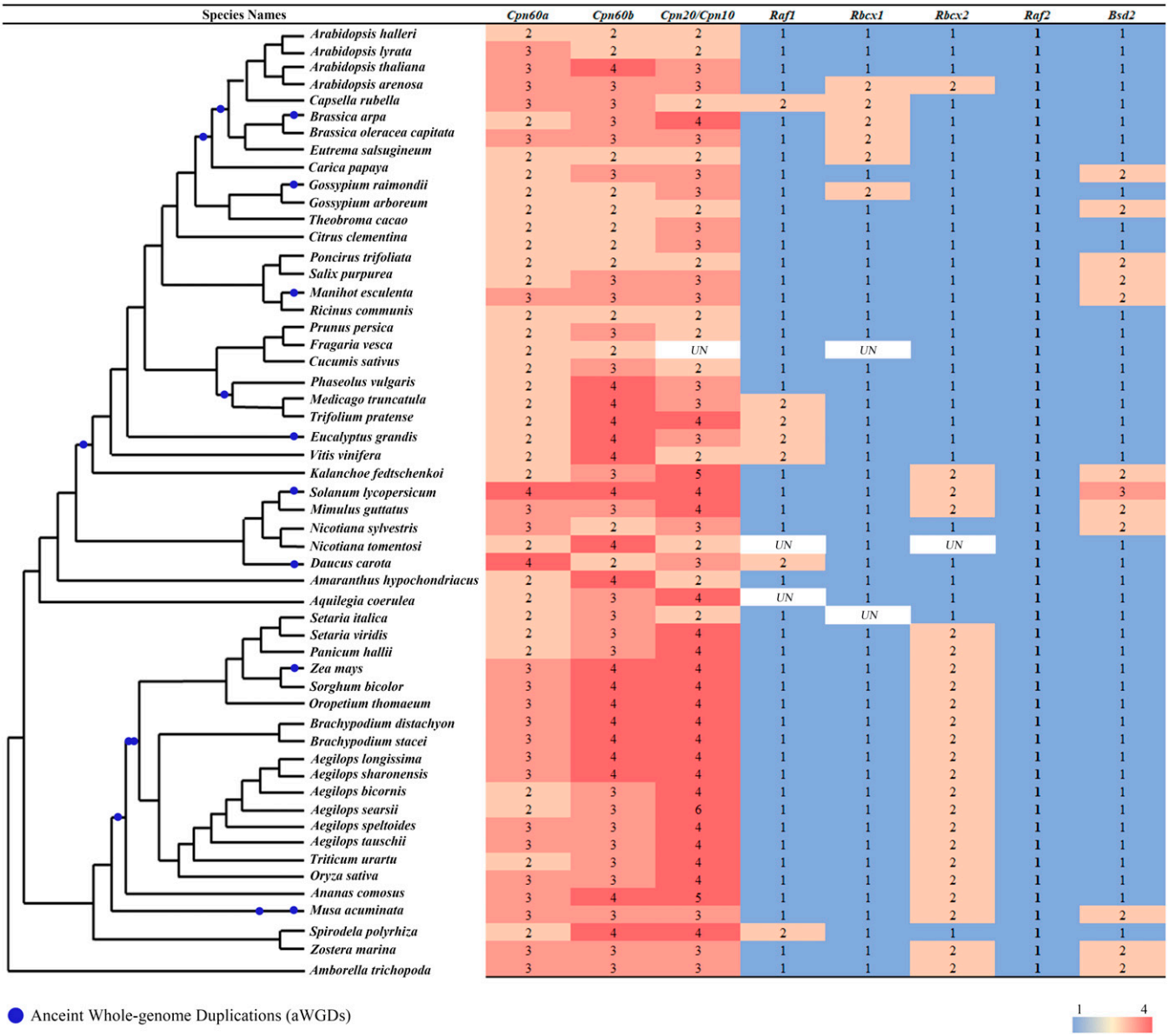

Fig. 2. Copy numbers of chaperonin and chaperon genes in representative sequenced diploid angiosperm species. Sequenced diploid angiosperm species are listed according to their phylogenetic relationships (revised from the dendrogram tabulated in Phytozome V12.1). In each cell of the table, copy numbers of chaperonin (in Cpn60α to Cpn20/Cpn10 columns) and chaperone (in Raf1 to Bsd2 columns) genes in respective species are specified and their relative abundance are denoted in gradient colors. UN denotes the undetermined cases, possibly reflecting missed or unannotated copies. Those ancient WGDs (aWGDs) are denoted in the respective branch or node . Similar numbers of multiple-copy *cpn60α*, *cpn60β*, and *cpn20*/*cpn10* genes across species are statistically tested by paired Wilcoxon rank sum tests, $P > 0.05$.

chloroplast *rbcL* (and its encoded LSU) and the biparentally inherited *rbcS* genes encoding the SSU (16, 17, 19, 21, 23). At the genomic level, this coordination involves paternal-to-maternal gene conversion of *rbcS* homeologs, which results in maternal-like, paternal *rbcS* copies (16, 17, 19, 21, 23). It is established that the folding process involves significant protein structural alterations in LSU oligomers, whereas the assembly process completes the polymerization of individual folded LSUs (8). Given these different effects of folding and assembly cofactors on LSUs during RuBisCo biogenesis, we hypothesized that chaperonins working at the folding stage would display more evidence of genomic cytonuclear coordination (more paternal-to-maternal gene conversions) than chaperones functioning during the later assembly stage.

To test this hypothesis, we characterized gene conversions within chaperonin/chaperone-encoding homeologs for genes encoding auxiliary proteins that function during either stage (assembly stage was subdivided into assembly-1 and assembly-2) (28). Nonsynonymous gene conversions were determined by aligning translated open reading frames of diploid chaperonin/chaperone homologs with their respective subgenomic chaperonin/chaperone homeologs in the allopolyploids. Intersubgenomic homeologous single nucleotide polymorphism (SNP) changes leading to amino acid changes in the mature protein regions (regions encoding N-terminal transit peptide regions were excluded) were identified as nonsynonymous converted SNPs (16, 17, 23). To quantify and compare this phenomenon within and among different chaperonin/chaperone genes, nonsynonymous converted SNPs and their corresponding amino acid changes (example as illustrated in *SI Appendix*, Fig. 1; also *Materials and Methods*) were normalized by the total number of polymorphic amino acid positions within each full-length sequence alignment. These normalized paternal-to-maternal and

maternal-to-paternal amino acid conversions were defined as positive and negative genomic cytonuclear coordination signals (GCCSs), respectively. Positive and negative GCCSs were averaged for each chaperone family acting in both the folding and assembly stages, for each allopolyploid studied.

As tabulated (*SI Appendix*, Table 2) and summarized (Fig. 3), the protein factor having the highest positive average GCCS values varied among genes and species (e.g., *cpn20* in *Gossypium hirsutum* and *Nicotiana tabacum* and *cpn60β* in *Arabidopsis suecica*, *Triticum turgidum*, and *Triticum aestivum*). In addition, the average positive GCCSs exhibited an attenuation trend, which involved higher GCCSs of folding-stage chaperonins than those of assembly-stage chaperones in most sampled allopolyploids (Fig. 3 *A* and *B*; Wilcoxon rank sum test, $P < 0.05$). Notably, there were some exceptional cases with negative GCCSs or maternal-to-paternal conversion for some chaperonin and chaperones in some sampled allopolyploids, including *cpn60α* in *N. tabacum*, *cpn60β* and *cpn10* in *G. hirsutum*, *rbcX* in *N. tabacum*, and *raf2* in *T. aestivum* (D-subgenomic portion) (Fig. 3 *A* and *B*). However, since the directionality of the attenuation trend is too common to be by chance, the overall attenuated positive GCCCs during RuBisCo biogenesis were still supported.

To further explore the nonsynonymous gene conversions in chaperonin and chaperone genes, we aligned the diploid chaperonin and chaperone homologs with their respective homeologs in all sampled allopolyploids (*SI Appendix*, Fig. 1). For Cpn20, with significantly high positive GCCSs in *G. hirsutum* and *N. tabacum*, GCCS signals mainly occur around the protein interface and adjacent regions that interact with LSUs (*SI Appendix*, Fig. 1 *A* and *B*; 32), where nonsynonymous substitutions between parental diploid homologs were detected. For these same species, other chaperonins/chaperones had limited or no GCCSs (including Cpn10, Raf1, RbcX, and Bsd2), which corresponded to parental diploids where homologs lacked nonsynonymous substitutions in the LSU interface regions (28, 30, 33–36). For Cpn60β, with a significant negative GCCS in *G. hirsutum*, its GCCSs were detected at a certain position (the 315th aligned amino acid position in *SI Appendix*, Fig. 1C). For *Triticum*, positive paternal-to-maternal GCCSs were detected in the paternal homeolog of *Cpn60β*; whereas *Cpn20* did not have nonsynonymous substitutions in the interface regions with LSU and did not display any GCCS (*SI Appendix*, Fig. 1*D*). Together, our results suggest that the protein compositional difference of chaperonins and chaperones in their interface regions with LSUs could have provided a selective force in driving the fixation of nonsynonymous gene conversions following allopolyploidy.

Attenuated Cytonuclear Transcriptomic Responses to Allpolyploidy. Biased maternal homeologous transcriptional expression has previously been observed for some genes in allopolyploids (16–20, 22). Here, we quantified the ratio of maternal vs. paternal homoelogous expression (evaluated in fragments per kilobase of transcript per million mapped reads) for each pair of chaperonin and chaperone homeologs and termed this ratio as transcriptomic cytonuclear coordination signal (TCCS) (Fig. 4). This ratio was statistically tested for departures from equivalence using DESeq2 (*Materials and Methods*). The results showed that 1) although the TCCSs varied among chaperonin/chaperone genes at the folding and assembly stages in the same allopolyploid species, there was at least one folding stage chaperonin gene displaying statistically significant positive transcriptomic maternal bias (significant positive TCCS) in all allopolyploid species, which included the chaperonin *cpn60β*, *cpn20*, and/or *cpn10* genes (Fig. 4). However, there were occasional significant TCCSs of chaperone at the assembly stage in *Triticum* allopolyploids (Fig. 4); 2) the genes showing the highest bias were the folding stage chaperonins *cpn20* and *cpn60β* (Fig. 4); 3) in most cases, significant transcriptomic maternal biases were higher during the folding stage and lower to nonexistent during the assembly stage (Fig. 4), when comparing TCCSs of paired folding chaperonin vs. assembly chaperone genes with statistical significant bias (exceptional cases involve Cpn20-1 vs. Raf2 in *T. turgidum* ssp. *dicoccoides*, Cpn20-2 vs. Raf1 in *T. aestivum* A-subgenome portion). Thus, Cpn20 and Cpn60β chaperonins at the folding stage appear to exhibit similar attenuated cytonuclear evolutionary responses to allopolyploidy at both the genomic and transciptomic levels (Figs. 3 and 4).

Immediate Versus Evolved Genomic and Transcriptional Cytonuclear Responses to Allpolyploidy. An interesting follow-up question is about the timing of genomic and transcriptomic cytonuclear responses to the shock of genome merger and doubling. Are these established immediately with allopolyploidization or are they evolved responses? To address this question, we characterized genomic and transcriptomic responses in the synthetic allotetraploids *A. suecica* (738) and *G. hirsutum* (2(A2D1)) and two synthetic allotetraploid wheat lines (AADD and S^LS^LAA), using similar analyses as described above (Figs. 3*C* and 5).

As for the genomic responses, we detected both positive and negative GCCSs within synthesized allopolyploids. Specifically, for the synthetic AADD wheat and the 2(A2D1) cotton allotetraploids, the nonsynonymous averaged positive GCCSs of all chaperonins and chaperons were mostly the same. However, within synthesized *Arabidopsis* and S^LS^LAA wheat allotetraploids, positive GCCSs at the assembly stage of RuBisCo biogenesis were higher than those at the folding stage (Fig. 3*C*). As expected, in synthetic allopolyploids, there was no attenuation in positive genomic responses as observed in natural allopolyploids (Fig. 3 *A* and *B*). In contrast, there were even notable negative GCCSs for many chaperones at the folding and assembly stages in the synthetic allopolyploids (Fig. 3*C*), suggesting that gene conversions in both directions occurred within the initial generations after allopolyploid synthesis. Our observations not only support the attenuated patterns of cytonuclear responses in natural allopolyploids but also demonstrate they are evolved responses.

With respect to transcriptomic responses, one might expect that the combination of *cis* and *trans* factors from divergent diploids into a common nucleus might disrupt progenitor expression patterns. Here, being similar as those analyzed natural allopolyploids, we characterized and compared the TCCSs of cofactors at folding and assembly stages in artificial (synthesized) allopolyploids (Fig. 5). Notably, although the chaperonin/chaperones with transcriptional maternal bias in each stage were occasionally different in synthetic and natural allopolyploids (Fig. 5; i.e., Cpn60β vs. Cpn20 in synthetic and natural *Gossypium* allopolyploids), the TCCSs with statistically significant maternal biases were higher during the folding than the assembly stage in most synthesized allopolyploids, except Cpn20-3 vs. RbcX-1 in synthesized AADD wheat. Such a result is consistent with the aforementioned observations for natural allopolyploids (Figs. 4 and 5). The implication of these results is that transcriptional cytonuclear responses to genome merger and doubling can be immediate, or in the early generations, with the onset of allopolyploidization.

Discussion

The well-characterized protein components of the folding and assembly processes during RuBisCo biogenesis provide an excellent

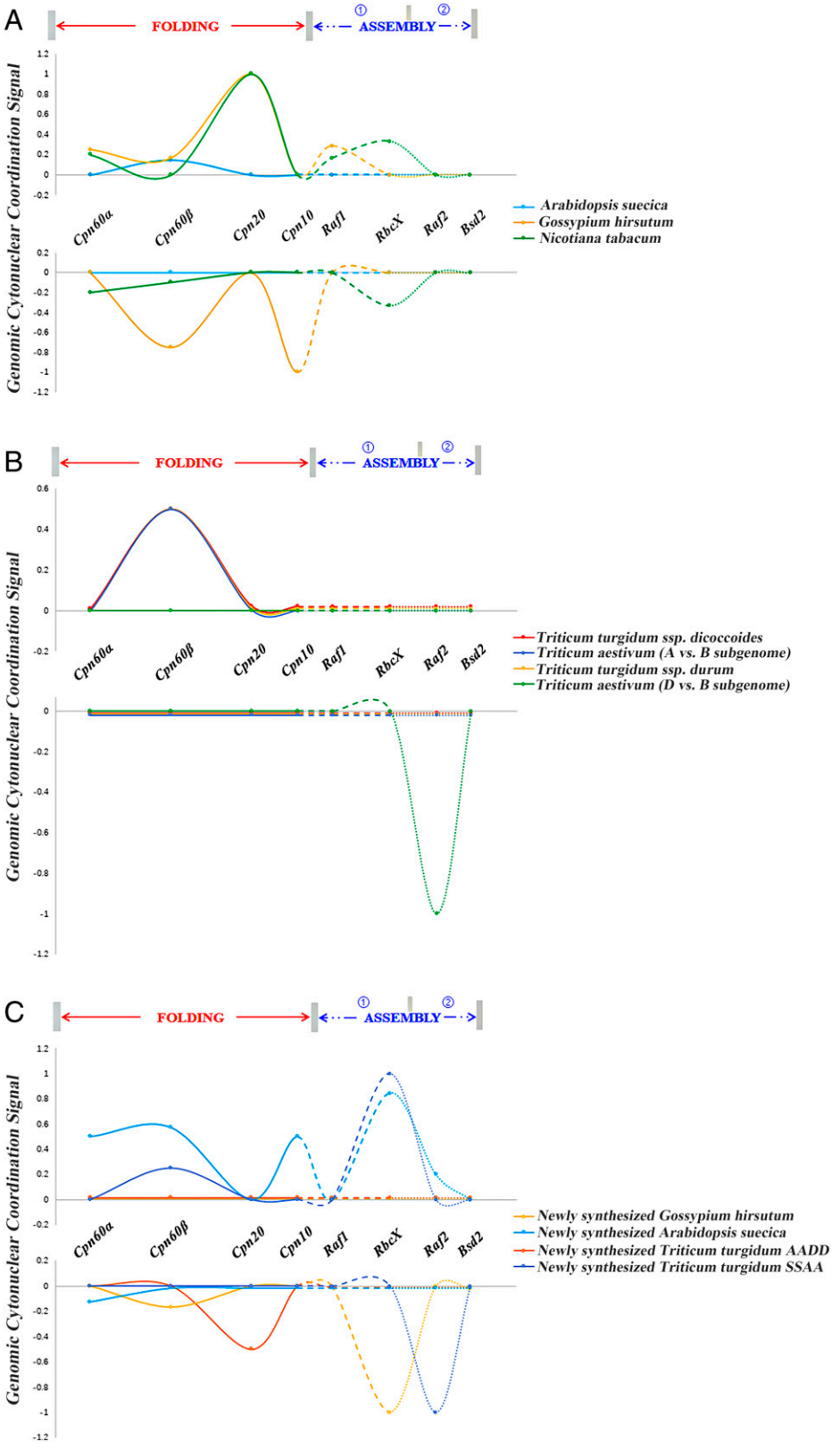

Fig. 3. Averaged GCCSs of chaperonin and chaperone genes in the folding and assembly processes during RuBisCo biogenesis in representative allopolyploids. Chaperonin and chaperone genes are labeled on the x-axis, which are arranged in accordance with their functional position in folding and assembly stage (denoted at the top of each panel); on the *y* axis, positive and negative averaged GCCSs are shown. The natural allopolyploids *A. suecica*, *G. hirsutum*, and *N. tabacum* (in *A*), the wild and domesticated *T. turgidum* subspecies and the paternal A- and D-subgenomes of *T. aestivum* (in *B*), and newly synthesized allopolyploids (in *C*) are compared and summarized. Signals of chaperonin genes at folding stages are linked with solid lines, while those of chaperone genes at the assembly substages 1 and 2 are within-group linked using dashed lines and dotted lines, respectively.

opportunity for exploration of the cytonuclear consequences of genome merger and doubling, where the nuclear genome is doubled yet in a context of plastid and mitochondrial genomes from only one of the progenitor genomes. Here, we elucidate the cytonuclear genomic and transcriptomic responses of chaperonin and chaperone genes across a wide range of different

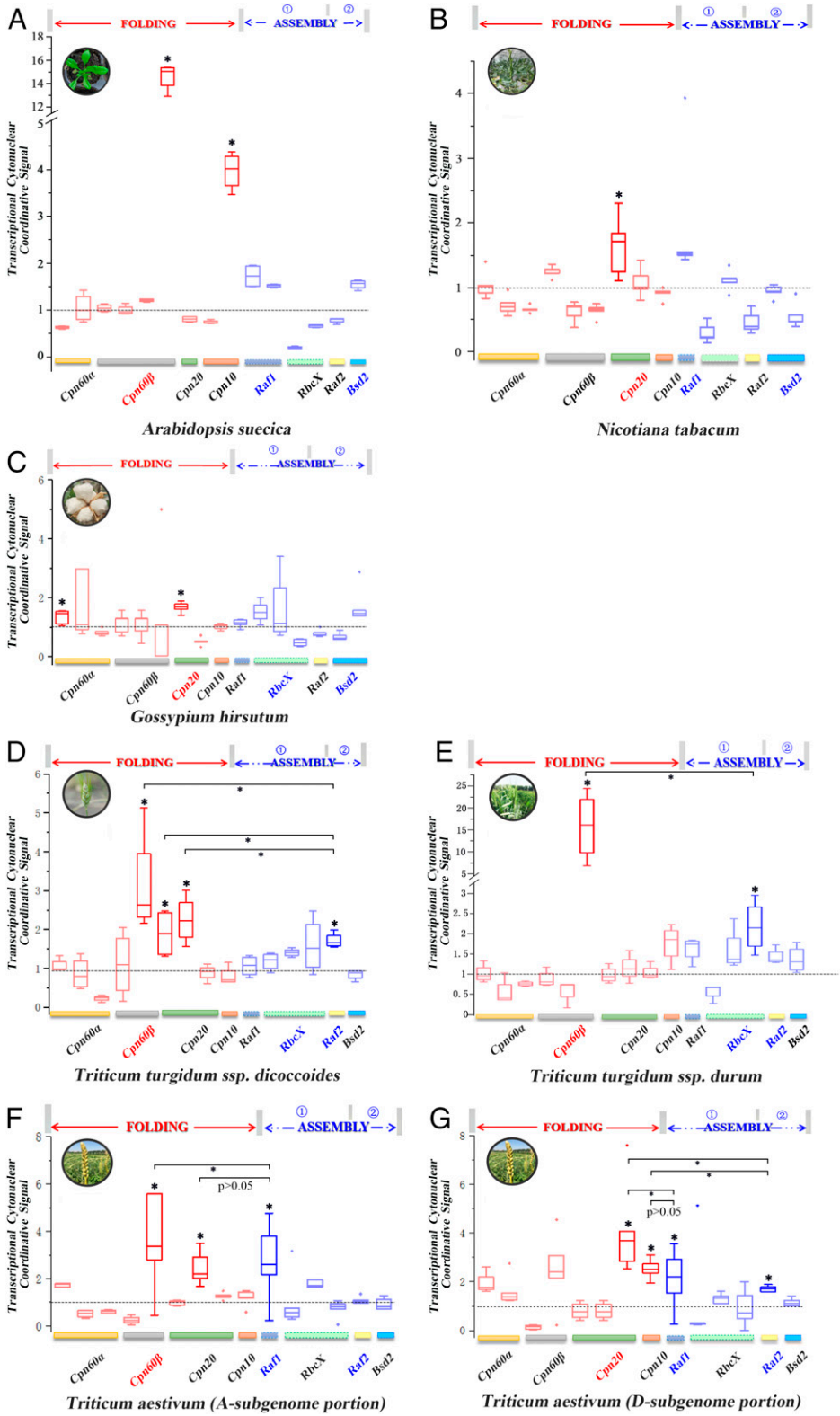

Fig. 4. TCCSs of each chaperonin and chaperone gene in folding and assembly processes during RuBisCo biogenesis in representative natural allopolyploids. Chaperonin and chaperone genes functioning in folding and assembly stages (denoted in red and blue at the top of each panel) are arranged along the x-axis; corresponding TCCSs for each homeolog pair, estimated as the expression ratio of maternal vs. paternal chaperonin/chaperone gene homeologs, are summarized in boxplots. Statistically significant TCCS boxes (biased maternal expression is determined in DESeq2 and depicted above the horizontal nonbiased homeologous expression dotted line; *Materials and Methods*) are denoted by asterisks above boxes. Genes in folding (red) and assembly (blue) stage with significant and nonsignificant TCCSs are summarized in dark and light boxes, respectively. As for species in *A* to *C* (corresponding to *A. suecica*, *N. tabacum*, and *G. hirsutum*), since there is no significant TCCS detected in any assembly chaperone, the comparison of TCCSs between folding chaperonin and assembly chaperone are naturally accepted as being significant; as for species in *D* to *G* (corresponding to wild and domesticated *T. turgidum* subspecies and the paternal A- and D-subgenomes of *T. aestivum*), iterative comparisons of TCCSs are made between paired cofactors with significant bias at folding and/or assembly stage (pairs are linked by solid horizontal line). Respective statistical significant (in asterisk) and nonsignificant ($P > 0.05$) comparisons are denoted beneath the pairing line.

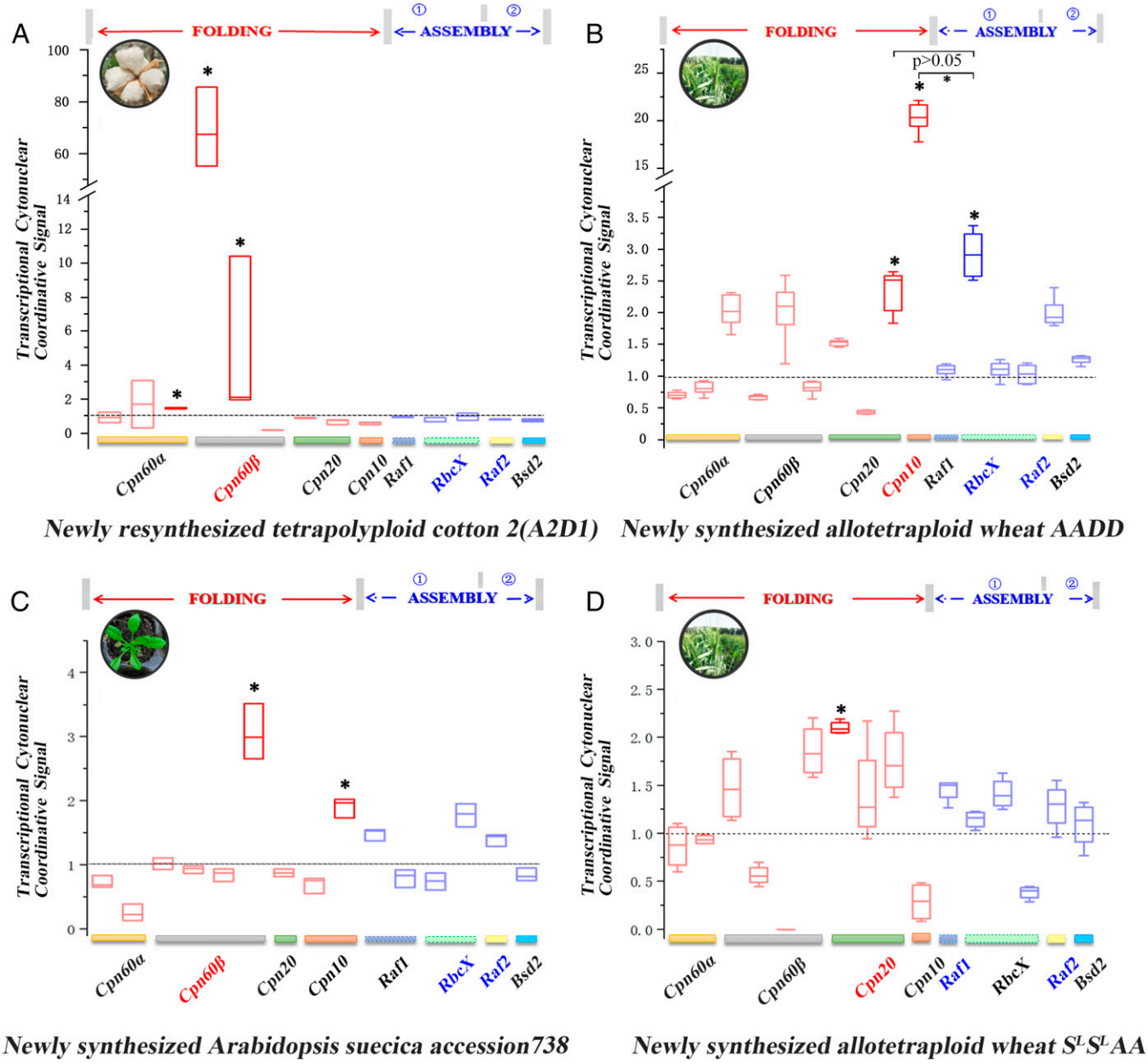

Fig. 5. TCCSs of each chaperonin and chaperone gene in the folding and assembly processes of RuBisCo biogenesis as revealed in synthetic allopolyploids. All symbols, labels, and annotations are the same as those in Fig. 4. The newly synthesized allopolyploids, including 2(A2D1) cotton, AADD allotetraploid wheat, *A. suecica* accession 738, and S^LS^LAA allotetraploid wheat, are illustrated in *A*–*D*, respectively.

allopolyploid systems, as well as in synthetic allopolyploids and along a temporal gradient with respect to the steps of RuBisCo biogenesis. We show that in four different genera across the angiosperm phylogenetic spectrum, there are similar patterns of apparent cytonuclear evolutionary responses to allopolyploidy. These include differences in paternal vs. maternal gene conversion in the predicted direction based on maternal parentage and alterations in homeolog expression from the two progenitor diploids that also favor the maternal parent. An observation here is the attenuated cytonuclear response, where the maximum effect is observed for proteins acting early during RuBisCo biogenesis, with diminishing or attenuated evolutionary responses for later stages. Our study points to the possibility that this is a more widespread and hitherto unrecognized feature of cytonuclear evolution.

Fate of WGD Chaperonin/Chaperone Gene Duplicates in RuBisCo Biogenesis Can Be Explained by Dosage Balance and Dominant Negative Mutation. It has been established that all modern diploid flowering plants are paleopolyploids, having experienced multiple rounds of ancient WGD followed by recurrent diploidization (including angiosperms in Fig. 2; 24, 37–40). One of the key mechanisms thought to explain the variable patterns of duplicate gene retention following WGD is the gene dosage balance hypothesis (31, 41, 42), which postulates that after WGD, genes encoding component subunits of specific protein complexes are preferentially maintained in appropriate dosages to maintain balanced interactions within the protein complex (31, 41–43, 44, 45). Although gene balance is unlikely to entirely explain variable patterns of duplicate gene loss and retention (43, 44, 46–48), this framework

offers an appropriate lens for mosaic nuclear- and chloroplast-encoded protein complexes as well (44); because the number of chloroplasts increases after polyploidy (49–52), the ratio of chloroplast-encoded genes vs. nuclear-encoded genes in the same mosaic complex are often to remain relative stable, perhaps accounting for at least part of the maintenance of duplicate genes for some gene families (53, 54).

Given the increased number of chloroplast genomes following both ancient and recent allopolyploidy (55), we suggest that the maintenance of dosage balance might play an important role in the retention of multiple copies of *cpn60α*, *cpn60β*, and *cpn20/cpn10* (Fig. 2), counteracting the concurrent diploidization processes (56, 57). Copy numbers for other chaperones in the assembly process (including Raf1, RbcX1, RbcX2, Raf2, and Bsd2; Fig. 2) appear to have been mostly reduced following WGDs, thus requiring explanations apart from dosage balance, including a dominant-negative mutation model (44, 46), and variable constraints based on position in the pathway or network (58, 59).

In the present example, the coexistence of multiple- or single-copy RuBisCo biogenesis genes may be related to the well-understood protein stoichiometries (3, 28, 30, 33, 35, 36), in which the ratio of LSU vs. chaperonin/chaperon proteins in each functional unit is known to change from 1:1 during folding to 2:1 during assembly (Fig. 1). Accordingly, the increased dosage of LSUs after WGD is consistent with a 1:1 ratio of LSU:chaperonin at the folding stage given the doubling of nuclear genomes. The 2:1 ratio at the assembly stage could release the selection from increased LSU dosage or cause less constraint for the genes in the latter RuBisCo biogenesis pathway, under which the dominant-negative mutation mainly directs returning to single copy for those chaperone protein-encoding genes after WGD. Such a supposed mechanism still needs further experimental and/or evolutionary confirmations.

Cytonuclear Evolutionary Responses to Allopolyploidy Operate at Different Paces. Previous work has established two common cytonuclear responses to allopolyploidy, i.e., paternal-to-maternal gene conversion and maternally biased gene expression at the level of the transcriptome. Most of this work has involved RuBisCo, due to the asymmetry in the location of its protein-encoding components (16–23). By extending this analysis to include auxiliary proteins involved in RuBisCo biogenesis, we asked whether 1) the footprints of cytonuclear evolution that often are apparent for *rbcS* also are evident for chaperonin and chaperone genes that participate in RuBisCo folding and assembly. We further asked whether 2) the pace of apparent cytonuclear evolutionary responses to allopolyploidy varies among these protein participants.

With respect to the first question, we show both positive genomic and transcriptional cytonuclear coordination, as reflected by frequent intergenomic paternal-to-maternal conversions and maternally biased expression among diverse allopolyploids, but intriguingly, these responses are unequal during RuBisCo biogenesis, becoming attenuated from early to later during protein assembly (Fig. 3 and 4). This diminution of an apparent cytonuclear evolutionary response may reflect the intimacy of interaction of each chaperonin and chaperone with the LSU. If we assume a Cytonuclear Coordination Olympic Game for the relevant chaperonins and chaperones, Cpn20 or Cpn60β at the folding step win both the genomic and transcriptional marathon race, where marathon implies a long-term evolutionary outcome following allopolyploidy. Noting that the chaperonins and chaperones studied here are also involved in folding of other protein cytonuclear coencoded protein complexes (i.e., NADH complex in the electron transport chain; 57), the possibility arises that the evolutionary patterns we observe reflect cytonuclear pressure from these other complexes in addition to RuBisCo. In other words, additional organellar component subunits could also be judges in the cytonuclear Olympic competitions. It will be of interest to explore this possibility in future studies.

As for the second question, by characterizing and comparing genomic and transcriptional responses among natural and artificial synthetic allopolyploids, we discovered that apparent cytonuclear evolutionary responses are attenuated during RuBisCo biogenesis (Fig. 4), even immediately with the onset of allopolyploidy (Fig. 5). In an analogy to the above marathon race, Cpn20 or Cpn60β at the folding step win the transcriptional 100-meter dash (Cpn20 or Cpn60β harboring the highest maternal biased expression), where dash conveys an immediate evolutionary response to allopolyploidy. Although the underlying molecular mechanism for this diminution in genomic or transcription bias for proteins facilitating RuBisCo biogenesis remains uncharacterized, we speculate that this reflects transcriptional feedback effects from the maternally inherited plastid to the nucleus, e.g., via reprogrammed epigenomic modification (60), divergence in promoter regions of parental homeologs (61), or differences in binding affinity of transcription factors to homeologous promoters (62).

In contrast to transcriptional cytonuclear coordination, genomic changes during the initial generations of allopolyploidy appear to be more random with few gene conversions and in either direction (Fig. 3). These few but apparently random gene conversions could reflect 1) genomic damage and subsequent homeologous repair (63, 64); 2) the lack of subgenomic recombination suppressive genes in the diploid parents, such as the absence of Ph1 gene in AADD and S^LS^LAA wheat (65), or the presence of a ph1-like allele in diploid *A. thaliana* (66) with potential modifications of the meiotic machinery in the *A. arenosa* subgenome in synthetic *A. suecica* (67, 68), allowing pairing and recombination of homeologous chromosomes; or 3) reprogrammed epigenomic landscape caused by allopolyploidy, e.g., in allopolyploid cotton (69), which could promote homeologous pairing and recombination as well. Taken together, our data support the notion that cytonuclear coevolution is a slow process. Future insights into the pace of cytonuclear evolution are likely to derive from comparisons of multiple allopolyploids of varying age, either within genera (for example, *Nicotiana*; 70) or from additional, broader surveys among angiosperms.

Implications for the Structural Interaction Interface between LSU and Cpn60 Chaperonin and the Functioning Stage of the Raf2 Chaperone. Understanding the intermediate folding and assembly complex involving LSUs and their interacting auxiliary cofactors is essential for interpreting molecular evolution in the interaction face between component subunits (28, 30, 32, 33, 35, 36). Because the intermediate complexes are not easily expressed or purified for protein structural analyses (6), the interaction interface between Cpn60β and LSU and the functioning stage of the Raf2 cofactor interacting with LSUs in RuBisCo biogenesis have remained elusive. Here, we augment more traditional molecular protein–protein interaction assays with evolutionary considerations. Specifically, we note that Cpn60β has a domain with clusters of significant positive GCCSs, as might be expected if these function at the interface region with LSU during folding (Fig. 1 and *SI Appendix*, Fig. 1). With respect to the undetermined functioning location of cofactors interacting with LSUs

in complex, perhaps less compelling, but suggestive, is the case for Raf2, where our interpretation that its interaction with LSUs and function during assembly is evolutionarily supported by its single-copy composition in all sampled diploid plant species (Fig. 2) and limited genomic cytonuclear coordination (Fig. 3) similar to other assembly chaperones (Fig. 1). Evidence testing these suggestions awaits biophysical or other analyses that are more direct.

Materials and Methods

Plant Materials. Four representative monocot and eudicot allopolyploid lineages, including both natural and synthesized allopolyploids and diploid parents, were selected for our cytonuclear analyses (*SI Appendix, Materials and Methods*), which include *Arabidopsis* (71), *Gossypium* (72), *Nicotiana* (73), and *Triticum* (74).

Gene Sequence Retrieval, Cloning, and Sequencing. Chaperonin and chaperone gene family composition and gene sequences in different allopolyploid lineages were characterized using genomic searching. To validate the retrieved sequences, we designed homeolog-specific primers to complete PCR amplification and cloning for each chaperonin and chaperone homeolog coding DNA sequence in each allopolyploid lineage (*SI Appendix*, Table 1). Additionally, we also searched for respective homologs in 41 genome assemblies of representative diploid angiosperms deposited in Phytozome v12.1. Details of the technical procedure are described in *SI Appendix, Materials and Methods*.

Cytonuclear Coevolution at the Genomic and Transcriptomic Level. To quantify and compare genomic cytonuclear coordination at different stages in RuBisCo biogenesis, we defined the GCCS for each chaperonin and chaperone gene at the sequence level as the ratio of nonsynonymous, presumptively converted, amino acids vs. the total of nonsynonymous amino acid differences in each pair of parental mature proteins (*SI Appendix*, Table 2). Positive paternal-to-maternal and negative maternal-to-paternal intergenomic gene conversions were discriminated (*SI Appendix*, Table 2). To characterize possible cytonuclear coordination at the transcriptomic level, for each chaperonin and chaperone homeologous pair, we defined and calculated the TCCS as the subgenomic expression ratio of maternal vs. paternal chaperonin/chaperone gene homeologs in leaf RNA Sequencing (RNAseq) data (expression level estimated by reads per kilobase of transcript, per million mapped reads; *SI Appendix*, Table 3 and 4). Details of foregoing coevolutionary analyses are described in *SI Appendix, Materials and Methods*.

Data Availability. All study data are included in the article and/or supporting information.

ACKNOWLEDGMENTS. This work was supported by the National Natural Science Foundation of China (Grant no. 31970238). We appreciate the knowledge and trainings given by the course of Evolution Biology in Northeast Normal University. We are grateful for the technical supports and the data provided by Lili Jiang (lab technician) in Northeast Normal University.

Author affiliations: [a]Key Laboratory of Molecular Epigenetics of the Ministry of Education (MOE), Northeast Normal University, Changchun, Jilin Province, 130024, China; [b]Laboratory of Plant Epigenetics and Evolution, School of Life Science, Liaoning University, Shenyang, 110036, China; [c]Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA, 50011 and [d]Institute of Agricultural Biotechnology, Jilin Academy of Agricultural Sciences, Changchun, Jilin Province, 130024, China

Author contributions: B.L., Y.L., T.W., and L.G. designed research; C.L., B.D., X.M., X.Y., H.W., Y.D., Z.Z., J.F.W., X.L., Yanan Yu, Yiyang Yu, Y.L., and T.W. performed research; Y.D., J.F.W., and Y.L. contributed new reagents/analytic tools; C.L., B.D., X.M., X.Y., H.W., Y.D., Z.Z., J.F.W., X.L., Jinbin Wang, Yanan Yu, Yiyang Yu, T.W., and L.G. analyzed data; and C.L., B.D., X.M., B.L., J.F.W., Y.L., T.W., and L.G. wrote the paper.

1. R. J. Ellis, The most abundant protein in the world. *Trends Biochem. Sci.* **4**, 241–244 (1979).
2. J. A. Raven, Rubisco: Still the most abundant protein of Earth? *New Phytol.* **198**, 1–3 (2013).
3. H. Aigner *et al.*, Plant RuBisCo assembly in *E. coli* with five chloroplast chaperones including BSD2. *Science* **358**, 1272–1278 (2017).
4. A. Bracher, S. M. Whitney, F. U. Hartl, M. Hayer-Hartl, Biogenesis and metabolic maintenance of Rubisco. *Annu. Rev. Plant Biol.* **68**, 29–60 (2017).
5. M. Alcalde, *Directed Enzyme Evolution: Advances and Applications* (Springer, 2017).
6. R. H. Wilson, G. Thieulin-Pardo, F. U. Hartl, M. Hayer-Hartl, Improved recombinant expression and purification of functional plant Rubisco. *FEBS Lett.* **593**, 611–621 (2019).
7. M. Flecken *et al.*, Dual functions of a Rubisco activase in metabolic repair and recruitment to carboxysomes. *Cell* **183**, 457–473.e20 (2020).
8. M. Hayer-Hartl, F. U. Hartl, Chaperone machineries of Rubisco–The most abundant enzyme. *Trends Biochem. Sci.* **45**, 748–763 (2020).
9. A. K. Singh, D. Balchin, R. Imamoglu, M. Hayer-Hartl, F. U. Hartl, Efficient catalysis of protein folding by GroEL/ES of the obligate chaperonin substrate MetF. *J. Mol. Biol.* **432**, 2304–2318 (2020).
10. K. F. Müller, T. Borsch, K. W. Hilu, Phylogenetic utility of rapidly evolving DNA at high taxonomical levels: Contrasting matK, trnT-F, and rbcL in basal angiosperms. *Mol. Phylogenet. Evol.* **41**, 99–117 (2006).
11. P. A. Christin *et al.*, Evolutionary switch and genetic convergence on rbcL following the evolution of C4 photosynthesis. *Mol. Biol. Evol.* **25**, 2361–2368 (2008).
12. L. Sen *et al.*, Molecular evolution of rbcL in three gymnosperm families: Identifying adaptive and coevolutionary patterns. *Biol. Direct* **6**, 29 (2011).
13. E. Pichersky, R. Bernatzky, S. D. Tanksley, A. R. Cashmore, Evidence for selection as a mechanism in the concerted evolution of *Lycopersicon esculentum* (tomato) genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3880–3884 (1986).
14. K. Yamada, I. I. Davydov II, G. Besnard, N. Salamin, Duplication history and molecular evolution of the rbcS multigene family in angiosperms. *J. Exp. Bot.* **70**, 6127–6139 (2019).
15. R. J. Miller, "Evolution of the rbcS gene family in Solanaceae: Concerted evolution and gain and loss of introns, with a description of new statistical guidelines for determining the number of unique gene copies," PhD thesis, University of Washington, Seattle, WA (2014).
16. L. Gong *et al.*, The cytonuclear dimension of allopolyploid evolution: An example from cotton using rubisco. *Mol. Biol. Evol.* **29**, 3023–3036 (2012).
17. L. Gong, M. Olson, J. F. Wendel, Cytonuclear evolution of rubisco in four allopolyploid lineages. *Mol. Biol. Evol.* **31**, 2624–2636 (2014).
18. T. Sehrish, V. V. Symonds, D. E. Soltis, P. S. Soltis, J. A. Tate, Cytonuclear coordination is not immediate upon allopolyploid formation in *Tragopogon miscellus* (Asteraceae) allopolyploids. *PLoS One* **10**, e0144339 (2015).
19. X. Wang *et al.*, Cytonuclear variation of Rubisco in synthesized rice hybrids and allotetraploids. *Plant Genome* **10**, plantgenome2017.05.0041 (2017).
20. J. Ferreira de Carvalho *et al.*, Cytonuclear interactions remain stable during allopolyploid evolution despite repeated whole-genome duplications in Brassica. *Plant J.* **98**, 434–447 (2019).
21. C. Li *et al.*, Cytonuclear coevolution following homoploid hybrid speciation in *Aegilops tauschii*. *Mol. Biol. Evol.* **36**, 341–349 (2019).
22. Y. Zhai *et al.*, Nuclear-cytoplasmic coevolution analysis of RuBisCO in synthesized *Cucumis* allopolyploid. *Genes (Basel)* **10**, 869 (2019).
23. C. Li *et al.*, Coevolution in hybrid genomes: Nuclear-encoded Rubisco small subunits and their plastid-targeting translocons accompanying sequential allopolyploidy events in triticum. *Mol. Biol. Evol.* **37**, 3409–3422 (2020).
24. J. F. Wendel, The wondrous cycles of polyploidy in plants. *Am. J. Bot.* **102**, 1753–1756 (2015).
25. J. J. Doyle *et al.*, Evolutionary genetics of genome merger and doubling in plants. *Annu. Rev. Genet.* **42**, 443–461 (2008).
26. D. E. Soltis, C. J. Visger, D. B. Marchant, P. S. Soltis, Polyploidy: Pitfalls and paths to a paradigm. *Am. J. Bot.* **103**, 1146–1166 (2016).
27. P. S. Soltis, D. E. Soltis, *Polyploidy and Genome Evolution* (Springer, **vol. 665**, 2012).
28. L. Y. Xia *et al.*, Molecular basis for the assembly of RuBisCO assisted by the chaperone Raf1. *Nat. Plants* **6**, 708–717 (2020).
29. F. Chen *et al.*, The sequenced angiosperm genomes and genome databases. *Front Plant Sci* **9**, 418 (2018).
30. J. A. Birchler, R. A. Veitia, The gene balance hypothesis: Dosage effects in plants. *Methods Mol. Biol.* **1112**, 25–32 (2014).
31. A. Bracher, A. Starling-Windhof, F. U. Hartl, M. Hayer-Hartl, Crystal structure of a chaperone-bound assembly intermediate of form I Rubisco. *Nat. Struct. Mol. Biol.* **18**, 875–880 (2011).
32. C. Weiss, A. Bonshtien, O. Farchi-Pisanty, A. Vitlin, A. Azem, Cpn20: Siamese twins of the chaperonin world. *Plant Mol. Biol.* **69**, 227–238 (2009).
33. S. Saschenbrecker *et al.*, Structure and function of RbcX, an assembly chaperone for hexadecameric Rubisco. *Cell* **129**, 1189–1200 (2007).
34. T. Hauser, L. Popilka, F. U. Hartl, M. Hayer-Hartl, Role of auxiliary proteins in Rubisco biogenesis and function. *Nat. Plants* **1**, 15065 (2015).
35. T. Hauser *et al.*, Structure and mechanism of the Rubisco-assembly chaperone Raf1. *Nat. Struct. Mol. Biol.* **22**, 720–728 (2015).
36. Y. Jiao *et al.*, Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**, 97–100 (2011).
37. P. S. Soltis, D. B. Marchant, Y. Van de Peer, D. E. Soltis, Polyploidy and genome evolution in plants. *Curr. Opin. Genet. Dev.* **35**, 119–125 (2015).
38. Y. Van de Peer, E. Mizrachi, K. Marchal, The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**, 411–424 (2017).
39. J. F. Wendel, D. Lisch, G. Hu, A. S. Mason, The long and short of doubling down: Polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Curr. Opin. Genet. Dev.* **49**, 1–7 (2018).
40. J. A. Birchler, R. A. Veitia, Gene balance hypothesis: Connecting issues of dosage sensitivity across biological disciplines. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 14746–14753 (2012).
41. R. A. Veitia, D. R. Govindaraju, S. Bottani, J. A. Birchler, Aging: Somatic mutations, epigenetic drift and gene dosage imbalance. *Trends Cell Biol.* **27**, 299–310 (2017).
42. P. P. Edger, J. C. Pires, Gene and genome duplications: The impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Res.* **17**, 699–717 (2009).
43. J. F. Gout, D. Kahn, L. Duret; Paramecium Post-Genomics Consortium, The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. *PLoS Genet.* **6**, e1000944 (2010).
44. X. Shi *et al.*, The gene balance hypothesis: Epigenetics and dosage effects in plants. *Methods Mol. Biol.* **2093**, 161–171 (2020).

45. T. Makino, A. McLysaght, Ohnologs in the human genome are dosage balanced and frequently associated with disease. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9270–9274 (2010).
46. R. De Smet *et al.*, Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2898–2903 (2013).
47. M. D'Antonio, F. D. Ciccarelli, Modification of gene duplicability during the evolution of protein interaction network. *PLOS Comput. Biol.* **7**, e1002029 (2011).
48. J. J. Doyle, J. E. Coate, Polyploidy, the nucleotype, and novelty: The impact of genome doubling on the biology of the cell. *Int. J. Plant Sci.* **180**, 1–52 (2019).
49. M. J. Song, B. I. Potter, J. J. Doyle, J. E. Coate, Gene balance predicts transcriptional responses immediately following ploidy change in *Arabidopsis thaliana. Plant Cell* **32**, 1434–1448 (2020).
50. M. Fernandes Gyorfy *et al.*, Nuclear-cytoplasmic balance: Whole genome duplications induce elevated organellar genome copy number. *Plant J.* **108**, 219–230 (2021).
51. M. F. Gyorfy *et al.*, Nuclear-cytoplasmic balance: Whole genome duplications induce elevated organellar genome copy number. *Plant J.* 108(1):219-230. (2021).
52. B. J. Pogson, N. S. Woo, B. Förster, I. D. Small, Plastid signalling to the nucleus and beyond. *Trends Plant Sci.* **13**, 602–609 (2008).
53. T. Kleine, C. Voigt, D. Leister, Plastid signalling to the nucleus: Messengers still lost in the mists? *Trends Genet.* **25**, 185–192 (2009).
54. C. Oberprieler, M. Talianova, J. Griesenbeck, Effects of polyploidy on the coordination of gene expression between organellar and nuclear genomes in *Leucanthemum* Mill. (Compositae, Anthemideae). *Ecol. Evol.* **9**, 9100–9110 (2019).
55. J. A. Birchler, R. A. Veitia, The gene balance hypothesis: Implications for gene regulation, quantitative traits and evolution. *New Phytol.* **186**, 54–62 (2010).
56. Z. Li *et al.*, Patterns and processes of diploidization in land plants. *Annu. Rev. Plant Biol.* **72**, 387–410 (2021).
57. M. Bekaert, G. C. Conant, Copy number alterations among mammalian enzymes cluster in the metabolic network. *Mol. Biol. Evol.* **28**, 1111–1121 (2011).
58. M. D. Rausher, The evolution of genes in branched metabolic pathways. *Evolution* **67**, 34–48 (2013).
59. J. D. Rochaix, Regulation of photosynthetic electron transport. *Biochim. Biophys. Acta* **1807**, 375–383 (2011).
60. M. Borg *et al.*, Epigenetic reprogramming rewires transcription during the alternation of generations in *Arabidopsis*. *eLife* **10**, e61894 (2021).
61. P. J. Wittkopp, G. Kalay, Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* **13**, 59–69 (2011).
62. Y. Bao *et al.*, Unraveling cis and trans regulatory evolution during cotton domestication. *Nat. Commun.* **10**, 5399 (2019).
63. R. T. Gaeta, J. Chris Pires, Homoeologous recombination in allopolyploids: The polyploid ratchet. *New Phytol.* **186**, 18–28 (2010).
64. A. L. Grusz, E. M. Sigel, C. Witherup, Homoeologous chromosome pairing across the eukaryote phylogeny. *Mol. Phylogenet. Evol.* **117**, 83–94 (2017).
65. S. Griffiths *et al.*, Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. *Nature* **439**, 749–752 (2006).
66. T. Zheng *et al.*, CDKG1 protein kinase is essential for synapsis and male meiosis at high ambient temperature in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2182–2187 (2014).
67. J. D. Hollister *et al.*, Genetic adaptation associated with genome-doubling in autotetraploid *Arabidopsis arenosa*. *PLoS Genet.* **8**, e1003093 (2012).
68. L. Yant *et al.*, Meiotic adaptation to genome duplication in *Arabidopsis arenosa*. *Curr. Biol.* **23**, 2151–2156 (2013).
69. Z. J. Chen *et al.*, Genomic diversifications of five Gossypium allopolyploid species and their impact on cotton improvement. *Nat. Genet.* **52**, 525–533 (2020).
70. S. Dodsworth *et al.*, Repetitive DNA restructuring across multiple *Nicotiana* allopolyploidisation events shows a lack of strong cytoplasmic bias in influencing repeat turnover. *Genes (Basel)* **11**, 216 (2020).
71. M. Jakobsson *et al.*, A unique recent origin of the allotetraploid species *Arabidopsis suecica*: Evidence from nuclear DNA markers. *Mol. Biol. Evol.* **23**, 1217–1231 (2006).
72. J. F. Wendel, C. L. Brubaker, T. Seelanan, "The origin and evolution of Gossypium" in *Physiology of Cotton*. Dordrecht Ed,(Springer, 2010), pp. 1–18.
73. I. J. Leitch *et al.*, The ups and downs of genome size evolution in polyploid species of Nicotiana (Solanaceae). *Ann. Bot.* **101**, 805–814 (2008).
74. T. Marcussen *et al.*; International Wheat Genome Sequencing Consortium, Ancient hybridizations among the ancestral genomes of bread wheat. *Science* **345**, 1250092 (2014).