

TECHNICAL ADVANCE

Utilizing high-resolution ribosome profiling for the global investigation of gene expression in *Chlamydomonas*Vincent Leon Gotsmann¹, Michael Kien Yin Ting², Nadin Haase³, Sophia Rudorf³, Reimo Zoschke²  and Felix Willmund^{1,*†} ¹Molecular Genetics of Eukaryotes, RPTU Kaiserslautern-Landau, Paul-Ehrlich-Str. 23, 67663 Kaiserslautern, Germany,²Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany,³Institute of Cell Biology and Biophysics, Leibniz University Hanover, Herrenhäuser-Str. 2, 30419 Hanover, Germany

Received 29 June 2023; revised 20 November 2023; accepted 22 November 2023; published online 4 December 2023.

*For correspondence (e-mail willmund@staff.uni-marburg.de).

†Present address: Plant Physiology/Synmikro, University of Marburg, Karl-von-Frisch-Str. 14 35043, Marburg, Germany

SUMMARY

Ribosome profiling (Ribo-seq) is a powerful method for the deep analysis of translation mechanisms and regulatory circuits during gene expression. Extraction and sequencing of ribosome-protected fragments (RPFs) and parallel RNA-seq yields genome-wide insight into translational dynamics and post-transcriptional control of gene expression. Here, we provide details on the Ribo-seq method and the subsequent analysis with the unicellular model alga *Chlamydomonas reinhardtii* (*Chlamydomonas*) for generating high-resolution data covering more than 10 000 different transcripts. Detailed analysis of the ribosomal offsets on transcripts uncovers presumable transition states during translocation of elongating ribosomes within the 5' and 3' sections of transcripts and characteristics of eukaryotic translation termination, which are fundamentally distinct for chloroplast translation. In chloroplasts, a heterogeneous RPF size distribution along the coding sequence indicates specific regulatory phases during protein synthesis. For example, local accumulation of small RPFs correlates with local slowdown of *psbA* translation, possibly uncovering an uncharacterized regulatory step during PsbA/D1 synthesis. Further analyses of RPF distribution along specific cytosolic transcripts revealed characteristic patterns of translation elongation exemplified for the major light-harvesting complex proteins, LHCs. By providing high-quality datasets for all subcellular genomes and attaching our data to the *Chlamydomonas* reference genome, we aim to make ribosome profiles easily accessible for the broad research community. The data can be browsed without advanced bioinformatic background knowledge for translation output levels of specific genes and their splice variants and for monitoring genome annotation.

Keywords: ribosome profiling, next-generation sequencing, ribosomes, protein synthesis, translation, transcript accumulation, *Chlamydomonas reinhardtii*.

INTRODUCTION

Protein synthesis is the origin for the quantitative and qualitative determination of a proteome's diversity, which ultimately shapes the biochemical character of a cell. It is now clear that translation output does not strictly follow the changes of transcript abundance, but instead underlies strict regulatory control (Payne, 2015; Pechmann et al., 2013; Slobodin & Dikstein, 2020). The importance of this regulation is illustrated by the fact that bacteria invest approximately 50% of their available energy for protein biogenesis and that more than 10% of all yeast proteins

may be somehow involved in this process (Costanzo et al., 2000; Russell & Cook, 1995). Translational regulation is observed throughout an organism's life span and includes central regulatory circuits that control the cell cycle, tissue differentiation and development, acclimation to environmental changes and the integration of external signals (reviewed in Pechmann et al., 2013; Teixeira & Lehmann, 2019; Wang & Amoyel, 2022; Zoschke & Bock, 2018). Thus, understanding the regulatory principles and mechanisms of translation is a central task of modern biology. With an adaptation of next-generation

sequencing, ribosome profiling systematically assays the status of cellular translation (Ingolia et al., 2009). The technology is based on the observation that ribosomes occupy precise sections of a translated mRNA (Wolin & Walter, 1988). These mRNA sections (termed ribosome footprints or ribosome-protected fragments, RPFs) are extracted by nucleolytic removal of unoccupied mRNA in cell lysates or isolated ribosomes. RPFs are then identified through next-generation sequencing and mapping to a reference transcriptome or genome. Thus, ribosome profiling (termed Ribo-seq in this study) benefits from the same dynamic range as RNA sequencing and provides a rich and precise positional information of translating ribosomes from the physiological state when cells were harvested (Brar & Weissman, 2015; Ingolia, 2014; Ingolia et al., 2009). Parallel sequencing of all expressed transcripts (termed RNA-seq in this study) allows the comparison of RNA accumulation with translation and thereby determining the fraction of translated transcripts, which is often referred to as 'translation efficiency'. Assuming that each RPF represents a translating ribosome and thus a synthesized nascent polypeptide, mRNAs with high average RPF coverage can be considered as highly translated transcripts if initiation is the limiting process. In addition, local over-representation of specific RPFs within a specific protein-coding sequence (CDS) are commonly interpreted as regions of slow or even halted translation (Brar & Weissman, 2015; Ingolia, 2014; Ingolia et al., 2009). Compared with other techniques, ribosome profiling goes far beyond the targeted assays of radio-labeling of nascent polypeptides or polysome loading for specific mRNAs, and regularly provides much deeper dataset information than proteomic analyses via mass spectrometry (Brar & Weissman, 2015; Ingolia, 2014). Yet, ribosome profiling faces the challenges of distinguishing active translation events from non-translating (paused or stalled) ribosomes. However, high-quality Ribo-seq datasets, with populations of precise read length distribution allow the determination of active translation events if RPFs are aligned relative to the ribosomal Peptidyl-site (P-site). This reveals a 3-nucleotide periodic movement of translating ribosomes along a translated CDS, as shown by an over-accumulation of reads containing the first nucleotide of a decoded codon in the P-site. Thus, it is possible to distinguish ribosome pausing/stalling from translation and therefore even enables *ab initio* annotation of uncharacterized protein-coding sequences (Calviello et al., 2016; Hsu et al., 2016; Ingolia et al., 2009).

Studying translational regulation in photosynthetic eukaryotes is of specific interest. These cells need to orchestrate gene expression between three semi-autonomous genomes (nuclear, chloroplast, and mitochondrial) and are specifically dependent on the quick and efficient acclimation to environmental changes such as temperature and light. Chloroplast gene expression seems

particularly dependent on co-translational regulation, which facilitates the fast adjustment of the photosynthesis machinery to environmental changes and the stoichiometric assembly of chloroplast multi-subunit complexes which often contain nuclear and chloroplast-encoded subunits (Eberhard et al., 2002; Willmund et al., 2023; Zoschke & Bock, 2018). Recently, ribosome profiling revealed the specific adjustments made to chloroplast gene expression during acclimation to different light and temperature regimes (Chotewutmontri & Barkan, 2018, 2020; Gao et al., 2022; Schuster et al., 2019; Trösch et al., 2022). In general, Ribo-seq has been applied to different plant species with variable gene coverage and quality, including *Chlamydomonas reinhardtii* (*Chlamydomonas* hereafter) (Chung et al., 2015, 2017; Wu et al., 2022), *Arabidopsis thaliana*, (Chen et al., 2022; Hsu et al., 2016; Juntawong et al., 2014; Liu et al., 2013; Lukoszek et al., 2016; Merchante et al., 2015), maize (Chotewutmontri & Barkan, 2018; Lei et al., 2015), and other crop plants (Chiu et al., 2022; Wu, Song, et al., 2019; Wu, Zinshteyn, et al., 2019; Yang et al., 2020, 2021). However, a systematic comparison of translation between the three subcellular genomes is rare. While the earlier *Chlamydomonas* Ribo-seq studies did not focus on chloroplast gene expression (focusing on RNA-induced gene silencing and ciliogenesis; (Chung et al., 2017; Wu et al., 2022)), we previously investigated chloroplast translation with high-resolution microarray (Trösch et al., 2018, 2022). The approach is fast and cost-efficient and primarily focuses on chloroplast-encoded transcripts, whereas the limited resolution does not allow for the study of mechanistical details of translation initiation and elongation.

Here, we aimed to generate high-quality Ribo-seq data with sub-codon resolution for the systematic and mechanistic comparison of translation in the three *Chlamydomonas* compartments, namely cytosol, chloroplast, and mitochondrion. Furthermore, our data will support the *Chlamydomonas* research community: By directly linking our data to genes of the most recent *Chlamydomonas* genome v6 (Craig et al., 2022) we have made it readily accessible for specific genes of interest through the genome portal.

RESULTS AND DISCUSSION

Establishing super-resolution ribosome profiling for *Chlamydomonas reinhardtii*

Ribosome profiling was performed from logarithmically grown *Chlamydomonas* cultures that were kept under mixotrophic conditions and moderate light (see Experimental Procedures). To avoid ribosome run-off (Trösch et al., 2018, 2022), the elongation inhibitors chloramphenicol (CAP) and cycloheximide (CHX) (stalling 70S and 80S ribosomes, respectively) were added just before harvest, to keep the exposure time to the minimum. For nucleolytic

digestion of RNA that is not protected by a ribosome, we applied RNase I, consistent with most Ribo-seq studies on other organisms, avoiding possible sequence biases during digestion as reported for MNase (Dingwall et al., 1981), which improves three-nucleotide periodicity of the data (see below). Our previous chloroplast ribosome profiling approaches (using MNase) revealed that the best digestion (i.e., polysome to monosome dissociation) was achieved when ribosomes were purified prior to nuclease treatment (Trösch et al., 2018). RNase I digestion is usually performed at room temperature or 4°C. We thus compared the digestion on pre-purified ribosomes both at 4°C (condition i.) and 23°C (condition ii.) (Figure 1a). In addition, digestion in the cell lysate was tested. Here, the reaction was performed at 4°C to avoid complex disassembly and proteolytic degradation. For all digests, RNase I was applied at a concentration of 1 unit (U) per microgram (µg) RNA, which is similar to published Ribo-seq approaches performed in yeast and *Arabidopsis* (Chartron et al., 2016; Döring et al., 2017; Hsu et al., 2016). However, the RNase I concentration was far below the >9 U per µg RNA that was previously applied for *Chlamydomonas* (Chung et al., 2015). Importantly, all steps of the Ribo-seq protocol – including deletion of contaminating rRNA species and sequencing library generation – are highly comparable to a Ribo-seq protocol for *Arabidopsis* and tobacco that we established in parallel (Ting et al., 2023). Resulting sequencing libraries were corrected after genome mapping by deduplication based on unique molecular identifier tags (UMI tags).

Parallel to Ribo-seq, transcript accumulation was determined by sequencing whole-cell RNA samples on the same Illumina flow cell. Ideally, RNA samples are harvested just prior to ribosome profiling experiments from the same culture. While typical transcriptomic studies enrich mRNA pools using polyadenylation enrichment, we here sequenced fragmented small RNAs that are derived from the full set of transcripts including plastid transcripts. For chloroplast and mitochondrial transcripts, enrichment via polyadenylation is not comparable to the respective enrichment of cytosolic transcripts, since the plastid post-transcription modification was reported to serve as destabilization signal opposed to acting as stabilization signal of cytosolic transcripts (Gallaher et al., 2018; Komine et al., 2000).

Ribosome profiling and transcript samples were sequenced with a depth of 40 M and 20 M reads, respectively. A higher depth was selected for the Ribo-seq samples, due to the over-proportional content of contaminating rRNAs (as determined by pre-runs). The average read content of 25.7% mapped to annotated CDSs of the most recent *Chlamydomonas* reference genome v6.1 (Craig et al., 2022), 64% mapped to putative rRNA loci and 4.2% mapped to annotated untranslated regions

upstream or downstream of CDSs. The remaining reads mapped to non-annotated regions, likely deriving from rRNA-encoding loci. For digest condition i. and ii., abundant rRNA species were depleted by hybridization to biotinylated oligos and subsequent extraction. For condition iii., the same rRNA contaminants were additionally removed by RNase H treatment, which cleaves hybrids of the rRNA-oligo molecules (see Experimental Procedures). Despite the application of different rRNA depletion methods, the rRNA contamination was very similar in all samples. Most sequenced reads that corresponded to annotated CDS, mapped to unique loci within the genome. To avoid false information from reads with multiple matches (<1% of all reads), we set an arbitrary threshold removing all reads that matched to more than 10 positions within the genome. Of the Ribo-seq reads mapping to CDSs 97.77% mapped to transcripts of the nuclear genome, 2.15% mapped to chloroplast-encoded transcripts, and 0.07% to mitochondrial transcripts (averaged between i. and iii.). In contrast, RNA-seq reads mapping to CDSs had an average of 82.13% for nuclear transcripts, 17.59% for chloroplast transcripts, and 0.28% for mitochondrial transcripts, across replicates, indicating a higher proportion of accumulating chloroplast transcripts relative to their translation.

Transcript accumulation was determined by averaging the reads that mapped to a respective CDS normalized to the length of the CDS and the total read numbers of the dataset (Reads Per Kilobase per Million, RPKM), a parameter that is widely used for transcriptome analyses and accounts for a higher likelihood of measuring long RNA fragments in comparison with short sequences (Conesa et al., 2016). This is of particular importance for the sequencing of small and fragmented RNAs, as commonly accompanied with Ribo-seq. Gene-wise averaging of RPFs that mapped to a respective CDS and subsequent RPKM calculation corrects for local variations of RPF distribution and allows to get a proxy for the level of translation output per transcript (Ingolia, 2014; Ingolia et al., 2011).

RPKM values were highly comparable between the different nuclease digestion conditions i. to iii. ($r = 0.99$, Figure 1b) of the Ribo-seq data, suggesting that our RNase treatment conditions generally result in robust ribosome profiling data. However, different RNase treatments may affect coverage and resolution of the data. We thus determined the depth of the Ribo-seq approach. Low coverage Ribo-seq frequently leads to ribosome profiles that cover only the transcripts with the highest ribosome occupancy (i.e., translation output), while moderately or poorly translated CDSs may exhibit partial RPF coverage along the corresponding sequence, referred to as ‘gene body coverage’ (Figure 1c, left panel). In Ribo-seq with adequate depth, many CDSs ought to display adequate gene body coverage (right panel, Figure 1c). The depth was determined for all

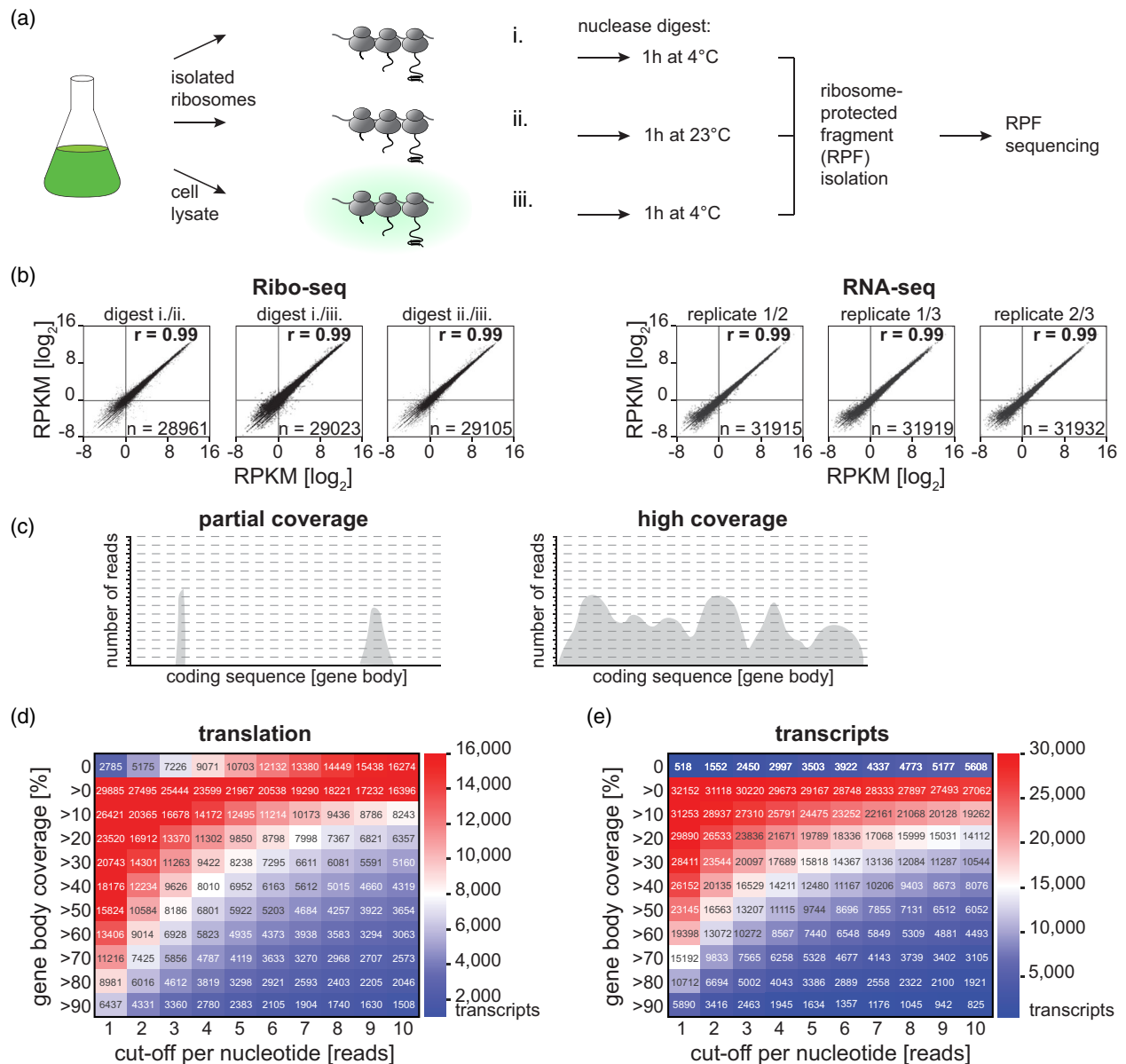


Figure 1. Ribo-seq provides a broad coverage of expressed genes.

(a) Illustration of the experimental setup, testing nucleolytic conditions. Polysome purification and subsequent RNase I digest for 1 h at 4°C (i.) or 1 h at 23°C (ii.) were performed from *Chlamydomonas* samples, respectively. Alternatively, cell lysate was directly supplemented with RNase I (iii.).

(b) Scatter plots representing the reproducibility of Ribo-seq (left panels) and RNA-seq experiments (right panels) based on RPKM values. Pearson correlation r -values and number of quantified CDS are given.

(c) Schematic representation of the concept of gene body coverage (left panel: incomplete coverage that declines strongly with an increasing read cutoff; right panel: good coverage that is robust against increasing read cutoffs).

(d) and (e) Cumulative heat maps visualize the number of genes (given in the respective panel) covered to a certain minimal extent (rows) by a certain minimal number of reads (columns) for ribosome RPFs (c) and transcripts (d). Ribo-seq data shown are derived from the RNase I digest in lysate (condition iii.). The corresponding data for the other treatments are shown in Supporting Figure S1.

Chlamydomonas transcripts of the v6.1 genome (32 670 transcripts considering the products of gene copies). For digest condition iii., more than 10 000 transcripts had at least 50% of their sequence covered with two or more RPFs, while 6437 transcripts had over 90% of their CDS covered with at least one RPF (Figure 1d). 1508 transcripts are

covered over >90% with an average of more than 10 RPFs/nt, hence representing the group of transcripts with the most complete gene body coverage combined with a high basal coverage (see also Figure 4b). Categorizing the transcripts in this way allows for the assessment of the general quality of a dataset, as well as the selection of

transcripts of similar quality for specific analyses without having to rely on expression values alone. The gene body coverage was highest for digest condition iii. (Figure 1d and Figure S1) and comparably good for the coverage of RNA fragments for the RNA-seq experiment (Figure 1e). The read length distribution of the RPFs mapping to annotated CDSs of the nuclear genome is within the size range of previously published Ribo-seq data from plant species. In *Arabidopsis* or maize, RNase I treatment typically results in RPFs with a length between 28 and 31 nt (Chotewutmontri & Barkan, 2016; Hsu et al., 2016; Merchante et al., 2015), which is also comparable to yeast and mammalian cells (Ingolia et al., 2012). Our digestion of purified ribosomes at 4°C resulted in RPFs with a predominant length of 31 nt. Digest condition ii. caused a broader distribution of RPFs between 28 and 30 nt, while digestion in cell lysate resulted in a sharp peak of RPFs with a length of 30 nt (Figure 2a). Such a sharp peak at 30 nt is typically found in Ribo-seq from mammalian cells when harvested in the presence of CHX (Sharma et al., 2021; Wu, Song, et al., 2019; Wu, Zinshteyn, et al., 2019). The sharp peak is even more noticeable if only those RPFs are considered that map to the CDS (Figure S2a). This is most apparent for the size distribution mapping to the CDS of the eight mitochondrial transcripts. Here, RPF accumulation peaks at a size of 33 nt (Figure S2a). The larger size of mitochondrial RPFs was described before and could stem from the bulkiness of the membrane-associated ribosomes (Rooijers et al., 2013). Whereas cytosolic and mitochondrial ribosomes produced sharp RPF peaks, the RPF size distribution of chloroplast transcripts is less defined and may be the result of diverse translational states, as previously reported for maize and *Arabidopsis* (Chotewutmontri & Barkan, 2016; Fujita et al., 2019; Gawronski et al., 2018) (see below). The sharpest peak for chloroplast RPFs is seen under condition ii., albeit the predominant RPFs of 28 nt are found under all treatment conditions (Figure 2a). While consistent with the observations in *Arabidopsis*, maize and tobacco (Chotewutmontri & Barkan, 2016; Gawronski et al., 2018; Ting et al., 2023), our RPF size distribution varies from the previous Ribo-seq study with *Chlamydomonas*, which reported a general shift toward smaller RPF sizes (26–27 nt), maybe as consequence of their harsh RNase I treatment (Chung et al., 2015).

A hallmark of high-resolution Ribo-seq data is the triplet-wise 3-nt periodicity of RPFs that represents slow-down of ribosome movement as ribosomes decode a respective codon (Ingolia, 2016). To calculate periodicity, it is necessary to determine the 5'-offset for every RPF length species, which is the number of bases located 5' to the P-site (the translational state of polypeptide bond formation) of a respective RPF. This is usually done by considering all RPFs that fully enclose an annotated start codon and determining the frequency of all occurring 5'-offsets

for every RPF length separately. The 5'-offset most frequently occurring for a specific RPF length is then set to be the 5'-offset of this species. This strategy is based on the fact that RPF occupancy of transcripts usually increases significantly at the borders between 5'-untranslated regions and CDS, with the start codon often having an exceptionally high coverage because of the slow process of translation initiation (Figure S2b). The 5'-offsets determined in this way are used to calculate the exact P-site position and reading frame of every RPF. By this, 3-nt periodicity can be nicely visualized in meta-plots combining all genes while 3-nt plots of individual transcripts can be noisy, depending on translation levels. The exact 5'-offset varies in an organism- and protocol-specific manner and is usually 12–13 nt. Alternatively, the P-site position has also been determined by calculating the 3'-offset of the ribosomal aminoacyl site (A-site), applying the same strategy to RPFs mapping to annotated stop codons.

With the majority of cytosolic RPFs having one read length (Figure 2a), all three RNase I digestion conditions yielded a clear frame preference of 'frame 0' for nucleus-encoded transcripts, meaning that the base triplet found at the P-site of a RPF represents the majority of RPFs in frame with the annotated CDS. Again, lysate digestion (iii.) resulted in the best frame preference values (>68% of RPFs in frame 0, 19.4% in frame 1 and 12.5% in frame 2), suggesting that this condition yields Ribo-seq data of the best quality (Figure 2b; Figure S2b). Due to the differences between 80S and 70S ribosomes, 3-nt periodicity needs to be determined separately for chloroplast translation. For chloroplast transcripts, the frame preference was less obvious, albeit it was clearest for digest condition iii. (Figure 2c; Figure S2b). Considering the overall quality of sample iii. (which had a similar degree of rRNA contamination as the other two samples and a high correlation with the other samples at the RPKM level), it can be concluded that RNase H-mediated rRNA depletion has no adverse effects on the Ribo-seq sample quality and seems to be similarly effective.

To get a better insight into the RPF offsets and periodicity, different lengths of the RPFs were plotted against their offset frequencies upon P-site and A-site alignment for the dataset of condition iii. (Figure 2d). The triplet-wise movements of the 5'-ends with an offset of 12 nt during initiation are clearly visible for cytosolic RPFs and most prominent for the 30 nt RPFs of digestion conditions iii., consistent with the observation in other eukaryotic organisms (Lauria et al., 2018). Likewise, triplet-wise movements of the 3'-ends are also detectable and are in frame with the stop codon. Here, a 9 nt long 3'-offset seems to be the predominant form prior to termination with the stop codon in the A-site of the ribosome. However, due to the 12 nt 5'-offset and 3 nt occupied by the P- and A-sites respectively, a 3'-offset of 12 nt was expected. This discrepancy could be a result of CHX treatment, which locks ribosomes in a

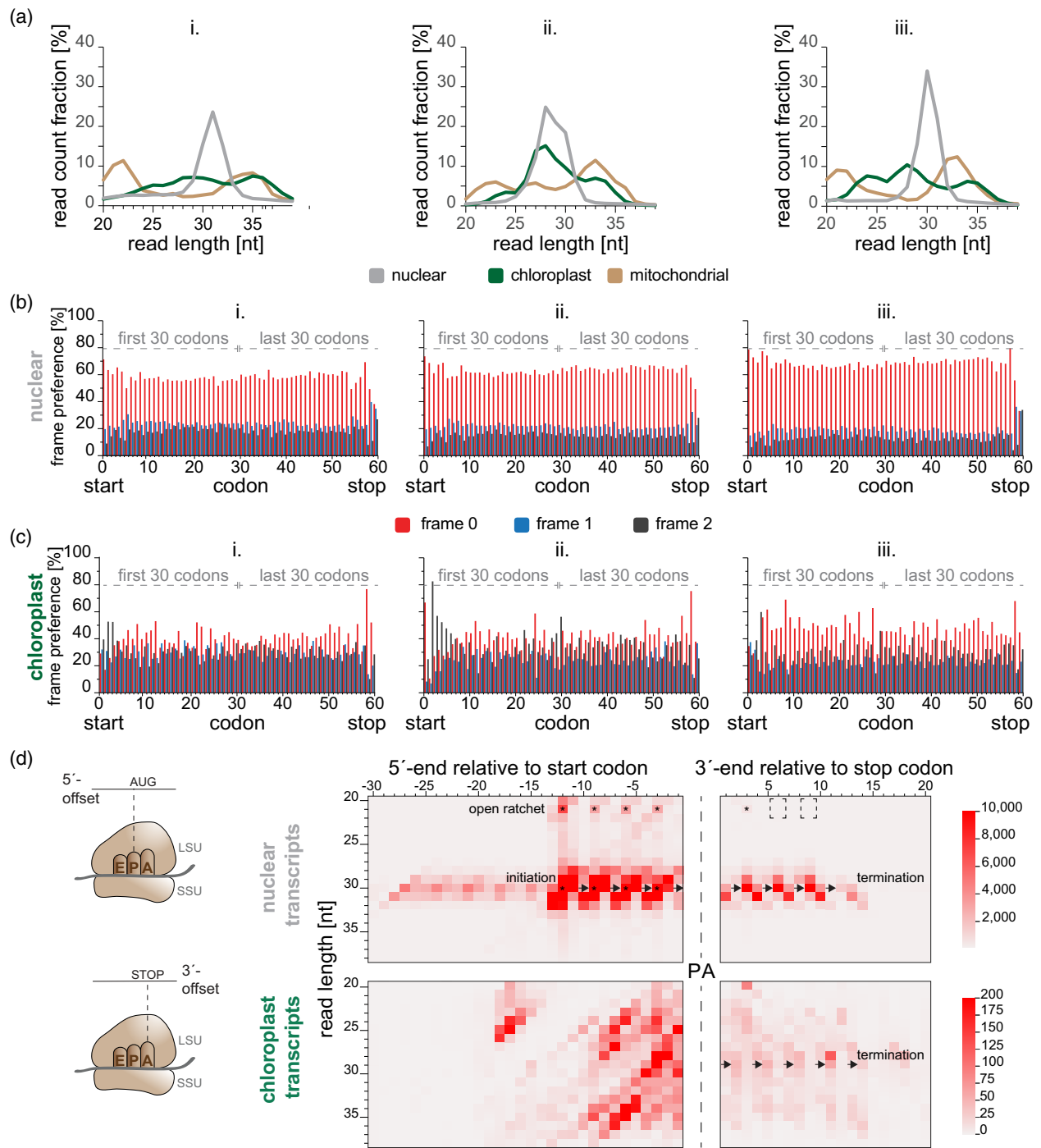


Figure 2. Tri-nucleotide periodicity of cellular translation.

(a) Length distribution of RPFs for each organelle from the three different nuclease treatments tested (i. and ii.: digestion of purified ribosomes at 4°C and 23°C, respectively, iii.: digestion in lysate at 4°C).

(b) and (c) metagene relative frame preference of nuclear and chloroplast-encoded transcripts per codon for all three nuclease treatments. Codons 1–30 represent the first 30 codons of each transcript while codons 31–60 represent the last 30 codons of each transcript. Frame 0 corresponds to AUG in the ribosomal P-site.

(d) Left side: cartoon of the ribosome with small (SSU) and large (LSU) subunits. Heatmaps represent normalized 5'-P-site offset and 3'-A-site offset counts for nucleus- and chloroplast-mapped RPFs that fully enclose a start or stop codon. Columns represent 5'/3'-RPF end positions relative to the start or stop codon while rows represent the RPF lengths. The shades of red are proportional to the number of RPF ends mapping onto a tile. Arrows indicate the triplet-wise movement of ribosomes in early or late elongation state. Asterisks indicate the tiles harboring the most 5'- or 3'-ends of the 21 and 30 nt RPF species. Dashed boxes indicate the sudden absence of 21 nt RPFs for the last two codons before translation terminates.

pre-translocation conformation after peptide bond formation (Wu, Song, et al., 2019; Wu, Zinshteyn, et al., 2019) so that RPFs on non-stop codons may exceed the natural termination-caused accumulation on stop codons. Moreover, a slow termination process might arrest ribosomes on the codon just upstream of the stop, prolonging ribosome dwell time at this site. At the same time, ribosomes that already entered termination during sampling are likely to complete the process and dissociate from the transcript. CHX does not inhibit termination, which in turn would dramatically lower the number of RPFs mapped to stop codons with their A-site. Interestingly, RPFs with a length of 21 nt are detectable upon initiation (asterisks, Figure 2d). These fragments have been previously characterized as RPFs of ribosomes in an open-ratchet conformation during translocation, even accumulating as abundant species without CHX treatment in yeast (Lareau et al., 2014). During elongation, ribosomes undergo massive conformational changes by rotating the large subunit relative to the small subunit (Frank & Agrawal, 2000; Zhang et al., 2009). The fact that the RPFs produced from the open-ratchet conformation have the same 5'-offsets as the main RPF species despite being considerably shorter, suggests that the RNase exclusively removes nucleotides from the 3'-end of RPFs in this conformation (and could indicate that the center of the rotation is located toward the 5'-direction, viewed relative to the RPF). Cycloheximide was shown to prevent binding of the next aminoacylated tRNA in the A-site during peptide bond formation (Schneider-Poetsch et al., 2010) and thus stabilizes ribosomes in a specific stage during the elongation cycle, in which RPFs of 28–30 nt are the prevalent form. Interestingly, human 80S ribosomes are not exclusively arrested in one conformation upon CHX treatment (Sharma et al., 2021), comparable with the situation seen here for *Chlamydomonas* 80S ribosomes (Figure 2d, upper panel). These 20–22 nt RPFs might derive from a post-translocation ribosome population (Lareau et al., 2014). The presence of both species demonstrates active movement of ribosomes along the CDS. Interestingly, the frame preference for frame 0 of the stop-minus-2 codon is the highest in the whole metagene periodicity plot (79.6% Figure 2b), which could denote a short stalling at this codon before translocation. However, for the next codons the frame 0 preference drops considerably to 55.8%, especially in favor of frame 1 (36.2%) and then to 33.1% at the stop codon, clearly indicating that translation dynamics change toward the termination step. Moreover, these characteristics are accompanied by a continuous decline of absolute numbers of reads mapping to these codons on the metagene level (for digest iii: stop-2: 19890, stop-1: 6786, stop: 522).

The prokaryotic-type chloroplast RPFs 5'-offsets appear more diverse than their cytosolic counterparts, as evident by their multiple clusters on the heat map

(Figure 2d, bottom panel). These could resemble populations of initiation complexes in different states or in complex with transcript-specific RNA processing factors and *trans*-acting factors that interact with ribosomes (Westrich et al., 2021). Without clear criteria on how to distinguish and interpret these clusters, definition of 5'-offsets is problematic, as many RPF lengths have multiple 5'-offset frequencies of comparable amplitude for many RPF lengths. However, the chloroplast 3'-offsets seem less scattered and reveal a main RPF species of 28 nt length with 11 nt 3'-offset. This is in accordance with the RPF length distribution of all chloroplast RPFs, where the 28 nt species also represents the main species, although the distribution is much broader than for the cytosolic RPFs. Like the cytosolic offset heat map, a triplet-wise movement of ribosomes can also be observed for the terminating chloroplast ribosomes in the 3'-offset heatmap for 29 nt RPFs, albeit with much lower clarity. Altogether, the observations made for chloroplast offsets and RPF length distribution might indicate a certain heterogeneity of the chloroplast ribosome pool, which is an interpretation that would be in line with the large number of proteins that interact with chloroplast ribosomes found in our previous study (Westrich et al., 2021) and the large number of RNA-binding proteins known or suspected to regulate the translation of specific chloroplast transcripts (Nickelsen et al., 2014). This is remarkable considering the very small number of transcripts encoded on the chloroplast genome (72). Together, this suggests that translation in the chloroplast of *Chlamydomonas* is heavily influenced by transcript-specific features rather than following a universal scheme (Zoschke & Bock, 2018).

Differences in translation output between cytosolic and organellar transcripts

We next compared the translation levels of the organellar versus nuclear transcripts. *Chlamydomonas* cells contain on average 83 plastid genomes within the large cup-shaped chloroplast occupying about 40% of the cellular volume. The mitochondrial genomes exist in 130 copies, also greatly outnumbering the one haploid nuclear genome (Gallagher et al., 2018). Despite the rather low contribution of organellar RPFs relative to the total RPF pool, both chloroplast and mitochondrial transcripts had higher RPKM values when compared to the average RPKM values of nuclear genes (Figure 3a). A similar trend is seen for transcript levels (Figure 3b), suggesting that organellar genes are expressed at considerably higher levels compared with most nuclear genes (Forsythe et al., 2022).

With the high reproducibility among the three nuclease treatment conditions, we chose to treat these datasets as independent replicates for further analyses of transcript accumulation and translation output. A direct comparison between translation output and transcript accumulation is

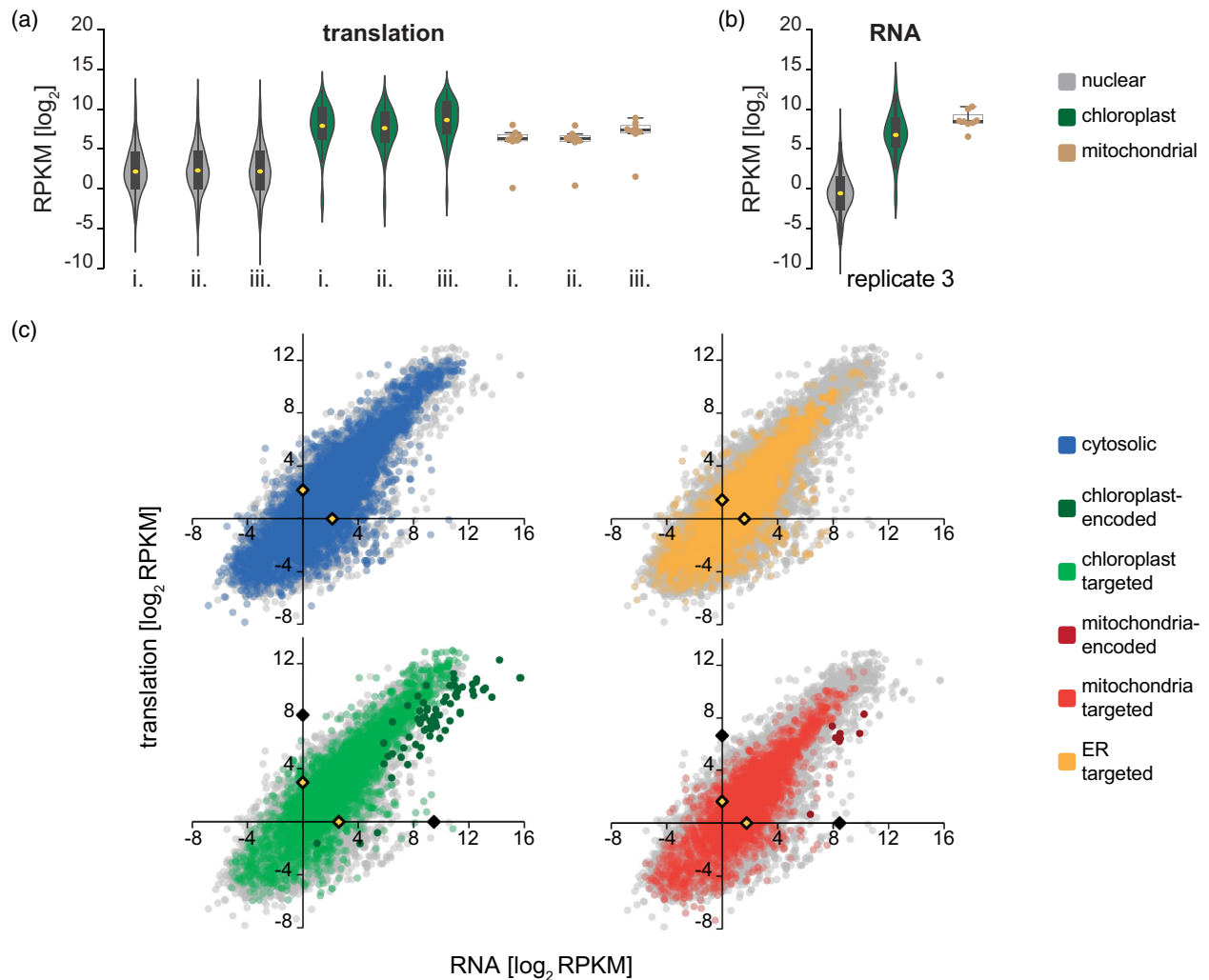


Figure 3. Chloroplast-encoded proteins are highly expressed.

(a) Violin plots or box plots (due to the low number of mitochondrial genes) of the RPKM distribution for genes of each sub-cellular genome, following the three different RNase I treatments (i. and ii.: digest of purified ribosomes at 4°C and 23°C, respectively, iii.: digest in lysate at 4°C). Values with a RPKM cutoff <2 (log₂) are excluded.

(b) Violin plots or dot plots, displaying distribution of RPKM (reads per kilobase of exon per million reads mapped) for RNA values with a cut-off >2 (log₂) for each subcellular cell genome in replicate 3 exemplary. Median values are shown in yellow.

(c) Scatterplots of the averaged Ribo-seq and RNA-seq (RPKM) data for different subcellular-localized proteins. Coloring indicates subcellular localization and origin of the encoded proteins. Rectangles indicate median values (yellow fill for nuclear-encoded transcripts, solid fill for organellar-encoded transcripts).

often referred to as translation efficiency (TE) and normalizes RPK values over transcript accumulation. Contrasting translation of the *Chlamydomonas* genes over transcript data in a scatter plot and highlighting their subcellular localization showed that nuclear transcripts encoding chloroplast-localized proteins have higher expression values when compared to proteins of cytosolic or endoplasmic reticulum localization (Figure 3c; Figure S3), again supporting the notion that many chloroplast proteins are expressed at high levels. Interestingly, most nucleus-encoded transcripts showed similar levels between transcript accumulation and translation (distributed along the diagonal, Figure 3c). In contrast, chloroplast- and

mitochondria-encoded transcripts showed higher levels of transcript accumulation, when compared to translation, which agrees with the published data of profound potential translational regulation within the chloroplast (Sun & Zerges, 2015; Trösch et al., 2018; Zoschke & Bock, 2018). Remarkably, translational regulation also seems to be predominant for the few mitochondrial genes.

Column plots of the individual chloroplast transcript and translation values show that even the weakest expressed known chloroplast transcripts have log₂ RPKM values >3, which are within the median range of nucleus-encoded transcripts (Figures 4a versus 3a). Not surprisingly, transcripts that encode the core photosystem I

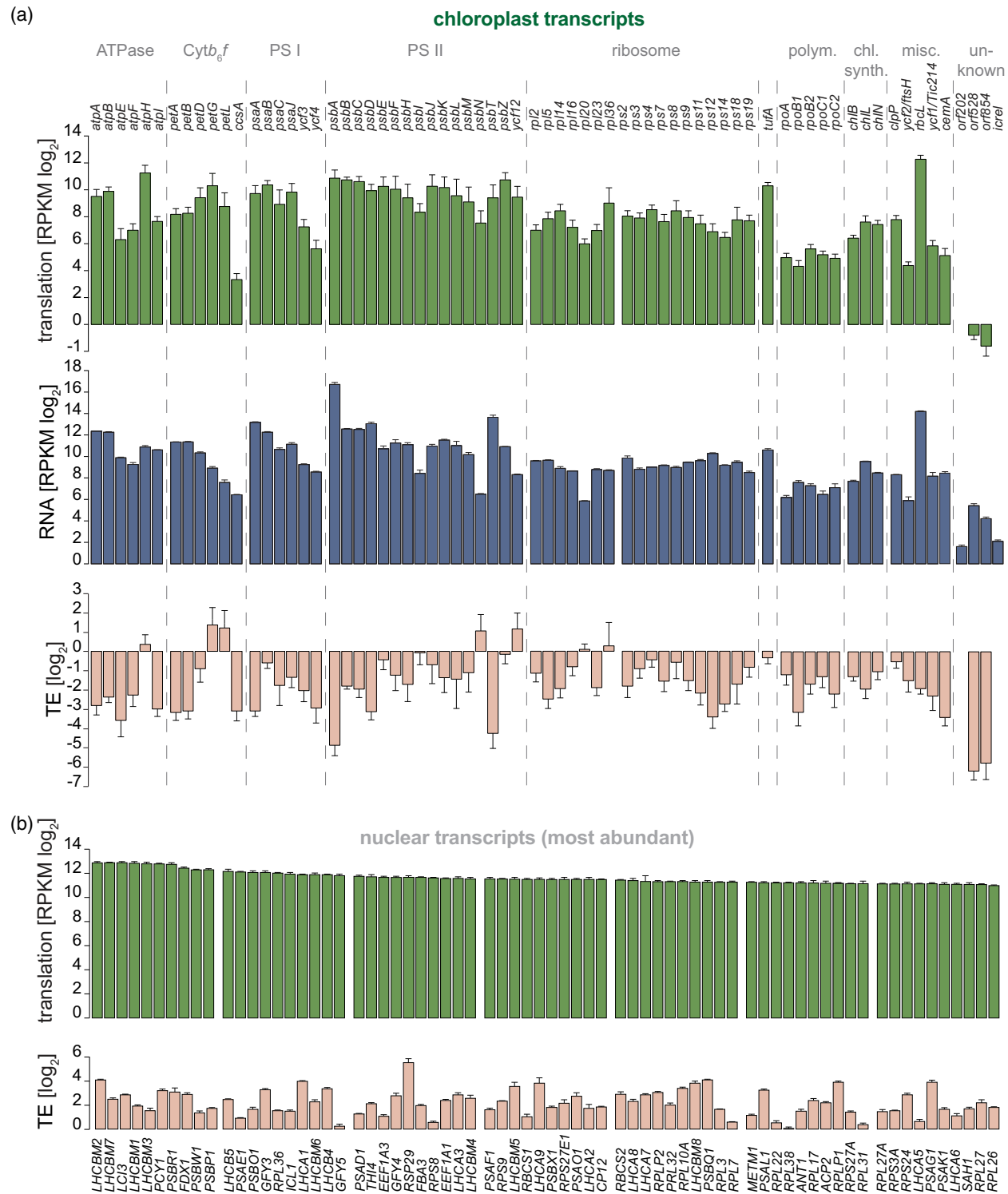


Figure 4. Translation output of chloroplast- and nucleus-encoded transcripts. (a) Mean read intensity/TE ratio (log₂ RPKM of each respective coding sequence) of chloroplast-encoded transcripts for translation (Ribo-seq), transcripts (RNA-seq) translation efficiency (TE, translation over transcript RPKM values). The transcripts are grouped by function of the encoded proteins (Cytb₆f = Cytochrome b₆f complex, PS I = Photosystem I, PS II = Photosystem II, polym. = RNA polymerase, chl. synth. = chlorophyll synthesis, misc. = miscellaneous). (b) Log₂ RPKM values of the 70 most highly expressed nucleus-encoded genes on the translation level. Only the longest splice variants were considered. Respective translation efficiency for these transcripts is given below. For comparison, transcripts with the top 70 highest TE values are given in Figure S4a. Error bars indicate standard deviation of the three experiments (treatment i. to iii. for Ribo-seq, and biological replicate 1–3 for RNA-seq) and are shown unidirectionally to improve readability.

subunits PsbA and PsbB, the photosystem II core subunits PsbA/D1, PsbB/CP47, PsbC/CP43, PsbD/D2, the Rubisco large subunit, RbcL, the CF₀ ATPase subunit AtpH, and the translation elongation factor TufA show the highest translation levels, consistent with our previous chloroplast ribosome profiling experiments and the examination of protein accumulation levels (Schroda et al., 2015; Trösch et al., 2018). We further detected transcripts of the uncharacterized chloroplast Open Reading Frames (ORF) *orf202* (positioned within the transposon Wendy I), *orf854* (positioned within the transposon Wendy II), *orf528*, and *i-crel* (Gallaher et al., 2018). However, translation output was low or even undetectable (for *orf202* and *i-crel*), suggesting that these transcripts are barely translated in *Chlamydomonas* cultures, at least under standard conditions (Figure 4a). The TE ratios of the chloroplast-encoded transcripts again show different principles of gene expression between nuclear- and chloroplast-encoded genes. Chloroplast-encoded transcripts have lower translation output values relative to transcript accumulation values (non-log RPKM ratios <1) when compared to nuclear-encoded TE values (Figure 4a versus 4b). This supports the hypothesis of vast post-transcriptional control in expression of chloroplast genes (Willmund et al., 2023; Zoschke & Bock, 2018) and that the long-living chloroplast transcripts serve as a buffering capacity to quickly respond to environmental changes, as seen, for example, seen during light and temperature acclimation and most prominently for PsbA/D1 synthesis (Chotewutmontri & Barkan, 2018; Schuster et al., 2019; Trösch et al., 2022).

We next determined translation output levels of nucleus-encoded transcripts, only considering the longest transcript variant if multiple splice forms existed. Of the transcripts with the highest translation output, 51 protein products have a known or predicted chloroplast localization and 47 are present in the nucleus or cytosol (with ~2500 chloroplast-localized proteins encoded by the 17 000 nuclear genes, *Chlamydomonas* genome v6.1), again showing the high relative accumulation levels of chloroplast proteins (Figure S4a). Among those, transcripts of the light-harvesting complex proteins (LHCs) have the highest translation output, consistent with their high protein accumulation levels in plant cells (Dall'Osto et al., 2015). Generally, LHC translation showed comparable RPKM translation output values, while transcript accumulation was rather variable. For the transcripts of the highlight-induced photoprotective LHCSR3.1 and LHCSR3.3 proteins, we could not determine translation profiles (Figure S4b), suggesting that these proteins are not synthesized under the investigated growth conditions. In fact, both proteins are responsible for nonphotochemical quenching and mainly accumulate upon growth under high light of >400 $\mu\text{mol photons m}^{-2}$ (Bonente et al., 2012). Besides photosynthesis-related transcripts, the list with top-most translated transcripts

contained transcripts encoding numerous ribosomal proteins of the 80S complex and elongation factors (Figure 4b). Other abundantly translated proteins are the putative acetate transporter GYF3 (RPKM 4285, rank 14), the Isocitrate lyase (ICL1, RPKM 3918, rank 16), a key enzyme of the glyoxylate cycle (Plancke et al., 2014), the thiazole biosynthesis enzyme (THI4, RPKM 3402, rank 22) (Moulin et al., 2013) or the Methionine Adenosyl-transferase (METM1, RPKM 2438, rank 51), confirming their central role in *Chlamydomonas* cells. Plotting of transcripts with the highest TE values showed an over-representation of histone transcripts, suggesting their robust translation in *Chlamydomonas* cells under logarithmic growth (Figure S4c).

Exploiting individual ribosome profiles to mine for putative regulatory principles

Local elongation rates are known to vary considerably during translation along a transcript, depending on tRNA availability, mRNA secondary structure, the amino acid sequence of the resulting polypeptide and co-translational folding (Gloge et al., 2014; Kramer et al., 2019). With the robust gene body coverage over *Chlamydomonas* transcripts (Figure 1d), we aimed to learn more from the dynamics of translation elongation, derived from the RPF distribution over individual coding sequences (ribosome profiles). Ribosome profiles were highly comparable between all tested RNase I treatment conditions, as exemplified by median *Pearson* correlations between datasets ii. and iii. of ~0.62 for nuclear-encoded transcripts and >0.85 for chloroplast-encoded transcripts. Although RNA-seq read distributions over individual CDS were even more comparable (median of *r*-value 0.87 for nucleus-encoded and 0.99 for chloroplast-encoded transcripts) among the RNA-seq datasets, they showed no correlation (median *r*-value close to zero) to the respective ribosome profiles (Figure 5a; Figure S5). Certainly, methodologically driven differences between Ribo-seq and RNA-seq cannot be fully excluded, but it can be assumed that the ribosome profiles reflect transcript-specific translation dynamics and are likely not the result of sequencing biases.

The correlation coefficients between two experiments for the ribosome profiles of individual transcripts (both *Pearson* and *Spearman*) stably range between 0.7 and 1 for the majority of ribosome profiles that have an average coverage >100 reads per nucleotide, suggesting that low correlation coefficients are the consequence of sequencing depth limitations and thus limited observability of lowly translated transcripts rather than the true absence of a characteristic translation dynamic of these transcripts (Figure S6). Importantly, more than 10 000 transcripts have *r*-values >0.7 (Figure 5b), again showing the high reproducibility between the RNase I digest conditions for transcripts with moderate and high RPF coverage (Figure S6). In fact,

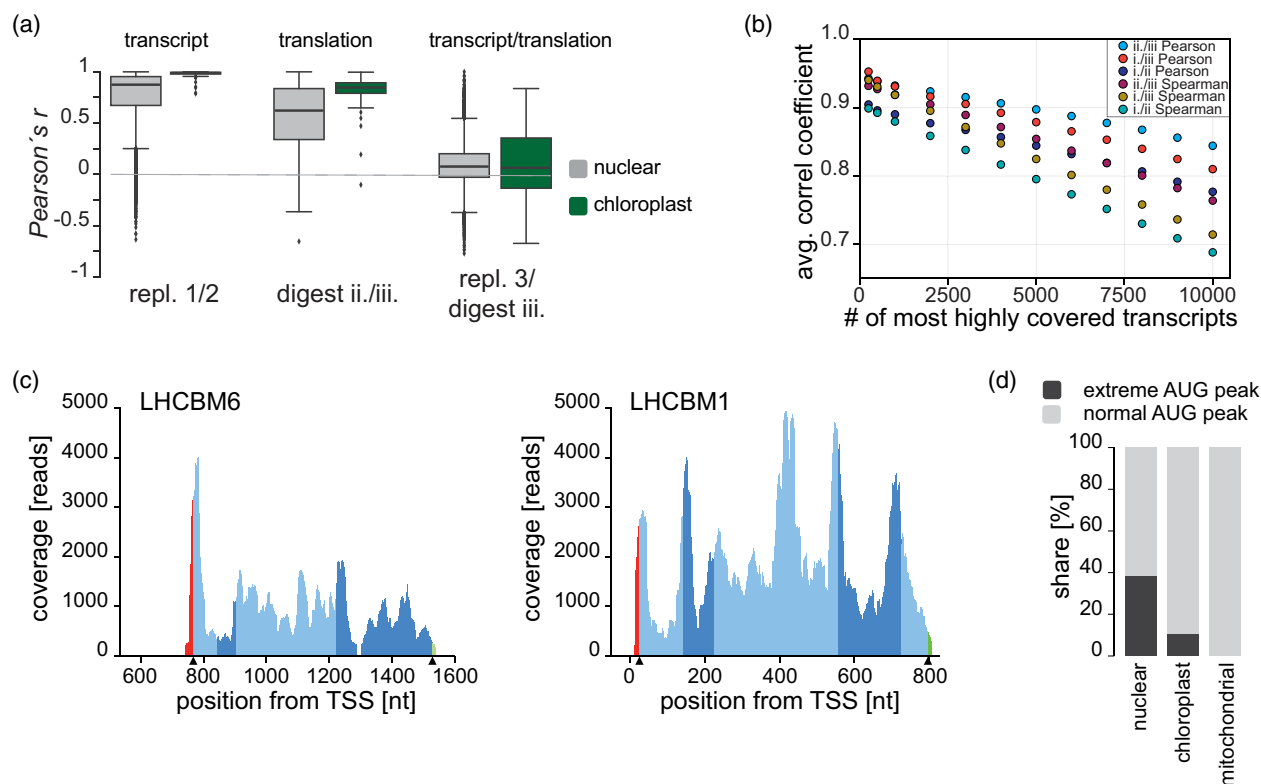


Figure 5. Putative translational regulation of individual transcripts.

(a) Box plots representing the distribution of Pearson's r correlation coefficients comparing the RNA gene body coverage and ribosome profiles of all chloroplast- and nucleus-encoded transcripts between replicates (R1 and R2 for RNA) and between two types of RNase I treatments (ii. and iii. for translation). (b) Average *Pearson* and *Spearman* correlation coefficients depend on the number n of considered transcripts. Transcripts were ordered by descending RPF count per transcript. The n most highly covered transcripts were selected, and correlation coefficients were computed for each transcript in this subset. The average was plotted. For correlation coefficients of individual transcripts see Figure S6a. (c) Ribosome profiles of LHCBM6 serves as an exemplary profile with an extreme initiation peak (before the 800 nt mark), while LHCBM1 is an example of a profile with an even RPF distribution. Red parts of the profile represent coverage within the 5'-UTRs, green parts represent 3'-UTRs and alternating blue shades represent exons within the CDS. Black arrows on the x-axis indicate start and stop codons. TSS = transcription start site. (d) Stacked column chart representing the fraction of ribosome profiles containing extreme initiation peaks (defined as transcripts with an average coverage in the first 7 codons exceeding by at least four times the average coverage of the whole CDS). Nucleus-, chloroplast-, and mitochondria-encoded transcripts are plotted separately containing only transcripts with a gene body coverage of >70% in the Ribo-seq. Transcripts with extreme initiation peak are 3346 out of 8757 for nuclear, 6 out of 66 for chloroplasts and 0 for mitochondria. Further information on the occurrence of abnormal initiation peaks is provided in Figure S7.

the r -value distribution shows that these direct comparisons of read coverage per transcript might be well suited for selecting good cutoff ranges for a deeper analysis of ribosome profiles.

Ribosome profiles of the highly translated *LHC* transcripts showed some remarkable differences between the individual *LHC* species. Efficiently controlling the expression of *LHC* genes is essential for regulating the antenna size of the photosynthesis machinery in order to capture photons under low irradiance conditions (increased expression) or avoid over-excitation and thus photodamage under high irradiance conditions (reduced expression) (Dall'Osto et al., 2015). Of the transcripts encoding the 21 different *LHC* members serving Photosystem I or II (LHCA and LHCB, respectively), 20 were found among the top 100 highest translated transcripts (not considering splice variants). Only the transcript encoding LHCB7, a rather newly

discovered protein containing the unusual number of four transmembrane domains (Klimmek et al., 2006) had low RPKM values of ~15 (Figure S4b). The variations in *LHC* transcript levels result in considerably different TE values (Figure S4b), suggesting the translational regulation may be achieved by diverse strategies for expression of the *LHC* species. In fact, the ribosome profiles vary profoundly among transcripts that encode the different LHCS. For example, *LHCBM6* showed an exceptionally strong peak of RPFs covering the first 7 codons, which exceeds the average RPF density over the remaining CDS by more than 4-fold (Figure 5c). While 'initiation peaks' are commonly accumulated over the first few codons of a CDS, their amplitudes are usually within the dynamic range of the remaining CDS. Such high peaks, in contrast, could point to a regulatory mechanism stalling ribosomes during or shortly after initiation. Indeed, *LHCBM6* was shown to be

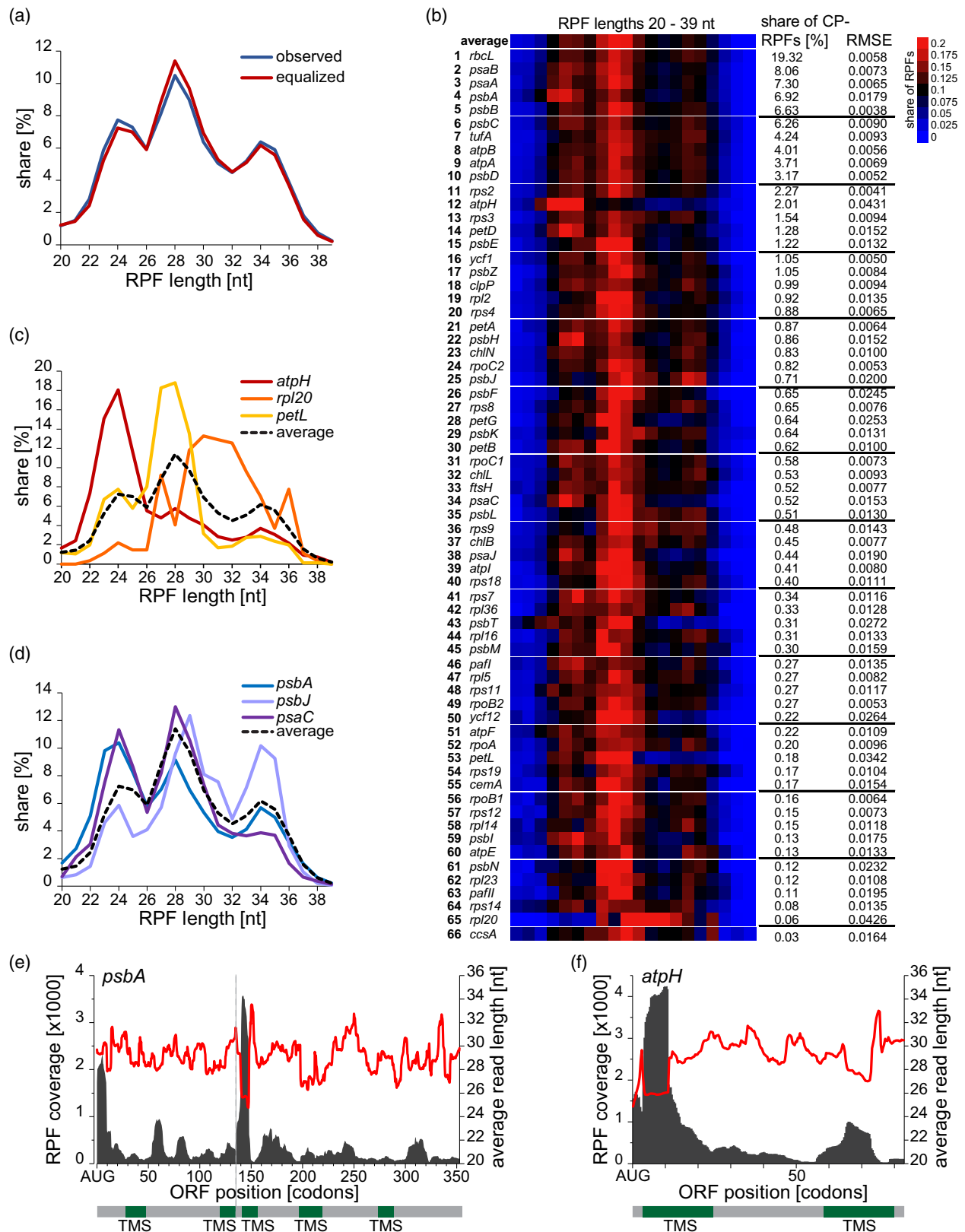


Figure 6. Chloroplast ribosomes generate characteristic ribosome-protected fragments.

(a) Size distribution of RPFs that map the CDS of chloroplast-encoded transcripts for nuclease treatment conditions iii. (observed). Equalized size distribution shows the normalized size distribution for each chloroplast transcript, neglecting possible weights from the most abundant transcripts.

(b) Heatmap presenting the proportion of the different RPF sizes for each chloroplast transcript. Transcripts are ranked based on their overall proportion of RPFs relative to all measured chloroplast RPFs. RMSE (Root Mean Square Error) refers to the deviation of the respective size distribution from the average size distribution (top row). RMSE is defined as the square root of the mean of squared differences between each transcript's RPF length distribution and the average RPF length distribution.

(c) Size distribution of chloroplast-encoded transcripts with a high degree of deviation relative to the average size distribution.

(d) Additional size distribution curves of subunits in the photosynthesis complex.

(e) and (f) Ribosome profile of *psbA* and *atpH*, respectively. Red curve shows the RPF sizes, accumulating within the respective CDS section, black curves indicate RPF coverage per nt. Positions of transmembrane segments (TMS) are given below the graph.

translationally repressed under moderate light conditions by a co-translational regulator called NAB1 (Nucleic Acid Binding protein). NAB1 itself is tuned by a regulatory circuit that senses carbon dioxide supply levels in *Chlamydomonas*, for precise adjustments of the antenna composition of photosystem II (Blifernez-Klassen et al., 2021; Mussgnug et al., 2005). Similarly, *LHCBM4* is recognized by NAB1, also presenting a strong RPF peak over the initiation codon, whereas *LHCBM1* (not a target of NAB1) has no over-proportional initiation peak (Figure 5c), despite sharing a 75% amino acid sequence identity with *LHCBM6*. Translational regulation during *LHC* expression is not unique to *Chlamydomonas*, as *Lhcb4.2*, *Lhcb4.3*, and *Lhcb6* are also subject to post-transcriptional regulation in *Arabidopsis* (Floris et al., 2013). We searched all transcripts that had a gene body coverage of at least 70% (8757 nucleus-encoded, 66 chloroplast-encoded, and 6 mitochondria-encoded) for the occurrence of comparable initiation peaks (average coverage of the first seven codons at least four times higher than the average coverage of the remaining CDS). Around 38% of selected nucleus-encoded transcripts had comparable RPF accumulation peaks near the initiation codon, possibly hinting at more co-translational regulation in the cytosol than previously anticipated. In contrast, only 10% (six transcripts) of the chloroplast-encoded genes showed such peaks (Figure 5d). However, it cannot be excluded that these peaks present artifacts of CHX treatment of the *Chlamydomonas* cultures during harvest. Upon establishment of the Ribo-seq protocol, several reports demonstrated that pre-incubation of yeast cultures with CHX may create a bias by inducing increased RPF accumulation at initiation codons, which could be the consequence of stalled elongation but unaffected initiation through CHX treatment (Gerashchenko & Gladyshev, 2014; Ingolia et al., 2009; Sharma et al., 2021). However, a recent direct comparison of CHX-treated and untreated conditions demonstrated that CHX causes no initiation peak bias in human cells, and that even the strong accumulation in yeast might derive from slightly altered protocols between laboratories (Sharma et al., 2021). For *Chlamydomonas*, CHX biases should result in general initiation peaks, but those peaks are only found in a fraction of

transcripts, independent of their translation activity (e.g., *LHCs*). Thus, varying accumulation of RPFs around the initiation codon may be an interesting indicator for altered initiation regulation between transcripts in the cytosol.

We further assayed whether the RPFs of various sizes, ranging from 24 to 34 nt, are randomly distributed across all chloroplast CDS or if a respective RPF size is transcript-specific. In general, the dominant RPF size is 28 nt long and the trimodal size distribution appears to be widely present for the translation of most chloroplast transcripts, even if the over-representation of the highly translated transcripts (i.e., *rbcl*, *psbA*) is normalized out (Figure 6a). While present for most chloroplast transcripts, accumulation of the different RPF sizes varies between transcripts with some over-representation of smaller or larger RPFs, respectively. For example, *psbA*, *psbH*, *psaC*, *psaJ*, *petD*, and *atpH* have an additional strong peak at 24 nt. In contrast, *rpl20* and *psbJ* have a shift toward a dominant population of RPFs with a size >30 nt (Figure 6b–d). Strikingly, the small RPFs within *psbA* and *atpH* show an interesting pattern with increased accumulation in the regions of highest RPF coverage. For *psbA*, the 24 nt RPFs are most prominent in a section between codon 135 and 150, which is just downstream of the transcript encoding the second transmembrane segment (TMS, Figure 6e). It appears as if elongation severely slows down within this region, favoring the extraction of small RPFs, possibly resulting from ribosomes in an 'open-ratchet' formation. Alternatively, the smaller reads may result from local binding of a regulatory factor to the transcript, although the small RPFs have similar features as the longer RPFs. Interestingly, we just recently reported that deletion of the One-Helix-Protein 2, OHP2, locally affects *psbA* translation between initiation and the coding sequence of the 3rd TMS. However, ribosome occupancy downstream of the 3rd TMS appeared normal in the mutant. In addition, the absence of OHP2 results in a rapid degradation of nascent PsbA/D1 protein (Wang et al., 2023). Thus, the observed local translation slowdown might be an important stage for PsbA/D1 biogenesis and cofactor integration and OHP2 might be the driver of this regulation. Similarly, RPF accumulation peaks over the CDS of the first TMS of *atpH*, mostly deriving

from fragments with a size of 24–25 nt, again suggesting that translation elongation is reduced here (Figure 6e). It will be interesting to uncover in future experiments if the local translation slowdown is accompanied or caused by binding of regulatory factors.

Suitability of ribosome profiles for improving *Chlamydomonas* genome annotation

Lastly, we aimed to make our ribosome profiling data available to a broad scientific community to easily examine translation output and ribosome dwell times for transcripts of interest. Generally, Ribo-seq data requires advanced bioinformatic processing and publicly accessible ribosome profiles are rare. Therefore, we have linked our Ribo-seq data to the latest *Chlamydomonas* genome CC-4532 v6.1 (<https://phytozome-next.jgi.doe.gov>) (Craig et al., 2022), which covers all translated transcripts. This will contribute to validating and improving existing gene models, as well as aiding in the interpretation of predicted transcript variants and correcting possible errors. A brief introduction, how these profiles can be accessed is given in Figure S8. The *Chlamydomonas* genome version 6.1 corrected for missing genes that were absent in previous versions and significantly expanded the number of alternatively spliced transcripts (Craig et al., 2022). In previous genome versions, several genes were erroneously annotated, disregarding upstream start codons that could possibly be the genuine initiation site of translation (Craig et al., 2022; Cross, 2015). In the new version, the first in-frame AUG codon is typically recognized as the initiation codon, in accordance with the scanning model of initiating ribosomes (Hinnebusch, 2011). Systematic comparison of the reference start site with the onset of RPF coverage confirmed that start codon assignment has been clearly improved between genome versions 5.6 and 6.1. However, we detected gene models in which the first AUG does not appear to be the genuine initiation site. For instance, transcripts encoding LHCA4 (Cre10.g452050) do not initiate from the first AUG (Figure 7a) and were correctly annotated in the previous genome versions. Another example is PRPL3 (Cre10.g417700) (Figure 7b). Additionally, *PDI2* (Cre01.g033550) is an example of how ribosome profiles can assist in improving genome annotation and unveiling the expression of the primary transcript in the presence of multiple splice variants (Figure 7c). For *PDI2*, the ribosome profile starts with a prominent peak upstream of the annotated start codon located within the annotated 5'-UTR while the annotated start codon is poorly covered compared with the remaining CDS. Here, initiation from the non-cognate UUG could be an explanation (Cao & Slavoff, 2020), which is located in-frame with the annotated coding sequence at two positions within the area covered by the peak. Alternatively, these RPFs could point to the presence of upstream Open

Reading Frames for *PDI2*. A detailed systematic screen for frequencies and statistical significance of the usage of non-canonical initiation sites within *Chlamydomonas* cells would be highly fascinating and has been conducted for other organisms before (Brar & Weissman, 2015). However, due to the required bioinformatic tools and the need for independent validation we were not able to perform this screen at this point. Yet, our available data will be suitable to individually browse for similar non-canonical initiation events. *PDI2* also represents a good example for determining the dominantly translated transcript variant. Six splice variants exist for *PDI2*. Transcripts 3 and 6 are shorter than 1,2,4 and 5, but RPF coverage is homogeneous over the long transcripts, suggesting that 3 and 6 are not the dominant versions. Transcript 1 possesses a short sequence stretch between exon 3 and 4 (379 nt downstream of the transcription start site), which is spliced out in transcript versions 2–4 (Figure 7c). For this stretch, we did not detect RPFs. Therefore, translation of this longer transcript 1 is unlikely under the growth conditions examined. Moreover, transcripts 2 and 4 have a non-homogeneous ribosome profile at the end of exon 3. Here, RPF abundance drops exactly at the exon border suggesting that our sequenced RPFs result from a different transcript variant (transcript 5, see 'smooth transition' in Figure 7c), which has an insertion at this section. Thus, transcript 5 has the most consistent profile and might hence be the prevalently expressed variant in *Chlamydomonas* cells grown under standard laboratory conditions. In future, intelligent modeling and peak deconvolution algorithms could be potentially applied to Ribo-seq data to estimate the stoichiometric ratio of different splice variants' translation output.

Concluding remarks

Taken together, ribosome profiling is a fascinating tool for advancing our understanding of post-transcriptional regulation of gene expression. Our Ribo-seq approach has the resolution and depth to deeply study translation of the nuclear and chloroplast transcripts in *Chlamydomonas* cells. Our data suggest transcript-specific regulatory phases in elongation, which seem unique for chloroplast translation. Furthermore, a highly comparable Ribo-seq protocol was established for *Arabidopsis* and tobacco by our team, which will help to directly compare translation between different land plants and *Chlamydomonas* (Ting et al., 2023). *Chlamydomonas* is a well-suited organism for understanding system-level changes in gene expression throughout the cell cycle or following environmental changes. Ribosome profiles can be specifically helpful for a first understanding of the many transcripts that have not been experimentally investigated before.

Visualizing ribosome profiles can help to uncover the dominant transcript variants that are translated under

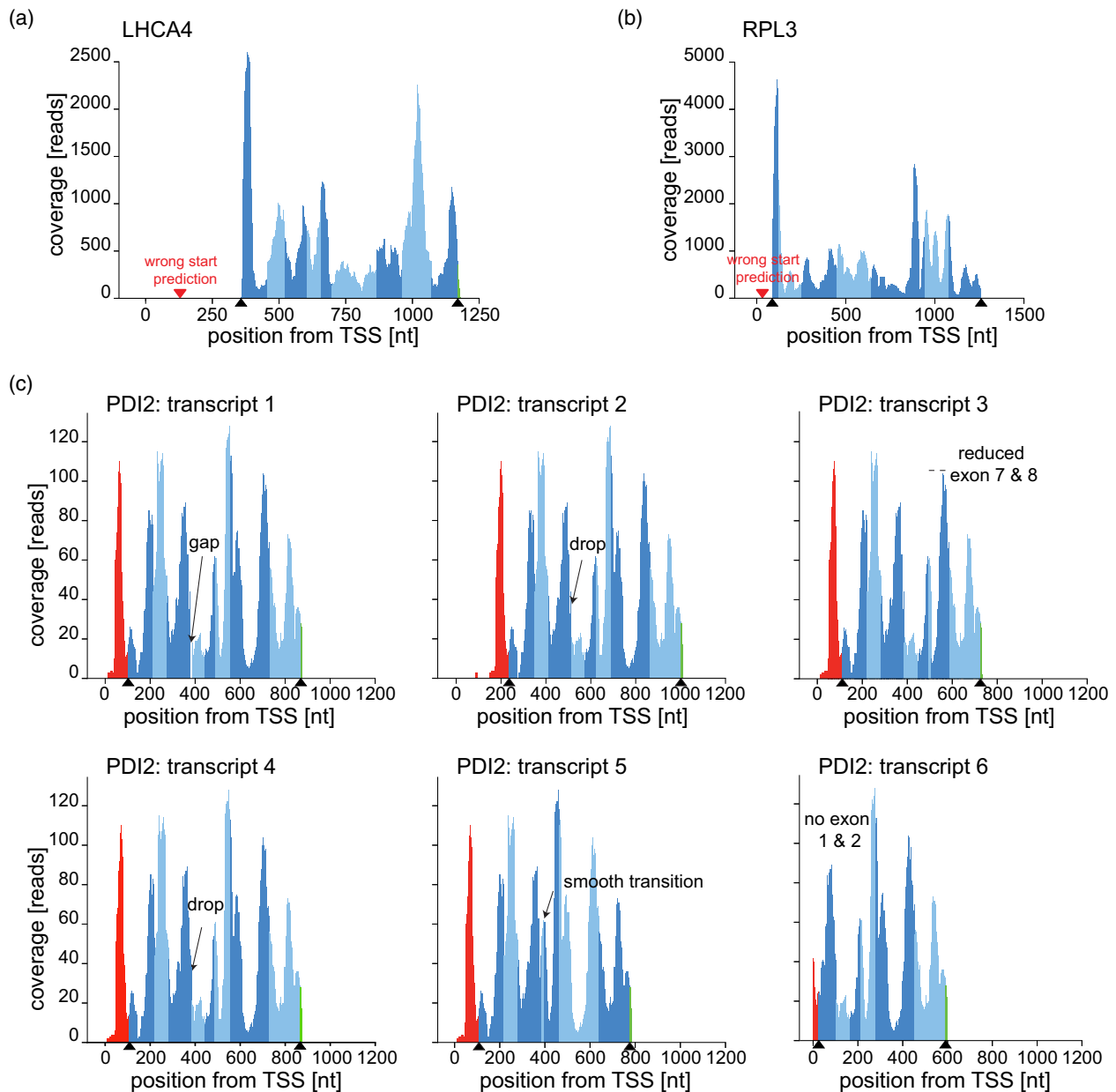


Figure 7. Utilizing ribosome profiles for coding sequence annotation.

Exemplary ribosome profiles for determining initiation sites and alternatively spliced transcripts. Red parts of the profiles represent coverage within the 5'-UTRs, green parts represent 3'-UTRs and alternating blue shades represent exons within the CDS. Black arrows on the x-axis indicate start and stop codons. (a) and (b) ribosome profiles of LHCA4 and RPL3 as an example of incorrect start codon prediction, respectively. The red arrow indicates the annotated start codon while the black arrows indicate true start and stop codons.

(c) Ribosome profiles of PDI2 (Cre01.g033550) transcript variants demonstrating how ribosome profiles can be utilized to identify putative non-cognate start codon initiation and estimate the dominance of transcript variants. Note that AUG was labeled as 'true' initiation site, however, an upstream non-cognate start codon might be the correct initiation site.

standard laboratory conditions. Certainly, additional experimental approaches are required to understand unexpected features such as translation initiation from non-cognate start sites, RPFs along upstream open reading frames, mechanism of translational regulation, and spatial organization of protein synthesis.

EXPERIMENTAL PROCEDURES

Cell growth and harvest

For each experiment, *Chlamydomonas reinhardtii* CC-1690 cells were grown mixotrophically in Tris-Acetate Phosphate (TAP) medium (Kropat et al., 2011) to mid-log phase (4 to 5×10^6 cells

per ml) under constant light of $80 \text{ mmol m}^{-2} \text{ sec}^{-1}$ (MASTER LED-tube HF 1200 mm UO 16 W830 T8 and 16 W840 T8, Philips) and 25°C on a rotary shaker at 130 rpm. Immediately before harvest, $100 \mu\text{g/ml}$ chloramphenicol (CAP) and cycloheximide (CHX) was added and cells were rapidly chilled by pouring through a funnel containing -80°C cold silicon ice cubes to quickly reach sample temperatures of 4°C . Subsequently, cells were pelleted by 2 min centrifugation at 4000 g (Beckman Coulter, Avanti) and 4°C , washed in ice-cold TKM+ buffer (50 mM acetate-buffered Tris pH 8.0, 150 mM KCl, 10 mM MgCl_2 , $100 \mu\text{g/ml}$ CAP, $100 \mu\text{g/ml}$ CHX) and pelleted again using the same centrifugation conditions. Afterward, the pellet was resuspended in freezing buffer (TKM+ supplemented with 100 mM phenyl methyl sulfonyl fluoride (PMSF) and 16% of $25\times$ Roche protease inhibitor cocktail) by slow pipetting in $1/2000$ of the initial culture volume and flash frozen by dripping the cell suspension into liquid nitrogen. Frozen cells were stored at -80°C until further use.

Ribosome-protected fragment generation and RNA sampling

Frozen cells were ground using a bead mill (Retch, MM400) with nitrogen-cooled steel containers and 15 mm beads for two times 2 min at 27 Hz frequency and cooling in liquid nitrogen between rounds. The cell powder was mixed with equal volumes of $2\times$ concentrated lysis buffer ($2\times$ concentrated TKM+ buffer supplemented with 2 mM dithiothreitol (DTT), 2% Triton X-100 and 20% sucrose) and incubated for 5 min at 4°C before pelleting cell debris at 8000 g (Beckman Coulter, Avanti), 4°C for 10 min. For nucleolytic digest conditions i. and ii., lysate was layered on a 2.5 ml 64% sucrose cushion prepared in TKM+ buffer supplemented with 1 mM DTT and ultracentrifuged for 3 h at 33000 g , Beckman Coulter, Brea, CA, USA) and 4°C in a Ti70 rotor. The resulting polysome pellet was briefly rinsed with $500 \mu\text{l}$ ice-cold diethyl pyro carbonate (DEPC)-treated water and incubated overnight on ice to fully resuspended in $100 \mu\text{l}$ ice-cold ribosome buffer ($2\times$ concentrated TKM+ buffer diluted with $25\times$ concentrated Roche protease inhibitor cocktail to a total amount of 16% , 1 mM PMSF and 0.1% Triton X-100). Remaining insoluble debris were removed by centrifugation (Hermle 2216 MK) for 1 min at 1500 g before addition of 1 U of Ambion™ RNase I (Thermo Fisher, Waltham, MA, USA) per μg RNA and 0.134 U TURBO DNase (Thermo Fisher, Waltham, MA, USA). Nucleolytic digest was performed for 1 h at 4°C (condition i.) or 23°C (condition ii.) and stopped by addition of 0.4 U per applied unit of RNase I of SUPERase-In RNase Inhibitor (Thermo Fisher, Waltham, MA, USA). For condition iii., 1 U of Ambion™ RNase I (Thermo Fisher, Waltham, MA, USA) per μg RNA and $0.134 \mu\text{l}$ per μl lysate of TURBO DNase (Thermo Fisher, Waltham, MA, USA) was directly added to the lysate and incubated for 1 h at 4°C . The digest was stopped by addition of 0.4 U of SUPERase-In RNase Inhibitor (Thermo Fisher, Waltham, MA, USA) per applied unit of RNase I. The lysate was centrifuged again for 10 min at 8000 g (Hermle 2216 MK) and 4°C to remove any cell debris that may have precipitated during the nuclease digest. For all conditions, monosomes were pelleted through a $750 \mu\text{l}$ 30% sucrose cushion prepared in TKM+ buffer supplemented with 1 mM DTT for 30 min in a S150AT rotor at 72000 rpm and 4°C (Sorvall Discovery M150 SE). The resulting ribosome pellets were resuspended in $100 \mu\text{l}$ of ice-cold ribosome buffer, then supplemented with 15 mM ethylene diamine tetra-acetic acid (EDTA), pH 8.0 and instantly mixed with $750 \mu\text{l}$ TRIzol reagent (Invitrogen, Thermo Fisher, Waltham, MA, USA) and incubated at room temperature for 5 min. The solution was mixed with $150 \mu\text{l}$ chloroform and incubated for 2 min at room temperature before 15 min

centrifugation at 20000 g and 4°C . The supernatant was mixed with $3 \mu\text{l}$ GlycoBlue co-precipitant (Thermo Fisher, Waltham, MA, USA), $1:10$ volume of 3 M sodium acetate, pH 5.5, 1 volume isopropanol and incubated overnight at -20°C . The precipitated RNA was pelleted by 1 h centrifugation at 20000 g and 4°C (Hermle 2216 MK), briefly washed with ice-cold 80% ethanol and centrifuged again for 30 min at the same parameters. The resulting RNA pellet was briefly dried and dissolved in $11 \mu\text{l}$ of nuclease-free water.

For RNA-seq samples, 10 ml of culture were harvested by 2 min centrifugation at 4000 g and room temperature (before harvesting the culture for RPF isolation), shock-freezing in liquid nitrogen and storage at -80°C . Samples were lysed in prewarmed (50°C) lysis buffer (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 20 mM EDTA pH 8.0 and 2% SDS) and incubated for 2 min at 50°C . Immediately after incubation, $500 \mu\text{l}$ of TRIzol (Thermo Fisher, Waltham, MA, USA) reagent were added and the mixture was incubated for 5 min at room temperature. The suspension was mixed with $200 \mu\text{l}$ chloroform and incubated for another 5 min at room temperature before 3 min centrifugation at 12000 g and room temperature (Hermle 2216 MK). The nucleic acid-containing phase was transferred to a fresh tube and mixed with 1.5 volumes of ethanol. RNA was purified by the Macherey Nagel NucleoSpin RNA Plant kit.

Ribosome-protected fragment extraction and rRNA depletion

$50 \mu\text{g}$ RNA from the ribosomal extractions were diluted to a volume of $20 \mu\text{l}$ with nuclease-free water, mixed with $2\times$ formamide RNA loading buffer (90% deionized formamide, 20 mM Tris-HCl, pH 7.5, 0.5 M EDTA, 0.04% bromophenol blue in ethanol) and denatured for 90 sec at 80°C . The samples were instantly put on ice and then loaded onto 12% urea-TBE-gels (90 mM Tris, 9 mM boric acid, 2 mM EDTA, 12% $19:1$ 40% acrylamide-bisacrylamide mix, 8 M urea). Gels were run at 200 V for approximately 1 h and incubated for 5 min with SYBR Gold nucleic acid stain (Thermo Fisher, Waltham, MA, USA), diluted $1:10000$ in TBE buffer (90 mM Tris, 9 mM boric acid, 2 mM EDTA). RPF were excised from the gels in a size range from 20 to 40 nt according to a pre-stained DynaMarker for small RNA (BioDynamics). Gel pieces were crushed and incubated overnight with $400 \mu\text{l}$ RPF elution buffer (300 mM sodium acetate, pH 5.5, 0.25% sodium dodecyl sulfate, and 1 mM EDTA, pH 8.0) at room temperature. The eluate was collected, mixed with $3 \mu\text{l}$ GlycoBlue co-precipitant, 1 volume of isopropanol and incubated overnight at -20°C . The precipitated RNA was pelleted by 1 h centrifugation at 20000 g (Hermle 2216 MK) and 4°C , briefly washed with ice-cold 70% ethanol and centrifuged again for 30 min at the same parameters. The resulting RNA pellet was briefly dried by air and dissolved in $20 \mu\text{l}$ of nuclease-free water. Our pilot Ribo-seq runs had contaminating rRNA levels between 80 and 90%, mainly deriving from the cytosolic 5.8S rRNA and the chloroplast 23S rRNA. To further reduce the amount of contaminating rRNA species, we generated anti-sense biotinylated oligos that allow for the specific depletion of these species. For digest condition iii., we additionally tested RNase H treatment, which cleaves hybrids of the anti-sense oligos and the target rRNA contaminants. Other than the standard procedure, we performed rRNA depletion prior to dual linker ligation and were thus able to provide a higher input of RPFs in the sample, relative to contaminating rRNAs, which then allows for lowering the number of PCR amplification steps during final sequencing library generation. More specifically, abundant rRNA fragments were depleted by incubating $100\text{--}200 \text{ ng}$ of ribosome RPFs with $1.19 \mu\text{l}$ of $100 \mu\text{M}$

biotinylated rRNA-oligo depletion mix (consisting of the oligonucleotides: AATATGCGTTCAAAGATTTCGATGATTCACG, TAGCTCTA GAATTACTACGGTTATCCGAGTA, TACCCGACGCTGAGGCAGACA TGCTCTTGG, GATTCGTGAAGTTATCATGATTCACCGCA, ACGGG ATGAATCTCAGTGGATCGTAGCA, CGATCTAGCCGTCTTAGAGCT AGAAGCAGG) with 4 µl formamide, 1 µl 20× concentrated hybridization buffer (300 mM sodium citrate, 3 M NaCl), 2 µl 0.5 M EDTA, pH 8.0 and in a total volume of 20 µl. The mixture was hybridized in a thermocycler (FastGene UltraCycler, Nippon Genetics, Dueren, Germany) by heating up to 80°C for 5 min and subsequent cooling in 5°C steps for 2 min each until 35°C was reached in the device. Afterward, 60 µl of hybridization buffer, supplemented with 20% formamide was added to the tube and the sample was depleted twice for 15 min using 15 µl of Dynabeads MyOne Strep-tavidin T1 magnetic beads (Invitrogen, Thermo Fisher, Waltham, MA, USA) freshly washed and prepared according to the manufacturer's manual. The supernatant was mixed with 3 µl GlycoBlue co-precipitant, 2.5 volumes of ethanol and incubated overnight at −20°C. RPFs were pelleted by 1 h centrifugation at 20,000 *g* and 4°C, briefly washed with ice-cold 70% ethanol and centrifuged again for 10 min at the same parameters. The resulting pellet was briefly dried by air and dissolved in 41 µl of nuclease-free water. Excess oligos were removed by 30 min DNase digest at 37°C upon addition of 5 µl 10× TURBO DNase buffer, 2 µl TURBO DNase and 2 µl SUPERase-In RNase Inhibitor (both Invitrogen, Thermo Fisher, Waltham, MA, USA). After digest, RPFs were mixed with 150 µl of 100 mM NaCl and purified again by phenol-chloroform precipitation and final precipitation overnight at −20°C using 2.5 volumes of ethanol and 2 µl GlycoBlue co-precipitant. RPFs were pelleted by 1 h centrifugation at 20,000 *g* and 4°C, briefly washed with ice-cold 70% ethanol and centrifuged again for 10 min at the same parameters. The resulting pellet was briefly dried by air and dissolved in 36.5 µl of nuclease-free water.

For sequencing library preparation, ribosome RPFs were denatured at 65°C for 5 min in a thermocycler (FastGene UltraCycler, Nippon Genetics, Dueren, Germany) and immediately transferred to ice. The sample was then mixed with 5 µl 10× T4-polynucleotide kinase buffer A, 2.5 µl T4-polynucleotide kinase (Thermo Fisher, Waltham, MA, USA), 2 µl SUPERase-In RNase Inhibitor and incubated for 10 min at 37°C. Subsequently, 5 µl of 10 mM ATP was added, and the sample was incubated for additional 30 min at 37°C. After phosphorylation, the RPFs were mixed with 150 µl 100 mM NaCl and purified again by phenol-chloroform precipitation and final precipitation overnight at −20°C using 2.5 volumes of ethanol and 2 µl GlycoBlue co-precipitant. RPFs were pelleted by 1 h centrifugation at 20,000 *g* and 4°C, briefly washed with ice-cold 70% ethanol and centrifuged again for 10 min at the same parameters. The resulting pellet was briefly dried and dissolved in 11.5 µl of nuclease-free water. The whole sample was used as input for the NEXTFLEX Small RNA-seq Kit v3 (Perkin Elmer, Waltham, MA, USA) to generate libraries according to the kit's manual. For RNA-seq, the purified RNA was later used as input for the Zymo-Seq RiboFree Total RNA library kit for library preparation according to the manufacturer's manual.

Processing of raw sequencing data

Demultiplexed FASTQ files were derived from an Illumina Next-Seq 550 system with >52 M reads for i., >53 M reads for ii., and > 49 M reads for iii., respectively, and 21–23 M reads for the RNA-seq sample at a read length of 75 nt. Samples were processed through a custom-made pipeline in form of a Linux shell script calling several bioinformatic tools and custom Python scripts. First operation of the pipeline was removing 3' adapter

sequences using cutadapt for RNA-seq and Ribo-seq, respectively (Martin, 2011) with parameters [—minimum length = 9, —a=‘TGGAATTCTCGGGTGCCAAGG’]. After removal of 3' adapter sequences, the 3' and 5' unique molecular identifier tags (UMI tags) were removed from each read's sequence and shifted to the rear of the header row by UMI tools (Smith et al., 2017) with parameters [extract – bc-pattern = NNNN] for 5' UMI tags and [extract – bc-pattern = NNNN – 3prime] for 3' UMI tags. Remaining reads were filtered for sequence length using cut adapt again with parameters [—minimum length = 20 –maximum length = 39] to remove any read that is unlikely to represent a real RPF. After size filtering, a custom-made Python script corrected the UMI tags that were previously shifted to the read headers in the form [header line_NNNN_NNNN] to [header line_NNNNNNNNN], allowing UMI tools in a later step to deduplicate the dataset based on both UMI tags. In a first mapping step, reads were aligned to a set of *C. reinhardtii* non-coding RNA FASTA sequences provided by ENSEMBL plants together with the genome versions 5.5 and v6.1 (ncRNAs are not annotated in version 6.1) to remove contaminating RNA species using STAR with parameters [—runMode alignReads – outSAMtype BAM SortedByCoordinate – outFilter-MultimapNmax 20 – outFilterMismatchNmax 3 – outReadsUnmapped Fastx – alignIntronMax 7000 – twopassMode Basic]. Reads that did not align to the ncRNA set were copied to a new FASTQ file that was mapped against the *C. reinhardtii* genome version 6.1 using STAR (Dobin et al., 2013) with parameters [—runMode alignReads – outSAMtype BAM SortedByCoordinate – outFilter-MismatchNoverLmax 0.1 – outReadsUnmapped Fastx – alignIntronMax 8000 – outSAMmultNmax 1 – outMultimapperOrder Random]. After genome mapping, the resulting BAM file was deduplicated using UMI tools' dedup function and afterward indexed using samtools' index function (Li et al., 2009).

Data processing and statistical analysis

All analyses on the NGS datasets including calculations of gene body coverage, periodicity, P-site offsets, organellar read length distributions, ribosome profiles, and gene expression metrics were performed using an own, object-oriented Python module 'Ribo-Seq functions' (short: RSf, Supporting Methods S1) that makes extensive use of the packages pysam 3.8.13/HTSLib (<https://github.com/pysam-developers/pysam>), pandas 1.4.3 (McKinney, 2010), NumPy 1.20.3 (Harris et al., 2020), matplotlib 3.3.4 (Hunter, 2007), seaborn 0.11.2 (Waskom, 2021) and the Python 3.8.13 standard library. We developed RSf for interactive analysis of Ribo-seq data in Jupyter Notebooks to enable the user to tailor an analysis to the specific needs of the experiment. The module works by importing Ribo-seq data as BAM file and a genome database that can be generated from a genomic annotation as GFF3-file. The module provides high-level functions for genome-wide calculation of all Ribo-seq-specific metrics described below. The module is provided in Supporting Methods S1 (Ribo-Seq_functions.py). For RPKM calculations in Ribo-seq datasets, only reads mapping to the annotated CDS of a transcript were considered. Alternatively, we calculate counts per million values for the interested reader (Data S1). For any calculations on RNA-seq datasets, reads mapping to the whole transcript were considered and the full length of the mature transcript was used for RPKM calculations. Translation efficiency values were simply calculated by dividing an CDSs RPKM value from a Ribo-seq dataset by the respective transcript's RPKM value from the affiliated RNA-seq dataset.

For the 5'-P-site / 3'-A-site offset calculations, all reads fully enclosing either an annotated start codon (5'-P-site offset) or an

annotated stop codon (for 3'-A-site offset) were taken into consideration separately for the nuclear and the chloroplast genome. These reads were sorted according to the length of their alignment to the genome, which was considered as read length to avoid miscalculations due to clipped or masked sequences. For every read length species, the position of the reads 5' end or 3' end relative to the position of the first base of the start codon (for 5'-P-site offsets) or the last base of the stop codon (for 3'-A-site offsets) was determined as the offset. Afterward, the most frequently occurring offset for a read length species was determined to be the true offset of this species and was used in the P-site calculation. To display the results of the offset calculations, the count matrix of both 5'-P-site offsets and 3'-A-site offsets were normalized separately to the total number of counts each and plotted as heatmap.

For calculation of the P-sites first position of reads mapping to nucleus-encoded CDSs, all reads mapping to a CDS were filtered from the dataset and for every read, the 5'-most mapping position of the read's alignment + the 5'-P-site offset determined for the reads alignment length was determined to be the first P-site position in case of transcripts encoded on the + strand. For reads encoded on the minus strand, the first P-site position was determined to be the 3'-most mapping position of the RPF alignment minus the determined 5'-P-site offset. In all cases, splicing was considered and whenever a spliced read was found, the calculation was adjusted according to the number of genomic positions spanned by the splice junction of the read. For reads mapping to chloroplast-encoded transcripts on the + strand, the first P-site position was determined to be the 3'-most mapping position - (3'-A-site offset +5). For reads mapping to chloroplast-encoded transcripts on the - strand, the first P-site position was calculated to be the 5'-most mapping position + (3'-A-site offset +5). The metagene P-site profiles of nucleus-encoded transcripts were calculated considering only the 2000 most expressed transcripts of a dataset, according to their RPKM values. If multiple splice variants of the same gene occurred in the list, only the longest variant (annotated as transcript 1 in the genome) was considered for the calculation to avoid analyzing the same reads multiple times. The profiles were constructed by calculating the first position of the P-site for all reads mapping to the first 30 and the last 30 codons of considered transcript CDSs. For a ribosome RPF to be in frame with the annotated CDS, the first position of the P-site must be located on the first base of a codon. According to this logic, the calculated P-site positions were categorized into 'in-frame', '+1 shifted', and '+2 shifted' and summed up for all codons of considered transcript CDSs. To calculate the frame preference per codon, the counts in each category in a codon were divided by the sum of all categories for every codon separately. For calculation of the chloroplast metagene P-site profile, all protein-coding chloroplast transcripts were considered.

The extension of gene body coverage versus a respective cut-off was calculated as the portion of an CDS that was covered by the number of reads equal to or greater than the respective cutoff. This procedure was applied to all annotated CDSs using increasing cutoffs from 1 to 10 reads to monitor the stepwise decrease in gene body coverage. For every applied cut-off, the CDSs were categorized as being covered over >10% of their length to >90% of their length in 10% steps and CDSs falling into every category were counted and displayed in a heat map to estimate both completeness and strength of CDS coverage in the dataset.

To calculate the read length distribution, all reads mapping to the respective genome were filtered and the length of their alignment to the genome was taken as read length. After counting the number of all possible read lengths for each genome

separately, these numbers were normalized by the total number of reads mapping to the respective genome. For read biotype distribution, the reads were further categorized if they map to an annotated coding sequence, 5'- or 3'-UTR or intergenic region. To handle reads overlapping two regions of different category, for example, 5'-UTR and coding sequence as seen regularly in initiation peaks, the read was assigned the category that had the larger share on the read's alignment. In rare cases, when a read mapped to different categories in two different splice variants of the same transcript, the read was assigned to both categories.

For correlation analysis of ribosome profiles, data were extracted by accessing the `.cds_coverage` attribute of the respective transcripts for direct comparison and subsequent conversion of outcomes into dictionaries of the pandas Series objects. Pearson's correlation coefficient between both Series was calculated using the Python package SciPy 1.11.2 (`scipy.stats.pearsonr(x = Series 1, y = Series 2)`) (Virtanen et al., 2020). For directly comparing ribosome profiles, Pearson and Spearman correlation coefficients were calculated in R using the `cor()` function. For calculation of the root mean square error between RPF length distributions of chloroplast transcripts and the average chloroplast RPF length distribution, the respective distributions were calculated using the `FPlength_dist()` function of RSf. RSME was then calculated using the Python package scikit-learn 1.3.0 (`sklearn.metrics.mean_squared_error(y_true = Transcripts distribution, y_pred = Average distribution, squared = False)`) (Pedregosa et al., 2011).

ACKNOWLEDGEMENTS

We thank Fabian Ries, Claudia Herkt and Peter Emelin for critical discussion of the data and the manuscript. Martin Simon, Jaro Schmitt, and Günter Kramer for help with next-generation sequencing. We further thank Jeremy Phillips and David Goodstein from the Joint Genome Institute for making our data available to the public on the Phytozome website.

AUTHOR CONTRIBUTIONS

VLG performed ribosome profiling experiments, bioinformatic data analyses, and wrote parts of the manuscript. MKYT and RZ helped with setting up ribosome profiling. NH and SR conducted correlation analyses. FW designed experiments, analyzed data, and wrote the manuscript.

FUNDING INFORMATION

This work was supported by the Deutsche Forschungsgemeinschaft grant 437345987 (WI 3477/3-1, ZO 302/5-1, RU 2266/2-1 to F.W., R.Z. and S.R., respectively) and the Forschungsschwerpunkt BioComp to F.W.

CONFLICT OF INTEREST

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

All gene accession numbers are given in the Data S1 and are in accordance with the latest *Chlamydomonas* genome v6.1 annotation (<https://phytozome-next.jgi.doe.gov>) or the GenBank/EMBL data libraries. The full Ribo-seq and RNA-seq data are accessible via <https://www.ncbi.nlm.nih.gov/biosample/37486409>; <https://www.ncbi.nlm.nih.gov/>

biosample/37486410; <https://www.ncbi.nlm.nih.gov/biosample/37486411>; <https://www.ncbi.nlm.nih.gov/biosample/37486412>; <https://www.ncbi.nlm.nih.gov/biosample/37486413>.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Features of ribosome protected fragment coverage.

Figure S2. RPF length distribution and periodicity.

Figure S3. Chloroplast proteins have overall higher translation of their transcripts.

Figure S4. Transcript accumulation, translation output and translation efficiency of nuclear genes.

Figure S5. Correlation between nuclease digestion conditions, transcript accumulation and translation.

Figure S6. Correlation analysis of RPF count between nuclease digest conditions.

Figure S7. Detection of abnormal initiation peaks in the three genomes of *Chlamydomonas reinhardtii*.

Figure S8. Guideline how to access ribosome profiles in the *Chlamydomonas* genome portal.

Dataset S1. List of accession numbers, gene body coverage, RPKM and CPM of Ribo-seq and RNA-seq experiments.

Methods. Code of the object-oriented Python module “Ribo-Seq functions”.

REFERENCES

- Blifern-Klassen, O., Berger, H., Mittmann, B.G.K., Klassen, V., Schelleter, L., Buchholz, T. et al. (2021) A gene regulatory network for antenna size control in carbon dioxide-deprived *Chlamydomonas reinhardtii* cells. *Plant Cell*, **33**, 1303–1318.
- Bonente, G., Pippa, S., Castellano, S., Bassi, R. & Ballottari, M. (2012) Acclimation of *Chlamydomonas reinhardtii* to different growth irradiances. *The Journal of Biological Chemistry*, **287**, 5833–5847.
- Brar, G.A. & Weissman, J.S. (2015) Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nature Reviews. Molecular Cell Biology*, **16**, 651–664.
- Calviello, L., Mukherjee, N., Wyler, E., Zaubner, H., Hirsekorn, A., Selbach, M. et al. (2016) Detecting actively translated open reading frames in ribosome profiling data. *Nature Methods*, **13**, 165–170.
- Cao, X. & Slavoff, S.A. (2020) Non-AUG start codons: expanding and regulating the small and alternative ORFeome. *Experimental Cell Research*, **391**, 111973.
- Chartron, J.W., Hunt, K.C. & Frydman, J. (2016) Cotranslational signal-independent SRP preloading during membrane targeting. *Nature*, **536**, 224–228.
- Chen, H., Alonso, J.M. & Stepanova, A.N. (2022) A Ribo-Seq method to study genome-wide translational regulation in plants. *Methods in Molecular Biology*, **2494**, 61–98.
- Chiu, C.W., Li, Y.R., Lin, C.Y., Yeh, H.H. & Liu, M.J. (2022) Translation initiation landscape profiling reveals hidden open-reading frames required for the pathogenesis of tomato yellow leaf curl Thailand virus. *Plant Cell*, **34**, 1804–1821.
- Chotewutmontri, P. & Barkan, A. (2016) Dynamics of chloroplast translation during chloroplast differentiation in maize. *PLoS Genetics*, **12**, e1006106.
- Chotewutmontri, P. & Barkan, A. (2018) Multilevel effects of light on ribosome dynamics in chloroplasts program genome-wide and *psbA*-specific changes in translation. *PLoS Genetics*, **14**, e1007555.
- Chotewutmontri, P. & Barkan, A. (2020) Light-induced *psbA* translation in plants is triggered by photosystem II damage via an assembly-linked autoregulatory circuit. *Proceedings of the National Academy of Sciences of the United States of America*, **117**, 21775–21784.
- Chung, B.Y., Deery, M.J., Groen, A.J., Howard, J. & Baulcombe, D.C. (2017) Endogenous miRNA in the green alga *Chlamydomonas* regulates gene expression through CDS-targeting. *Nature Plants*, **3**, 787–794.
- Chung, B.Y., Hardcastle, T.J., Jones, J.D., Irigoyen, N., Firth, A.E., Baulcombe, D.C. et al. (2015) The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for Ribo-seq data analysis. *RNA*, **21**, 1731–1745.
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A. et al. (2016) A survey of best practices for RNA-seq data analysis. *Genome Biology*, **17**, 13.
- Costanzo, M.C., Hogan, J.D., Cusick, M.E., Davis, B.P., Fancher, A.M., Hodges, P.E. et al. (2000) The yeast proteome database (YPD) and *Caenorhabditis elegans* proteome database (WormPD): comprehensive resources for the organization and comparison of model organism protein information. *Nucleic Acids Research*, **28**, 73–76.
- Craig, R.J., Gallaher, S.D., Shu, S., Salome, P., Jenkins, J.W., Blaby-Haas, C.E. et al. (2022) The *Chlamydomonas* genome project, version 6: reference assemblies for mating type *plus* and *minus* strains reveal extensive structural mutation in the laboratory. *Plant Cell*, **35**.
- Cross, F.R. (2015) Tying down loose ends in the *Chlamydomonas* genome: functional significance of abundant upstream open reading frames. *G3 (Bethesda)*, **6**, 435–446.
- Dall’Osto, L., Bressan, M. & Bassi, R. (2015) Biogenesis of light harvesting proteins. *Biochimica et Biophysica Acta*, **1847**, 861–871.
- Dingwall, C., Lomonosoff, G.P. & Laskey, R.A. (1981) High sequence specificity of micrococcal nuclease. *Nucleic Acids Research*, **9**, 2659–2673.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S. et al. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
- Döring, K., Ahmed, N., Riemer, T., Suresh, H.G., Vainshtein, Y., Habich, M. et al. (2017) Profiling Ssb-nascent chain interactions reveals principles of Hsp70-assisted folding. *Cell*, **170**, 298–311.e20.
- Eberhard, S., Drapier, D. & Wollman, F.A. (2002) Searching limiting steps in the expression of chloroplast-encoded proteins: relations between gene copy number, transcription, transcript abundance and translation rate in the chloroplast of *Chlamydomonas reinhardtii*. *The Plant Journal: for Cell And Molecular Biology*, **31**, 149–160.
- Floris, M., Bassi, R., Robaglia, C., Alborese, A. & Lanet, E. (2013) Post-transcriptional control of light-harvesting genes expression under light stress. *Plant Molecular Biology*, **82**, 147–154.
- Forsythe, E.S., Grover, C.E., Miller, E.R., Conover, J.L., Arick, M.A., 2nd, Chavarro, M.C.F. et al. (2022) Organellar transcripts dominate the cellular mRNA pool across plants of varying ploidy levels. *Proceedings of the National Academy of Sciences of the United States of America*, **119**, e2204187119.
- Frank, J. & Agrawal, R.K. (2000) A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature*, **406**, 318–322.
- Fujita, T., Kurihara, Y. & Iwasaki, S. (2019) The plant Translatome surveyed by ribosome profiling. *Plant & Cell Physiology*, **60**, 1917–1926.
- Gallaher, S.D., Fitz-Gibbon, S.T., Strenkert, D., Purvine, S.O., Pellegrini, M. & Merchant, S.S. (2018) High-throughput sequencing of the chloroplast and mitochondrion of *Chlamydomonas reinhardtii* to generate improved de novo assemblies, analyze expression patterns and transcript speciation, and evaluate diversity among laboratory strains and wild isolates. *The Plant Journal*, **93**, 545–565.
- Gao, Y., Thiele, W., Saleh, O., Scossa, F., Arabi, F., Zhang, H. et al. (2022) Chloroplast translational regulation uncovers nonessential photosynthesis genes as key players in plant cold acclimation. *Plant Cell*, **34**, 2056–2079.
- Gawronski, P., Jensen, P.E., Karpinski, S., Leister, D. & Scharff, L.B. (2018) Plastid ribosome pausing is induced by multiple features and is linked to protein complex assembly. *Plant Physiology*, **176**, 2557–2569.
- Gerashchenko, M.V. & Gladyshev, V.N. (2014) Translation inhibitors cause abnormalities in ribosome profiling experiments. *Nucleic Acids Research*, **42**, e134.
- Gloge, F., Becker, A.H., Kramer, G. & Bukau, B. (2014) Co-translational mechanisms of protein maturation. *Current Opinion in Structural Biology*, **24**, 24–33.
- Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D. et al. (2020) Array programming with NumPy. *Nature*, **585**, 357–362.

- Hinnebusch, A.G. (2011) Molecular mechanism of scanning and start codon selection in eukaryotes. *Microbiology and Molecular Biology Reviews*, **75**, 434–467.
- Hsu, P.Y., Calviello, L., Wu, H.L., Li, F.W., Rothfels, C.J., Ohler, U. *et al.* (2016) Super-resolution ribosome profiling reveals unannotated translation events in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, E7126–E7135.
- Hunter, J. (2007) Matplotlib: a 2D graphics environment. *Computing in Science & Engineering*, **9**, 90–95.
- Ingolia, N.T. (2014) Ribosome profiling: new views of translation, from single codons to genome scale. *Nature Reviews. Genetics*, **15**, 205–213.
- Ingolia, N.T. (2016) Ribosome footprint profiling of translation throughout the genome. *Cell*, **165**, 22–33.
- Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M. & Weissman, J.S. (2012) The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nature Protocols*, **7**, 1534–1550.
- Ingolia, N.T., Ghaemmaghami, S., Newman, J.R. & Weissman, J.S. (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, **324**, 218–223.
- Ingolia, N.T., Lareau, L.F. & Weissman, J.S. (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, **147**, 789–802.
- Juntawong, P., Girke, T., Bazin, J. & Bailey-Serres, J. (2014) Translational dynamics revealed by genome-wide profiling of ribosome footprints in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, E203–E212.
- Klimmek, F., Sjödin, A., Noutsos, C., Leister, D. & Jansson, S. (2006) Abundantly and rarely expressed *Lhc* protein genes exhibit distinct regulation patterns in plants. *Plant Physiology*, **140**, 793–804.
- Komine, Y., Kwong, L., Anguera, M.C., Schuster, G. & Stern, D.B. (2000) Polyadenylation of three classes of chloroplast RNA in *Chlamydomonas reinhardtii*. *RNA*, **6**, 598–607.
- Kramer, G., Shiber, A. & Bukau, B. (2019) Mechanisms of Cotranslational maturation of newly synthesized proteins. *Annual Review of Biochemistry*, **88**, 337–364.
- Kropat, J., Hong-Hermesdorf, A., Casero, D., Ent, P., Castruita, M., Pellegrini, M. *et al.* (2011) A revised mineral nutrient supplement increases biomass and growth rate in *Chlamydomonas reinhardtii*. *The Plant Journal*, **66**, 770–780.
- Lareau, L.F., Hite, D.H., Hogan, G.J. & Brown, P.O. (2014) Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *eLife*, **3**, e01257.
- Lauria, F., Tebaldi, T., Bernabo, P., Groen, E.J.N., Gillingwater, T.H. & Viero, G. (2018) riboWaltz: optimization of ribosome P-site positioning in ribosome profiling data. *PLoS Computational Biology*, **14**, e1006169.
- Lei, L., Shi, J., Chen, J., Zhang, M., Sun, S., Xie, S. *et al.* (2015) Ribosome profiling reveals dynamic translational landscape in maize seedlings under drought stress. *The Plant Journal*, **84**, 1206–1218.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. *et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
- Liu, M.J., Wu, S.H., Wu, J.F., Lin, W.D., Wu, Y.C., Tsai, T.Y. *et al.* (2013) Translational landscape of photomorphogenic *Arabidopsis*. *Plant Cell*, **25**, 3699–3710.
- Lukoszek, R., Feist, P. & Ignatova, Z. (2016) Insights into the adaptive response of *Arabidopsis thaliana* to prolonged thermal stress by ribosomal profiling and RNA-Seq. *BMC Plant Biology*, **16**, 221.
- Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal*, **2011**(17), 3.
- McKinney, W. (2010) Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, 56–61. Available from: <https://doi.org/10.25080/Majora-92bf1922-00a>
- Merchante, C., Brumos, J., Yun, J., Hu, Q., Spencer, K.R., Enriquez, P. *et al.* (2015) Gene-specific translation regulation mediated by the hormone-signaling molecule EIN2. *Cell*, **163**, 684–697.
- Moulin, M., Nguyen, G.T., Scaife, M.A., Smith, A.G. & Fitzpatrick, T.B. (2013) Analysis of *Chlamydomonas* thiamin metabolism in vivo reveals riboswitch plasticity. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 14622–14627.
- Mussnug, J.H., Wobbe, L., Elles, I., Claus, C., Hamilton, M., Fink, A. *et al.* (2005) NAB1 is an RNA binding protein involved in the light-regulated differential expression of the light-harvesting antenna of *Chlamydomonas reinhardtii*. *Plant Cell*, **17**, 3409–3421.
- Nickelsen, J., Bohne, A.-V. & Westhoff, P. (2014) Chloroplast gene expression - translation. *Plastid Biology*, **5**, 49–78.
- Payne, S.H. (2015) The utility of protein and mRNA correlation. *Trends in Biochemical Sciences*, **40**, 1–3.
- Pechmann, S., Willmund, F. & Frydman, J. (2013) The ribosome as a hub for protein quality control. *Molecular Cell*, **49**, 411–421.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. *et al.* (2011) Scikit-learn: machine learning in python. *Journal of Machine Learning Research*, **12**, 2825–2830.
- Plancke, C., Vigeolas, H., Hohner, R., Roberty, S., Emonds-Alt, B., Larosa, V. *et al.* (2014) Lack of isocitrate lyase in *Chlamydomonas* leads to changes in carbon metabolism and in the response to oxidative stress under mixotrophic growth. *The Plant Journal*, **77**, 404–417.
- Rooijers, K., Loayza-Puch, F., Nijtmans, L.G. & Agami, R. (2013) Ribosome profiling reveals features of normal and disease-associated mitochondrial translation. *Nature Communications*, **4**, 2886.
- Russell, J.B. & Cook, G.M. (1995) Energetics of bacterial growth: balance of anabolic and catabolic reactions. *Microbiological Reviews*, **59**, 48–62.
- Schneider-Poetsch, T., Ju, J., Eyler, D.E., Dang, Y., Bhat, S., Merrick, W.C. *et al.* (2010) Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. *Nature Chemical Biology*, **6**, 209–217.
- Schroda, M., Hemme, D. & Mühlhaus, T. (2015) The *Chlamydomonas* heat stress response. *The Plant Journal*, **82**, 466–480.
- Schuster, M., Gao, Y., Schöttler, M.A., Bock, R. & Zoschke, R. (2019) Limited responsiveness of chloroplast gene expression during acclimation to high light in tobacco. *Plant Physiology*, **182**, 424–435.
- Sharma, P., Wu, J., Nilges, B.S. & Leidel, S.A. (2021) Humans and other commonly used model organisms are resistant to cycloheximide-mediated biases in ribosome profiling experiments. *Nature Communications*, **12**, 5094.
- Slobodin, B. & Dikstein, R. (2020) So close, no matter how far: multiple paths connecting transcription to mRNA translation in eukaryotes. *EMBO Reports*, **21**, e50799.
- Smith, T., Heger, A. & Sudbery, I. (2017) UMI-tools: modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Research*, **27**, 491–499.
- Sun, Y. & Zerges, W. (2015) Translational regulation in chloroplasts for development and homeostasis. *Biochimica et Biophysica Acta*, **1847**, 809–820.
- Teixeira, F.K. & Lehmann, R. (2019) *Translational control during developmental transitions*. Cold Spring Harb Perspect Biol, p. 11. Available from: <https://doi.org/10.1101/cshperspect.a032987>
- Ting, M.K.Y., Gao, Y., Barahimipour, R., Ghandour, R., Liu, J., Martinez-Seidel, F. *et al.* (2023) Guidelines for performing ribosome profiling in plants with structural analysis of rRNA fragments. *BioRxiv*. Available from: <https://doi.org/10.1101/2023.11.16.567332>
- Trösch, R., Barahimipour, R., Gao, Y., Badillo-Corona, J.A., Gotsmann, V.L., Zimmer, D. *et al.* (2018) Commonalities and differences of chloroplast translation in a green alga and land plants. *Nature Plants*, **4**, 564–575.
- Trösch, R., Ries, F., Westrich, L.D., Gao, Y., Herkt, C., Hopstädter, J. *et al.* (2022) Fast and global reorganization of the chloroplast protein biogenesis network during heat acclimation. *Plant Cell*, **34**, 1075–1099.
- Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D. *et al.* (2020) SciPy 1.0: fundamental algorithms for scientific computing in python. *Nature Methods*, **17**, 261–272.
- Wang, F., Dischinger, K., Westrich, L.D., Meindl, I., Egid, F., Trosch, R. *et al.* (2023) ONE-HELIX PROTEIN 2 is not required for the synthesis of photosystem II subunit D1 in *Chlamydomonas*. *Plant Physiology*, **191**, 1612–1633.
- Wang, R. & Amoyel, M. (2022) mRNA translation is dynamically regulated to instruct stem cell fate. *Frontiers in Molecular Biosciences*, **9**, 863885.
- Waskom, M. (2021) Seaborn: statistical data visualization. *Journal of Open Source Software*, **6**, 3021.
- Westrich, L.D., Gotsmann, V.L., Herkt, C., Ries, F., Kazek, T., Trosch, R. *et al.* (2021) The versatile interactome of chloroplast ribosomes revealed by affinity purification mass spectrometry. *Nucleic Acids Research*, **49**, 400–415.

- Willmund, F., Hauser, C. & Zerges, W. (2023) Translation and protein synthesis in the chloroplast. In: Grossman, A.R. & Wollman, F.A. (Eds.) *The Chlamydomonas Sourcebook*, Vol. 2, third edition. London: Academic Press Elsevier, pp. 467–508.
- Wolin, S.L. & Walter, P. (1988) Ribosome pausing and stacking during translation of a eukaryotic mRNA. *The EMBO Journal*, **7**, 3559–3569.
- Wu, C.C., Zinshteyn, B., Wehner, K.A. & Green, R. (2019) High-resolution ribosome profiling defines discrete ribosome elongation states and translational regulation during cellular stress. *Molecular Cell*, **73**, 959–970.e955.
- Wu, D., Huang, J., Zhu, H., Chen, Z., Chai, Y., Ke, J. *et al.* (2022) Ciliogenesis requires sphingolipid-dependent membrane and axoneme interaction. *Proceedings of the National Academy of Sciences of the United States of America*, **119**, e2201096119.
- Wu, H.L., Song, G., Walley, J.W. & Hsu, P.Y. (2019) The tomato translational landscape revealed by transcriptome assembly and ribosome profiling. *Plant Physiology*, **181**, 367–380.
- Yang, X., Cui, J., Song, B., Yu, Y., Mo, B. & Liu, L. (2020) Construction of high-quality Rice ribosome footprint library. *Frontiers in Plant Science*, **11**, 572237.
- Yang, X., Song, B., Cui, J., Wang, L., Wang, S., Luo, L. *et al.* (2021) Comparative ribosome profiling reveals distinct translational landscapes of salt-sensitive and -tolerant rice. *BMC Genomics*, **22**, 612.
- Zhang, W., Dunkle, J.A. & Cate, J.H. (2009) Structures of the ribosome in intermediate states of ratcheting. *Science*, **325**, 1014–1017.
- Zoschke, R. & Bock, R. (2018) Chloroplast translation: structural and functional organization, operational control and regulation. *Plant Cell*, **30**, 745–770.