Wierzbicki Andrzej (Orcid ID: 0000-0002-5713-1306)



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

V. Miguel Palomar[1], Sarah Jaksich[1], Sho Fujii[1, 2, 3], 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

[3]Department of Biology, Faculty of Agriculture and Life Science, Hirosaki University, Hirosaki, 036-8561, Japan

*corresponding author wierzbic@umich.edu, +1 734 647 6841

Short title: Binding pattern of plastid-encoded RNA polymerase

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/tpj.15882

## SUMMARY

Plastids contain 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 chloroplast 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 levels of RNA accumulation, which demonstrates the impact of PEP on chloroplast gene expression. Presented data are available through a publicly available Plastid Genome Visualization Tool (Plavisto) at https://plavisto.mcdb.lsa.umich.edu/.


## SIGNIFICANCE STATEMENT

Plastids contain their own transcriptional machinery, which is essential for photosynthesis. Genome-wide pattern of plastid-encoded RNA polymerase binding to DNA shows the importance of transcription for plastid gene expression.

## INTRODUCTION

Plastids have endosymbiotic origin and 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 thaliana* 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 in early 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 of 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 and are responsible for recruiting PEP to gene promoters (Chi *et al.*, 2015; Lysenko, 2007; Feklístov *et al.*, 2014). Six SIG isoforms have been identified in Arabidopsis. Although they have partially redundant functions, the 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), but 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 the recruitment of PEP to specific genes.

We established an improved method to study protein-DNA interactions in plastids and utilized 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. The data generated in this work can be accessed through Plastid Genome Visualization Tool (Plavisto) at https://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). We designed the ptChIP-seq protocol to maximize capture of protein-DNA interactions using 4% formaldehyde, unlike most applications in the nuclear genome (Davis *et al.*, 2011; Zaidi *et al.*, 2017). To demonstrate the specificity of ptChIP-seq, we first compared different crosslinking durations. For this purpose, we used 14-days-old plants expressing HA-tagged

pTAC12/HEMERA (Galvão *et al.*, 2012). pTAC12 is a PEP-associated factor (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 14-day-old Col-0 wild-type plants and a commercially available polyclonal antibody against the β subunit of PEP (RpoB; Fig. 1B, Fig. S1B, Fig. S1C). No enriched 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 the 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), and an expected band pattern on a western blot using the αRpoB antibody (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 at the same loci, as expected, and further supports a high specificity of ptChIP-seq.

A critical element of ChIP is a proper negative control. Genotypes that do not express the epitope, capture 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 less practical for αRpoB ptChIP-seq as the *rpoB* mutant is non-autotrophic (Allison *et al.*, 1996). Because of that, αRpoB ptChIP-seq signal may be compared to input samples instead, which is a common practice in studies of nuclear DNA-binding proteins (Flensburg *et al.*, 2014). The strong correlation between RpoB and pTAC12 ptChIP-seq experiments (Fig. 2A, Fig. S2A) indicates that input serves as a good negative control and 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 the 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. We then examined if PEP binding to DNA is affected in a mutant defective in SIG2, a sigma factor known to control the expression of 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 previously reported substantial reductions 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* and *psbA* were reduced to 0.50 and 0.41 of Col-0, which is consistent with previously reported small effects of *sig2* on the accumulation of their RNA products (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 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 a prior low resolution 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 different 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.

Next, we 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, 18% 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, and subsets of genes encoding ribosomal proteins or 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 interpreted as an indication that the transcriptional machinery may adopt four preferred levels of activity driven by the different strengths of various promoters.

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

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

Regression analysis of *sig2* compared to Col-0 wild-type revealed that ptChIP-seq signals in Col-0 wild-type and *sig2* are significantly correlated with a slope of 0.75 (Fig. 4B). This indicates that in *sig2*, most genes have a consistent, moderate reduction of PEP binding (Fig. 4B). Only a few genes had PEP binding reduced to much greater extents than indicated by the genome-wide trend. These included four previously studied tRNA genes (Fig. 2B, Fig. 4AD),

*trnL2* (Fig. 4B, Fig. S4A), and one protein-coding gene, *ndhC* (Fig. 4B, Fig. S4A). However, PEP binding was unchanged on *rbcL* (Fig. 4B, Fig. S4A).

A similar pattern was observed in *sig6* where PEP binding was strongly reduced throughout the genome. PEP binding in *sig6* and Col-0 wild-type remained significantly correlated, but the slope of the regression line was reduced to 0.45 (Fig. 4C). Only one gene had an almost complete loss of PEP binding, the tRNA gene, *trnI*, located in the inverted repeat (Fig. 4CE). Consistently, accumulation of the *trnI* tRNA product was significantly reduced in *sig6* (Fig. S4B). These results suggest that SIG2 and SIG6 have specific impacts on a limited number of genes and weaker but broad impacts on PEP occupancy throughout the genome.

**PEP preferentially binds to gene promoters**

Previous studies identified preferential binding of PEP to two promoters of photosystem II genes (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. The method we used (TSS RNA-seq) is similar to a previously established protocol in barley and relies on distinct 5' end phosphorylation status of primary and processed transcripts (Zhelyazkova *et al.*, 2012). The positions and relative quantities of 5' ends identified using TSS RNA-seq were consistent with previously reported Terminome-seq data (Castandet *et al.*, 2019) (Fig. S5ABC). Of note, we found that some of 5' ends identified by TSS RNA-seq did not perfectly match with those identified by primer extension or nuclease protection assays, probably due to methodological differences (Fig. S5C). We then compared RpoB ptChIP-seq signal with known PEP promoter locations and the positions of 5' ends of mRNAs. 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 are in close proximity to transcription start sites. We observed similar preferential PEP binding to other known promoters including *psaA* and *rbcL* (Fig. 5B, Fig. S5C). When averaged over all promoters that have been previously identified in Arabidopsis (Fig. S5CD), PEP binding and TSS RNA-seq signals were strongly enriched close to gene promoters (Fig. 5C). These results indicate that strong binding of PEP may be a general property of gene promoters.

**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 the steady-state levels of primary transcripts were significantly correlated (Fig. 6A). This suggests that differences in the steady-state levels of RNA may to some extent be explained by differences in PEP binding to DNA.

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), consistent with ribodepletion of RNA samples prior to the library prep (Thieffry *et al.*, 2020) 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). A $R^2$ of 0.36 indicates that about a third of RNA-seq variance may be predicted by the level of PEP binding to DNA. The remainder 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 has a significant impact on the steady-state levels of RNA and 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 the signal preserved by crosslinking is specific.

Our protocol can capture the pattern of PEP-DNA interactions in both enriched mature chloroplasts and intact seedlings with similar specificity, which is important as transcriptional regulation is particularly relevant 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). However, the dynamic range of ptChIP-seq is an order of magnitude higher than in a 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 and 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 inconsistent with a simplistic view of the model assuming full transcription of the plastid genome (Shi *et al.*, 2016).

The intensity of PEP binding to DNA throughout the plastid genome is not distributed normally and instead shows four preferred levels. These four levels could be attributed to specific cis-acting elements and regulatory proteins bound to those elements. They could also be reflected by preferred interactions between the core PEP complex and accessory proteins, or

even could be an indication that the genome exists in four favored structural states, which would be reminiscent of the situation in the nucleus (Roudier *et al.*, 2011).

A strong preferential binding of PEP within gene promoters, around transcription start sites, is consistent with the concept of RNA polymerase pausing on gene promoters, which is well established in bacteria (Landick, 2006; Landick, 2021). While it has been previously shown on the *psbEFLJ* operon (Ding *et al.*, 2019), our results indicate that pausing may be a general property of PEP. The role of PEP pausing in transcription and gene regulation remains unresolved.

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

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

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

tRNAs with strong and likely direct effects 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 PEP to gene regulation in plastids**

The significant correlation between PEP binding to DNA determined by ptChIP-seq and the steady-state levels of RNA determined by RNA-seq allows for 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 the reduction of PEP recruitment in sigma factor mutants is comparable to previously reported reductions of RNA levels.

Our observations about the contribution of PEP binding to DNA 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.

## EXPERIMENTAL PROCEDURES

**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 0.5X 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 crosslink DNA to proteins, a final concentration of 4% formaldehyde was applied to the amount of chloroplast corresponding to 200 µg of chlorophyll and 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**

The 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 subjected 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 αHA 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 crosslinked 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 a 20 µl reaction volume. The reaction was stopped by addition of 1 µl of 100mM EDTA. Subsequently, digested RNA was purified with acidic buffer-saturated phenol, washed with 70% ethanol, and resuspended in 10 µl of water. Purified RNA (~10 ng) 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.2.4.5 (Langmead and Salzberg, 2012). Read counts on defined genomic regions were determined using samtools v.1.15.1 and bedtools v.2.30.0 (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. TSS RNA-seq reads were mapped using the local alignment option to account for nucleotides added at 5' ends during the template-switching library prep.

**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; Shen, 2001; Demarsy *et al.*, 2012). Regions around the -35 (TTGACA) and -10 (TATAAT) consensus motifs were identified and are shown in Figure S5C.

**Immunoblot analysis**

Total proteins were extracted from 14-day-old plants 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, the indicated amount of proteins was separated by SDS-PAGE. For RpoB detection we used the same antibody as described for ptChIP-seq, and a secondary anti-rabbit IgG conjugated with horseradish-peroxidase (Thermo, catalog PI314), visualized using chemiluminescence reagents (ECL Prime Western Blotting Detection Reagent, Amersham) and blue films (Kodak).

**tRNA quantification**

Total RNA was isolated from 50 fresh 4-day-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 https://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).

## SHORT LEGENDS FOR SUPPORTING INFORMATION

**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.

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

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

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

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

## REFERENCES

**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.

**Barkan, A.** (2011) Expression of plastid genes: organelle-specific elaborations on a prokaryotic scaffold. *Plant Physiol*, **155**, 1520–1532.

**Barkan, A.** (2009) Genome-Wide Analysis of RNA-Protein Interactions in Plants. In D. A. Belostotsky, ed. *Plant Systems Biology*. Methods in Molecular Biology™. Totowa, NJ: Humana Press, pp. 13–37. Available at: https://doi.org/10.1007/978-1-60327-563-7_2 [Accessed October 28, 2021].

**Bock, R.** (2007) Structure, function, and inheritance of plastid genomes. In R. Bock, ed. *Cell and Molecular Biology of Plastids*. Topics in Current Genetics. Berlin, Heidelberg: Springer, pp. 29–63. Available at: https://doi.org/10.1007/4735_2007_0223 [Accessed November 29, 2021].

**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.

**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.

**Castandet, B., Germain, A., Hotto, A.M. and Stern, D.B.** (2019) Systematic sequencing of chloroplast transcript termini from Arabidopsis thaliana reveals >200 transcription initiation sites and the extensive imprints of RNA-binding proteins and secondary structures. *Nucleic Acids Research*, **47**, 11889–11905.

**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.

**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.

**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.

**Demarsy, E., Buhr, F., Lambert, E. and Lerbs-Mache, S.** (2012) Characterization of the plastid-specific germination and seedling establishment transcriptional programme. *Journal of Experimental Botany*, **63**, 925–939.

**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.

**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.

**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.

**Feklístov, A., Sharon, B.D., Darst, S.A. and Gross, C.A.** (2014) Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu Rev Microbiol*, **68**, 357–376.

**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.

**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.

**Flensburg, C., Kinkel, S.A., Keniry, A., Blewitt, M.E. and Oshlack, A.** (2014) A comparison of control samples for ChIP-seq of histone modifications. *Frontiers in Genetics*, **5**. Available at: https://www.frontiersin.org/article/10.3389/fgene.2014.00329 [Accessed March 25, 2022].

**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.

**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.

**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.

**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.

**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.

**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.

**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.

**Ishizaki, Y., Tsunoyama, Y., Hatano, K., et al.** (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.

**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.

**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.

**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.

**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.

**Landick, R.** (2006) The regulatory roles and mechanism of transcriptional pausing. *Biochem Soc Trans*, **34**, 1062–1066.

**Landick, R.** (2021) Transcriptional Pausing as a Mediator of Bacterial Gene Regulation. *Annu Rev Microbiol*, **75**, 291–314.

**Langmead, B. and Salzberg, S.L.** (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.

**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.

**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.

**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.

**Lysenko, E.A.** (2007) Plant sigma factors and their role in plastid transcription. *Plant Cell Rep*, **26**, 845–859.

**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.

**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.

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

**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.

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

**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.

**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.

**Pfalz, J. and Pfannschmidt, T.** (2013) Essential nucleoid proteins in early chloroplast development. *Trends in Plant Science*, **18**, 186–194.

**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.

**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.

**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.

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

**Roudier, F., Ahmed, I., Bérard, C., et al.** (2011) Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *EMBO J*, **30**, 1928–1938.

**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.

**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.

**Shen, Y.** (2001) RNA Binding-Proteins Interact Specifically with the Arabidopsis Chloroplast psbA mRNA 5' Untranslated Region in a Redox-Dependent Manner. *Plant and Cell Physiology*, **42**, 1071–1078.

**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.

**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.

**Shimmura, S., Nozoe, M. and Shiina, T.** (2008) Evolution of the Light Responsive psbD Promoter in Chloroplast. In J. F. Allen, E. Gantt, J. H. Golbeck, and B. Osmond, eds. *Photosynthesis. Energy from the Sun*. Dordrecht: Springer Netherlands, pp. 1193–1197. Available at: http://link.springer.com/10.1007/978-1-4020-6709-9_260 [Accessed December 14, 2021].

**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.

**Stern, D.B., Goldschmidt-Clermont, M. and Hanson, M.R.** (2010) Chloroplast RNA metabolism. *Annu Rev Plant Biol*, **61**, 125–155.

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

**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.

**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.

**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.

**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.

**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.

**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.

**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.

**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.

**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.

## FIGURE LEGENDS

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

A. Optimization of formaldehyde crosslinking duration 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 duration 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 into groups by the functions of their products into protein-coding, rRNA genes, or 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 and show averages from three independent replicates, error bars indicate standard deviations and the 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. The 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 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 indicate 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 function and show averages from three independent replicates. Error bars indicate standard deviations. Blue line represents the

linear regression model. Red line represents values that are 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 indicate 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 locations of transcription start sites on promoters previously shown to be bound by PEP.

B. Preferential binding of PEP and locations of 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 positions of the annotated promoters. 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 that were 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 replicate. Mean value from three biological replicates is shown. Light blue ribbon corresponds to standard deviation. Mean TSS RNA-seq signal was calculated from both strands in 10 bp genomic bins for each biological replicate. Brown dots correspond to mean values from three biological replicates. Error bars indicate standard deviations. Grey vertical line indicates aligned position of the annotated promoter.

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

A. RpoB ptChIP-seq enrichments and TSS RNA-seq signals are significantly 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 location. 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-day-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.

Figure 1

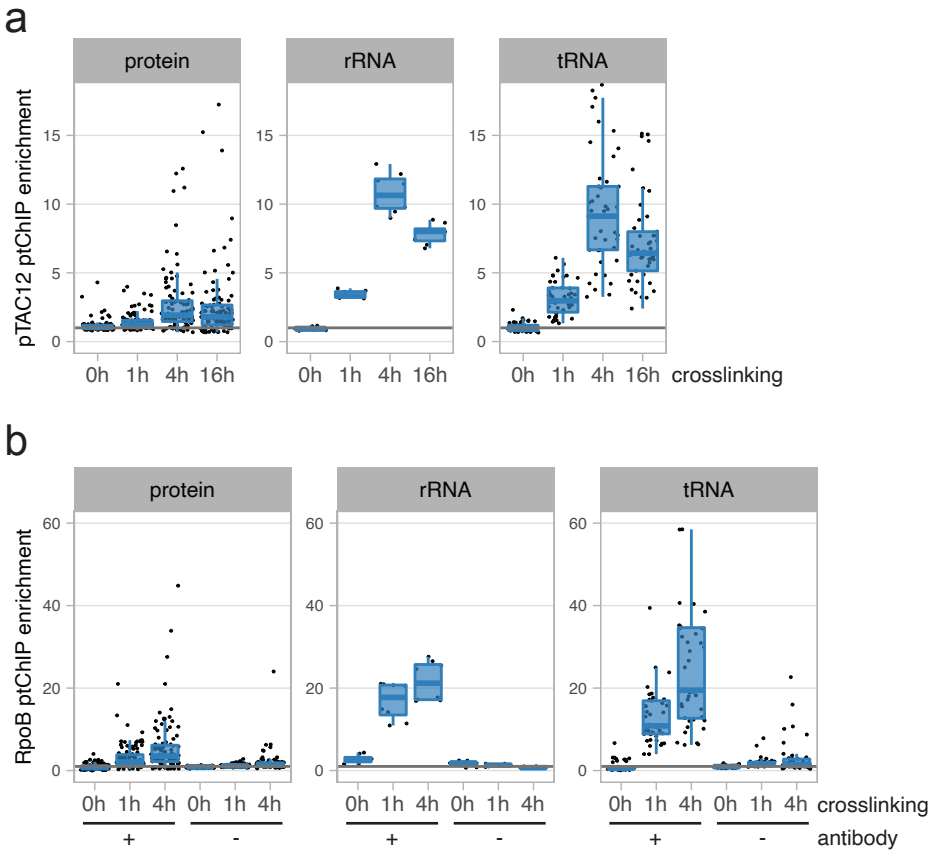

Figure 2

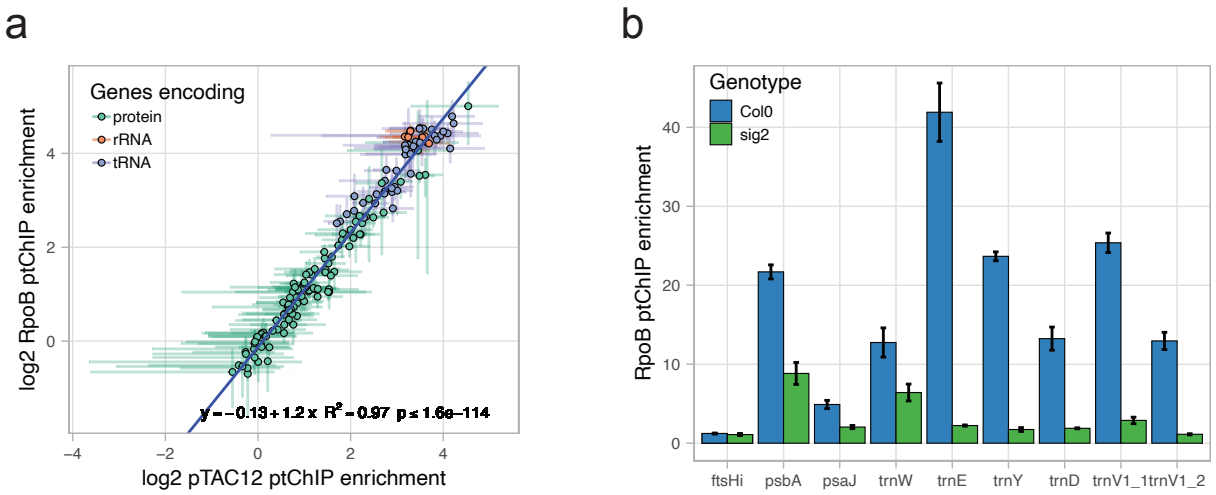

Figure 3

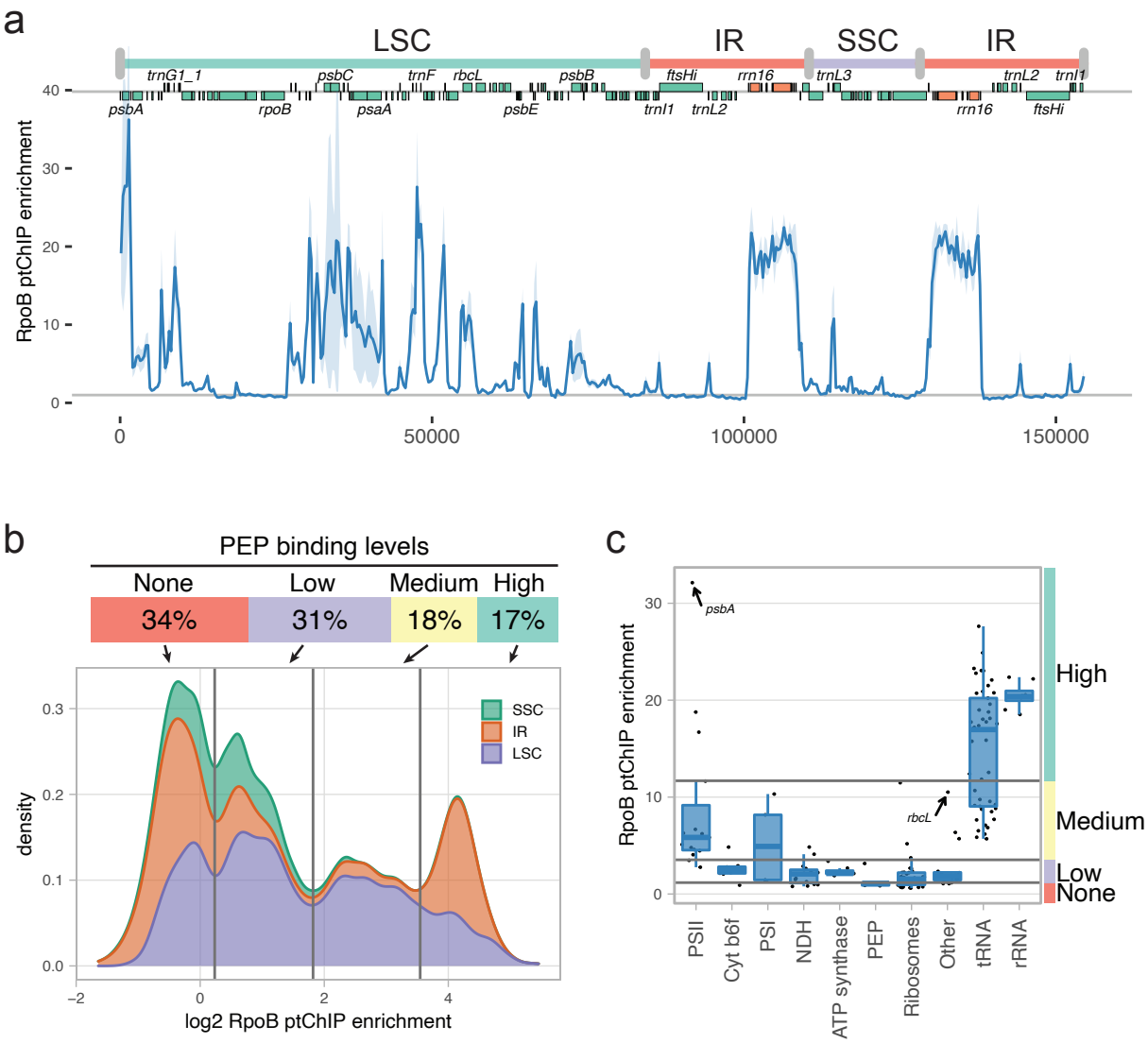

Figure 4

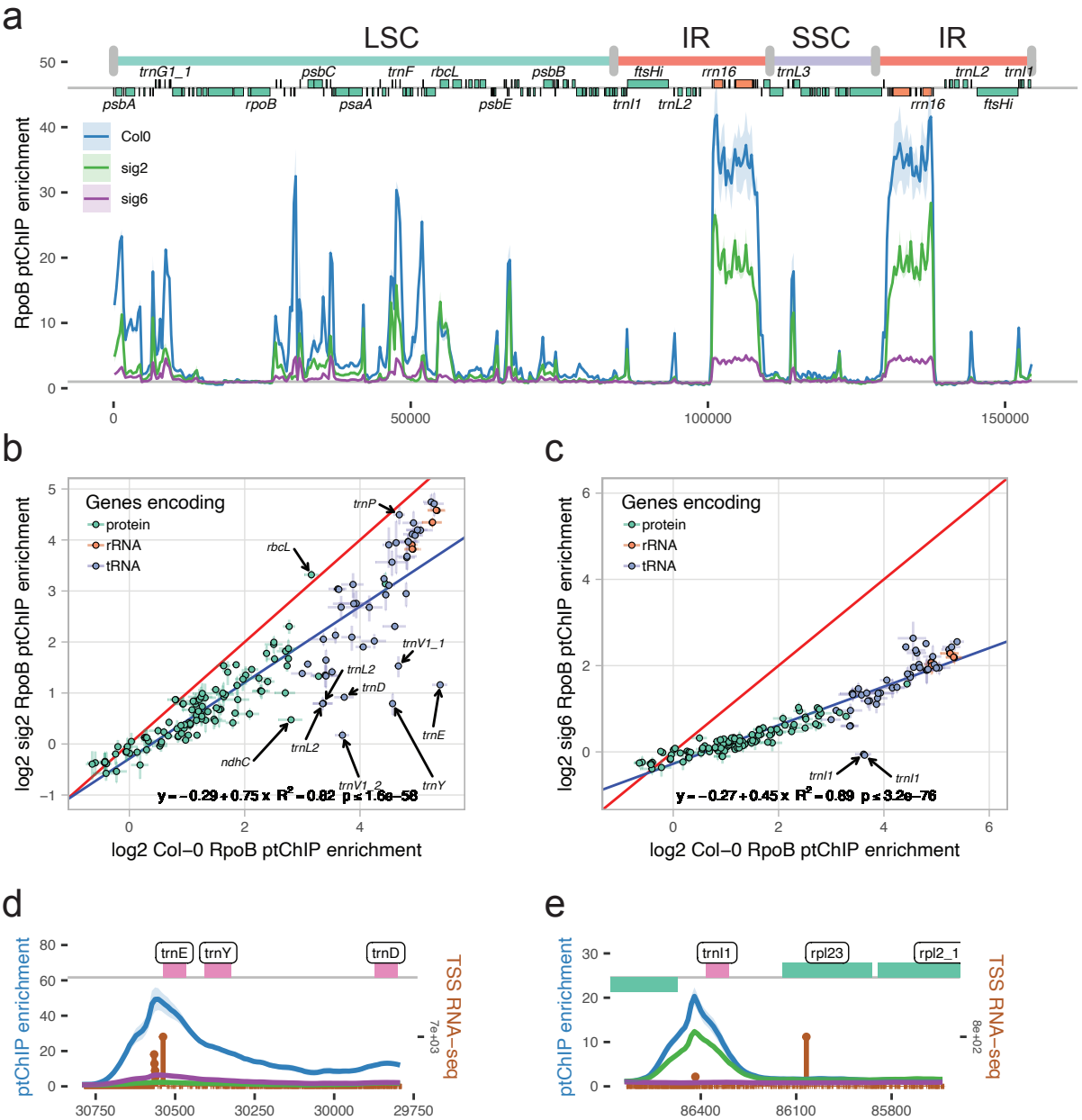

Figure 5

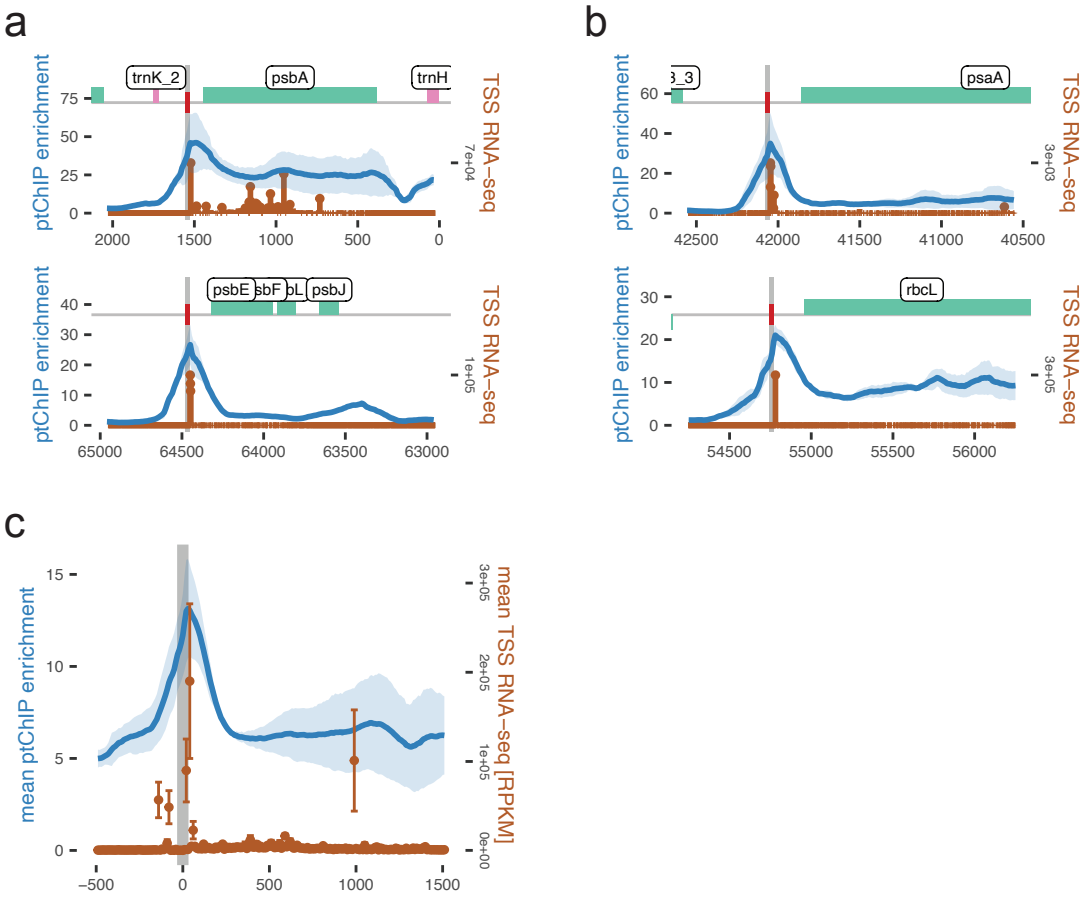

Figure 6