High-resolution map of plastid encoded polymerase binding patterns demonstrates a major role of transcription in chloroplast gene expression

V. Miguel Palomar[1], Sarah Jaksich[1], Sho Fujii[1, 2], Jan Kuciński[1] and Andrzej T. Wierzbicki[1, *]

[1]Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI 48109, USA

[2]Department of Botany, Graduate School of Science, Kyoto University, Kyoto, 606-8502, Japan

*corresponding author wierzbic@umich.edu

Short title: Binding pattern of plastid encoded RNA polymerase

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plcell/pages/General-Instructions) is: Andrzej T Wierzbicki (wierzbic@umich.edu).

ABSTRACT

Plastids are endosymbiotic organelles containing their own genomes, which are transcribed by two types of RNA polymerases. One of those enzymes is a bacterial-type, multi-subunit polymerase encoded by the plastid genome. The plastid encoded RNA polymerase (PEP) is required for efficient expression of genes encoding proteins involved in photosynthesis. Despite the importance of PEP, its DNA binding locations have not been studied on the genome-wide scale at high resolution. We established a highly specific approach to detect the genome-wide pattern of PEP binding to chloroplasts DNA using ptChIP-seq. We found that in mature *Arabidopsis thaliana* chloroplasts, PEP has a complex DNA binding pattern with preferential association at genes encoding rRNA, tRNA, and a subset of photosynthetic proteins. Sigma factors SIG2 and SIG6 strongly impact PEP binding to a subset of tRNA genes and have more moderate effects on PEP binding throughout the rest of the genome. PEP binding is commonly enriched on gene promoters, around transcription start sites. Finally, the levels of PEP binding to DNA are correlated with the levels of RNA accumulation, which allowed estimating the quantitative contribution of transcription to RNA accumulation.

INTRODUCTION

Plastids are endosymbiotic organelles, which contain their own genomes derived from a cyanobacterial ancestor. Plastid genomes are relatively small, containing between 120 and 160 kb of DNA and encoding typically between 100 and 120 genes (Bock, 2007). The Arabidopsis plastid genome encodes 120 genes in 154,478 bp of DNA (Sato et al., 1999). Most plastid encoded proteins and non-coding RNAs are components of gene expression machinery or photosynthetic enzyme complexes (Bock, 2007). The remainder of the complex plastid proteome is encoded by the nuclear genome and transported into plastids post-translationally (Bock, 2007).

Plastid genomes are transcribed by two types of RNA polymerases. The nuclear-encoded RNA polymerase (NEP) is a phage-type, single-subunit enzyme. NEP transcribes mostly housekeeping genes and is most active early in chloroplast development (Ortelt and Link, 2021; Pfannschmidt et al., 2015). The plastid encoded RNA polymerase (PEP) is a bacterial-type enzyme with four core subunits (α, β, β', and β'') encoded by the plastid genome (*rpoA*, *rpoB*, *rpoC1*, and *rpoC2*, respectively). It transcribes mostly genes encoding photosynthetic proteins,

such as subunits for photosystems and the Rubisco large subunit (RbcL), and is the predominant RNA polymerase in mature chloroplasts (Ortelt and Link, 2021; Pfannschmidt et al., 2015).

Similar to bacterial RNA polymerase, nuclear-encoded sigma factors (SIG) are required for PEP activity by recruiting PEP to gene promoters (Chi et al., 2015; Lysenko, 2007). Six SIG isoforms have been identified in Arabidopsis. Although they have partially redundant functions, loss of SIG2 and SIG6, but not other sigma factors, broadly decreased the mRNA levels of PEP-transcribed genes and impaired chloroplast development in seedlings, indicating the importance of these two sigma factors in chloroplast biogenesis (Woodson et al., 2013). The major targets of SIG2 and SIG6 are considered to be tRNA coding genes and photosynthesis-related genes, respectively (Ishizaki et al., 2005; Kanamaru et al., 2001). A group of peripheral PEP components, pTAC or PAP proteins, is also important for PEP activity (Pfalz and Pfannschmidt, 2013; Pfannschmidt et al., 2015).

Plastid transcription has been studied using run-on experiments designed to assay specific genes in spinach (Deng et al., 1987), potato (Valkov et al., 2009), barley (Krupinska and Apel, 1989; Melonek et al., 2010), tobacco (Krause et al., 2000; Legen et al., 2002; Shiina et al., 1998) and Arabidopsis (Isono et al., 1997; Tsunoyama et al., 2004). Chromatin immunoprecipitation (ChIP) is another approach that has been used to estimate the patterns of transcription by determining the DNA binding pattern of an RNA polymerase. It has been performed in tobacco using epitope-tagged RpoA and genome-wide detection of DNA using a microarray. This method achieved an average spatial resolution of 716 bp, which limits obtained insights to the scale of individual genes (Finster et al., 2013). In Arabidopsis, PEP binding to DNA has been assayed on a limited number of specific loci (Ding et al., 2019; Hanaoka et al., 2012; Yagi et al., 2012) and the genome-wide pattern of PEP activity remains unknown.

Existing run-on and DNA binding data demonstrated substantial differences in transcription and PEP association with DNA between various plastid genes (Deng et al., 1987; Finster et al., 2013). The impact of PEP activity on changes in gene expression in response to developmental or environmental cues is variable with evidence for gene regulation occurring with or without changes in transcription rates (Isono et al., 1997; Krupinska and Apel, 1989; Shiina et al., 1998). Additionally, gene expression in plastids is strongly influenced by posttranscriptional processes including RNA processing and translation (Barkan, 2011; Stern et al., 2010). The impact of transcription on plastid gene regulation remains only partially

understood because existing data inform about PEP transcription on limited numbers of loci or with low resolution. Therefore, the pattern of PEP activity within individual genes or operons is known on just a few loci. Moreover, the relationship between transcription and RNA accumulation is unknown on the genome-wide scale. It is also not known how individual sigma factors contribute to recruiting PEP to specific genes.

We established an improved method to study protein-DNA interactions in plastids and applied it to determine the genome-wide pattern of PEP binding to DNA. We confirmed that PEP has a complex pattern of DNA binding and found that PEP binding is the strongest at rRNA and tRNA genes. Sigma factors SIG2 and SIG6 have dual impacts on PEP binding to specific tRNA genes and to the remainder of the genome. PEP associates with a substantial subset of gene promoters and the levels of PEP binding are correlated with steady-state levels of RNA accumulation. Presented data are available through a publicly available Plastid Genome Visualization Tool (Plavisto) at http://plavisto.mcdb.lsa.umich.edu.

RESULTS

Genome-wide detection of PEP binding to chloroplast DNA

To detect interactions between PEP and specific sequences within the plastid genome, we adapted a nuclear ChIP-seq protocol (Rowley et al., 2013) for use with chloroplasts. We refer to this method as plastid ChIP-seq (ptChIP-seq). A critical step of ChIP is crosslinking with formaldehyde, which covalently preserves protein-DNA interactions (Hoffman et al., 2015). The ptChIP-seq protocol was designed to maximize capture of protein-DNA interactions and, unlike most applications in the nuclear genome, uses 4% formaldehyde (Davis et al., 2011; Zaidi et al., 2017). To demonstrate the specificity of ptChIP-seq, we first compared different lengths of time of the crosslinking reaction. For this purpose, we used 14 days old plants expressing HA-tagged pTAC12/HEMERA (Galvão et al., 2012). pTAC12 is one of PEP-associated factors (Pfalz et al., 2006), which binds at least a subset of PEP-transcribed loci (Pfalz et al., 2015). Crosslinking for 4 hours resulted in the highest signal to noise levels, compared to 1h and 16h (Fig. 1A and Fig. S1A). This was especially visible on tRNA and rRNA genes, where ptChIP-seq signal was the strongest (Fig. 1A and Fig. S1A). We obtained similar results performing ptChIP-seq using chloroplasts enriched from Col-0 wild-type plants and a commercially available polyclonal antibody against the β subunit of PEP (RpoB; Fig. 1B, Fig. S1B). No signal was observed in non-crosslinked controls (Fig. 1A and Fig. 1B), which indicates that unlike other related protocols (Barkan, 2009; Newell and Gray, 2010), ptChIP-seq only captures protein-nucleic acid interactions that have been preserved by crosslinking.

ptChIP-seq with αRpoB antibody is specific

Reliance of ptChIP-seq signal on formaldehyde crosslinking (Fig. 1AB) offers one line of evidence that this method is specific. Additionally, specificity of ptChIP-seq is supported by the lack of signal enrichment in controls without an antibody (Fig. 1B, Fig. S1B). To further test ptChIP specificity, we compared ptChIP-seq using αRpoB antibody in Col-0 wild-type to ptChIP-seq using αHA antibody in plants expressing pTAC12-HA (Galvão et al., 2012). Obtained ptChIP-seq enrichments were highly and significantly correlated between the two experiments when analyzed on annotated genes (Fig. 2A) or bins distributed throughout the

entire plastid genome (Fig. S2). This indicates that pTAC12 and RpoB bind the same loci, as expected, and further supports high specificity of ptChIP-seq.

A critical element of ChIP is a proper negative control. A genotype not expressing the epitope captures most sources of non-specific signal. For αHA ptChIP-seq in plants expressing pTAC12-HA, Col-0 wild-type serves as a proper negative control. Such a control is however much more difficult to obtain for αRpoB ptChIP-seq as the *rpoB* mutant is non-autotrophic (Allison et al., 1996). Because of that, αRpoB ptChIP-seq signal may instead be compared to input samples. The strong correlation between RpoB and pTAC12 ptChIP-seq experiments (Fig. 2A, Fig. S2A) indicates that input serves as a good negative control and that the αRpoB antibody may be used for ptChIP-seq.

To further test if αRpoB ptChIP-seq is specific, we investigated αRpoB ptChIP-seq signal on a NEP-transcribed negative control locus *ftsHi*/*ycf2* (Swiatecka-Hagenbruch et al., 2007). No enrichment was observed on *ftsHi* in Col-0 wild type (Fig. 2B), which indicates that ptChIP-seq is specific. Then we examined if PEP binding to DNA is affected in a mutant defective in SIG2, a sigma factor known to affect specific genes in Arabidopsis seedlings (Chi et al., 2015; Lerbs-Mache, 2011). We performed αRpoB ptChIP-seq in 4-day-old seedlings of Col-0 wild-type and *sig2* mutant. Because plastids are difficult to isolate from seedlings at this growth stage, we applied 4% formaldehyde for 4 h to the intact seedlings to capture protein-DNA interaction. RpoB enrichment on DNA in 4-day-old wild type seedlings was well correlated with that observed in 14-day-old seedlings (Fig. S2B), indicating that ptChIP can be applied to different developmental stages. We further analyzed the mean ptChIP-seq enrichment on a subset of loci that have previously been assayed for changes in RNA accumulation in *sig2* (Hanaoka et al., 2003; Kanamaru et al., 2001; Nagashima et al., 2004; Privat et al., 2003). ptChIP-seq enrichment was strongly reduced on *trnE*, *trnY*, *trnD*, and *trnV* (Fig. 2B), which is consistent with reported substantial reduction of steady state levels of those tRNAs in *sig2* (Hanaoka et al., 2003; Kanamaru et al., 2001; Nagashima et al., 2004; Privat et al., 2003). ptChIP-seq enrichment on *trnW* was reduced to 0.53 of Col-0, which is consistent with a small effect of *sig2* on the accumulation of its tRNA product (Kanamaru et al., 2001). ptChIP-seq enrichment on *psbA* was reduced to 0.44 of Col-0, consistently with a relatively small impact of *sig2* on the accumulation of its mRNA (Kanamaru et al., 2001). ptChIP-seq enrichment was also reduced on *psaJ*, which is consistent with prior RNA accumulation data (Nagashima et al.,

2004). This indicates that αRpoB ptChIP results in *sig2* are generally consistent with prior RNA accumulation studies. Together, these results indicate that ptChIP-seq with αRpoB antibody is highly specific and may be used to assay the interactions of PEP with DNA.

Complex pattern of PEP binding to DNA

Analysis of PEP binding across the plastid genome revealed a complex pattern of occupancy with preferential binding to genes encoding rRNA, tRNA, and some protein-coding genes (Fig. 3A). Most PEP binding is present within the inverted repeats (IR) where rRNA genes are located and in the large single copy region (LSC). The small single copy region (SSC) had little PEP binding (Fig. 3A). These observations are consistent with prior ChIP-chip study in tobacco (Finster et al., 2013) and assays on a limited number of loci in Arabidopsis (Yagi et al., 2012). Interestingly, we observed over a 20-fold dynamic range of ptChIP-seq signals between regions with various levels of PEP binding (Fig. 3A). This is also consistent with prior reports of PEP binding (Finster et al., 2013; Yagi et al., 2012) and transcription (Deng et al., 1987) and confirms that the levels of transcription may be greatly variable between individual genes.

We next tested if various levels of PEP binding are equally likely throughout the plastid genome. Distribution of ptChIP enrichment levels was multimodal with four peaks corresponding to no detectable PEP binding and three levels of PEP presence on the genome (Fig. 3B). Thirty four percent of the genome had no PEP binding, 31% had a low level of PEP, 19% had a medium level of PEP, and 17% had a high level of PEP (Fig. 3B). Focusing on annotated genes only, no PEP was detected on the *rpoB* operon, a subset of genes encoding ribosomal proteins, and a subset of genes encoding NDH subunits (Fig. 3C, Fig. S3A). Low and medium levels of PEP were detected on most genes encoding photosynthetic proteins and a subset of tRNA genes (Fig. 3C, Fig. S3A). High levels of PEP were found on the remaining tRNA genes, rRNA genes, and three genes encoding photosystem II subunits, including *psbA* (Fig. 3C, Fig. S3A). The multimodal distribution of PEP binding to DNA may be speculatively interpreted as an indication that the transcriptional machinery may adopt four distinct functional states corresponding to the four preferred levels of PEP binding.

Together, these results show a complex pattern of PEP binding to the plastid genome and suggest the presence of at least four preferred functional states of plastid transcriptional machinery.

198

199 **Dual impact of *sig2* and *sig6* mutants on PEP binding**

200 Out of six sigma factors in Arabidopsis, only SIG2 and SIG6 are essential for proper chloroplast
201 development (Chi et al., 2015; Lerbs-Mache, 2011). To determine the impact of those two sigma
202 factors on PEP binding to DNA, we analyzed RpoB ptChIP-seq in *sig2* and *sig6* across the entire
203 genome. The pattern of PEP binding was disrupted throughout the genome with most genes
204 showing partial reduction in both mutants (Fig. 4A). The *sig6* mutant had a much stronger effect
205 than *sig2* mutant on most PEP-bound loci (Fig. 4A)

206 Regression analysis of *sig2* compared to Col-0 wild type revealed that ptChIP-seq signals
207 in Col-0 wild type and *sig2* are significantly correlated with a slope of 0.75 (Fig. 4B). This
208 indicates that in *sig2*, most genes have a consistent, moderate reduction of PEP binding (Fig.
209 4B). Only a few genes had PEP binding reduced to much greater extents than indicated by the
210 genome-wide trend. These included four previously studied tRNA genes (Fig. 2B, Fig. 4AD),
211 *trnL2* (Fig. 4B, Fig. S4A), and one protein-coding gene *ndhC* (Fig. 4B, Fig. S4A). However, PEP
212 binding was unchanged on *rbcL* (Fig. 4B, Fig. S4A).

213 A similar pattern was observed in *sig6* where PEP binding was strongly reduced
214 throughout the genome. PEP binding in *sig6* and Col-0 wild type remained significantly
215 correlated, but the slope of the regression line was reduced to 0.47 (Fig. 4C). Only one gene had
216 an almost complete loss of PEP binding, the tRNA gene, *trnI*, located in the inverted repeat (Fig.
217 4CE). Consistently, accumulation of the *trnI* tRNA product was significantly reduced in *sig6*
218 (Fig. S4B).

219 Decreased PEP binding in *sig2* and *sig6* mutants may be due to lower abundance of PEP.
220 To test this possibility, we assessed the protein accumulation of a PEP subunit, RpoC1 (Fig.
221 S4C). The accumulation levels of this subunit were slightly lower in both *sig2* and *sig6*
222 compared to the Col-0 wild type (Fig. S4C), which indicates a small reduction of PEP
223 abundance. These results suggest that SIG2 and SIG6 have specific impacts on limited numbers
224 of genes together with weaker but broad impacts on PEP occupancy throughout the genome.

225

226 **PEP preferentially binds to gene promoters**

227 Previous studies identified preferential binding of PEP to two promoters of photosystem II genes
228 (Ding et al., 2019; Yagi et al., 2012). To determine if promoter binding is a more general

property of PEP, we performed RNA-seq designed to identify triphosphorylated 5' ends of primary transcripts (Zhelyazkova et al., 2012), or TSS RNA-seq, and compared it with the ptChIP-seq signal on PEP-transcribed genes with known promoter locations. Consistent with prior findings (Ding et al., 2019; Yagi et al., 2012), PEP binding was strongly enriched on annotated *psbA* and *psbEFLJ* promoters (Fig. 5A). These promoters also had strong TSS RNA-seq signals (Fig. 5A), which confirms that peaks of PEP binding coincide with transcription start sites. We observed similar preferential PEP binding to other known promoters including *psaA* and *rbcL* (Fig. 5B, Fig. S5A). When averaged over all promoters previously identified in Arabidopsis (Fig. S5AB), PEP binding and TSS RNA-seq signals were strongly enriched on gene promoters (Fig. 5C). These results indicate that preferential binding to gene promoters may be a general property of PEP.

PEP binding is correlated with steady state levels of RNA

Posttranscriptional regulation has a major impact on plastid gene expression (Barkan, 2011). Therefore, the levels of PEP binding to DNA may have a limited correlation with steady state levels of RNA. To test this prediction, we split the genome into 250nt bins and counted average TSS RNA-seq read counts and ptChIP-seq enrichments from three replicates of both experiments. Inverted Repeat regions were not included because highly structured regions within mature rRNAs may inhibit Terminator exonuclease. PEP binding to DNA and steady state levels of primary transcripts were significantly correlated (Fig. 6A). This suggests that differences in steady state levels of RNA may to some extent be explained by differences in PEP transcription.

Transcription start sites are located outside of annotated genes. Consistently, there are many genomic bins which have high levels of PEP binding to DNA and low levels of TSS RNA-seq (Fig. 6A). To overcome this limitation of TSS RNA-seq, we used a previously published RNA-seq dataset from 14-day-old plants (Thieffry et al., 2020) and compared the steady state level of total RNA to the level of PEP binding to DNA on annotated genes. RNA-seq levels on rRNA and tRNA genes were not correlated with PEP binding (Fig. S6A), probably due to ribodepletion of RNA samples prior to the library prep and highly structured tRNAs being poor substrates for library production. However, annotated protein-coding genes had a significant correlation between RNA-seq and ptChIP-seq signals (Fig. 6B). R^2 of 0.3 indicates that about a third of RNA-seq variance may be predicted by the level of PEP binding to DNA. The reminder

of RNA-seq variance is likely affected by RNA processing and degradation, and a subset of PEP-transcribed mRNAs, such as *psbA*, *rbcL*, and *petB*, appears to be particularly stable (Fig. 6B). Together, these results indicate that PEP binding to DNA, which may be interpreted as a proxy for the level of transcription, has a significant impact on the steady state levels of RNA and possibly, more generally, on gene expression.

DISCUSSION

Detection of protein-DNA interactions in plastids

Several lines of evidence support the specificity of ptChIP-seq in detecting PEP binding to plastid DNA. A key feature of this protocol is efficient formaldehyde crosslinking of enriched chloroplasts combined with stringent immunoprecipitation conditions. Together, they allow detection of only protein-DNA interactions that have been preserved by crosslinking. This is in stark difference to some other protein-nucleic acid interaction studies in plastids (Barkan, 2009; Newell and Gray, 2010). While a long formaldehyde treatment may increase the risk of crosslinking artifacts (Walker et al., 2020), our data clearly demonstrate no RpoB ptChIP-seq signal on loci that are not transcribed by PEP, such as the *rpoB* operon and the coding region of *ftsHi*/*ycf2* (Hajdukiewicz et al., 1997; Swiatecka-Hagenbruch et al., 2007). This indicates that signal preserved by crosslinking is very likely to be specific.

It is also important to note that our protocol can capture the pattern of PEP-DNA interactions in both enriched mature chloroplasts and intact seedlings with similar specificity. Transcriptional regulation is particularly important in the early stages of chloroplast differentiation (Pfannschmidt et al., 2015). Considering the difficulty of plastid isolation at early growth stages, our ptChIP-seq protocol for intact seedlings can be a powerful approach to explore the mechanism of transcriptional regulation in developing chloroplasts.

ptChIP-seq with the αRpoB antibody shows a high dynamic range of PEP binding to DNA, which is consistent with prior observations on small subsets of specific loci in Arabidopsis and tobacco (Finster et al., 2013; Yagi et al., 2012). The dynamic range of ptChIP-seq is however an order of magnitude higher than previously reported ChIP-chip study of RpoA in tobacco (Finster et al., 2013). This indicates that ptChIP-seq may substantially expand our understanding of PEP transcription and overall regulation of plastid gene expression.

Complexity of PEP transcription

The pattern of PEP binding to DNA is complex, which illustrates that variable levels of protein production are at least partially caused by variable levels of PEP transcription. Although the involvement of NEP transcription remains unknown, the complex pattern of PEP binding is not

297 consistent with a simplistic view of the model assuming full transcription of the plastid genome
298 (Shi et al., 2016).

299 Intensity of PEP binding to DNA throughout the plastid genome is not distributed
300 normally and instead shows four preferred levels. We speculate that these levels correspond to
301 four preferred functional states of PEP. These four states could be caused by specific cis-acting
302 elements and regulatory proteins bound to those elements. They could also be reflected by
303 preferred interactions between the core PEP complex and accessory proteins, or even could be an
304 indication that the genome exists in four favored structural states, which would be reminiscent of
305 the situation in the nucleus (Roudier et al., 2011).

306 A strong preferential binding of PEP within gene promoters, around transcription start
307 sites, is consistent with the concept of RNA polymerase pausing on gene promoters (Landick,
308 2006). While it has been previously shown on the *psbEFLJ* operon (Ding et al., 2019), our
309 results indicate that pausing may be a general property of PEP. The role of PEP pausing in
310 transcription or gene regulation remains unresolved.

311

312 **Role of SIG2 and SIG6 in PEP recruitment**

313 PEP is recruited to its promoters by binding of sigma factors in a sequence specific manner.
314 Among six nuclear-encoded sigma factors in Arabidopsis, SIG1, SIG3, SIG4 and SIG5 have
315 highly specific functions and are not required for proper chloroplast development. In contrast,
316 SIG2 and SIG6 have more general impacts on plastid transcription and are required for early
317 chloroplast development (Börner et al., 2015; Chi et al., 2015; Puthiyaveetil et al., 2021).

318 Strong impacts of *sig2* and *sig6* on a limited number of tRNA genes accompanied by a
319 genome-wide moderate reduction of PEP binding could be explained by complex patterns of
320 SIG2 and SIG6 binding specificities. In this scenario, strong non-redundant impacts on a few
321 targets would be accompanied by weaker non-redundant but still specific roles on all remaining
322 PEP-transcribed genes. We, however, propose a simpler explanation where SIG2 and SIG6 have
323 a dual impact on PEP binding by a combination of direct and indirect mechanisms. In this model,
324 both SIG2 and SIG6 directly and non-redundantly only impact limited numbers of mostly tRNA
325 genes. Then, tRNA deficiencies negatively impact plastid translation and to some extent, PEP
326 production, which leads to consistent reductions of PEP occupancy throughout the genome. This
327 model is consistent with the observation that *sig2* showed a stronger decrease of tRNA compared

to mRNA for photosynthesis-associated genes (Kanamaru et al., 2001) and also explains why *sig2* and *sig6* mutants recover from early developmental defects and produce fully functional chloroplasts later in development (Ishizaki et al., 2005; Privat et al., 2003).

All tRNAs strongly affected in *sig2* or *sig6* mutants lose PEP binding not only throughout their transcribed sequences but also on their promoters. This result suggests that pausing is tightly coupled with transcription elongation in plastids.

Contribution of transcription to gene regulation in plastids

Significant correlation between PEP binding to DNA determined by ptChIP-seq and steady state levels of RNA determined by RNA-seq allows estimation of the contribution of PEP binding to RNA accumulation. We estimate that about 30% of RNA accumulation can be explained by PEP binding. PEP binding may be interpreted as a proxy for transcription rates, although it should be noted that PEP elongation rates remain unknown and may be variable between various genes. Several prior run on and ChIP studies also suggest that transcription has a significant role in gene regulation (Deng et al., 1987; Finster et al., 2013). It is also supported by the observation that reduction of PEP recruitment in sigma factor mutants is comparable to previously reported reductions of RNA levels. However, this is to our knowledge the first study providing a genome-wide quantitative estimate of the relationship between transcription and RNA levels.

Our observations about the contribution of transcription to gene regulation apply to the variability between genes. Changes in transcription between developmental or environmental conditions have mixed impacts on gene expression (Isono et al., 1997; Krupinska and Apel, 1989; Shiina et al., 1998). The presented approach may be used in future studies to uncover the contribution of PEP to condition-dependent plastid gene regulation.

METHODS

Plant materials and growth conditions

Arabidopsis thaliana wild-type Columbia-0 (Col-0) ecotype was used in all analyses. We used the following genotypes: *sig2-2* (SALK_045706) (Woodson et al., 2012), *sig6-1* (SAIL_893_C09) (Ishizaki et al., 2005) and pTAC12-HA (*HMR::HA*/*hmr-5*) transgenic line (Galvão et al., 2012). For experiments with 14-day-old plants, seeds were stratified in darkness at 4ºC for 48 hours and grown on soil at 22ºC under white LED light (100 µmol m^{-2} s^{-1}) in 16h/8h day/night cycle. For experiments with 4-day-old plants, seeds were stratified in darkness at 4ºC for 48 hours and grown on 0.5 X MS plates (0.215% MS salts, 0.05% MES-KOH pH 5.7, 0.65% Agar) for four days at 22ºC under constant white LED light (50 µmol m^{-2} s^{-1}).

Chloroplast enrichment and crosslinking

Chloroplasts from 14-day-old seedlings were enriched following the protocol described by Nakatani and Barber with minor modifications (Nakatani and Barber, 1977). In brief, 5 grams of rosette leaves were harvested and rinsed 3 times with ultra-pure water to eliminate soil debris and homogenized in chloroplast enrichment buffer (0.33 M Sorbitol, 30 mM HEPES-KOH (pH 7.5), 0.001% β-mercaptoethanol) using a blender. The homogenate was filtered through two layers of Miracloth, and the flow-through was centrifuged at 1500 g for 5 minutes at 4ºC. The pelleted chloroplasts were resuspended in 1 ml of chloroplast enrichment buffer. Chlorophyll concentration was determined by resuspending 10 µl of the chloroplast fraction in 1 ml 80% acetone and measuring its absorbance at 652 nm as reported (Inskeep and Bloom, 1985). To cross-link DNA to proteins, 4% final concentration of formaldehyde was applied to the amount of chloroplast corresponding to 200 µg of chlorophyll followed by incubation at 4ºC for 4h unless indicated otherwise. Formaldehyde was quenched by diluting the chloroplasts 5 times in the chloroplast enrichment buffer containing 125 mM glycine, followed by chloroplast pelleting at 1500 g at 4ºC. For experiments using 4-day-old seedlings, whole seedlings were vacuum infiltrated with 4% formaldehyde for 10 min as reported previously (Rowley et al., 2013) and incubated for 4h at 4ºC.

ptChIP-seq

ptChIP-seq protocol was based on a previously published nuclear ChIP protocol (Rowley et al., 2013). In brief, enriched chloroplasts from 14-day-old plants corresponding to 50 µg chlorophyll were subject to *in vitro* crosslinking. Alternatively, 50 4-day-old seedlings were crosslinked *in vivo*, flash-frozen, homogenized in lysis buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS) and filtered through two layers of Miracloth. Obtained samples were sonicated to achieve DNA fragments ranging from 200 nt to 300 nt using a QSonica Q700 sonicator. The fragmented samples were incubated overnight with 1 µg of monoclonal anti-HA antibody (Invitrogen catalog number 26183) or 5 µg of polyclonal anti-RpoB antibody (PhytoAB catalog number Phy1239) with 40 µl Protein G Dynabeads (Invitrogen catalog number 10004D) or 60 µl Protein A Dynabeads (Invitrogen catalog number 10002D) respectively. After incubation, the beads were washed, and DNA was eluted, and reverse cross-linked as described (Rowley et al., 2013). High throughput sequencing libraries were prepared as reported (Bowman et al., 2013) and sequenced using an Illumina NovaSeq 6000 S4 flow-cell with 150x150 paired-end configuration at the University of Michigan Advanced Genomics Core.

TSS RNA-seq

Five micrograms of RNA isolated from enriched chloroplasts was digested with 6U of Terminator exonuclease (Lucigen catalog number TER51020) for 1h at 30ºC in 20µl reaction volume. The reaction was stopped by addition of 1µl of 100mM EDTA. Subsequently, digested RNA was purified with acidic buffer-saturated phenol, washed with 70% ethanol, and resuspended in 10µl of water. Purified RNA (~10ng) was submitted to the University of Michigan Biomedical Research Core where libraries were generated with SMARTer Stranded Total RNA-Seq Kit-Pico Input Mammalian and sequenced. Modifications to the original protocol include: no RNA fragmentation, no rRNA depletion and inclusion of size selection (~200bp fragments) of the libraries before sequencing.

Data analysis

The obtained raw sequencing reads were trimmed using trim_galore v.0.4.1, and mapped to the TAIR10 Arabidopsis plastid genome (www.arabidopsis.org) using Bowtie2 v.2.2.8 (Langmead and Salzberg, 2012). For ptChIP-seq reads were PCR de-duplicated using PICARD tools. Read counts on defined genomic regions were determined using bedtools v.2.25.0 (Quinlan and Hall,

2010). ptChIP-seq signals on annotated genes were calculated by dividing RPM normalized read counts from αHA or αRpoB ptChIP-seq by RPM normalized read counts from αHA ptChIP-seq in Col-0 wild type or input samples, respectively. ptChIP-seq enrichments on annotated genes were calculated by dividing signal levels on individual genes by the median signal level on genes in the *rpoB* operon, which is not transcribed by PEP and represents background signal levels. ptChIP-seq enrichments on genomic bins were calculated by dividing signal levels on individual bins by the signal level on the entire *rpoB* operon.

Identification of promoters

PEP promoters in the plastid genome were identified based on previous studies representing promoter sequences in Arabidopsis (Favory et al., 2005; Fey et al., 2005; Hanaoka et al., 2003; Hoffer and Christopher, 1997; Ishizaki et al., 2005; Kanamaru et al., 2001; Liere et al., 1995; Nagashima et al., 2004; Privat et al., 2003; Shimmura et al., 2008; Sriraman et al., 1998; Swiatecka-Hagenbruch et al., 2007; Zghidi et al., 2007). Regions between -35 (TTGACA) and -10 (TATAAT) consensus motives were identified and are shown in Figure S5.

Immunoblot analysis

Total proteins were extracted from 4-day-old seedlings by incubating homogenized samples in the sample buffer (20 mM Tris-HCl (pH 6.8), 3% β-mercaptoethanol, 2.5% sodium dodecyl sulfate, 10% sucrose) with cOmplete protease inhibitor cocktail (Roche) for 1.5 h at room temperature. After removing debris by centrifugation, 20 µg proteins were separated by SDS-PAGE. To detect RpoC1, monoclonal anti-RpoC1 antibody (PhytoAB catalog number PHY1904) and anti-mouse IgG antibody conjugated with horseradish peroxidase (Amersham catalog number NA931) were used as the primary and secondary antibodies, respectively. Protein bands were visualized using chemiluminescence reagents (ECL Prime Western Blotting Detection Reagent, Amersham) and an ImageQuant LAS 4000 imager.

tRNA quantification

Total RNA was isolated from 50 fresh 4 days-old seedlings from wild type Col-0 and *sig6-1* mutant plants using Trizol following the manufacturer protocol. Following DNaseI digestion, 500 and 1000 ng of RNA were used to generate primer specific (for tRNA-Ile-CAU) and polyA

cDNA, respectively, using the SuperScript III First Strand Kit following the instructions for highly secondary structured templates. Real-time PCR was performed using the KAPA Sybr Green 2x kit with the following primers: Ath_Actin2_FWD 5'GAGAGATTCAGATGCCCAGAAGTC3' , Ath_Actin2_REV 5'TGGATTCCAGCAGCTTCCA3', tRNA-Ile-CAU_FWD 5'ATCCATGGCTGAATGGTTAAAGCG3', tRNA-Ile-CAU-REV 5'CATCCAGTAGGAATTGAACCTACGA3'. Results were analyzed using the ddCT method.

Accession numbers

The sequencing data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE192568. Sequencing data presented in this study are available through a dedicated publicly available Plastid Genome Visualization Tool (Plavisto) at http://plavisto.mcdb.lsa.umich.edu.

Acknowledgements

This work was supported by a grant from the National Science Foundation (MCB 1934703) to A.T.W. S.F. was supported by grants from the Japanese Society for the Promotion of Science (19J01779, 20K15819). The pTAC12-HA (*HMR::HA*/*hmr-5*) transgenic line was kindly provided by Meng Chen (University of California, Riverside).

Author contributions

V.M.P., S.J, S.F. and J.K. performed research. V.M.P, S.F. and A.T.W analyzed data. A.T.W. wrote the paper.

Supplemental data

Figure S1. Detection of PEP binding to DNA using ptChIP-seq. Individual biological replicates of data shown in Fig. 1.

Figure S2. Specificity of ptChIP-seq with αRpoB antibody.

Figure S3. Complex pattern of PEP binding to plastid DNA. Individual biological replicates of data shown in Fig. 3C.

Figure S4. Dual impact of *sig2* and *sig6* mutants on PEP binding to plastid DNA.

476 **Figure S5.** PEP promoters identified in Arabidopsis.

477 **Figure S6.** Correlation of PEP binding with steady state levels of RNA.

478 **Table S1.** High throughput sequencing datasets generated in this study.

479 **Table S2.** Annotation of plastid-encoded genes in Arabidopsis used in this study.

480

481

FIGURES

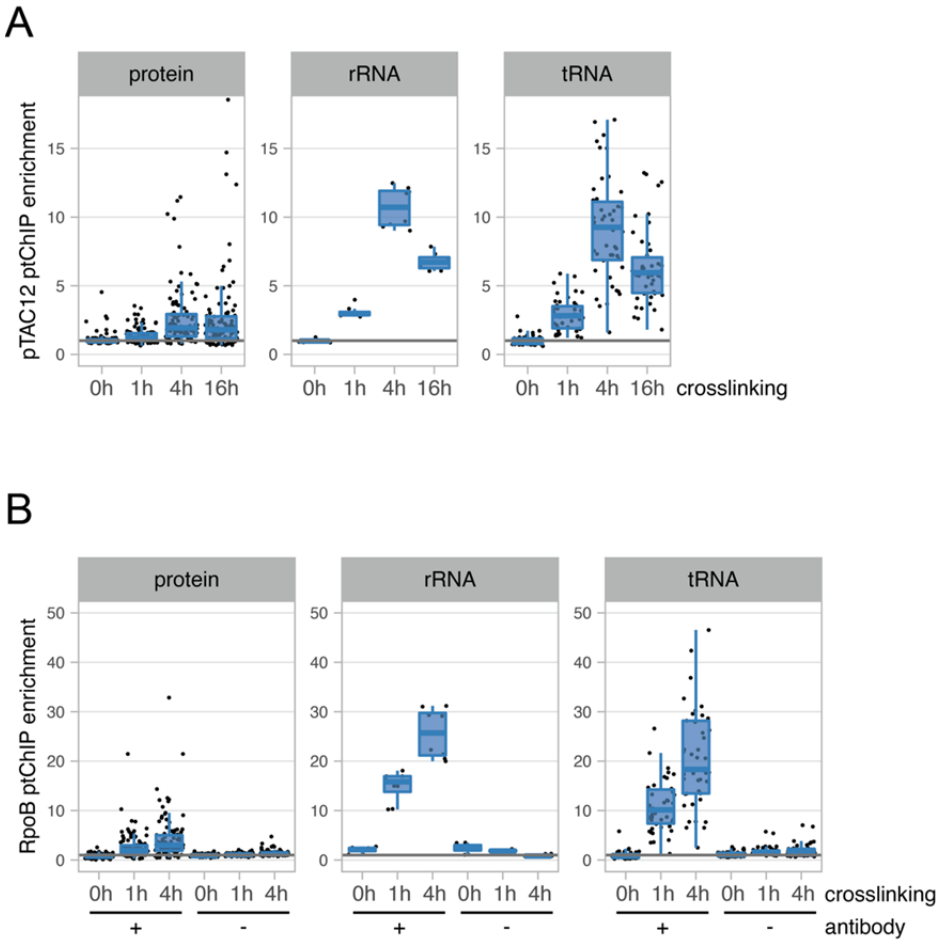


Figure 1. Detection of PEP binding to DNA using ptChIP-seq.

A. Optimization of formaldehyde crosslinking time in ptChIP-seq. ptChIP-seq was performed using αHA antibody in plants expressing pTAC12-HA (Galvão et al., 2012) with no crosslinking or crosslinking of enriched chloroplasts with 4% formaldehyde for 1h, 4h and 16h. ptChIP-seq signals on annotated genes were calculated by dividing RPM normalized read counts from αHA ptChIP-seq in pTAC12-HA by RPM normalized read counts from αHA ptChIP-seq in Col-0 wild type.

B. Optimization of formaldehyde crosslinking time and negative controls in ptChIP-seq. ptChIP-seq was performed using αRpoB antibody in Col-0 wild type plants with no crosslinking or crosslinking of enriched chloroplasts with 4% formaldehyde for 1h and 4h. ptChIP-seq was performed with and without the αRpoB antibody. ptChIP-seq signals on annotated genes were calculated by dividing RPM normalized read counts from

αRpoB ptChIP-seq in Col-0 wild type by RPM normalized read counts from input samples.

In A and B ptChIP-seq enrichments were calculated by dividing signal level on individual genes by the median signal level on genes in the *rpoB* operon, which is not transcribed by PEP and represents background signal levels. Genes were divided by the functions of their products into protein-coding, rRNA genes and tRNA genes. Average enrichments from two or three independent biological replicates are shown. Individual biological replicates are shown in Fig. S1.

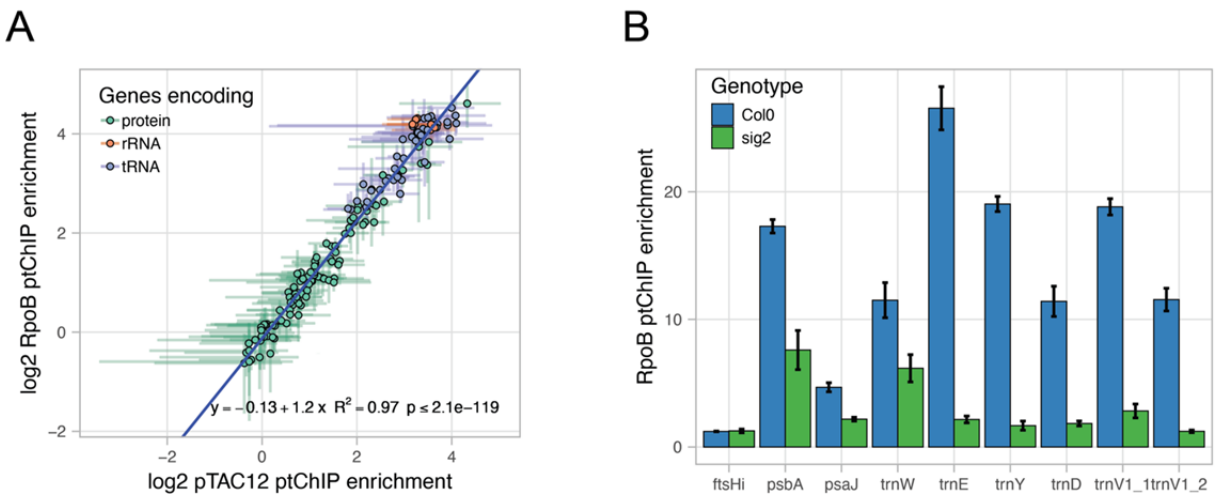


Figure 2. Specificity of ptChIP-seq with αRpoB antibody.

A. pTAC12 and RpoB ptChIP-seq signals are highly correlated. Enrichment levels on annotated genes were compared between ptChIP-seq experiments using αHA antibody in plants expressing pTAC12-HA and using αRpoB antibody in Col-0 wild type plants. Data points are color-coded by function of the corresponding genes and show averages from three independent replicates. Error bars indicate standard deviations. Blue line represents the linear regression model.

B. RpoB ptChIP-seq signal is reduced in *sig2* mutant on genes known to be affected by SIG2. Enrichment levels of ptChIP-seq using αRpoB antibody in Col-0 wild type and *sig2* were calculated on individual genes. Bars show averages from three independent biological replicates. Error bars indicate standard deviations.

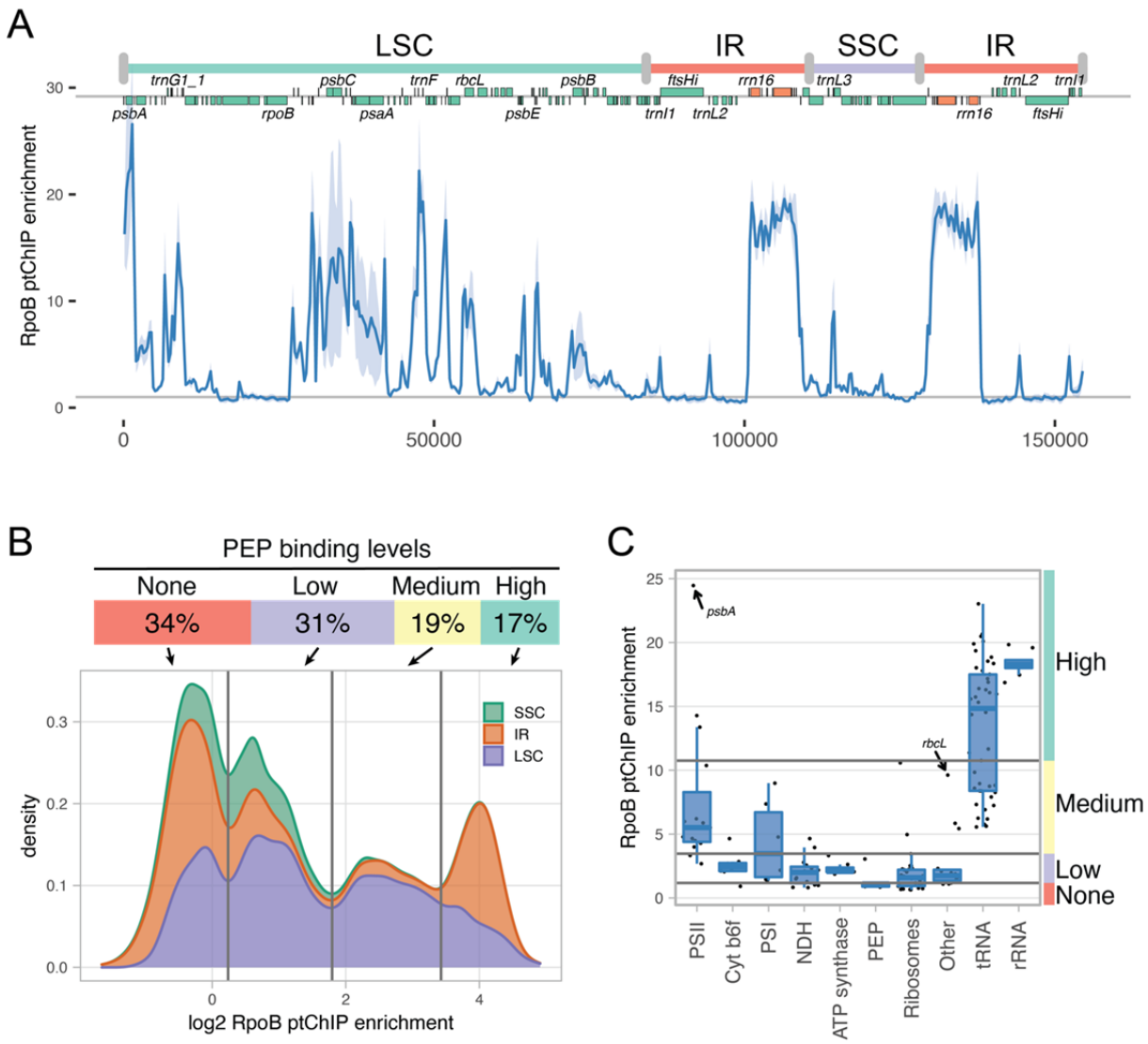


Figure 3. Complex pattern of PEP binding to plastid DNA.

A. Genome-wide pattern of PEP binding to DNA. Signal enrichment from ptChIP-seq using αRpoB antibody in Col-0 wild type plants was calculated in 50 bp genomic bins and plotted throughout the entire plastid genome. Genome annotation including genomic regions, positions of annotated genes and names of selected individual genes are provided on top of the plot. Average enrichments from three independent biological replicates are shown. Light blue ribbon indicates standard deviation.

B. Four preferred levels of PEP binding to DNA. Density plot of signal enrichments of ptChIP-seq using αRpoB antibody in Col-0 wild type plants. Average enrichments in 50 bp genomic bins from three independent biological replicates were analyzed in the SSC,

IR and LSC. PEP binding level groups were determined by positions of local minima on the density plot. Percentages on top indicate fraction of all genomic bins assigned to a particular group.

C. PEP binding to DNA of genes classified by the function of their products. Enrichment levels of ptChIP-seq using αRpoB antibody in Col-0 wild type were plotted on annotated genes split by the functions of gene products (Chotewutmontri and Barkan, 2018). PEP binding level groups are indicated on the right. Data points show averages from three independent biological replicates. Independent replicates are shown in Fig. S3.

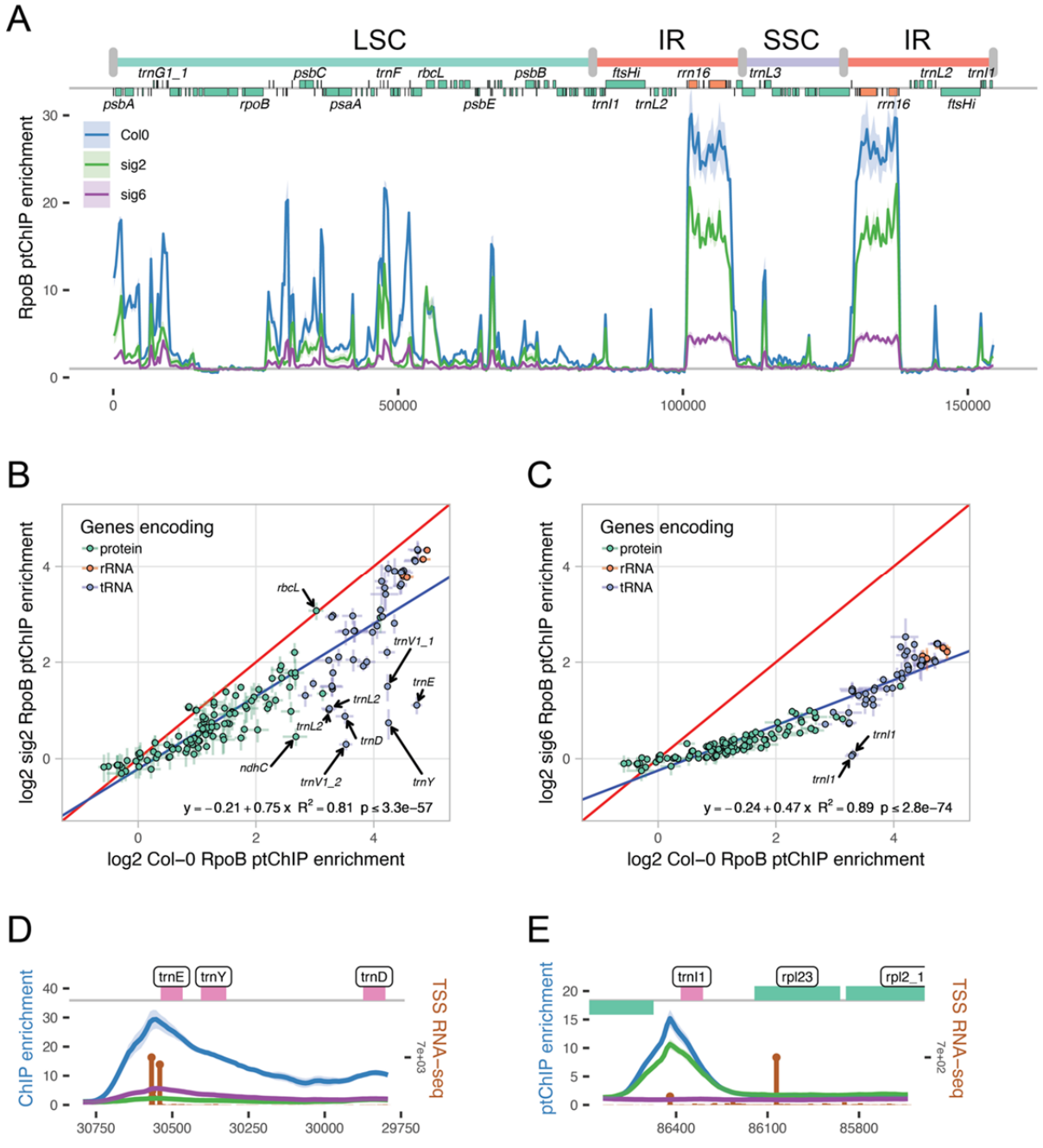


Figure 4. Dual impact of *sig2* and *sig6* mutants on PEP binding to plastid DNA.

A. Genome-wide impact of *sig2* and *sig6* on of PEP binding to DNA. Signal enrichments from ptChIP-seq using αRpoB antibody in Col-0 wild type, *sig2* and *sig6* plants were calculated in 50 bp genomic bins and plotted throughout the entire plastid genome. Genome annotation including genomic regions, positions of annotated genes, and names of selected individual genes are provided on top of the plot. Average enrichments from three independent biological replicates are shown. Ribbons indicates standard deviations.

B. Dual impact of *sig2* on PEP binding. Enrichment levels on annotated genes from ptChIP-seq using αRpoB antibody were compared between Col-0 wild type and *sig2* plants.

C. Dual impact of *sig6* on PEP binding. Enrichment levels on annotated genes from ptChIP-seq using αRpoB antibody were compared between Col-0 wild type and *sig6* plants.

In B and C data points are color-coded by the function of the corresponding genes and show averages from three independent replicates. Error bars indicate standard deviations. Blue line represents the linear regression model. Red line represents values equal between both genotypes.

D. Reduction of PEP binding to DNA in *sig2* and *sig6* on the *trnEYD* operon.

E. Reduction of PEP binding to DNA in *sig2* and *sig6* on *trnI1*.

In D and E signal enrichments from ptChIP-seq using αRpoB antibody in Col-0 wild type, *sig2* and *sig6* plants were calculated in 10 bp genomic bins and plotted at the relevant locus. Color coding of ptChIP-seq data corresponds to data shown in Fig. 4A. Average enrichments from three independent biological replicates are shown. Ribbons indicates standard deviations. Brown vertical lines indicate sense strand data from three combined replicates of TSS RNA-seq. Genome annotation is shown on top.

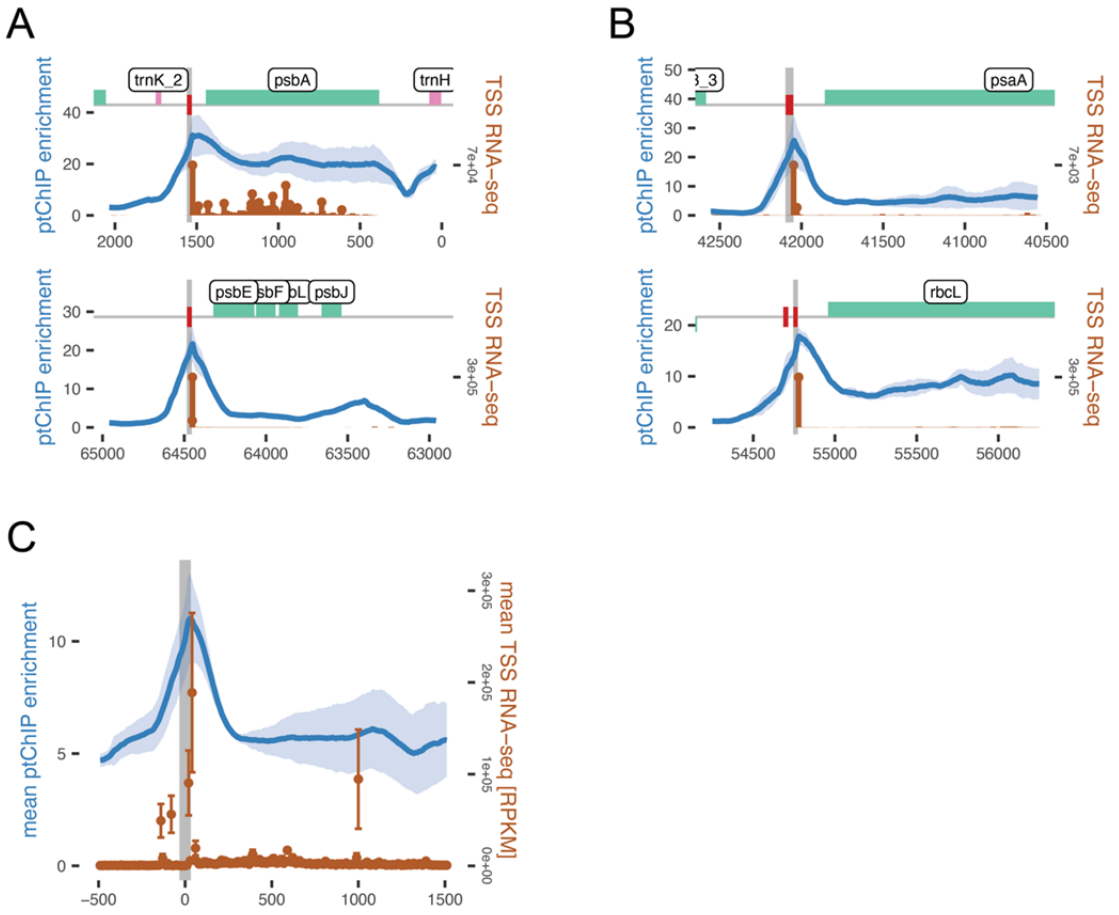


Figure 5. Preferential binding of PEP to gene promoters.

A. Binding of PEP and transcription start sites on promoters previously shown to be bound by PEP.

B. Preferential binding of PEP and transcription start sites on other promoters.

In A and B signal enrichment from ptChIP-seq using αRpoB antibody in Col-0 wild type was calculated in 10 bp genomic bins and plotted at the relevant loci. Average enrichments from three independent biological replicates are shown. Light blue ribbons indicate standard deviations. Brown vertical lines indicate sense strand data from three combined replicates of TSS RNA-seq. Grey vertical line indicates position of the annotated promoter. Genome annotation is shown on top.

C. Average binding of PEP and transcription start sites on all promoters found in Arabidopsis. Genomic regions surrounding all PEP promoters we identified in prior studies in Arabidopsis were aligned and mean ptChIP-seq enrichment using αRpoB antibody in Col-0 wild type was calculated in 10 nt genomic bins for each biological

578 replicate. Mean value from three biological replicates is shown. Light blue ribbon
579 corresponds to standard deviation. Mean TSS RNA-seq signal was calculated from both
580 strands in 10 bp genomic bins for each biological replicate. Brown dots correspond to
581 mean values from three biological replicates. Error bars indicate standard deviations.
582 Grey vertical line indicates aligned position of the annotated promoter.
583

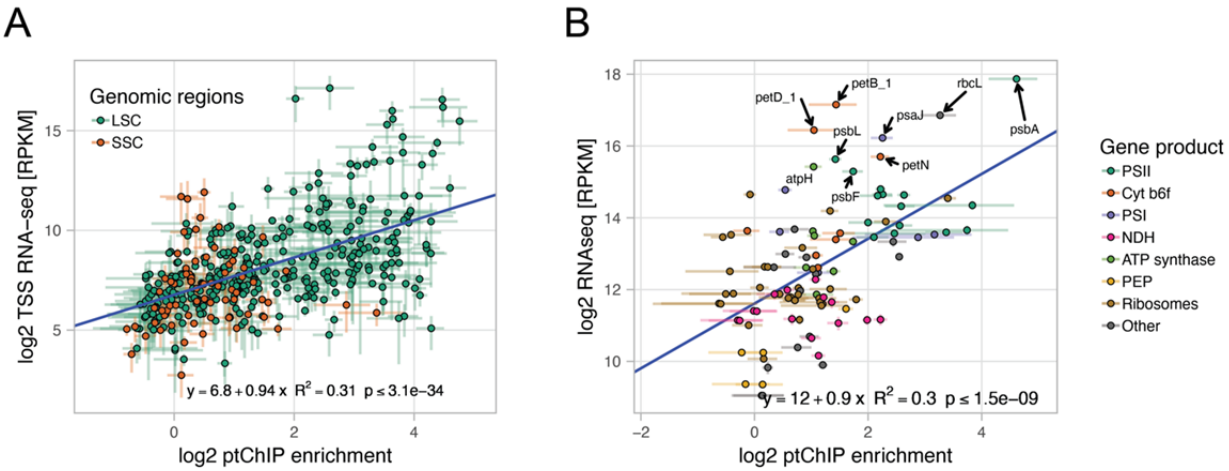


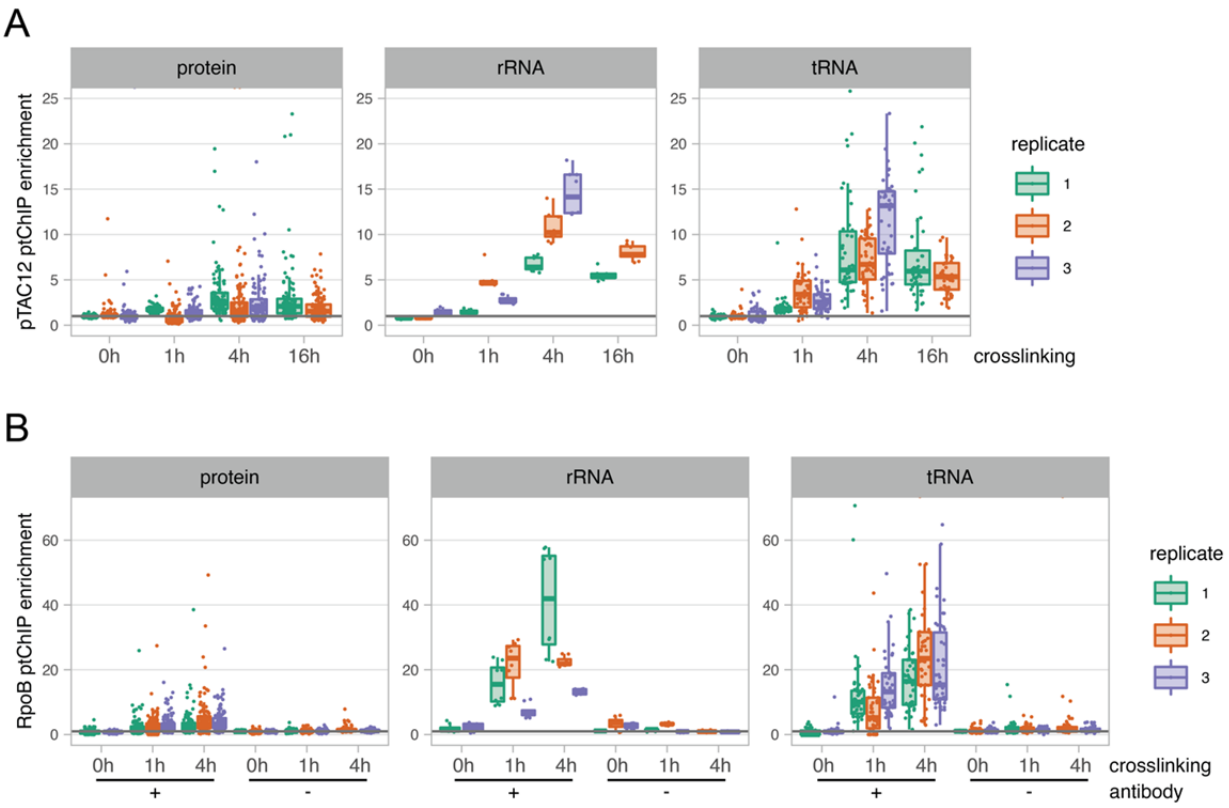
Figure 6. Correlation of PEP binding with steady state levels of RNA.

A. RpoB ptChIP-seq enrichments and TSS RNA-seq signals are highly correlated. Enrichment levels of RpoB ptChIP-seq and RPKM normalized TSS RNA-seq signals in Col-0 wild type were compared on 250 bp genomic bins including LSC and SSC. Data points are color-coded by locations. Blue line represents the linear regression model.

B. RpoB ptChIP-seq enrichments and total RNA-seq signals are highly correlated on protein coding genes. Enrichment levels of RpoB ptChIP-seq in 14 days old Col-0 wild type plants (this study) and RPKM normalized total RNA-seq signals from a similar developmental stage (Thieffry et al., 2020) were compared on annotated protein coding genes. Data points are color-coded by the function of the corresponding genes and show averages from three independent replicates. Error bars indicate standard deviations. Blue line represents the linear regression model.

598 **SUPPLEMENTAL FIGURES**

599



600

601 **Figure S1.** Detection of PEP binding to DNA using ptChIP-seq. Individual biological replicates
602 of data shown in Fig. 1.

603 A. Optimization of formaldehyde crosslinking time in ptChIP-seq. Individual biological
604 replicates of ptChIP-seq performed using αHA antibody in plants expressing pTAC12-
605 HA shown in Fig. 1A.

606 B. Optimization of formaldehyde crosslinking time and negative controls in ptChIP-seq.
607 Individual biological replicates of ptChIP-seq using αRpoB antibody in Col-0 wild type
608 plants shown in Fig. 1B.

609

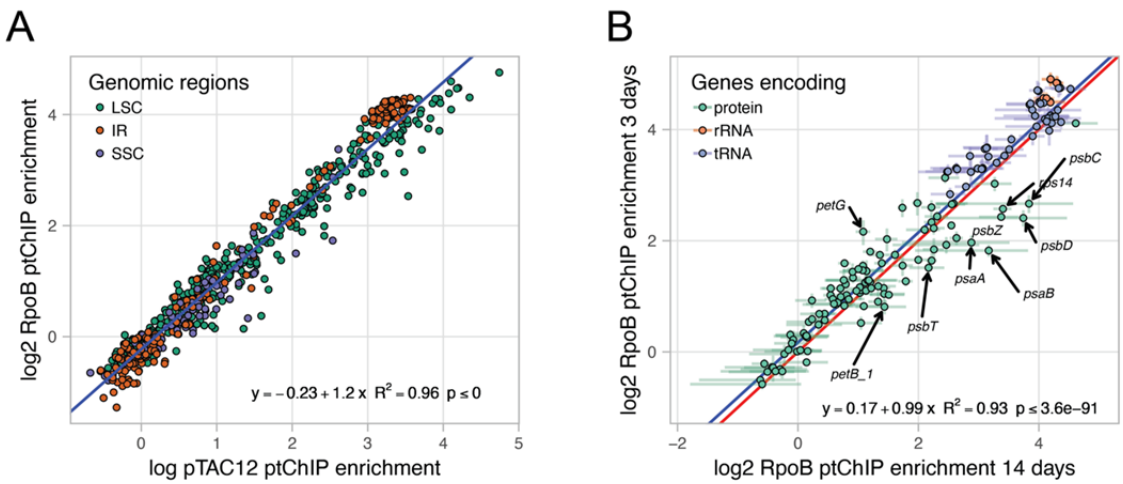


Figure S2. Specificity of ptChIP-seq with αRpoB antibody.

A. pTAC12 and RpoB ptChIP-seq signals are highly correlated. Enrichment levels on 250 bp genomic bins were compared between ptChIP-seq experiments using αHA antibody in plants expressing pTAC12-HA and using αRpoB antibody in Col-0 wild type plants. Data points are color-coded by locations of bins within the LSC, IR or SSC. Blue line represents the linear regression model.

B. RpoB ptChIP-seq signals are highly correlated between 4-day-old plants and 14-day-old plants. Enrichment levels on annotated genes from ptChIP-seq using αRpoB antibody in Col-0 wild type plants were compared between 4-day-old and 14-day-old plants. Data points are color-coded by the function of the corresponding genes and show averages from three independent replicates. Error bars indicate standard deviations. Blue line represents the linear regression model.

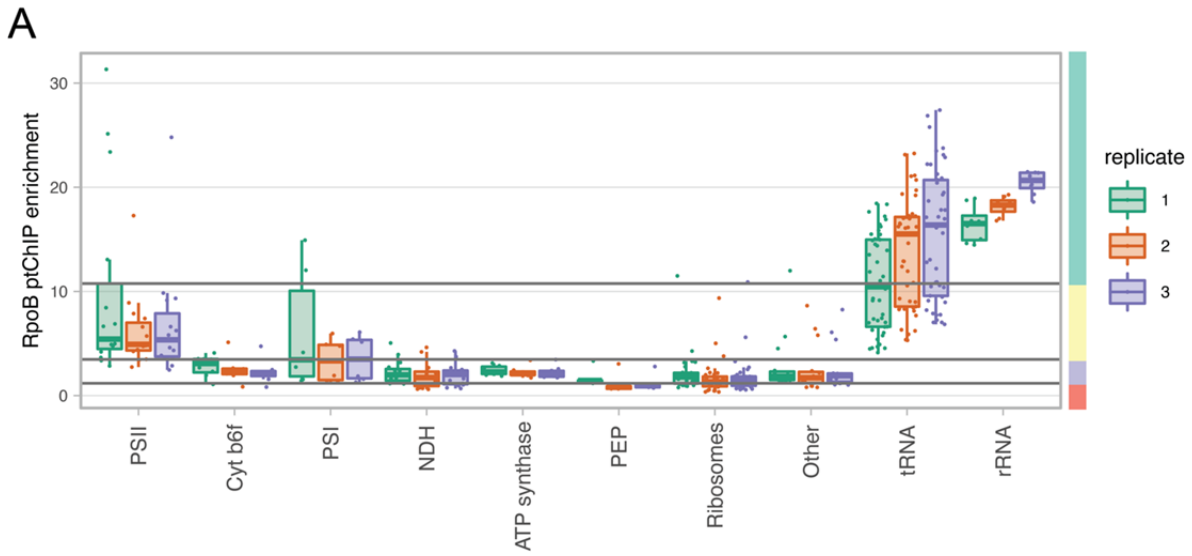


Figure S3. Complex pattern of PEP binding to plastid DNA. Individual biological replicates of data shown in Fig. 3C.

A. PEP binding to DNA of genes classified by the function of their products. Enrichment levels of ptChIP-seq using αRpoB antibody in Col-0 wild type were plotted on annotated genes split by the function of gene products (Chotewutmontri and Barkan, 2018). PEP binding level groups are indicated on the right. Colors indicate data from independent replicates.

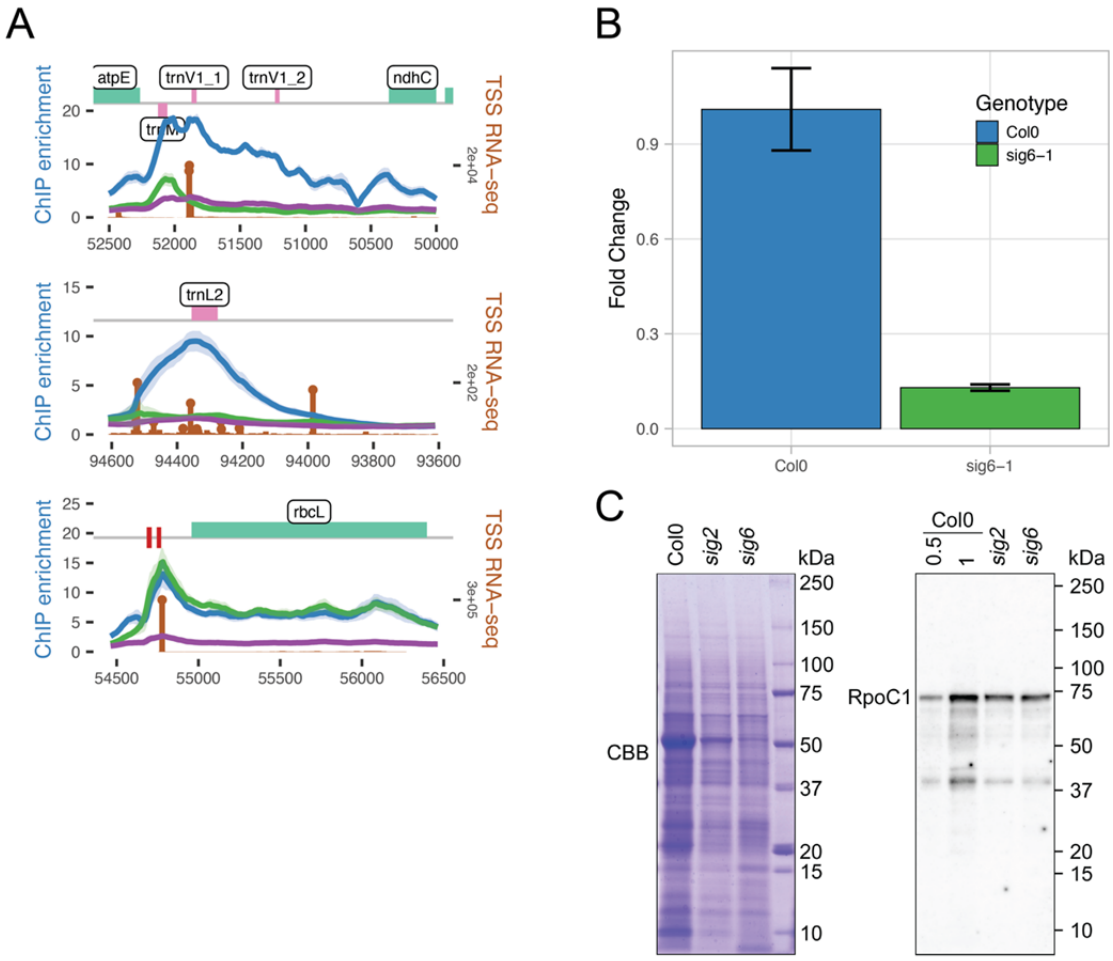


Figure S4. Dual impact of *sig2* and *sig6* mutants on PEP binding to plastid DNA.

A. Reduction of PEP binding to DNA in *sig2* and *sig6* on *trnV*, *trnL* and *rbcL*. Signal enrichments from ptChIP-seq using αRpoB antibody in Col-0 wild type, *sig2* and *sig6* plants were calculated in 10 bp genomic bins and plotted at the relevant locus. Color-coding of ptChIP-seq data corresponds to data shown in Fig. 4A. Average enrichments from three independent biological replicates are shown. Ribbons indicates standard deviations. Brown vertical lines indicate sense strand data from three combined replicates of TSS RNA-seq. Genome annotation is shown on top.

B. Reduction of the tRNA product of *trnI1* (tRNA-Ile-CAU) in the *sig6* mutant. Accumulation of the tRNA was assayed using RT-qPCR. Average and standard deviation from three biological replicates are shown.

C. Abundance of the RpoC1 protein in *sig2* and *sig6*. 20 µg of total proteins were loaded. Gel stained with Coomassie brilliant blue (CBB) is shown as a loading control.

A

Gene	Sequence	Position relative to translation start site	Ref.
psbA	TTTACATTGG**TTGACA**TGGCTATATAAGTCATGT----**TATACT**GTTTCATAACAAGCTCTCAA	−66	1
trnQ	TCATTTTGAA**TTGACG**AACAACCAATAGGAATAT----**TACTCT**TTTAGTAGTCTTGAATAAAA	−8	2
atpH	TTGGTTTACA**TTATGT**AAGAAACACTTGTATA------**TGTGAT**ATTTGATATTGCCTAGGTAT	−403	3
atpI	AGGGCCGTTC**TAGCTA**TATACAAAATTCTTGATT----**AATAAT**AAGATAAATAATCAATTTT	−215	4
petN	TCCCTTTTTT**TTGACT**CTGCACCAGTGATTTCAC----**TATTAT**TAGTGAACAATAATGGAATA	−39	5
trnE	TTTCATTATA**TTGACA**ATTTAAAAAAACTGATCA----**TACTAT**GATCATAGTATGATGGCGGT	−12	2
psbD	ACCTAACCCA**TCGAAT**CATGACTATATCCAC-------**TATTCT**GATATTCAAATTCGATAGAG	−934	6,7
psbD	GAAACTCTCA**TTTACA**GTTTCCTATAATTTTATT----**TAAATA**TTGAATTAAATATAAATAAG	−527	6,8
psbD	CATCCGACAA**TTCATG**ATTTAGATTCAACTACT-----**TATACT**TATTAATAAACTAATAGCAA	−244	7
psbD	ACAAATTGAG**TTGATC**CGTTTACCTAAGTAAGGACCAA**TAAAAT**CAAAAATTTTGATCTTCGAA	−173	6,8
psaA	AAAGGGTCCG**TTGAGC**ACCCTATGGATATGT-------**CATAAT**AGATCCGAACACTTGCCCCA	−175	5,9
rps4	TTTATATATA**TTGTAT**ATACAATAAAAAAAATCTG---**TATATT**GAGAGAATATATCTTCTTAC	−109	4
trnV1	AATATTAATC**TTGACA**AGAAATTATCTACATGA-----**TAAGAT**AAAATGTGTATCACAAACAC	−2	7
trnM	TTTATTGGTA**TTGCTT**AGAAATAATATTGGATT-----**TATAAT**CCTATCGATGTGATAGGTAT	−31	2
atpB	TTTTGAATCT**TTGACT**TTTTATAAATCCAAATATT---**TAATAT**TAAAATTAAAATATTAATAA	−508	5
atpB	AAAATCACTC**TTGACA**GTAATATCTGTTGTATA-----**TGTAAA**TCCTAGATATGACAATATGC	−451	4
rbcL	AAGTATTAGG**TTGCGC**TATACATATGAAAGAATA----**TACAAT**AATGATGTATTTGGCGAATC	−164	10
ycf4	TTTTAAAAAC**TTCGAT**TCGATAAGACCCATTTGG----**TAGAAT**GTTGTATAACACATAGATTC	−314	11
psbE	TCAAATTGCG**TTGCTG**TGTCAGAAGAAGGATAGC----**TATACT**GATTCGGTAGACTCTAAAAA	−112	5
psaJ	CTTAATTACA**TGTACA**TCTGTAATTATATATAT-----**TACTAT**ATATAATGTAATACAATAAA	−23	12
psbB	TTAACCCCCA**TTGCAT**ATTGGTACTTATCGGATA----**TAGAAT**AGATCCGCTTCCCTTTTTTT	−158	5
psbN	TCTGTTGACT**TTGTAT**ACCATTCCGTTGTAAA------**TAAATG**ATCTTAGCATAGATCCGGTG	−22	3
rrn16	GCTCGTGGGA**TTGACG**TGAGGGGGTAGGGGTAGC----**TATATT**TCTGGGAGCGAACTCCATGC	−102	13
ndhF	GTGGAAATCT**TTGTTC**TATTCTTAATATATGTA-----**TATAAA**TTATTGTAATATGGAATTAT	−305	11
tic214	ATTTCTATAT**ATGGAA**AGTTGCAAAATCATCA------**TATAAT**AATCCAGAAATTGAAATAGA	−20	4

BOLD	−35 and −10 motives
UNDERLINE	Reported transcription start site
RED BOLD	5′ ends of RNA
BLUE BACKGROUND	Peak of RpoB enrichment

B

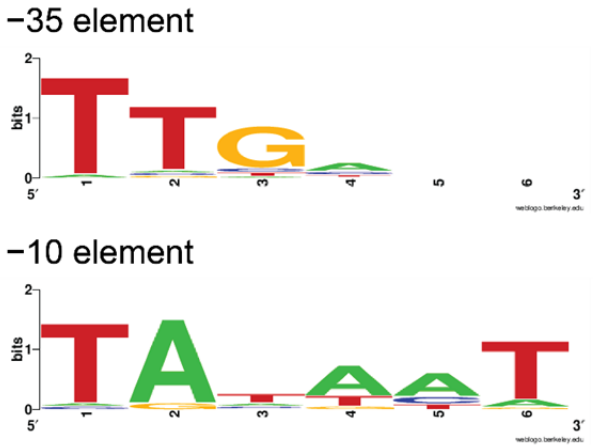


Figure S5. PEP promoters identified in Arabidopsis.

A. Alignment of PEP promoters. Positions of the last nucleotides (20 nt from the −10 element) relative to translation start sites are indicated. Transcription start sites reported previously and identified with TSS RNA-seq in this study are represented with underlines and red colors, respectively. Blue background indicates the position with highest RpoB

653 enrichment downstream of each −10 element. [1]Liere et al., 1995; [2]Kanamaru et al., 2001;
654 [3]Zghidi et al., 2007; [4]Swiatecka-Hagenbruch et al., 2007; [5]Ishizaki et al., 2005; [6]Hoffer
655 and Christopher, 1997; [7]Hanaoka et al., 2003; [8]Shimmura et al., 2008; [9]Fey et al., 2005;
656 [10]Hakimi et al., 2000; [11]Favory et al., 2005; [12]Nagashima et al., 2004; [13]Sriraman et al.,
657 1998.
658 B. Sequence logos for −35 and −10 elements of PEP promoters listed in A. Images were
659 generated by using WebLogo (https://weblogo.berkeley.edu/logo.cgi) with default
660 settings.
661

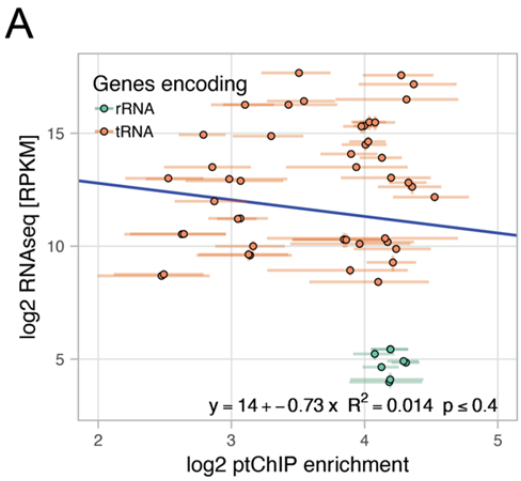


Figure S6. Correlation of PEP binding with steady state levels of RNA.

A. No correlation on rRNA and tRNA genes between RpoB ptChIP-seq enrichments and total RNA-seq signals. Enrichment levels of RpoB ptChIP-seq in 14 days old Col-0 wild type plants (this study) and RPKM normalized total RNA-seq signals from a similar developmental stage (Thieffry et al., 2020) were compared on annotated rRNA and tRNA genes. Data points are color-coded by function of the corresponding genes and show averages from three independent replicates. Error bars indicate standard deviations. Blue line represents the linear regression model.

673 **Table S1.** High throughput sequencing datasets generated in this study.

Experiment	Genotype and treatment	Repl.	Numbers of paired end sequencing reads			Plastid genome coverage	GEO accession
			Raw	Mapped to whole genome	Mapped to plastid genome		
ptChIP-seq in Col-0	αRpoB ptChIP Col-0, 14do IP	1	4054606	2531764	847535	1645.9	GSM5751803
		2	2882681	2107372	769067	1493.6	GSM5751804
		3	2433915	1827970	1066030	2070.3	GSM5751805
	αRpoB ptChIP Col-0, 14do input	1	2658090	1972573	653441	1269.0	GSM5751806
		2	2198631	1616715	932056	1810.1	GSM5751807
		3	4179457	3763419	2011669	3906.7	GSM5751808
ptChIP-seq in sigma factor mutants	αRpoB ptChIP Col-0, 4do IP	1	3178254	2042275	1749678	3397.9	GSM5751809
		2	4290851	2553815	2002541	3889.0	GSM5751810
		3	4022836	2726824	2146437	4168.5	GSM5751811
	αRpoB ptChIP Col-0, 4do input	1	3054715	2696902	1508241	2929.1	GSM5751812
		2	3512602	3155100	1613083	3132.7	GSM5751813
		3	2213601	1982579	1184063	2299.5	GSM5751814
	αRpoB ptChIP *sig2-2*, 4do IP	1	5124923	1814658	991452	1925.4	GSM5751815
		2	5461169	2611987	1665262	3234.0	GSM5751816
		3	5449395	3213878	2128564	4133.7	GSM5751817
	αRpoB ptChIP *sig2-2*, 4do input	1	3275658	2857270	1851446	3595.6	GSM5751818
		2	2835440	2522089	1608186	3123.2	GSM5751819
		3	2321023	2069124	1430554	2778.2	GSM5751820
	αRpoB ptChIP *sig6-1*, 4do IP	1	2270376	1312654	625240	1214.2	GSM5751821
		2	4898113	2548998	1180339	2292.3	GSM5751822
		3	4337287	2693064	1581682	3071.7	GSM5751823
	αRpoB ptChIP *sig6-1*, 4do input	1	3409410	2921964	1779439	3455.7	GSM5751824
		2	2733495	2375248	1197919	2326.4	GSM5751825
		3	3314878	2911594	1863476	3618.9	GSM5751826
ptChIP xlinking optimization αHA	αHA ptChIP Col-0, 14do IP, 0h xlinking	1	920851	56960	32956	64.0	GSM5751827
		2	1795101	867714	506490	983.6	GSM5751831
		3	1257078	5319	2663	5.2	GSM5751835
	αHA ptChIP Col-0, 14do IP, 1h xlinking	1	1400807	325523	18444	35.8	GSM5751829
		2	1371708	951186	26754	52.0	GSM5751833
		3	1454561	464395	10597	20.6	GSM5751836
	αHA ptChIP Col-0, 14do IP, 4h xlinking	1	863247	444108	17164	33.3	GSM5751830
		2	1097756	513773	20826	40.4	GSM5751834
		3	1378762	704513	153955	299.0	GSM5751837
	αHA ptChIP Col-0, 14do IP, 16h xlinking	1	1280235	455508	57230	111.1	GSM5751828
		2	859243	401668	20593	40.0	GSM5751832
	αHA ptChIP pTAC12-HA, 14do IP, 0h xlinking	1	1044721	133718	78968	153.4	GSM5751838
		2	861576	93290	49282	95.7	GSM5751842
		3	1274177	61517	37733	73.3	GSM5751846
	αHA ptChIP pTAC12-HA, 14do IP, 1h xlinking	1	1523025	785809	142068	275.9	GSM5751840
		2	921045	543108	161582	313.8	GSM5751844
		3	1075036	784516	44905	87.2	GSM5751847
	αHA ptChIP pTAC12-HA, 14do IP, 4h xlinking	1	1865224	846911	455825	885.2	GSM5751841
		2	2386408	1844646	1399279	2717.5	GSM5751845
		3	1305128	278014	168917	328.0	GSM5751848
	αHA ptChIP pTAC12-HA, 14do IP, 16h xlinking	1	1773357	1118771	736914	1431.1	GSM5751839
		2	3405511	2380146	2101294	4080.8	GSM5751843
ptChIP xlinking optimization αRpoB	αRpoB ptChIP Col-0, 14do IP, 0h xlinking	1	101564	1274	516	1.0	GSM5751849
		3	1311600	6967	3699	7.2	GSM5751854
	αRpoB ptChIP	1	1290652	669663	404378	785.3	GSM5751850

	Col-0, 14do IP, 1h xlinking	2	2344827	2301	748	1.5	GSM5751852
		3	2118397	241528	79914	155.2	GSM5751855
	αRpoB ptChIP Col-0, 14do IP, 4h xlinking	1	1494376	581952	373502	725.4	GSM5751851
		2	1407565	77717	40178	78.0	GSM5751853
		3	1989369	251422	197843	384.2	GSM5751856
	αRpoB ptChIP Col-0, 14do input, 0h xlinking	1	1946086	1773603	1186320	2303.9	GSM5751857
		2	1314300	584617	410968	798.1	GSM5751860
		3	1904283	361640	234984	456.3	GSM5751863
	αRpoB ptChIP Col-0, 14do input, 1h xlinking	1	2017339	921339	663423	1288.4	GSM5751858
		2	2353928	44686	25616	49.7	GSM5751861
		3	2630442	1601690	973938	1891.4	GSM5751864
	αRpoB ptChIP Col-0, 14do input, 4h xlinking	1	857691	416462	313340	608.5	GSM5751859
		2	2235821	14520	8878	17.2	GSM5751862
		3	2041195	90834	54473	105.8	GSM5751865
	noAB ptChIP Col-0, 14do IP, 0h xlinking	1	1390386	302040	206679	401.4	GSM5751866
		2	1609455	6691	2996	5.8	GSM5751868
		3	1729661	191269	119759	232.6	GSM5751871
	noAB ptChIP Col-0, 14do IP, 1h xlinking	1	2003662	113136	4017	7.8	GSM5751867
		2	1846817	200328	4218	8.2	GSM5751869
		3	2264948	113027	2912	5.7	GSM5751872
	noAB ptChIP Col-0, 14do IP, 4h xlinking	2	2119540	229291	16608	32.3	GSM5751870
		3	3974954	210751	8665	16.8	GSM5751873
TSS RNA-seq	TSS RNA-seq Col-0, 14do	1	37708807	31520792	7916770	15374.7	GSM5751874
		2	24176101	19826820	8907118	17297.9	GSM5751875
		3	21521488	16265941	5854662	11370.0	GSM5751876

677 **Table S2.** Annotation of plastid-encoded genes in Arabidopsis used in this study.

chr	start	end	strand	type	name	long name	gene ID
Pt	4	76	-	tRNA	trnH	tRNA-His-GUG	ATCG00010
Pt	383	1444	-	protein	psbA	PSII D1 protein	ATCG00020
Pt	1717	1751	-	tRNA	trnK_2	tRNA-Lys-UUU	ATCG00030
Pt	2056	3570	-	protein	matK	Splicing factor	ATCG00040
Pt	4311	4347	-	tRNA	trnK_1	tRNA-Lys-UUU	ATCG00030
Pt	5084	5283	-	protein	rps16_2	30S Ribosomal protein CS16	ATCG00050
Pt	6149	6188	-	protein	rps16_1	30S Ribosomal protein CS16	ATCG00050
Pt	6616	6687	-	tRNA	trnQ	tRNA-Gln-UUG	ATCG00060
Pt	7017	7202	+	protein	psbK	PSII K protein	ATCG00070
Pt	7583	7693	+	protein	psbI	PSII I protein	ATCG00080
Pt	7785	7872	+	tRNA	trnS1	tRNA-Ser-GCU	ATCG00090
Pt	8646	8668	+	tRNA	trnG1_1	tRNA-Gly-UCC	ATCG00100
Pt	9383	9431	+	tRNA	trnG1_2	tRNA-Gly-UCC	ATCG00100
Pt	9590	9661	+	tRNA	trnR1	tRNA-Arg-UCU	ATCG00110
Pt	9938	11461	-	protein	atpA	ATP synthase subunit CF1 alpha	ATCG00120
Pt	11529	11938	-	protein	atpF_2	ATP synthase subunit CFo I	ATCG00130
Pt	12654	12798	-	protein	atpF_1	ATP synthase subunit CFo I	ATCG00130
Pt	13262	13507	-	protein	atpH	ATP synthase subunit CFo III	ATCG00140
Pt	14021	14770	-	protein	atpI	ATP synthase subunit CFo IV	ATCG00150
Pt	15013	15723	-	protein	rps2	30S Ribosomal protein CS2	ATCG00160
Pt	15938	20068	-	protein	rpoC2	PEP subunit beta prime prime	ATCG00170
Pt	20251	21861	-	protein	rpoC1_2	PEP subunit beta prime	ATCG00180
Pt	22653	23084	-	protein	rpoC1_1	PEP subunit beta prime	ATCG00180
Pt	23111	26329	-	protein	rpoB	PEP subunit beta	ATCG00190
Pt	27373	27443	+	tRNA	trnC	tRNA-Cys-GCA	ATCG00200
Pt	28089	28178	+	protein	petN	Cytochrome b6/f complex subunit N	ATCG00210
Pt	28707	28811	-	protein	psbM	PSII M protein	ATCG00220
Pt	29801	29874	-	tRNA	trnD	tRNA-Asp-GUC	ATCG00230
Pt	30323	30406	-	tRNA	trnY	tRNA-Tyr-GUA	ATCG00240
Pt	30466	30538	-	tRNA	trnE	tRNA-Glu-UUC	ATCG00250
Pt	31369	31440	+	tRNA	trnT1	tRNA-Thr-GGU	ATCG00260
Pt	32711	33772	+	protein	psbD	PSII D2 protein	ATCG00270
Pt	33720	35141	+	protein	psbC	PSII CP43 protein	ATCG00280
Pt	35312	35403	-	tRNA	trnS2	tRNA-Set-UGA	ATCG00290
Pt	35751	35939	+	protein	psbZ	PSII Z protein	ATCG00300
Pt	36490	36560	+	tRNA	trnG2	tRNA-Gly-GCC	ATCG00310
Pt	36704	36777	-	tRNA	trnfM	tRNA-fMet-CAU	ATCG00320
Pt	36938	37240	-	protein	rps14	30S Ribosomal protein CS14	ATCG00330
Pt	37375	39579	-	protein	psaB	PSI PsaB protein	ATCG00340
Pt	39605	41857	-	protein	psaA	PSI PsaA protein	ATCG00350
Pt	42584	42736	-	protein	ycf3_3	PSI assembly	ATCG00360
Pt	43524	43751	-	protein	ycf3_2	PSI assembly	ATCG00360
Pt	44466	44591	-	protein	ycf3_1	PSI assembly	ATCG00360
Pt	44827	44913	+	tRNA	trnS3	tRNA-Ser-GGA	ATCG00370
Pt	45223	45828	-	protein	rps4	30S Ribosomal protein CS4	ATCG00380
Pt	46213	46285	-	tRNA	trnT2	tRNA-Thr-UGU	ATCG00390
Pt	46894	46928	+	tRNA	trnL1_1	tRNA-Leu-UAA	ATCG00400
Pt	47441	47490	+	tRNA	trnL1_2	tRNA-Leu-UAA	ATCG00400
Pt	48175	48247	+	tRNA	trnF	tRNA-Phe-GAA	ATCG00410
Pt	48677	49153	-	protein	ndhJ	NDH complex subunit J	ATCG00420
Pt	49257	49934	-	protein	ndhK	NDH complex subunit K	ATCG00430
Pt	50001	50363	-	protein	ndhC	NDH complex subunit C	ATCG00440
Pt	51199	51233	-	tRNA	trnV1_2	tRNA-Val-UAC	ATCG00450
Pt	51833	51871	-	tRNA	trnV1_1	tRNA-Val-UAC	ATCG00450
Pt	52056	52128	+	tRNA	trnM	tRNA-Met-CAU	ATCG00460
Pt	52265	52663	-	protein	atpE	ATP synthase subunit CF1 epsilon	ATCG00470

Pt	52660	54156	-	protein	atpB	ATP synthase subunit CF1 beta	ATCG00480
Pt	54958	56397	+	protein	rbcL	Rubisco large subunit	ATCG00490
Pt	57075	58541	+	protein	accD	Acetyl-CoA carboxylase beta subunit	ATCG00500
Pt	59247	59360	+	protein	psaI	PSI I protein	ATCG00510
Pt	59772	60326	+	protein	ycf4	PSI assembly	ATCG00520
Pt	60741	61430	+	protein	cemA	Inner envelope protein	ATCG00530
Pt	61657	62619	+	protein	petA	Cytochrome b6/f complex cytochrome f	ATCG00540
Pt	63538	63660	-	protein	psbJ	PSII J protein	ATCG00550
Pt	63804	63920	-	protein	psbL	PSII L protein	ATCG00560
Pt	63942	64061	-	protein	psbF	PSII cytochrome b559 alpha subunit	ATCG00570
Pt	64071	64322	-	protein	psbE	PSII cytochrome b559 beta subunit	ATCG00580
Pt	65712	65807	+	protein	petL	Cytochrome b6/f complex subunit L	ATCG00590
Pt	65998	66111	+	protein	petG	Cytochrome b6/f complex subunit G	ATCG00600
Pt	66229	66302	-	tRNA	trnW	tRNA-Trp-CCA	ATCG00610
Pt	66490	66563	-	tRNA	trnP	tRNA-Pro-UGG	ATCG00620
Pt	66929	67063	+	protein	psaJ	PSI J protein	ATCG00630
Pt	67488	67688	+	protein	rpl33	50S ribosomal protein CL33	ATCG00640
Pt	67917	68222	+	protein	rps18	30S ribosomal protein CS18	ATCG00650
Pt	68512	68865	-	protein	rpl20	50S ribosomal protein CL20	ATCG00660
Pt	69611	69724	-	protein	rps12_1	30S ribosomal protein CS12	ATCG00065
Pt	69910	70137	-	protein	clpP_3	ATP-dependent Clp protease	ATCG00670
Pt	70653	70944	-	protein	clpP_2	ATP-dependent Clp protease	ATCG00670
Pt	71812	71882	-	protein	clpP_1	ATP-dependent Clp protease	ATCG00670
Pt	72371	73897	+	protein	psbB	PSII CP47 protein	ATCG00680
Pt	74082	74183	+	protein	psbT	PSII T protein	ATCG00690
Pt	74249	74380	-	protein	psbN	PSII N protein	ATCG00700
Pt	74485	74706	+	protein	psbH	PSII H protein	ATCG00710
Pt	74841	74846	+	protein	petB_1	Cytochrome b6/f complex cytochrome b6	ATCG00720
Pt	75651	76292	+	protein	petB_2	Cytochrome b6/f complex cytochrome b6	ATCG00720
Pt	76481	76488	+	protein	petD_1	Cytochrome b6/f complex subunit IV	ATCG00730
Pt	77198	77672	+	protein	petD_2	Cytochrome b6/f complex subunit IV	ATCG00730
Pt	77901	78890	-	protein	rpoA	PEP subunit alpha	ATCG00740
Pt	78960	79376	-	protein	rps11	30S ribosomal protein CS11	ATCG00750
Pt	79489	79602	-	protein	rpl36	50S ribosomal protein CL36	ATCG00760
Pt	80068	80472	-	protein	rps8	30S ribosomal protein CS8	ATCG00770
Pt	80696	81064	-	protein	rpl14	50S ribosomal protein CL14	ATCG00780
Pt	81189	81587	-	protein	rpl16_2	50S ribosomal protein CL16	ATCG00790
Pt	82644	82652	-	protein	rpl16_1	50S ribosomal protein CL16	ATCG00790
Pt	82826	83482	-	protein	rps3	30S ribosomal protein CS3	ATCG00800
Pt	83467	83949	-	protein	rpl22	50S ribosomal protein CL22	ATCG00810
Pt	84005	84283	-	protein	rps19	30S ribosomal protein CS19	ATCG00820
Pt	84337	84771	-	protein	rpl2_2	50S ribosomal protein CL2	ATCG00830
Pt	85454	85843	-	protein	rpl2_1	50S ribosomal protein CL2	ATCG00830
Pt	85862	86143	-	protein	rpl23	50S ribosomal protein CL23	ATCG00840
Pt	86312	86385	-	tRNA	trnI1	tRNA-Ile-CAU	ATCG00850
Pt	86474	93358	+	protein	ftsHi	AAA-protease	ATCG00860
Pt	94276	94356	-	tRNA	trnL2	tRNA-Leu-CAA	ATCG00880
Pt	94941	95702	-	protein	ndhB_2	NDH complex subunit B	ATCG00890
Pt	96388	97164	-	protein	ndhB_1	NDH complex subunit B	ATCG00890
Pt	97478	97945	-	protein	rps7	30S ribosomal protein CS7	ATCG00900
Pt	97999	98024	-	protein	rps12_3	30S ribosomal protein CS12	ATCG00905
Pt	98562	98793	-	protein	rps12_2	30S ribosomal protein CS12	ATCG00905
Pt	100709	100780	+	tRNA	trnV2	tRNA-Val-GAC	ATCG00910
Pt	101012	102502	+	rRNA	rrn16	16S rRNA	ATCG00920
Pt	102801	102837	+	tRNA	trnI2_1	tRNA-Ile-GAU	ATCG00930
Pt	103567	103601	+	tRNA	trnI2_2	tRNA-Ile-GAU	ATCG00930
Pt	103665	103702	+	tRNA	trnA_1	tRNA-Ala-UGC	ATCG00940
Pt	104504	104538	+	tRNA	trnA_2	tRNA-Ala-UGC	ATCG00940
Pt	104691	107500	+	rRNA	rrn23	23S rRNA	ATCG00950

Pt	107599	107701	+	rRNA	rrn4.5	4.5S rRNA	ATCG00960
Pt	107949	108069	+	rRNA	rrn5	5S rRNA	ATCG00970
Pt	108302	108375	+	tRNA	trnR2	tRNA-Arg-ACG	ATCG00980
Pt	109013	109084	-	tRNA	trnN	tRNA-Asn-GUU	ATCG00990
Pt	109405	110436	+	protein	tic214	TIC214 Partial	ATCG01000
Pt	110398	112638	-	protein	ndhF	NDH complex subunit F	ATCG01010
Pt	113449	113607	+	protein	rpl32	50S ribosomal protein CL32	ATCG01020
Pt	114270	114349	+	tRNA	trnL3	tRNA-Leu-UAG	ATCG01030
Pt	114461	115447	+	protein	ccsA	cytochrome c6 biosynthesis protein	ATCG01040
Pt	115665	117167	-	protein	ndhD	NDH complex subunit D	ATCG01050
Pt	117318	117563	-	protein	psaC	PSI PsaC protein	ATCG01060
Pt	117804	118109	-	protein	ndhE	NDH complex subunit E	ATCG01070
Pt	118377	118907	-	protein	ndhG	NDH complex subunit G	ATCG01080
Pt	119244	119762	-	protein	ndhI	NDH complex subunit I	ATCG01090
Pt	119847	120376	-	protein	ndhA_2	NDH complex subunit A	ATCG01100
Pt	121457	122009	-	protein	ndhA_1	NDH complex subunit A	ATCG01100
Pt	122011	123192	-	protein	ndhH	NDH complex subunit H	ATCG01110
Pt	123296	123562	-	protein	rps15	30S ribosomal protein CS15	ATCG01120
Pt	123884	129244	-	protein	tic214	Translocon subunit TIC214	ATCG01130
Pt	129565	129636	+	tRNA	trnN	tRNA-Asn-GUU	ATCG01140
Pt	130274	130347	-	tRNA	trnR2	tRNA-Arg-ACG	ATCG01150
Pt	130580	130700	-	rRNA	rrn5	5S rRNA	ATCG01160
Pt	130948	131050	-	rRNA	rrn4.5	4.5S rRNA	ATCG01170
Pt	131149	133958	-	rRNA	rrn23	23S rRNA	ATCG01180
Pt	134111	134145	-	tRNA	trnA_2	tRNA-Ala-UGC	ATCG01190
Pt	134947	134984	-	tRNA	trnA_1	tRNA-Ala-UGC	ATCG01190
Pt	135048	135082	-	tRNA	trnI2_2	tRNA-Ile-GAU	ATCG01200
Pt	135812	135848	-	tRNA	trnI2_1	tRNA-Ile-GAU	ATCG01200
Pt	136147	137637	-	rRNA	rrn16	16S rRNA	ATCG01210
Pt	137869	137940	-	tRNA	trnV2	tRNA-Val-GAC	ATCG01220
Pt	139856	140087	+	protein	rps12_2	30S ribosomal protein CS12	ATCG01230
Pt	140625	140650	+	protein	rps12_3	30S ribosomal protein CS12	ATCG01230
Pt	140704	141171	+	protein	rps7	30S ribosomal protein CS7	ATCG01240
Pt	141485	142261	+	protein	ndhB_1	NDH complex subunit B	ATCG01250
Pt	142947	143708	+	protein	ndhB_2	NDH complex subunit B	ATCG01250
Pt	144293	144373	+	tRNA	trnL2	tRNA-Leu-CAA	ATCG01260
Pt	145291	152175	-	protein	ftsHi	AAA-protease	ATCG01280
Pt	152264	152337	+	tRNA	trnI1	tRNA-Ile-CAU	ATCG01290
Pt	152506	152787	+	protein	rpl23	50S ribosomal protein CL2	ATCG01300
Pt	152806	153195	+	protein	rpl2_1	50S ribosomal protein CL2	ATCG01310
Pt	153878	154312	+	protein	rpl2_2	50S ribosomal protein CL2	ATCG01310

678
679
680

681

Parsed Citations

Allison, L.A., Simon, L.D., and Maliga, P. (1996). Deletion of rpoB reveals a second distinct transcription system in plastids of higher plants. EMBO J 15: 2802–2809.
Google Scholar: Author Only Title Only Author and Title

Barkan, A. (2011). Expression of plastid genes: organelle-specific elaborations on a prokaryotic scaffold. Plant Physiol 155: 1520–1532.
Google Scholar: Author Only Title Only Author and Title

Barkan, A. (2009). Genome-Wide Analysis of RNA-Protein Interactions in Plants. In Plant Systems Biology, D.A. Belostotsky, ed, Methods in Molecular BiologyTM. (Humana Press: Totowa, NJ), pp. 13–37.
Google Scholar: Author Only Title Only Author and Title

Bock, R. (2007). Structure, function, and inheritance of plastid genomes. In Cell and Molecular Biology of Plastids, R. Bock, ed, Topics in Current Genetics. (Springer: Berlin, Heidelberg), pp. 29–63.
Google Scholar: Author Only Title Only Author and Title

Börner, T., Aleynikova, A.Y., Zubo, Y.O., and Kusnetsov, V.V. (2015). Chloroplast RNA polymerases: Role in chloroplast biogenesis. Biochim Biophys Acta 1847: 761–769.
Google Scholar: Author Only Title Only Author and Title

Bowman, S.K., Simon, M.D., Deaton, A.M., Tolstorukov, M., Borowsky, M.L., and Kingston, R.E. (2013). Multiplexed Illumina sequencing libraries from picogram quantities of DNA. BMC Genomics 14: 466.
Google Scholar: Author Only Title Only Author and Title

Chi, W., He, B., Mao, J., Jiang, J., and Zhang, L. (2015). Plastid sigma factors: Their individual functions and regulation in transcription. Biochim Biophys Acta 1847: 770–778.
Google Scholar: Author Only Title Only Author and Title

Chotewutmontri, P. and Barkan, A. (2018). Multilevel effects of light on ribosome dynamics in chloroplasts program genome-wide and psbA-specific changes in translation. PLoS Genet 14: e1007555.
Google Scholar: Author Only Title Only Author and Title

Davis, S.E., Mooney, R.A., Kanin, E.I., Grass, J., Landick, R., and Ansari, A.Z. (2011). Mapping E. coli RNA polymerase and associated transcription factors and identifying promoters genome-wide. Methods Enzymol 498: 449–471.
Google Scholar: Author Only Title Only Author and Title

Deng, X.W., Stern, D.B., Tonkyn, J.C., and Gruissem, W. (1987). Plastid run-on transcription. Application to determine the transcriptional regulation of spinach plastid genes. J Biol Chem 262: 9641–9648.
Google Scholar: Author Only Title Only Author and Title

Ding, S., Zhang, Y., Hu, Z., Huang, X., Zhang, B., Lu, Q., Wen, X., Wang, Y., and Lu, C. (2019). mTERF5 Acts as a Transcriptional Pausing Factor to Positively Regulate Transcription of Chloroplast psbEFLJ. Mol Plant 12: 1259–1277.
Google Scholar: Author Only Title Only Author and Title

Favory, J.-J., Kobayashi, M., Tanaka, K., Peltier, G., Kreis, M., Valay, J.-G., and Lerbs-Mache, S. (2005). Specific function of a plastid sigma factor for ndhF gene transcription. Nucleic Acids Research 33: 5991–5999.
Google Scholar: Author Only Title Only Author and Title

Fey, V., Wagner, R., Braütigam, K., Wirtz, M., Hell, R., Dietzmann, A., Leister, D., Oelmüller, R., and Pfannschmidt, T. (2005). Retrograde Plastid Redox Signals in the Expression of Nuclear Genes for Chloroplast Proteins of Arabidopsis thaliana. Journal of Biological Chemistry 280: 5318–5328.
Google Scholar: Author Only Title Only Author and Title

Finster, S., Eggert, E., Zoschke, R., Weihe, A., and Schmitz-Linneweber, C. (2013). Light-dependent, plastome-wide association of the plastid-encoded RNA polymerase with chloroplast DNA. Plant J 76: 849–860.
Google Scholar: Author Only Title Only Author and Title

Galvão, R.M., Li, M., Kothadia, S.M., Haskel, J.D., Decker, P.V., Van Buskirk, E.K., and Chen, M. (2012). Photoactivated phytochromes interact with HEMERA and promote its accumulation to establish photomorphogenesis in Arabidopsis. Genes Dev. 26: 1851–1863.
Google Scholar: Author Only Title Only Author and Title

Hajdukiewicz, P.T.J., Allison, L.A., and Maliga, P. (1997). The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. EMBO J 16: 4041–4048.
Google Scholar: Author Only Title Only Author and Title

Hanaoka, M., Kanamaru, K., Takahashi, H., and Tanaka, K. (2003). Molecular genetic analysis of chloroplast gene promoters dependent on SIG2, a nucleus-encoded sigma factor for the plastid-encoded RNA polymerase, in Arabidopsis thaliana. Nucleic

Acids Res 31: 7090–7098.
Google Scholar: Author Only Title Only Author and Title

Hanaoka, M., Kato, M., Anma, M., and Tanaka, K. (2012). SIG1, a Sigma Factor for the Chloroplast RNA Polymerase, Differently Associates with Multiple DNA Regions in the Chloroplast Chromosomes in Vivo. International Journal of Molecular Sciences 13: 12182–12194.
Google Scholar: Author Only Title Only Author and Title

Hoffer, P.H. and Christopher, D.A. (1997). Structure and Blue-Light-Responsive Transcription of a Chloroplast psbD Promoter from Arabidopsis thaliana. Plant Physiology 115: 213–222.
Google Scholar: Author Only Title Only Author and Title

Hoffman, E.A., Frey, B.L., Smith, L.M., and Auble, D.T. (2015). Formaldehyde crosslinking: a tool for the study of chromatin complexes. J Biol Chem 290: 26404–26411.
Google Scholar: Author Only Title Only Author and Title

Inskeep, W.P. and Bloom, P.R. (1985). Extinction coefficients of chlorophyll a and B in n,n-dimethylformamide and 80% acetone. Plant Physiol 77: 483–485.
Google Scholar: Author Only Title Only Author and Title

Ishizaki, Y., Tsunoyama, Y., Hatano, K., Ando, K., Kato, K., Shinmyo, A., Kobori, M., Takeba, G., Nakahira, Y., and Shiina, T. (2005). A nuclear-encoded sigma factor, Arabidopsis SIG6, recognizes sigma-70 type chloroplast promoters and regulates early chloroplast development in cotyledons: AtSIG6 general sigma factor in chloroplasts. The Plant Journal 42: 133–144.
Google Scholar: Author Only Title Only Author and Title

Isono, K., Niwa, Y., Satoh, K., and Kobayashi, H. (1997). Evidence for transcriptional regulation of plastid photosynthesis genes in Arabidopsis thaliana roots. Plant Physiol 114: 623–630.
Google Scholar: Author Only Title Only Author and Title

Kanamaru, K., Nagashima, A., Fujiwara, M., Shimada, H., Shirano, Y., Nakabayashi, K., Shibata, D., Tanaka, K., and Takahashi, H. (2001). An Arabidopsis sigma factor (SIG2)-dependent expression of plastid-encoded tRNAs in chloroplasts. Plant Cell Physiol 42: 1034–1043.
Google Scholar: Author Only Title Only Author and Title

Krause, K., Maier, R.M., Kofer, W., Krupinska, K., and Herrmann, R.G. (2000). Disruption of plastid-encoded RNA polymerase genes in tobacco: expression of only a distinct set of genes is not based on selective transcription of the plastid chromosome. Mol Gen Genet 263: 1022–1030.
Google Scholar: Author Only Title Only Author and Title

Krupinska, K. and Apel, K. (1989). Light-induced transformation of etioplasts to chloroplasts of barley without transcriptional control of plastid gene expression. Mol Gen Genet 219: 467–473.
Google Scholar: Author Only Title Only Author and Title

Landick, R. (2006). The regulatory roles and mechanism of transcriptional pausing. Biochem Soc Trans 34: 1062–1066.
Google Scholar: Author Only Title Only Author and Title

Langmead, B. and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods 9: 357–359.
Google Scholar: Author Only Title Only Author and Title

Legen, J., Kemp, S., Krause, K., Profanter, B., Herrmann, R.G., and Maier, R.M. (2002). Comparative analysis of plastid transcription profiles of entire plastid chromosomes from tobacco attributed to wild-type and PEP-deficient transcription machineries. Plant J 31: 171–188.
Google Scholar: Author Only Title Only Author and Title

Lerbs-Mache, S. (2011). Function of plastid sigma factors in higher plants: regulation of gene expression or just preservation of constitutive transcription? Plant Mol Biol 76: 235–249.
Google Scholar: Author Only Title Only Author and Title

Liere, K., Kestermann, M., Müller, U., and Link, G. (1995). Identification and characterization of the Arabidopsis thaliana chloroplast DNA region containing the genes psbA, trnH and rps19. Curr Genet 28: 128–130.
Google Scholar: Author Only Title Only Author and Title

Lysenko, E.A. (2007). Plant sigma factors and their role in plastid transcription. Plant Cell Rep 26: 845–859.
Google Scholar: Author Only Title Only Author and Title

Melonek, J., Mulisch, M., Schmitz-Linneweber, C., Grabowski, E., Hensel, G., and Krupinska, K. (2010). Whirly1 in chloroplasts associates with intron containing RNAs and rarely co-localizes with nucleoids. Planta 232: 471–481.
Google Scholar: Author Only Title Only Author and Title

Nagashima, A., Hanaoka, M., Motohashi, R., Seki, M., Shinozaki, K., Kanamaru, K., Takahashi, H., and Tanaka, K. (2004). DNA

microarray analysis of plastid gene expression in an Arabidopsis mutant deficient in a plastid transcription factor sigma, SIG2. Biosci Biotechnol Biochem 68: 694–704.
Google Scholar: Author Only Title Only Author and Title

Nakatani, H.Y. and Barber, J. (1977). An improved method for isolating chloroplasts retaining their outer membranes. Biochim Biophys Acta 461: 500–512.
Google Scholar: Author Only Title Only Author and Title

Newell, C.A. and Gray, J.C. (2010). Binding of lac repressor-GFP fusion protein to lac operator sites inserted in the tobacco chloroplast genome examined by chromatin immunoprecipitation. Nucleic Acids Res 38: e145.
Google Scholar: Author Only Title Only Author and Title

Ortelt, J. and Link, G. (2021). Plastid Gene Transcription: An Update on Promoters and RNA Polymerases. Methods Mol Biol 2317: 49–76.
Google Scholar: Author Only Title Only Author and Title

Pfalz, J., Holtzegel, U., Barkan, A., Weisheit, W., Mittag, M., and Pfannschmidt, T. (2015). ZmpTAC12 binds single-stranded nucleic acids and is essential for accumulation of the plastid-encoded polymerase complex in maize. New Phytol 206: 1024–1037.
Google Scholar: Author Only Title Only Author and Title

Pfalz, J., Liere, K., Kandlbinder, A., Dietz, K.-J., and Oelmüller, R. (2006). pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. Plant Cell 18: 176–197.
Google Scholar: Author Only Title Only Author and Title

Pfalz, J. and Pfannschmidt, T. (2013). Essential nucleoid proteins in early chloroplast development. Trends in Plant Science 18: 186–194.
Google Scholar: Author Only Title Only Author and Title

Pfannschmidt, T., Blanvillain, R., Merendino, L., Courtois, F., Chevalier, F., Liebers, M., Grübler, B., Hommel, E., and Lerbs-Mache, S. (2015). Plastid RNA polymerases: orchestration of enzymes with different evolutionary origins controls chloroplast biogenesis during the plant life cycle. J. Exp. Bot. 66: 6957–6973.
Google Scholar: Author Only Title Only Author and Title

Privat, I., Hakimi, M.-A., Buhot, L., Favory, J.-J., and Mache-Lerbs, S. (2003). Characterization of Arabidopsis plastid sigma-like transcription factors SIG1, SIG2 and SIG3. Plant Mol Biol 51: 385–399.
Google Scholar: Author Only Title Only Author and Title

Puthiyaveetil, S., McKenzie, S.D., Kayanja, G.E., and Ibrahim, I.M. (2021). Transcription initiation as a control point in plastid gene expression. Biochim Biophys Acta Gene Regul Mech 1864: 194689.
Google Scholar: Author Only Title Only Author and Title

Quinlan, A.R. and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842.
Google Scholar: Author Only Title Only Author and Title

Roudier, F. et al. (2011). Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. EMBO J 30: 1928–1938.
Google Scholar: Author Only Title Only Author and Title

Rowley, M.J., Böhmdorfer, G., and Wierzbicki, A.T. (2013). Analysis of long non-coding RNAs produced by a specialized RNA polymerase in Arabidopsis thaliana. Methods 63: 160–169.
Google Scholar: Author Only Title Only Author and Title

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., and Tabata, S. (1999). Complete structure of the chloroplast genome of Arabidopsis thaliana. DNA Res 6: 283–290.
Google Scholar: Author Only Title Only Author and Title

Shi, C., Wang, S., Xia, E.-H., Jiang, J.-J., Zeng, F.-C., and Gao, L.-Z. (2016). Full transcription of the chloroplast genome in photosynthetic eukaryotes. Sci Rep 6: 30135.
Google Scholar: Author Only Title Only Author and Title

Shiina, T., Allison, L., and Maliga, P. (1998). rbcL Transcript levels in tobacco plastids are independent of light: reduced dark transcription rate is compensated by increased mRNA stability. Plant Cell 10: 1713–1722.
Google Scholar: Author Only Title Only Author and Title

Shimmura, S., Nozoe, M., and Shiina, T. (2008). Evolution of the Light Responsive psbD Promoter in Chloroplast. In Photosynthesis. Energy from the Sun, J.F. Allen, E. Gantt, J.H. Golbeck, and B. Osmond, eds (Springer Netherlands: Dordrecht), pp. 1193–1197.
Google Scholar: Author Only Title Only Author and Title

Sriraman, P., Silhavy, D., and Maliga, P. (1998). Transcription from Heterologous rRNA Operon Promoters in Chloroplasts Reveals Requirement for Specific Activating Factors1. Plant Physiology 117: 1495–1499.
Google Scholar: Author Only Title Only Author and Title

Stern, D.B., Goldschmidt-Clermont, M., and Hanson, M.R. (2010). Chloroplast RNA metabolism. Annu Rev Plant Biol 61: 125–155.
Google Scholar: Author Only Title Only Author and Title

Swiatecka-Hagenbruch, M., Liere, K., and Börner, T. (2007). High diversity of plastidial promoters in Arabidopsis thaliana. Mol Genet Genomics 277: 725–734.
Google Scholar: Author Only Title Only Author and Title

Thieffry, A., Vigh, M.L., Bornholdt, J., Ivanov, M., Brodersen, P., and Sandelin, A. (2020). Characterization of Arabidopsis thaliana Promoter Bidirectionality and Antisense RNAs by Inactivation of Nuclear RNA Decay Pathways. Plant Cell 32: 1845–1867.
Google Scholar: Author Only Title Only Author and Title

Tsunoyama, Y., Ishizaki, Y., Morikawa, K., Kobori, M., Nakahira, Y., Takeba, G., Toyoshima, Y., and Shiina, T. (2004). Blue light-induced transcription of plastid-encoded psbD gene is mediated by a nuclear-encoded transcription initiation factor, AtSig5. Proc Natl Acad Sci U S A 101: 3304–3309.
Google Scholar: Author Only Title Only Author and Title

Valkov, V.T., Scotti, N., Kahlau, S., Maclean, D., Grillo, S., Gray, J.C., Bock, R., and Cardi, T. (2009). Genome-wide analysis of plastid gene expression in potato leaf chloroplasts and tuber amyloplasts: transcriptional and posttranscriptional control. Plant Physiol 150: 2030–2044.
Google Scholar: Author Only Title Only Author and Title

Walker, D.M., Freddolino, P.L., and Harshey, R.M. (2020). A Well-Mixed E. coli Genome: Widespread Contacts Revealed by Tracking Mu Transposition. Cell 180: 703-716.e18.
Google Scholar: Author Only Title Only Author and Title

Woodson, J.D., Perez-Ruiz, J.M., Schmitz, R.J., Ecker, J.R., and Chory, J. (2013). Sigma factor-mediated plastid retrograde signals control nuclear gene expression. Plant J 73: 1–13.
Google Scholar: Author Only Title Only Author and Title

Yagi, Y., Ishizaki, Y., Nakahira, Y., Tozawa, Y., and Shiina, T. (2012). Eukaryotic-type plastid nucleoid protein pTAC3 is essential for transcription by the bacterial-type plastid RNA polymerase. Proc Natl Acad Sci U S A 109: 7541–7546.
Google Scholar: Author Only Title Only Author and Title

Zaidi, H., Hoffman, E.A., Shetty, S.J., Bekiranov, S., and Auble, D.T. (2017). Second-generation method for analysis of chromatin binding with formaldehyde-cross-linking kinetics. J Biol Chem 292: 19338–19355.
Google Scholar: Author Only Title Only Author and Title

Zghidi, W., Merendino, L., Cottet, A., Mache, R., and Lerbs-Mache, S. (2007). Nucleus-encoded plastid sigma factor SIG3 transcribes specifically the psb N gene in plastids. Nucleic Acids Research 35: 455–464.
Google Scholar: Author Only Title Only Author and Title

Zhelyazkova, P., Sharma, C.M., Förstner, K.U., Liere, K., Vogel, J., and Börner, T. (2012). The primary transcriptome of barley chloroplasts: numerous noncoding RNAs and the dominating role of the plastid-encoded RNA polymerase. Plant Cell 24: 123–136.
Google Scholar: Author Only Title Only Author and Title