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

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@bio.uni-kl.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 in plant cells. 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. Furthermore, drastic differences between cytosolic and chloroplast offset distributions underline the different nature of the chloroplast translation mechanism. Analyses of RPF distribution along specific 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. These features will be easily accessible for the research community since we attached our ribosome profiling data to the most recent Chlamydomonas reference genome.

## Introduction

Translation is accomplished via ribosomes, highly conserved macromolecular ribonucleoprotein machines that decode genetic information into linear polypeptide chains

46 resulting in an almost unlimited diversity of proteins. Protein synthesis is the origin for the
47 quantitative and qualitative determination of a proteome´s diversity, which ultimately
48 shapes the biochemical character of a cell. It is now clear that translation output does not
49 strictly follow the changes of transcript abundance, but instead underlies strict regulatory
50 control (Pechmann et al., 2013; Payne, 2015; Slobodin and Dikstein, 2020). The
51 importance of this regulation is illustrated by the fact that bacteria invest about 50% of
52 their available energy for protein biogenesis and that more than 10% of all yeast proteins
53 may be somehow involved in this process (Russell and Cook, 1995; Costanzo et al.,
54 2000). Translational regulation is observed throughout an organism's life span and
55 includes central regulatory circuits that control the cell cycle, tissue differentiation and
56 development, acclimation to environmental changes and the integration of external
57 signals (reviewed in Pechmann et al., 2013; Zoschke and Bock, 2018; Teixeira and
58 Lehmann, 2019; Wang and Amoyel, 2022). Thus, understanding the regulatory principles
59 and mechanisms of translation is a central task of modern biology. With the tremendous
60 advances in next generation sequencing technologies, it is now possible to systematically
61 assay the status of cellular translation by a method called ribosome profiling (Ribo-seq)
62 (Ingolia et al., 2009). The technology is based on the observation that ribosomes occupy
63 precise sections of a translated mRNA (Wolin and Walter, 1988). These mRNA sections
64 (termed ribosome footprints or Ribosome Protected Fragments, RPFs) are extracted by
65 nucleolytic removal of un-occupied mRNA in cell lysates or isolated ribosomes. RPFs are
66 then identified through next generation sequencing and mapping to a reference
67 transcriptome or genome. Thus, Ribo-seq benefits from the same dynamic range as RNA-
68 seq, and provides a rich and precise positional information of translating ribosomes from
69 the physiological state when cells were harvested (Ingolia et al., 2009; Ingolia, 2014; Brar

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

91 Ribo-seq has been applied to different plant species with variable gene coverage
92 and quality, including *Chlamydomonas reinhardtii* (Chlamydomonas hereafter) (Chung et
93 al., 2015), *Arabidopsis thaliana*, (Liu et al., 2013; Juntawong et al., 2014; Merchante et

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

116 **Results and Discussion**

117 *Establishing super resolution ribosome profiling for Chlamydomonas reinhardtii*

118 Ribosome profiling was performed from logarithmically grown Chlamydomonas cultures
119 that were kept under mixotrophic conditions and moderate light (see Materials and
120 Methods). Our previous experience with targeted chloroplast ribosome profiling (Trösch
121 et al., 2018; Trösch et al., 2022) showed that addition of the elongation inhibitors
122 chloramphenicol (CAP) and cycloheximide (CHX) (stalling 70S and 80S ribosomes,
123 respectively) is crucial to prevent ribosome run-off during harvesting (see Materials and
124 Methods). However, by short incubation of the inhibitors just during harvest, we reduced
125 the exposure time to the minimum. For RPF generation by nucleolytic digest of RNA that
126 is not protected by a ribosome, we applied RNase I at room temperature (23 °C) or 4°C,
127 consistent with most Ribo-seq studies on other organisms (Ingolia et al., 2012). Our
128 previous chloroplast ribosome profiling approaches (using MNase) revealed that the best
129 digestion (i.e. polysome to monosome dissociation) was achieved when ribosomes were
130 purified prior nuclease treatment (Trösch et al., 2018). We thus compared RNase I
131 digestion on pre-purified ribosomes both at 4°C (condition i.) and 23°C (condition ii.) with
132 samples containing whole cell lysates (condition iii.) (Figure 1A). RNase I was applied at
133 a concentration of 1 unit (U) per micro gram (µg) RNA, which is similar to published Ribo-
134 seq approaches performed in yeast and Arabidopsis (Chartron et al., 2016; Hsu et al.,
135 2016; Döring et al., 2017). However, the RNase I concentration was far below the >9 U
136 per µg RNA that was previously applied for Chlamydomonas (Chung et al., 2015). By
137 using the lowest possible concentrations of RNase I, experimental costs can be reduced.
138 Moreover, over digestion of ribosomes may cause an over-proportional content of
139 contaminating rRNA fragments, which then dominate sequencing libraries. In the previous
140 Chlamydomonas Ribo-seq study by Chung et al. (2015), contaminating rRNA fragments
141 accounted for >90% of all sequenced reads. Our pilot runs had contaminating rRNA levels

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

157 Parallel to Ribo-seq, transcript accumulation was determined by sequencing whole
158 cell RNA samples on the same platform. Ideally, RNA samples are harvested just prior to
159 ribosome profiling experiments from the same culture. While typical transcriptomic studies
160 enrich mRNA pools by polyadenylation-enrichment, we here sequenced fragmented small
161 RNAs that are derived from the full set of transcripts including plastid transcripts, which
162 frequently have no or short polyadenylation signal sequences and are thus poorly
163 enriched by standard protocols (Komine et al., 2000; Gallaher et al., 2018).

164 Ribosome profiling and transcript samples were sequenced with a depth of 40 M
165 and 20 M reads, respectively, which on average resulted in 25.7% of reads mapping to

166 annotated CDSs of the most recent Chlamydomonas reference genome v6.1 (Craig et al.,
167 2022). On average, 64% mapped to putative rRNA loci and 4.2% mapped to annotated
168 untranslated regions upstream or downstream of CDSs. Despite the application of
169 different rRNA depletion methods, the rRNA contamination was very similar in all samples.
170 Of the sequenced reads that mapped annotated CDS a minor fraction (<1%) matched to
171 more than 10 positions and were removed from further analyses. Of the reads mapping
172 to CDSs, 96.24% mapped to transcripts of the nuclear genome, 3.63% mapped to
173 chloroplast-encoded transcripts, and 0.12% to mitochondrial transcripts.

174 Transcript accumulation was determined by averaging reads per CDS, relative to
175 the length of each CDS and the total read numbers of the dataset (Reads Per Kilobase
176 per Million, RPKM), a parameter that is widely used for transcriptome analyses and
177 accounts for a higher likelihood of measuring RNA fragments of long versus short
178 sequences (Conesa et al., 2016). This is of particular importance for the sequencing of
179 small and fragmented RNAs, as commonly accompanied with Ribo-seq. For Ribo-seq,
180 averaging RPFs across CDS has been previously used to correct for local variations of
181 RPF distribution to determine the level of translation output per transcript (Ingolia et al.,
182 2011; Ingolia, 2014). However, the short RPF molecules represent absolute counts of
183 ribosome-protected RNA segments in a sample. Since the number of ribosomes bound to
184 a transcript should majorly be determined by the transcripts initiation and translation rate
185 instead of their length, RPKM normalization may distort the data, especially by over-
186 representing read counts for transcripts considerably shorter than 1000 nt and under-
187 representing very long transcripts (Ingolia et al., 2009). Thus, translation output was
188 determined by normalizing the RPF counts per CDS relative to total read depth of the
189 sequencing dataset (Count Per Million, CPM), placing the expression values into a

190 ribosome-centric context that represents the proportion of ribosomes per transcripts
191 relative to the total ribosome pool. This is consistent with recent Ribo-seq studies (Hsu et
192 al., 2016). *Pearson* correlation coefficients between CPM values and transcript length is
193 near to zero in all samples and proves the absence of any linear monotonic relationship
194 between these variables. Thus, the effects of transcript length on the read count, if
195 existent, are much weaker than other influences and therefore negligible (Supplemental
196 Figure S1A).

197 Based on CPM and RPKM values, respectively, the three nuclease digestion
198 conditions (i. to iii.) showed high reproducibility between the Ribo-seq and RNA-seq
199 experiments, respectively (r = 0.98 and r = 0.99, Figure 1B). This agrees with previous
200 observations that Ribo-seq tolerates variations in RNase digest (Gerashchenko and
201 Gladyshev, 2017).

202 We next aimed to determine the depth of the Ribo-seq approach. Low coverage
203 Ribo-seq frequently results in ribosome profiles covering only the transcripts with highest
204 ribosome occupancy (i.e. translation output), whereas moderate or lowly translated CDS
205 may have partial RPF coverage along the respective sequence (here termed gene body
206 coverage, Figure 1C, left panel). For Ribo-seq with a good depth, many CDS should show
207 decent gene body coverage (Figure 1C, right panel). The depth was determined for all
208 Chlamydomonas transcripts of the v6 genome (32670 transcripts considering the products
209 of gene copies). For digest condition iii., more than 10,000 transcripts had at least 50% of
210 their CDSs sequence covered with two or more RPFs, while 6437 transcripts had over
211 90% of their CDS covered with at least one RPF (Figure 1D). 1508 transcripts are covered
212 over >90% with an average of more than 10 RPFs/nt, hence representing the group of
213 transcripts with the most complete gene body coverage combined with a high basal

214 coverage (see also Figure 4B). Categorizing the transcripts in this way allows for the
215 assessment of the general quality of a data set, as well as the selection of transcripts of
216 similar quality for specific analyses without having to rely on expression values alone. The
217 gene body coverage was highest for digest condition iii. (Figure 1D and Supplemental
218 Figure S1B and S1C), and comparably good for the coverage of RNA fragments for the
219 RNA-seq experiment (Figure 1E). Of note, even transcripts with low CPM correlated well
220 between experimental conditions, demonstrating good reproducibility that is comparable
221 to the pioneer experiments performed with yeast (Figure 1B) (Ingolia et al., 2009).

222 The read length distribution of the RPFs mapping to annotated CDSs of the nuclear
223 genome displayed high comparability to published Ribo-seq approaches with other
224 organisms. RNase I treatment typically results in RPFs with a length between 28 and 30
225 nt. Here, digest of purified ribosomes at 4°C resulted in RPFs with a predominant length
226 of 31 nt. Digest condition ii. caused a broader distribution of RPFs between 28 and 30 nt,
227 while digest in cell lysate resulted in a sharp peak of RPFs with a length of 30 nt (Figure
228 2A). Such a sharp peak at 30 nt is typically found in Ribo-seq from mammalian cells when
229 harvested in the presence of CHX (Wu et al., 2019; Sharma et al., 2021). The RPF size
230 distribution of the 8 mitochondrial transcripts needs to be considered with caution due to
231 the low coverage but appears bimodal with peaks at 22 nt and 33 nt under all conditions.
232 Similar size distributions were observed from previously described mitochondrial Ribo-
233 seq and could stem from the bulkiness of the membrane associated ribosomes (Rooijers
234 et al., 2013). For chloroplast transcripts, the RPF size distribution is less defined and may
235 be the result of diverse translational states, as previously reported for maize and
236 Arabidopsis (Chotewutmontri and Barkan, 2016; Gawronski et al., 2018; Fujita et al.,
237 2019). The sharpest peak for chloroplast RPFs is seen under condition ii., albeit the

238 predominant RPFs of 28 nt is found under all treatment conditions (Figure 2A). While
239 consistent with the observations in other organisms, our RPF size distribution varies from
240 the previous Ribo-seq data with Chlamydomonas, which reported a general shift towards
241 a smaller RPF sizes, maybe as consequence from their harsh RNase I treatment (Chung
242 et al., 2015).

243 A hallmark of high-resolution Ribo-seq data is the triplet-wise 3-nt periodicity
244 of RPFs that represents slowdown of ribosome movement as ribosomes decode a
245 respective codon (Ingolia, 2016). To calculate periodicity, it is necessary to determine the
246 5'-offset for every RPF length species, which is the number of bases located 5' to the P-
247 site (the translational state of polypeptide bond formation) of a respective RPF. This is
248 usually done by considering all RPFs that fully enclose an annotated start codon and
249 determining the frequency of all occurring 5'-offsets for every RPF length separately. The
250 5'-offset, most frequently occurring for a specific RPF length, is then declared to be the
251 true 5'-offset of this species. This strategy is based on the fact that RPF occupancy of
252 transcripts usually inclines dramatically at the borders between 5'-untranslated regions
253 and CDS, with the start codon often having an exceptionally high coverage, which is due
254 to the slow process of translation initiation. The 5'-offsets determined in this way are used
255 to calculate the exact P-site position and reading frame of every RPF. By this, 3-nt
256 periodicity can be nicely visualized in meta-plots combining all genes while 3-nt plots of
257 individual transcripts can be noisy, depending on translation levels. The exact 5'-offset
258 varies in an organism- and protocol-specific manner and is usually 12 to 13 nt.
259 Alternatively, the P-site position has also been determined by calculating the 3'-offset of
260 the ribosomal Aminoacyl-site (A-site), applying the same strategy to RPFs mapping to
261 annotated stop codons.

262 With the majority of cytosolic RPFs with one read length (Figure 2A), all three
263 RNase I digest conditions yielded a clear frame preference of “frame 0” for nucleus-
264 encoded transcripts, meaning that the base triplet found at the P-site of a RPF is for the
265 for the majority of RPFs in-frame with the annotated CDS. Again, lysate digest (iii.)
266 resulted in the best frame preference values (>68% of RPFs in frame 0, 19.4% in frame 1
267 and 12.5% in frame 2), suggesting that this condition yields Ribo-seq data of the best
268 quality (Figure 2B). Due to the differences between 80S and 70S ribosomes, 3-nt
269 periodicity needs to be determined separately for chloroplast translation. For chloroplast
270 transcripts, the frame preference was less obvious, albeit it was clearest for digest
271 condition iii. (Figure 2C). Considering the overall quality of sample iii. (the similar degree
272 of rRNA contamination between all three samples and the high correlation with the other
273 samples on CPM level), it can be concluded that RNase H mediated rRNA depletion has
274 no adverse effects on the Ribo-seq sample quality and seems to be similarly effective.
275 To get a better insight into the RPF offsets and periodicity, different lengths of the
276 RPFs were plotted against their offset frequencies upon P-site and A-site alignment for
277 the dataset of condition iii. (Figure 2D). The triplet-wise movements of the 5´ends with an
278 offset of 12 nt during initiation is clearly visible for cytosolic footprints and most prominent
279 for the 30 nt RPFs of digestion conditions iii., consistent with the observation in other
280 eukaryotic organisms (Lauria et al., 2018). Likewise, triplet-wise movements of the 3´-
281 ends are also detectable and are in frame with the stop codon. Here, a 9 nt long 3´-offset
282 seems the predominant form prior to termination with the stop codon in the A-site of the
283 ribosome. However, due to the 12 nt 5’-offset and 3 nt occupied by the P- and A-sites
284 respectively, a 3’-offset of 12 nt was expected. This discrepancy could be a result of CHX
285 treatment, which locks ribosomes in a pre-translocation conformation after peptide bond

286 formation (Wu et al., 2019) so that RPFs on non-stop codons may exceed the natural
287 termination-caused accumulation on stop-codons. Moreover, a slow termination process
288 might arrest ribosomes on the codon just upstream of the stop, prolonging ribosome dwell
289 time at this site. At the same time, ribosomes that already entered termination at the time
290 of sampling are likely to complete the process and dissociate from the transcript. CHX
291 does not inhibit termination, which in turn would dramatically lower the number of FPs
292 mapped to stop codons with their A-site. Interestingly, RPFs with a length of 21 nt are
293 detectable upon initiation (asterisks, Figure 2D). These fragments have been previously
294 characterized as footprints of ribosomes in an open ratchet conformation during
295 translocation, even accumulating as abundant species without CHX treatment in yeast
296 (Lareau et al., 2014). During elongation, ribosomes undergo massive conformational
297 changes by rotating the large subunit relative to the small subunit (Frank and Agrawal,
298 2000; Zhang et al., 2009). The fact that the RPFs produced from the open-ratchet
299 conformation have the same 5'-offsets as the main RPF species despite being
300 considerably shorter, suggests that the RNase exclusively removes nucleotides from the
301 3´-end of RPFs in this conformation (and could indicate that the center of the rotation is
302 located towards the 5'-direction, seen relative to the RPF). Cycloheximide was shown to
303 prevent binding of the next aminoacylated tRNA in the A-site during peptide bond
304 formation (Schneider-Poetsch et al., 2010) and thus stabilizes ribosomes in a specific
305 stage during the elongation cycle, in which RPFs of 28-30 nt are the prevalent form.
306 Interestingly, human 80S ribosomes are not exclusively arrested in one conformation
307 upon CHX treatment (Sharma et al., 2021), comparable to the situation seen here for
308 Chlamydomonas 80S ribosomes (Figure 2D, upper panel). These 20-22 nt RPFs might
309 derive from a post-translocation ribosome species (Lareau et al., 2014). The presence of

310 both species demonstrates active movement of ribosomes along the CDS. The absence
311 of the “open-ratchet” 21 nt footprints starting from two peptide-bond formations prior to
312 termination, may support the idea from above that ribosomes near stop codons are
313 paused, which can prevent the collision between ribosomes (Figure 2D upper panel,
314 dashed box). Interestingly, the frame preference for frame 0 of the stop-minus-2 codon is
315 the highest in the whole metagene periodicity plot (79.6% Figure 2B), which could denote
316 a short stalling at this codon before translocation. However, for the next codons the frame
317 0 preference drops considerably to 55.8%, especially in favor of frame 1 (36.2%) and then
318 to 33.1% at the stop codon, clearly indicating that translation dynamics change towards
319 the termination step. Moreover, these characteristics are accompanied by a striking
320 continuous decline of absolute numbers of reads mapping to these codons on the meta-
321 gene level (for digest iii: stop-2: 19890, stop-1: 6786, stop: 522).

322 The prokaryotic-type chloroplast RPFs 5’-offsets seem to be much more diverse
323 than their cytosolic counterparts, apparent by their multiple clusters on the heat map
324 (Figure 2D, bottom panel). These might resemble populations of initiation complexes in
325 different states or in complex with transcript specific RNA processing factors and *trans*-
326 acting factors that interact with ribosomes (Westrich et al., 2021). Without clear criteria
327 how to distinguish and interpret these clusters, 5’-offset definition is problematic, since
328 multiple 5’-offset frequencies of comparable amplitude exist for many RPF lengths.
329 However, the chloroplast 3’-offsets seem less scattered and reveals a main RPF species
330 of 28 nt length with 11 nt 3’-offset. This is in accordance with the RPF length distribution
331 of all chloroplast RPFs, where the 28 nt species also represents the main species,
332 although the distribution is much broader than for the cytosolic RPFs. Similar to the
333 cytosolic offset heat map, a triplet-wise movement of ribosomes can also be observed for

334 the terminating chloroplast ribosomes in the 3'-offset heatmap for 29 nt RPFs, albeit with
335 much lower clarity. Altogether, the observations made for chloroplast offsets and RPF
336 length distribution might indicate a certain heterogeneity of the chloroplast ribosome pool,
337 which is an interpretation that would be in line with the large number of proteins that
338 interact with chloroplast ribosome found in our previous study (Westrich et al., 2021) and
339 the large number of RNA-binding proteins known or suspected to regulate the translation
340 of specific chloroplasts transcripts (Nickelsen et al., 2014). This is remarkable considering
341 the very small number of transcripts encoded on the chloroplast genome (72). Together,
342 this might suggest that translation in the chloroplast of Chlamydomonas is heavily
343 influenced by transcript-specific features rather than following a universal scheme
344 (Zoschke and Bock, 2018).

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

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

355 With the high reproducibility of Ribo-seq and RNA-seq data, we chose to average
356 the datasets for estimating the levels of transcript accumulation and translation output. A
357 direct comparison between translation output and transcript accumulation is frequently

358 referred to as translation efficiency (TE) and normalizes RPF values over transcript
359 accumulation. Contrasting translation over transcript data of the Chlamydomonas genes
360 in a scatter plot and highlighting their subcellular localization showed that nuclear
361 transcripts encoding chloroplast-localized proteins have higher expression values when
362 compared to proteins of cytosolic or Endoplasmic Reticulum localization (Figure 3C),
363 again supporting the notion that many chloroplast proteins are expressed at high levels.
364 Interestingly, most nucleus-encoded transcripts showed similar levels between transcript
365 accumulation and translation (distributed along the diagonal, Figure 3C). In contrast,
366 chloroplast- and mitochondria-encoded transcripts showed higher transcript accumulation
367 levels, when compared to translation, which agrees with the published data of profound
368 potential translational regulation within the chloroplast (Sun and Zerges, 2015; Trösch et
369 al., 2018; Zoschke and Bock, 2018). Remarkably, translational regulation seems also
370 predominant for the few mitochondrial genes.

371 Column plots of the individual chloroplast transcript and translation values show
372 that even the weakest expressed known chloroplast transcripts have CPM values ~30,
373 which is within the median range of nucleus-encoded transcripts (Figure 4A versus 3A).
374 Not surprisingly, transcripts that encode the core photosystem I subunits PsaA and PsaB,
375 the photosystem II core subunits PsbA/D1,PsbB/CP47, PsbC/CP43, PsbD/D2, the
376 Rubisco large subunit, RbcL, and the translation elongation factor TufA show the highest
377 translation levels, consistent with our previous chloroplast ribosome profiling experiments
378 and the examination of protein accumulation levels (Schroda et al., 2015; Trösch et al.,
379 2018). We further detected transcripts of the uncharacterized chloroplast Open Reading
380 Frames (ORF) *orf202* (positioned within the transposon Wendy I), *orf854* (positioned
381 within the transposon Wendy II), *orf528*, and *i-crel* (Gallaher et al., 2018), however,

382 translation output was low or even undetectable (for *i-crel*), suggesting that these
383 transcripts are barely translated in Chlamydomonas cultures, at least under standard
384 conditions (Figure 4A). The direct TE ratios of the chloroplast transcripts again show the
385 imbalance between chloroplast transcript accumulation and translation output, suggesting
386 that the long-living chloroplast transcripts present a buffering capacity for the fast and
387 efficient adjustment of translation levels during environmental changes, as for example
388 seen during light and temperature acclimation and most prominent for PsbA/D1 synthesis
389 (Chotewutmontri and Barkan, 2018; Schuster et al., 2019; Trösch et al., 2022).

390 We next determined translation output levels of nucleus-encoded transcripts, only
391 considering the longest transcript variant if multiple splice forms existed. Of the transcripts
392 with the highest translation output, 54 protein products have a known or predicted
393 chloroplast localization and only 34 are present in the nucleus or cytosol, again showing
394 the high accumulation levels of chloroplast proteins (Supplemental Figure S2A). Amongst
395 those, transcripts of the light harvesting complex proteins (LHCs) have the highest
396 translation output, consistent with their high protein accumulation levels in plant cells
397 (Dall'Osto et al., 2015). Generally, *LHC* translation showed comparable CPM translation
398 output values, whereas transcript accumulation was rather variable. For the transcripts of
399 the high-light induced photoprotective LHCSR3.1 and LHCSR3.3 proteins, we could not
400 determine translation profiles (Supplemental Figure S2B), suggesting that these proteins
401 are not synthesized under the investigated growth conditions (Bonente et al., 2012).
402 Besides photosynthesis related transcripts, the list with top-most translation further
403 contained transcripts encoding proteins such as the Isocitrate lyase (ICL1, CPM 4656,
404 rank 5), a key enzyme of the glyoxylate cycle (Plancke et al., 2014), the thiazole
405 biosynthesis enzyme (THI4, CPM 3230, rank 13) (Moulin et al., 2013) or the Methionine

406 Adenosyl-transferase (METM1, CPM 2856, rank 26), confirming their central role in
407 Chlamydomonas cells. Interestingly, the $B_{12}$-independent methionine synthase enzyme,
408 METE1, seems to be higher expressed, when compared to the $B_{12}$-dependent METH1
409 isoform (CMP 1655 versus 990) (Bunbury et al., 2020). METE1 is less common than
410 METH1 within the algal clade and makes these species independent form external $B_{12}$
411 supply (Bertrand et al., 2012).

412 *Detection of open reading frames by ribosome profiling facilitate Chlamydomonas*
413 *genome annotation*

414 Local elongation rates are known to vary considerably during translation along a transcript,
415 depending on tRNA availability, mRNA secondary structure, the amino acid sequence of
416 the resulting polypeptide and co-translational folding (Gloge et al., 2014; Kramer et al.,
417 2019). With the robust gene body coverage over Chlamydomonas transcripts (Figure 1D),
418 we aimed to learn more from the dynamics of translation elongation, derived from the RPF
419 distribution over individual coding sequences (ribosome profiles). Ribosome profiles were
420 highly comparable between the three RNase I treatment conditions i.-iii., as indicated by
421 median *Pearson* correlations between data sets ii. and iii. of ~0.62 for nucleus-encoded
422 transcripts and >0.85 for chloroplast-encoded transcripts. Although RNA-seq read
423 distributions over individual CDS were even more comparable (median of *r*-value 0.87 for
424 nucleus-encoded and 0.99 for chloroplast-encoded transcripts) among the RNA-seq data
425 sets, they showed no correlation (median *r*-value close to zero) to the respective ribosome
426 profiles (Figure 5A and Supplemental Figure S3A). Certainly, methodological-driven
427 differences between Ribo-seq and RNA-seq cannot be fully excluded, but it can be
428 assumed that the ribosome profiles image transcript-specific translation dynamics and are
429 likely not the result of sequencing biases.

430 The correlation coefficients for specific transcripts comparing two experiments
431 (both *Pearson* and *Spearman*) stably range between 0.7 and 1 for the majority of ribosome
432 profiles that have an average coverage >100 reads per nucleotide, suggesting that low
433 correlation coefficients are the consequence of sequencing depth limitations and thus
434 limited observability of lowly translated transcripts rather than the true absence of a
435 characteristic translation dynamic of these transcripts (Supplemental Figure S4).
436 Importantly, more than 10,000 transcripts have *r*-values >0.7 (Figure 5B), again showing
437 the high reproducibility between the RNase I digest conditions for transcripts with
438 moderate and high RPF coverage (Supplemental Figure S4). In fact, the *r*-value
439 distribution shows that these direct comparisons of read coverage per transcript might be
440 well suited for selecting good cut-off ranges for a deeper analyses of ribosome profiles. In
441 many previous studies, simply minimum CPM values (e.g. 1) were taken as cut-off.

442 Ribosome profiles of the highly translated *LHC* transcripts showed some
443 remarkable differences between the individual *LHC* species. Efficiently controlling the
444 expression of *LHC* genes is essential for regulating the antenna size of the photosynthesis
445 machinery in order to capture photons under low irradiance conditions (increased
446 expression) or avoid over-excitation and thus photodamage under high irradiance
447 conditions (reduced expression) (Dall'Osto et al., 2015). Of the transcripts encoding the
448 21 different LHC members serving Photosystem I or II (LHCA and LHCB, respectively),
449 20 were found amongst the top 100 highest translated transcripts (not considering splice
450 variants). Only the transcript encoding LHCB7, a rather newly discovered protein
451 containing the unusual number of 4 transmembrane domains (Klimmek et al., 2006) had
452 low CPM values of ~16 (Supplemental Figure S2B). The variations in *LHC* transcript levels
453 result in considerably different TE values (Supplemental Figure S2B), suggesting that

454 translational regulation might be achieved by different strategies for expression of the *LHC*
455 species. Indeed, the ribosome profiles are considerably different between transcripts
456 encoding the different LHCs. For example, *LHCBM6* showed an exceptionally strong peak
457 of RPFs covering the first 7 codons, which exceeds the average RPF density over the
458 remaining CDS by more than 4-fold (Figure 5C). While "initiation peaks" are commonly
459 accumulating over the first few codons of a CDS, their amplitudes are usually within the
460 dynamic range of the remaining CDS. Such high peaks, in contrast, could point to a
461 regulative mechanism stalling ribosomes during or shortly after initiation. Indeed,
462 *LHCBM6* was shown to be translationally repressed under moderate light conditions by a
463 co-translational regulator termed NAB1 (Nucleic Acid Binding protein). NAB1 itself is
464 tuned by a regulatory circuit sensing carbon dioxide supply levels in Chlamydomonas, for
465 precisely adjusting the antenna composition of photosystem II (Mussgnug et al., 2005;
466 Blifernez-Klassen et al., 2021). Similarly, *LHCBM4* is recognized by NAB1, also
467 presenting a strong RPF peak over the initiation codon, while LHCBM1 (not a target of
468 NAB1) has no over-proportional initiation peak (Figure 5C), despite sharing a 75% amino
469 acid sequence identity with LHCBM6. Translational regulation during *LHC* expression is
470 not unique to Chlamydomonas, as *Lhcb4.2*, *Lhcb4.3* and *Lhcb6* are also subject to post-
471 transcriptional regulation in Arabidopsis (Floris et al., 2013). We searched all transcripts
472 that had a gene body coverage of at least 70% (8757 nucleus-encoded, 66 chloroplast-
473 encoded, 6 mitochondria-encoded) for the occurrence of comparable initiation peaks
474 (average coverage of the first seven codons at least four times higher than the average
475 coverage of the remaining CDS). Around 38% of selected nucleus-encoded transcripts
476 had comparable RPF accumulation peaks near the initiation codon, possibly hinting to
477 more co-translational regulation in the cytosol than previously anticipated. In contrast, only

478 10% (6 transcripts) of the chloroplast-encoded genes showed such peaks (Figure 5D).
479 However, it cannot be excluded that these peaks present artifacts of CHX treatment of the
480 Chlamydomonas cultures during harvest. Upon establishment of the Ribo-seq protocol,
481 several reports demonstrated that pre-incubation of yeast cultures with CHX may create
482 a bias by inducing increased RPF accumulation at initiation codons, which could be the
483 consequence of stalled elongation but unaffected initiation through CHX treatment (Ingolia
484 et al., 2009; Gerashchenko and Gladyshev, 2014; Sharma et al., 2021). However, a recent
485 direct comparison of CHX-treated and untreated conditions demonstrated that CHX
486 causes no initiation-peak bias in human cells, and that even the strong accumulation in
487 yeast might derive from slightly altered protocols between laboratories (Sharma et al.,
488 2021). For Chlamydomonas, CHX biases should result in general initiation peaks, but
489 these peaks are only found in a fraction of transcripts, independent of their translation
490 activity (e.g. *LHCs*). Thus, varying accumulation of RPFs around the initiation codon may
491 be an interesting indicator for altered initiation regulation between transcripts in the
492 cytosol.

493 Lastly, we aimed to exploit our Ribo-seq data for surveying correct CDS
494 annotations and transcript splice variants within the recently released Chlamydomonas
495 genome, version 6.1 (Craig et al., 2022). In previous genome versions, several genes
496 were falsely annotated, ignoring upstream start codons that were possibly the genuine
497 initiation site of translation (Cross, 2015; Craig et al., 2022). In the new version, the first
498 in-frame ATG codon was generally considered as the initiation codon, in accordance with
499 the scanning model of initiating ribosomes (Hinnebusch, 2011). Systematic comparison
500 of the reference start site with the onset of RPF coverage, confirmed that start codon
501 assignment has been clearly improved between genome version 5.6 and 6.1. However,

502 we detected gene models in which the first ATG appears not to be the genuine initiation
503 site. For example, transcripts encoding LHCA4 (Cre10.g452050) do not initiate from the
504 first ATG (Figure 6A) and were correctly annotated in the previous genome versions.
505 Another example is PRPL3 (Cre10.g417700) (Figure 6B). In addition, we determined
506 some gene models in which RPF accumulation did not correlate with the cognate ATG
507 start codon, possibly hinting to a translation start from non-cognate start sites as
508 exemplified by the *PDI2* CDS (Cre01.g033550, Figure 6C). In this case, the ribosome
509 profile starts with a prominent peak upstream of the annotated start codon located within
510 the annotated 5'-UTR while the annotated start codon is poorly covered compared with
511 the remaining CDS. Here, initiation from the non-cognate TTG could be an explanation
512 (Cao and Slavoff, 2020), which is located in-frame with the annotated coding sequence at
513 two positions within the area covered by the peak. Ribo-seq was already used before to
514 mine for non-cognate start sites in other organisms (Brar and Weissman, 2015).
515 Alternatively, these RPFs could point to the presence of upstream Open Reading Frames
516 for *PDI2*. *PDI2* represents a good example for detecting alternative splicing. Transcript 1
517 of *PDI2*, possesses a RPF gap between exon 3 and 4 (379 nt downstream of the
518 transcription start site), which is spliced out in transcript versions 2-4 (Figure 6C). In
519 addition, a sudden drop of RPFs is seen at the end of exon 3 in transcript 2 and 4. Based
520 on distribution of RPFs and the transition of ribosome profiles, transcript 5 has the most
521 consistent profile and might hence be the prevalently expressed variant in
522 Chlamydomonas cells, grown under standard laboratory conditions. In future, intelligent
523 modelling and peak deconvolution algorithms could be potentially applied to Ribo-seq
524 data to estimate the stoichiometric ratio of different splice variants' translation output.
525 *Concluding remarks*

526 Taken together, ribosome profiling is a fascinating tool for advancing our understanding
527 of post-transcriptional regulation of gene expression. Our Ribo-seq approach has the
528 resolution and depth to deeply study translation of the nuclear and chloroplast transcripts
529 in Chlamydomonas cells. Furthermore, Chlamydomonas is a well-suited organism for
530 understanding system level changes of gene expression throughout the cell cycle or
531 following environmental changes. Ribosome profiles can be of specific help, for a first
532 understanding of the many transcripts, which were not experimentally investigated before.
533 To make our data easily accessible in the research community, we have now linked our
534 Ribo-seq data to the recent Chlamydomonas genome CC-4532 v6.1 (https://phytozome-
535 next.jgi.doe.gov), thereby allowing individuals to conveniently browse the ribosome
536 profiles of their genes of interest. With the new genome release, the number of
537 alternatively spliced transcripts have been significantly expanded. Visualization of
538 ribosome profiles may help to uncover the dominant transcript variants, which are
539 translated under standard laboratory conditions. Certainly, additional experimental
540 approaches are required to understand unexpected features such as translation initiation
541 from non-cognate start sites, RPFs along upstream open reading frames, mechanism of
542 translational regulation, and spatial organization of protein synthesis.

543 **Materials and Methods**

544 *Cell growth and harvest*

545 *Chlamydomonas reinhardtii* CC-1690 cells were grown mixotrophically in Tris-Acetate
546 Phosphate (TAP) medium (Kropat et al., 2011) to mid-log phase (4 to $5x10^6$ cells per mL)
547 under constant light of 80 mmol $m^{-2}$ $s^{-1}$ (MASTER LEDtube HF 1,200 mm UO 16W830 T8
548 and 16W840 T8, Philips) and 25°C on a rotary shaker. Immediately before harvest, 100
549 µg/mL chloramphenicol (CAP) and cycloheximide (CHX) was added and cells were rapidly

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

559 *Ribosome profiling and RNA sampling*

560 Frozen cells were ground using a bead mill with nitrogen-cooled steel containers and
561 beads for two times 2 min at 27 Hz frequency and cooling in liquid nitrogen between
562 rounds. The cell powder was mixed with equal volumes of 2x concentrated lysis buffer (2x
563 concentrated TKM+ buffer supplemented with 2 mM dithiotreitol (DTT), 2% Triton X-100
564 and 20% sucrose) and incubated for 5 min at 4°C before pelleting cell debris at 8,000 g,
565 4°C for 10 min. For nucleolytic digest conditions i. and ii., lysate was layered on a 2.5 mL
566 64% sucrose cushion prepared in TKM+ buffer supplemented with 1 mM DTT and
567 ultracentrifuged for 3 h at 60,000 rpm and 4°C in a Ti70 rotor. The resulting polysome
568 pellet was briefly rinsed with 500 µL ice-cold diethyl pyro carbonate (DEPC)-treated water
569 and resuspended overnight in 100 µL ice-cold ribosome buffer (2x concentrated TKM+
570 buffer diluted with 25x concentrated Roche protease inhibitor cocktail to a total amount of
571 16%, 1 mM PMSF and 0.1% Triton X-100). Remaining insoluble debris were removed by
572 centrifuged for 1 min at 1,500 g before addition of 1 U of RNase I (Ambion) per µg RNA
573 and 0.134 U TURBO DNase (Invitrogen). Nucleolytic digest was performed for 1 h at 4°C

574 (condition i.) or 23°C (condition ii.) and stopped by addition of 0.4 U per applied unit of
575 RNase I of SUPERase•In RNase Inhibitor (Invitrogen). For condition iii., 1 U of Ambion
576 RNase I (Invitrogen) per µg RNA and 0.134 µL per µL lysate of TURBO DNase (Invitrogen)
577 was directly added to the lysate and incubated for 1 h at 4°C. The digest was stopped by
578 addition of 0.4 U of SUPERase•In RNase Inhibitor (Invitrogen) per used unit of RNase I.
579 The lysate was centrifuged again for 10 min at 8,000 g and 4°C to remove any cell debris
580 that may have precipitated during the nuclease digest. For all conditions, monosomes
581 were pelleted through a 750 µL 30% sucrose cushion prepared in TKM+ buffer
582 supplemented with 1 mM DTT for 30 min in a S150AT rotor at 72,000 rpm and 4°C. The
583 resulting ribosome pellets were resuspended in 100 µL of ice-cold ribosome buffer, then
584 supplemented with 15 mM ethylene diamine tetra-acetic acid (EDTA), pH 8.0 and instantly
585 mixed with 750 µL Invitrogen TRIzol reagent (Invitrogen) and incubated at room
586 temperature for 5 min. The solution was mixed with 150 µL chloroform and incubated for
587 2 min at room temperature before 15 min centrifugation at 20,000 g and 4°C. The
588 supernatant was mixed with 3 µL GlycoBlue co-precipitant (Invitrogen), 1:10 volume of 3
589 M sodium acetate, pH 5.5, 1 volume isopropanol and incubated overnight at -20°C. The
590 precipitated RNA was pelleted by 1 h centrifugation at 20,000 g and 4°C, briefly washed
591 with ice-cold 80% ethanol and centrifuged again for 30 min at the same parameters. The
592 resulting RNA pellet was briefly dried and dissolved in 11 µL of nuclease-free water.
593 For RNA-seq samples, 10 mL of culture were harvested by 2 min centrifugation at
594 4,000 g at room temperature (before harvesting the culture for footprint isolation), shock-
595 freezing in liquid nitrogen and storage at -80°C. Samples were lysed in prewarmed (50°C)
596 lysis buffer (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 20 mM EDTA pH 8.0 and 2% SDS)
597 and incubated for 2 min at 50°C. Immediately after incubation, 500 µL of TRIzol

598 (Invitrogen) reagent were added and the mixture was incubated for 5 min at room
599 temperature. The suspension was mixed with 200 µL chloroform and incubated for
600 another 5 min at room temperature before 3 min centrifugation at 12,000 g and room
601 temperature. The nucleic acid containing phase was transferred to a fresh tube and mixed
602 with 1.5 volumes of ethanol. RNA was purified by the Macherey Nagel NucleoSpin RNA
603 Plant kit.

604 *Ribosome footprint extraction and rRNA depletion*

605 50 µg RNA from the ribosomal extractions were diluted to a volume of 20 µL with nuclease-
606 free water, mixed with 2x formamide RNA loading buffer (90% deionized formamide, 20
607 mM Tris-HCl, pH 7.5, 0.5 M EDTA, 0.04% bromophenol blue in ethanol) and denatured
608 for 90 s at 80°C. The samples were instantly put on ice and then loaded onto 12% urea-
609 TBE-gels (90 mM Tris, 9 mM boric acid, 2 mM EDTA, 12% 19:1 40% acrylamide-
610 bisacrylamide mix, 8 M urea). Gels were run at 200 V for approximately 1 h and incubated
611 with SYBR Gold nucleic acid stain (Invitrogen), diluted 1:10,000 in TBE buffer (90 mM
612 Tris, 9 mM boric acid, 2 mM EDTA). Ribosome footprints were excised from the gels in a
613 size range from 20 – 40 nt according to a RNA size marker. Gel pieces were crushed and
614 incubated overnight with 400 µL footprint elution buffer (300 mM sodium acetate, pH 5.5,
615 0.25% sodium dodecyl sulfate and 1 mM EDTA, pH 8.0) at room temperature. The eluate
616 was collected, mixed with 3 µL GlycoBlue co-precipitant, 1 volume of isopropanol and
617 incubated overnight at -20°C. The precipitated RNA was pelleted by 1 h centrifugation at
618 20,000 g and 4°C, briefly washed with ice-cold 70% ethanol and centrifuged again for 30
619 min at the same parameters. The resulting RNA pellet was briefly dried and dissolved in
620 20 µL of nuclease-free water. For ribosome footprint samples, abundant rRNA fragments
621 were depleted by incubating 100 - 200 ng of ribosome footprints with 1.19 µL of 100 µM

622 biotinylated rRNA-oligo depletion mix (consisting of the oligonucleotides:
623 AATATGCGTTCAAAGATTCGATGATTCACG,
624 TAGCTCTAGAATTACTACGGTTATCCGAGTA,
625 TACCCGACGCTGAGGCAGACATGCTCTTGG,
626 GATTCGTGAAGTTATCATGATTCACCGCA,
627 ACGGGATGAATCTCAGTGGATCGTAGCA,
628 CGATCTAGCCGTCTTAGAGCTAGAAGCAGG) with 4 µL formamide, 1 µL 20x
629 concentrated hybridization buffer (300 mM sodium citrate, 3 M NaCl), 2 µL 0.5 M EDTA,
630 pH 8.0 and in a total volume of 20 µL. The mixture was hybridized in a thermocycler by
631 heating up to 80°C for 5 min and subsequent cooling in 5°C steps for 2 min each until
632 35°C was reached in the device. Afterwards, 60 µL of hybridization buffer, supplemented
633 with 20% formamide was added to the tube and the sample was depleted twice for 15 min
634 using 15 µL of Dynabeads MyOne Streptavidin T1 magentic beads (Invitrogen) freshly
635 washed and prepared according to the manufacturers manual. The supernatant was
636 mixed with 3 µL GlycoBlue co-precipitant, 2.5 volumes of ethanol and incubated overnight
637 at -20°C. Footprints were pelleted by 1 h centrifugation at 20,000 g and 4°C, briefly
638 washed with ice-cold 70% ethanol and centrifuged again for 10 min at the same
639 parameters. The resulting pellet was briefly dried and dissolved in 41 µL of nuclease-free
640 water. Excess oligos were removed by 30 min DNase digest at 37°C upon addition of 5
641 µL 10x TURBO DNase buffer, 2 µL Invitrogen TURBO DNase and 2 µL SUPERase•In
642 RNase Inhibitor (Invitrogen). After digest, footprints were mixed with 150 µL of 100 mM
643 NaCl and purified again by phenol-chloroform precipitation and final precipitation
644 overnight at -20°C using 2.5 volumes of ethanol and 2 µL GlycoBlue co-precipitant.
645 Footprints were pelleted by 1 h centrifugation at 20,000 g and 4°C, briefly washed with

646 ice-cold 70% ethanol and centrifuged again for 10 min at the same parameters. The
647 resulting pellet was briefly dried and dissolved in 36.5 µL of nuclease-free water.
648 For sequencing library preparation, ribosome footprints were denatured at 65°C for 5 min
649 in a thermocycler and immediately transferred to ice. The sample was then mixed with 5
650 µL 10 x T4-polynucleotide kinase buffer A, 2.5 µL T4-polynucleotide kinase (Thermo
651 Scientific), 2 µL SUPERase•In RNase Inhibitor and incubated for 10 min at 37°C.
652 Subsequently, 5 µL of 10 mM ATP was added and the sample was incubated for additional
653 30 min at 37°C. After phosphorylation, the footprints were mixed with 150 µL 100 mM
654 NaCl and purified again by phenol-chloroform precipitation and final precipitation
655 overnight at -20°C using 2.5 volumes of ethanol and 2 µL GlycoBlue co-precipitant.
656 Footprints were pelleted by 1 h centrifugation at 20,000 g and 4°C, briefly washed with
657 ice-cold 70% ethanol and centrifuged again for 10 min at the same parameters. The
658 resulting pellet was briefly dried and dissolved in 11.5 µL of nuclease-free water. The
659 whole sample was used as input for the NEXTFLEX Small RNA-seq Kit v3 (Perkin Elmer)
660 to generate libraries according to the kit's manual. For RNA-seq, the purified RNA was
661 later used as input for the Zymo-Seq RiboFree Total RNA library kit for library preparation
662 according to the manufacturer's manual.
663 *Processing of raw sequencing data*
664 Demultiplexed FASTQ files were derived from an Illumina NextSeq 550 system aiming for
665 40 M reads per footprint sample and 20 M reads per RNA sample at a read length of 75
666 nt. Samples were processed through a custom-made pipeline in form of a Linux shell
667 script calling several bioinformatic tools and custom Python scripts. First operation of the
668 pipeline was removing 3' adapter sequences using cutadapt for RNA-seq and Ribo-seq,
669 respectively (Martin, 2011) with parameters [--minimum-length=9, -

670 a="TGGAATTCTCGGGTGCCAAGG" ]. After removal of 3' adapter sequences, the 3' and
671 5' unique molecular identifier tags (UMI tags) were removed from each read's sequence
672 and shifted to the rear of the header row by UMI tools (Smith et al., 2017) with parameters
673 [extract --bc-pattern=NNNN] for 5' UMI tags and [extract --bc-pattern=NNNN –3prime] for
674 3' UMI tags. Remaining reads were filtered for sequence length using cutadapt again with
675 parameters [--minimum length=20 --maximum-length=39] to remove any read that is
676 unlikely to represent a real ribosome footprint. After size filtering, a custom-made Python
677 script corrected the UMI-tags that were previously shifted to the read headers in the form
678 [header line_NNNN_NNNN] to [header line_NNNNNNNN], allowing UMI tools in a later
679 step to deduplicate the data set based on both UMI tags. In a first mapping step, reads
680 were aligned to a set of *C. reinhardtii* non-coding RNA FASTA sequences provided by
681 ENSEMBL plants together with the genome version 5.5 to remove contaminating RNA
682 species using STAR with parameters [--runMode alignReads --outSAMtype BAM
683 SortedByCoordinate --outFilterMultimapNmax 20 --outFilterMismatchNmax 3 --
684 outReadsUnmapped Fastx --alignIntronMax 7000 --twopassMode Basic]. Reads that did
685 not align to the ncRNA set were copied to a new FASTQ file that was mapped against the
686 *C. reinhardtii* genome version 6.1 using STAR (Dobin et al., 2013) with parameters [--
687 runMode alignReads –outSAMtype BAM
688 SortedByCoordinate --outFilterMismatchNoverLmax 0.1 --outReadsUnmapped Fastx --
689 alignIntronMax 8000 –outSAMmultNmax 1 –outMultimapperOrder Random]. After
690 genome mapping, the resulting BAM file was deduplicated using UMI-tools' dedup
691 function and afterwards indexed using samtools' index function (Li et al., 2009).
692 *Data processing*

693 All analyses on the NGS data sets including calculations of gene body coverage,
694 periodicity, P-site offsets, organellar read length distributions, ribosome profiles and gene
695 expression metrics were performed using a custom-made object-oriented Python module
696 that makes extensive use of the packages pysam/HTSlib (https://github.com/pysam-
697 developers/pysam), pandas (McKinney, 2010), NumPy (Harris et al., 2020), SciPy
698 (Virtanen et al., 2020), matplotlib (Hunter, 2007) and seaborn (Waskom, 2021) among
699 multiple others. For CPM and RPKM calculations in Ribo-seq data sets, only reads
700 mapping to the annotated CDS of a transcript were considered. Although we are
701 convinced that RPKM is not a suitable measure to quantify gene expression on the
702 translatome level, we also calculated RPKM values for the interested reader based on the
703 length of the annotated ORF of a transcript only (Supplemental Dataset). For any
704 calculations on RNA-seq data sets, reads mapping to the whole transcript were
705 considered and also the full length of the mature transcript was used for RPKM
706 calculations. Translation efficiency values were simply calculated by dividing an CDSs
707 CPM value from a footprint data set by the respective transcripts RPKM value from the
708 affiliated RNA-seq data set.

709 For the 5'-P-site / 3'-A-site offset calculations, all reads fully enclosing either an
710 annotated start codon (5'-P-site offset) or an annotated stop codon (for 3'-A-site offset)
711 were taken into consideration separately for the nuclear and the chloroplast genome.
712 These reads were sorted according to the length of their alignment to the genome, which
713 was considered as read length to avoid miscalculations due to clipped or masked
714 sequences. For every read length species, the position of the reads 5' end or 3' end
715 relative to the position of the first base of the start codon (for 5'-P-site offsets) or the last
716 base of the stop codon (for 3'-A-site offsets) was determined as the offset. Afterwards, the

717 most frequently occurring offset for a read length species was determined to be the true
718 offset of this species and was used in the P-site calculation. To display the results of the
719 offset calculations, the count matrix of both 5'-P-site offsets and 3'-A-site offsets were
720 normalized separately to the total number of counts each and plotted as heatmap.

721 For calculation of the P-sites first position of reads mapping to nucleus-encoded
722 CDSs, all reads mapping to a CDS were filtered from the data set and for every read, the
723 5'-most mapping position of the read's alignment + the 5'-P-site offset determined for the
724 reads alignment length was determined to be the first P-site position in case of transcripts
725 encoded on the + strand. For reads encoded on the minus strand, the first P-site position
726 was determined to be the 3'-most mapping position of the RPF alignment minus the
727 determined 5'-P-site offset. In all cases, splicing was considered and whenever a spliced
728 read was found, the calculation was adjusted according to the number of genomic
729 positions spanned by the splice junction of the read. For reads mapping to chloroplast-
730 encoded transcripts on the + strand, the first P-site position was determined to be the 3'-
731 most mapping position – (3'-A-site offset + 5). For reads mapping to chloroplast-encoded
732 transcripts on the - strand, the first P-site position was calculated to be the 5'-most
733 mapping position + (3'-A-site offset + 5). The metagene P-site profiles of nucleus-encoded
734 transcripts were calculated considering only the 2,000 most expressed transcripts of a
735 dataset, according to their CPM values. If multiple splice variants of the same gene
736 occurred in the list, only the longest variant (annotated as transcript 1 in the genome) was
737 considered for the calculation to avoid analyzing the same reads multiple times. The
738 profiles were constructed by calculating the first position of the P-site for all reads mapping
739 to the first 30 and the last 30 codons of considered transcript CDSs. For a ribosome
740 footprint to be in-frame with the annotated CDS, the first position of the P-site must be

741 located on the first base of a codon. According to this logic, the calculated P-site positions
742 were categorized into “in-frame”, “+1 shifted” and “+2 shifted” and summed up for all
743 codons of considered transcript CDSs. To calculate the frame preference per codon, the
744 counts in each category in a codon were divided by the sum of all categories for every
745 codon separately. For calculation of the chloroplast metagene P-site profile, all protein-
746 coding chloroplast transcripts were considered.

747 The extension of gene body coverage versus a respective cut-off was calculated
748 as the portion of an CDS that was covered by the number of reads equal to or greater
749 than the respective cut-off. This procedure was applied to all annotated CDSs using
750 increasing cut-offs from 1 to 10 reads to monitor the stepwise decrease of gene-body
751 coverage. For every applied cut-off, the CDSs were categorized as being covered over
752 >10% of their length to >90% of their length in 10% steps and CDSs falling into every
753 category were counted and displayed in a heat map to estimate both completeness and
754 strength of CDS coverage in the data set.

755 To calculate the read length distribution, all reads mapping to the respective
756 genome were filtered and the length of their alignment to the genome was taken as read
757 length. After counting the number of all possible read lengths for each genome separately,
758 these numbers were normalized by the total number of reads mapping to the respective
759 genome. For read biotype distribution, the reads were further categorized if they map to
760 an annotated coding sequence, 5’- or 3’-UTR or intergenic region. To handle reads
761 overlapping two regions of different category, for example 5’-UTR and coding sequence
762 as seen regularly in initiation peaks, the read was assigned the category that had the
763 larger share on the read’s alignment. In rare cases, when a read mapped to different

764 categories in two different splice variants of the same transcript, the read was assigned to
765 both categories.

766 **Supplemental data**

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

768 **Supplemental Figure S2:** Transcript accumulation, translation output and translation
769 efficiency of nuclear genes.

770 **Supplemental Figure S3:** Correlation between nuclease digestion conditions, transcript
771 accumulation and translation.

772 **Supplemental Figure S4:** Correlation analysis of RPF count between nuclease digest
773 conditions.

774 **Supplemental Figure S5:** Detection of abnormal initiation peaks in the three genomes of
775 *Chlamydomonas reinhardtii.*

776 **Supplemental Dataset.** List of gene body coverage, RPKM and CPM of Ribo-seq and
777 RNA-seq experiments.

778 **Acknowledgements**

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

784 **Funding**

785 This work was supported by the Deutsche Forschungsgemeinschaft grant 437345987 (WI
786 3477/3-1, ZO 302/5-1, RU 2266/2-1 to F.W., R.Z. and S.R., respectively) and the
787 Forschungsschwerpunkt BioComp to F.W.

788 **Author contribution:**

789 V.L.G. performed ribosome profiling experiments, bioinformatic data analyses and wrote
790 parts of the manuscript. M.K.Y.T. and R.Z. helped with setting up ribosome profiling, N.H.
791 and S.R. conducted correlation analyses. F.W. designed experiments, analyzed data, and
792 wrote the manuscript.

793 **Data availability**

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

798 **Competing interests**

799 The authors declare no competing interests.

800 **References**

801 **Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C, Jaffe DB, Nusbaum**
802 **C, Gnirke A** (2011) Analyzing and minimizing PCR amplification bias in Illumina
803 sequencing libraries. Genome Biol **12:** R18
804 **Bertrand EM, Allen AE, Dupont CL, Norden-Krichmar TM, Bai J, Valas RE, Saito MA**
805 (2012) Influence of cobalamin scarcity on diatom molecular physiology and
806 identification of a cobalamin acquisition protein. Proc Natl Acad Sci U S A **109:**
807 E1762-1771
808 **Blifernez-Klassen O, Berger H, Mittmann BGK, Klassen V, Schelletter L, Buchholz**
809 **T, Baier T, Soleimani M, Wobbe L, Kruse O** (2021) A gene regulatory network

810 for antenna size control in carbon dioxide-deprived *Chlamydomonas reinhardtii*
811 cells. Plant Cell **33:** 1303-1318
812 **Bonente G, Pippa S, Castellano S, Bassi R, Ballottari M** (2012) Acclimation of
813 *Chlamydomonas reinhardtii* to different growth irradiances. J Biol Chem **287:** 5833-
814 5847
815 **Brar GA, Weissman JS** (2015) Ribosome profiling reveals the what, when, where and
816 how of protein synthesis. Nat Rev Mol Cell Biol **16:** 651-664
817 **Bunbury F, Helliwell KE, Mehrshahi P, Davey MP, Salmon DL, Holzer A, Smirnoff N,**
818 **Smith AG** (2020) Responses of a Newly Evolved Auxotroph of Chlamydomonas
819 to B(12) Deprivation. Plant Physiol **183:** 167-178
820 **Calviello L, Mukherjee N, Wyler E, Zauber H, Hirsekorn A, Selbach M, Landthaler M,**
821 **Obermayer B, Ohler U** (2016) Detecting actively translated open reading frames
822 in ribosome profiling data. Nat Methods **13:** 165-170
823 **Cao X, Slavoff SA** (2020) Non-AUG start codons: Expanding and regulating the small
824 and alternative ORFeome. Exp Cell Res **391:** 111973
825 **Chartron JW, Hunt KC, Frydman J** (2016) Cotranslational signal-independent SRP
826 preloading during membrane targeting. Nature **536:** 224-228
827 **Chen H, Alonso JM, Stepanova AN** (2022) A Ribo-Seq Method to Study Genome-Wide
828 Translational Regulation in Plants. Methods Mol Biol **2494:** 61-98
829 **Chiu CW, Li YR, Lin CY, Yeh HH, Liu MJ** (2022) Translation initiation landscape profiling
830 reveals hidden open-reading frames required for the pathogenesis of tomato yellow
831 leaf curl Thailand virus. Plant Cell **34:** 1804-1821
832 **Chotewutmontri P, Barkan A** (2016) Dynamics of chloroplast translation during
833 chloroplast differentiation in maize. PLoS Genet **12:** e1006106
834 **Chotewutmontri P, Barkan A** (2018) Multilevel effects of light on ribosome dynamics in
835 chloroplasts program genome-wide and *psbA*-specific changes in translation.
836 PLoS Genet **14:** e1007555
837 **Chotewutmontri P, Barkan A** (2020) Light-induced *psbA* translation in plants is triggered
838 by photosystem II damage via an assembly-linked autoregulatory circuit. Proc Natl
839 Acad Sci U S A **117:** 21775-21784
840 **Chung BY, Hardcastle TJ, Jones JD, Irigoyen N, Firth AE, Baulcombe DC, Brierley I**
841 (2015) The use of duplex-specific nuclease in ribosome profiling and a user-friendly
842 software package for Ribo-seq data analysis. RNA **21:** 1731-1745
843 **Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A,**
844 **Szczesniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A** (2016) A survey of
845 best practices for RNA-seq data analysis. Genome Biol **17:** 13
846 **Costanzo MC, Hogan JD, Cusick ME, Davis BP, Fancher AM, Hodges PE, Kondu P,**
847 **Lengieza C, Lew-Smith JE, Lingner C, Roberg-Perez KJ, Tillberg M, Brooks**
848 **JE, Garrels JI** (2000) The yeast proteome database (YPD) and *Caenorhabditis*
849 *elegans* proteome database (WormPD): comprehensive resources for the
850 organization and comparison of model organism protein information. Nucleic Acids
851 Res **28:** 73-76
852 **Craig RJ, Gallaher SD, Shu S, Salome P, Jenkins JW, Blaby-Haas CE, Purvine SO,**
853 **O'Donnell S, Barry K, Grimwood J, Strenkert D, Kropat J, Daum C, Yoshinaga**
854 **Y, Goodstein DM, Vallon O, Schmutz J, Merchant SS** (2022) The
855 Chlamydomonas Genome Project, version 6: reference assemblies for mating type
856 *plus* and *minus* strains reveal extensive structural mutation in the laboratory. Plant
857 Cell

858 **Cross FR** (2015) Tying Down Loose Ends in the *Chlamydomonas* Genome: Functional
859 Significance of Abundant Upstream Open Reading Frames. G3 (Bethesda) **6:** 435-
860 446
861 **Dall'Osto L, Bressan M, Bassi R** (2015) Biogenesis of light harvesting proteins. Biochim
862 Biophys Acta **1847:** 861-871
863 **Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson**
864 **M, Gingeras TR** (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics
865 **29:** 15-21
866 **Döring K, Ahmed N, Riemer T, Suresh HG, Vainshtein Y, Habich M, Riemer J, Mayer**
867 **MP, O'Brien EP, Kramer G, Bukau B** (2017) Profiling Ssb-Nascent Chain
868 Interactions Reveals Principles of Hsp70-Assisted Folding. Cell **170:** 298-311 e220
869 **Eberhard S, Drapier D, Wollman FA** (2002) Searching limiting steps in the expression
870 of chloroplast-encoded proteins: relations between gene copy number,
871 transcription, transcript abundance and translation rate in the chloroplast of
872 *Chlamydomonas reinhardtii*. The Plant journal : for cell and molecular biology **31:**
873 149-160
874 **Floris M, Bassi R, Robaglia C, Alboresi A, Lanet E** (2013) Post-transcriptional control
875 of light-harvesting genes expression under light stress. Plant Mol Biol **82:** 147-154
876 **Forsythe ES, Grover CE, Miller ER, Conover JL, Arick MA, 2nd, Chavarro MCF, Leal-**
877 **Bertioli SCM, Peterson DG, Sharbrough J, Wendel JF, Sloan DB** (2022)
878 Organellar transcripts dominate the cellular mRNA pool across plants of varying
879 ploidy levels. Proc Natl Acad Sci U S A **119:** e2204187119
880 **Frank J, Agrawal RK** (2000) A ratchet-like inter-subunit reorganization of the ribosome
881 during translocation. Nature **406:** 318-322
882 **Fu Y, Wu PH, Beane T, Zamore PD, Weng Z** (2018) Elimination of PCR duplicates in
883 RNA-seq and small RNA-seq using unique molecular identifiers. BMC Genomics
884 **19:** 531
885 **Fujita T, Kurihara Y, Iwasaki S** (2019) The Plant Translatome Surveyed by Ribosome
886 Profiling. Plant Cell Physiol **60:** 1917-1926
887 **Gallaher SD, Fitz-Gibbon ST, Strenkert D, Purvine SO, Pellegrini M, Merchant SS**
888 (2018) High-throughput sequencing of the chloroplast and mitochondrion of
889 *Chlamydomonas reinhardtii* to generate improved de novo assemblies, analyze
890 expression patterns and transcript speciation, and evaluate diversity among
891 laboratory strains and wild isolates. Plant J **93:** 545-565
892 **Gao Y, Thiele W, Saleh O, Scossa F, Arabi F, Zhang H, Sampathkumar A, Kuhn K,**
893 **Fernie A, Bock R, Schottler MA, Zoschke R** (2022) Chloroplast translational
894 regulation uncovers nonessential photosynthesis genes as key players in plant cold
895 acclimation. Plant Cell **34:** 2056-2079
896 **Gawronski P, Jensen PE, Karpinski S, Leister D, Scharff LB** (2018) Plastid ribosome
897 pausing is induced by multiple features and is linked to protein complex assembly.
898 Plant Physiol
899 **Gerashchenko MV, Gladyshev VN** (2014) Translation inhibitors cause abnormalities in
900 ribosome profiling experiments. Nucleic Acids Res **42:** e134
901 **Gerashchenko MV, Gladyshev VN** (2017) Ribonuclease selection for ribosome profiling.
902 Nucleic Acids Res **45:** e6
903 **Gloge F, Becker AH, Kramer G, Bukau B** (2014) Co-translational mechanisms of protein
904 maturation. Curr Opin Struct Biol **24:** 24-33

905 **Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D,**
906 **Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk**
907 **MH, Brett M, Haldane A, Del Rio JF, Wiebe M, Peterson P, Gerard-Marchant**
908 **P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE**
909 (2020) Array programming with NumPy. Nature **585:** 357-362
910 **Hinnebusch AG** (2011) Molecular mechanism of scanning and start codon selection in
911 eukaryotes. Microbiol Mol Biol Rev **75:** 434-467, first page of table of contents
912 **Hsu PY, Calviello L, Wu HL, Li FW, Rothfels CJ, Ohler U, Benfey PN** (2016) Super-
913 resolution ribosome profiling reveals unannotated translation events in
914 Arabidopsis. Proc Natl Acad Sci U S A
915 **Hunter J** (2007) Matplotlib: A 2D Graphics Environment. Computing in Science &
916 Engineering **9:** 90-95
917 **Ingolia NT** (2014) Ribosome profiling: new views of translation, from single codons to
918 genome scale. Nat Rev Genet **15:** 205-213
919 **Ingolia NT** (2016) Ribosome Footprint Profiling of Translation throughout the Genome.
920 Cell **165:** 22-33
921 **Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS** (2012) The ribosome
922 profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-
923 protected mRNA fragments. Nat Protoc **7:** 1534-1550
924 **Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS** (2009) Genome-wide
925 analysis in vivo of translation with nucleotide resolution using ribosome profiling.
926 Science **324:** 218-223
927 **Ingolia NT, Lareau LF, Weissman JS** (2011) Ribosome profiling of mouse embryonic
928 stem cells reveals the complexity and dynamics of mammalian proteomes. Cell
929 **147:** 789-802
930 **Juntawong P, Girke T, Bazin J, Bailey-Serres J** (2014) Translational dynamics revealed
931 by genome-wide profiling of ribosome footprints in *Arabidopsis*. Proc Natl Acad Sci
932 U S A **111:** E203-212
933 **Klimmek F, Sjodin A, Noutsos C, Leister D, Jansson S** (2006) Abundantly and rarely
934 expressed *Lhc* protein genes exhibit distinct regulation patterns in plants. Plant
935 Physiol **140:** 793-804
936 **Komine Y, Kwong L, Anguera MC, Schuster G, Stern DB** (2000) Polyadenylation of
937 three classes of chloroplast RNA in *Chlamydomonas reinhadtii*. RNA **6:** 598-607
938 **Kramer G, Shiber A, Bukau B** (2019) Mechanisms of Cotranslational Maturation of
939 Newly Synthesized Proteins. Annu Rev Biochem **88:** 337-364
940 **Kropat J, Hong-Hermesdorf A, Casero D, Ent P, Castruita M, Pellegrini M, Merchant**
941 **SS, Malasarn D** (2011) A revised mineral nutrient supplement increases biomass
942 and growth rate in *Chlamydomonas reinhardtii*. Plant J **66:** 770-780
943 **Lareau LF, Hite DH, Hogan GJ, Brown PO** (2014) Distinct stages of the translation
944 elongation cycle revealed by sequencing ribosome-protected mRNA fragments.
945 Elife **3:** e01257
946 **Lauria F, Tebaldi T, Bernabo P, Groen EJN, Gillingwater TH, Viero G** (2018) riboWaltz:
947 Optimization of ribosome P-site positioning in ribosome profiling data. PLoS
948 Comput Biol **14:** e1006169
949 **Lei L, Shi J, Chen J, Zhang M, Sun S, Xie S, Li X, Zeng B, Peng L, Hauck A, Zhao H,**
950 **Song W, Fan Z, Lai J** (2015) Ribosome profiling reveals dynamic translational
951 landscape in maize seedlings under drought stress. Plant J **84:** 1206-1218

952 **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G,**
953 **Durbin R, Genome Project Data Processing S** (2009) The Sequence
954 Alignment/Map format and SAMtools. Bioinformatics **25:** 2078-2079
955 **Liu MJ, Wu SH, Wu JF, Lin WD, Wu YC, Tsai TY, Tsai HL, Wu SH** (2013) Translational
956 landscape of photomorphogenic *Arabidopsis*. Plant Cell **25:** 3699-3710
957 **Lukoszek R, Feist P, Ignatova Z** (2016) Insights into the adaptive response of
958 *Arabidopsis thaliana* to prolonged thermal stress by ribosomal profiling and RNA-
959 Seq. BMC Plant Biol **16:** 221
960 **Martin M** (2011) Cutadapt removes adapter sequences from high-throughput sequencing
961 reads. 2011 **17:** 3
962 **McKinney W** (2010) Data Structures for Statistical Computing in Python. *In* Proceedings
963 of the 9th Python in Science Conference, pp 56-61
964 **Merchante C, Brumos J, Yun J, Hu Q, Spencer KR, Enriquez P, Binder BM, Heber S,**
965 **Stepanova AN, Alonso JM** (2015) Gene-specific translation regulation mediated
966 by the hormone-signaling molecule EIN2. Cell **163:** 684-697
967 **Moulin M, Nguyen GT, Scaife MA, Smith AG, Fitzpatrick TB** (2013) Analysis of
968 *Chlamydomonas* thiamin metabolism in vivo reveals riboswitch plasticity. Proc Natl
969 Acad Sci U S A **110:** 14622-14627
970 **Mussgnug JH, Wobbe L, Elles I, Claus C, Hamilton M, Fink A, Kahmann U,**
971 **Kapazoglou A, Mullineaux CW, Hippler M, Nickelsen J, Nixon PJ, Kruse O**
972 (2005) NAB1 is an RNA binding protein involved in the light-regulated differential
973 expression of the light-harvesting antenna of *Chlamydomonas reinhardtii*. Plant
974 Cell **17:** 3409-3421
975 **Nickelsen J, Bohne A-V, Westhoff P** (2014) Chloroplast gene expression - translation.
976 49-78
977 **Payne SH** (2015) The utility of protein and mRNA correlation. Trends Biochem Sci **40:** 1-
978 3
979 **Pechmann S, Willmund F, Frydman J** (2013) The ribosome as a hub for protein quality
980 control. Mol Cell **49:** 411-421
981 **Plancke C, Vigeolas H, Hohner R, Roberty S, Emonds-Alt B, Larosa V, Willamme R,**
982 **Duby F, Onga Dhali D, Thonart P, Hiligsmann S, Franck F, Eppe G, Cardol P,**
983 **Hippler M, Remacle C** (2014) Lack of isocitrate lyase in *Chlamydomonas* leads to
984 changes in carbon metabolism and in the response to oxidative stress under
985 mixotrophic growth. Plant J **77:** 404-417
986 **Rooijers K, Loayza-Puch F, Nijtmans LG, Agami R** (2013) Ribosome profiling reveals
987 features of normal and disease-associated mitochondrial translation. Nat Commun
988 **4:** 2886
989 **Russell JB, Cook GM** (1995) Energetics of bacterial growth: balance of anabolic and
990 catabolic reactions. Microbiol Rev **59:** 48-62
991 **Schneider-Poetsch T, Ju J, Eyler DE, Dang Y, Bhat S, Merrick WC, Green R, Shen**
992 **B, Liu JO** (2010) Inhibition of eukaryotic translation elongation by cycloheximide
993 and lactimidomycin. Nat Chem Biol **6:** 209-217
994 **Schroda M, Hemme D, Mühlhaus T** (2015) The *Chlamydomonas* heat stress response.
995 Plant J **82:** 466-480
996 **Schuster M, Gao Y, Schöttler MA, Bock R, Zoschke R** (2019) Limited Responsiveness
997 of Chloroplast Gene Expression during Acclimation to High Light in Tobacco. Plant
998 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**

**Trösch R, Barahimipour R, Gao Y, Badillo-Corona JA, Gotsmann VL, Zimmer D, Mühlhaus T, Zoschke R, Willmund F** (2018) Commonalities and differences of chloroplast translation in a green alga and land plants. Nat Plants **4:** 564-575

**Trösch R, Ries F, Westrich LD, Gao Y, Herkt C, Hoppstädter J, Heck-Roth J, Mustas M, Scheuring D, Choquet Y, Räschle M, Zoschke R, Willmund F** (2022) Fast and global reorganization of the chloroplast protein biogenesis network during heat acclimation. Plant Cell **34:** 1075-1099

**Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat I, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P, SciPy C** (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods **17:** 261-272

**Wang R, Amoyel M** (2022) mRNA Translation Is Dynamically Regulated to Instruct Stem Cell Fate. Front Mol Biosci **9:** 863885

**Waskom M** (2021) Seaborn: statistical data visualization. Journal of Open Source Software **6:** 3021

**Westrich LD, Gotsmann VL, Herkt C, Ries F, Kazek T, Trosch R, Armbruster L, Mühlenbeck JS, Ramundo S, Nickelsen J, Finkemeier I, Wirtz M, Storchova Z, Räschle M, Willmund F** (2021) The versatile interactome of chloroplast ribosomes revealed by affinity purification mass spectrometry. Nucleic Acids Res **49:** 400-415

**Wolin SL, Walter P** (1988) Ribosome pausing and stacking during translation of a eukaryotic mRNA. EMBO J **7:** 3559-3569

**Wu CC, Zinshteyn B, Wehner KA, Green R** (2019) High-Resolution Ribosome Profiling Defines Discrete Ribosome Elongation States and Translational Regulation during Cellular Stress. Mol Cell **73:** 959-970 e955

**Wu HL, Song G, Walley JW, Hsu PY** (2019) The Tomato Translational Landscape Revealed by Transcriptome Assembly and Ribosome Profiling. Plant Physiol **181:** 367-380

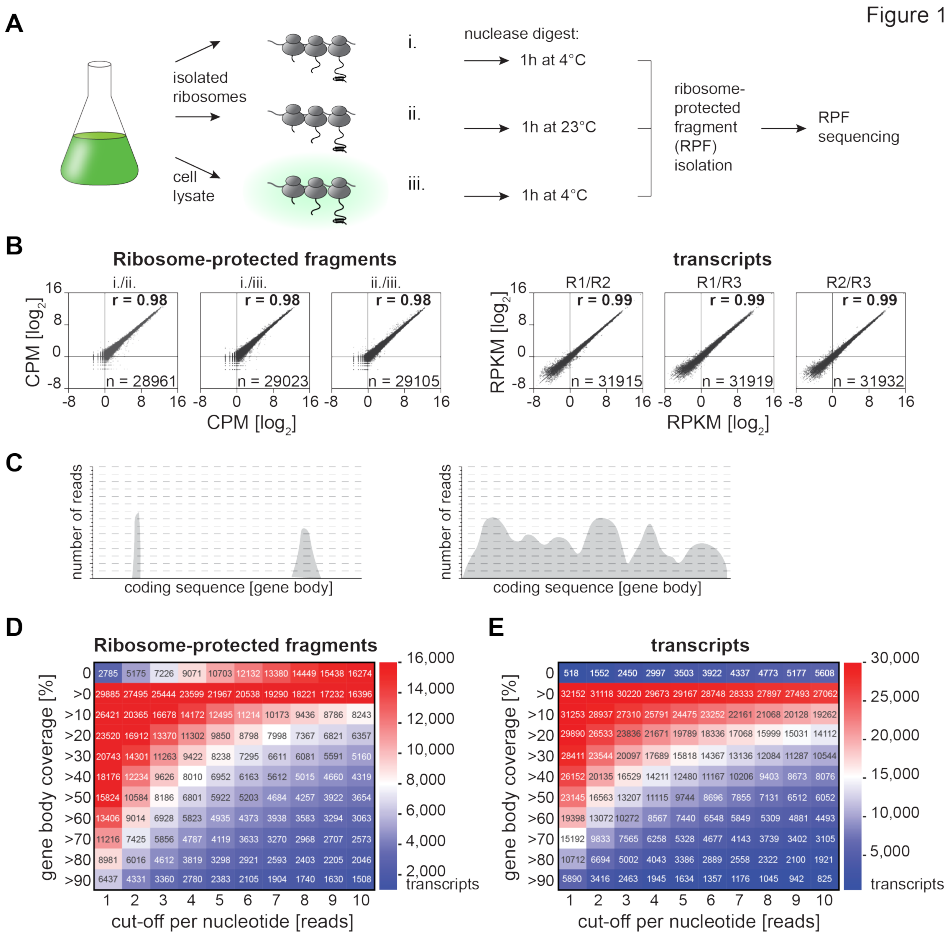
**Yang X, Cui J, Song B, Yu Y, Mo B, Liu L** (2020) Construction of High-Quality Rice Ribosome Footprint Library. Front Plant Sci **11:** 572237

**Yang X, Song B, Cui J, Wang L, Wang S, Luo L, Gao L, Mo B, Yu Y, Liu L** (2021) Comparative ribosome profiling reveals distinct translational landscapes of salt-sensitive and -tolerant rice. BMC Genomics **22:** 612

**Zhang W, Dunkle JA, Cate JH** (2009) Structures of the ribosome in intermediate states of ratcheting. Science **325:** 1014-1017

1047 **Zoschke R, Bock R** (2018) Chloroplast Translation: Structural and Functional
1048 Organization, Operational Control and Regulation. Plant Cell **30:** 745-770
1049

1050 **Figures and legends**




1051

1052 **Figure 1: Ribo-seq provides a broad coverage of expressed genes**
1053 (A) Illustration of the experimental setup, testing nucleolytic conditions. Polysome
1054 purification and subsequent RNase I digest for 1h at 4°C (i.) or 1h at 23°C (ii.) was
1055 performed from Chlamydomonas samples, respectively. Alternatively, cell lysate was
1056 directly supplemented with RNase I (iii.). (B) Scatter plots representing the reproducibility
1057 of Ribo-seq (left panels) and RNA-seq experiments (right panels). Shown are the R-values
1058 and CPM/RPKM values. (C) Schematic representation of the concept of gene body
1059 coverage (left panel: incomplete coverage that strongly declines with an increasing read
1060 cut-off; right panel: good coverage that is robust against increasing read cut-offs). (D) and
1061 (E) Cumulative heatmaps visualizing the number of genes (given in the respective panel)
1062 covered to a certain minimal extent (rows) by a certain minimal number of reads (columns)
1063 for ribosome footprints (C) and transcripts (D). Ribo-seq data shown are derived from the
1064 RNase I digest in lysate (condition iii.). The respective data for the other treatments are
1065 shown in Supplemental Figure 1.

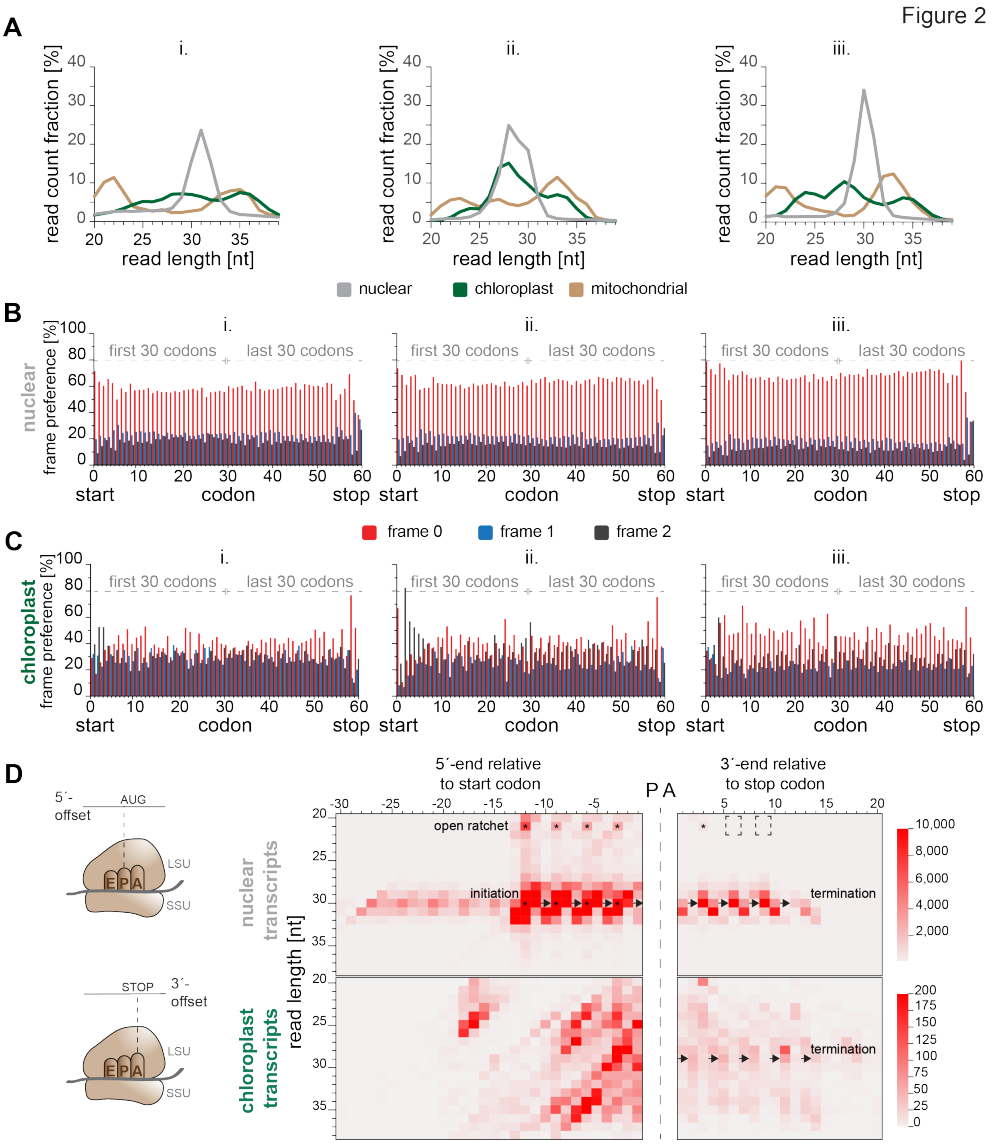



**Figure 2: Tri-nucleotide periodicity of cellular translation**
(A) Length distribution of RPFs for each organelle from the three different nuclease treatments tested (i. and ii.: digest of purified ribosomes at 4°C and 23 °C, respectively, iii.: digest in lysate at 4°C). (B) and (C) metagene relative 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) 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.

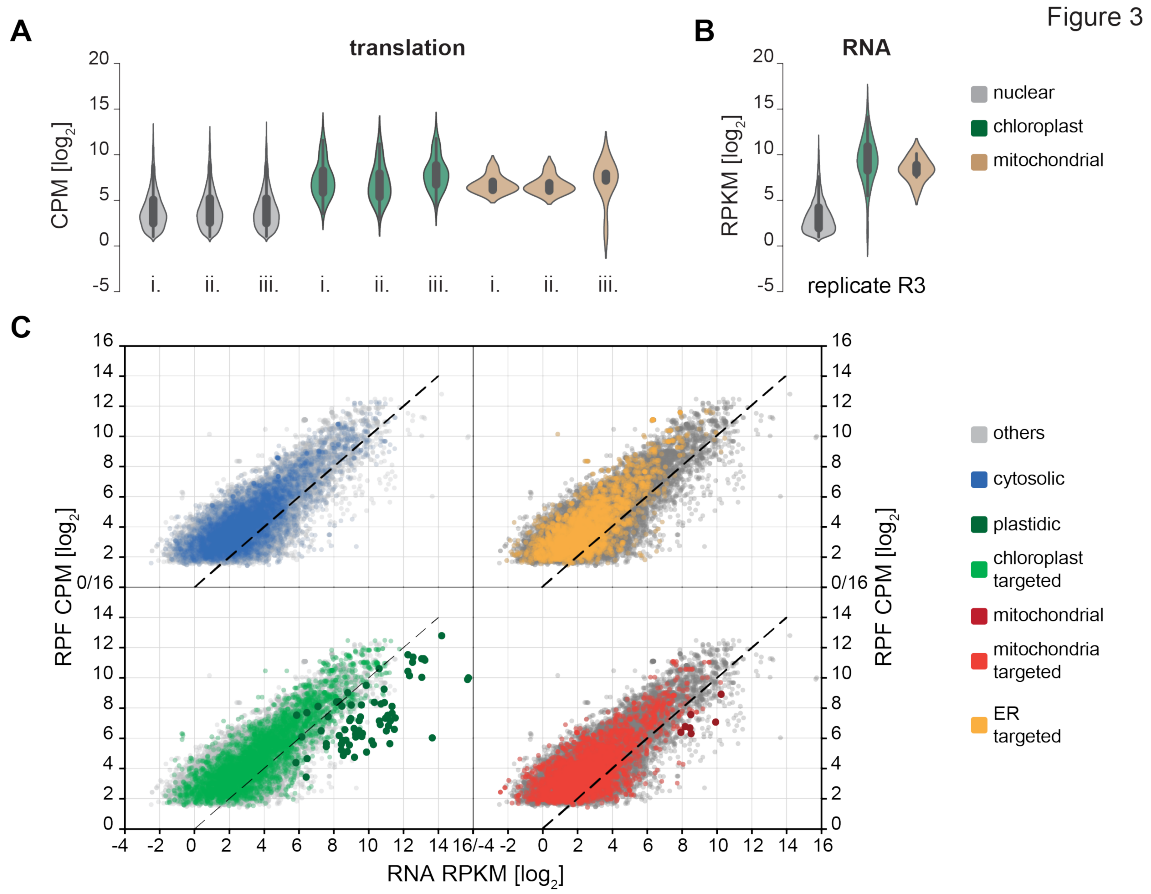



**Figure 3: Chloroplast-encoded proteins are highly expressed**
(A) Violin plots of the RPF data, displaying the CPM (count per million reads) distribution for genes of each plant cell genome, following the three different RNase I treatments (i. and ii.: digest of purified ribosomes at 4°C and 23°C, respectively, iii.: digest in lysate at 4°C). Values with a CPM 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 (CPM) and RNA-seq (RPKM) data for different subcellular-localized proteins. Coloring indicates subcellular localization and origin of the encoded proteins.

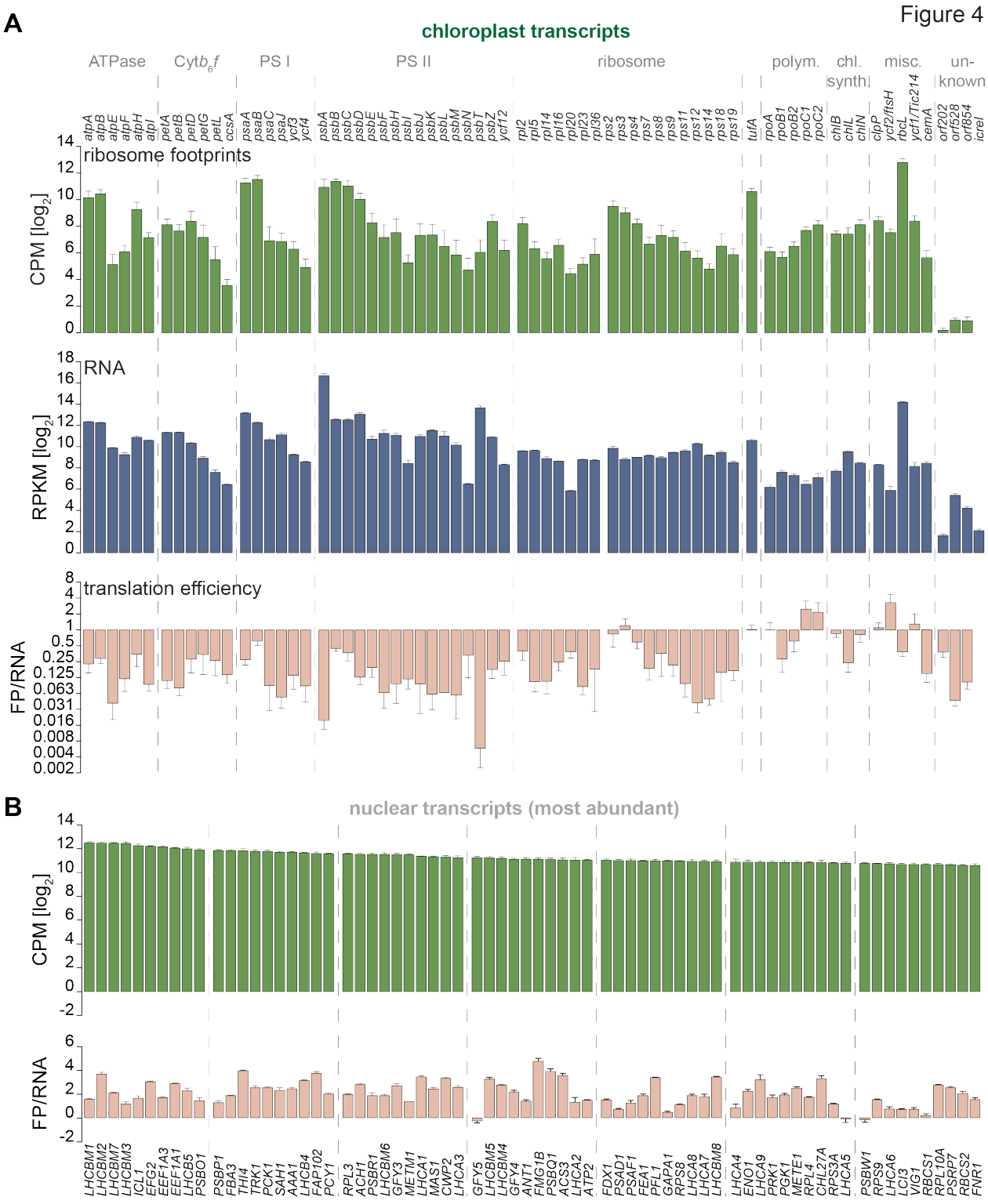
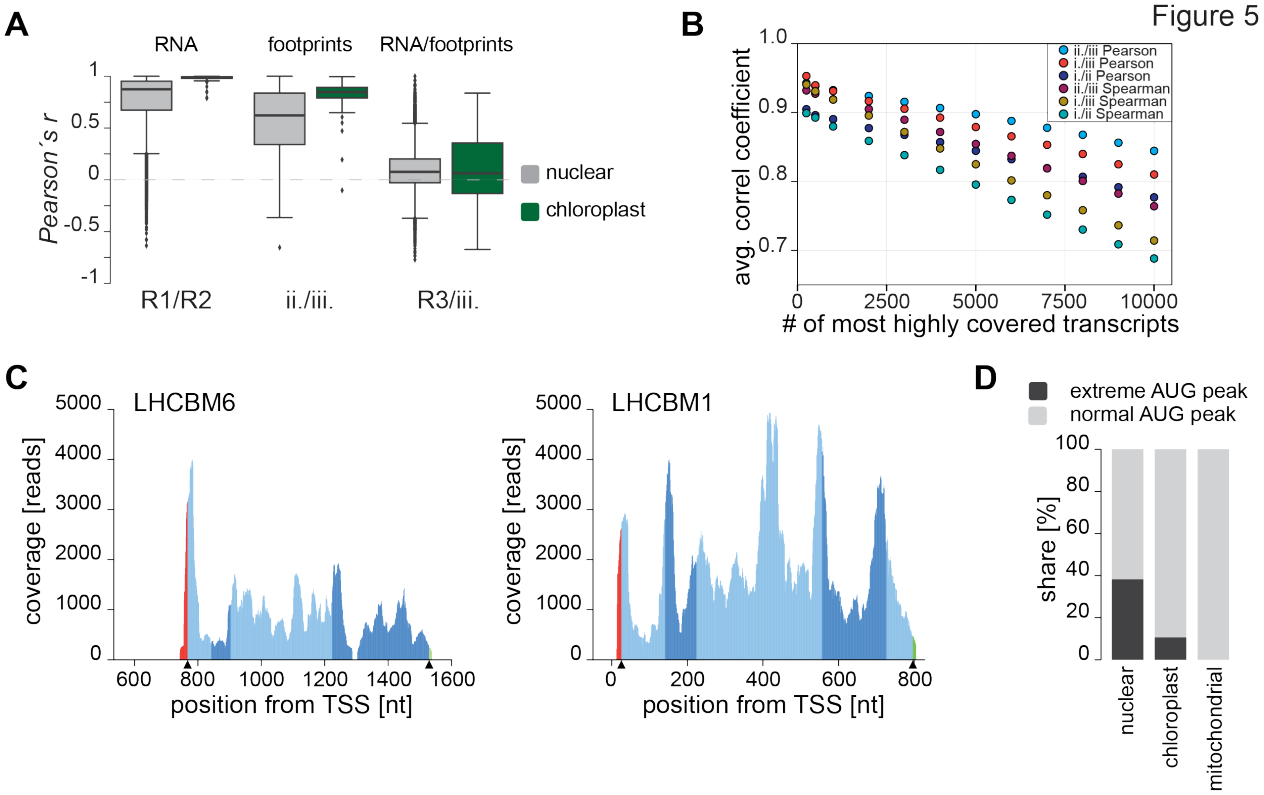



**Figure 4: Translation output of chloroplast- and nucleus-encoded transcripts**
(A) Mean read intensity/TE ratio (per respective coding sequence) of chloroplast-encoded transcripts on the level of RPF, RNA and translation efficiency. The transcripts are grouped by function of the encoded proteins (Cyt*b*$_6$*f*= cytochrome *b*$_6$*f* complex, PS I = photosystem I, PS II = photosystem II, polym. = RNA polymerase, chl. synth. = chlorophyll

1102 synthesis, misc. = miscellaneous). (B) Values of the 70 most highly expressed nucleus-
1103 encoded genes on the footprint level. Only the longest splice variants were considered.
1104 Respective translation efficiency for these transcripts is given below. For comparison,
1105 transcripts with the top 70 highest TE values are given in Supplementary Figure S2C.
1106 Error bars indicate standard deviation of the three experiments (treatment i. to iii. for
1107 ribosome footprint, and biological replicate 1-3 for RNA-seq), and are shown unidirectional
1108 to improve readability.
1109

1110




1111

1112 **Figure 5: Putative translational regulation of individual transcripts**
1113 (A) Boxplots representing the distribution of Pearson's r correlation coefficients comparing
1114 the RNA gene body coverage and ribosome profiles of all chloroplast- and nucleus-
1115 encoded transcripts between replicates (R1 and R2 for RNA) and between two types of
1116 RNase I treatments (ii and iii for ribosome footprints). (B) Average *Pearson* and *Spearman*
1117 correlation coefficients depending on the number n of considered transcripts. Transcripts
1118 were ordered by descending RPF count per transcript, n most highly covered transcripts
1119 were selected, correlation coefficients were computed for each transcript in this subset,
1120 and the average was plotted. For correlation coefficients of individual transcripts see
1121 Supplemental Figure S4A. (C) Ribosome profiles of LHCBM1 exemplary for a profile with
1122 even footprint distribution and LHCBM6 as an example for a profile with an extreme
1123 initiation peak. Red parts of the profile represent coverage within the 5'-UTRs, green parts
1124 represent 3'-UTRs and alternating blue shades represent exons within the CDS. Black
1125 arrows on the x-axis indicate start and stop codons. (D) Stacked column chart
1126 representing the fraction of ribosome profiles containing extreme initiation peaks (defined
1127 as transcripts with an average coverage in first 7 codons exceeding by at least 4 times
1128 the average coverage of the whole CDS). Nucleus-, chloroplast- and mitochondria-
1129 encoded transcripts are plotted separately containing only transcripts with a gene body
1130 coverage of >70% in the footprint dataset (transcripts with extreme initiation peak are 3346
1131 out of 8757 for nuclear, 7 out of 66 for chloroplast and 0 out of 7 for mitochondria).

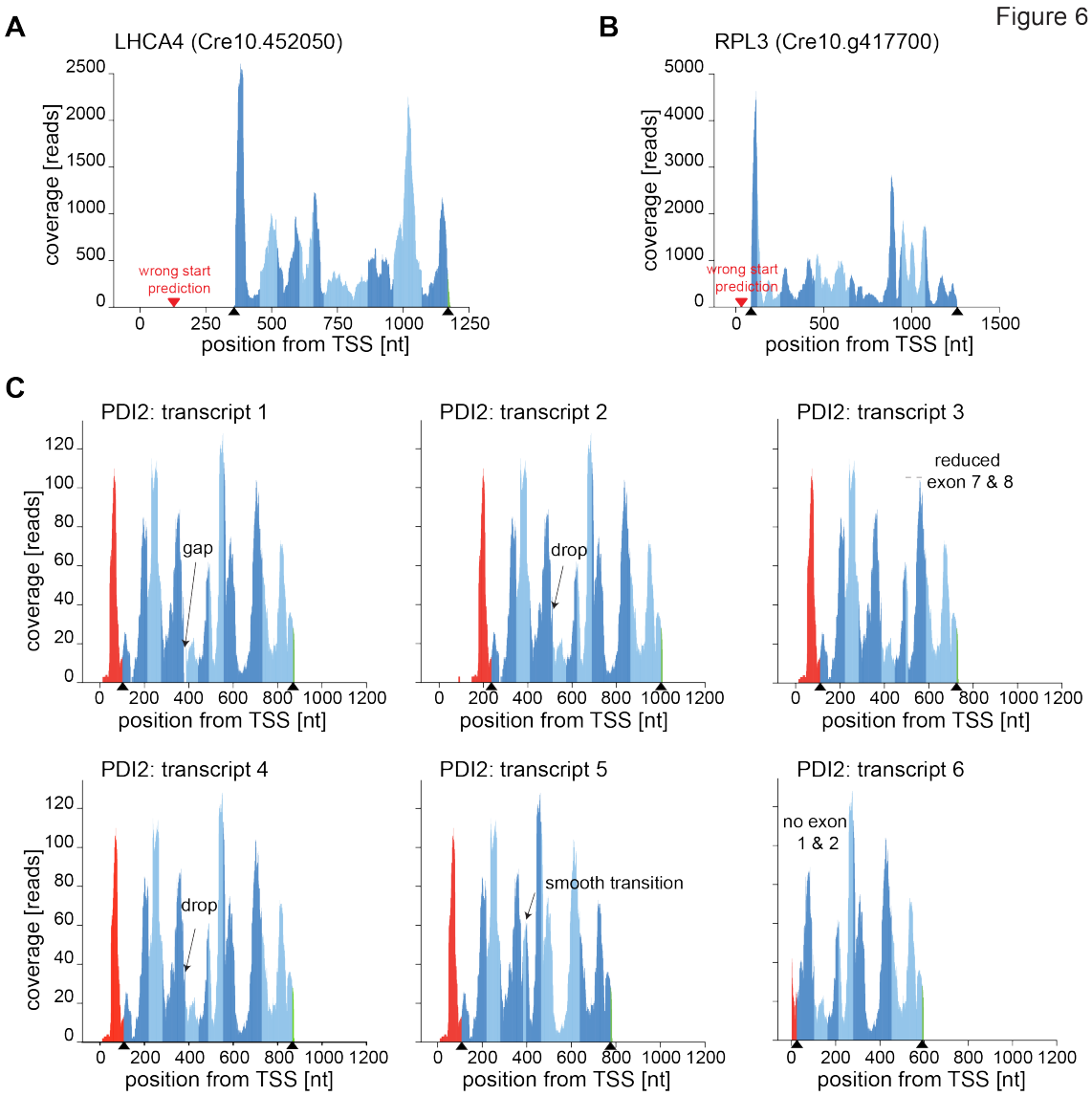
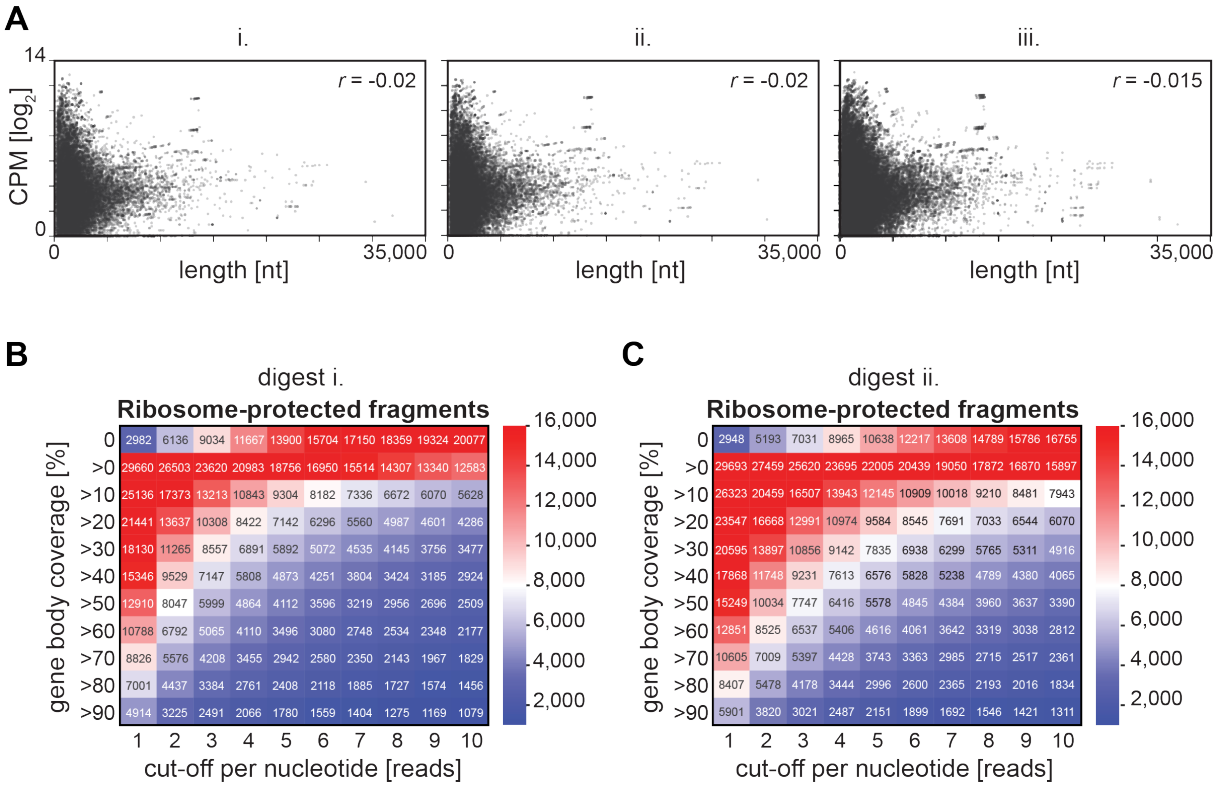



**Figure 6: Utilizing ribosome profiles for coding sequence annotation**
Exemplary ribosome profiles for determining initiation sites and alternatively spliced transcripts. Red parts of the profiles represent coverage within the 5'-UTRs, green parts represent 3'-UTRs and alternating blue shades represent 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.

1145 **Supplemental Figures and Legends:**

1146

1147

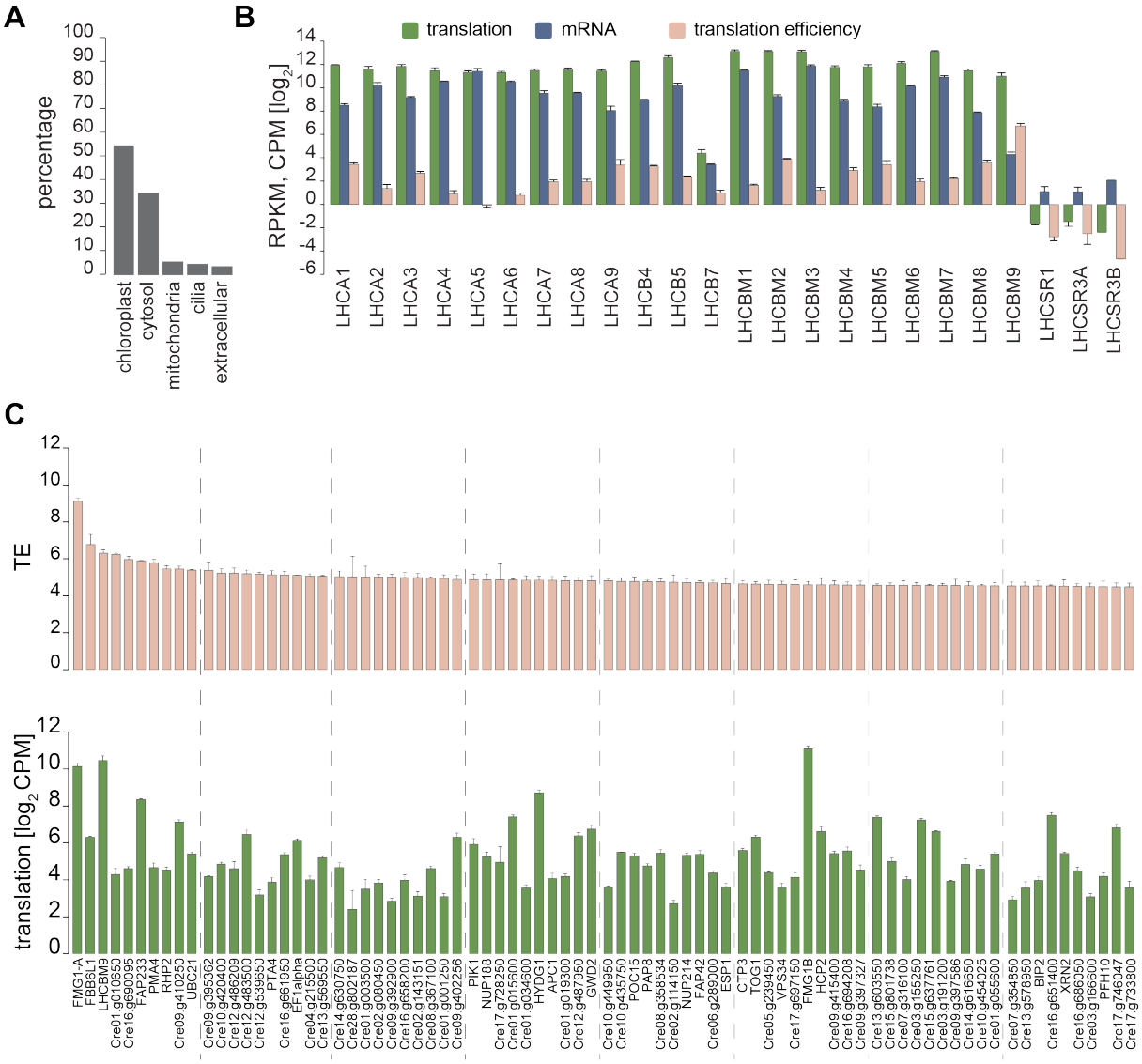



1148

1149

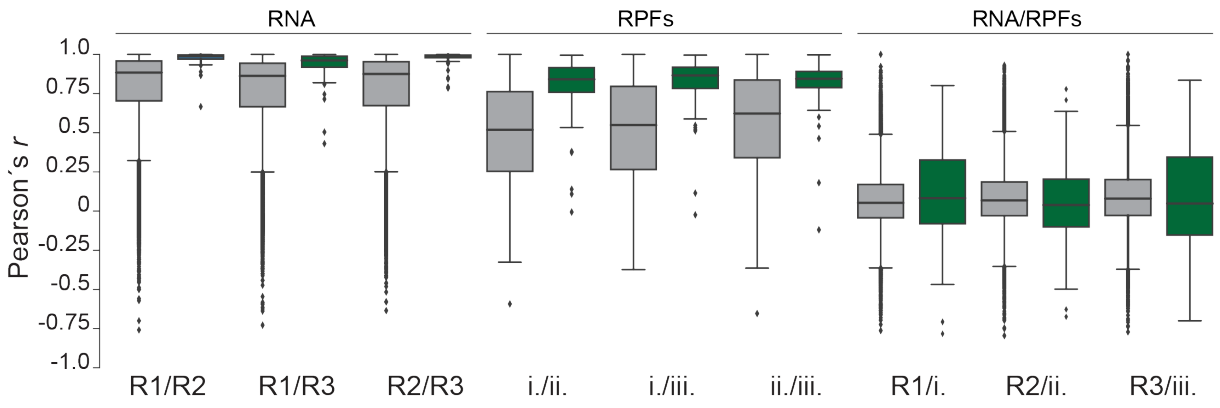
1150 **Supplemental Figure S1: Features of ribosome protected fragment coverage**
1151 Supporting Figure 1. (A) scatter plots representing CPM values of transcripts compared
1152 to their length for all three Ribo-seq data sets. R represents the Pearson correlation
1153 coefficient between both variables for each data set separately. (B) and (C) heatmaps
1154 representing transcript numbers categorized by their extent of gene body coverage in
1155 dependence of specific read count cut-offs for Ribo-seq data sets i. and ii.
1156

1157



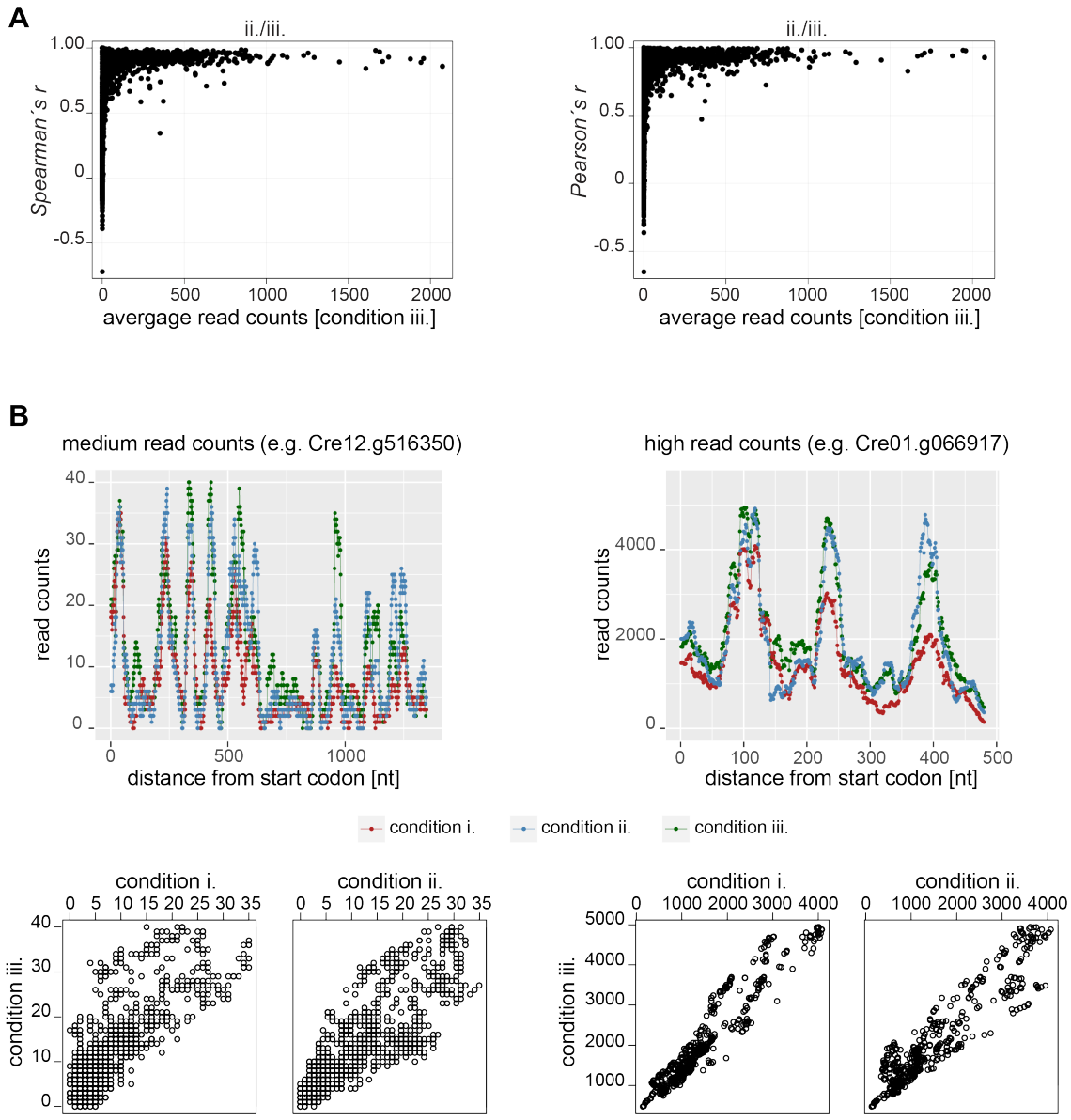

**Supplemental Figure S2: Transcript accumulation, translation output and translation efficiency of nuclear genes**

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 80 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.




**Supplemental Figure S3: Correlation between nuclease digestion conditions, transcript accumulation and translation**

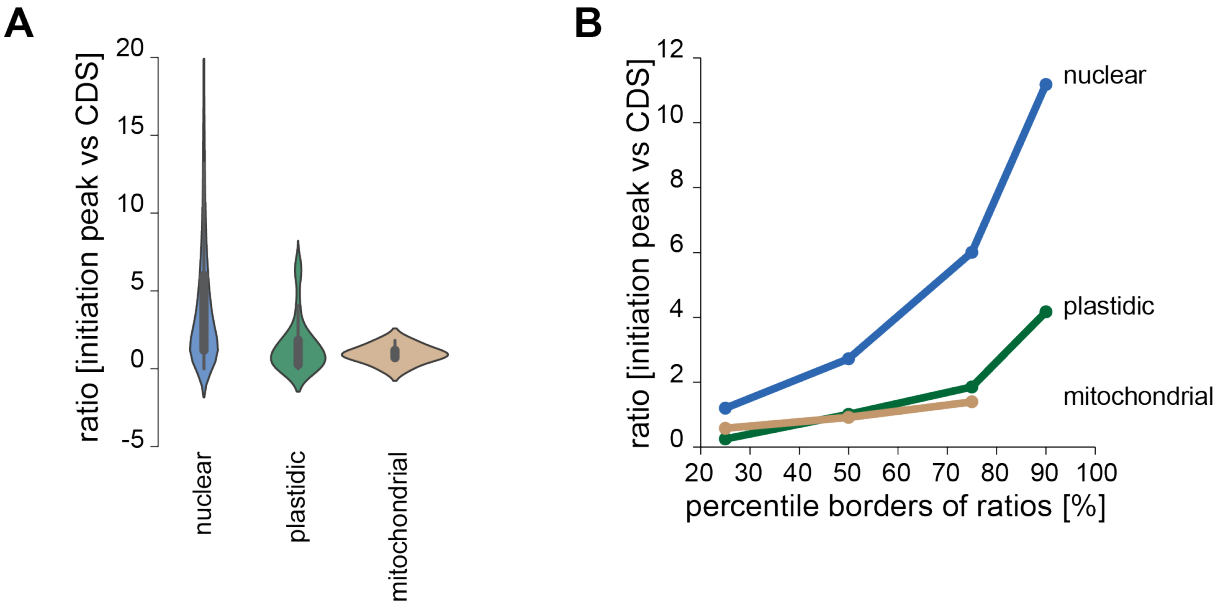
Supporting Figure 5. Boxplots representing the distribution of Pearson correlation coefficients between ribosome profiles and RNA-seq read distribution of the same transcripts each in different data sets. Boxes represent the 0.25 to 0.5 and the 0.5 to 0.75 quartiles, lines represent the median values 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 S4: 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.

1201




1202

1203

1204 **Supplemental Figure S5: Detection of abnormal initiation peaks in the three**
1205 **genomes of *Chlamydomonas reinhardtii***
1206 Supporting Figure 5. (A) Violin plots representing the distribution of ratios between
1207 average per nucleotide coverage within the first seven codons and the average per
1208 nucleotide coverage of the remaining CDS for each organelle's transcripts separately.
1209 Thick lines within the violins represent the interquartile range while thin lines represent the
1210 whiskers with a cut-off of 1.5 times the interquartile range. (B) Scatter plot with lines
1211 representing the increase of the percentile borders of the distributions shown in (A). In
1212 both cases, only transcripts with a gene-body coverage of at least 70% were considered
1213 (8757 nucleus-encoded, 66 chloroplast-encoded, 6 mitochondrial-encoded) to minimize
1214 the risk of misinterpretations due to low or absent coverage in either of both regions.
1215 Calculation of a 90$^{th}$-percentile border for mitochondrial transcripts was not possible due
1216 to the low number of transcripts in consideration.