Transgene insertion into the plastid genome alters expression of adjacent native chloroplast genes at the transcriptional and translational levels

Rabea Ghandour[†] (iD), Yang Gao[†] (iD), Josephin Laskowski (iD), Rouhollah Barahimipour (iD), Stephanie Ruf (iD), Ralph Bock (iD) and Reimo Zoschke* (iD)

Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany

Summary

In plant biotechnology and basic research, chloroplasts have been used as chassis for the expression of various transgenes. However, potential unintended side effects of transgene insertion and high-level transgene expression on the expression of native chloroplast genes are often ignored and have not been studied comprehensively. Here, we examined expression of the chloroplast genome at both the transcriptional and translational levels in five transplastomic tobacco (*Nicotiana tabacum*) lines carrying the identical *aadA* resistance marker cassette in diverse genomic positions. Although none of the lines exhibits a pronounced visible phenotype, the analysis of three lines that contain the *aadA* insertion in different locations within the *petL-petG-psaJ-rpl33-rps18* transcription unit demonstrates that transcriptional read-through from the *aadA* resistance marker is unavoidable, and regularly causes overexpression of downstream sense-oriented chloroplast genes at the transcriptional and translational levels. Investigation of additional lines that harbour the *aadA* intergenically and outside of chloroplast transcription units revealed that expression of the resistance marker can also cause antisense effects by interference with transcription/transcript accumulation and/or translation of downstream antisense-oriented genes. In addition, we provide evidence for a previously suggested role of genomically encoded tRNAs in chloroplast transcription termination and/or transcript processing. Together, our data uncover principles of neighbouring effects of chloroplast transgenes and suggest general strategies for the choice of transgene insertion sites and expression elements to minimize unintended consequences of transgene expression on the transcription and translation of native chloroplast genes.

Introduction

Chloroplasts are endosymbiotic descendants of formerly free-living cyanobacteria (Sagan, 1967). During endosymbiont–host co-evolution, the genome of chloroplasts was dramatically reduced by gene transfer to the nuclear genome and loss of redundant genes (Bock and Timmis, 2008; Martin *et al.*, 2002). However, chloroplasts still contain a small genome that encodes approximately 120 genes, most of which are essential for photosynthesis or gene expression (Bock, 2007). Chloroplast genes are expressed by a gene expression machinery of prokaryotic origin (including, e.g. a bacteria-derived RNA polymerase and 70S-like ribosomes) that is complemented by factors and mechanisms of non-bacterial origin which were established during host–endosymbiont co-evolution (e.g. phage-type RNA polymerases, RNA splicing and RNA editing) (Barkan, 2011). Despite their strongly reduced genome size, chloroplasts harbour a gene expression system of remarkable capacity. This is exemplified by expression of the large subunit of Rubisco (RbcL), one of the most abundant proteins in the biosphere (Bar-On and Milo, 2019; Ellis, 1979), from a single-copy chloroplast gene. The enormous expression levels of *rbcL* and several other plastid-encoded photosynthesis-related genes are achieved by (i) the presence of numerous genome copies per chloroplast (Sakamoto

and Takami, 2018) in conjunction with multiple chloroplasts per cell, (ii) efficient transcription (Börner *et al.*, 2015), (iii) high mRNA abundances coupled with low mRNA turnover (Stern *et al.*, 2010) and (iv) efficient translation (Zoschke and Bock, 2018).

Protocols for transformation of the *Chlamydomonas* (*Chlamydomonas reinhardtii*) and tobacco (*Nicotiana tabacum*) chloroplast genomes were established more than 30 years ago (Boynton *et al.*, 1988; Svab *et al.*, 1990a), and have been refined and extended to numerous other species since then (Yu *et al.*, 2020). Expressing transgenes from the chloroplast genome offers several advantages. First, transgene insertion is achieved via homologous recombination and therefore enables targeting of transgenes to selected genomic regions (Svab *et al.*, 1990a). Second, transgenes can be expressed at extremely high levels from the chloroplast genome (Bock and Warzecha, 2010; Daniell, 2006). Third, expression of chloroplast transgenes is transgenerationally stable due to the lack of epigenetic gene-silencing mechanisms in plastids (Bock, 2015). Fourth, in most crop plants, plastids are maternally inherited, thereby virtually excluding chloroplast transgenes from pollen and leading to superior transgene containment (Ruf *et al.*, 2007; Svab and Maliga, 2007). Fifth, due to the bacterial origin of chloroplasts, transgenes that encode enzymes of metabolic pathways can be

co-expressed in operon-like arrangements, enabling efficient pathway engineering (Fuentes *et al.*, 2018). Finally, the large diversity of chloroplast metabolites offers flexible access to substrates and co-factors for metabolic engineering approaches (Fuentes *et al.*, 2018). Over the last decades, these characteristics greatly attracted biotechnologists to express transgenes in chloroplasts.

Numerous transgenes have been inserted into various regions of the chloroplast genome of the model plant tobacco, leading to successful transgene expression. The most commonly inserted transgene in chloroplast biotechnology and basic research is the *aadA* selection marker, which encodes an aminoglycoside-3″-adenylyl transferase that confers resistance against the antibiotics streptomycin and spectinomycin (Svab *et al.*, 1990b). The *aadA* marker is regularly linked to other transgenes of interest to select transplastomic lines on antibiotic-containing medium. The *aadA* gene has often been equipped with strong promoters and translation elements to confer high resistance levels to transplastomic plants. However, the effects of inserting *aadA* (or other transgenes) on the expression of native chloroplast genes have not been analysed comprehensively. It has been observed previously that in transplastomic plants with extremely high transgene expression levels, RbcL protein accumulation was reduced whereas *rbcL* transcript abundance was not affected, which led to the conclusion that high-level transgene expression causes exhaustion of the chloroplast translation machinery (Oey *et al.*, 2009). In some cases, effects of transgene insertion on editing or transcript accumulation of native chloroplast RNAs have been observed and were assigned to antisense transcript accumulation or transcriptional read-through (Hegeman *et al.*, 2005; Lutz *et al.*, 2007; Verhounig *et al.*, 2010; Zhou *et al.*, 2007, 2008). However, due to the biotechnological focus of these studies, alterations in chloroplast gene expression were not systematically analysed and possible effects at the level of translation were not studied at all.

Here, we applied a targeted ribosome and transcriptome profiling approach to globally investigate chloroplast gene expression in transplastomic plants that contain the same *aadA* cassette in a set of 'neutral' genomic locations. We show that, despite the absence of pronounced visible phenotypes in all analysed lines, the *aadA* insertion affected the expression of adjacent native chloroplast genes. Our results reveal that transcriptional read-through is almost unavoidable with currently used expression elements, reemphasizing the apparent lack of efficient transcription termination in chloroplasts. The systematic gene expression analysis of three different *aadA* insertion sites within the *petL-petG-psaJ-rpl33-rps18* transcription unit demonstrates that transcriptional read-through occurs regularly, is largely independent of the insertion site, leads to transcript overaccumulation from downstream sense-oriented genes and causes higher translation output of sense-oriented reading frames. Furthermore, our results indicate that native tRNAs that are located downstream of transgenes reduce transcriptional read-through. Finally, our data suggest that antisense transcription caused by read-through from the resistance marker leads to reduced transcript accumulation and translational activity of downstream antisense-oriented native chloroplast genes. Based on these data, we propose general guidelines for the choice of insertion sites and expression elements with the goal of designing transplastomic plants that efficiently express transgenes without affecting the expression of neighbouring native chloroplast genes.

Results

aadA insertion at different locations in the *petL-petG-psaJ-rpl33-rps18* transcription unit causes transcriptional and translational overexpression of downstream sense-oriented genes

In countless transplastomic studies, the *aadA* selection cassette was inserted in diverse regions of the chloroplast genome. However, possible effects of transgene insertion on the expression of adjacent endogenous plastid genes have only rarely been considered. To systematically analyse the effects of *aadA* expression on neighbouring chloroplast genes, specific insertion lines with wild-type-like growth behaviour were chosen for a chloroplast genome-wide analysis of translation and transcript accumulation (Figure 1a). As expected (based on the selection criterion for this collection), none of the transplastomic lines displayed a pronounced visible phenotype. Also, the major photosynthetic complexes accumulated normally under the chosen standard growth condition (Figure 1b,c; Figure S1). The only exception was PetB, a subunit of the cytochrome b_6f complex, which exhibited a significantly reduced protein accumulation in the pRB8c line (Figures 1 and S1), consistent with a previous report (Loiacono *et al.*, 2019). The molecular cause for reduced cytochrome b_6f levels in pRB8c will be described below. By choosing insertion lines with wild-type-like growth behaviour, we aimed to reveal direct impacts of transgene insertion on chloroplast gene expression and largely exclude secondary effects. The latter may originate, for example, from disturbed photosynthesis and/or chloroplast gene expression, which can trigger a plethora of regulatory responses in plastid transcription, post-transcriptional RNA processing and translation. It should be noted that all below-described transplastomic lines harbour the identical *aadA* cassette driven by the same expression elements, namely the *Nicotiana tabacum* (tobacco) *rrn16* promoter, a synthetic *rbcL*-derived Shine-Dalgarno sequence (Svab and Maliga, 1993) and the *psbA* 3′ UTR.

In the course of our search for suitable and available mutants, the *petL-petG-psaJ-rpl33-rps18* transcription unit caught our attention. This transcription unit contains the three non-essential genes *petL*, *psaJ* and *rpl33* whose expression is dispensable for autotrophic growth under standard conditions as evidenced by the wild-type-like phenotypes of the respective knockout mutants generated by *aadA* insertion (Rogalski *et al.*, 2008; Schöttler *et al.*, 2007a,b; Schwenkert *et al.*, 2007; Figure 1a; Appendix S1). These mutants enabled us to examine the primary effects of *aadA* insertion at different sites within a specific chloroplast transcription unit.

The *petL*, *psaJ* and *rpl33* mutants were grown along with wild-type plants under fully controlled conditions to the four-true-leaf stage (Figure 1a) and aerial parts of the plants were used for microarray-based ribosome and transcriptome profiling as described previously (Trösch *et al.*, 2018). In the selected early developmental stage, the biogenesis of photosynthetic complexes is in full operation and, therefore, plastid gene expression is expected to be limiting for chloroplast biogenesis. Consequently, insertion and expression of the *aadA* transgene can be expected to have the strongest direct effects on the expression of neighbouring genes at this growth stage. Similarly, potential consequences on the expression of other, distantly located, chloroplast genes may more likely be observed in early developmental stages (caused, e.g., by exhaustion of
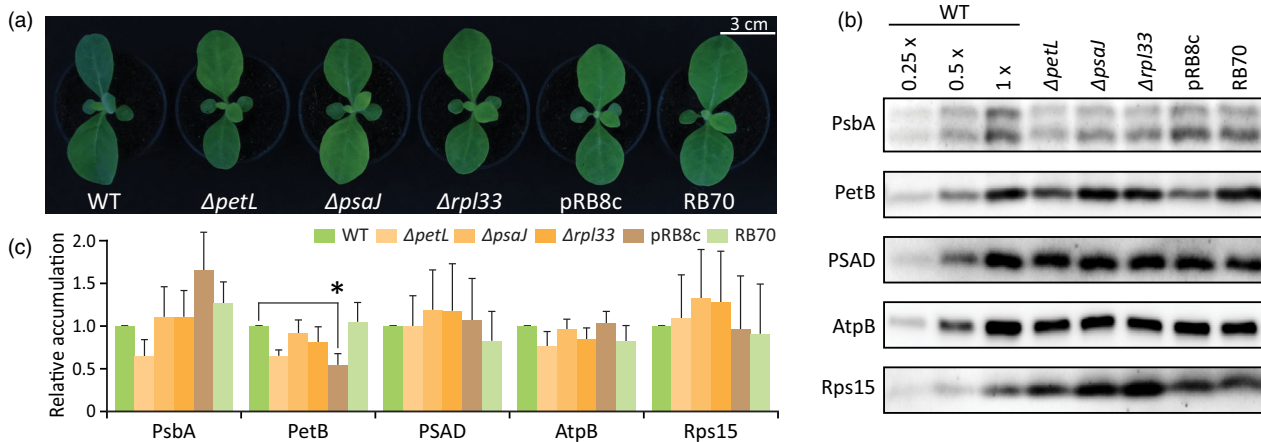
Figure 1 Examination of visible phenotypes and the accumulation of major chloroplast protein complexes in the analysed transplastomic lines. (a) Wild-type-like phenotypes of the analysed transplastomic lines. Transplastomic and wild-type control plants (denominations given below each plant) were grown in standard growth conditions (see Experimental procedures) to the four-true-leaf stage at which the aerial parts were harvested for ribosome and transcriptome profiling. Note that a previously unnoticed (Schöttler et al., 2007b), very mild pale green phenotype was repeatedly observed for *ΔpsaJ* mutants in early developmental stages (disappearing in later growth stages), whereas all other transplastomic lines were completely indistinguishable from the wild type. (b) Accumulation of proteins of the photosynthetic machinery and the ribosome in the transplastomic lines. Immunoblot analyses of PsbA (photosystem II), PetB (cytochrome b_6f complex), PSAD (photosystem I), AtpB (ATP synthase) and Rps15 (small subunit of the chloroplast ribosome) protein accumulation in transplastomic and wild-type plants. The experiment was performed with three independent biological replicates and a representative examination is shown (all replicate results are provided in Figure S1). A dilution series of the wild-type sample is loaded on the left (dilutions as indicated on top of the blots). (c) Quantification of immunoblot signal intensities shown in (b) and Figure S1. The band intensities were quantified for three independent biological replicates using Image Lab software (Bio-Rad) and normalized to the intensity of the undiluted wild-type sample. Error bars indicate the standard deviation of results obtained from three independent biological replicates. Asterisk marks a statistically significant difference ($P < 0.01$, two-sided Student's t-test).

the gene expression capacity or depletion of specific *trans*-factors by the duplicated *cis*-elements used to drive *aadA* expression).

Transcript accumulation and translation output were assessed chloroplast genome-wide by microarray-based ribosome and transcriptome profiling (Trösch et al., 2018) in the three mutant lines *ΔpetL*, *ΔpsaJ* and *Δrpl33* in comparison to wild-type plants that were grown side-by-side (Figures 1–3; Data S1). In brief, cell lysates were ribonuclease-treated to degrade mRNA stretches that are not covered by translating ribosomes, thereby disintegrating polysomes to monosomes, which were subsequently purified through a sucrose cushion. The mRNA footprints of these monosomes (ribosome footprints) were extracted, size-selected, labelled with fluorescent dyes (Cy3 and Cy5 for mutant and wild-type samples, respectively) and hybridized to custom microarrays which contain overlapping oligonucleotide probes that cover all chloroplast reading frames (Trösch et al., 2018). In addition, total RNA was extracted from the very same cell lysates, fragmented, similarly labelled with fluorescent dyes and hybridized to duplicate microarrays (for details see Experimental procedures). For all of the analysed lines, the examined three biological replicates showed very high reproducibility (average Pearson's *r* values >0.95 for ribosome footprint and mRNA data; Figure S2; Data S1). After normalization, we determined average ribosome footprint abundance (as proxy of the translation output) and mRNA abundance (i.e. transcript accumulation) for each chloroplast reading frame before calculating relative changes between mutant and wild-type plants. Footprint abundances in a given reading frame reflect translation output that provides a proxy for protein synthesis levels because each elongating ribosome leaves exactly one footprint.

The *ΔpetL* mutant was created via gene disruption by inserting the *aadA* cassette downstream of the *petL* start codon (Fiebig et al., 2004; Schöttler et al., 2007a). The knockout of the small PetL subunit is known to cause mildly reduced accumulation of the cytochrome b_6f complex (Schöttler et al., 2007a), a consequence that was also visible in our protein gel blot analyses by an approximately 35% reduced PetB accumulation (Figure 1b,c, Figure S1). In the *ΔpetL* mutant, a large part of the *petL* reading frame is located downstream of the *aadA* cassette (Figure 2a; Appendix S1), and *petL* translation is expected to terminate at the first in-frame stop codon within the *aadA* cassette. Hence, except for the first five codons located upstream of the *aadA* insertion site, *petL* translation should be knocked out (Figure 2b). In line with this assumption, our ribosome profiling data revealed an on average more than threefold reduced ribosome coverage of the *petL* reading frame (Figure 3a). By contrast, due to the inherent leakiness of transcription termination in chloroplasts (e.g. Sharwood et al., 2011a; Stern and Gruissem, 1987), transcription that initiates at the strong 16S rRNA promoter of the *aadA* resistance marker continues downstream of the transgene insertion site and causes higher transcript accumulation of large parts of *petL* (Figure 2b). This leads to a more than eightfold overaccumulation of the *petL* transcript (Figure 3a). Similarly, the adjacent *petG* gene that is located downstream of both the *aadA* cassette and *petL* is significantly transcriptionally overexpressed (more than fivefold; Figure 2b, Figure 3a). By contrast, the transcripts of the *psaJ*, *rpl33* and *rps18* genes that are encoded further downstream (and separated from *petG* by two tRNAs encoded on the opposite DNA strand) only mildly overaccumulate (Figure 3a). Nevertheless, the transcript overaccumulation of *petG*, *psaJ*, *rpl33* and *rps18* was accompanied by an increase in translation output

for these reading frames (~4-fold, ~1.3-fold, ~3.5-fold and 2.5-fold, respectively, and significant for *petG*, *rpl33* and *rps18*; Figure 3a). In addition, the *petN* reading frame, which is distantly located in the genome, displayed a significantly (more than twofold) reduced coverage with ribosomes in the *ΔpetL* mutant (Figure 3).

Similar to *petL*, *psaJ* expression was knocked out by inserting the *aadA* cassette in sense orientation close to the start codon into the *psaJ* reading frame (Schöttler *et al.*, 2007b). This leads to premature *psaJ* translation termination within the *aadA* cassette and the majority of the *psaJ* reading frame remains untranslated (Figure 2a; Appendix S1). In fact, downstream of the *psaJ* start codon, only two *psaJ* codons are translated before ribosomes enter the *aadA* cassette. Hence, *psaJ* translation should be virtually knocked out (Figure 2c), an assumption that was confirmed by the approximately eightfold lower coverage of the *psaJ* reading frame with ribosomes (Figure 3b). By contrast, at the transcript level, large parts of the *psaJ* reading frame overaccumulate due to transcriptional read-through from the inserted *aadA* transgene (Figure 2c). This leads to a significant (almost threefold) increase in *psaJ* transcript levels (Figure 3b). Likewise, the downstream genes *rpl33* and *rps18* are significantly overexpressed at the RNA level (about fourfold and almost fivefold, respectively), which is accompanied by a significantly increased translation output of the *rpl33* and *rps18* reading frames (more than sixfold and almost fivefold, respectively). Apart from these direct effects of *aadA* insertion, transcript accumulation and translation output of the distantly located *rbcL* gene is significantly and more than twofold reduced in the *ΔpsaJ* mutant (Figure 3c), which, however, did not cause a substantial alteration in RbcL protein accumulation (Figure S1).

In the *Δrpl33* mutant, the *aadA* insertion deletes a large portion of the *rpl33* reading frame beginning directly after the start codon (Rogalski *et al.*, 2008) (Figure 2a; Appendix S1). Accordingly, probes in the deleted region show abolished transcript accumulation, whereas increased transcript levels were detected downstream of the removed area, again reflecting transcriptional read-through from *aadA* (Figure 2d). This leads to a more than twofold (but not significant) increase in *rpl33* transcript accumulation (Figure 3c). By contrast, *rpl33* translation output is significantly and almost fivefold decreased because translation is knocked out within the deleted region and, due to premature termination, also downstream of the *aadA* insertion (Figure 3c). In addition, the transcript level of *rps18* is more than sevenfold increased (Figure 2d, Figure 3c) due to transcriptional read-through. The *rps18* transcript overaccumulation, in turn, causes an almost fivefold elevated *rps18* translation output (Figure 3c). Furthermore, the *Δrpl33* mutant displays significantly (more than twofold) increased *rps15* transcript levels and a significantly (more than threefold) reduced *rbcL* translation output (Figure 3c), which, however, did not cause a substantially altered accumulation of these proteins (Figures 1 and S1). Both genes are located in genomic positions that are distant from *rpl33*.

The three knockout mutants of the non-essential genes *petL*, *psaJ* and *rpl33* have the common effect of *rps18* overexpression (Figure 3). Hence, one may wonder, if the described mild molecular defects could be explained by *rps18* overexpression rather than *petL*, *psaJ* or *rpl33* knockout. However, this can almost certainly be ruled out because the previous detailed characterization of defects in these mutants suggests that the observed molecular deficiencies in *ΔpetL*, *ΔpsaJ* and *Δrpl33* are

rather distinct with clear links to the respective complexes the knocked-out proteins reside in, namely the cytochrome b_6f complex (PetL), photosystem I (PsaJ) and the chloroplast ribosome (Rpl33) (Fiebig *et al.*, 2004; Rogalski *et al.*, 2008; Schöttler *et al.*, 2007a,b; Schwenkert *et al.*, 2007). In addition, our analysis of the accumulation of the chloroplast small ribosomal subunit protein S15 (Figures 1b,c and S1) and the examination of the abundance of the mature 16S rRNA (Figure 4), both of which can serve as a proxy for the accumulation of the chloroplast 30S ribosomal subunit, did not reveal a significant effect of *rps18* overexpression on chloroplast ribosome levels in the three mutant lines. However, interestingly the transplastomic *ΔpetL*, *ΔpsaJ* and *Δrpl33* lines displayed a significant overaccumulation of the most abundant 16S rRNA precursor transcript (Figure 4).

Altogether, our results demonstrate that *aadA* insertions in different positions of a chloroplast transcription unit consistently cause transcript overaccumulation and increased translation output of downstream sense-oriented genes. These effects are largely attributable to transcriptional read-through, and should be taken into consideration when interpreting the effects of targeted chloroplast gene knockouts by transgene insertion.

aadA insertion in intergenic regions can mildly affect the expression of downstream antisense genes at the translational and transcript levels

Insertion of the *aadA* marker into chloroplast reading frames is usually done in fundamental research to study gene functions by reverse genetics. It can be expected to cause alterations in the expression of adjacent native genes, especially within the same (polycistronic) transcription unit. For the expression of chloroplast transgenes in biotechnology, typically intergenic insertion sites are chosen. They are located outside of transcription units and are considered neutral in that they are expected to not appreciably affect the expression of endogenous plastid genes. Two such neutral insertion sites were initially tested in the pRB8c and RB70 transplastomic tobacco lines (Bock *et al.*, 1994; Krech *et al.*, 2013; Loiacono *et al.*, 2019; Ruf *et al.*, 2001). The RB70 insertion site has subsequently been commonly used in both biotechnology and basic research (e.g. Fuentes *et al.*, 2016; Krech *et al.*, 2013; Qu *et al.*, 2018; Ruf *et al.*, 2001, 2019; Stanbekova *et al.*, 2021; Wurbs *et al.*, 2007), including studies that yield exceptionally high transgene expression levels (Oey *et al.*, 2009).

The pRB8c transplastome contains the *aadA* marker gene in the intergenic spacer between the *psbE-psbF-psbL-psbJ* (sense) transcription unit and the *psaI-ycf4-ycf10-petA* (antisense) transcription unit (Figure 5a; Appendix S1). Despite the wild-type-like growth behaviour of pRB8c plants at standard growth conditions (Bock *et al.*, 1994), this transplastomic line has been previously shown to display a substantial reduction in cytochrome b_6f complex content (Loiacono *et al.*, 2019), a finding that was confirmed by our protein gel blot analysis (Figures 1b,c and S1). Interestingly, our ribosome and transcriptome profiling data show that the *aadA* insertion causes mildly reduced transcript accumulation (1.4-fold) and ribosome coverage (1.9-fold) for *petA* (Figures 5b and 6), the gene encoding the essential cytochrome *f* subunit of the cytochrome b_6f complex. The *petA* gene is located in antisense direction next to the *aadA* insertion site. The detected almost twofold reduction in *petA* translation output (Figure 6) can fully account for the decrease in cytochrome b_6f complex accumulation of the pRB8c line observed here (Figures 1b,c and S1) and in a previous study (Loiacono *et al.*, 2019),
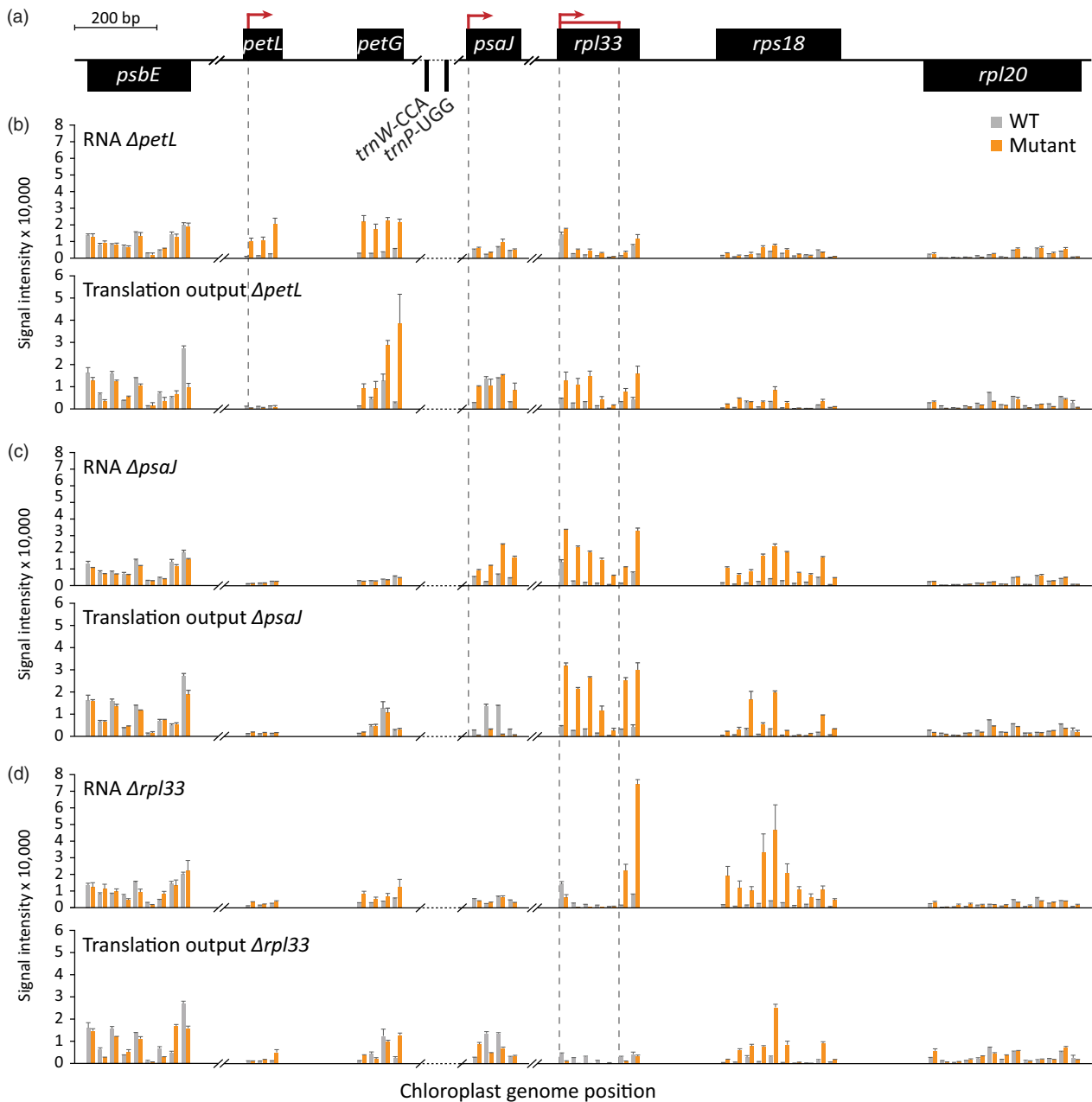
Figure 2 Altered transcript levels and ribosome footprint abundances in protein-coding regions downstream of the *aadA* insertion sites in the *ΔpetL*, *ΔpsaJ* and *Δrpl33* mutants. (a) Gene map illustrating the insertion site positions and transcription direction of the *aadA* resistance marker cassette in *ΔpetL*, *ΔpsaJ* and *Δrpl33* mutants (from left to right; note that each of the mutants contains only one *aadA* insertion within the respective reading frame). Black boxes indicate protein-coding regions (gene names are given within the boxes). Thin black boxes represent tRNA genes (with names and anticodons given). Red lines and horizontal red arrows depict *aadA* cassette insertion sites and indicate the direction of transcription originating from the promoter of the *aadA* cassette, respectively. Note that in *Δrpl33* a large part of the *rpl33* reading frame that is located between the two vertical red lines was replaced by the *aadA* insertion. Genes above and below the thin horizontal line are transcribed from left to right and from right to left, respectively. For clearer visualization, large regions between protein-coding genes were removed (depicted by broken horizontal lines in the gene map). Sequences flanking the *aadA* insertion sites are provided in Appendix S1. (b-d) Bar plots representing the normalized relative signal intensities of transcript accumulation (RNA; top diagrams) and ribosome footprints (translation output; bottom diagrams) for probes in protein-coding regions that are located upstream and downstream of the *aadA* insertion in the (b) *ΔpetL*, (c) *ΔpsaJ* and (d) *Δrpl33* mutants in comparison to wild-type plants. Signal intensities for wild-type and mutant plants are represented by grey and orange bars, respectively. Plots are scaled with the gene map shown in (a). Grey dashed vertical lines connect the *aadA* insertion sites shown in the map with the respective position in the bar plot. Error bars denote standard deviations of three biological replicates. Abscissae are broken at the same positions as in the gene map in (a).

and may be explained by antisense transcript effects that originate from *aadA* read-through transcription. In addition, the translation output of the *petL* transcript was significantly and

more than twofold up-regulated (Figure 6). *petL* encodes a small cytochrome b_6f complex subunit whose knockout does not affect tobacco cytochrome b_6f complex accumulation in young plants
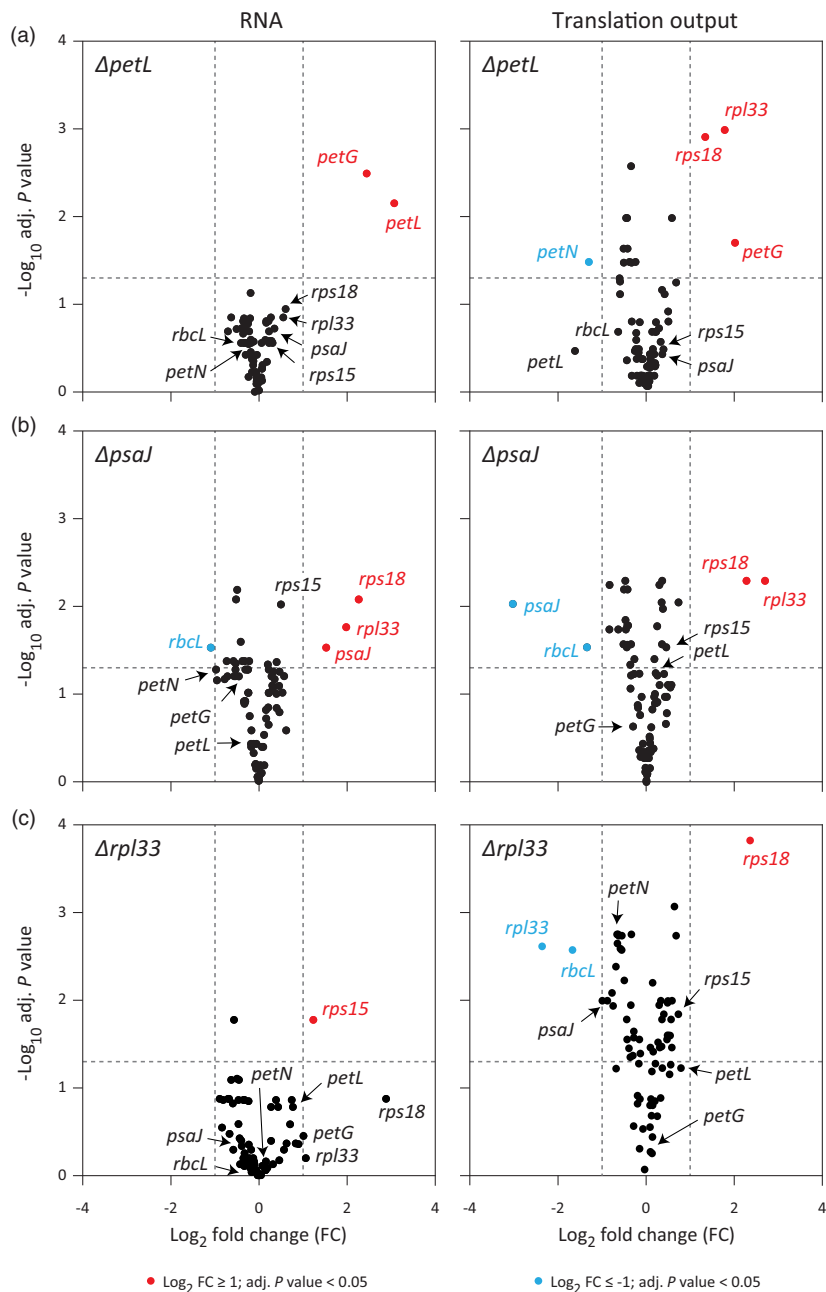
Figure 3 Chloroplast genome-wide analysis of gene expression effects caused by *aadA* insertion in the *ΔpetL*, *ΔpsaJ* and *Δrpl33* mutants. Volcano plots representing comparative analyses of relative changes in transcript accumulation (left) and translation output (right) of chloroplast protein-coding genes in (a) *ΔpetL*, (b) *ΔpsaJ* and (c) *Δrpl33* mutants compared to wild-type plants. The negative decimal logarithm of the adjusted *P* values is plotted against the log$_2$ fold-change ratios of the average relative transcript and ribosome footprint abundances per reading frame (data are based on three biological replicates). The vertical and horizontal grey dashed lines represent the cut-offs for significant differential expression: genes with log$_2$ fold change ≥1 and ≤−1 and adjusted *P* value <0.05 were considered as significantly differentially expressed and are labelled with red and blue dots, respectively, and equally coloured denominations. Chloroplast genes that are not significantly differentially expressed are represented by black dots (genes that flank the *aadA* insertion site or show a more than twofold change in one of the three mutants are represented by their name in black). Values for all genes are provided in Data S1.

under standard growth conditions (Gao *et al.*, 2022; Schöttler *et al.*, 2007a; Schwenkert *et al.*, 2007). Hence, altered *petL* expression is unlikely to explain reduced cytochrome b_6f complex accumulation in pRB8c (leaving aside that *petL* translation output is increased, whereas cytochrome b_6f complex accumulation is decreased). Since the *petL* gene is genomically distantly located to the *aadA* insertion site, its elevated translation output cannot be a direct consequence of *aadA* expression. Beside these effects, our data show more than twofold reduced levels of *ycf10* translation output and almost twofold reduced *ycf10* transcript
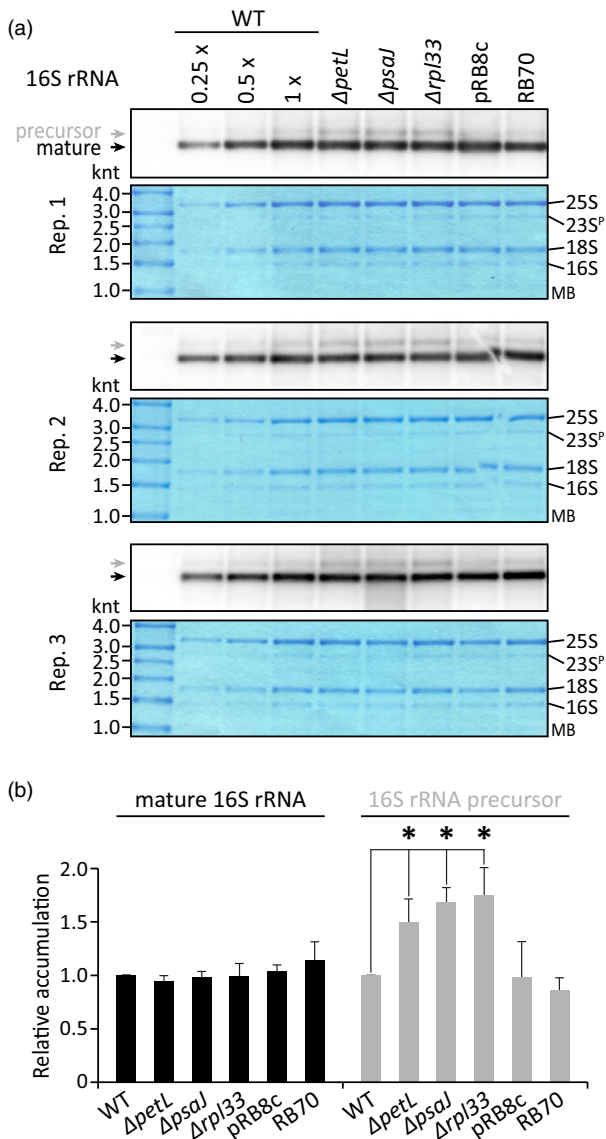
Figure 4 Accumulation of plastid 16S rRNA and its major unprocessed precursor in the transplastomic lines. (a) RNA gel blot analysis of 16S rRNA accumulation in the transplastomic lines. The experiment was performed with three independent biological replicates (Rep.; obtained from three individual plants). The major band represents the mature 16S rRNA (black arrow). In addition, a larger unprocessed 16S rRNA precursor is detectable with the applied PCR probe (grey arrow). A dilution series is shown for the wild-type plants on the left. Pictures of methylene blue staining (MB) of all rRNA species are shown as control for equal loading below each 16S rRNA detection image (23S^P, chloroplast 23S rRNA precursor). Sizes of marker bands are indicated on the left side in kilonucleotides (knt). (b) Quantification of 16S rRNA (left) and its precursor (right) accumulation. The band intensities were quantified using Image Lab software (Bio-Rad) and normalized to the intensity of the undiluted wild-type samples. Error bars indicate the standard deviation of results obtained from three independent biological replicates. Asterisks indicate statistically significant differences ($P < 0.01$, two-sided Student's t test).

accumulation levels in pRB8c (Figures 5b and 6). In tobacco, the non-nonessential *ycf10* gene is located upstream of *petA* (Rolland *et al*., 1997; Trinh *et al*., 2021), and its reduced expression is likely

also attributable to read-through antisense transcription that originates from the *aadA* cassette. Similar but overall milder effects were observed for the *ycf4* gene that is located upstream of *ycf10* (Figures 5b and 6). By contrast, the expression of *psaI*, which is the most upstream-located gene in the *psaI-ycf4-ycf10-petA* gene cluster, is virtually unaffected (Figures 5b and 6).

In the RB70 line, the *aadA* cassette was inserted between the two tRNA-encoding genes *trnfM* (sense) and *trnG* (antisense) (Figure 7a; Appendix S1). Potential consequences of *aadA* insertion in the transplastomic RB70 line are of particular interest because the same insertion site (with identical direction of transgene expression) has been frequently used for diverse applications in basic research and chloroplast biotechnology (e.g. Fuentes *et al*., 2016; Krech *et al*., 2013; Qu *et al*., 2018; Ruf *et al*., 2001, 2019; Stanbekova *et al*., 2021; Wurbs *et al*., 2007). Our chloroplast ribosome profiling and transcriptome analyses revealed that not a single chloroplast gene displays significantly (*P* value <0.05) or substantially (log$_2$ fold change >1 or <−1) altered expression in RB70 plants at either the level of transcript accumulation or ribosome coverage (Figure 8). These findings validate that the *aadA* insertion site in RB70 can be considered as neutral or at least nearly neutral.

Together, our results demonstrate that, depending on the genomic region, intergenic *aadA* insertion can have mild effects on transcript accumulation and/or translation of adjacent antisense-oriented endogenous chloroplast genes that are located downstream of the *aadA* insertion site.

Discussion

Previous work in yeast has documented that even targeted manipulations of genomic loci by homologous recombination (e.g. directed knockout of genes by insertion of a transgene) can have pronounced consequences for transcription and translation of adjacent genes (e.g. Atias *et al*., 2016; Egorov *et al*., 2021), a phenomenon also referred to as neighbouring-gene effect (Baryshnikova and Andrews, 2012; Ben-Shitrit *et al*., 2012). In both eukaryotes and bacteria, clustered genes can co-evolve to fine-tune their expression levels, possibly suggesting that natural neighbouring-gene effects are widespread (e.g. Ghanbarian and Hurst, 2015; Steinrueck and Guet, 2017). However, possible effects of transgene insertion and expression on neighbouring genes in plastid genomes have not been comprehensively studied. In this work, we systematically analysed chloroplast genome expression in transplastomic lines that contain identical *aadA* resistance marker cassettes in different genomic locations (Figures 2, 5 and 7; Appendix S1). Our results substantiate the previous observation that transcriptional read-through from the *aadA* marker gene cassette is probably inevitable (e.g. Lutz *et al*., 2007), which is certainly due to the naturally inefficient termination of chloroplast transcription (e.g. Sharwood *et al*., 2011a; Stern and Gruissem, 1987) as evidenced by ineffective transcription termination of native plastid genes and the common post-transcriptional generation of mature chloroplast transcript ends (e.g. del Campo, 2009; Ji *et al*., 2019; Stern *et al*., 2010; Stern and Gruissem, 1987; Zhelyazkova *et al*., 2012a). Consequently, regions downstream of the *aadA* cassette are regularly overexpressed at the RNA level. We observed two possible consequences of this transcriptional read-through: (i) overaccumulation of transcripts from sense-oriented neighbouring genes, which usually also causes higher translation output of these genes (Figures 2, 3 and 9), or (ii)
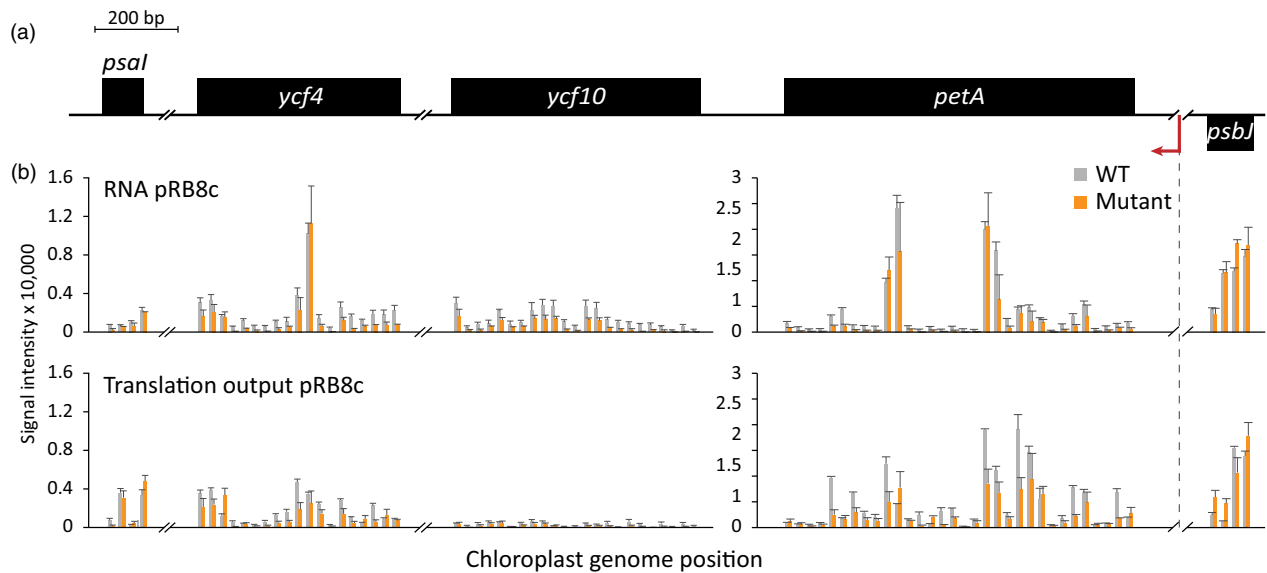
(a)

(b)

Figure 5 Mild reduction in ribosome coverage of *petA*, *ycf10* and *ycf4*, and transcript accumulation of *ycf10* induced by intergenic antisense-directed *aadA* insertion downstream of the *psaI-ycf4-ycf10-petA* transcription unit. (a) Gene map representing the intergenic insertion site position and transcription direction of the *aadA* resistance marker cassette in the pRB8c line (Bock *et al.*, 1994; Loiacono *et al.*, 2019). For better illustration, large intergenic regions were removed (indicated by broken horizontal lines). Sequences flanking the *aadA* insertion site are provided in Appendix S1. All other details are as described in Figure 2a. (b) Bar plots illustrating the normalized relative signal intensities of transcript accumulation (RNA; top diagram) and ribosome footprints (translation output; bottom diagram) for probes in protein-coding regions upstream and downstream of the intergenic *aadA* insertion site. Note that, due to the substantially higher expression levels of *petA* and *psbJ*, the ordinates for the diagrams shown on the left and on the right have different scales. All other labelling details are identical to Figure 2b-d.
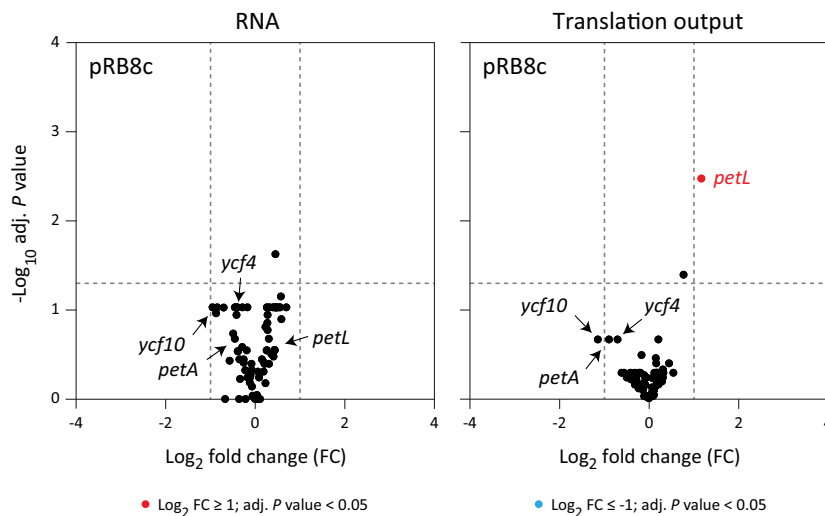

Figure 6 Global analysis of chloroplast gene expression in a transplastomic line with intergenic antisense-directed *aadA* insertion downstream of the *psaI-ycf4-ycf10-petA* transcription unit. Comparative analyses of relative changes in transcript accumulation (left) and translation output (right) in pRB8c (Bock *et al.*, 1994; Loiacono *et al.*, 2019) and wild-type plants (details as described in Figure 3). Values for all genes are provided in Data S1.

reduced expression of antisense-oriented genes at both the transcriptional and translational levels (Figures 5–9).

Transcriptional read-through from the *aadA* resistance marker causing transcript overaccumulation of downstream regions was previously observed in a number of cases (e.g. Fu *et al.*, 2021; Lu *et al.*, 2013; Lutz *et al.*, 2007; Verhounig *et al.*, 2010; Zhang *et al.*, 2015; Zhou *et al.*, 2007, 2008) but the consequences on the translational level remained virtually unexplored. Here, we have shown that transcriptional read-through originating from different locations of the *aadA* cassette within the *petL-psaJ-rpl33-rps18* transcription unit regularly causes elevated translation output for downstream sense-oriented genes (Figures 2 and 3). The generality of this finding is further substantiated by the previous finding that transcriptional read-through into the *psbD-psbC-psbZ* transcription unit causes transcript overaccumulation and, in addition, leads to considerably elevated translation output (Fu *et al.*, 2021). Hence, we conclude that elevated translation output of downstream, sense-oriented reading frames is a common consequence of transcriptional read-through from highly expressed chloroplast transgenes.

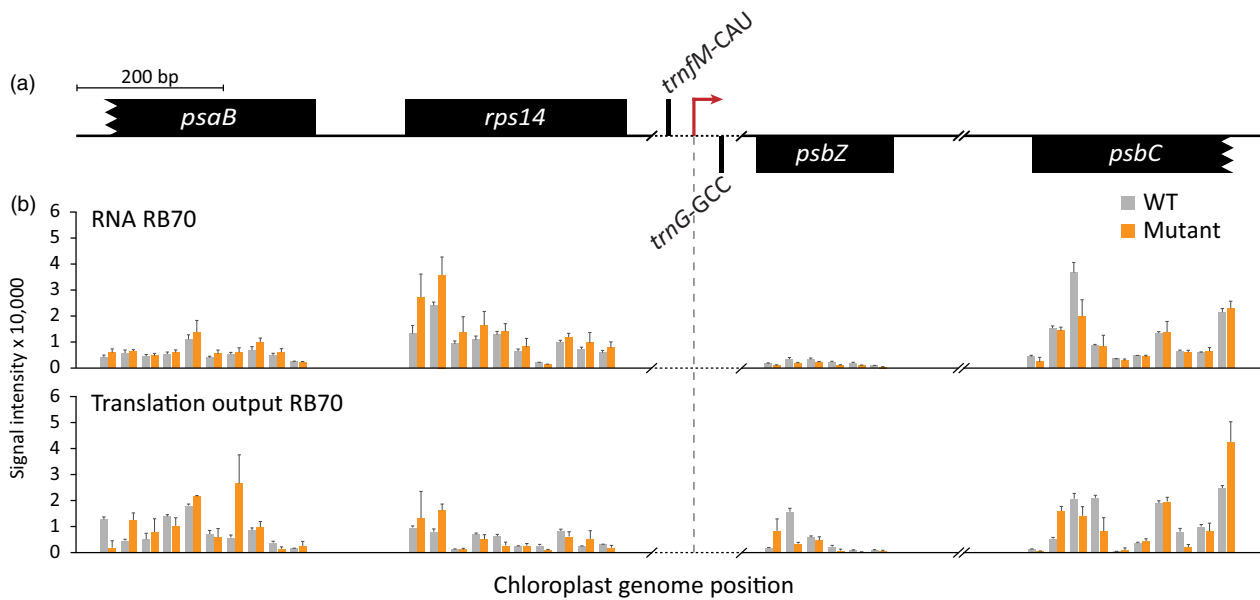Interestingly, in the *ΔpetL* mutant, we observed that the translation output of *rpl33* and *rps18* that are located

Figure 7 Intergenic antisense-directed *aadA* insertion downstream of *trnG*-GCC does not substantially alter expression of neighbouring reading frames. (a) Gene map of the intergenic *aadA* insertion site position and transcription direction in the transplastomic RB70 line (Krech *et al.*, 2013; Ruf *et al.*, 2001). To simplify visualization, large regions between protein-coding genes were removed (designated by broken horizontal lines). All other details are as described in Figure 2a. Sequences flanking the *aadA* insertion site are provided in Appendix S1. (b) Bar plots representing the normalized relative signal intensities of transcript accumulation (RNA; top diagram) and ribosome footprints (translation output; bottom diagram) for probes in protein-coding regions upstream and downstream of the intergenic *aadA* cassette insertion site. Labelling details are as in Figure 2b-d.

Figure 8 Genome-wide analysis reveals virtually unaltered chloroplast gene expression in a transplastomic line with intergenic antisense-directed *aadA* insertion downstream of *trnG*-GCC. Comparative examination of relative changes in transcript accumulation (left) and translation output (right) in RB70 and wild-type plants (details as described in Figure 3). Values for all genes are provided in Data S1.
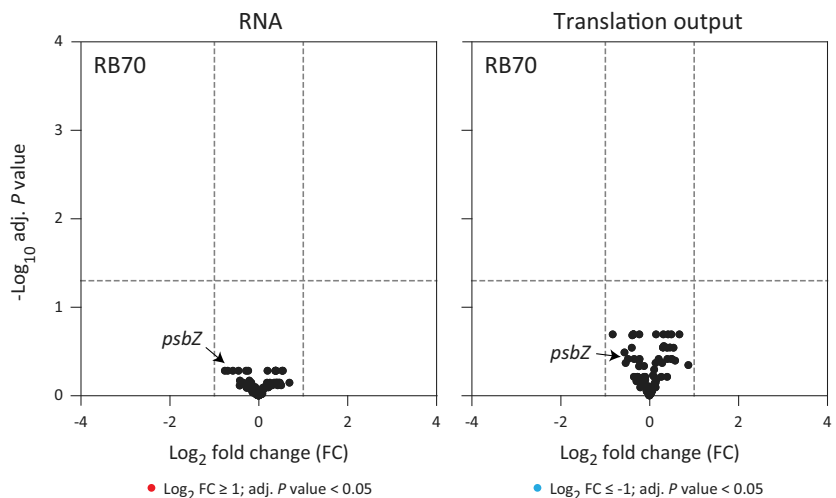
downstream of the *aadA* resistance marker was substantially stronger induced than expected by the change in their transcript accumulation (less than twofold and statistically not significantly increased transcript levels, but significantly and more than threefold and twofold elevated translation output, respectively; Figure 3a). In bacteria, translation of adjacent reading frames in polycistronic mRNAs is often coupled in that translation of the upstream coding sequence activates the translation of downstream reading frame(s) (Jackson *et al.*, 2007; Yamamoto *et al.*, 2016). Translational coupling of adjacent reading frames was also observed in plastid translation (Adachi *et al.*, 2012; Fu *et al.*, 2021; Martin Avila *et al.*, 2016; Yukawa and Sugiura, 2008; Zoschke *et al.*, 2013) and has been exploited in chloroplast

biotechnology (Martin Avila *et al.*, 2016). Hence, in addition to the transcriptional overexpression of genes downstream of *aadA*, the strongly translated *aadA* reading frame may also activate the translation of downstream reading frames that are located on the same mRNA by translational coupling or by increasing the local ribosome density. Therefore, translational coupling could potentially explain the discrepancy observed between the increase in translation output and transcript levels for *rpl33* and *rps18* in the *ΔpetL* mutant.

Apart from the effects caused by sense-directed read-through from the *aadA* cassette, we also observed antisense effects (Figures 5–9). These include reduced transcript accumulation and decreased translation output for genes that are located
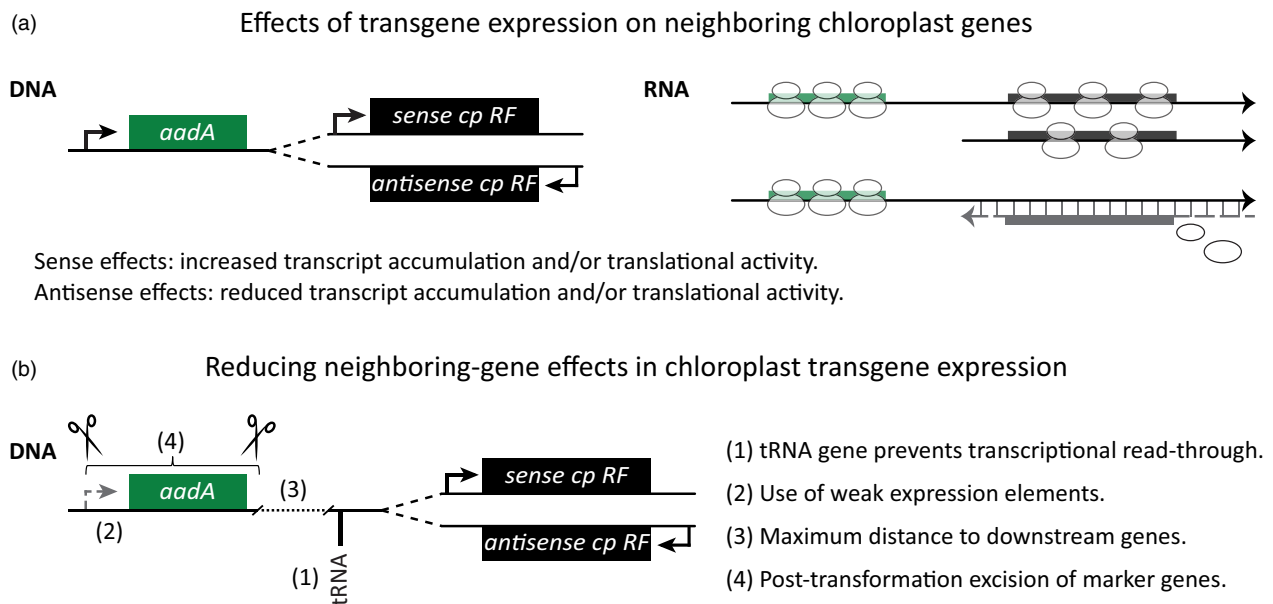
(a)

Effects of transgene expression on neighboring chloroplast genes

Sense effects: increased transcript accumulation and/or translational activity.
Antisense effects: reduced transcript accumulation and/or translational activity.

(b)

Reducing neighboring-gene effects in chloroplast transgene expression

(1) tRNA gene prevents transcriptional read-through.

(2) Use of weak expression elements.

(3) Maximum distance to downstream genes.

(4) Post-transformation excision of marker genes.

Figure 9 Observed effects of transgene expression on downstream resident chloroplast genes and strategies to alleviate such neighbouring-gene effects. (a) Transgene expression enhances expression of downstream sense-oriented endogenous chloroplast genes and inhibits expression of downstream antisense-oriented genes. Left panel (DNA): genomic arrangement of the *aadA* transgene (green box representing the *aadA* reading frame) located upstream of a native chloroplast gene that is either sense oriented or antisense oriented. In both cases, the *aadA* gene is connected by a dashed line with the downstream genomic region (black box above or below the horizontal line depicting native chloroplast (cp) reading frames (RF) in sense or antisense direction). Black arrows indicate promoters with arrowheads indicating the direction of transcription. Right panel (RNA): the resulting transcripts (horizontal lines) and their translation are shown (upper part: sense arrangement, lower part: antisense arrangement). Arrowheads indicate the 5′-to-3′ orientation of the transcripts. Translating ribosomes are depicted by two connected white ellipses covering the reading frames (filled rectangles). The dashed grey horizontal line indicates reduced transcript accumulation from a downstream gene in antisense orientation. Complementary base paring between RNA molecules is shown by vertical grey lines connecting the sense and antisense transcripts and is suggested to cause inhibition of translation (indicated by the two unassembled ribosomal subunits). (b) Strategies to alleviate the neighbouring-gene effects which are described in (a). Four strategies (1–4) to reduce neighbouring-gene effects are schematically shown on the left and described in the legend on the right. Note that these approaches are not mutually exclusive and can also be combined. The grey arrowed dashed line depicts a weak promoter (exemplifying weak expression elements). Scissors symbolize strategies for marker gene excision. Broken horizontal line connected with a dotted line represents a large intergenic region that is not shown. Other details are as in (a).

downstream of the *aadA* cassette in antisense direction. Antisense transcription is widespread in all domains of life and has been assigned to various molecular mechanisms such as transcriptional and translational regulation as well as control of RNA stability, although, especially in prokaryotes, the specific roles of particular antisense transcripts often remained obscure (e.g. Georg and Hess, 2018; Lasa *et al*., 2012; Layer and Weil, 2009). In plastids, numerous native antisense transcripts have been described (e.g. Georg *et al*., 2010; Sharwood *et al*., 2011b; Zghidi-Abouzid *et al*., 2011; Zhelyazkova *et al*., 2012b), and there is evidence for antisense transcripts being involved in, or interfering with, chloroplast RNA processing (e.g. Sharwood *et al*., 2011a; Zghidi-Abouzid *et al*., 2011). Additionally, it is known that transcriptional interference commonly occurs in convergent transcriptional processes in eukaryotes and bacteria (Shearwin *et al*., 2005), and this may also happen in chloroplasts (Ji *et al*., 2019). Together, these effects might explain the reduced accumulation of transcripts that are encoded in antisense orientation downstream of the *aadA* gene (e.g. *ycf10*, *petA* and *ycf4* in pRB8c; Figure 5). Furthermore, elevated chloroplast antisense transcript levels can inhibit translation of sense-oriented reading frames, which was proposed to be caused by formation of double-stranded RNA that masks the reading frame and thereby affects translational activity

(Sharwood *et al*., 2011a). This effect might cause the reduced translation output of reading frames that are located in antisense direction downstream of the *aadA* gene, especially in those cases where the consequences of antisense transcription are stronger on the translational level than on the transcriptional level (*ycf10*, *petA* and *ycf4* in pRB8c; Figure 5). Our data show that the ribosome coverage of *petA*, *ycf10* and *ycf4* in pRB8c is mostly uniformly reduced throughout the whole reading frame. This finding suggests that translational inhibition by antisense transcripts is predominantly caused at the initiation level or evenly affects elongation across the entire reading frame (Figures 5 and 7; Data S1).

Most of the effects of *aadA* transgene expression are related to transcriptional read-through. This raises the question whether there is any potential route towards better control of transcriptional read-through originating from the *aadA* gene. In the mitochondria of animals, tRNA sequences in primary transcripts serve as transcript processing sites and tRNA genes are also involved in transcription termination (reviewed in D'Souza and Minczuk, 2018; Guja and Garcia-Diaz, 2012). Interestingly, some chloroplast tRNAs were shown to terminate transcription *in vitro* (Gruissem *et al*., 1986; Stern and Gruissem, 1987), and there is evidence that chloroplast tRNAs can serve as transcript processing sites in longer precursor RNAs (e.g. Zhou *et al*., 2007, 2015), which

is reminiscent of tRNA functions as transcription terminators in animal mitochondria (Ji *et al.*, 2019). Our data suggest that transcriptional read-through effects can alter the expression of chloroplast genes that are up to ~2 kb (or even more) downstream of the *aadA* cassette (*ycf10* and *ycf4* in pRB8c, Figure 5). The distance between *aadA* and *psbZ* in RB70 is only ~0.5 kb and that between *aadA* and *psaJ* in Δ*petL* is only ~1.3 kb, but nonetheless, expression of *psbZ* and *psaJ* is not significantly affected by *aadA* insertion (Figures 2b, 3, 7b and 8). Remarkably, in both cases, tRNAs are located in antisense direction between *aadA* and the next gene downstream (*trnG*-GCC upstream of *psbZ*, and *trnW*-CCA and *trnP*-UGG upstream of *psaJ*; Figures 2a and 7a). This may indicate that antisense-located tRNAs efficiently terminate chloroplast transcription. Alternatively, the secondary structures that antisense tRNA transcripts fold into (which often are similar to the cloverleaf-type secondary structures the sense transcripts of tRNA genes form) may enable efficient *aadA* transcript 3′ end processing by 3′-5′ exonucleases and thereby prevent accumulation of downstream read-through transcripts (Figure 9). Interestingly, similar to the RB70 insertion site (Ruf *et al.*, 2001), another commonly used insertion site (within the chloroplast rRNA transcription unit) is flanked by tRNAs (De Cosa *et al.*, 2001). The very successful use of these two insertion sites in chloroplast biotechnology may be, at least in part, explainable by limited neighbouring-gene effects that are due to reduced read-through effects by tRNA-mediated transcription termination.

Interestingly, we also observed a few effects of transgene insertion on the expression of distantly located genes. These effects cannot be a direct consequence of *aadA* expression. Such *trans* effects are evident for *rps15* and *petL* overexpression in the Δ*rpl33* and pRB8c lines as well as down-regulation of *rbcL* and *petN* expression in the Δ*psaJ*, Δ*rpl33* and Δ*petL* mutants, respectively (Figures 3 and 6). A possible explanation could be the operation of regulatory circuits that couple the expression of the knocked-out gene with the expression of the distantly located gene. It is, for instance, tempting to speculate that the increased *rps15* transcript accumulation in Δ*rpl33* could reflect a potential regulatory connection between *rpl33* and *rps15* gene expression. Indeed, the previously observed synthetic lethality of a double knockout of these two non-essential genes may be interpreted as genetic evidence for such a regulatory link (Ehrnthaler *et al.*, 2014). In the same way, *petA* and *petL* expression (in pRB8c), and *petL/petG* and *petN* expression (in Δ*petL*) may be coupled based on the co-location of their gene products in the cytochrome b_6f complex (Figures 3 and 6). Similarly, we found that *rps18* overexpression in the Δ*petL*, Δ*psaJ* and Δ*rpl33* transplastomic lines correlates with the overaccumulation of the major 16S rRNA precursor transcript (Figure 4). This may suggest a connection between *rps18* expression levels and 16S rRNA processing, which would be, to the best of our knowledge, unprecedented in bacterial ribosome assembly. However, the attenuated 16S rRNA processing did not substantially affect the accumulation of the chloroplast small ribosomal subunit as evidenced by wild-type-like levels of the mature 16S rRNA and the Rps15 protein (Figures 1 and 4; Figure S1). Another reason for *trans* effects of *aadA* insertion could be given by regulatory responses of secondary nature to a mildly disturbed chloroplast function. This possibly explains the reduced *rbcL* expression in the Δ*psaJ* and Δ*rpl33* mutants (Figure 3), both of which have mild photosynthesis defects due to slightly impaired PSI and ribosome functions, respectively (Rogalski *et al.*, 2008; Schöttler *et al.*, 2007b; Schwenkert *et al.*, 2007).

In sum, our work reported here reveals principles of how transgene expression from the chloroplast genome affects the expression of neighbouring endogenous genes (Figure 9a). Changes in gene expression are usually interpreted in the context of changes in cellular homeostasis. However, this relationship between disturbed gene expression and altered cellular homeostasis can be broken, for example, by the action of additional limiting factors, the redundancy in regulatory networks or post-translational mechanisms (e.g. dynamics in proteostasis). In line with this notion, reduced *rbcL* expression in the Δ*psaJ* and Δ*rpl33* mutants and increased *rps18* expression in the Δ*petL*, Δ*psaJ* and Δ*rpl33* mutants did not cause altered Rubisco and ribosome accumulation, respectively, or any visible mutant phenotype in the growth conditions tested (Figures 1 and 3; Figure S1). The operation of additional layers of regulation, however, does not rule out potential effects in specific growth conditions, developmental stages or tissues. On the other hand, the mildly impaired *petA* expression in pRB8c did cause reduced PetA protein accumulation (Figures 1, 6 and S1), thus providing an example of a direct effect on cellular homeostasis. However, independent of the ultimate consequences of altered gene expression, the principles of chloroplast neighbouring-gene effects identified in the course of this work enable us to derive some guidelines for the design of chloroplast transformation constructs that minimize the risk of undesired neighbouring-gene effects (Figure 9b). First, since tRNAs prevent transcriptional read-through (or at least RNA accumulation from downstream regions), transgenes should preferably be inserted upstream of chloroplast tRNA genes. Second, the use of weak expression elements for the resistance marker reduces the effects of transcriptional read-through (e.g. Wu *et al.*, 2022). Third, transgenes should be inserted at maximum distance to downstream genes to dampen read-through effects on gene expression. Fourth, methods for post-transformation excision of marker genes are available (Lutz *et al.*, 2006; Martin Avila *et al.*, 2016) and can be used to omit transgene expression effects on neighbouring genes, at least in the final stage of generating transplastomic lines. Fifth, recent progress with marker-free manipulation of the plastid genome and ectopic chloroplast transgene expression provide additional options to avoid neighbouring-gene effects altogether (Maliga, 2022). Finally, known 'neutral' insertion sites should be preferred for transgene insertion. A good example is the RB70 insertion site, which shows negligible side effects on chloroplast gene expression (Figure 8) and was successfully used in a number of transplastomic studies (e.g. Fuentes *et al.*, 2016; Krech *et al.*, 2013; Qu *et al.*, 2018; Ruf *et al.*, 2001, 2019; Stanbekova *et al.*, 2021; Wurbs *et al.*, 2007). The RB70 insertion site has also been efficiently used in co-transformation experiments to engineer distantly located sites in the chloroplast genome, where neighbouring-gene effects must be avoided to obtain clean mutants (Krech *et al.*, 2013; Torabi *et al.*, 2014).

Experimental procedures

Plant material and growth conditions

Plants were grown on soil at 16 h light (150 μmol m^{-2} s^{-1})/8 h dark at 22 °C/18 °C and 70% humidity. The aerial part of the plants was harvested 21 days after sowing when the seedlings were in the four-true-leaf stage (Figure 1a), immediately snap-frozen in liquid nitrogen and stored at −80 °C until further use. The genotypes of the transplastomic lines were confirmed by *aadA*-spanning PCR (primers listed in Table S1). All analysed

transplastomic lines contain the identical *aadA* cassette (derived from vector pZS197), including identical expression elements: the tobacco 16S rDNA promoter, the *psbA* terminator region and an *rbcL*-derived Shine-Dalgarno sequence (Svab and Maliga, 1993).

Ribosome profiling

Ribosome footprints and total RNA were prepared, labelled and hybridized as described previously (Gao *et al.*, 2022; Schuster *et al.*, 2020; Trösch *et al.*, 2018). Briefly, 4 μg ribosome footprint RNA and 3.5 μg total RNA from transplastomic and wild-type plants were chemically labelled with Cy3 and Cy5, respectively, and hybridized onto identical custom microarrays as described (Trösch *et al.*, 2018). These microarrays contain 50-nt probes that cover chloroplast protein-coding regions and UTRs in high density (one 50mer probe starting every 30 nt within these regions; each probe is represented in four technical replicate spots on each microarray). Transcript and ribosome profiling data were processed as described in Schuster *et al.* (2020) and Gao *et al.* (2022). In brief, background-subtracted probe signals lower than 100 were set to zero. Probes that showed signal saturation were excluded from further analysis. Signal intensities of all probes covering protein-coding regions (in all experiments and replicates) were normalized to the average signal intensities of probes covering protein-coding regions in the three wild-type datasets (derived from competitive hybridization with RNA and footprints from the *ΔpsaJ* mutant, respectively). Thereby, technical variations (caused by, e.g., different labelling efficiencies of Cy3 and Cy5) were eliminated and, consequently, conclusive comparisons of all transplastomic-line-derived and wild-type datasets were enabled. The average values and standard deviations of three biological replicates for these normalized signal intensities of probes within protein-coding regions were plotted according to their genomic position to reflect local differences in transcript accumulation or translation output between wild-type and transplastomic plants (shown in Figures 2, 5 and 7). Next, we calculated for each chloroplast reading frame average signal intensity values based on the signal intensities of all probes located in a given reading frame. This enabled us to determine the fold-change differences in comparison to wild-type data shown in Figures 3, 6 and 8. Genes with 50% or less probes showing above-background signal were excluded from the downstream analysis (i.e. *P* value calculations), and are accordingly omitted from the volcano plots shown in Figures 3, 6 and 8. Significances of changes in gene expression were calculated by Student's *t*-test and the derived *P* values were adjusted for multiple testing according to Benjamini and Hochberg (1995). *P* values are shown in Data S1. Genes with twofold or more than twofold change in either direction (i.e. values that are ≥1 or ≤−1 after $\log_2$ transformation) and adjusted *P* value <0.05 were considered as significantly differentially expressed, and are labelled with red or blue dots in the shown volcano plots (for up- and down-regulated genes, respectively).

Immunoblot and RNA gel blot analyses

For immunoblot analyses, total cellular proteins were isolated from aerial parts of transplastomic and wild-type plants. 4 μg total protein were size-separated by polyacrylamide gel electrophoresis and blotted onto nitrocellulose membranes as described previously (Gao *et al.*, 2022). The membranes were subsequently incubated with antibodies against PsbA, PetB, PSAD, AtpB and Rps15 obtained from Agrisera (Vännäs, Sweden;

product No. (working dilution)): AS10704 (1:10 000), AS03034 (1:5000), AS09461 (1:1000), AS05085 (1:10 000) and AS14 2779 (1:1000). Immunodetection was performed using ECL Plus detection reagents (GE Healthcare, Chicago, IL, USA) and documented with Syngene G:BOX Chemi XT4 (SynOptics). Signal intensities were quantified using the Image Lab software (Bio-Rad).

For RNA gel blot analyses, total RNA was extracted from aerial parts of transplastomic and wild-type plants with the TRIzol reagent (Invitrogen) following the manufacturer's instructions. After denaturation, 300 ng of purified total RNA was separated in 1.2% (w/v) agarose gels and then transferred to Hybond-N nylon membranes (GE Healthcare). Equal loading was verified by methylene blue staining. A PCR probe against the 16S rRNA was amplified from genomic tobacco DNA with the primers listed in Table S1 and 50-ng PCR product was labelled with [alpha-P32]-dCTP using the Multiprime DNA labelling system (GE Healthcare). After denaturing the probe for 5 min at 95 °C, hybridization was performed at 65°C in Church buffer (0,5 M sodium phosphate (pH 7,0), 7% (w/v) SDS and 1 mM EDTA), followed by two washes at 65 °C for 10 min each in 1x SSC (0.15 M sodium chloride and 0.015 M sodium citrate)/0.5% SDS and a third wash for 10 min in 0.5x SSC/0.1% SDS. In order to prevent saturation effects for the detection of the highly abundant chloroplast 16S rRNA, 10 times cold probe (non-radiolabelled probe) was added to the hybridization reaction. The signals were detected by exposing PhosphorImager screens (GE Healthcare) to the radio-labelled membranes. Signals were quantified using the Image Lab software (Bio-Rad).

Accession numbers

Tobacco microarray design was based on the GenBank library under accession number Z00044.2. All discussed genes are annotated under this GenBank library accession number.

Conflict of interest

The authors have no conflicts of interest to declare.

Data availability statement

All data that support the findings of this study are provided within the manuscript or the Supporting Information.

References

Adachi, Y., Kuroda, H., Yukawa, Y. and Sugiura, M. (2012) Translation of partially overlapping *psbD-psbC* mRNAs in chloroplasts: the role of 5′-processing and translational coupling. *Nucleic Acids Res.* **40**, 3152–3158.

Atias, N., Kupiec, M. and Sharan, R. (2016) Systematic identification and correction of annotation errors in the genetic interaction map of *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **44**, e50.

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

Bar-On, Y.M. and Milo, R. (2019) The global mass and average rate of rubisco. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 4738–4743.

Baryshnikova, A. and Andrews, B. (2012) Neighboring-gene effect: a genetic uncertainty principle. *Nat. Methods* **9**, 343.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300.

Ben-Shitrit, T., Yosef, N., Shemesh, K., Sharan, R., Ruppin, E. and Kupiec, M. (2012) Systematic identification of gene annotation errors in the widely used yeast mutation collections. *Nat. Methods* **9**, 373–378.

Bock, R. (2007) *Cell and Molecular Biology of Plastids*. Berlin: Springer Verlag.

Bock, R. (2015) Engineering plastid genomes: methods, tools, and applications in basic research and biotechnology. *Annu. Rev. Plant Biol.* **66**, 211–241.

Bock, R., Kössel, H. and Maliga, P. (1994) Introduction of a heterologous editing site into the tobacco plastid genome: the lack of RNA editing leads to a mutant phenotype. *EMBO J.* **13**, 4623–4628.

Bock, R. and Timmis, J.N. (2008) Reconstructing evolution: gene transfer from plastids to the nucleus. *Bioessays* **30**, 556–566.

Bock, R. and Warzecha, H. (2010) Solar-powered factories for new vaccines and antibiotics. *Trends Biotechnol.* **28**, 246–252.

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.

Boynton, J.E., Gillham, N.W., Harris, E.H., Hosler, J.P., Johnson, A.M., Jones, A.R., Randolph-Anderson, B.L. *et al.* (1988) Chloroplast transformation in *Chlamydomonas* with high velocity microprojectiles. *Science* **240**, 1534–1538.

Daniell, H. (2006) Production of biopharmaceuticals and vaccines in plants via the chloroplast genome. *Biotechnol. J.* **1**, 1071–1079.

De Cosa, B., Moar, W., Lee, S.B., Miller, M. and Daniell, H. (2001) Overexpression of the *Bt cry2Aa2* operon in chloroplasts leads to formation of insecticidal crystals. *Nat. Biotechnol.* **19**, 71–74.

del Campo, E.M. (2009) Post-transcriptional control of chloroplast gene expression. *Gene Regul Syst Bio* **3**, 31–47.

D'Souza, A.R. and Minczuk, M. (2018) Mitochondrial transcription and translation: overview. *Essays Biochem.* **62**, 309–320.

Egorov, A.A., Alexandrov, A.I., Urakov, V.N., Makeeva, D.S., Edakin, R.O., Kushchenko, A.S., Gladyshev, V.N. *et al.* (2021) A standard knockout procedure alters expression of adjacent loci at the translational level. *Nucleic Acids Res.* **49**, 11134–11144.

Ehrnthaler, M., Scharff, L.B., Fleischmann, T.T., Hasse, C., Ruf, S. and Bock, R. (2014) Synthetic lethality in the tobacco plastid ribosome and its rescue at elevated growth temperatures. *Plant Cell* **26**, 765–776.

Ellis, R.J. (1979) The most abundant protein in the world. *Trends Biochem. Sci.* **4**, 241–244.

Fiebig, A., Stegemann, S. and Bock, R. (2004) Rapid evolution of RNA editing sites in a small non-essential plastid gene. *Nucleic Acids Res.* **32**, 3615–3622.

Fu, H.Y., Ghandour, R., Ruf, S., Zoschke, R., Bock, R. and Schöttler, M.A. (2021) The availability of neither D2 nor CP43 limits the biogenesis of photosystem II in tobacco. *Plant Physiol.* **185**, 1111–1130.

Fuentes, P., Armarego-Marriott, T. and Bock, R. (2018) Plastid transformation and its application in metabolic engineering. *Curr. Opin. Biotechnol.* **49**, 10–15.

Fuentes, P., Zhou, F., Erban, A., Karcher, D., Kopka, J. and Bock, R. (2016) A new synthetic biology approach allows transfer of an entire metabolic pathway from a medicinal plant to a biomass crop. *Elife* **5**, e13664.

Gao, Y., Thiele, W., Saleh, O., Scossa, F., Arabi, F., Zhang, H., Sampathkumar, A. *et al.* (2022) Chloroplast translational regulation uncovers nonessential photosynthesis genes as key players in plant cold acclimation. *Plant Cell* **34**, 2056–2079.

Georg, J. and Hess, W.R. (2018) Widespread antisense transcription in prokaryotes. *Microbiol. Spectr.* **6**, 191–210.

Georg, J., Honsel, A., Voss, B., Rennenberg, H. and Hess, W.R. (2010) A long antisense RNA in plant chloroplasts. *New Phytol.* **186**, 615–622.

Ghanbarian, A.T. and Hurst, L.D. (2015) Neighboring genes show correlated evolution in gene expression. *Mol. Biol. Evol.* **32**, 1748–1766.

Gruissem, W., Greenberg, B.M., Zurawski, G. and Hallick, R.B. (1986) Chloroplast gene expression and promoter identification in chloroplast extracts. *Methods Enzymol.* **118**, 253–270.

Guja, K.E. and Garcia-Diaz, M. (2012) Hitting the brakes: termination of mitochondrial transcription. *Biochim. Biophys. Acta* **1819**, 939–947.

Hegeman, C.E., Halter, C.P., Owens, T.G. and Hanson, M.R. (2005) Expression of complementary RNA from chloroplast transgenes affects editing efficiency of transgene and endogenous chloroplast transcripts. *Nucleic Acids Res.* **33**, 1454–1464.

Jackson, R.J., Kaminski, A. and Pöyry, T.A.A. (2007) Coupled termination-reinitiation events in mRNA translation. In *Translational Control in Biology and Medicine*(Mathews, M.B., Sonenberg, N. and Hershey, J.W.B., eds), pp. 197–223. New York: Cold Spring Harbor Laboratory Press.

Ji, D., Manavski, N., Meurer, J., Zhang, L. and Chi, W. (2019) Regulated chloroplast transcription termination. *Biochim. Biophys. Acta Bioenerg.* **1860**, 69–77.

Krech, K., Fu, H.Y., Thiele, W., Ruf, S., Schöttler, M.A. and Bock, R. (2013) Reverse genetics in complex multigene operons by co-transformation of the plastid genome and its application to the open reading frame previously designated *psbN*. *Plant J.* **75**, 1062–1074.

Lasa, I., Toledo-Arana, A. and Gingeras, T.R. (2012) An effort to make sense of antisense transcription in bacteria. *RNA Biol.* **9**, 1039–1044.

Layer, J.H. and Weil, P.A. (2009) Ubiquitous antisense transcription in eukaryotes: novel regulatory mechanism or byproduct of opportunistic RNA polymerase? *F1000 Biol. Rep.* **1**, 33.

Loiacono, F.V., Thiele, W., Schöttler, M.A., Tillich, M. and Bock, R. (2019) Establishment of a heterologous RNA editing event in chloroplasts. *Plant Physiol.* **181**, 891–900.

Lu, Y., Rijzaani, H., Karcher, D., Ruf, S. and Bock, R. (2013) Efficient metabolic pathway engineering in transgenic tobacco and tomato plastids with synthetic multigene operons. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E623–E632.

Lutz, K.A., Azhagiri, A.K., Tungsuchat-Huang, T. and Maliga, P. (2007) A guide to choosing vectors for transformation of the plastid genome of higher plants. *Plant Physiol.* **145**, 1201–1210.

Lutz, K.A., Svab, Z. and Maliga, P. (2006) Construction of marker-free transplastomic tobacco using the Cre-loxP site-specific recombination system. *Nat. Protoc.* **1**, 900–910.

Maliga, P. (2022) Engineering the plastid and mitochondrial genomes of flowering plants. *Nat. Plants* **8**, 996–1006.

Martin Avila, E., Gisby, M.F. and Day, A. (2016) Seamless editing of the chloroplast genome in plants. *BMC Plant Biol.* **16**, 168.

Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D. *et al.* (2002) Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12246–12251.

Oey, M., Lohse, M., Kreikemeyer, B. and Bock, R. (2009) Exhaustion of the chloroplast protein synthesis capacity by massive expression of a highly stable protein antibiotic. *Plant J.* **57**, 436–445.

Qu, Y., Legen, J., Arndt, J., Henkel, S., Hoppe, G., Thieme, C., Ranzini, G. *et al.* (2018) Ectopic transplastomic expression of a synthetic *matK* gene leads to cotyledon-specific leaf variegation. *Front. Plant Sci.* **9**, 1453.

Rogalski, M., Schöttler, M.A., Thiele, W., Schulze, W.X. and Bock, R. (2008) Rpl33, a nonessential plastid-encoded ribosomal protein in tobacco, is required under cold stress conditions. *Plant Cell* **20**, 2221–2237.

Rolland, N., Dorne, A.J., Amoroso, G., Sültemeyer, D.F., Joyard, J. and Rochaix, J.D. (1997) Disruption of the plastid *ycf10* open reading frame affects uptake of inorganic carbon in the chloroplast of *Chlamydomonas. EMBO J.* **16**, 6713–6726.

Ruf, S., Forner, J., Hasse, C., Kroop, X., Seeger, S., Schollbach, L., Schadach, A. *et al.* (2019) High-efficiency generation of fertile transplastomic *Arabidopsis* plants. *Nat. Plants* **5**, 282–289.

Ruf, S., Hermann, M., Berger, I.J., Carrer, H. and Bock, R. (2001) Stable genetic transformation of tomato plastids and expression of a foreign protein in fruit. *Nat. Biotechnol.* **19**, 870–875.

Ruf, S., Karcher, D. and Bock, R. (2007) Determining the transgene containment level provided by chloroplast transformation. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 6998–7002.

Sagan, L. (1967) On the origin of mitosing cells. *J. Theor. Biol.* **14**, 255–274.

Sakamoto, W. and Takami, T. (2018) Chloroplast DNA Dynamics: Copy Number, Quality Control and Degradation. *Plant Cell Physiol.* **59**, 1120–1127.

Schöttler, M.A., Flügel, C., Thiele, W. and Bock, R. (2007a) Knock-out of the plastid-encoded PetL subunit results in reduced stability and accelerated leaf age-dependent loss of the cytochrome b_6f complex. *J. Biol. Chem.* **282**, 976–985.

Schöttler, M.A., Flügel, C., Thiele, W., Stegemann, S. and Bock, R. (2007b) The plastome-encoded PsaJ subunit is required for efficient photosystem I excitation, but not for plastocyanin oxidation in tobacco. *Biochem. J.* **403**, 251–260.

Schuster, M., Gao, Y., Schöttler, M.A., Bock, R. and Zoschke, R. (2020) Limited responsiveness of chloroplast gene expression during acclimation to high light in tobacco. *Plant Physiol.* **182**, 424–435.

Schwenkert, S., Legen, J., Takami, T., Shikanai, T., Herrmann, R.G. and Meurer, J. (2007) Role of the low-molecular-weight subunits PetL, PetG, and PetN in assembly, stability, and dimerization of the cytochrome b_6f complex in tobacco. *Plant Physiol.* **144**, 1924–1935.

Sharwood, R.E., Halpert, M., Luro, S., Schuster, G. and Stern, D.B. (2011a) Chloroplast RNase J compensates for inefficient transcription termination by removal of antisense RNA. *RNA* **17**, 2165–2176.

Sharwood, R.E., Hotto, A.M., Bollenbach, T.J. and Stern, D.B. (2011b) Overaccumulation of the chloroplast antisense RNA AS5 is correlated with decreased abundance of 5 S rRNA in vivo and inefficient 5 S rRNA maturation in vitro. *RNA* **17**, 230–243.

Shearwin, K.E., Callen, B.P. and Egan, J.B. (2005) Transcriptional interference - a crash course. *Trends Genet.* **21**, 339–345.

Stanbekova, G., Beisenov, D., Nizkorodova, A., Iskakov, B. and Warzecha, H. (2021) Production of the sheep pox virus structural protein SPPV117 in tobacco chloroplasts. *Biotechnol. Lett.* **43**, 1475–1485.

Steinrueck, M. and Guet, C.C. (2017) Complex chromosomal neighborhood effects determine the adaptive potential of a gene under selection. *Elife* **6**, e25100.

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

Stern, D.B. and Gruissem, W. (1987) Control of plastid gene expression: 3′ inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. *Cell* **51**, 1145–1157.

Svab, Z., Hajdukiewicz, P. and Maliga, P. (1990a) Stable transformation of plastids in higher plants. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 8526–8530.

Svab, Z., Harper, E.C., Jones, J.D. and Maliga, P. (1990b) Aminoglycoside-3″-adenyltransferase confers resistance to spectinomycin and streptomycin in *Nicotiana tabacum. Plant Mol. Biol.* **14**, 197–205.

Svab, Z. and Maliga, P. (1993) High-frequency plastid transformation in tobacco by selection for a chimeric *aadA* gene. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 913–917.

Svab, Z. and Maliga, P. (2007) Exceptional transmission of plastids and mitochondria from the transplastomic pollen parent and its impact on transgene containment. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 7003–7008.

Torabi, S., Umate, P., Manavski, N., Plöchinger, M., Kleinknecht, L., Bogireddi, H., Herrmann, R.G. *et al.* (2014) PsbN is required for assembly of the photosystem II reaction center in *Nicotiana tabacum. Plant Cell* **26**, 1183–1199.

Trinh, M.D.L., Hashimoto, A., Kono, M., Takaichi, S., Nakahira, Y. and Masuda, S. (2021) Lack of plastid-encoded Ycf10, a homolog of the nuclear-encoded DLDG1 and the cyanobacterial PxcA, enhances the induction of non-photochemical quenching in tobacco. *Plant Direct* **5**, e368.

Trösch, R., Barahimipour, R., Gao, Y., Badillo-Corona, J.A., Gotsmann, V.L., Zimmer, D., Mühlhaus, T. *et al.* (2018) Commonalities and differences of chloroplast translation in a green alga and land plants. *Nat Plants* **4**, 564–575.

Verhounig, A., Karcher, D. and Bock, R. (2010) Inducible gene expression from the plastid genome by a synthetic riboswitch. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 6204–6209.

Wu, M., Dong, Y., Zhang, Q., Li, S., Chang, L., Loiacono, F.V., Ruf, S. *et al.* (2022) Efficient control of western flower thrips by plastid-mediated RNA interference. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2120081119.

Wurbs, D., Ruf, S. and Bock, R. (2007) Contained metabolic engineering in tomatoes by expression of carotenoid biosynthesis genes from the plastid genome. *Plant J.* **49**, 276–288.

Yamamoto, H., Wittek, D., Gupta, R., Qin, B., Ueda, T., Krause, R., Yamamoto, K. *et al.* (2016) 70S-scanning initiation is a novel and frequent initiation mode of ribosomal translation in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E1180–E1189.

Yu, Y., Yu, P.C., Chang, W.J., Yu, K. and Lin, C.S. (2020) Plastid transformation: how does it work? Can it be applied to crops? What can it offer? *Int. J. Mol. Sci.* **21**, 4854.

Yukawa, M. and Sugiura, M. (2008) Termination codon-dependent translation of partially overlapping *ndhC-ndhK* transcripts in chloroplasts. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 19550–19554.

Zghidi-Abouzid, O., Merendino, L., Buhr, F., Malik Ghulam, M. and Lerbs-Mache, S. (2011) Characterization of plastid *psbT* sense and antisense RNAs. *Nucleic Acids Res.* **39**, 5379–5387.

Zhang, J., Khan, S.A., Hasse, C., Ruf, S., Heckel, D.G. and Bock, R. (2015) Pest control. Full crop protection from an insect pest by expression of long double-stranded RNAs in plastids. *Science* **347**, 991–994.

Zhelyazkova, P., Hammani, K., Rojas, M., Voelker, R., Vargas-Suarez, M., Börner, T. and Barkan, A. (2012a) Protein-mediated protection as the predominant mechanism for defining processed mRNA termini in land plant chloroplasts. *Nucleic Acids Res.* **40**, 3092–3105.

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

Zhou, F., Badillo-Corona, J.A., Karcher, D., Gonzalez-Rabade, N., Piepenburg, K., Borchers, A.M., Maloney, A.P. *et al.* (2008) High-level expression of human immunodeficiency virus antigens from the tobacco and tomato plastid genomes. *Plant Biotechnol. J.* **6**, 897–913.

Zhou, F., Karcher, D. and Bock, R. (2007) Identification of a plastid intercistronic expression element (IEE) facilitating the expression of stable translatable monocistronic mRNAs from operons. *Plant J.* **52**, 961–972.

Zhou, W., Karcher, D., Fischer, A., Maximova, E., Walther, D. and Bock, R. (2015) Multiple RNA processing defects and impaired chloroplast function in plants deficient in the organellar protein-only RNase P enzyme. *PLoS One* **10**, e0120533.

Zoschke, R. and Bock, R. (2018) Chloroplast translation: structural and functional organization, operational control, and regulation. *Plant Cell* **30**, 745–770.

Zoschke, R., Watkins, K.P. and Barkan, A. (2013) A rapid ribosome profiling method elucidates chloroplast ribosome behavior *in vivo. Plant Cell* **25**, 2265–2275.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Data S1. Ribosome profiling and transcript profiling data and statistics.

Table S1. Primers used for genotyping and 16S rRNA gel blot analysis.

Appendix S1. Detailed information about *aadA* insertion sites of the analysed transplastomic lines.

Figure S1. Immunoblot analyses of the accumulation of the photosynthetic machinery and the ribosome. Immunoblot analyses of PsbA (photosystem II), PetB (cytochrome b6f complex), PSAD (photosystem I), AtpB (ATP synthase), and Rps15 (small subunit of the chloroplast ribo-some) in three independent biological replicates (Rep.; obtained from three individual plants). Note that a representative replicate is shown in Figure 1b. The Ponceau S-stained membrane is shown to illustrate the equal accumulation of RbcL. Other details are as described for Figure 1b. The quantification that is shown in Figure 1c is based on the here presented three biological replicates.

Figure S2. Reproducibility of transcript abundance and ribosome footprint data between representative biological replicates. Data were obtained from (a) WT, (b) ΔpetL, (c) ΔpsaJ, (d) Δrpl33, (e) pRB8c, and (f) RB70 samples. Large plots: average transcript accumulation (RNA, left) and ribosome footprint abundances (FP, right) for each plastid reading frame from the two biological replicates 1 and 2 (data provided in Data S1) were plotted in log10 scale against each other for any of the indicated genotypes. Pear-son's r and Anova's *p* values (in nEm non-superscript format for n•10m, providing the significance of the regression) are given in each of the diagrams. Small inset plots: transcript accumulation (RNA, left) and ribosome footprint abundances (FP, right) for each probe that is located in a plastid reading frame from the two biological replicates 1 and 2 (data provided in Data S1) were plotted in log10 scale against each other for each of the indicated genotypes. Pearson's *r* and Anova's *p* values (in nEm non-superscript format for n•10m, providing the significance of the regression) are given in each of the diagrams. For further comparisons of the three analysed independent biological repli-cates, see Data S1.