Utilizing high resolution ribosome profiling for the global investigation of gene expression in *Chlamydomonas*

Vincent Leon Gotsmann[1], Michael Kien Yin Ting[2], Nadin Haase[3], Sophia Rudorf[3], Reimo Zoschke[2], and Felix Willmund[1]*

Author's institution(s)/affiliation(s):

[1] Molecular Genetics of Eukaryotes, RPTU Kaiserslautern-Landau, Paul-Ehrlich-Str. 23, 67663 Kaiserslautern, Germany.

[2] Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany.

[3] Institute of Cell Biology and Biophysics, Leibniz University Hanover, Herrenhäuser-Str. 2, 30419 Hanover, Germany.

***Corresponding author:** Felix Willmund (willmund@staff.uni-marburg.de)

Keywords: ribosome profiling, next generation sequencing, ribosomes, protein synthesis, translation, transcript accumulation

Short title: Cell-wide ribosome profiling in *Chlamydomonas*

Abstract

Ribosome profiling (Ribo-seq) is a powerful method for the deep analysis of translation mechanisms and regulatory circuits during gene expression. Here, we established an optimized and high resolution Ribo-seq protocol for the unicellular model alga *Chlamydomonas reinhardtii* (Chlamydomonas). Comparing different nuclease treatments for the extraction and sequencing of ribosome-protected fragments (RPFs) and parallel RNA-seq, provided deep insight into translational dynamics and post-transcriptional control of gene expression, thoroughly covering more than 10,000 different transcripts. Our high quality Ribo-seq protocol captures the 3-nucleotide movement of elongating ribosomes along nuclear and chloroplast 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 features of eukaryotic translation termination. These offsets reveal drastic differences between the nature of cytosolic and chloroplast translation mechanisms. Chloroplast translation is further characterized by heterogenous RPF size distribution. We found that local accumulation of small RPFs correlates with local slowdown of *psbA* translation, possibly revealing an uncharacterized regulator 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. Moreover, our Ribo-seq data can be utilized to survey coding sequence annotations and the expression preference of alternatively spliced transcripts in Chlamydomonas. We made these features easily accessible for the research community by attaching our Ribo-seq data to the most recent Chlamydomonas reference genome.

45 **Introduction**

46 Translation is accomplished via ribosomes, highly conserved macromolecular
47 ribonucleoprotein machines that decode genetic information into linear polypeptide chains
48 resulting in an almost unlimited diversity of proteins. Protein synthesis is the origin for the
49 quantitative and qualitative determination of a proteome´s diversity, which ultimately
50 shapes the biochemical character of a cell. It is now clear that translation output does not
51 strictly follow the changes of transcript abundance, but instead underlies strict regulatory
52 control (Pechmann et al., 2013; Payne, 2015; Slobodin and Dikstein, 2020). The
53 importance of this regulation is illustrated by the fact that bacteria invest about 50% of
54 their available energy for protein biogenesis and that more than 10% of all yeast proteins
55 may be somehow involved in this process (Russell and Cook, 1995; Costanzo et al.,
56 2000). Translational regulation is observed throughout an organism's life span and
57 includes central regulatory circuits that control the cell cycle, tissue differentiation and
58 development, acclimation to environmental changes and the integration of external
59 signals (reviewed in Pechmann et al., 2013; Zoschke and Bock, 2018; Teixeira and
60 Lehmann, 2019; Wang and Amoyel, 2022). Thus, understanding the regulatory principles
61 and mechanisms of translation is a central task of modern biology. With the tremendous
62 advances in next generation sequencing technologies, it is now possible to systematically
63 assay the status of cellular translation by a method called ribosome profiling (Ribo-seq)
64 (Ingolia et al., 2009). The technology is based on the observation that ribosomes occupy
65 precise sections of a translated mRNA (Wolin and Walter, 1988). These mRNA sections
66 (termed ribosome footprints or Ribosome Protected Fragments, RPFs) are extracted by
67 nucleolytic removal of un-occupied mRNA in cell lysates or isolated ribosomes. RPFs are
68 then identified through next generation sequencing and mapping to a reference

69 transcriptome or genome. Thus, Ribo-seq benefits from the same dynamic range as RNA-
70 seq, and provides a rich and precise positional information of translating ribosomes from
71 the physiological state when cells were harvested (Ingolia et al., 2009; Ingolia, 2014; Brar
72 and Weissman, 2015). Parallel sequencing of all expressed transcripts via RNA-seq
73 allows the comparison of RNA accumulation with translation and thereby to determine the
74 fraction of translated transcripts, which is often referred to as “translation efficiency”.
75 Assuming that each RPF represents a translating ribosome and thus a synthesized
76 nascent polypeptide, mRNAs with high average RPF coverages can be considered as
77 highly translated transcripts. In addition, local over-representation of specific RPFs within
78 a specific protein-coding sequence (CDS) are commonly interpreted as regions of slow or
79 even halted translation (Ingolia et al., 2009; Ingolia, 2014; Brar and Weissman, 2015).
80 Compared to other techniques, ribosome profiling goes far beyond the targeted assays of
81 radio-labelling of nascent polypeptides or polysome-loading for specific mRNAs, and
82 regularly provides much deeper dataset information than proteomic analyses via mass-
83 spectrometry (Ingolia, 2014; Brar and Weissman, 2015). Yet, ribosome profiling faces the
84 challenges to distinguish active translation events from non-translating (paused or stalled)
85 ribosomes. However, high-quality Ribo-seq datasets, with populations of precise read-
86 length distribution allow the determination of active translation events if RPFs are aligned
87 relative to the ribosomal Peptidyl-site (P-site). This reveals a 3-nucleotide periodic
88 movement of translating ribosome along a translated CDS by over-accumulation of reads
89 containing the first nucleotide of a decoded codon in the P-site. Thus, it is possible to
90 distinguish ribosome pausing/stalling from translation and therefore even enables *ab initio*
91 annotation of uncharacterized protein-coding sequences (Ingolia et al., 2009; Calviello et
92 al., 2016; Hsu et al., 2016).

93 Ribo-seq has been applied to different plant species with variable gene coverage
94 and quality, including *Chlamydomonas reinhardtii* (Chlamydomonas hereafter) (Chung et
95 al., 2015), *Arabidopsis thaliana*, (Liu et al., 2013; Juntawong et al., 2014; Merchante et
96 al., 2015; Hsu et al., 2016; Lukoszek et al., 2016; Chen et al., 2022), maize (Lei et al.,
97 2015; Chotewutmontri and Barkan, 2018), and other crop plants (Wu et al., 2019; Yang et
98 al., 2020; Yang et al., 2021; Chiu et al., 2022). Studying translational regulation in plants
99 is of specific interest, due to the interplay of three semi-autonomous genomes (nuclear,
100 chloroplast and mitochondrial), for which many aspects are not fully understood to date.
101 Chloroplast gene expression seems particularly dependent on co-translational regulation,
102 which facilitates the fast adjustment of the photosynthesis machinery to environmental
103 changes and the stoichiometric assembly of chloroplast multi-subunit complexes which
104 often contain nuclear and chloroplast encoded subunits (Eberhard et al., 2002; Zoschke
105 and Bock, 2018; Fujita et al., 2019). Recently, ribosome profiling revealed the specific
106 adjustments made to chloroplast gene expression during acclimation to different light and
107 temperature regimes (Chotewutmontri and Barkan, 2018; Schuster et al., 2019;
108 Chotewutmontri and Barkan, 2020; Gao et al., 2022; Trösch et al., 2022). We have
109 previously applied a high-resolution microarray approach for the fast and cost-efficient
110 analysis of chloroplast translation in Chlamydomonas, which allowed for the direct
111 comparison of transcript accumulation and translation output between the alga and land
112 plants (Trösch et al., 2018; Trösch et al., 2022). We now established Ribo-seq for the
113 deep analyses of the three Chlamydomonas genomes. Ribo-seq had been described for
114 Chlamydomonas before, however, with limited depth and analysis (Chung et al., 2015).
115 However, we aimed to optimize the method to obtain a deeper coverage of genes, with
116 super-resolution tri-nucleotide periodicity for both nucleus- and chloroplast-encoded

117 genes and to provide a publicly accessible dataset aligned to the most recent
118 Chlamydomonas genome (Craig et al., 2022).

119 **Results and Discussion**

120 *Establishing super resolution ribosome profiling for Chlamydomonas reinhardtii*

121 Ribosome profiling was performed from logarithmically grown Chlamydomonas cultures
122 that were kept under mixotrophic conditions and moderate light (see Materials and
123 Methods). Our previous experience with targeted chloroplast ribosome profiling (Trösch
124 et al., 2018; Trösch et al., 2022) showed that addition of the elongation inhibitors
125 chloramphenicol (CAP) and cycloheximide (CHX) (stalling 70S and 80S ribosomes,
126 respectively) is crucial to prevent ribosome run-off during harvesting (see Materials and
127 Methods). However, by short incubation of the inhibitors just during harvest, we reduced
128 the exposure time to the minimum. For nucleolytic digest of RNA that is not protected by
129 a ribosome, we applied RNase I, consistent with most Ribo-seq studies on other
130 organisms (Ingolia et al., 2012). Our previous chloroplast ribosome profiling approaches
131 (using MNase) revealed that the best digestion (i.e. polysome to monosome dissociation)
132 was achieved when ribosomes were purified prior nuclease treatment (Trösch et al.,
133 2018). We thus compared RNase I digestion on pre-purified ribosomes both at 4°C
134 (condition i.) and 23°C (condition ii.) with samples containing whole cell lysates (condition
135 iii.) (Figure 1A). RNase I was applied at a concentration of 1 unit (U) per micro gram (µg)
136 RNA, which is similar to published Ribo-seq approaches performed in yeast and
137 Arabidopsis (Chartron et al., 2016; Hsu et al., 2016; Döring et al., 2017). However, the
138 RNase I concentration was far below the >9 U per µg RNA that was previously applied for
139 Chlamydomonas (Chung et al., 2015). By using the lowest possible concentrations of
140 RNase I, experimental costs can be reduced. Moreover, over digestion of ribosomes may

141 cause an over-proportional content of contaminating rRNA fragments, which then
142 dominate sequencing libraries. In the previous Chlamydomonas Ribo-seq study by Chung
143 et al. (2015), contaminating rRNA fragments accounted for >90% of all sequenced reads.
144 Our pilot runs had contaminating rRNA levels between 80 and 90%, mainly deriving from
145 the cytosolic 5.8S rRNA and the chloroplast 23S rRNA. To further reduce the amount of
146 contaminating rRNA species, we generated anti-sense biotinylated oligos that allow for
147 the specific depletion of these species. For digest condition iii., we additionally tested
148 RNase H treatment, which cleaves hybrids of the anti-sense oligos and the target rRNA
149 contaminants. Other than the standard procedure, we performed rRNA depletion prior to
150 dual linker ligation and were thus able to provide a higher input of RPFs in the sample,
151 relative to contaminating rRNAs, which then allows for lowering the number of PCR
152 amplification steps during final sequencing library generation. Importantly, all these steps
153 were similar to a Ribo-seq protocol that we established for Arabidopsis and tobacco in
154 parallel (Ting et al., 2023). The final PCR amplification step of the sequencing library
155 preparation is known to be a major source of bias by over-representing certain reads (Aird
156 et al., 2011). Read duplicates were defined as reads that share the same sequence and
157 the same unique molecular identifier tags (UMI tags), four random bases located between
158 a RPF and the sequencing adapters at both of its termini (Fu et al., 2018). Resulting
159 sequencing libraries were thus corrected by deduplicating the final data sets after the
160 genome mapping step.

161 Parallel to Ribo-seq, transcript accumulation was determined by sequencing whole
162 cell RNA samples on the same platform. Ideally, RNA samples are harvested just prior to
163 ribosome profiling experiments from the same culture. While typical transcriptomic studies
164 enrich mRNA pools by polyadenylation-enrichment, we here sequenced fragmented small

165 RNAs that are derived from the full set of transcripts including plastid transcripts, which
166 frequently have no or short polyadenylation signal sequences and are thus poorly
167 enriched by standard protocols (Komine et al., 2000; Gallaher et al., 2018).

168 Ribosome profiling and transcript samples were sequenced with a depth of 40 M
169 and 20 M reads, respectively, which on average resulted in 25.7% of reads mapping to
170 annotated CDSs of the most recent Chlamydomonas reference genome v6.1 (Craig et al.,
171 2022). On average, 64% mapped to putative rRNA loci and 4.2% mapped to annotated
172 untranslated regions upstream or downstream of CDSs. Despite the application of
173 different rRNA depletion methods, the rRNA contamination was very similar in all samples.
174 Of the sequenced reads that mapped annotated CDS a minor fraction (<1%) matched to
175 more than 10 positions and were removed from further analyses. Of the reads mapping
176 to CDSs, 96.24% mapped to transcripts of the nuclear genome, 3.63% mapped to
177 chloroplast-encoded transcripts, and 0.12% to mitochondrial transcripts.

178 Transcript accumulation was determined by averaging reads per CDS, relative to
179 the length of each CDS and the total read numbers of the dataset (Reads Per Kilobase
180 per Million, RPKM), a parameter that is widely used for transcriptome analyses and
181 accounts for a higher likelihood of measuring RNA fragments of long versus short
182 sequences (Conesa et al., 2016). This is of particular importance for the sequencing of
183 small and fragmented RNAs, as commonly accompanied with Ribo-seq. Averaging RPFs
184 across the respective CDS corrects for local variations of RPF distribution and allows to
185 get a proxy for the level of translation output per transcript (Ingolia et al., 2011; Ingolia,
186 2014). For better comparability and consistent with previous Ribo-seq studies, RPKM was
187 also determined for Ribo-seq data.

188 Based on RPKM values, the three nuclease digestion conditions (i. to iii.) showed
189 high reproducibility between the Ribo-seq and RNA-seq experiments, respectively (r =
190 0.99, Figure 1B). This agrees with previous observations that Ribo-seq tolerates variations
191 in RNase digest (Gerashchenko and Gladyshev, 2017).

192 We next aimed to determine the depth of the Ribo-seq approach. Low coverage
193 Ribo-seq frequently results in ribosome profiles covering only the transcripts with highest
194 ribosome occupancy (i.e. translation output), whereas moderate or lowly translated CDS
195 may have partial RPF coverage along the respective sequence (here termed gene body
196 coverage, Figure 1C, left panel). For Ribo-seq with a good depth, many CDS should show
197 decent gene body coverage (Figure 1C, right panel). The depth was determined for all
198 Chlamydomonas transcripts of the v6 genome (32670 transcripts considering the products
199 of gene copies). For digest condition iii., more than 10,000 transcripts had at least 50% of
200 their CDSs sequence covered with two or more RPFs, while 6437 transcripts had over
201 90% of their CDS covered with at least one RPF (Figure 1D). 1508 transcripts are covered
202 over >90% with an average of more than 10 RPFs/nt, hence representing the group of
203 transcripts with the most complete gene body coverage combined with a high basal
204 coverage (see also Figure 4B). Categorizing the transcripts in this way allows for the
205 assessment of the general quality of a data set, as well as the selection of transcripts of
206 similar quality for specific analyses without having to rely on expression values alone. The
207 gene body coverage was highest for digest condition iii. (Figure 1D and Supplemental
208 Figure S1), and comparably good for the coverage of RNA fragments for the RNA-seq
209 experiment (Figure 1E).

210 The read length distribution of the RPFs mapping to annotated CDSs of the nuclear
211 genome displayed high comparability to published Ribo-seq approaches with other

212 organisms. RNase I treatment typically results in RPFs with a length between 28 and 30
213 nt. Here, digest of purified ribosomes at 4°C resulted in RPFs with a predominant length
214 of 31 nt. Digest condition ii. caused a broader distribution of RPFs between 28 and 30 nt,
215 while digest in cell lysate resulted in a sharp peak of RPFs with a length of 30 nt (Figure
216 2A). Such a sharp peak at 30 nt is typically found in Ribo-seq from mammalian cells when
217 harvested in the presence of CHX (Wu et al., 2019; Sharma et al., 2021). The sharp peak
218 is even more prominent if only those RPFs are considered that map to the CDS
219 (Supplemental Figure S2A). This is most evident for the size distribution mapping to the
220 CDS of the 8 mitochondrial transcripts. Here, RPF accumulations peaks at a size of 33 nt
221 (Supplemental Figure S2A). The larger size of mitochondrial RPFs was described before
222 and could stem from the bulkiness of the membrane associated ribosomes (Rooijers et
223 al., 2013). Whereas, cytosolic and mitochondrial ribosomes produced sharp RPF peaks,
224 the RPF size distribution of chloroplast transcripts is less defined and may be the result of
225 diverse translational states, as previously reported for maize and Arabidopsis
226 (Chotewutmontri and Barkan, 2016; Gawronski et al., 2018; Fujita et al., 2019) (see
227 below). The sharpest peak for chloroplast RPFs is seen under condition ii., albeit the
228 predominant RPFs of 28 nt is found under all treatment conditions (Figure 2A). While
229 consistent with the observations in other organisms, our RPF size distribution varies from
230 the previous Ribo-seq data with Chlamydomonas, which reported a general shift towards
231 smaller RPF sizes, maybe as consequence from their harsh RNase I treatment (Chung et
232 al., 2015).

233 A hallmark of high-resolution Ribo-seq data is the triplet-wise 3-nt periodicity
234 of RPFs that represents slowdown of ribosome movement as ribosomes decode a
235 respective codon (Ingolia, 2016). To calculate periodicity, it is necessary to determine the

236 5’-offset for every RPF length species, which is the number of bases located 5’ to the P-
237 site (the translational state of polypeptide bond formation) of a respective RPF. This is
238 usually done by considering all RPFs that fully enclose an annotated start codon and
239 determining the frequency of all occurring 5’-offsets for every RPF length separately. The
240 5’-offset, most frequently occurring for a specific RPF length, is then declared to be the
241 true 5’-offset of this species. This strategy is based on the fact that RPF occupancy of
242 transcripts usually inclines dramatically at the borders between 5’-untranslated regions
243 and CDS, with the start codon often having an exceptionally high coverage, which is due
244 to the slow process of translation initiation (Supplemental Figure S2B). The 5’-offsets
245 determined in this way are used to calculate the exact P-site position and reading frame
246 of every RPF. By this, 3-nt periodicity can be nicely visualized in meta-plots combining all
247 genes while 3-nt plots of individual transcripts can be noisy, depending on translation
248 levels. The exact 5’-offset varies in an organism- and protocol-specific manner and is
249 usually 12 to 13 nt. Alternatively, the P-site position has also been determined by
250 calculating the 3’-offset of the ribosomal Aminoacyl-site (A-site), applying the same
251 strategy to RPFs mapping to annotated stop codons.

252 With the majority of cytosolic RPFs with one read length (Figure 2A), all three
253 RNase I digest conditions yielded a clear frame preference of “frame 0” for nucleus-
254 encoded transcripts, meaning that the base triplet found at the P-site of a RPF is for the
255 for the majority of RPFs in-frame with the annotated CDS. Again, lysate digest (iii.)
256 resulted in the best frame preference values (>68% of RPFs in frame 0, 19.4% in frame 1
257 and 12.5% in frame 2), suggesting that this condition yields Ribo-seq data of the best
258 quality (Figure 2B and Supplemental Figure S2B). Due to the differences between 80S
259 and 70S ribosomes, 3-nt periodicity needs to be determined separately for chloroplast

260 translation. For chloroplast transcripts, the frame preference was less obvious, albeit it
261 was clearest for digest condition iii. (Figure 2C and Supplemental Figure S2B).
262 Considering the overall quality of sample iii. (the similar degree of rRNA contamination
263 between all three samples and the high correlation with the other samples on RPKM level),
264 it can be concluded that RNase H mediated rRNA depletion has no adverse effects on the
265 Ribo-seq sample quality and seems to be similarly effective.

266 To get a better insight into the RPF offsets and periodicity, different lengths of the
267 RPFs were plotted against their offset frequencies upon P-site and A-site alignment for
268 the dataset of condition iii. (Figure 2D). The triplet-wise movements of the 5´ends with an
269 offset of 12 nt during initiation is clearly visible for cytosolic RPFs and most prominent for
270 the 30 nt RPFs of digestion conditions iii., consistent with the observation in other
271 eukaryotic organisms (Lauria et al., 2018). Likewise, triplet-wise movements of the 3´-
272 ends are also detectable and are in frame with the stop codon. Here, a 9 nt long 3´-offset
273 seems the predominant form prior to termination with the stop codon in the A-site of the
274 ribosome. However, due to the 12 nt 5'-offset and 3 nt occupied by the P- and A-sites
275 respectively, a 3'-offset of 12 nt was expected. This discrepancy could be a result of CHX
276 treatment, which locks ribosomes in a pre-translocation conformation after peptide bond
277 formation (Wu et al., 2019) so that RPFs on non-stop codons may exceed the natural
278 termination-caused accumulation on stop-codons. Moreover, a slow termination process
279 might arrest ribosomes on the codon just upstream of the stop, prolonging ribosome dwell
280 time at this site. At the same time, ribosomes that already entered termination during
281 sampling are likely to complete the process and dissociate from the transcript. CHX does
282 not inhibit termination, which in turn would dramatically lower the number of FPs mapped
283 to stop codons with their A-site. Interestingly, RPFs with a length of 21 nt are detectable

284 upon initiation (asterisks, Figure 2D). These fragments have been previously
285 characterized as RPFs of ribosomes in an open ratchet conformation during translocation,
286 even accumulating as abundant species without CHX treatment in yeast (Lareau et al.,
287 2014). During elongation, ribosomes undergo massive conformational changes by
288 rotating the large subunit relative to the small subunit (Frank and Agrawal, 2000; Zhang
289 et al., 2009). The fact that the RPFs produced from the open-ratchet conformation have
290 the same 5'-offsets as the main RPF species despite being considerably shorter, suggests
291 that the RNase exclusively removes nucleotides from the 3´-end of RPFs in this
292 conformation (and could indicate that the center of the rotation is located towards the 5'-
293 direction, viewed relative to the RPF). Cycloheximide was shown to prevent binding of the
294 next aminoacylated tRNA in the A-site during peptide bond formation (Schneider-Poetsch
295 et al., 2010) and thus stabilizes ribosomes in a specific stage during the elongation cycle,
296 in which RPFs of 28-30 nt are the prevalent form. Interestingly, human 80S ribosomes are
297 not exclusively arrested in one conformation upon CHX treatment (Sharma et al., 2021),
298 comparable to the situation seen here for Chlamydomonas 80S ribosomes (Figure 2D,
299 upper panel). These 20-22 nt RPFs might derive from a post-translocation ribosome
300 species (Lareau et al., 2014). The presence of both species demonstrates active
301 movement of ribosomes along the CDS. The absence of the "open-ratchet" 21 nt RPFs
302 starting from two peptide-bond formations prior to termination, may support the idea from
303 above that ribosomes near stop codons are paused, which can prevent the collision
304 between ribosomes (Figure 2D upper panel, dashed box). Interestingly, the frame
305 preference for frame 0 of the stop-minus-2 codon is the highest in the whole metagene
306 periodicity plot (79.6% Figure 2B), which could denote a short stalling at this codon before
307 translocation. However, for the next codons the frame 0 preference drops considerably to

308 55.8%, especially in favor of frame 1 (36.2%) and then to 33.1% at the stop codon, clearly
309 indicating that translation dynamics change towards the termination step. Moreover, these
310 characteristics are accompanied by a continuous decline of absolute numbers of reads
311 mapping to these codons on the meta-gene level (for digest iii: stop-2: 19890, stop-1:
312 6786, stop: 522).

313 The prokaryotic-type chloroplast RPFs 5'-offsets seem to be much more diverse
314 than their cytosolic counterparts, apparent by their multiple clusters on the heat map
315 (Figure 2D, bottom panel). These might resemble populations of initiation complexes in
316 different states or in complex with transcript specific RNA processing factors and *trans*-
317 acting factors that interact with ribosomes (Westrich et al., 2021). Without clear criteria
318 how to distinguish and interpret these clusters, 5'-offset definition is problematic, since
319 multiple 5'-offset frequencies of comparable amplitude exist for many RPF lengths.
320 However, the chloroplast 3'-offsets seem less scattered and reveals a main RPF species
321 of 28 nt length with 11 nt 3'-offset. This is in accordance with the RPF length distribution
322 of all chloroplast RPFs, where the 28 nt species also represents the main species,
323 although the distribution is much broader than for the cytosolic RPFs. Similar to the
324 cytosolic offset heat map, a triplet-wise movement of ribosomes can also be observed for
325 the terminating chloroplast ribosomes in the 3'-offset heatmap for 29 nt RPFs, albeit with
326 much lower clarity. Altogether, the observations made for chloroplast offsets and RPF
327 length distribution might indicate a certain heterogeneity of the chloroplast ribosome pool,
328 which is an interpretation that would be in line with the large number of proteins that
329 interact with chloroplast ribosome found in our previous study (Westrich et al., 2021) and
330 the large number of RNA-binding proteins known or suspected to regulate the translation
331 of specific chloroplasts transcripts (Nickelsen et al., 2014). This is remarkable considering

332 the very small number of transcripts encoded on the chloroplast genome (72). Together,
333 this might suggest that translation in the chloroplast of Chlamydomonas is heavily
334 influenced by transcript-specific features rather than following a universal scheme
335 (Zoschke and Bock, 2018).

336 *Differences in translation output between cytosolic and organellar transcripts*

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

346 With the high reproducibility of Ribo-seq and RNA-seq data, we chose to average
347 the datasets for estimating the levels of transcript accumulation and translation output. A
348 direct comparison between translation output and transcript accumulation is frequently
349 referred to as translation efficiency (TE) and normalizes RPF values over transcript
350 accumulation. Contrasting translation over transcript data of the Chlamydomonas genes
351 in a scatter plot and highlighting their subcellular localization showed that nuclear
352 transcripts encoding chloroplast-localized proteins have higher expression values when
353 compared to proteins of cytosolic or Endoplasmic Reticulum localization (Figure 3C),
354 again supporting the notion that many chloroplast proteins are expressed at high levels.
355 Interestingly, most nucleus-encoded transcripts showed similar levels between transcript

356 accumulation and translation (distributed along the diagonal, Figure 3C). In contrast,
357 chloroplast- and mitochondria-encoded transcripts showed higher transcript accumulation
358 levels, when compared to translation, which agrees with the published data of profound
359 potential translational regulation within the chloroplast (Sun and Zerges, 2015; Trösch et
360 al., 2018; Zoschke and Bock, 2018). Remarkably, translational regulation seems also
361 predominant for the few mitochondrial genes.

362 Column plots of the individual chloroplast transcript and translation values show
363 that even the weakest expressed known chloroplast transcripts have RPKM values ~10,
364 which is within the median range of nucleus-encoded transcripts (Figure 4A versus 3A).
365 Not surprisingly, transcripts that encode the core photosystem I subunits PsaA and PsaB,
366 the photosystem II core subunits PsbA/D1,PsbB/CP47, PsbC/CP43, PsbD/D2, the
367 Rubisco large subunit, RbcL, the CF_0 ATPase subunit AtpH, and the translation elongation
368 factor TufA show the highest translation levels, consistent with our previous chloroplast
369 ribosome profiling experiments and the examination of protein accumulation levels
370 (Schroda et al., 2015; Trösch et al., 2018). We further detected transcripts of the
371 uncharacterized chloroplast Open Reading Frames (ORF) *orf202* (positioned within the
372 transposon Wendy I), *orf854* (positioned within the transposon Wendy II), *orf528*, and *i-*
373 *crel* (Gallaher et al., 2018), however, translation output was low or even undetectable (for
374 *orf2020* and *i-crel*), suggesting that these transcripts are barely translated in
375 Chlamydomonas cultures, at least under standard conditions (Figure 4A). The direct TE
376 ratios of the chloroplast transcripts again show the imbalance between chloroplast
377 transcript accumulation and translation output, suggesting that the long-living chloroplast
378 transcripts present a buffering capacity for the fast and efficient adjustment of translation
379 levels during environmental changes, as for example seen during light and temperature

380 acclimation and most prominent for PsbA/D1 synthesis (Chotewutmontri and Barkan,
381 2018; Schuster et al., 2019; Trösch et al., 2022).

382 We next determined translation output levels of nucleus-encoded transcripts, only
383 considering the longest transcript variant if multiple splice forms existed. Of the transcripts
384 with the highest translation output, 51 protein products have a known or predicted
385 chloroplast localization and 47 are present in the nucleus or cytosol, again showing the
386 high accumulation levels of chloroplast proteins (Supplemental Figure S3A). Amongst
387 those, transcripts of the light harvesting complex proteins (LHCs) have the highest
388 translation output, consistent with their high protein accumulation levels in plant cells
389 (Dall'Osto et al., 2015). Generally, *LHC* translation showed comparable RPKM translation
390 output values, whereas transcript accumulation was rather variable. For the transcripts of
391 the high-light induced photoprotective LHCSR3.1 and LHCSR3.3 proteins, we could not
392 determine translation profiles (Supplemental Figure S3B), suggesting that these proteins
393 are not synthesized under the investigated growth conditions (Bonente et al., 2012).
394 Besides photosynthesis related transcripts, the list with top-most translation contained
395 transcripts encoding numerous ribosomal proteins of the 80S complex and elongation
396 factors (Figure 4B). Other abundantly translated proteins are the putative acetate
397 transporter GYF3 (RPKM 4285, rank 14), the Isocitrate lyase (ICL1, RPKM 3918, rank
398 16), a key enzyme of the glyoxylate cycle (Plancke et al., 2014), the thiazole biosynthesis
399 enzyme (THI4, RPKM 3402, rank 22) (Moulin et al., 2013) or the Methionine Adenosyl-
400 transferase (METM1, RPKM 2438, rank 51), confirming their central role in
401 Chlamydomonas cells. Plotting of transcripts with the highest TE values showed an over
402 representation of histone transcripts, suggesting their robust translation in
403 Chlamydomonas cells under logarithmic growth (Supplemental Figure S3C).

404 *Exploiting individual ribosome profiles to mine for putative regulatory principles*

405 Local elongation rates are known to vary considerably during translation along a transcript,
406 depending on tRNA availability, mRNA secondary structure, the amino acid sequence of
407 the resulting polypeptide and co-translational folding (Gloge et al., 2014; Kramer et al.,
408 2019). With the robust gene body coverage over Chlamydomonas transcripts (Figure 1D),
409 we aimed to learn more from the dynamics of translation elongation, derived from the RPF
410 distribution over individual coding sequences (ribosome profiles). Ribosome profiles were
411 highly comparable between the three RNase I treatment conditions i.-iii., as indicated by
412 median *Pearson* correlations between data sets ii. and iii. of ~0.62 for nucleus-encoded
413 transcripts and >0.85 for chloroplast-encoded transcripts. Although RNA-seq read
414 distributions over individual CDS were even more comparable (median of *r*-value 0.87 for
415 nucleus-encoded and 0.99 for chloroplast-encoded transcripts) among the RNA-seq data
416 sets, they showed no correlation (median *r*-value close to zero) to the respective ribosome
417 profiles (Figure 5A and Supplemental Figure S4). Certainly, methodological-driven
418 differences between Ribo-seq and RNA-seq cannot be fully excluded, but it can be
419 assumed that the ribosome profiles reflect transcript-specific translation dynamics and are
420 likely not the result of sequencing biases.

421 The correlation coefficients for specific transcripts comparing two experiments
422 (both *Pearson* and *Spearman*) stably range between 0.7 and 1 for the majority of ribosome
423 profiles that have an average coverage >100 reads per nucleotide, suggesting that low
424 correlation coefficients are the consequence of sequencing depth limitations and thus
425 limited observability of lowly translated transcripts rather than the true absence of a
426 characteristic translation dynamic of these transcripts (Supplemental Figure S5).
427 Importantly, more than 10,000 transcripts have *r*-values >0.7 (Figure 5B), again showing

428 the high reproducibility between the RNase I digest conditions for transcripts with
429 moderate and high RPF coverage (Supplemental Figure S5). In fact, the *r*-value
430 distribution shows that these direct comparisons of read coverage per transcript might be
431 well suited for selecting good cut-off ranges for a deeper analyses of ribosome profiles. In
432 many previous studies, simply minimum RPKM values (e.g. 1) were taken as cut-off.

433 Ribosome profiles of the highly translated *LHC* transcripts showed some
434 remarkable differences between the individual *LHC* species. Efficiently controlling the
435 expression of *LHC* genes is essential for regulating the antenna size of the photosynthesis
436 machinery in order to capture photons under low irradiance conditions (increased
437 expression) or avoid over-excitation and thus photodamage under high irradiance
438 conditions (reduced expression) (Dall'Osto et al., 2015). Of the transcripts encoding the
439 21 different LHC members serving Photosystem I or II (LHCA and LHCB, respectively),
440 20 were found amongst the top 100 highest translated transcripts (not considering splice
441 variants). Only the transcript encoding LHCB7, a rather newly discovered protein
442 containing the unusual number of 4 transmembrane domains (Klimmek et al., 2006) had
443 low RPKM values of ~15 (Supplemental Figure S3B). The variations in *LHC* transcript
444 levels result in considerably different TE values (Supplemental Figure S3B), suggesting
445 that translational regulation might be achieved by different strategies for expression of the
446 *LHC* species. Indeed, the ribosome profiles are considerably different between transcripts
447 encoding the different LHCs. For example, *LHCBM6* showed an exceptionally strong peak
448 of RPFs covering the first 7 codons, which exceeds the average RPF density over the
449 remaining CDS by more than 4-fold (Figure 5C). While “initiation peaks” are commonly
450 accumulating over the first few codons of a CDS, their amplitudes are usually within the
451 dynamic range of the remaining CDS. Such high peaks, in contrast, could point to a

452 regulative mechanism stalling ribosomes during or shortly after initiation. Indeed,
453 *LHCBM6* was shown to be translationally repressed under moderate light conditions by a
454 co-translational regulator termed NAB1 (Nucleic Acid Binding protein). NAB1 itself is
455 tuned by a regulatory circuit sensing carbon dioxide supply levels in Chlamydomonas, for
456 precisely adjusting the antenna composition of photosystem II (Mussgnug et al., 2005;
457 Blifernez-Klassen et al., 2021). Similarly, *LHCBM4* is recognized by NAB1, also
458 presenting a strong RPF peak over the initiation codon, whereas LHCBM1 (not a target of
459 NAB1) has no over-proportional initiation peak (Figure 5C), despite sharing a 75% amino
460 acid sequence identity with LHCBM6. Translational regulation during *LHC* expression is
461 not unique to Chlamydomonas, as *Lhcb4.2*, *Lhcb4.3* and *Lhcb6* are also subject to post-
462 transcriptional regulation in Arabidopsis (Floris et al., 2013). We searched all transcripts
463 that had a gene body coverage of at least 70% (8757 nucleus-encoded, 66 chloroplast-
464 encoded, 6 mitochondria-encoded) for the occurrence of comparable initiation peaks
465 (average coverage of the first seven codons at least four times higher than the average
466 coverage of the remaining CDS). Around 38% of selected nucleus-encoded transcripts
467 had comparable RPF accumulation peaks near the initiation codon, possibly hinting to
468 more co-translational regulation in the cytosol than previously anticipated. In contrast, only
469 10% (6 transcripts) of the chloroplast-encoded genes showed such peaks (Figure 5D).
470 However, it cannot be excluded that these peaks present artifacts of CHX treatment of the
471 Chlamydomonas cultures during harvest. Upon establishment of the Ribo-seq protocol,
472 several reports demonstrated that pre-incubation of yeast cultures with CHX may create
473 a bias by inducing increased RPF accumulation at initiation codons, which could be the
474 consequence of stalled elongation but unaffected initiation through CHX treatment (Ingolia
475 et al., 2009; Gerashchenko and Gladyshev, 2014; Sharma et al., 2021). However, a recent

476 direct comparison of CHX-treated and untreated conditions demonstrated that CHX
477 causes no initiation-peak bias in human cells, and that even the strong accumulation in
478 yeast might derive from slightly altered protocols between laboratories (Sharma et al.,
479 2021). For Chlamydomonas, CHX biases should result in general initiation peaks, but
480 these peaks are only found in a fraction of transcripts, independent of their translation
481 activity (e.g. *LHCs*). Thus, varying accumulation of RPFs around the initiation codon may
482 be an interesting indicator for altered initiation regulation between transcripts in the
483 cytosol.

484 We further assayed, if the RPFs of various sizes, ranging from 24 to 34 nt, are
485 randomly distributed across all chloroplast CDS or if a respective RPF size is transcript
486 specific. In general, the dominant RPF size is 28 nt long and the trimodal size distribution
487 appears widely present for translation of most chloroplast transcripts, even if the
488 overrepresentation of the highly translated transcripts (i.e. *rbcL, psbA)* is normalized out
489 (Figure 6A). While present for most chloroplast transcripts, accumulation of the different
490 RPFs sizes varies between transcripts with some over-representation of smaller or larger
491 RPFs, respectively. For example, *psbA, psbH, psaC, psaJ, petD* and *atpH* have an
492 additional strong peak at 24 nt. In contrast, *rpl20* and *psbJ* have a shift towards a dominant
493 population of RPFs with a size >30 nt (Figure 6B-D). Strikingly, the small RPFs within
494 *psbA* and *atpH* show in interesting patter with increased accumulation in the regions of
495 highest RPF coverage. For *pbsA,* the 24 nt RPFs are most prominent in a section between
496 codon 135 and 150, which is just downstream of the transcript encoding the second
497 transmembrane segment (TMS, Figure 6E). It appears as if elongation severely slows
498 down within this region, favoring the extraction of small RPFs, possibly resulting from
499 ribosomes in an “open-ratchet” formation. Alternatively, the smaller reads may result from

500 local binding of a regulatory factor to the transcript, although the small RPFs have similar
501 features as the longer RPFs. Interestingly, we just recently reported that deletion of the
502 One-Helix-Protein 2, OHP2, locally affects *psbA* translation between initiation and the
503 coding sequence of the 3rd TMS. However, ribosome occupancy downstream of the 3rd
504 TMS appeared normal in the mutant. In addition, the absence of OHP2 results in a rapid
505 degradation of nascent PsbA/D1 protein (Wang et al., 2023). Thus, the observed local
506 translation slow-down might be an important stage for PsbA/D1 biogenesis and co-factor
507 integration and OHP2 might be the driver of this regulation. Similarly, RPF accumulation
508 peaks over the CDS of the first TMS of *atpH,* mostly deriving from fragments with a size
509 of 24-25 nt, again suggesting that translation elongation is reduced here (Figure 6E). It
510 will be interesting, to uncover in future experiments if the local translation slowdown is
511 accompanied or caused by binding of regulatory factors.

512 *Suitability of ribosome profiles for improving Chlamydomonas genome annotation*

513 Lastly, we aimed to exploit our Ribo-seq data for surveying correct CDS
514 annotations and transcript splice variants within the recently released Chlamydomonas
515 genome, version 6.1 (Craig et al., 2022). In previous genome versions, several genes
516 were falsely annotated, ignoring upstream start codons that were possibly the genuine
517 initiation site of translation (Cross, 2015; Craig et al., 2022). In the new version, the first
518 in-frame ATG codon was generally considered as the initiation codon, in accordance with
519 the scanning model of initiating ribosomes (Hinnebusch, 2011). Systematic comparison
520 of the reference start site with the onset of RPF coverage, confirmed that start codon
521 assignment has been clearly improved between genome version 5.6 and 6.1. However,
522 we detected gene models in which the first ATG appears not to be the genuine initiation
523 site. For example, transcripts encoding LHCA4 (Cre10.g452050) do not initiate from the

524 first ATG (Figure 7A) and were correctly annotated in the previous genome versions.
525 Another example is PRPL3 (Cre10.g417700) (Figure 7B). In addition, we determined
526 some gene models in which RPF accumulation did not correlate with the cognate ATG
527 start codon, possibly hinting to a translation start from non-cognate start sites as
528 exemplified by the *PDI2* CDS (Cre01.g033550, Figure 7C). In this case, the ribosome
529 profile starts with a prominent peak upstream of the annotated start codon located within
530 the annotated 5'-UTR while the annotated start codon is poorly covered compared with
531 the remaining CDS. Here, initiation from the non-cognate TTG could be an explanation
532 (Cao and Slavoff, 2020), which is located in-frame with the annotated coding sequence at
533 two positions within the area covered by the peak. Ribo-seq was already used before to
534 mine for non-cognate start sites in other organisms (Brar and Weissman, 2015).
535 Alternatively, these RPFs could point to the presence of upstream Open Reading Frames
536 for *PDI2*. *PDI2* represents a good example for detecting alternative splicing. Transcript 1
537 of *PDI2*, possesses a RPF gap between exon 3 and 4 (379 nt downstream of the
538 transcription start site), which is spliced out in transcript versions 2-4 (Figure 7C). In
539 addition, a sudden drop of RPFs is seen at the end of exon 3 in transcript 2 and 4. Based
540 on distribution of RPFs and the transition of ribosome profiles, transcript 5 has the most
541 consistent profile and might hence be the prevalently expressed variant in
542 Chlamydomonas cells, grown under standard laboratory conditions. In future, intelligent
543 modelling and peak deconvolution algorithms could be potentially applied to Ribo-seq
544 data to estimate the stoichiometric ratio of different splice variants' translation output.

545 *Concluding remarks*

546 Taken together, ribosome profiling is a fascinating tool for advancing our understanding
547 of post-transcriptional regulation of gene expression. Our Ribo-seq approach has the

548 resolution and depth to deeply study translation of the nuclear and chloroplast transcripts
549 in Chlamydomonas cells. Furthermore, a highly comparable Ribo-Seq protocol was
550 established for Arabidopsis and tobacco by our team, which will help to directly compare
551 translation between different land plants and Chlamydomonas (Ting et al., 2023).
552 Chlamydomonas is a well-suited organism for understanding system level changes of
553 gene expression throughout the cell cycle or following environmental changes. Ribosome
554 profiles can be of specific help, for a first understanding of the many transcripts, which
555 were not experimentally investigated before. To make our data easily accessible in the
556 research community, we have now linked our Ribo-seq data to the recent
557 Chlamydomonas genome CC-4532 v6.1 (https://phytozome-next.jgi.doe.gov), thereby
558 allowing individuals to conveniently browse the ribosome profiles of their genes of interest.
559 With the new genome release, the number of alternatively spliced transcripts have been
560 significantly expanded. Visualization of ribosome profiles may help to uncover the
561 dominant transcript variants, which are translated under standard laboratory conditions.
562 Certainly, additional experimental approaches are required to understand unexpected
563 features such as translation initiation from non-cognate start sites, RPFs along upstream
564 open reading frames, mechanism of translational regulation, and spatial organization of
565 protein synthesis.

566 **Materials and Methods**

567 *Cell growth and harvest*

568 *Chlamydomonas reinhardtii* CC-1690 cells were grown mixotrophically in Tris-Acetate
569 Phosphate (TAP) medium (Kropat et al., 2011) to mid-log phase (4 to $5x10^6$ cells per mL)
570 under constant light of 80 mmol m^{-2} s^{-1} (MASTER LEDtube HF 1,200 mm UO 16W830 T8
571 and 16W840 T8, Philips) and 25°C on a rotary shaker. Immediately before harvest, 100

572 µg/mL chloramphenicol (CAP) and cycloheximide (CHX) was added and cells were rapidly
573 chilled by pouring over -80°C cold silicon ice cubes until 4°C culture temperature was
574 reached. Subsequently, cells were pelleted by 2 min centrifugation at 4,000 g and 4°C,
575 washed in ice-cold TKM+ buffer (50 mM acetate buffered Tris pH 8.0, 150 mM KCl, 10
576 mM $MgCl_2$, 100 µg/mL CAP, 100 µg/mL CHX) and pelleted again using the same
577 centrifugation conditions. Afterwards, the pellet was resuspended in freezing buffer (TKM+
578 supplemented with 100 mM phenyl methyl sulfonyl fluoride (PMSF) and 16% of 25x Roche
579 protease inhibitor cocktail) by slow pipetting in 1/2000 of the initial culture volume and
580 flash frozen by dripping the cell suspension into liquid nitrogen. Frozen cells were stored
581 at -80°C until further use.

582 *Ribosome profiling and RNA sampling*

583 Frozen cells were ground using a bead mill with nitrogen-cooled steel containers and
584 beads for two times 2 min at 27 Hz frequency and cooling in liquid nitrogen between
585 rounds. The cell powder was mixed with equal volumes of 2x concentrated lysis buffer (2x
586 concentrated TKM+ buffer supplemented with 2 mM dithiotreitol (DTT), 2% Triton X-100
587 and 20% sucrose) and incubated for 5 min at 4°C before pelleting cell debris at 8,000 g,
588 4°C for 10 min. For nucleolytic digest conditions i. and ii., lysate was layered on a 2.5 mL
589 64% sucrose cushion prepared in TKM+ buffer supplemented with 1 mM DTT and
590 ultracentrifuged for 3 h at 60,000 rpm and 4°C in a Ti70 rotor. The resulting polysome
591 pellet was briefly rinsed with 500 µL ice-cold diethyl pyro carbonate (DEPC)-treated water
592 and resuspended overnight in 100 µL ice-cold ribosome buffer (2x concentrated TKM+
593 buffer diluted with 25x concentrated Roche protease inhibitor cocktail to a total amount of
594 16%, 1 mM PMSF and 0.1% Triton X-100). Remaining insoluble debris were removed by
595 centrifugation for 1 min at 1,500 g before addition of 1 U of RNase I (Ambion) per µg RNA

596 and 0.134 U TURBO DNase (Invitrogen). Nucleolytic digest was performed for 1 h at 4°C
597 (condition i.) or 23°C (condition ii.) and stopped by addition of 0.4 U per applied unit of
598 RNase I of SUPERase•In RNase Inhibitor (Invitrogen). For condition iii., 1 U of Ambion
599 RNase I (Invitrogen) per µg RNA and 0.134 µL per µL lysate of TURBO DNase (Invitrogen)
600 was directly added to the lysate and incubated for 1 h at 4°C. The digest was stopped by
601 addition of 0.4 U of SUPERase•In RNase Inhibitor (Invitrogen) per applied unit of RNase
602 I. The lysate was centrifuged again for 10 min at 8,000 g and 4°C to remove any cell debris
603 that may have precipitated during the nuclease digest. For all conditions, monosomes
604 were pelleted through a 750 µL 30% sucrose cushion prepared in TKM+ buffer
605 supplemented with 1 mM DTT for 30 min in a S150AT rotor at 72,000 rpm and 4°C. The
606 resulting ribosome pellets were resuspended in 100 µL of ice-cold ribosome buffer, then
607 supplemented with 15 mM ethylene diamine tetra-acetic acid (EDTA), pH 8.0 and instantly
608 mixed with 750 µL Invitrogen TRIzol reagent (Invitrogen) and incubated at room
609 temperature for 5 min. The solution was mixed with 150 µL chloroform and incubated for
610 2 min at room temperature before 15 min centrifugation at 20,000 g and 4°C. The
611 supernatant was mixed with 3 µL GlycoBlue co-precipitant (Invitrogen), 1:10 volume of 3
612 M sodium acetate, pH 5.5, 1 volume isopropanol and incubated overnight at -20°C. The
613 precipitated RNA was pelleted by 1 h centrifugation at 20,000 g and 4°C, briefly washed
614 with ice-cold 80% ethanol and centrifuged again for 30 min at the same parameters. The
615 resulting RNA pellet was briefly dried and dissolved in 11 µL of nuclease-free water.

616 For RNA-seq samples, 10 mL of culture were harvested by 2 min centrifugation at
617 4,000 g at room temperature (before harvesting the culture for RPF isolation), shock-
618 freezing in liquid nitrogen and storage at -80°C. Samples were lysed in prewarmed (50°C)
619 lysis buffer (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 20 mM EDTA pH 8.0 and 2% SDS)

620 and incubated for 2 min at 50°C. Immediately after incubation, 500 µL of TRIzol
621 (Invitrogen) reagent were added and the mixture was incubated for 5 min at room
622 temperature. The suspension was mixed with 200 µL chloroform and incubated for
623 another 5 min at room temperature before 3 min centrifugation at 12,000 g and room
624 temperature. The nucleic acid containing phase was transferred to a fresh tube and mixed
625 with 1.5 volumes of ethanol. RNA was purified by the Macherey Nagel NucleoSpin RNA
626 Plant kit.

627 *Ribosome protected fragment extraction and rRNA depletion*

628 50 µg RNA from the ribosomal extractions were diluted to a volume of 20 µL with nuclease-
629 free water, mixed with 2x formamide RNA loading buffer (90% deionized formamide, 20
630 mM Tris-HCl, pH 7.5, 0.5 M EDTA, 0.04% bromophenol blue in ethanol) and denatured
631 for 90 s at 80°C. The samples were instantly put on ice and then loaded onto 12% urea-
632 TBE-gels (90 mM Tris, 9 mM boric acid, 2 mM EDTA, 12% 19:1 40% acrylamide-
633 bisacrylamide mix, 8 M urea). Gels were run at 200 V for approximately 1 h and incubated
634 with SYBR Gold nucleic acid stain (Invitrogen), diluted 1:10,000 in TBE buffer (90 mM
635 Tris, 9 mM boric acid, 2 mM EDTA). RPF were excised from the gels in a size range from
636 20 – 40 nt according to a RNA size marker. Gel pieces were crushed and incubated
637 overnight with 400 µL RPF elution buffer (300 mM sodium acetate, pH 5.5, 0.25% sodium
638 dodecyl sulfate and 1 mM EDTA, pH 8.0) at room temperature. The eluate was collected,
639 mixed with 3 µL GlycoBlue co-precipitant, 1 volume of isopropanol and incubated
640 overnight at -20°C. The precipitated RNA was pelleted by 1 h centrifugation at 20,000 g
641 and 4°C, briefly washed with ice-cold 70% ethanol and centrifuged again for 30 min at the
642 same parameters. The resulting RNA pellet was briefly dried and dissolved in 20 µL of
643 nuclease-free water. For Ribo-seq samples, abundant rRNA fragments were depleted by

644 incubating 100 - 200 ng of ribosome RPFs with 1.19 µL of 100 µM biotinylated rRNA-oligo
645 depletion mix (consisting of the oligonucleotides:
646 AATATGCGTTCAAAGATTCGATGATTCACG,
647 TAGCTCTAGAATTACTACGGTTATCCGAGTA,
648 TACCCGACGCTGAGGCAGACATGCTCTTGG,
649 GATTCGTGAAGTTATCATGATTCACCGCA,
650 ACGGGATGAATCTCAGTGGATCGTAGCA,
651 CGATCTAGCCGTCTTAGAGCTAGAAGCAGG) with 4 µL formamide, 1 µL 20x
652 concentrated hybridization buffer (300 mM sodium citrate, 3 M NaCl), 2 µL 0.5 M EDTA,
653 pH 8.0 and in a total volume of 20 µL. The mixture was hybridized in a thermocycler by
654 heating up to 80°C for 5 min and subsequent cooling in 5°C steps for 2 min each until
655 35°C was reached in the device. Afterwards, 60 µL of hybridization buffer, supplemented
656 with 20% formamide was added to the tube and the sample was depleted twice for 15 min
657 using 15 µL of Dynabeads MyOne Streptavidin T1 magnetic beads (Invitrogen) freshly
658 washed and prepared according to the manufacturer's manual. The supernatant was
659 mixed with 3 µL GlycoBlue co-precipitant, 2.5 volumes of ethanol and incubated overnight
660 at -20°C. RPFs were pelleted by 1 h centrifugation at 20,000 g and 4°C, briefly washed
661 with ice-cold 70% ethanol and centrifuged again for 10 min at the same parameters. The
662 resulting pellet was briefly dried and dissolved in 41 µL of nuclease-free water. Excess
663 oligos were removed by 30 min DNase digest at 37°C upon addition of 5 µL 10x TURBO
664 DNase buffer, 2 µL Invitrogen TURBO DNase and 2 µL SUPERase•In RNase Inhibitor
665 (Invitrogen). After digest, RPFs were mixed with 150 µL of 100 mM NaCl and purified
666 again by phenol-chloroform precipitation and final precipitation overnight at -20°C using
667 2.5 volumes of ethanol and 2 µL GlycoBlue co-precipitant. RPFs were pelleted by 1 h

668 centrifugation at 20,000 g and 4°C, briefly washed with ice-cold 70% ethanol and
669 centrifuged again for 10 min at the same parameters. The resulting pellet was briefly dried
670 and dissolved in 36.5 µL of nuclease-free water.
671 For sequencing library preparation, ribosome RPFs were denatured at 65°C for 5 min in
672 a thermocycler and immediately transferred to ice. The sample was then mixed with 5 µL
673 10 x T4-polynucleotide kinase buffer A, 2.5 µL T4-polynucleotide kinase (Thermo
674 Scientific), 2 µL SUPERase•In RNase Inhibitor and incubated for 10 min at 37°C.
675 Subsequently, 5 µL of 10 mM ATP was added and the sample was incubated for additional
676 30 min at 37°C. After phosphorylation, the RPFs were mixed with 150 µL 100 mM NaCl
677 and purified again by phenol-chloroform precipitation and final precipitation overnight at -
678 20°C using 2.5 volumes of ethanol and 2 µL GlycoBlue co-precipitant. RPFs were pelleted
679 by 1 h centrifugation at 20,000 g and 4°C, briefly washed with ice-cold 70% ethanol and
680 centrifuged again for 10 min at the same parameters. The resulting pellet was briefly dried
681 and dissolved in 11.5 µL of nuclease-free water. The whole sample was used as input for
682 the NEXTFLEX Small RNA-seq Kit v3 (Perkin Elmer) to generate libraries according to
683 the kit's manual. For RNA-seq, the purified RNA was later used as input for the Zymo-Seq
684 RiboFree Total RNA library kit for library preparation according to the manufacturer's
685 manual.
686 *Processing of raw sequencing data*
687 Demultiplexed FASTQ files were derived from an Illumina NextSeq 550 system aiming for
688 40 M reads per Ribo-seq sample and 20 M reads per RNA-seq sample at a read length
689 of 75 nt. Samples were processed through a custom-made pipeline in form of a Linux shell
690 script calling several bioinformatic tools and custom Python scripts. First operation of the
691 pipeline was removing 3' adapter sequences using cutadapt for RNA-seq and Ribo-seq,

692 respectively (Martin, 2011) with parameters [--minimum-length=9, -
693 a="TGGAATTCTCGGGTGCCAAGG"]. After removal of 3' adapter sequences, the 3' and
694 5' unique molecular identifier tags (UMI tags) were removed from each read's sequence
695 and shifted to the rear of the header row by UMI tools (Smith et al., 2017) with parameters
696 [extract --bc-pattern=NNNN] for 5' UMI tags and [extract --bc-pattern=NNNN –3prime] for
697 3' UMI tags. Remaining reads were filtered for sequence length using cutadapt again with
698 parameters [--minimum length=20 --maximum-length=39] to remove any read that is
699 unlikely to represent a real RPF. After size filtering, a custom-made Python script
700 corrected the UMI-tags that were previously shifted to the read headers in the form
701 [header line_NNNN_NNNN] to [header line_NNNNNNNN], allowing UMI tools in a later
702 step to deduplicate the data set based on both UMI tags. In a first mapping step, reads
703 were aligned to a set of *C. reinhardtii* non-coding RNA FASTA sequences provided by
704 ENSEMBL plants together with the genome version 5.5 to remove contaminating RNA
705 species using STAR with parameters [--runMode alignReads --outSAMtype BAM
706 SortedByCoordinate --outFilterMultimapNmax 20 --outFilterMismatchNmax 3 --
707 outReadsUnmapped Fastx --alignIntronMax 7000 --twopassMode Basic]. Reads that did
708 not align to the ncRNA set were copied to a new FASTQ file that was mapped against the
709 *C. reinhardtii* genome version 6.1 using STAR (Dobin et al., 2013) with parameters [--
710 runMode alignReads –outSAMtype BAM
711 SortedByCoordinate --outFilterMismatchNoverLmax 0.1 --outReadsUnmapped Fastx --
712 alignIntronMax 8000 –outSAMmultNmax 1 –outMultimapperOrder Random]. After
713 genome mapping, the resulting BAM file was deduplicated using UMI-tools' dedup
714 function and afterwards indexed using samtools' index function (Li et al., 2009).
715 *Data processing*

716 All analyses on the NGS data sets including calculations of gene body coverage,
717 periodicity, P-site offsets, organellar read length distributions, ribosome profiles and gene
718 expression metrics were performed using a custom-made object-oriented Python module
719 that makes extensive use of the packages pysam/HTSlib (https://github.com/pysam-
720 developers/pysam), pandas (McKinney, 2010), NumPy (Harris et al., 2020), SciPy
721 (Virtanen et al., 2020), matplotlib (Hunter, 2007) and seaborn (Waskom, 2021) among
722 multiple others. For RPKM calculations in Ribo-seq data sets, only reads mapping to the
723 annotated CDS of a transcript were considered. Alternatively we calculate counts per
724 million values for the interested reader (Supplemental Dataset). For any calculations on
725 RNA-seq data sets, reads mapping to the whole transcript were considered and the full
726 length of the mature transcript was used for RPKM calculations. Translation efficiency
727 values were simply calculated by dividing an CDSs RPKM value from a Ribo-seq data set
728 by the respective transcripts RPKM value from the affiliated RNA-seq data set.

729 For the 5'-P-site / 3'-A-site offset calculations, all reads fully enclosing either an
730 annotated start codon (5'-P-site offset) or an annotated stop codon (for 3'-A-site offset)
731 were taken into consideration separately for the nuclear and the chloroplast genome.
732 These reads were sorted according to the length of their alignment to the genome, which
733 was considered as read length to avoid miscalculations due to clipped or masked
734 sequences. For every read length species, the position of the reads 5' end or 3' end
735 relative to the position of the first base of the start codon (for 5'-P-site offsets) or the last
736 base of the stop codon (for 3'-A-site offsets) was determined as the offset. Afterwards, the
737 most frequently occurring offset for a read length species was determined to be the true
738 offset of this species and was used in the P-site calculation. To display the results of the

739 offset calculations, the count matrix of both 5'-P-site offsets and 3'-A-site offsets were
740 normalized separately to the total number of counts each and plotted as heatmap.
741 For calculation of the P-sites first position of reads mapping to nucleus-encoded
742 CDSs, all reads mapping to a CDS were filtered from the data set and for every read, the
743 5'-most mapping position of the read's alignment + the 5'-P-site offset determined for the
744 reads alignment length was determined to be the first P-site position in case of transcripts
745 encoded on the + strand. For reads encoded on the minus strand, the first P-site position
746 was determined to be the 3'-most mapping position of the RPF alignment minus the
747 determined 5'-P-site offset. In all cases, splicing was considered and whenever a spliced
748 read was found, the calculation was adjusted according to the number of genomic
749 positions spanned by the splice junction of the read. For reads mapping to chloroplast-
750 encoded transcripts on the + strand, the first P-site position was determined to be the 3'-
751 most mapping position – (3'-A-site offset + 5). For reads mapping to chloroplast-encoded
752 transcripts on the - strand, the first P-site position was calculated to be the 5'-most
753 mapping position + (3'-A-site offset + 5). The metagene P-site profiles of nucleus-encoded
754 transcripts were calculated considering only the 2,000 most expressed transcripts of a
755 dataset, according to their RPKM values. If multiple splice variants of the same gene
756 occurred in the list, only the longest variant (annotated as transcript 1 in the genome) was
757 considered for the calculation to avoid analyzing the same reads multiple times. The
758 profiles were constructed by calculating the first position of the P-site for all reads mapping
759 to the first 30 and the last 30 codons of considered transcript CDSs. For a ribosome RPF
760 to be in-frame with the annotated CDS, the first position of the P-site must be located on
761 the first base of a codon. According to this logic, the calculated P-site positions were
762 categorized into "in-frame", "+1 shifted" and "+2 shifted" and summed up for all codons of

763 considered transcript CDSs. To calculate the frame preference per codon, the counts in
764 each category in a codon were divided by the sum of all categories for every codon
765 separately. For calculation of the chloroplast metagene P-site profile, all protein-coding
766 chloroplast transcripts were considered.

767 The extension of gene body coverage versus a respective cut-off was calculated
768 as the portion of an CDS that was covered by the number of reads equal to or greater
769 than the respective cut-off. This procedure was applied to all annotated CDSs using
770 increasing cut-offs from 1 to 10 reads to monitor the stepwise decrease of gene-body
771 coverage. For every applied cut-off, the CDSs were categorized as being covered over
772 >10% of their length to >90% of their length in 10% steps and CDSs falling into every
773 category were counted and displayed in a heat map to estimate both completeness and
774 strength of CDS coverage in the data set.

775 To calculate the read length distribution, all reads mapping to the respective
776 genome were filtered and the length of their alignment to the genome was taken as read
777 length. After counting the number of all possible read lengths for each genome separately,
778 these numbers were normalized by the total number of reads mapping to the respective
779 genome. For read biotype distribution, the reads were further categorized if they map to
780 an annotated coding sequence, 5'- or 3'-UTR or intergenic region. To handle reads
781 overlapping two regions of different category, for example 5'-UTR and coding sequence
782 as seen regularly in initiation peaks, the read was assigned the category that had the
783 larger share on the read's alignment. In rare cases, when a read mapped to different
784 categories in two different splice variants of the same transcript, the read was assigned to
785 both categories.

786 **Supplemental data**

787 **Supplemental Figure S1:** Features of ribosome protected fragment coverage.

788 **Supplemental Figure S2:** RPF length distribution and periodicity.

789 **Supplemental Figure S3:** Transcript accumulation, translation output and translation
790 efficiency of nuclear genes.

791 **Supplemental Figure S4:** Correlation between nuclease digestion conditions, transcript
792 accumulation and translation.

793 **Supplemental Figure S5:** Correlation analysis of RPF count between nuclease digest
794 conditions.

795 **Supplemental Figure S6:** Detection of abnormal initiation peaks in the three genomes of
796 *Chlamydomonas reinhardtii.*

797 **Supplemental Dataset.** List of accession numbers, gene body coverage, RPKM and
798 CPM of Ribo-seq and RNA-seq experiments.

799 **Acknowledgements**

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

805 **Author contribution:**

806 V.L.G. performed ribosome profiling experiments, bioinformatic data analyses and wrote
807 parts of the manuscript. M.K.Y.T. and R.Z. helped with setting up ribosome profiling, N.H.

808 and S.R. conducted correlation analyses. F.W. designed experiments, analyzed data, and
809 wrote the manuscript.

810 **Funding**

811 This work was supported by the Deutsche Forschungsgemeinschaft grant 437345987 (WI
812 3477/3-1, ZO 302/5-1, RU 2266/2-1 to F.W., R.Z. and S.R., respectively) and the
813 Forschungsschwerpunkt BioComp to F.W.

814 **Data availability**

815 All gene accession codes are given in the Supplemental Dataset and are in accordance
816 with the latest Chlamydomonas genome v6.1 annotation (https://phytozome-
817 next.jgi.doe.gov) or the GenBank/EMBL data libraries. The full Ribo-seq and RNA-seq
818 data are accessible via the Dataplant consortium.

819 **Competing interests**

820 The authors declare no competing interests.

821 **References**

822 **Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C, Jaffe DB, Nusbaum**
823 **C, Gnirke A** (2011) Analyzing and minimizing PCR amplification bias in Illumina
824 sequencing libraries. Genome Biol **12:** R18
825 **Blifernez-Klassen O, Berger H, Mittmann BGK, Klassen V, Schelletter L, Buchholz**
826 **T, Baier T, Soleimani M, Wobbe L, Kruse O** (2021) A gene regulatory network
827 for antenna size control in carbon dioxide-deprived *Chlamydomonas reinhardtii*
828 cells. Plant Cell **33:** 1303-1318
829 **Bonente G, Pippa S, Castellano S, Bassi R, Ballottari M** (2012) Acclimation of
830 *Chlamydomonas reinhardtii* to different growth irradiances. J Biol Chem **287:** 5833-
831 5847
832 **Brar GA, Weissman JS** (2015) Ribosome profiling reveals the what, when, where and
833 how of protein synthesis. Nat Rev Mol Cell Biol **16:** 651-664
834 **Calviello L, Mukherjee N, Wyler E, Zauber H, Hirsekorn A, Selbach M, Landthaler M,**
835 **Obermayer B, Ohler U** (2016) Detecting actively translated open reading frames
836 in ribosome profiling data. Nat Methods **13:** 165-170

Cao X, Slavoff SA (2020) Non-AUG start codons: Expanding and regulating the small and alternative ORFeome. Exp Cell Res **391:** 111973

Chartron JW, Hunt KC, Frydman J (2016) Cotranslational signal-independent SRP preloading during membrane targeting. Nature **536:** 224-228

Chen H, Alonso JM, Stepanova AN (2022) A Ribo-Seq Method to Study Genome-Wide Translational Regulation in Plants. Methods Mol Biol **2494:** 61-98

Chiu CW, Li YR, Lin CY, Yeh HH, Liu MJ (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 Genet **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 Genet **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. Proc Natl Acad Sci U S A **117:** 21775-21784

Chung BY, Hardcastle TJ, Jones JD, Irigoyen N, Firth AE, Baulcombe DC, Brierley I (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, Szczesniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A (2016) A survey of best practices for RNA-seq data analysis. Genome Biol **17:** 13

Costanzo MC, Hogan JD, Cusick ME, Davis BP, Fancher AM, Hodges PE, Kondu P, Lengieza C, Lew-Smith JE, Lingner C, Roberg-Perez KJ, Tillberg M, Brooks JE, Garrels JI (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 Res **28:** 73-76

Craig RJ, Gallaher SD, Shu S, Salome P, Jenkins JW, Blaby-Haas CE, Purvine SO, O'Donnell S, Barry K, Grimwood J, Strenkert D, Kropat J, Daum C, Yoshinaga Y, Goodstein DM, Vallon O, Schmutz J, Merchant SS (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

Cross FR (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. Biochim Biophys Acta **1847:** 861-871

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics **29:** 15-21

Döring K, Ahmed N, Riemer T, Suresh HG, Vainshtein Y, Habich M, Riemer J, Mayer MP, O'Brien EP, Kramer G, Bukau B (2017) Profiling Ssb-Nascent Chain Interactions Reveals Principles of Hsp70-Assisted Folding. Cell **170:** 298-311 e220

Eberhard S, Drapier D, Wollman FA (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, Alboresi A, Lanet E (2013) Post-transcriptional control of light-harvesting genes expression under light stress. Plant Mol Biol **82:** 147-154

Forsythe ES, Grover CE, Miller ER, Conover JL, Arick MA, 2nd, Chavarro MCF, Leal-Bertioli SCM, Peterson DG, Sharbrough J, Wendel JF, Sloan DB (2022) Organellar transcripts dominate the cellular mRNA pool across plants of varying ploidy levels. Proc Natl Acad Sci U S A **119:** e2204187119

Frank J, Agrawal RK (2000) A ratchet-like inter-subunit reorganization of the ribosome during translocation. Nature **406:** 318-322

Fu Y, Wu PH, Beane T, Zamore PD, Weng Z (2018) Elimination of PCR duplicates in RNA-seq and small RNA-seq using unique molecular identifiers. BMC Genomics **19:** 531

Fujita T, Kurihara Y, Iwasaki S (2019) The Plant Translatome Surveyed by Ribosome Profiling. Plant Cell Physiol **60:** 1917-1926

Gallaher SD, Fitz-Gibbon ST, Strenkert D, Purvine SO, Pellegrini M, Merchant SS (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. Plant J **93:** 545-565

Gao Y, Thiele W, Saleh O, Scossa F, Arabi F, Zhang H, Sampathkumar A, Kuhn K, Fernie A, Bock R, Schottler MA, Zoschke R (2022) Chloroplast translational regulation uncovers nonessential photosynthesis genes as key players in plant cold acclimation. Plant Cell **34:** 2056-2079

Gawronski P, Jensen PE, Karpinski S, Leister D, Scharff LB (2018) Plastid ribosome pausing is induced by multiple features and is linked to protein complex assembly. Plant Physiol

Gerashchenko MV, Gladyshev VN (2014) Translation inhibitors cause abnormalities in ribosome profiling experiments. Nucleic Acids Res **42:** e134

Gerashchenko MV, Gladyshev VN (2017) Ribonuclease selection for ribosome profiling. Nucleic Acids Res **45:** e6

Gloge F, Becker AH, Kramer G, Bukau B (2014) Co-translational mechanisms of protein maturation. Curr Opin Struct Biol **24:** 24-33

Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, Del Rio JF, Wiebe M, Peterson P, Gerard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE (2020) Array programming with NumPy. Nature **585:** 357-362

Hinnebusch AG (2011) Molecular mechanism of scanning and start codon selection in eukaryotes. Microbiol Mol Biol Rev **75:** 434-467, first page of table of contents

Hsu PY, Calviello L, Wu HL, Li FW, Rothfels CJ, Ohler U, Benfey PN (2016) Super-resolution ribosome profiling reveals unannotated translation events in Arabidopsis. Proc Natl Acad Sci U S A

Hunter J (2007) Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering **9:** 90-95

Ingolia NT (2014) Ribosome profiling: new views of translation, from single codons to genome scale. Nat Rev Genet **15:** 205-213

933 **Ingolia NT** (2016) Ribosome Footprint Profiling of Translation throughout the Genome.
934 Cell **165:** 22-33
935 **Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS** (2012) The ribosome
936 profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-
937 protected mRNA fragments. Nat Protoc **7:** 1534-1550
938 **Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS** (2009) Genome-wide
939 analysis in vivo of translation with nucleotide resolution using ribosome profiling.
940 Science **324:** 218-223
941 **Ingolia NT, Lareau LF, Weissman JS** (2011) Ribosome profiling of mouse embryonic
942 stem cells reveals the complexity and dynamics of mammalian proteomes. Cell
943 **147:** 789-802
944 **Juntawong P, Girke T, Bazin J, Bailey-Serres J** (2014) Translational dynamics revealed
945 by genome-wide profiling of ribosome footprints in *Arabidopsis*. Proc Natl Acad Sci
946 U S A **111:** E203-212
947 **Klimmek F, Sjodin A, Noutsos C, Leister D, Jansson S** (2006) Abundantly and rarely
948 expressed *Lhc* protein genes exhibit distinct regulation patterns in plants. Plant
949 Physiol **140:** 793-804
950 **Komine Y, Kwong L, Anguera MC, Schuster G, Stern DB** (2000) Polyadenylation of
951 three classes of chloroplast RNA in *Chlamydomonas reinhadtii*. RNA **6:** 598-607
952 **Kramer G, Shiber A, Bukau B** (2019) Mechanisms of Cotranslational Maturation of
953 Newly Synthesized Proteins. Annu Rev Biochem **88:** 337-364
954 **Kropat J, Hong-Hermesdorf A, Casero D, Ent P, Castruita M, Pellegrini M, Merchant**
955 **SS, Malasarn D** (2011) A revised mineral nutrient supplement increases biomass
956 and growth rate in *Chlamydomonas reinhardtii*. Plant J **66:** 770-780
957 **Lareau LF, Hite DH, Hogan GJ, Brown PO** (2014) Distinct stages of the translation
958 elongation cycle revealed by sequencing ribosome-protected mRNA fragments.
959 Elife **3:** e01257
960 **Lauria F, Tebaldi T, Bernabo P, Groen EJN, Gillingwater TH, Viero G** (2018) riboWaltz:
961 Optimization of ribosome P-site positioning in ribosome profiling data. PLoS
962 Comput Biol **14:** e1006169
963 **Lei L, Shi J, Chen J, Zhang M, Sun S, Xie S, Li X, Zeng B, Peng L, Hauck A, Zhao H,**
964 **Song W, Fan Z, Lai J** (2015) Ribosome profiling reveals dynamic translational
965 landscape in maize seedlings under drought stress. Plant J **84:** 1206-1218
966 **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G,**
967 **Durbin R, Genome Project Data Processing S** (2009) The Sequence
968 Alignment/Map format and SAMtools. Bioinformatics **25:** 2078-2079
969 **Liu MJ, Wu SH, Wu JF, Lin WD, Wu YC, Tsai TY, Tsai HL, Wu SH** (2013) Translational
970 landscape of photomorphogenic *Arabidopsis*. Plant Cell **25:** 3699-3710
971 **Lukoszek R, Feist P, Ignatova Z** (2016) Insights into the adaptive response of
972 *Arabidopsis thaliana* to prolonged thermal stress by ribosomal profiling and RNA-
973 Seq. BMC Plant Biol **16:** 221
974 **Martin M** (2011) Cutadapt removes adapter sequences from high-throughput sequencing
975 reads. 2011 **17:** 3
976 **McKinney W** (2010) Data Structures for Statistical Computing in Python. *In* Proceedings
977 of the 9th Python in Science Conference, pp 56-61
978 **Merchante C, Brumos J, Yun J, Hu Q, Spencer KR, Enriquez P, Binder BM, Heber S,**
979 **Stepanova AN, Alonso JM** (2015) Gene-specific translation regulation mediated
980 by the hormone-signaling molecule EIN2. Cell **163:** 684-697

Moulin M, Nguyen GT, Scaife MA, Smith AG, Fitzpatrick TB (2013) Analysis of *Chlamydomonas* thiamin metabolism in vivo reveals riboswitch plasticity. Proc Natl Acad Sci U S A **110:** 14622-14627

Mussgnug JH, Wobbe L, Elles I, Claus C, Hamilton M, Fink A, Kahmann U, Kapazoglou A, Mullineaux CW, Hippler M, Nickelsen J, Nixon PJ, Kruse O (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. 49-78

Payne SH (2015) The utility of protein and mRNA correlation. Trends Biochem Sci **40:** 1-3

Pechmann S, Willmund F, Frydman J (2013) The ribosome as a hub for protein quality control. Mol Cell **49:** 411-421

Plancke C, Vigeolas H, Hohner R, Roberty S, Emonds-Alt B, Larosa V, Willamme R, Duby F, Onga Dhali D, Thonart P, Hiligsmann S, Franck F, Eppe G, Cardol P, Hippler M, Remacle C (2014) Lack of isocitrate lyase in *Chlamydomonas* leads to changes in carbon metabolism and in the response to oxidative stress under mixotrophic growth. Plant J **77:** 404-417

Rooijers K, Loayza-Puch F, Nijtmans LG, Agami R (2013) Ribosome profiling reveals features of normal and disease-associated mitochondrial translation. Nat Commun **4:** 2886

Russell JB, Cook GM (1995) Energetics of bacterial growth: balance of anabolic and catabolic reactions. Microbiol Rev **59:** 48-62

Schneider-Poetsch T, Ju J, Eyler DE, Dang Y, Bhat S, Merrick WC, Green R, Shen B, Liu JO (2010) Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. Nat Chem Biol **6:** 209-217

Schroda M, Hemme D, Mühlhaus T (2015) The *Chlamydomonas* heat stress response. Plant J **82:** 466-480

Schuster M, Gao Y, Schöttler MA, Bock R, Zoschke R (2019) Limited Responsiveness of Chloroplast Gene Expression during Acclimation to High Light in Tobacco. Plant Physiol

Sharma P, Wu J, Nilges BS, Leidel SA (2021) Humans and other commonly used model organisms are resistant to cycloheximide-mediated biases in ribosome profiling experiments. Nat Commun **12:** 5094

Slobodin B, Dikstein R (2020) So close, no matter how far: multiple paths connecting transcription to mRNA translation in eukaryotes. EMBO Rep **21:** e50799

Smith T, Heger A, Sudbery I (2017) UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res **27:** 491-499

Sun Y, Zerges W (2015) Translational regulation in chloroplasts for development and homeostasis. Biochim Biophys Acta **1847:** 809-820

Teixeira FK, Lehmann R (2019) Translational Control during Developmental Transitions. Cold Spring Harb Perspect Biol **11**

Ting MKY, Gao Y, Barahimipour R, Ghandour R, Liu J, Martinez-Seidel F, Smirnova J, Gotsmann VL, Fischer A, Haydon MJ, Willmund F, Zoschke R (2023) Improved Plant Ribosome Profiling with Structural Assessment of rRNA Contaminants. manuscript submitted

1028 **Trösch R, Barahimipour R, Gao Y, Badillo-Corona JA, Gotsmann VL, Zimmer D,**
1029 **Mühlhaus T, Zoschke R, Willmund F** (2018) Commonalities and differences of
1030 chloroplast translation in a green alga and land plants. Nat Plants **4:** 564-575
1031 **Trösch R, Ries F, Westrich LD, Gao Y, Herkt C, Hoppstädter J, Heck-Roth J, Mustas**
1032 **M, Scheuring D, Choquet Y, Räschle M, Zoschke R, Willmund F** (2022) Fast
1033 and global reorganization of the chloroplast protein biogenesis network during heat
1034 acclimation. Plant Cell **34:** 1075-1099
1035 **Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D,**
1036 **Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M,**
1037 **Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E,**
1038 **Carey CJ, Polat I, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J,**
1039 **Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH,**
1040 **Pedregosa F, van Mulbregt P, SciPy C** (2020) SciPy 1.0: fundamental algorithms
1041 for scientific computing in Python. Nat Methods **17:** 261-272
1042 **Wang F, Dischinger K, Westrich LD, Meindl I, Egidi F, Trosch R, Sommer F, Johnson**
1043 **X, Schroda M, Nickelsen J, Willmund F, Vallon O, Bohne AV** (2023) ONE-
1044 HELIX PROTEIN 2 is not required for the synthesis of photosystem II subunit D1
1045 in Chlamydomonas. Plant Physiol
1046 **Wang R, Amoyel M** (2022) mRNA Translation Is Dynamically Regulated to Instruct Stem
1047 Cell Fate. Front Mol Biosci **9:** 863885
1048 **Waskom M** (2021) Seaborn: statistical data visualization. Journal of Open Source
1049 Software **6:** 3021
1050 **Westrich LD, Gotsmann VL, Herkt C, Ries F, Kazek T, Trosch R, Armbruster L,**
1051 **Mühlenbeck JS, Ramundo S, Nickelsen J, Finkemeier I, Wirtz M, Storchova Z,**
1052 **Räschle M, Willmund F** (2021) The versatile interactome of chloroplast ribosomes
1053 revealed by affinity purification mass spectrometry. Nucleic Acids Res **49:** 400-415
1054 **Wolin SL, Walter P** (1988) Ribosome pausing and stacking during translation of a
1055 eukaryotic mRNA. EMBO J **7:** 3559-3569
1056 **Wu CC, Zinshteyn B, Wehner KA, Green R** (2019) High-Resolution Ribosome Profiling
1057 Defines Discrete Ribosome Elongation States and Translational Regulation during
1058 Cellular Stress. Mol Cell **73:** 959-970 e955
1059 **Wu HL, Song G, Walley JW, Hsu PY** (2019) The Tomato Translational Landscape
1060 Revealed by Transcriptome Assembly and Ribosome Profiling. Plant Physiol **181:**
1061 367-380
1062 **Yang X, Cui J, Song B, Yu Y, Mo B, Liu L** (2020) Construction of High-Quality Rice
1063 Ribosome Footprint Library. Front Plant Sci **11:** 572237
1064 **Yang X, Song B, Cui J, Wang L, Wang S, Luo L, Gao L, Mo B, Yu Y, Liu L** (2021)
1065 Comparative ribosome profiling reveals distinct translational landscapes of salt-
1066 sensitive and -tolerant rice. BMC Genomics **22:** 612
1067 **Zhang W, Dunkle JA, Cate JH** (2009) Structures of the ribosome in intermediate states
1068 of ratcheting. Science **325:** 1014-1017
1069 **Zoschke R, Bock R** (2018) Chloroplast Translation: Structural and Functional
1070 Organization, Operational Control and Regulation. Plant Cell **30:** 745-770
1071

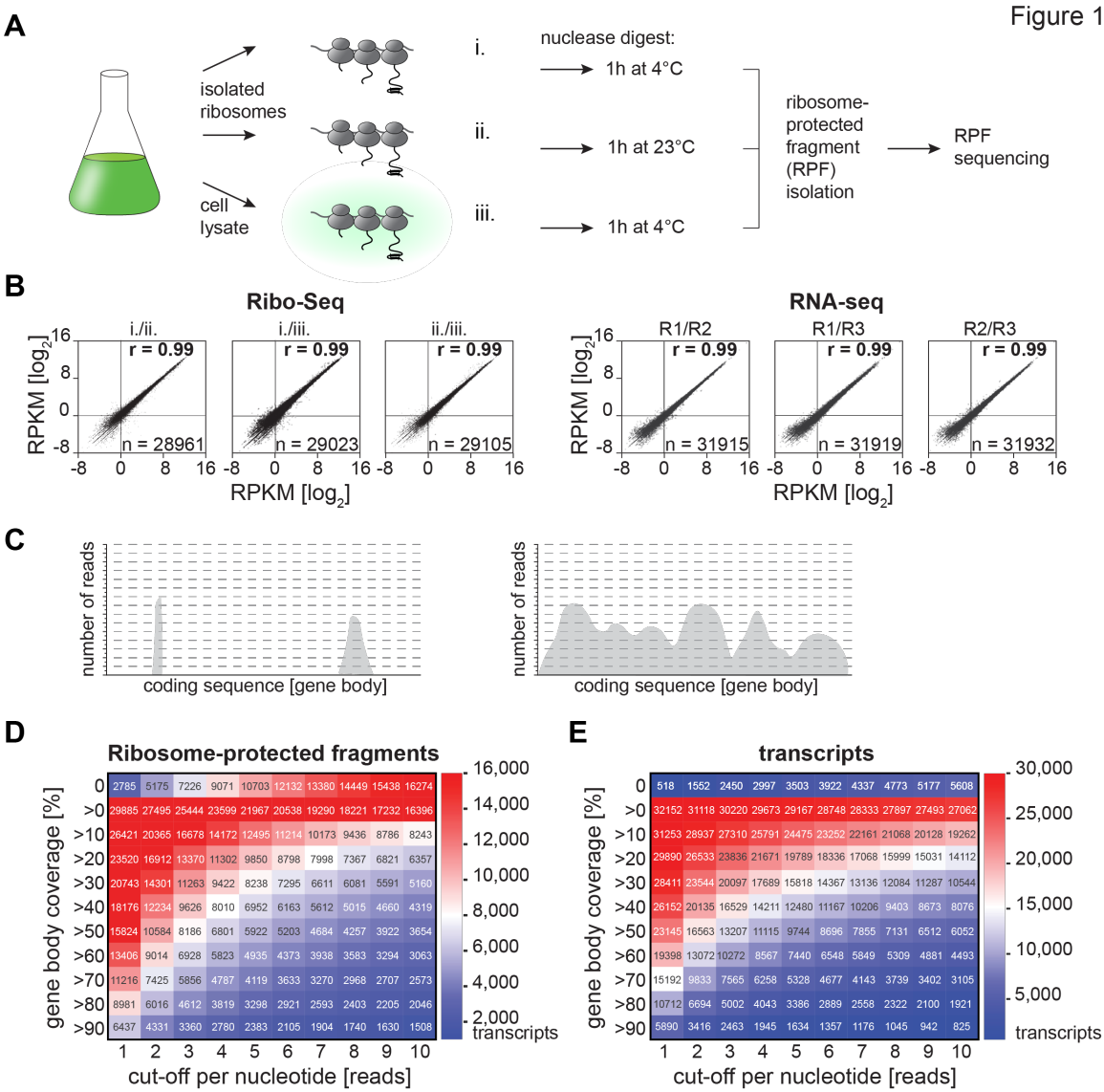


Figure 1: Ribo-seq provides a broad coverage of expressed genes
(A) Illustration of the experimental setup, testing nucleolytic conditions. Polysome purification and subsequent RNase I digest for 1h at 4°C (i.) or 1h at 23°C (ii.) was performed from Chlamydomonas samples, respectively. Alternatively, 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). Shown are the R-values and RPKM values. (C) Schematic representation of the concept of gene body coverage (left panel: incomplete coverage that strongly declines with an increasing read cut-off; right panel: good coverage that is robust against increasing read cut-offs). (D) and (E) Cumulative heatmaps visualizing 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 respective data for the other treatments are shown in Supplemental Figure 1.

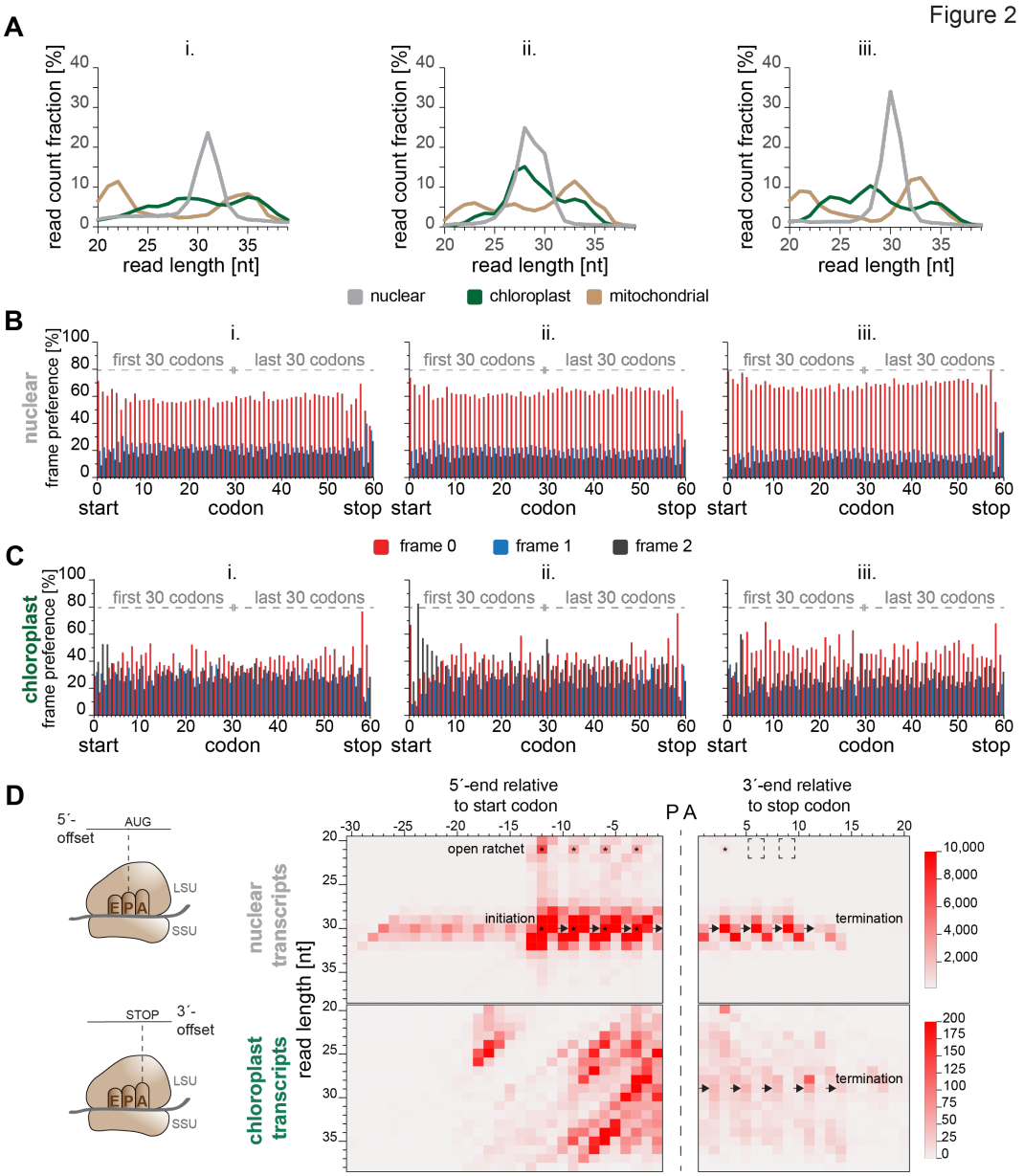


Figure 2: Tri-nucleotide periodicity of cellular translation

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

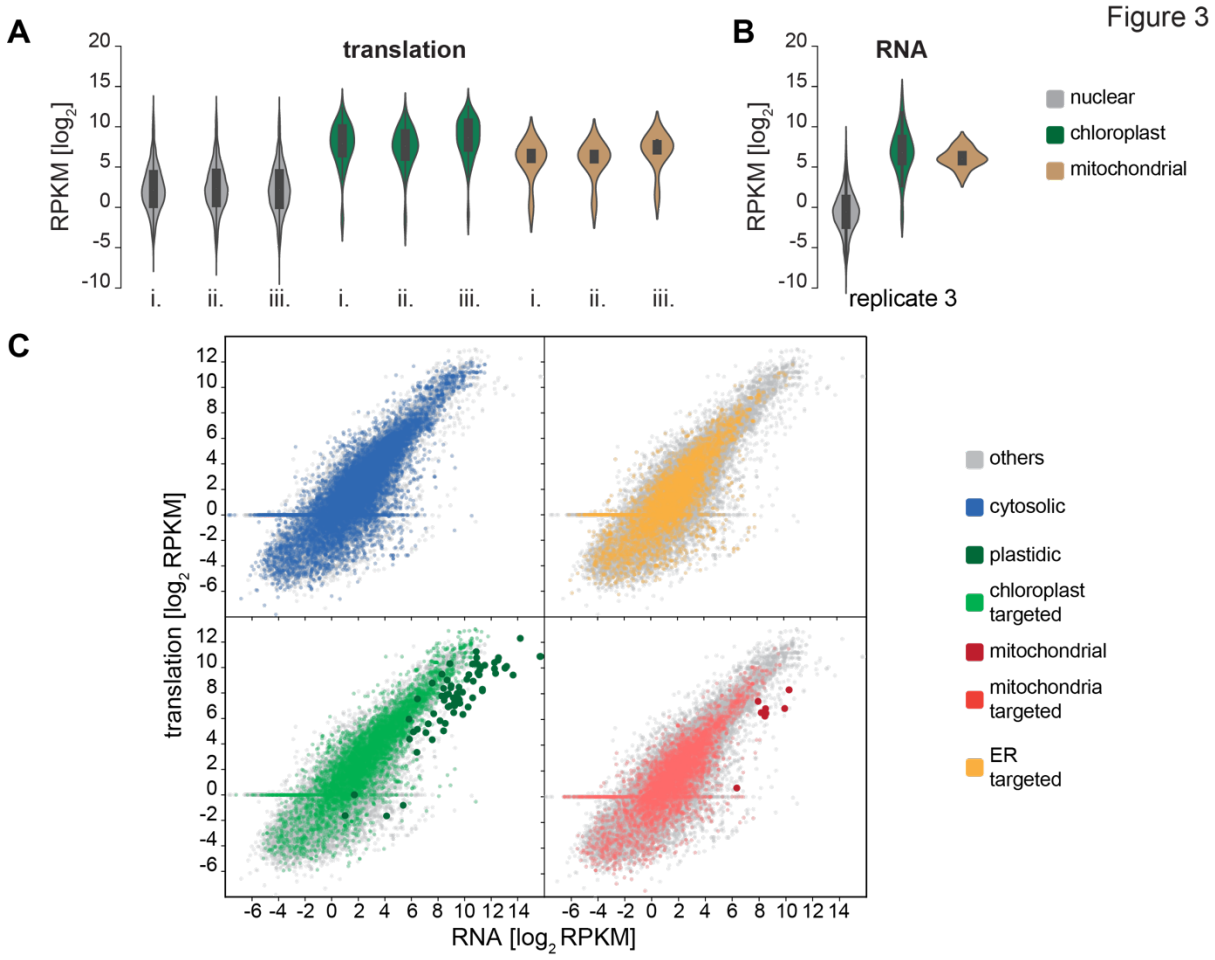


Figure 3: Chloroplast-encoded proteins are highly expressed
(A) Violin plots of the RPF data, displaying the RPKM (count per million reads) distribution for genes of each plant cell genome, following the three different RNase I treatments (i. and ii.: digest of purified ribosomes at 4°C and 23°C, respectively, iii.: digest in lysate at 4°C). Values with a RPKM cut-off <2 (log_2) are excluded. (B) Violin plots, displaying distribution of RPKM (reads per kilobase of exon per million reads mapped) for RNA values with a cut-off >2 (log_2) for each plant cell genome in replicate 3 exemplary. (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.

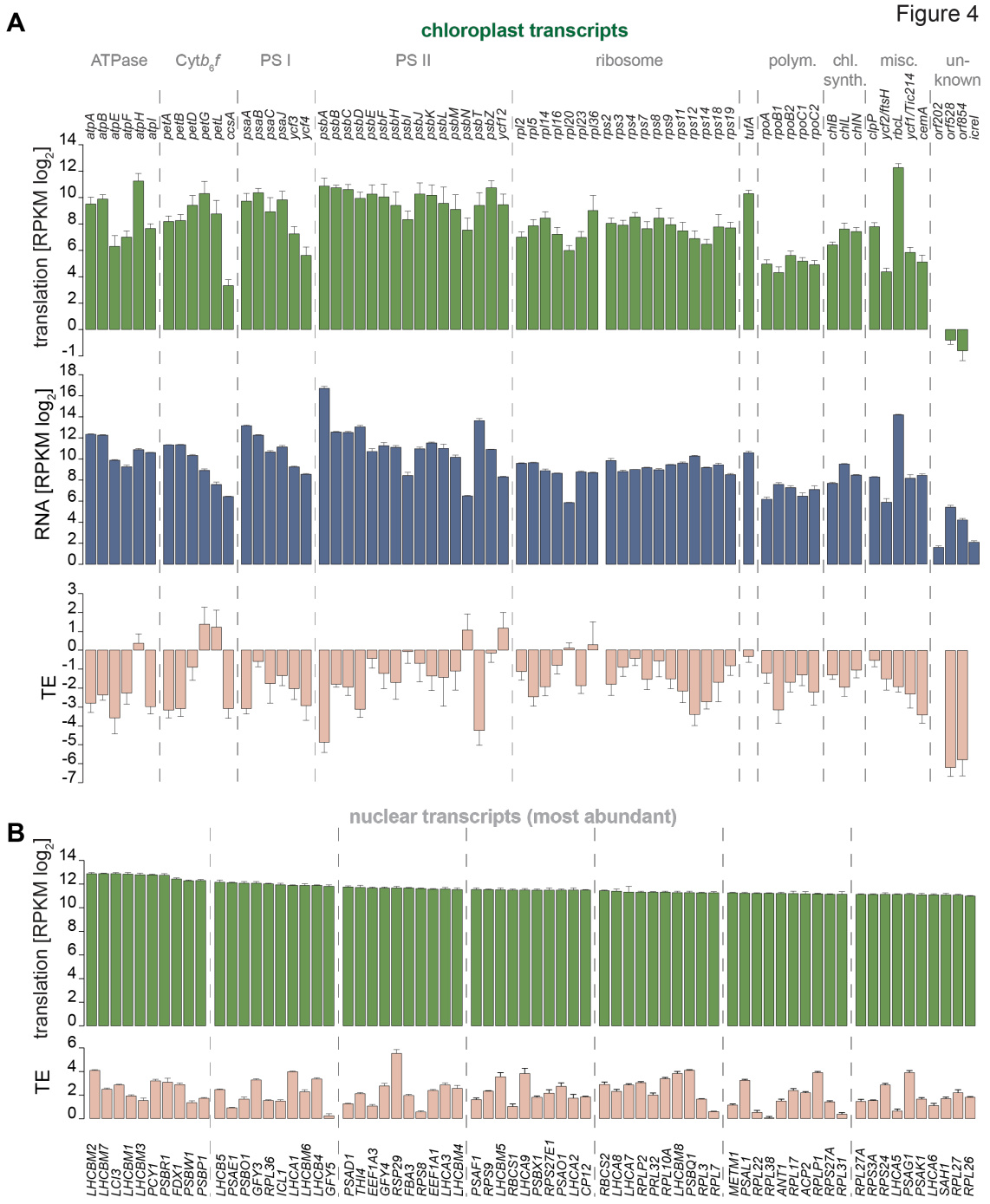


Figure 4: Translation output of chloroplast- and nucleus-encoded transcripts
(A) Mean read intensity/TE ratio (per respective coding sequence) of chloroplast-encoded transcripts on the level of RPF, RNA and translation efficiency (TE). The transcripts are grouped by function of the encoded proteins (Cytb_6f= cytochrome b_6f complex, PS I = photosystem I, PS II = photosystem II, polym. = RNA polymerase, chl. synth. = chlorophyll synthesis, misc. = miscellaneous). (B) 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 Supplementary Figure S3C. 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 unidirectional to improve readability.

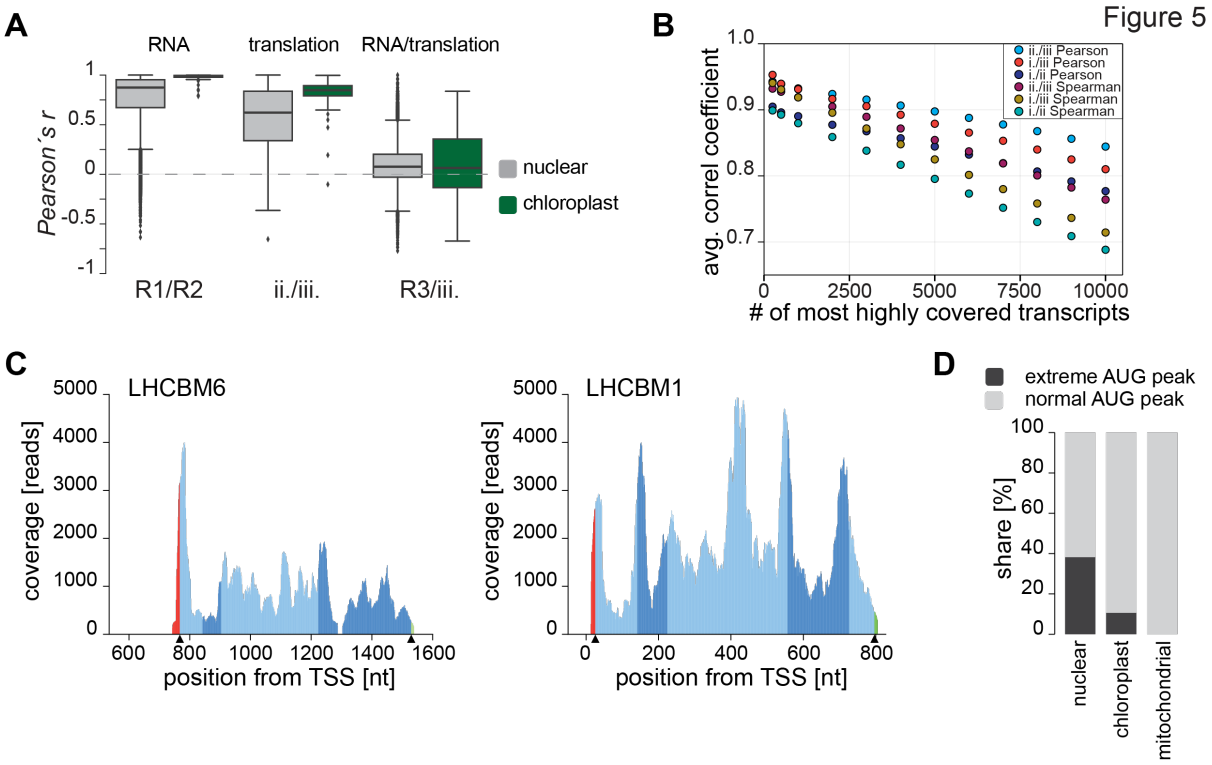


Figure 5: Putative translational regulation of individual transcripts
(A) Boxplots representing the distribution of Pearson's *r* correlation coefficients comparing the RNA gene body coverage and ribosome profiles of all chloroplast- and nucleus-encoded transcripts between replicates (R1 and R2 for RNA) and between two types of RNase I treatments (ii. and iii. for translation). (B) Average *Pearson* and *Spearman* correlation coefficients depending on the number n of considered transcripts. Transcripts were ordered by descending RPF count per transcript, n most highly covered transcripts were selected, correlation coefficients were computed for each transcript in this subset, and the average was plotted. For correlation coefficients of individual transcripts see Supplemental Figure S5A. (C) Ribosome profiles of LHCBM1 exemplary for a profile with even RPF distribution and LHCBM6 as an example for a profile with an extreme initiation peak. 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. (D) Stacked column chart representing the fraction of ribosome profiles containing extreme initiation peaks (defined as transcripts with an average coverage in first 7 codons exceeding by at least 4 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, 7 out of 66 for chloroplast and 0 out of 7 for mitochondria). Further information on occurrence of abnormal initiation peaks is provided in Supplemental Figure S6.

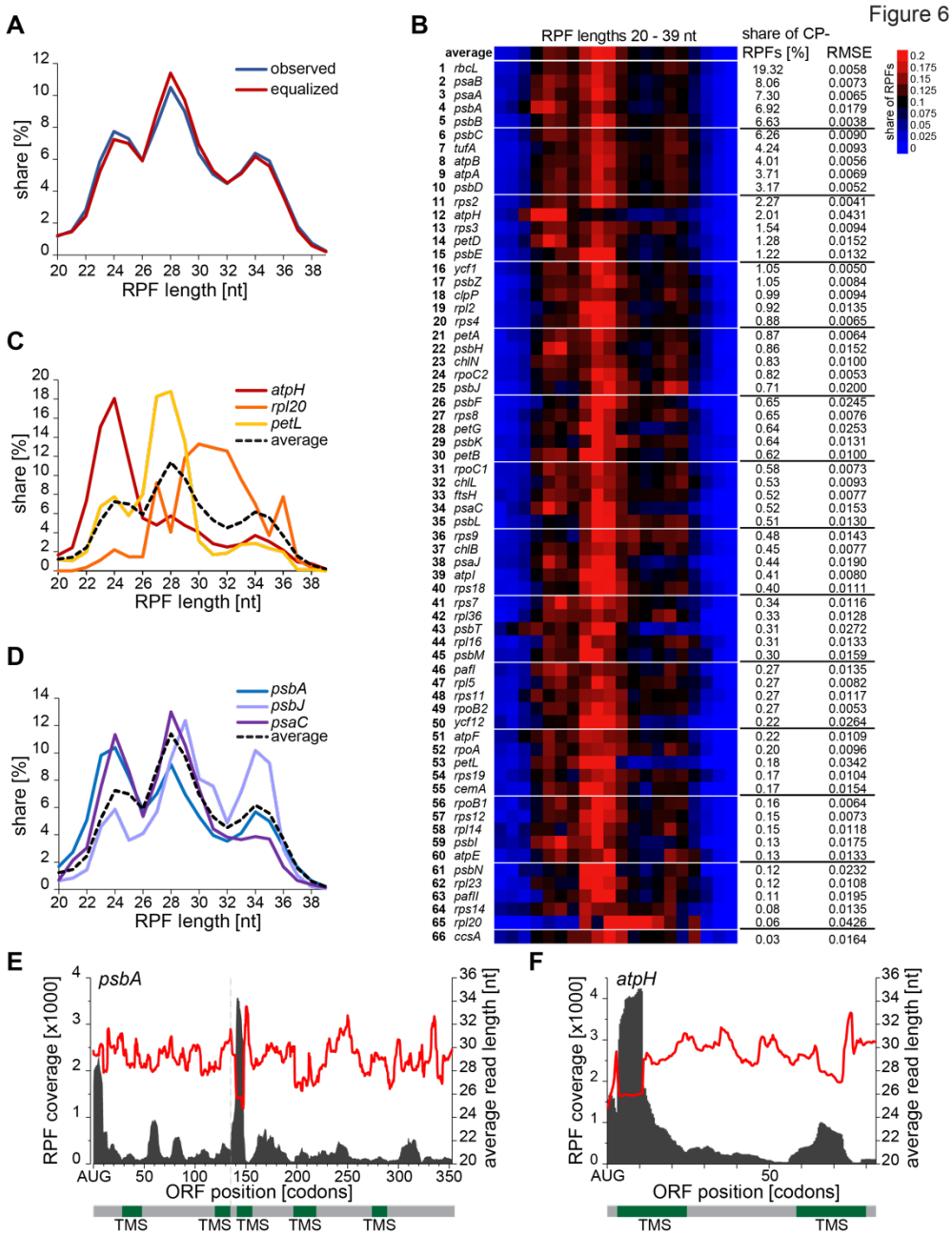


Figure 6: Chloroplast ribosomes generate characteristic ribosome protected fragments

(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 share of the different RPF sizes for each chloroplast transcript. Transcripts are ranked based on their overall share of RPFs relative to all measured chloroplast RPFs. RMSE (Root Mean Square Error) denotes the deviation of the respective size distribution relative to the average size distribution (top row). (C) Size distribution of chloroplast-encoded transcripts with high deviation relative to the average size distribution. (D) Additional size distribution curves of photosynthesis complex subunits. (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 RFP coverage per nt. Positions of transmembrane segments (TMS) are given below the graph.

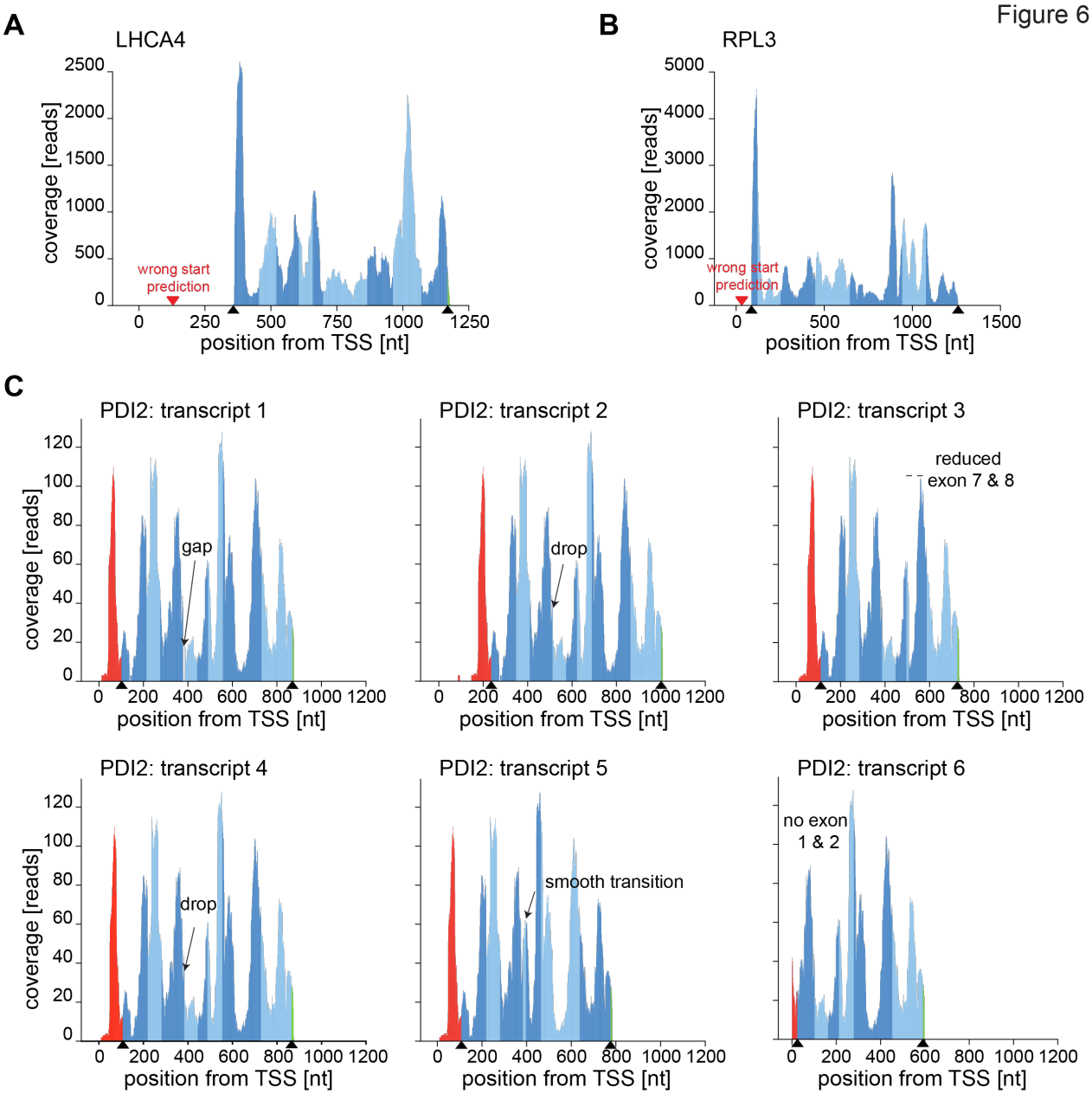


Figure 7: Utilizing ribosome profiles for coding sequence annotation
Exemplary ribosome profiles for determining initiation sites and alternatively spliced transcripts. Red parts of the profiles represent coverage within the 5'-UTRs, green parts represent 3'-UTRs and alternating blue shades represent 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 wrong 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 labelled as "true" initiation site, however, an upstream non-cognate start codon might be the correct initiation site.

SUPPLEMENTAL INFORMATION

Utilizing high resolution ribosome profiling for the global investigation of gene expression in *Chlamydomonas*

Vincent Leon Gotsmann[1], Michael Kien Yin Ting[2], Nadin Haase[3], Sophia Rudorf[3], Reimo Zoschke[2], and Felix Willmund[1]*

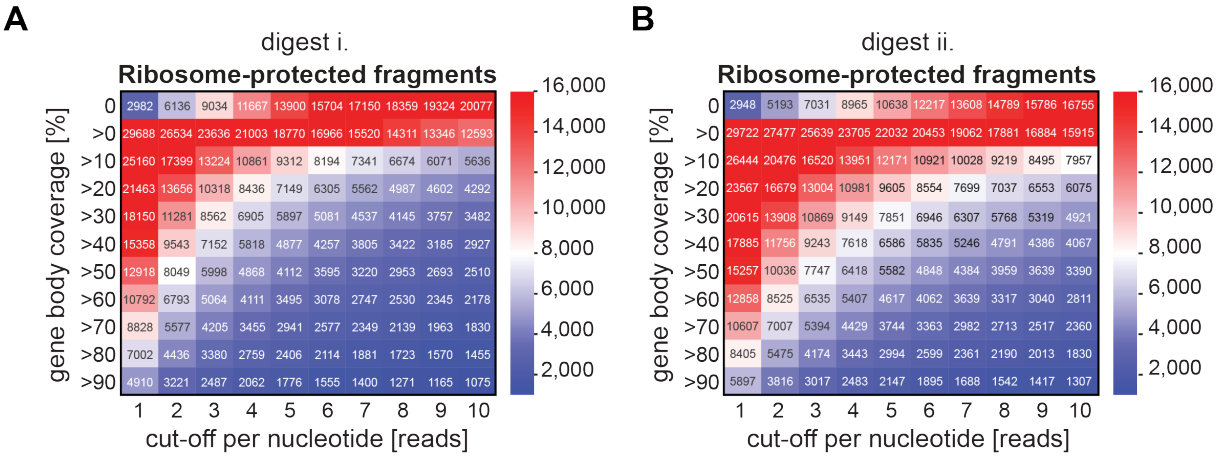
Author's institution(s)/affiliation(s):

[1] Molecular Genetics of Eukaryotes, University of Kaiserslautern, Paul-Ehrlich-Str. 23, 67663 Kaiserslautern, Germany.

[2] Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany.

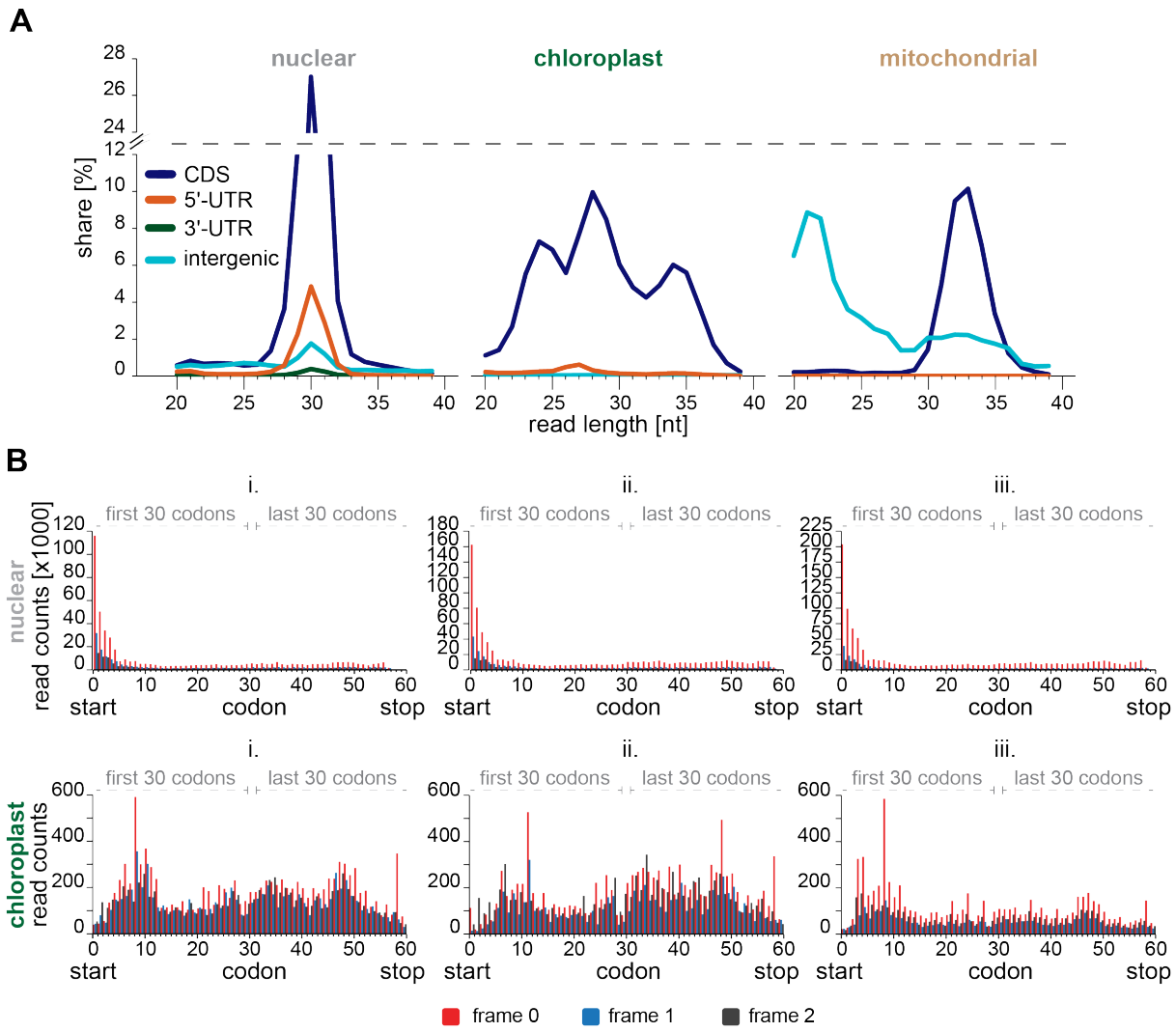
[3] Computational Biology, University of Hannover, Herrenhäuser-Str. 2, 30419 Hannover, Germany.

Supplemental Figures and Legends



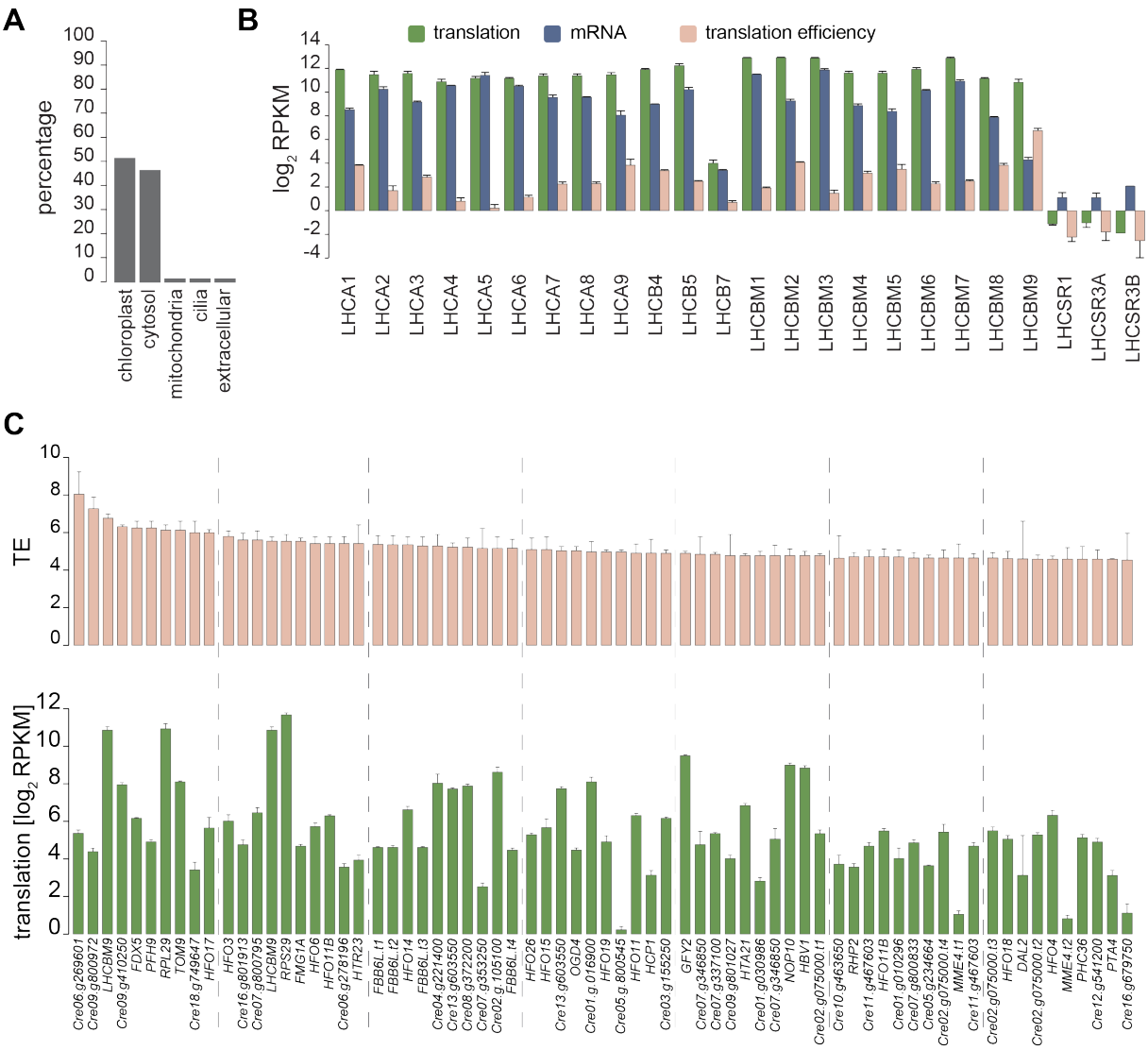
Supplemental Figure S1: Features of ribosome protected fragment coverage

Supporting Figure 1. (A) and (B) heatmaps representing transcript numbers categorized by their extent of gene body coverage in dependence of specific read count cut-offs for Ribo-seq data sets i. and ii.



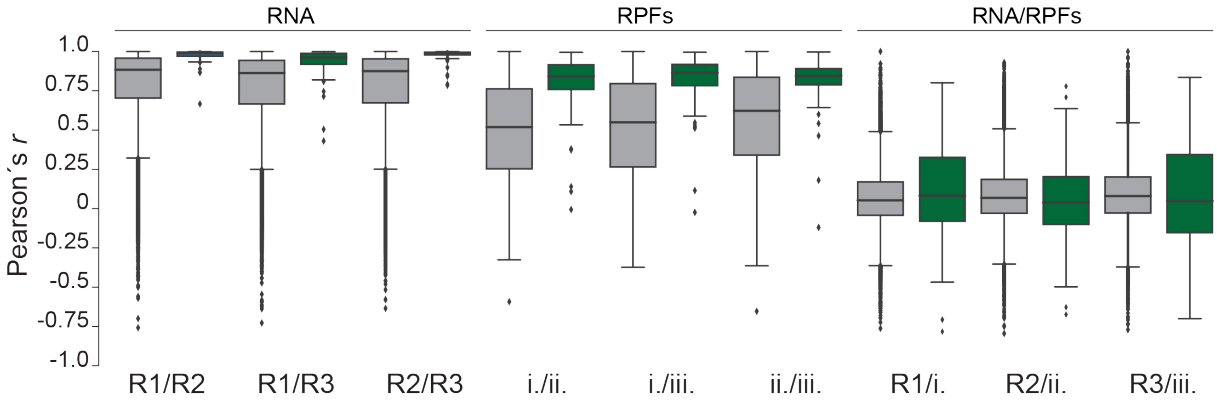
Supplemental Figure S2: RPF length distribution and periodicity

Supporting Figure 2. (A) Length distribution of RPFs for each organelle separated by their match to CDS, 3´- or 5´-UTR or intergenic regions (only data of digest iii. are shown here). (B) metagene RPF accumulation of nuclear and chloroplast-encoded transcripts per codon for all three nuclease treatments. Codons 1-30 represent the first 30 codons of each transcript, whereas codons 31-60 represent the last 30 codons of each transcript. Frame 0 corresponds to AUG in the ribosomal P-site.



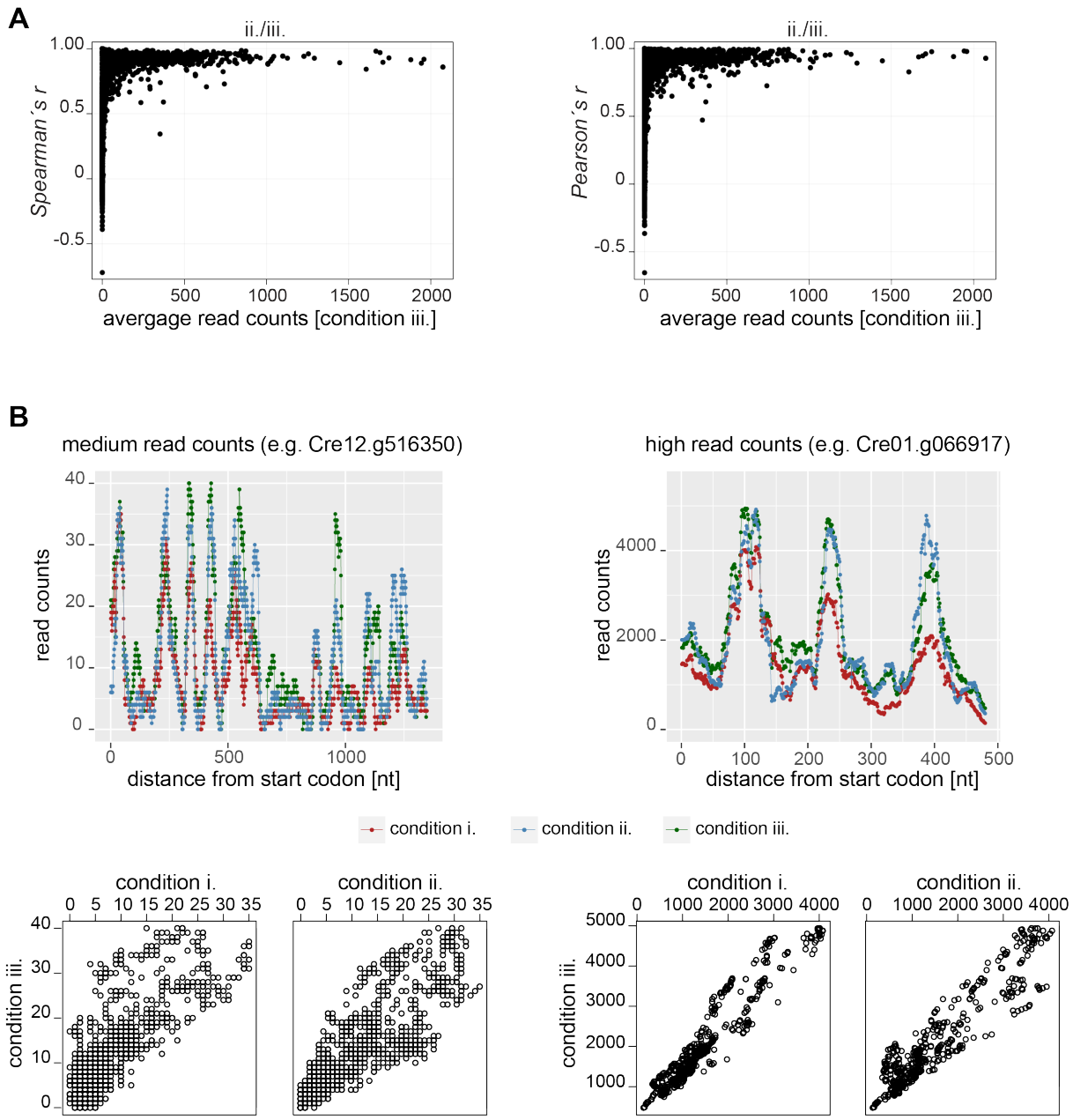
Supplemental Figure S3: Transcript accumulation, translation output and translation efficiency of nuclear genes

Supporting Figure 4. (A) Subcellular distribution of the top 100 expressed nucleus-encoded transcripts on the translation level. If multiple transcript variants of the same gene occurred in the list, only the longest occurring variant was considered. (B) Translation output, transcript abundance and translation efficiency of light harvesting protein encoding transcripts across all three Ribo-seq data sets, considering only the longest transcript variant, respectively. Mean values averaging all three Ribo-seq data are plotted, error bars denote standard deviations. (C) Translation efficiency and translation output of the 70 most efficiently translated nucleus-encoded transcripts. Mean values averaging all three Ribo-seq data are plotted, error bars denote standard deviations. If multiple transcript variants of the same gene occurred in the list, only the longest occurring variant was considered, unless stated in the legend.



Supplemental Figure S4: Correlation between nuclease digestion conditions, transcript accumulation and translation

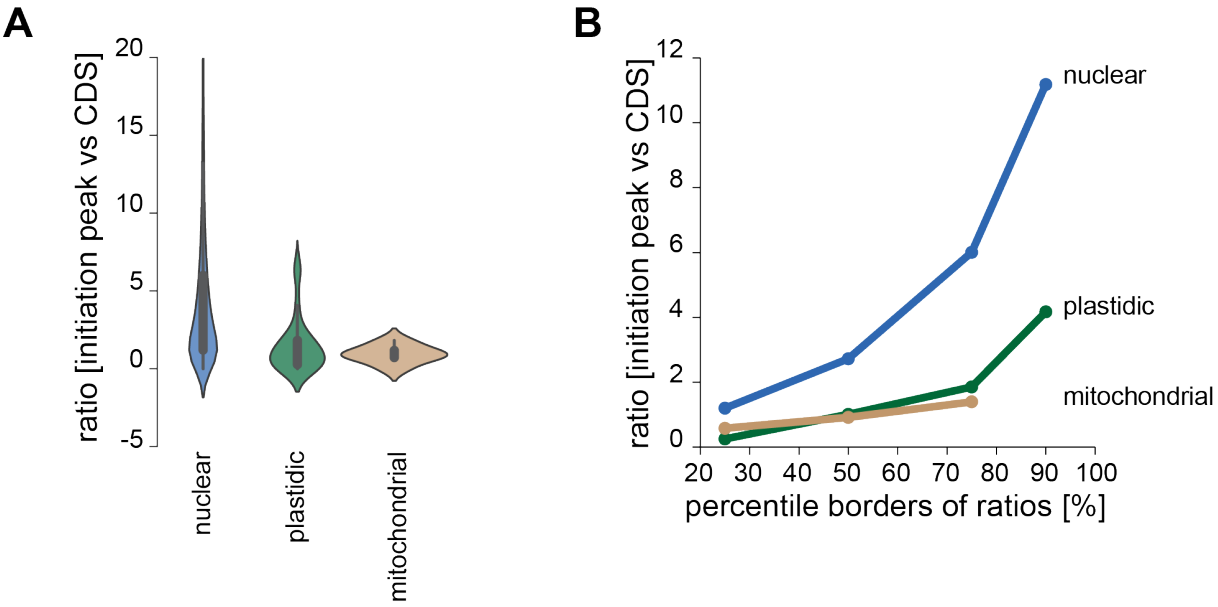
Supporting Figure 5. Boxplots representing the distribution of Pearson correlation coefficients between ribosome profiles and RNA-seq read distribution of the same transcripts each in different data sets. Boxes represent the 0.25 to 0.5 and the 0.5 to 0.75 quartiles, lines represent the median values and whiskers represent the remaining data. Outliers are defined as values being smaller or greater than the respective median ±1.5 times the interquartile range and are depicted as points R1/2/3 indicate different RNA-seq data sets.



Supplemental Figure S5: Correlation analysis of RPF count between nuclease digest conditions

Supporting Figure 5. (A) Relation between correlation coefficient comparing two experimental conditions and average read counts. For each transcript, counts were summed over all positions and then divided by the total number of nucleotides in the transcript. The correlation is high unless the average read counts are low. The distribution of *r*-values might be well suited for selecting good cut-off ranges for a deeper analyses of ribosome profiles. (B) Examples of ribosome profiles with medium and high read counts for all three experimental conditions. Scatter plot of

read counts for each nucleotide for comparison of two experimental conditions. Perfectly identical profiles would correspond to a straight line in the scatter plot.



Supplemental Figure S6: Detection of abnormal initiation peaks in the three genomes of *Chlamydomonas reinhardtii*

Supporting Figure 5. (A) Violin plots representing the distribution of ratios between average per nucleotide coverage within the first seven codons and the average per nucleotide coverage of the remaining CDS for each organelle's transcripts separately. Thick lines within the violins represent the interquartile range while thin lines represent the whiskers with a cut-off of 1.5 times the interquartile range. (B) Scatter plot with lines representing the increase of the percentile borders of the distributions shown in (A). In both cases, only transcripts with a gene-body coverage of at least 70% were considered (8757 nucleus-encoded, 66 chloroplast-encoded, 6 mitochondrial-encoded) to minimize the risk of misinterpretations due to low or absent coverage in either of both regions. Calculation of a 90th-percentile border for mitochondrial transcripts was not possible due to the low number of transcripts in consideration.