PNAS COMMENTARY

The recipe for cytonuclear interaction begins with a superabundance of plastid and mitochondrial mRNAs

Jeff J. Doyle[a,b,1]

Assembling protein complexes has been likened to cooking (1), and, as in any recipe, the individual subunits must be added in the correct ratio—stoichiometry—to cook up a good holoprotein. But, whereas ruining a cake does not damage the kitchen, a cell can be adversely affected by misassembly of key proteins and the buildup of unassembled subunits (2). The level of coordination required for proper transcription, translation, and assembly of proteins is impressive, and much remains to be understood about how it is achieved (3). Proportional synthesis of different subunits appears to be the rule for protein complexes in general (2), and there is a strong correlation between transcription and steady-state protein amount (4), so transcript levels of the individual subunits of a multimeric complex should be strongly correlated. But is this true for the handful of proteins encoded by the mitochondrial and plastid genomes that must interact with the thousands of nuclear-encoded proteins transported into the organelle (5)? In PNAS, Forsythe et al. (6) report the contribution of each organelle to the overall transcriptome, broken down by gene for the chloroplast and mitochondrial genomes and interacting nuclear partners, for 12 species of four flowering plant genera: *Arabidopsis* (which includes the small genome model, *Arabidopsis thaliana*), *Arachis* (peanut), *Gossypium* (cotton), and *Chenopodium* (quinoa).

Why was this information not already available? The problem is that conventional RNA sequencing (RNA-seq) takes advantage of the presence of the poly-A tail found on most nuclear messenger RNAs (mRNAs) to preferentially reverse transcribe them, leaving behind the ribosomal RNA (rRNA) that comprises up to 95% of total cellular RNA. But this also eliminates plant organellar mRNAs, for which a poly-A tail, if present at all, is often a marker for degradation (7). Forsythe et al. (6) circumvent this problem by using ribo-depleted RNA-seq, which, as the name implies, provides a picture of the total mRNA pool by selectively removing rRNA instead of preferentially amplifying mRNA (Fig. 1*A*).

Although nuclear genome sizes of diploid representatives of these genera vary more than 10-fold, the number of their nuclear genes is more constant (from 26,000 in *A. thaliana* to ~42,000 in *Arachis ipaensis*), as is typical for flowering plants (8). Although this is two orders of magnitude more genes than reside in the plastid or mitochondrial genome of any of these species, the apparent imbalance is compensated by the high ploidy of cells for plastid and mitochondrial genes. A mature *A. thaliana* leaf cell has on average 50 to 100 plastids, each with up to >2,000 nucleoids with multiple genome copies, for an average of over 3,000 plastid genomes per cell (9). With ~130 genes/plastid genome, if every gene in the nucleus and plastid were transcribed equally, this would result in a 10:1 ratio of plastid to nuclear transcripts (Fig. 1*B*). The situation is very different for mitochondria: mature leaves of *A. thaliana* have around 400 mitochondria, but, because the 58 genes of its mitochondrial genome often are not present in all mitochondria (10), there are likely fewer than 23,000 gene equivalents per leaf cell, which is roughly the same as the nuclear genome. But, of course, not all genes are transcribed equally from any genome, leaving open the question of the contribution of each genome to the total transcriptome.

Forsythe et al. (6) report that roughly three-quarters of the transcripts in the rRNA-depleted transcriptomes of mature leaf cells were of plastid origin in all 12 species, with most of the remaining fraction being nuclear encoded, well below the 10-fold difference in gene equivalents per cell (Fig. 1*B*). In contrast, although the mitochondrial genome is estimated to contribute roughly the same number of gene equivalents to a leaf cell as the nuclear genome, it produces fewer than 10% as many transcripts. Forsythe et al. (6) go on to show that all three genomes vary enormously in the degree to which their individual genes are transcribed, and that this differential transcription varies with function. For example, in all four genera, plastid-encoded genes involved in photosynthesis are expressed, on average, around 100-fold more than plastid-encoded genes involved in transcription. Nuclear genes involved in photosynthesis and transcription show the same relationship, but with 10-fold lower expression than their plastid counterparts in both cases (Fig. 1*C*). Intriguingly, genes from the mitochondrion behave very similarly to plastid genes and are much more highly expressed than their nuclear interactors. This suggests that, despite their different origins and biologies, the two cytoplasmic organelles interact with the nucleus in similar ways. Perhaps this is to be expected, given that the nucleus has assumed most of the responsibility for both the function and maintenance of these coexisting genomes (5). It will be interesting to extend studies beyond these eudicots, which diverged only around 120 million years ago, to determine whether correlated transcriptional responses across functional gene classes in the three genomes are underlain by

Author affiliations: [a]Plant Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853; and [b]Plant Breeding & Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853

Author contributions: J.J.D. wrote the paper.

The author declares no competing interest.

See companion article, "Organellar transcripts dominate the cellular mRNA pool across plants of varying ploidy levels," 10.1073/pnas.2204187119.

[1]Email: jjd5@cornell.edu.

Published August 9, 2022.

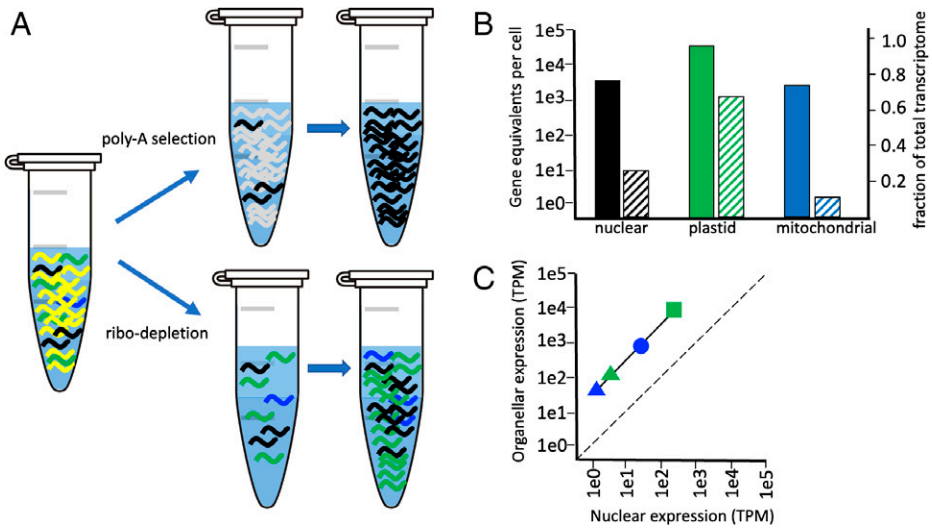

Fig. 1. (*A*) Enrichment of mRNA from total RNA. RNA isolated from plant leaf tissue (left tube) is dominated by rRNA (yellow wavy lines), swamping the much smaller amounts of the plastid (green), mitochondrial (blue), and nuclear (black) mRNAs of primary interest in RNA-seq experiments. Nuclear mRNAs have poly-A tails, making them accessible for amplification by conventional RNA-seq approaches, but the process excludes not only rRNA but also most plastid and mitochondrial mRNAs (top right tubes). Ribodepletion removes rRNA, allowing mRNAs from all three genomes to be amplified (bottom right tubes). (*B*) Comparison of predicted gene equivalents (genes/genome x genome number) per mature leaf cell for the three *A. thaliana* genomes (left scale, solid histograms) with the fraction of the mature leaf transcriptome measured by Forsythe et al. (6) for each genome (right scale, histograms with diagonal hatching). Although the plastome has 10× higher gene equivalents per cell relative to nuclear genome, plastid transcripts are only 3× more highly represented. Mitochondrial transcription is proportionally even lower. (*C*) The ratio of plastid (green) or mitochondrial (blue) to nuclear gene expression (transcripts per million) for genes grouped by functional class (triangle, transcription; circle, oxidative phosphorylation; square, photosynthesis) vary considerably but depart dramatically from a 1:1 ratio (dashed line). Linear relationships were found for all 12 plant species, with the same relative placement of functional classes (*Arabidopsis suecica* shown here).

common mechanisms throughout the green tree of life, and even, perhaps, in deeper branching lineages of the primary plastid endosymbiosis (e.g., red algae) and in independent secondary endosymbioses.

Unlike mitochondrial and plastid genes, many if not most plant nuclear genes belong to families of varying sizes due to duplications ranging from single gene events to whole genome duplications (WGD, polyploidy) (8). Increasing nuclear gene copy number could compensate for the higher expression of plastid or mitochondrial genes with which they interact, but Forsythe et al. (6) find that the average family size of *A. thaliana* nuclear genes encoding over 50 different plastid-interacting proteins and over 100 mitochondria-interacting proteins is only 1.2, far too low for effective compensation. This low copy number occurs despite the fact that all flowering plant genomes have experienced at least one WGD event, and the four genera studied here belong to lineages averaging three or four such events (11). But polyploidy is followed by gene loss as part of what one of the authors has elsewhere called "wondrous cycles of polyploidy" (12), and it has long been recognized that gene loss following duplication is not random. Dosage-sensitive genes—those predicted to interact either in coexpression networks or directly in multisubunit proteins—on average are retained in duplicate following polyploidy events, presumably because WGD preserves stoichiometry among interacting members (13). Plastid-targeted nuclear genes, however, disproportionately return to single copy in flowering plant genomes, regardless of the type of duplication, perhaps because polyploidy directly duplicates only the nuclear genome, disrupting stoichiometry for nuclear genes whose proteins interact with plastid-encoded proteins (14).

Genomes with ancient polyploid duplications reveal the long-term consequences of polyploidy, but as models are limited by the absence of diploid progenitors with which they can be compared. More can be learned about the process of polyploidization from the three species carefully chosen by Forsythe et al. (6) from each genus, comprising a polyploid and the two diploid species hypothesized to have contributed its subgenomes through recent (from <10,000 y to a few million years) hybridization and doubling (allopolyploidy). Change in gene copy number must be "felt" at the level of the transcriptome in order to have an effect on the proteome (15), so one possible outcome of polyploidy is that the stable organellar/nuclear mRNA ratios observed in diploid species might be reduced when the nuclear genome is doubled. However, Forsythe et al. (6) find, instead, that three of the allopolyploids show expression levels intermediate between those of their diploid relatives, and the fourth has higher levels of plastid expression. They also do not find consistent patterns of expression changes in polyploids for specific interacting nuclear–organellar gene pairs. Thus, it appears that any potentially negative effects of duplicating nuclear genes on cytonuclear interactions has been buffered at the transcriptomic level in these polyploids. This could be due to an increase in the number or size of organelles that accompanies the increased cell size observed in at least some cell types of polyploids (16). Supporting this idea, Fernandez Gyorfy et al. (17) have recently shown that organelle genome number is higher in polyploids of *A. thaliana* and wheat, and that this is an immediate effect of polyploidization seen even in laboratory-created polyploids.

The wide disparity between nuclear and organellar contributions to the transcriptome reported by Forsythe et al. (6), particularly for genes encoding interacting subunits, means that the recipe for proper organellar function must involve posttranscriptional processes. This is to be expected, given the dependence of protein abundance not only on mRNA

transcription rate and stability but on translation rate and protein half-life, all of which vary greatly and interact with environmental and intrinsic factors in ways that remain poorly understood even in yeast and humans (15). And, as Forsythe et al. (6) recognize, the "kitchens" in which all of this occurs are individual leaf cells comprising different cell types and states, and each having variable numbers of mitochondria and plastids, so bulk leaf transcriptomes do not tell the full story. Single-cell transcriptomic studies have revolutionized our understanding of intercellular mRNA abundance, and comparable proteomic methods could do the same for protein abundance (15). The tools are becoming available, even in plants (18), so … bon appetit!

ACKNOWLEDGMENTS. J.J.D.'s research is supported by NSF grant IOS-2127485.

1. N. A. Eckardt, Rubisco feedback loop: Control by epistasy of synthesis governs large subunit biosynthesis. *Plant Cell* **33**, 1407–1408 (2021).
2. J. C. Taggart, H. Zauber, M. Selbach, G.-W. Li, E. McShane, Keeping the proportions of protein complex components in check. *Cell Syst.* **10**, 125–132 (2020).
3. B. Slobodin, R. Dikstein, So close, no matter how far: Multiple paths connecting transcription to mRNA translation in eukaryotes. *EMBO Rep.* **21**, e50799 (2020).
4. Y. Liu, R. Aebersold, The interdependence of transcript and protein abundance: New data–new complexities. *Mol. Syst. Biol.* **12**, 856 (2016).
5. D. B. Sloan *et al.*, Cytonuclear integration and co-evolution. *Nat. Rev. Genet.* **19**, 635–648 (2018).
6. E. S. Forsythe *et al.*, Organellar transcripts dominate the cellular mRNA pool across plants of varying ploidy levels. *Proc. Natl. Acad. Sci. U.S.A.*, 10.1073/pnas.2204187119 (2022).
7. G. Schuster, D. Stern, RNA polyadenylation and decay in mitochondria and chloroplasts. *Prog. Mol. Biol. Transl. Sci.* **85**, 393–422 (2009).
8. N. Panchy, M. Lehti-Shiu, S. H. Shiu, Evolution of gene duplication in plants. *Plant Physiol.* **171**, 2294–2316 (2016).
9. S. Greiner *et al.*, Chloroplast nucleoids are highly dynamic in ploidy, number, and structure during angiosperm leaf development. *Plant J.* **102**, 730–746 (2020).
10. 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).
11. One Thousand Plant Transcriptones Initiative, One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679–685 (2019).
12. J. F. Wendel, The wondrous cycles of polyploidy in plants. *Am. J. Bot.* **102**, 1753–1756 (2015).
13. S. Tasdighian *et al.*, Reciprocally retained genes in the angiosperm lineage show the hallmarks of dosage balance sensitivity. *Plant Cell* **29**, 2766–2785 (2017).
14. 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).
15. C. Buccitelli, M. Selbach, mRNAs, proteins and the emerging principles of gene expression control. *Nat. Rev. Genet.* **21**, 630–644 (2020).
16. 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).
17. M. Fernandes Gyorfy *et al.*, Nuclear-cytoplasmic balance: Whole genome duplications induce elevated organellar genome copy number. *Plant J.* **108**, 219–230 (2021).
18. N. M. Clark, J. M. Elmore, J. W. Walley, To the proteome and beyond: Advances in single-cell omics profiling for plant systems. *Plant Physiol.* **188**, 726–737 (2022).