



Organellar transcripts dominate the cellular mRNA pool across plants of varying ploidy levels

Evan S. Forsythe^{a,1}, Corrinne E. Grover^{b,1}, Emma R. Miller^b, Justin L. Conover^b, Mark A. Arick II^c, M. Carolina F. Chavarro^{d,e}, Soraya C. M. Leal-Bertioli^d, Daniel G. Peterson^c, Joel Sharbrough^f, Jonathan F. Wendel^b, and Daniel B. Sloan^{a,2}

Edited by Douglas Soltis, University of Florida, Gainesville, FL; received March 12, 2022; accepted June 14, 2022

Mitochondrial and plastid functions depend on coordinated expression of proteins encoded by genomic compartments that have radical differences in copy number of organellar and nuclear genomes. In polyploids, doubling of the nuclear genome may add challenges to maintaining balanced expression of proteins involved in cytonuclear interactions. Here, we use ribo-depleted RNA sequencing (RNA-seq) to analyze transcript abundance for nuclear and organellar genomes in leaf tissue from four different polyploid angiosperms and their close diploid relatives. We find that even though plastid genomes contain <1% of the number of genes in the nuclear genome, they generate the majority (69.9 to 82.3%) of messenger RNA (mRNA) transcripts in the cell. Mitochondrial genes are responsible for a much smaller percentage (1.3 to 3.7%) of the leaf mRNA pool but still produce much higher transcript abundances per gene compared to nuclear genome. Nuclear genes encoding proteins that functionally interact with mitochondrial or plastid gene products exhibit mRNA expression levels that are consistently more than 10-fold lower than their organellar counterparts, indicating an extreme cytonuclear imbalance at the RNA level despite the predominance of equimolar interactions at the protein level. Nevertheless, interacting nuclear and organellar genes show strongly correlated transcript abundances across functional categories, suggesting that the observed mRNA stoichiometric imbalance does not preclude coordination of cytonuclear expression. Finally, we show that nuclear genome doubling does not alter the cytonuclear expression ratios observed in diploid relatives in consistent or systematic ways, indicating that successful polyploid plants are able to compensate for cytonuclear perturbations associated with nuclear genome doubling.

cytonuclear | mitochondrial | chloroplast | gene balance | polyploidy

The endosymbiotic acquisition of mitochondria and plastids was instrumental in the early evolution and diversification of eukaryotes (1, 2). These organelles still retain their own genomes, but they are highly reduced in gene content, owing, in part, to a combination of intracellular gene transfer to the nucleus and functional replacement via retargeting of existing eukaryotic genes (3–5). As a result, key functions in mitochondria and plastids, such as cellular respiration and photosynthesis, are now operated under divided genetic control, as they depend on intimate interactions between proteins encoded by different genomes within the cell (6, 7). There are now thousands of nuclear genes encoding proteins that are targeted to the mitochondria and plastids (8), known as *N*-mt and *N*-pt genes, respectively. Eukaryotes have evolved complex signaling mechanisms to coordinate organellar genome function with the expression, import, and turnover of *N*-mt and *N*-pt proteins (9–11). This regulatory task is complicated by the fact that nuclear, mitochondrial, and plastid genomes all have distinct transcription and translation systems (12).

Many of the protein subunits found in enzyme complexes that act in cellular respiration, photosynthesis, and other key organellar functions must assemble in equimolar ratios (Fig. 1). As such, disruption of the stoichiometric balance between *N*-mt/*N*-pt proteins and their cytoplasmic counterparts can have detrimental functional consequences (10, 13). Accordingly, eukaryotes must achieve this cytonuclear protein balance even though nuclear, mitochondrial, and plastid genomes differ wildly in copy number. For example, a typical diploid plant leaf cell contains two copies of the nuclear genome, dozens of copies of mitochondrial genome, and hundreds to thousands of copies of the plastid genome (14–17). The extent to which cytonuclear protein balance is regulated at transcriptional, translational, or posttranslational levels is not entirely clear, but nuclear and organellar mRNA transcripts can exist in imbalanced ratios (18–21). For example, the mitochondrial oxidative phosphorylation (OXPHOS) genes have higher messenger RNA (mRNA) abundances than their *N*-mt counterparts even though the corresponding proteins are typically found in equimolar ratios within these complexes (18). Differences in nuclear and organellar transcript abundances appear to be related

Significance

Plants cells must coordinate gene expression across three different genomic compartments (nuclear, chloroplast, and mitochondrial). Rampant nuclear genome duplication (polyploidy) across plant diversity may potentially disrupt this delicate balance. We performed genome-wide analyses of gene expression in related polyploids and diploids from four lineages of flowering plants. Despite the minimal gene content of chloroplast and mitochondrial genomes, their transcripts comprise most of the mRNA pool in leaf cells. This finding illustrates the extreme cytoplasmic-nuclear asymmetry in RNA expression that plants overcome to maintain balanced protein expression. Cytoplasmic-nuclear RNA expression ratios are similar across diverse functional gene classes and different ploidy levels, indicating that successfully established polyploids have compensated for perturbations in gene expression associated with nuclear genome doubling.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2022 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

See [online](#) for related content such as Commentaries.

¹E.S.F and C.E.G. contributed equally to this work.

²To whom correspondence may be addressed. Email: dan.sloan@colostate.edu.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2204187119/-/DCSupplemental>

Published July 19, 2022.

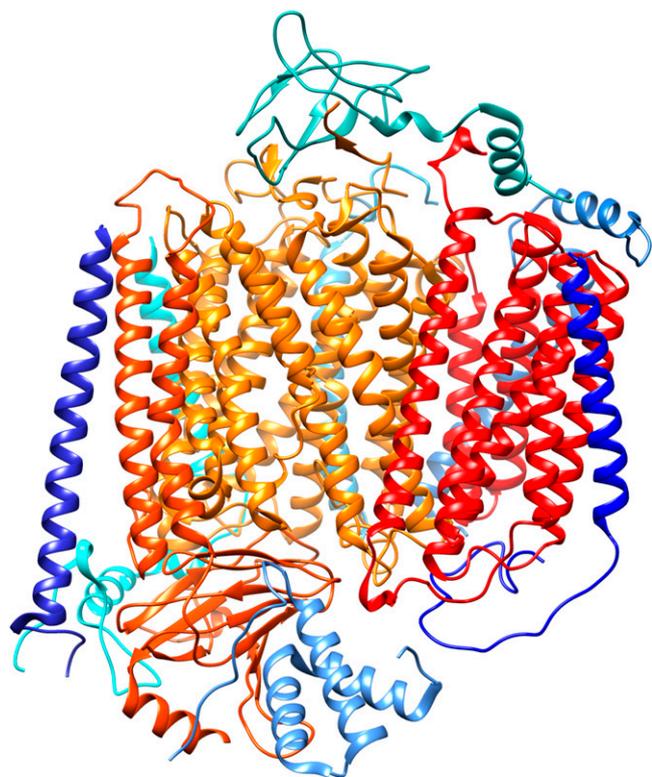


Fig. 1. Structure of cytochrome c oxidase from the angiosperm *Vigna radiata* (71). Mitochondrial-encoded subunits Cox1, Cox2, and Cox3 are shown in shades of orange to red. Nuclear-encoded subunits are shown in shades of blue. All subunits assemble in an equimolar ratio with each present in a single copy within the complex. The image was generated with Chimera v1.15 (72), using Protein Data Bank accession 7JRO.

to intrinsic properties of the genomes themselves, as organellar genes that have been functionally relocated to the nucleus exhibit reduced transcript abundance (19).

Doubling of the nuclear genome via polyploidy may further complicate the challenge of coordinating cytonuclear expression, as it has the potential to perturb the ratio of nuclear to organellar genome copies (22, 23). It appears that nuclear genome doubling elicits at least a partial compensatory increase in organellar genome copy number (15, 24–26), as polyploid cells often exhibit increased size and a higher number of organelles per cell (23). Whether the effects of nuclear genome doubling on cytonuclear stoichiometry are further stabilized by compensatory regulation of gene expression is less clear. A recent RNA sequencing (RNA-seq) analysis of synthetic *Arabidopsis thaliana* autopolyploids indicated that they were able to preserve mRNA expression ratios for interacting nuclear and organellar genes (24); however, patterns differed between the mitochondria and plastids. The abundance of both plastid and *N*-pt mRNAs was reduced in a correlated fashion in polyploids relative to other nuclear genes and the diploid progenitor, whereas mitochondrial and *N*-mt mRNA levels were constant or even upregulated in polyploids. A more detailed breakdown of specific functional complexes and pathways within the organelles was not performed in that study, but another recent analysis used qPCR in diploid and polyploid *Leucanthemum* species to investigate mRNA abundance for plastid and *N*-pt genes encoding subunits of two specific photosynthetic enzyme complexes: photosystem II and RuBisCO (26). That analysis found that the two complexes differed with respect to whether cytonuclear mRNA balance could be restored in response to observed shifts at the genomic level in polyploids.

Given the variation observed in the foregoing studies, there is clearly a need for more global analyses of nuclear and organellar expression across multiple polyploid systems. A challenge in performing such analyses is to accurately quantify both nuclear and organellar transcripts in a total-cellular RNA pool. RNA-seq offers a means to quantify expression on a genomewide basis. However, typical RNA-seq approaches in eukaryotes use selection of polyadenylated (polyA) transcripts to exclude the highly abundant ribosomal RNAs (rRNAs) that would otherwise overwhelm sequencing datasets, rendering these approaches inappropriate for assessing organellar transcription. Whereas polyadenylation is a nearly universal feature of mature nuclear mRNAs, it is not ubiquitous in organellar mRNA processing. For example, in plant mitochondria and plastids, polyadenylation can occur, but it acts predominantly as a degradation signal (27, 28). As such, the polyA⁺ fraction of plant organellar mRNAs may tend to reflect transcript turnover rather than total expression (24). rRNA depletion (ribo-depletion) is an alternative approach that can produce a less-biased estimate than polyA selection for overall mRNA abundance between nuclear and organellar genomes, as well as expression differences among genes within these genomes. This method employs hybrid probes that specifically bind to and remove rRNAs without excluding other transcripts that lack a polyA tail (29).

In this study, we performed RNA-seq with ribo-depletion in four different angiosperm polyploids (*Arabidopsis suecica*, *Arachis hypogea*, *Chenopodium quinoa*, and *Gossypium hirsutum*). Each of these species is the product of allopolyploidization between two diploids, and high-quality genome sequences are available for all of them (30–33). We also analyzed diploid relatives that serve as models of the progenitors of each allopolyploid. The resulting data reveal a highly imbalanced cytonuclear stoichiometry in plant cells that is dominated by plastid and mitochondrial mRNAs. In addition, we find that these ratios are largely preserved between polyploids and related diploids, indicating that polyploid plants often compensate for cytonuclear perturbations associated with nuclear genome doubling.

Results

Plastid Transcripts Represent the Dominant Fraction of the mRNA Pool in Plant Leaf Cells. We used ribo-depleted RNA-seq data and annotated nuclear, mitochondrial, and plastid protein-coding gene models to estimate mRNA abundance for each gene in terms of transcripts per million (TPM) (34, 35). Even though nuclear genomes contain tens of thousands of protein-coding genes compared to only dozens in organellar genomes, we found that organellar transcripts dominated the mRNA pool in plant leaf cells (Fig. 2*A* and *SI Appendix*, Figs. S1 and S2). In particular, the plastid genome accounted for the majority of total-cellular mRNA transcripts in leaf tissue from all 12 sampled species across four different angiosperm genera, with a mean of 76.2% and a range of 69.9 to 82.3% (Fig. 2*A* and *SI Appendix*, Table S1). Mitochondrial mRNAs represented a substantially smaller percentage (mean of 2.2% and range of 1.3 to 3.7%; *SI Appendix*, Table S1). Nevertheless, given the small number of genes in the mitochondrial genome, the average transcript abundance per gene is still much larger than for nuclear genes (Fig. 2*B*).

The abundance of mRNA transcripts varied dramatically across genes within the plastid and mitochondrial genomes, likely reflecting differences in rates of RNA polymerase binding and transcription (36). Photosynthetic genes were particularly dominant. In all species, the *psbA* gene, which encodes the D1 subunit of the photosystem II reaction center (37), exhibited the

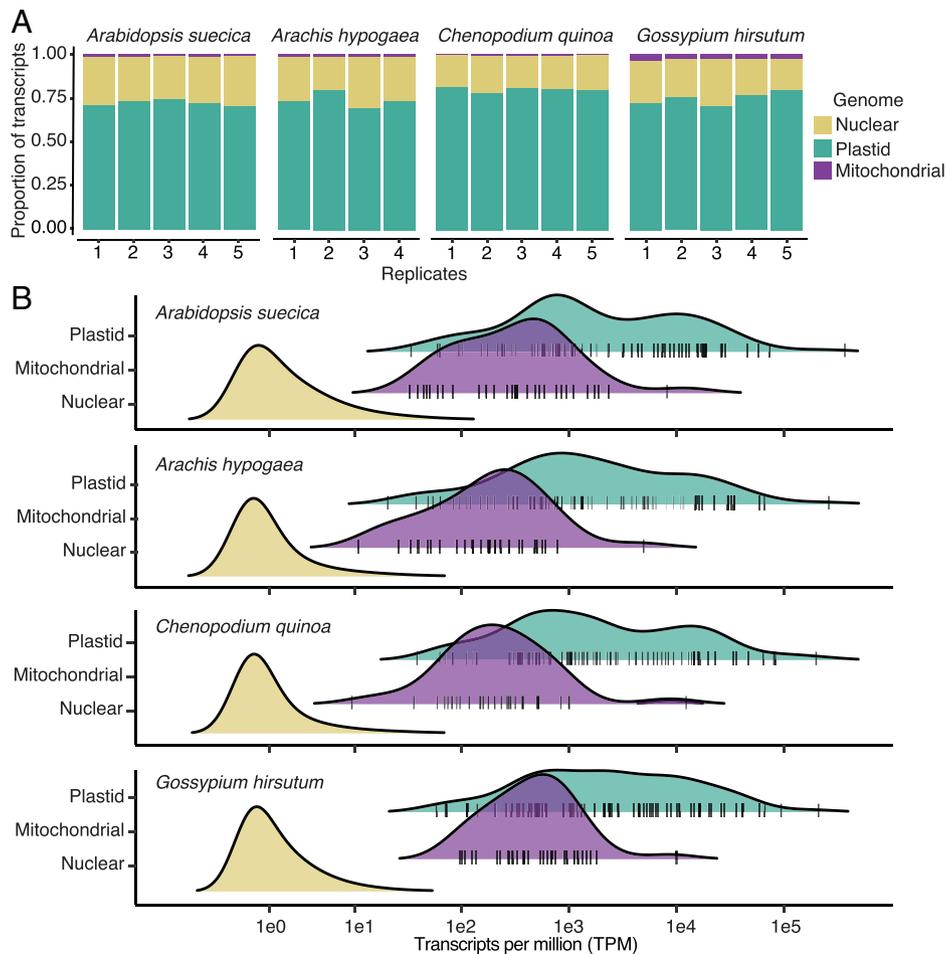


Fig. 2. Predominance of organellar transcripts in the mRNA pool of plant leaf cells. (A) Proportion of mRNA transcript by genomic compartment for each biological replicate in four allotetraploid species. (B) Density plots displaying frequency distributions for transcript abundances by gene for each genomic compartment in each species. These density plots were generated with the *ggridges* R package. Vertical tick marks underlying the plastid and mitochondrial distributions indicate individual gene values. Because of the smoothing effect of density plots, internal and flanking regions with no tick marks can still show nonzero density values. See *SI Appendix, Figs. S1 and S2* for corresponding data in diploids.

highest mRNA abundance, accounting for more than 25% of all mRNA transcripts in some samples (*SI Appendix, Table S2*). The other most highly expressed plastid genes all encode subunits of key photosynthetic enzyme complexes (cytochrome b6f complex, photosystem I, photosystem II, and RuBisCO; *SI Appendix, Table S2*). At the other extreme, the plastid genes with the lowest mRNA abundances generally encoded proteins involved in transcription (RpoB, RpoC1, and RpoC2) (38) or implicated in protein import (Ycf1 and Ycf2; *SI Appendix, Table S2*) (39, 40), but these still produced higher mRNA levels than most nuclear genes (Fig. 2B). In mitochondrial genomes, the genes encoding subunits of the major OXPHOS enzyme complexes exhibited the highest mRNA abundances; *atp9* had the highest expression level in all species (*SI Appendix, Table S3*). This gene encodes a subunit that is present in numerous copies that make up the membrane-bound c-ring within the F_0 portion of the mitochondrial ATP synthase complex (41). The lowest expression was observed for genes encoding proteins involved in cytochrome C biogenesis (CcmB, CcmC, CcmFc, CcmFn) (42), intron splicing (MatR) (43), and protein import (MttB; *SI Appendix, Table S3*) (44).

Consistent with the predominance of photosynthesis-related transcripts in the plastid mRNA pool, the most abundant nuclear mRNAs generally encoded subunits in key photosynthetic enzyme complexes, particularly photosystem II and RuBisCO. However, as detailed below, the mRNA abundance

for these nuclear-encoded subunits was far lower than for the corresponding plastid-encoded subunits that contribute to the same photosynthetic enzyme complexes.

Extreme Stoichiometric Imbalance between mRNAs from Organellar and Interacting Nuclear Genes.

For many of the major cytonuclear enzyme complexes in mitochondria and plastids, the constituent subunits must assemble in equimolar ratios. Examples include OXPHOS complexes I–IV; the ribosomes; the plastid acetyl CoA-carboxylase; and key photosynthetic complexes such as RuBisCO, photosystem I, and photosystem II (8). Given this relationship at the level of protein subunits, the simple expectation might be that there would be similar stoichiometric balance at the mRNA level. However, the massive transcript abundances observed for mitochondrial and plastid genes (Fig. 2) imply otherwise. To quantify nuclear expression for *N*-mt and *N*-pt genes involved in direct molecular interactions with organellar genomes or gene products, we analyzed major functional categories in the Cytonuclear Molecular Interactions Reference for *Arabidopsis* (CyMIRA) classification scheme (8). In order to have clear expectations for expression levels in diploids versus polyploids, we restricted this analysis to nuclear genes found in “quartets,” meaning that they were present in two homeologous copies in the focal allopolyploid and one copy in each of the two diploid progenitor models. To account for the fact that the gene copy number had doubled in the polyploids, we summed the

transcript abundance values for the two homeologs in polyploids and treated them as a single gene.

As predicted by the much higher TPM levels for organellar than for nuclear genes (Fig. 2), overall mRNA abundance was heavily tilted toward mitochondrial and plastid genes, even when compared to their *N*-mt and *N*-pt counterparts involved in direct molecular and functional interactions. The mRNA abundance per gene was more than 10-fold higher for mitochondrial and plastid genes than for their interacting nuclear counterparts in all 12 species and all eight of the tested CyMIRA functional categories; in some cases, these cytonuclear ratios even exceeded 100-fold, including the mitochondrial ribosome and OXPHOS system in all *Gossypium* species and the plastid Clp complex in six different species (Fig. 3A and *SI Appendix*, Figs. S3A and S4A).

Nuclear and Organellar Expression Levels Are Correlated across Functional Categories Despite Stoichiometric Imbalance. One possible explanation for the massive cytonuclear mRNA imbalance is that mitochondrial and plastid genes are simply transcribed in excess and decoupled from the functional requirements for the corresponding gene products within the cell. To test this possibility, we compared average TPM values for organellar genes versus interacting *N*-mt/*N*-pt genes across CyMIRA categories. In the absence of regulation at the transcriptional level in mitochondrial and plastid genomes, we would expect these values to be uncorrelated. To the contrary, this analysis revealed a strong positive relationship between organellar and nuclear mRNA abundance in each of the 12 species ($P \leq 0.001$ for each of the 12 species; Fig. 3B and *SI Appendix*, Figs. S3B and S4B). This finding represents clear evidence for cytonuclear coordination at the transcriptional level. The strong correlation across functional categories implies that expression levels in nuclear and organellar genomes both reflect cellular demands for these pathways despite the large observed stoichiometric imbalance that is biased toward mitochondrial and plastid mRNAs.

Cytonuclear Expression Ratios Remain Stable across Differing Nuclear Ploidy Levels. The above analysis shows that the baseline expectation in plants should not be for balanced cytonuclear stoichiometry in mRNA abundance. Nonetheless, the strong correlation between interacting nuclear and organellar genes means that the observed cytonuclear ratio stays in a relatively narrow range across functional categories that differ greatly in absolute expression levels. This observation raises the question as to whether these ratios are robust to changes in nuclear gene copy number that accompany polyploidy events. To address this question, we first looked at whether the percentage of mitochondrial and plastid transcripts in the total-cellular RNA pool differed systematically between diploids and polyploids. All else being equal, doubling the nuclear genome copy number would lead to a proportional increase in nuclear mRNAs and concomitant decrease in the relative abundance of mitochondrial and plastid mRNAs. However, we found little evidence for a consistent reduction in the proportion of organellar transcripts in polyploids.

We found that the four polyploids each had total plastid mRNA abundance levels that were either intermediate between their two diploid progenitors (*Arabidopsis*, *Arachis*, and *Gossypium*) or higher than both of their diploid progenitors (*Chenopodium*), on average (*SI Appendix*, Table S1). Therefore, none of the four systems exhibited the decline in plastid mRNA abundance relative to nuclear transcripts that might be expected to accompany nuclear genome doubling in the absence of any cytonuclear compensation, and there was only significant variation among species in plastid mRNA abundance in *Gossypium* (one-way ANOVA; $P = 0.0031$ in *Gossypium*; $P \geq 0.05$ in the three other genera). For mitochondrial mRNA abundance relative to nuclear transcripts, there were significant differences among species within both *Arabidopsis* ($P = 0.0042$) and *Chenopodium* ($P < 0.0001$), but in each case, the polyploid exhibited intermediate mitochondrial mRNA abundance compared to the two diploid progenitors, on average (*SI Appendix*, Table S1). For *Arachis* and *Gossypium*, the polyploid did exhibit lower relative abundance of mitochondrial

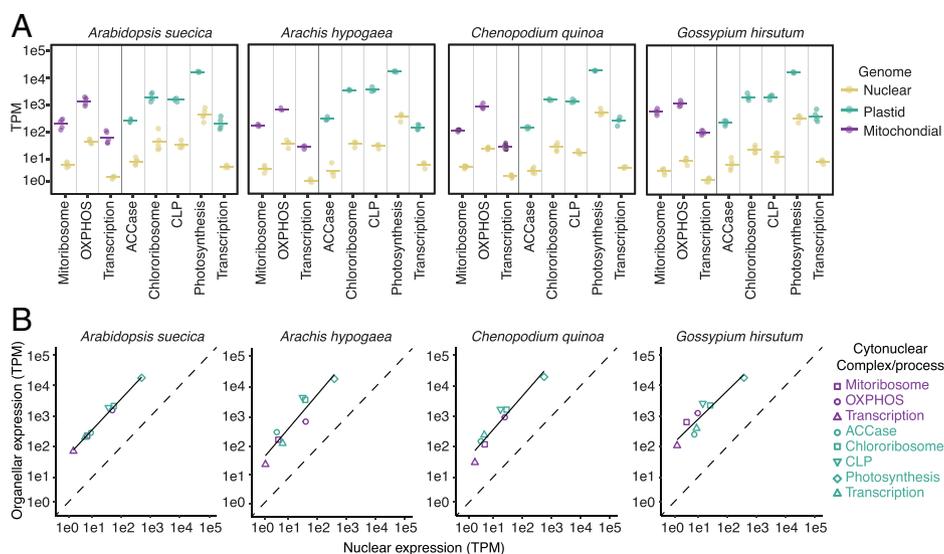


Fig. 3. Consistent bias in cytonuclear stoichiometry for mRNA abundance. (A) Transcript abundance for organellar genes and their interacting *N*-mt/*N*-pt counterparts broken down by functional categories according to the CyMIRA classification scheme (8) in four allotetraploid species. Reported values are averaged across genes within a category. Points represent biological replicates, and horizontal lines indicate the mean across replicates. In all cases, organellar mRNA abundance (measured as TPM) exceeds the corresponding nuclear mRNA abundance by at least 10-fold. (B) Correlation between organellar and nuclear mRNA abundances across CyMIRA categories, with the solid line representing a best fit linear model and the dashed line representing a 1:1 line. Despite the large imbalance toward organellar mRNAs, the cytonuclear ratio remains relatively consistent across functional categories expressed at very different levels, resulting in a strong positive correlation ($P < 0.001$ in all cases). For both sets of plots, the analysis of nuclear genes was limited to those found in homologous quartets (see *Materials and Methods*). Expression values for the two homeologs in the polyploid reference genome were summed for this analysis. See *SI Appendix*, Figs. S3 and S4 for corresponding data in diploids.

mRNAs than both diploids, but these differences were not significant ($P > 0.05$ in both cases; *SI Appendix, Table S1*).

Because the preceding analysis is based on total mRNA abundance in each organelle, it may be heavily influenced by individual genes with high expression levels, such as *psbA* and *atp9* in the plastids and mitochondria, respectively. Therefore, we also compared mRNA abundance across species at the level of individual mitochondrial and plastid genes (represented as \log_{10} TPM). In all cases, correlations in gene-specific expression levels between species were very strong (Fig. 4 and *SI Appendix, Fig. S5* and *SI Appendix, Table S4*). Patterns of variation between polyploids and diploids differed across genera, as well as between the mitochondria and plastids. In *Arabidopsis*, post hoc pairwise comparisons indicated that the polyploid *A. suecica* exhibited significantly higher gene-specific expression levels of plastid mRNAs compared to each of the two diploids ($P < 0.0001$), whereas it was intermediate between the two diploids for mitochondrial expression, with significantly higher values than the maternal diploid *A. thaliana* ($P < 0.0001$) but significantly lower than the paternal diploid *A. arenosa* ($P < 0.0001$) (*SI Appendix, Table S4*). In the other three genera, the polyploid had lower organellar mRNA abundance values than both diploids, although the difference between the polyploid *C. quinua* and the paternal diploid *C. suecicum* was not significant for mitochondrial transcripts ($P = 0.06$) (*SI Appendix, Table S4*). Overall, the departures from a one-to-one relationship between diploids and polyploids were very small and similar in magnitude to the differences between the two diploids (Fig. 4).

In addition to comparing mitochondrial and plastid mRNA abundance to total nuclear expression in polyploids versus diploids, we also investigated how it specifically related to expression of *N*-mt and *N*-pt genes. Because of the functional relationships between organellar genes and *N*-mt/*N*-pt genes, we would expect any perturbation in these cytonuclear expression ratios in polyploids to have the most significant fitness consequences. We compared the ratio of expression levels for organellar genes to their interacting *N*-mt/*N*-pt genes between species. In most cases, these ratios exhibited a significant positive correlation across functional categories (Fig. 5). Similar to what we observed in analyzing total nuclear expression (Fig. 4), we did not find that polyploids from different genera showed consistent deviations in these cytonuclear expression levels (Fig. 5). For example, in *Arabidopsis*, the polyploid exhibited higher ratios of organellar to *N*-mt/*N*-pt transcript abundance compared to its maternal diploid progenitor (Fig. 5*A*), but the opposite was generally true in *Gossypium* (Fig. 5*J*). In addition, levels of *N*-mt/*N*-pt transcript abundance compared to other nuclear genes were similar between diploids and polyploids (*SI Appendix, Table S5*). Overall, we do not find evidence that doubling the nuclear genome leads to consistent and long-term skewing of expression ratios toward nuclear genes.

Discussion

The Predominance of Organellar Transcripts in the Plant Leaf mRNA Pool. One of the most striking patterns to emerge from our quantification of mRNAs across three different genomic compartments in plant cells is the sheer abundance of organellar transcripts (Fig. 2). Most plant RNA-seq studies use polyA selection, leading to a large underrepresentation of organellar transcripts (24, 28) and making direct comparisons between organellar and nuclear expression levels infeasible. In contrast to plants, animal mitochondrial mRNAs can undergo widespread polyadenylation (45). Accordingly, standard RNA-seq datasets have been used to quantify the abundance of

mitochondrial versus nuclear mRNAs in humans, and our findings parallel observations that the 13 protein-coding genes in the human mitochondrial genome can account for ~30 to 50% of all mRNA transcripts in tissues with high metabolic activity such as heart and brain (20, 21).

Given that chloroplasts can contain ~80% of the protein molecules in leaf mesophyll cells (46), it is not surprising that mRNAs with photosynthetic and other plastid functions predominate in plant leaf tissue. However, of the thousands of angiosperm genes that encode proteins in the chloroplast proteome, fewer than 80 (<5%) of them are in the plastid genome (47, 48). Therefore, the fact that plastid transcripts constitute the majority of the cellular mRNA pool can be explained only by the much lower expression levels for *N*-pt genes compared to their plastid counterparts. Previous studies have indicated that organellar transcripts can outnumber nuclear counterparts that encode proteins involved in direct functional interactions (18, 19). Our genomewide analysis shows that this pattern extends across diverse functional categories with a wide range of expression levels (Fig. 3).

Eukaryotes typically have much higher copy numbers for organellar genomes than for the nuclear genome, and our analysis indicates that there is also highly imbalanced cytonuclear stoichiometry at the mRNA level. These imbalances contrast with the equimolar ratios observed among protein subunits in many key cytonuclear enzyme complexes (e.g., Fig. 1), but there are multiple reasons that translational demands may be higher for mitochondrial and plastid-encoded proteins. First, even in cases where nuclear and organellar subunits coassemble in a 1:1 ratio, they may be translated at different rates. For example, the fact that a remarkably high percentage of the entire mRNA pool in leaf tissue is dedicated to a single gene (*psbA*; *SI Appendix, Table S2*) presumably reflects its rapid protein turnover and correspondingly high rates of translation. Even though the RuBisCO subunits are famous for being some of the most abundant proteins on Earth (49), the photosystem II D1 subunit encoded by *psbA* can be translated at higher rates in the light because of the frequent replacement of subunits damaged in photoinhibition (50).

Second, some *N*-mt and *N*-pt genes have been duplicated, affecting the amount of translation that would be expected per gene. For example, the plastid-encoded protein RbcL and the *N*-pt protein RbcS assemble in a 1:1 stoichiometric ratio to form the RuBisCO enzyme complex. However, there is only one copy of the gene encoding RbcL in the plastid genome, whereas RbcS can be produced by different members of a multicopy gene family in the nucleus (51, 52). Therefore, the amount of protein production per gene may naturally be lower for nuclear-encoded RbcS, and such duplication events in the nucleus may contribute to imbalance measured on a per-gene basis (Fig. 3). However, even after summing TPM values for all nuclear *RbcS* paralogs, all 12 species in this study exhibited higher expression levels for the plastid-encoded *rbcL* gene, with a mean plastid-nuclear ratio of 6.8 (*SI Appendix, Table S6*). In a broader analysis of cytonuclear enzyme complexes in *A. thaliana*, most nuclear-encoded subunits are represented by only a single gene (*SI Appendix, Table S7*). The observed average gene family size of 1.2 is a small effect compared to the imbalance in cytonuclear expression, which is consistently greater than 10-fold (Fig. 3).

Third, the subunit composition of cytonuclear enzyme complexes deviates from equimolar stoichiometry in some cases. For example, the mitochondrial ATP synthase complex contains three copies of the Atp1 subunit and eight or more copies of the Atp9 subunit (53), both of which are mitochondrial encoded in plants. Likewise, the sole plastid-encoded component (ClpP1) of

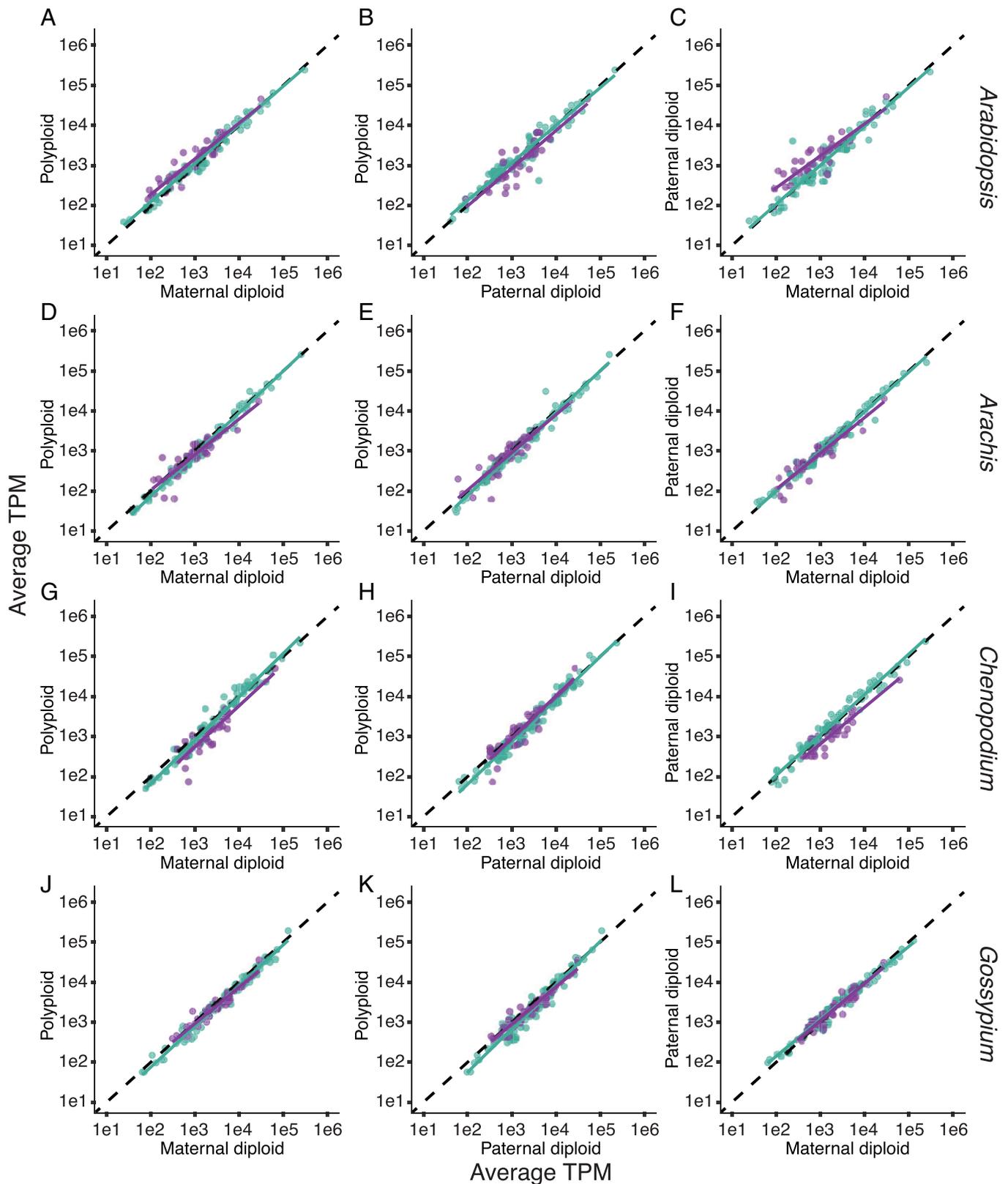


Fig. 4. Comparisons between congeneric species, including diploids and polyploids, in relative abundance of organellar mRNAs, as measured by TPM. The mitochondrial values are rescaled after removal of the plastid genes in each species, so they reflect only the relative abundance of mitochondrial and nuclear genes. Each point represents an average across biological replicates. The dashed line is the 1:1 line. Solid lines represent best-fit linear models for each organellar gene set separately. Each row corresponds to a different genus. (*Left*) Plots compare the polyploid against the maternal diploid. (*Middle*) Plots compare the polyploid against the paternal diploid. (*Right*) Plots compare the paternal diploid against the maternal diploid. R^2 values are available in *SI Appendix, Table S4*.

the caseinolytic protease complex is present in three copies, whereas the corresponding *N*-pt subunits can be present in anywhere from one to three copies (54). Such cases are expected to

increase demands for organellar expression. It also should be noted that although all the functional categories that we analyzed are characterized by cytonuclear interactions at a molecular level,

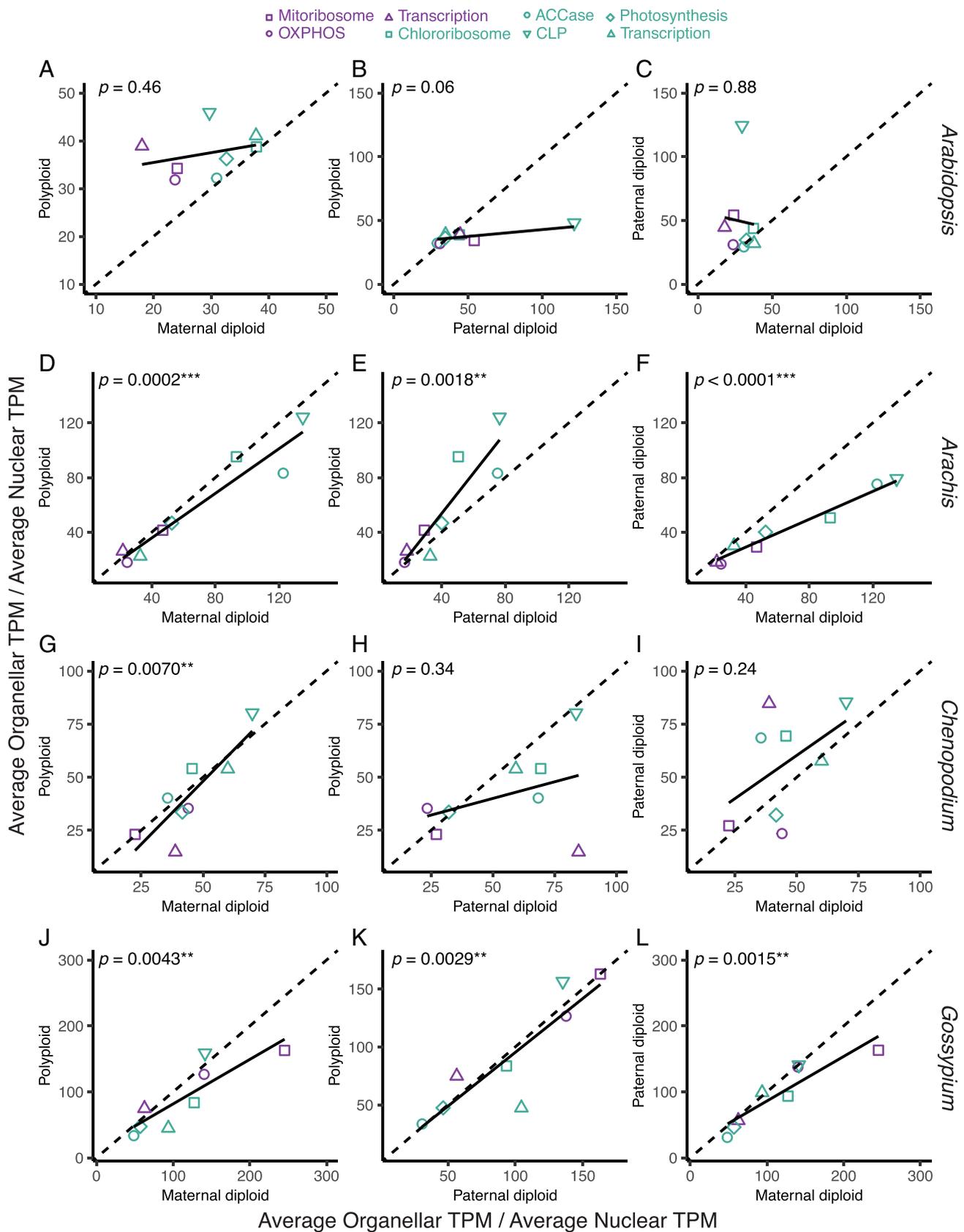


Fig. 5. Comparison of cytonuclear stoichiometry in diploids versus polyploids across functional classes. Scatter plots compare the ratio of organellar to nuclear mRNA abundance for individual CyMIRA categories between two congeneric species. Each point represents averages across biological replicates. The analysis of nuclear genes was limited to those found in homologous quartets (see *Materials and Methods*). Expression values for the two homeologs in the polyploid reference genome were summed for this analysis. The dashed line is the 1:1 line, and the solid line represents a best-fit linear model. Each row corresponds to a different genus. (*Left*) Plots compare the polyploid against the maternal diploid. (*Middle*) Plots compare the polyploid against the paternal diploid. (*Right*) Plots compare the paternal diploid against the maternal diploid.

some of them do not involve multisubunit protein complexes with clearly defined stoichiometric relationships. For example, the mitochondrial-encoded MatR protein from the CyMIRA mitochondrial transcription and transcript maturation functional category does not form a complex with the *N*-mt proteins in that category, which include various proteins involved in mitochondrial transcription and posttranscriptional modifications (8).

Although each of the preceding factors is relevant to cytonuclear expression, it is unlikely that they can collectively explain the enormous imbalance between organellar and nuclear mRNAs, which consistently exhibited a difference of one to two orders of magnitude across diverse functional categories (Fig. 3*A*). Therefore, we are still left with the question as to why plants (and other eukaryotes) maintain such imbalanced cytonuclear expression at the mRNA level. One possibility is that organellar genomes are simply less reliant on transcription as a means of gene regulation (19). While it is apparent that standing levels of mRNAs are broadly tuned to the demand for specific organellar functions (Fig. 3*B*), it is less clear whether plant organellar genomes rely heavily on transcriptional regulation in response to short-term environmental changes. For example, upon shifting from dark to light conditions, there is a dramatic change in ribosome occupancy for *psbA* transcripts observed within minutes but minimal change in *psbA* transcript abundance (50). More generally, a recent analysis of high-light stress on chloroplast proteostasis found that when rates of *N*-pt protein turnover increase, there is generally upregulation in mRNA abundance for the corresponding *N*-pt genes. In contrast, increased turnover of plastid-encoded proteins was rarely associated with an increase in the corresponding plastid mRNAs (55). These observations are consistent with the hypothesis that mitochondria and plastids are relatively lacking in precise, short-term transcriptional control and that the large standing population of organellar mRNAs offers a means for rapid regulatory responses at a posttranscriptional level.

One obvious difference between nuclear and organellar genomes is their copy number within the cell. Our estimates of transcript abundance “per gene” would be radically different if they were calculated per gene copy, given that plastid genomes are present in hundreds to thousands of copies per cell. Genome copy number, however, does not appear to be the sole determinant of the observed cytonuclear imbalance in mRNA abundance because mitochondrial-nuclear and plastid-nuclear ratios are similar to each other (Fig. 3) even though plastid genomes are present at much higher copy numbers than mitochondrial genomes in leaf tissue (14–17). Furthermore, all genes within the plastid genome generally have the same copy number (except for those present in twice as many copies due to the inverted repeat), and yet some plastid genes are expressed more than 1,000-fold higher than others, highlighting the effects of gene regulation (*SI Appendix, Table S2*). Nonetheless, the ratios between organellar and nuclear mRNA abundances would be smaller and even imbalanced toward nuclear expression in some cases when tabulated per gene copy.

An intriguing feature of plant mitochondrial genetics is that some mitochondria contain only partial mitochondrial genome copies or no mitochondrial DNA at all (16). Intermittent mitochondrial fusion events may provide a means to restore complete genome content (56), but it is still unclear how these organelles retain function when they are lacking a full set of genes. Our observations raise the possibility that the abundance of mitochondrial mRNAs (Fig. 2*B*) may serve as a buffer to sustain protein expression through periods where the underlying genes are not available to be transcribed, especially given the potentially long half-lives of plant mitochondrial mRNAs (57).

The Relative Stability of Cytonuclear mRNA Stoichiometry across Changes in Nuclear Ploidy. Previous work in these allopolyploids has found that homeologs in both nuclear subgenomes are expressed at similar levels for most cytonuclear interaction functional categories (58). As such, doubling of the nuclear genome has the potential to alter the balance of cytonuclear gene expression, at least in principle (22). However, our analyses generally did not find consistent patterns of divergence in cytonuclear stoichiometry at the mRNA level between polyploids and related diploids, indicating that successful polyploids maintain cytonuclear stoichiometry at the transcriptional level following whole-genome doubling. This result may not be surprising as the twofold increase in nuclear genome copy number is relatively small in scale compared to the existing imbalances in cytonuclear stoichiometry (Fig. 3). It is possible that potential perturbations in cytonuclear stoichiometry are alleviated by cellular and genomic responses even without changes to regulatory elements that control gene expression. In particular, elevated nuclear ploidy is associated with increases in cell size, number of organelles per cell, and the number of organellar genome copies per cell, resulting in at least partial compensation that stabilizes cytonuclear ratios at the DNA level (23). Notably, at least some of these responses can be observed immediately in laboratory-generated polyploids (15, 24).

Subsequent evolutionary responses in polyploid lineages may further finetune cytonuclear stoichiometry through regulatory changes in gene expression. For example, genomewide changes in regulatory elements have been identified as a response to selection on stoichiometric balance following polyploidization (59). The long-term pattern of nuclear gene loss in polyploids also provides evidence for selection to maintain balanced cytonuclear expression, as *N*-mt and *N*-pt genes have been found to differ significantly in retention rates relative to the rest of the nuclear gene set (60–63). The use of natural allopolyploids in this study means that our polyploid-diploid comparisons reflect a combination of immediate developmental changes and longer-term evolution in response to nuclear genome doubling. Regardless of ploidy, divergence between any pair of species may lead to changes in the stoichiometry of cytonuclear gene expression, which is illustrated by the fact that observed differences between pairs of diploids were often comparable in magnitude to differences between polyploids and diploids in our dataset (Figs. 4 and 5 and *SI Appendix, Tables S4 and S8*). This variation highlights the importance of replicated sampling across multiple independent diploid-polyploid systems.

Although the larger number of organellar genome copies per cell in polyploids (15, 24–26) suggests that an increase in total organellar mRNA abundance contributes to the stability of cytonuclear expression ratios, it remains an open question as to whether there is also some suppression of nuclear expression after genome doubling in these systems. Therefore, an important area for future research is to characterize absolute levels of organellar and nuclear gene expression across different ploidies, including at the single-cell level. Standard RNA-seq measures only relative gene expression, but it is possible to use RNA-seq to estimate changes in absolute expression levels per cell by quantifying the number of cells used in RNA extractions and applying spike-in controls during the sequencing process (64, 65). The use of bulk tissue samples also has the disadvantage of averaging over large numbers of cells. As such, single-cell RNA-seq techniques may provide additional resolution in assessing the relationships between organellar and nuclear transcriptomes. Given the increased cell size in polyploids, total transcriptome and proteome sizes are expected to scale allometrically with

ploidy (66, 67), but how these scaling relationships affect many key features of cytonuclear interactions has yet to be explored in polyploid plants.

Materials and Methods

Generation of RNA-Seq Data and Mapping to Reference Genomes. The sampling of diploids and polyploids from *Arabidopsis*, *Arachis*, *Chenopodium*, and *Gossypium* is described elsewhere, including a summary of the history of allopolyploidization in each lineage, plant growth conditions, RNA extraction, sequencing, and read mapping (58). Briefly, total-cellular leaf RNA was extracted from each species, ribo-depleted, and then sequenced as PE150 on a NovaSeq. 6000 S4 flow cell (average of 51 million read pairs per library). Reads from all three species (one allopolyploid and diploid models of both the maternal and paternal progenitors) within each genus were mapped to the annotated protein-coding sequences from the polyploid nuclear genome and from the organellar genomes of a representative member of the genus using Kallisto (68), as described elsewhere (58). We analyzed five biological replicates for each species with the exceptions of *Arachis hypogaea* and *C. suecicum*. For those species, one of the five replicates was excluded either because a preliminary clustering analysis failed to group the sample with other replicates from that species, raising doubts about sample identity (*C. suecicum*), or because of poor library quality and a low read mapping rate (*Arachis hypogaea*). For *Arabidopsis arenosa*, a total of seven biological replicates were available; however, we included only the first five to balance the sampling of the other species.

Estimation of Gene Expression Levels. Gene-specific TPM values for each sample were generated with Kallisto by using *kallisto quant*. Annotation quality has the potential to affect cytonuclear expression ratios in at least two significant ways. First, nuclear genomes contain numerous insertions of mitochondrial and plastid sequences (known as numts and nupts, respectively) (69). In the case of recent insertions, true mitochondrial and plastid reads can map equally well to numt/nupt pseudogenes or to the actual organellar genomes, resulting in an underestimate of organellar expression. These cases were accounted for prior to read mapping by excluding numts and nupts that were annotated as protein-coding genes from each nuclear reference (58). Second, because we employed a ribo-depletion strategy rather than polyA selection, sequencing produced large quantities of reads from noncoding RNAs (ncRNAs) that lack polyA tails and would be excluded from typical mRNA-seq libraries (70), such as signal recognition particle RNAs and U1 and U2 spliceosomal RNAs. In some cases, the nuclear references contained annotations of short protein-coding genes corresponding to these ncRNAs, whose high expression and short length led to extremely high TPM values. Because these high TPM ncRNAs could distort our estimates of nuclear mRNA expression, we manually screened all loci with values of $\geq 1,000$ TPM to identify protein-coding annotations that corresponded to highly expressed ncRNAs. These loci were retained in the reference for initial mapping to avoid mismapping of reads to other loci, but we subsequently excluded them and rescaled TPM estimates for the rest of the reference gene set so that the values summed to one million. Similarly, for certain analyses, we wanted to compare the relative abundance of nuclear mRNAs to plastid mRNAs by themselves or to mitochondrial mRNAs by themselves without the confounding effects of expression variation in the other organellar genome, so we excluded the relevant gene set and rescaled TPM values accordingly.

Identification of Homeologous “Quartets” and Assignment of CyMIRA Functional Categories. For direct comparisons of nuclear gene expression between polyploids and related diploids, we used the set of previously

identified gene “quartets” that were present as two homeologous copies in an allopolyploid and a single copy in each of the corresponding diploid relatives (58). Because read mapping was always performed against the polyploid reference genome, we summed TPM values from the two homeologs in each quartet for each species for subsequent comparisons of expression. By summing these values, we were able to account for any mismapping of diploid reads to the other subgenome due to sequence similarity between homeologs and to standardize for the difference in gene copy number between diploids and polyploids. For cross-species comparisons (Fig. 5 and *SI Appendix, Fig. S5*), we excluded nuclear genes not assigned to quartets (58) and rescaled TPM estimates for the remaining gene sets accordingly. For consistency across analyses investigating cytonuclear expression ratios across CyMIRA categories, we also used this rescaled dataset for the analyses in Fig. 3 and *SI Appendix, Figs. S3 and S4*.

To assign nuclear genes to functional categories based on predicted targeting to mitochondria and plastids, we used the CyMIRA classification scheme (8) and a previously defined pipeline (63) (https://github.com/jsharbrough/CyMIRA_gene_classification). This classification scheme further assigns genes involved in direct molecular interactions with organellar transcripts or gene products to specific functional categories or enzymatic complexes (Figs. 3 and 5). Although this pipeline generates CyMIRA classifications for all genes, we limited subsequent analyses to genes in homeologous quartets. Organellar genes were manually curated and assigned to the appropriate CyMIRA functional categories (*SI Appendix, Tables S2 and S3*).

Statistical Analyses. Statistical analyses were carried out in R v3.5.0 or v4.0.5. Simple linear models (Figs. 3 and 5) and one-way ANOVAs were implemented with the *lm* and *aov* functions, respectively. For analysis of gene-specific expression levels across species, we used a mixed-model approach with species and gene as fixed effects and individual plant (replicate) as a random effect. This model was fit with the *lmer* function (*lme4* package) and evaluated with the *Anova* function (*car* package) using type III sums of squares. Post hoc comparisons between species pairs were conducted with the *emmeans* function, using a Holm procedure for the Bonferroni correction to account for multiple pairwise comparisons. These gene-specific models were applied after exclusion of genes from the other organellar genome and rescaling of TPM values as described above. Data visualizations were generated with the *ggplot2* and *ggridges* packages. R code used to perform statistical analyses is available via https://github.com/EvanForsythe/Cytonuclear_RNAseq.

Data Availability. Previously published data were used for this work (58).

ACKNOWLEDGMENTS. This work was supported by grants from the National Science Foundation (IOS-1829176 and IOS-2145811) and the New Mexico Institute of Mining and Technology.

Author affiliations: ^aDepartment of Biology, Colorado State University, Fort Collins, CO 80523; ^bDepartment of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA 50010; ^cInstitute for Genomics, Biocomputing & Biotechnology, Mississippi State University, Mississippi State, MS 39762; ^dInstitute of Plant Breeding, Genetics and Genomics, Athens, GA 30602; ^eBayer Crop Science, Chesterfield, MO 63017; and ^fDepartment of Biology, New Mexico Institute of Mining and Technology, Socorro, NM 87801

Author contributions: E.S.F., C.E.G., J.L.C., J.S., J.F.W., and D.B.S. designed research; E.S.F., C.E.G., E.R.M., J.L.C., M.C.F.C., S.C.M.L.-B., J.S., and D.B.S. performed research; M.A.A., M.C.F.C., S.C.M.L.-B., and D.G.P. contributed new reagents/analytic tools; E.S.F., C.E.G., J.L.C., M.A.A., D.G.P., J.S., and D.B.S. analyzed data; and D.B.S. wrote the paper.

1. A. J. Roger, S. A. Muñoz-Gómez, R. Kamikawa, The origin and diversification of mitochondria. *Curr. Biol.* **27**, R1177–R1192 (2017).
2. S. J. Sibbald, J. M. Archibald, Genomic insights into plastid evolution. *Genome Biol. Evol.* **12**, 978–990 (2020).
3. J. N. Timmis, M. A. Ayliffe, C. Y. Huang, W. Martin, Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **5**, 123–135 (2004).
4. D. B. Sloan *et al.*, Cytonuclear integration and co-evolution. *Nat. Rev. Genet.* **19**, 635–648 (2018).
5. I. G. Johnston, B. P. Williams, Evolutionary inference across eukaryotes identifies specific pressures favoring mitochondrial gene retention. *Cell Syst.* **2**, 101–111 (2016).
6. D. M. Rand, R. A. Haney, A. J. Fry, Cytonuclear coevolution: The genomics of cooperation. *Trends Ecol. Evol.* **19**, 645–653 (2004).
7. S. Greiner, R. Bock, Tuning a ménage à trois: Co-evolution and co-adaptation of nuclear and organellar genomes in plants. *BioEssays* **35**, 354–365 (2013).
8. E. S. Forsythe, J. Sharbrough, J. C. Havid, J. M. Warren, D. B. Sloan, CyMIRA: The cytonuclear molecular interactions reference for *Arabidopsis*. *Genome Biol. Evol.* **11**, 2194–2202 (2019).
9. J. D. Woodson, J. Chory, Coordination of gene expression between organellar and nuclear genomes. *Nat. Rev. Genet.* **9**, 383–395 (2008).
10. G.-Z. Wu, R. Bock, GUN control in retrograde signaling: How GENOMES UNCOUPLED proteins adjust nuclear gene expression to plastid biogenesis. *Plant Cell* **33**, 457–474 (2021).
11. A. Mottis, S. Herzig, J. Auwerx, Mitochondrial communication: Shaping health and disease. *Science* **366**, 827–832 (2019).
12. I. D. Small, O. Rackham, A. Filipovska, Organelle transcriptomes: Products of a deconstructed genome. *Curr. Opin. Microbiol.* **16**, 652–658 (2013).

13. J. Song, J. M. Herrmann, T. Becker, Quality control of the mitochondrial proteome. *Nat. Rev. Mol. Cell Biol.* **22**, 54–70 (2021).
14. S. Greiner *et al.*, Chloroplast nucleoids are highly dynamic in ploidy, number, and structure during angiosperm leaf development. *Plant J.* **102**, 730–746 (2020).
15. M. Fernandes Gyorffy *et al.*, Nuclear-cytoplasmic balance: Whole genome duplications induce elevated organellar genome copy number. *Plant J.* **108**, 219–230 (2021).
16. T. Preuten *et al.*, Fewer genes than organelles: Extremely low and variable gene copy numbers in mitochondria of somatic plant cells. *Plant J.* **64**, 948–959 (2010).
17. R. Zoschke, K. Lieke, T. Börner, From seedling to mature plant: *Arabidopsis* plastidial genome copy number, RNA accumulation and transcription are differentially regulated during leaf development. *Plant J.* **50**, 710–722 (2007).
18. B. Nabholz, H. Ellegren, J. B. Wolf, High levels of gene expression explain the strong evolutionary constraint of mitochondrial protein-coding genes. *Mol. Biol. Evol.* **30**, 272–284 (2013).
19. J. C. Havird, D. B. Sloan, The roles of mutation, selection, and expression in determining relative rates of evolution in mitochondrial vs. nuclear genomes. *Mol. Biol. Evol.* **33**, 3042–3053 (2016).
20. M. F. Perez, P. Sarkies, Malignancy and NF- κ B signalling strengthen coordination between expression of mitochondrial and nuclear-encoded oxidative phosphorylation genes. *Genome Biol.* **22**, 328 (2021).
21. T. R. Mercer *et al.*, The human mitochondrial transcriptome. *Cell* **146**, 645–658 (2011).
22. J. Sharbrough, J. L. Conover, J. A. Tate, J. F. Wendel, D. B. Sloan, Cytonuclear responses to genome doubling. *Am. J. Bot.* **104**, 1277–1280 (2017).
23. 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).
24. J. E. Coate, W. M. Schreyer, D. Kum, J. J. Doyle, Robust cytonuclear coordination of transcription in nascent *Arabidopsis thaliana* autopolyploids. *Genes (Basel)* **11**, 134 (2020).
25. C. M. Bowman, Copy numbers of chloroplast and nuclear genomes are proportional in mature mesophyll cells of *Triticum* and *Aegilops* species. *Planta* **167**, 264–274 (1986).
26. 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). Accessed 6 May 2022.
27. J. Kuhn, U. Tengler, S. Binder, Transcript lifetime is balanced between stabilizing stem-loop structures and degradation-promoting polyadenylation in plant mitochondria. *Mol. Cell. Biol.* **21**, 731–742 (2001).
28. B. Castandet, A. M. Hotto, S. R. Strickler, D. B. Stern, ChloroSeq, an optimized chloroplast RNA-Seq bioinformatic pipeline, reveals remodeling of the organellar transcriptome under heat stress. *G3 (Bethesda)* **6**, 2817–2827 (2016).
29. W. Zhao *et al.*, Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics* **15**, 419 (2014).
30. D. E. Jarvis *et al.*, The genome of *Chenopodium quinoa*. *Nature* **542**, 307–312 (2017).
31. D. J. Bertoli *et al.*, The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nat. Genet.* **51**, 877–884 (2019).
32. R. Burns *et al.*, Gradual evolution of allopolyploidy in *Arabidopsis suecica*. *Nat. Ecol. Evol.* **5**, 1367–1381 (2021).
33. G. Hu *et al.*, "Evolution and diversity of the cotton genome" in *Cotton Precision Breeding*, M. Rahman, Y. Zafar, and T. Zhang, Eds. (Springer, 2021), pp. 25–78.
34. B. Li, C. N. Dewey, RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
35. G. P. Wagner, K. Kin, V. J. Lynch, Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* **131**, 281–285 (2012).
36. M. Palomar, S. Jaksich, S. Fujii, J. Kucinski, A. Wierzbicki, High-resolution map of plastid encoded polymerase binding patterns demonstrates a major role of transcription in chloroplast gene expression. *bioRxiv [Preprint]* (2022). 2022.2001.2018.476797. Accessed 6 May 2022.
37. B. Andersson, E.-M. Aro, "Photodamage and D1 protein turnover in photosystem II" in *Regulation of Photosynthesis*, E.-M. Aro and B. Andersson, Eds. (Springer, 2001), pp. 377–393.
38. G. Serino, P. Maliga, RNA polymerase subunits encoded by the plastid *rpo* genes are not shared with the nucleus-encoded plastid enzyme. *Plant Physiol.* **117**, 1165–1170 (1998).
39. S. Kikuchi *et al.*, Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science* **339**, 571–574 (2013).
40. S. Kikuchi *et al.*, A Ycf2-FtsHi heteromeric AAA-ATPase complex is required for chloroplast protein import. *Plant Cell* **30**, 2677–2703 (2018).
41. I. Wittig, H. Schägger, Structural organization of mitochondrial ATP synthase. *Biochim. Biophys. Acta - Bioenergetics* **1777**, 592–598 (2008).
42. P. Giegé, J. M. Grienenberger, G. Bonnard, Cytochrome c biogenesis in mitochondria. *Mitochondrion* **8**, 61–73 (2008).
43. L. D. Sultan *et al.*, The reverse transcriptase/RNA maturase protein MatR is required for the splicing of various group II introns in Brassicaceae mitochondria. *Plant Cell* **28**, 2805–2829 (2016).
44. C. Carrie, S. Weißenberger, J. Soll, Plant mitochondria contain the protein translocase subunits TatB and TatC. *J. Cell Sci.* **129**, 3935–3947 (2016).
45. A. Bratic *et al.*, Mitochondrial polyadenylation is a one-step process required for mRNA integrity and tRNA maturation. *PLoS Genet.* **12**, e1006028 (2016).
46. B. Heinemann, P. Künzler, H. Eubel, H.-P. Braun, T. M. Hildebrandt, Estimating the number of protein molecules in a plant cell: Protein and amino acid homeostasis during drought. *Plant Physiol.* **185**, 385–404 (2021).
47. K. J. van Wijk, S. Baginsky, Plastid proteomics in higher plants: Current state and future goals. *Plant Physiol.* **155**, 1578–1588 (2011).
48. T. A. Ruhlman, R. K. Jansen, "The plastid genomes of flowering plants" in *Chloroplast Biotechnology*, P. Maliga, Ed. (Springer, 2014), pp. 3–38.
49. Y. M. Bar-On, R. Milo, The global mass and average rate of rubisco. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 4738–4743 (2019).
50. P. Chotewumtontri, A. Barkan, Multilevel effects of light on ribosome dynamics in chloroplasts program genome-wide and psbA-specific changes in translation. *PLoS Genet.* **14**, e1007555 (2018).
51. S. Schwarte, R. Tiedemann, A gene duplication/loss event in the rubulose-1,5-bisphosphate-carboxylase/oxygenase (rubisco) small subunit gene family among accessions of *Arabidopsis thaliana*. *Mol. Biol. Evol.* **28**, 1861–1876 (2011).
52. L. Gong, M. Olson, J. F. Wendel, Cytonuclear evolution of rubisco in four allopolyploid lineages. *Mol. Biol. Evol.* **31**, 2624–2636 (2014).
53. W. Kühnbrandt, K. M. Davies, Rotary ATPases: A new twist to an ancient machine. *Trends Biochem. Sci.* **41**, 106–116 (2016).
54. P. D. B. Olinares, J. Kim, J. I. Davis, K. J. van Wijk, Subunit stoichiometry, evolution, and functional implications of an asymmetric plant plastid ClpP/R protease complex in *Arabidopsis*. *Plant Cell* **23**, 2348–2361 (2011).
55. L. Li *et al.*, Enzymes degraded under high light maintain proteostasis by transcriptional regulation in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2121362119 (2022).
56. K. Giannakis, J. M. Chustecki, I. Johnston, Encounter networks from collective mitochondrial dynamics support the emergence of effective mtDNA genomes in plant cells. *bioRxiv [Preprint]* (2021). 2021.2010.2001.462720.
57. A. Germain, S. H. Kim, R. Gutierrez, D. B. Stern, Ribonuclease II preserves chloroplast RNA homeostasis by increasing mRNA decay rates, and cooperates with polynucleotide phosphorylase in 3' end maturation. *Plant J.* **72**, 960–971 (2012).
58. C. E. Grover *et al.*, Variation in cytonuclear expression accommodation among allopolyploid plants. *bioRxiv [Preprint]* (2022). 2022.2003.2010.483551. Accessed 6 May 2022.
59. G. B. Gillard *et al.*, Comparative regulomics supports pervasive selection on gene dosage following whole genome duplication. *Genome Biol.* **22**, 103 (2021).
60. 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).
61. Z. Li *et al.*, Gene duplicability of core genes is highly consistent across all angiosperms. *Plant Cell* **28**, 326–344 (2016).
62. 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).
63. J. Sharbrough *et al.*, Global patterns of subgenome evolution in organelle-targeted genes of six allotetraploid angiosperms. *Mol. Biol. Evol.* **39**, msac074 (2022).
64. External RNA Controls Consortium, Proposed methods for testing and selecting the ERCC external RNA controls. *BMC Genomics* **6**, 150 (2005).
65. C. J. Visger, G. K. S. Wong, Y. Zhang, P. S. Soltis, D. E. Soltis, Divergent gene expression levels between diploid and autotetraploid *Tolmiea* relative to the total transcriptome, the cell, and biomass. *Am. J. Bot.* **106**, 280–291 (2019).
66. G. Yahya *et al.*, Scaling of cellular proteome with ploidy. *bioRxiv [Preprint]* (2021). 2021.2005.2006.442919.
67. J. E. Coate, J. J. Doyle, Quantifying whole transcriptome size, a prerequisite for understanding transcriptome evolution across species: An example from a plant allopolyploid. *Genome Biol. Evol.* **2**, 534–546 (2010).
68. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
69. E. Hazkani-Covo, R. M. Zeller, W. Martin, Molecular poltergeists: Mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genet.* **6**, e1000834 (2010).
70. S. Zhao, Y. Zhang, R. Gamini, B. Zhang, D. von Schack, Evaluation of two main RNA-seq approaches for gene quantification in clinical RNA sequencing: PolyA+ selection versus rRNA depletion. *Sci. Rep.* **8**, 4781 (2018).
71. M. Maldonado, F. Guo, J. A. Letts, Atomic structures of respiratory complex III₂, complex IV, and supercomplex III₂-IV from vascular plants. *eLife* **10**, e62047 (2021).
72. E. F. Pettersen *et al.*, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).