NOVEL TOOLS FOR PLASTID ENGINEERING IN HIGHER PLANTS

by

QIGUO YU

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Plant Biology

Written under the direction of

Dr. Pal Maliga

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

January 2020

ABSTRACT OF THE DISSERTATION

Novel Tools for Plastid Engineering in Higher Plants

By QIGUO YU

Dissertation Director:

Dr. Pal Maliga

Plastid genome engineering holds great promise in many biotechnological applications, including production of vaccines and antibodies, conferring insect tolerance by expressing dsRNAs, implanting novel metabolic pathways and improving photosynthetic efficiency and so on. However, the plastid transformation technology is not yet available in most crop plants.

I was part of the group that developed a system for plastid transformation in *Arabidopsis thaliana*, a Brassicaceae species, that is potentially applicable to the important oilseed crop canola (*Brassica napus*) and vegetable brassicas such as broccoli, cauliflower and cabbage (*Brassica oleracea*). Efficient plastid transformation is enabled by eliminating a duplicate fatty acid biosynthetic pathway in chloroplasts.

High level expression of recombinant proteins is desirable, but it may cripple the plant growth. To achieve on demand expression of plastid transgenes, I was involved in a project that exploited the use of an engineered PPR (pentatricopeptide)

RNA-binding protein for post-transcriptional regulation of chloroplast genes. One application is ethanol-inducible expression of chloroplast proteins by activating mRNA translation when the protein is needed. Thus, the transplastomic plants can be grown full size in the absence of the inducer and expression is turned on during the protein production phase. The second application is tuber-specific expression of plastid genes where plastid gene expression is normally very law. When the engineered PPR variant is expressed from a tuber-specific *patatin* promoter, a 60-fold increase over 0.02%, the maximum protein yield achieved to date, has been obtained in the potato tuber.

Plastids represent an appealing target for synthetic biology applications, where predictable protein output is of paramount importance. I designed synthetic operons from which reporter gene expression could be obtained in tobacco chloroplasts over a large dynamic range by post-transcriptional regulation, using the right combination of RNA binding proteins, RNA-binding protein binding sites and RNA processing elements.

ACKNOWLEDGEMENT

First of all, I would like to express my most profound appreciation to my project supervisor Prof. Pal Maliga. His continuous guidance and constructive criticism have helped me mature into an independent researcher rapidly. His expertise, immense knowledge, and insightful thoughts in our discussions have inspired me to continue pursuing science. His generosity to support me attend numerous conferences has broadened my perspective of research dramatically. It is my highest honor to join his lab.

I am also greatly indebted to our collaboration partner Prof. Alice Barkan. Her groundbreaking research in chloroplast biogenesis leads to a fantastic joint project that I am truly honored to participate in. Her dedication and brilliance motivate and encourage me to do better science unremittingly.

My heartfelt gratitude also goes to the members of the Maliga group. Special thanks to Dr. Zora Svab for organizing so many holiday parties (Easter, July 4th, Thanksgiving) and always prepared the fantastic food! Also, Lisa LaManna, Dr. Aki Matsuoka, Dr. Yuan Zhang, Dr. Massimo Bosacchi, Dr. Csanad Gurdon, Dr. Tarinee Tungsuchat-Huang for being such a great group to work with. All these funny, beautiful, and unforgettable time that I hang out with each of you during the

holiday celebrations, conference trips, daily lunch, already deeply embedded in my memory. We will keep in touch in the future!

I also would like to express my sincere gratitude to fellows from Dr. Juan Dong group, Dr. Andrea Gallavotti group, Dr. Joachim Messing group, for not only help me troubleshoot numerous technical problems but also provide me valuable suggestions in many aspects, from research thoughts to career development. I am lucky to have you guys being both my colleagues and friends.

A really huge thank you goes to the administration and technical staff (Maryann Bythell, Marylynn Bianca, Jill Wachter, Erin Sorge, Daja Bryant, Randall Newman, Shammara Scott, Nalini Kaul), and greenhouse staff (Themios Chionis) at Waksman Institute of Microbiology and Department of Plant Biology for providing necessary resources, and thus, ensure us throw ourselves into research.

I am deeply appreciated to my thesis committee, Dr. Andrea Gallavotti, Dr. Juan Dong, and Dr. Bryce Nickels for generously offering the time, discussions, guidance and, goodwill throughout the whole process.

Last but not least, I want to give my genuine gratitude to my family and friends. Without their unconditional love, encouragement, understanding, and support, I certainly cannot reach up to this point. Thank you for always being my strong spiritual support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

INTRODUCTION

Plant cells contain DNA in three cellular compartments: the nucleus, plastids, and mitochondria. Plastids derive from an endosymbiotic event that evolved by conversion of an engulfed cyanobacterium into a cellular organelle. Of the thousands of genes in the cyanobacterial ancestor, only approximately 100 genes are retained in the plastid genome. The plastid genetic system maintained many prokaryotic features. Now it depends on about ~3, 000 nuclear genes, the products of which are imported into the chloroplast.

The advantages of expressing transgenes in the plastid genome include[1,2]:

(1) High expression levels. Plastids organelles carry 1, 000 to 10, 000 copies of the ~150-kb circular genome. The target protein may accumulate up to about 70% of total soluble protein.

(2) Predictable gene expression. The target gene is integrated into the chloroplast genome by homologous recombination, thus transgenes from multiple events are expressed at the same level.

(3) Operons expression. Polycistronic chloroplast transcription units resemble bacterial operons, simplifying the incorporation of heterologous biochemical pathways.

(4) Transgene containment. In most species, the plastid genomes are transmitted maternally, providing a natural barrier to spreading transgenes by pollen.

(5) Plastid transformation enables the improvement of plant productivity by engineering the photosynthetic machinery, the production of pharmaceuticals and industry enzymes.

Plastid transformation is not available in *Arabidopsis thaliana*, the most advanced plant model species. The first objective of my research was to develop plastid transformation in *Arabidopsis thaliana*, so that basic research can benefit from the many nuclear mutants in studies of plastid function (Chapters 1 and 2). Constitutive high-level expression of transgenes is undesirable, because expression of the transgene places metabolic burden on the plants. The second objective of my research was to develop regulated chloroplast expression systems, so that gene expression is turned on when and where it is needed (Chapters 3 and 4). The third objective of my research was to develop regulatory elements to control the plastid gene expression across a dynamic range (Chapter 5 and 6). These new research tools will facilitate synthetic biology applications in higher plants.

Development of plastid transformation in Brassicaceae - Chapters 1 and 2

Plastid transformation is routine in tobacco (*Nicotiana tabacum*) but has been 100-fold less frequent in *Arabidopsis thaliana*, preventing its use in plastid biology. We hypothesized that plastid transformation efficiency should increase in the *acc2* background, because when ACC2 is absent, fatty acid biosynthesis becomes dependent on translation of the plastid-encoded ACC β-carboxylase subunit. I bombarded *ACC2*-defective Arabidopsis leaves with a vector carrying a

selectable spectinomycin resistance (*aadA*) gene and *gfp*, encoding the green fluorescence protein GFP and found that plastid transformation efficiency is increased 100-fold. Although I obtained transplastomic events, we did not obtain fertile transplastomic plants in the Slavice and Columbia *acc2* knockout backgrounds. The Wassiliewaskija (Ws-2) and RLD accessions are known to efficiently respond to growth regulators by shoot induction but are tolerant to the spectinomycin due to the presence of a functional ACC2. I generated *acc2* knockout line using CRISPR/Cas9 in Ws-2 and RLD and documented higher plastid transformation efficiency that their respective wild type backgrounds. The results of my research have been described in two publications[3,4].

Exploring the potential of PPR proteins for the design of regulatory systems for plastid transgenes – Chapters 3 and 4

Plastid organelles carry 1000 to 10000 copies of the ~150-kb circular genome. The target protein may accumulate up to about 20% of total soluble protein. However, constitutive expression of the transgene compromises the plant growth. The objective of this project is to develop a regulated chloroplast transgene expression system, so that the gene expression is turned on when it is needed. I employed a modified PPR10 RNA-binding protein that facilitates translation from a modified RNA binding site. In the absence of the modified PPR10* protein, the mRNA of interest is not translated, because the native PPR10 protein does not bind the mutant

target site. When the PPR10* is expressed from a nuclear gene, PPR10* binds to the mutant target site and the mRNA of interest is translated. To test the utility of this system, I expressed the modified PPR10* from a tuber specific promoter. When the PPR10 variant is expressed from the tuber-specific *patatin* promoter, GFP accumulated up to 1.3% of total soluble protein, a 60-fold increase over 0.02%, the maximum protein yield achieved to date in potato tuber. This regulatory system enables an increase in transgene expression in non-photosynthetic plastids without interfering with chloroplast gene expression in leaves [5]. Simultaneously, in collaboration with Dr. Alice Barkan's laboratory (University of Oregon), an analogous system was developed to control chloroplast transgene expression with an ethanol inducible promoter driving the engineered PPR10* expression in the nucleus. When expressed from an ethanol;-inducible nuclear transgene, the engineered PPR10 protein stimulates expression of chloroplast transgenes up to ~40-fold, with maximal protein abundance approaching that of Rubisco[6].

Dicistronic operons for predicable protein output in chloroplasts – Chapter 5 and 6

The plastid is becoming an appealing chassis for the synthetic biology applications. However, it is still lagging behand than that in bacteria or yeast. One of the major challenges standing in the way is the shortage of well-defined regulatory expression elements. I designed dicistronic operons in which a constant 1:1 mRNA stoichiometric ratio is maintained between the ORFs by expressing them on an

unprocessed mRNA. I found the protein accumulation is dependent on the translation efficiency of the individual ORFs[7]. Furthermore, we placed a series of *atpH* 5' UTRs in the intergenic region between the ORFs and obtained different levels of transgene expression, indicating that native chloroplast *cis*-elements can be utilized to regulate plastid transgene expression across large dynamic range in synthetic operons[8].

References

1   Maliga, P. in *Genomics of Chloroplasts and Mitochondria* Vol. 35 (eds R. Bock & V. Knoop) 393-414 (Springer, 2012).

2   Maliga, P. & Bock, R. Plastid biotechnology: food, fuel and medicine for the 21st century. *Plant Physiol.* **155**, 1501-1510, doi:pp.110.170969 [pii] 10.1104/pp.110.170969 (2011).

3   Yu, Q., LaManna, L., Kelly, M. E., Lutz, K. A. & Maliga, P. New Tools for Engineering the Arabidopsis Plastid Genome. *Plant Physiol.* **181**, 394-398 (2019).

4   Yu, Q., Lutz, K. A. & Maliga, P. Efficient plastid transformation in Arabidopsis. *Plant Physiol.* **175**, 186-193 (2017).

5   Yu, Q., Barkan, A. & Maliga, P. Engineered RNA-binding protein for transgene activation in non-green plastids. *Nat Plants* **5**, 486-490, doi:10.1038/s41477-019-0413-0 (2019).

6   Rojas, M., Yu, Q., Williams-Carrier, R., Maliga, P. & Barkan, A. Engineered PPR proteins as inducible switches to activate the expression of chloroplast transgenes. *Nat Plants* **5**, 505-511 (2019).

7   Yu, Q., Huang, T. T., Verma, K., Radler, M. R. & Maliga, P. Dicistronic operons as novel marker system for chloroplast engineering and building blocks for plant synthetic biology. *In preparation* (2019).

8   Yu, Q., Huang, T. T., Ioannou, A., Barkan, A. & Maliga, P. Posttranscriptional regulation tunes the output of cistrons in synthetic chloroplast operons. *In preparation* (2019).

CHAPTER 1

Efficient Plastid Transformation in *ACC2*-Defective Arabidopsis

Abstract

Plastid transformation is routine in tobacco, but 100-fold less frequent in Arabidopsis, preventing its use in plastid biology. A recent study revealed that null mutations in *ACC2*, encoding a plastid-targeted acetyl-CoA-carboxylase, cause hypersensitivity to spectinomycin. We hypothesized that plastid transformation efficiency should increase in the *acc2* background, because when ACC2 is absent, fatty acid biosynthesis becomes dependent on translation of the plastid-encoded ACC β-Carboxylase subunit. We bombarded *ACC2*-defective Arabidopsis leaves with a vector carrying a selectable spectinomycin resistance (*aadA*) gene and *gfp*, encoding the green fluorescence protein GFP. Spectinomycin resistant clones were identified as green cell clusters on a spectinomycin medium. Plastid transformation was confirmed by GFP accumulation from the second open reading frame of a polycistronic mRNA, that would not be translated in the cytoplasm. We obtained one to two plastid transformation events per bombarded sample in spectinomycin hypersensitive Slavice (Sav-0) and Columbia *acc2* knockout backgrounds, an approximately 100-fold enhanced plastid transformation frequency. Sav-0 and Columbia are accessions in which plant regeneration is uncharacterized or difficult to obtain. A practical system for Arabidopsis plastid transformation will be obtained by creating an *ACC2* null background in a regenerable Arabidopsis accession. The

recognition that the duplicated ACCase in Arabidopsis is an impediment to plastid transformation provides a rational template to implement plastid transformation in related recalcitrant crops.

Introduction

Plastids are semi-autonomous plant organelles with thousands of copies of the approximately 155-kb genome localized in 10 to 100 plastids per cell. The plastid genome of higher plants encodes about one hundred genes, the products of which assemble with approximately 3,000 nucleus-encoded proteins to form the plastid transcription and translation machinery and carry out complex metabolic functions, including photosynthesis, and fatty acid and amino acid biosynthesis. Plastid transformation is routine only in tobacco (*Nicotiana tabacum*) [1,2], but reproducible protocols for plastid transformation have also been described in tomato [3], potato [4], lettuce [5,6] and soybean [7]. Still, the technology is available in only a relatively small number of crops. *Arabidopsis thaliana*, the most widely used model plant is one of the species that is recalcitrant to plastid transformation. In *Arabidopsis*, only 2 transplastomic events were identified in 201 samples [8], a sample size that would have yielded approximately 200 events in tobacco using the technology available in 1998. Until now the reasons for the low efficiency in Arabidopsis were not understood.

Spectinomycin, the agent for selection of transplastomic events, binds to the 16S ribosomal RNA, blocking translation on the prokaryotic type 70S plastid ribosomes

[9,10] inhibiting greening and shoot regeneration in tissue culture cells [1]. When the plastid genome is transformed with the *aadA* gene encoding aminoglycoside-3"-adenylyltransferase (AAD), the AAD protein modifies the antibiotic such that it no longer binds to the 16S rRNA and translation proceeds, enabling greening. Tobacco, when cultured on a spectinomycin medium, bleaches and proliferates at a slow rate due to inhibition of plastid translation. Transplastomic tobacco cells are identified by the ability to green and regenerate shoots in spectinomycin-containing tissue culture medium. In contrast, *Brassica napus*, a close relative of *Arabidopsis thaliana*, bleaches but continues to proliferate as albino shoots on a spectinomycin medium in the absence of chloroplast ribosomes [11]. Two major studies by Parker et al. [12,13] revealed the existence of rare Arabidopsis accessions, in which plastids are extremely sensitive to spectinomycin. Seeds of most accessions in the study germinated on spectinomycin and developed into small albino seedlings or rosettes, including RLD (Reduced number of Long Days, Rschew), the accession used in the 1998 plastid transformation experiment. However, seeds from hypersensitive accessions germinated but did not develop beyond the cotyledonary stage. Genetic analysis revealed that spectinomycin sensitivity of hypersensitive accessions is due to mutations in the *ACC2* nuclear gene. The *ACC2* gene produces the homomeric acetyl-CoA-carboxylase (ACCase) that is imported into plastids, and in tissue culture partially duplicates the function of heteromeric ACCase, one subunit of which is encoded in the plastid *accD* gene (Fig. 1-1A). When plastid translation is blocked by spectinomycin, no heteromeric ACCase is made, and the homomeric

enzyme enables a limited amount of fatty acid biosynthesis to occur, thereby reducing the impact of the absence of heteromeric enzyme in culture, making spectinomycin selection inefficient. In the absence of a functional *ACC2* gene, fatty acid biosynthesis is dependent on the availability of heteromeric ACCase enzyme, the β-carboxylase subunit of which is translated on plastid ribosomes (Fig. 1-1B). We hypothesized that inefficient plastid transformation in our original study was due to tolerance of *Arabidopsis* to spectinomycin and that transformation of hypersensitive mutants defective in *ACC2* function should result in efficient recovery of transplastomic clones. We report here that the efficiency of plastid transformation in the *acc2* background is increased approximately 100-fold and comparable to that of tobacco, confirming our hypothesis.

Results

Plastid Transformation with vector pATV1 and Identification of Transplastomic Events

The plastid transformation vector pATV1 targets insertion upstream of the *trnV* gene in the inverted repeat region of the plastid genome (Fig. 1-2A). Vector pATV1 carries a dicistronic operon, in which the first open reading frame (ORF) encodes the *aadA* spectinomycin resistance gene and the second ORF encodes a green fluorescence protein (GFP) (Fig. 1-2A). Polycistronic mRNAs are not translated on the eukaryotic-type 80S ribosomes in the cytoplasm, thus accumulation of GFP in chloroplasts in spectinomycin-resistant clones indicates plastid transformation.

Plastid transformation was carried out in the Col-0 (Columbia) accession and the Columbia *ACC2* T-DNA insertion line *acc2-1* (SALK_148966C), which was shown by Parker et al. [12] to be hypersensitive to spectinomycin. We also evaluated plastid transformation efficiency in the Sav-0 (Slavice) accession that was the most sensitive to spectinomycin in the study of Parker et al. [12]. The Sav-0 *ACC2* gene carries 15 missense mutations, however the hypersensitivity to spectinomycin is thought to be due to one mutation (G135E) that alters a conserved residue immediately preceding the biotin carboxylase domain [13]. Plants were grown aseptically on Arabidopsis Revised Medium with 5% sucrose (ARM5 medium) (Fig. 1-3A); leaves for plastid transformation were harvested from plants grown under aseptic conditions and placed on ARMI media (see Methods). The leaf tissue was bombarded with gold particles coated with vector DNA. After two days, the leaves were stamped with a stack of razor blades, cut into 1 cm^2 pieces and transferred onto the same medium (ARMI) containing spectinomycin (100 mg/L; Fig. 1-3B) to facilitate preferential replication of plastids containing transformed ptDNA copies. The ARMI medium induces division of the leaf cells and formation of colorless, embryogenic callus. After 7-10 days of selection on ARMI medium, spectinomycin selection was continued on the ARMIIr medium, which induces greening. Since spectinomycin prevents greening of wild-type cells, only spectinomycin-resistant cells formed green calli. Visible green cell clusters appeared within 21 to 40 days

on the selective ARMIIr medium (Fig. 1-3C). Illumination of plates with UV light revealed intense fluorescence of GFP in the green calli (Fig. 1-3D).

In the wild-type Col-0 sample (4 bombarded plates), no transplastomic event was found. We obtained 8 events in 5 bombarded plates using leaf tissue in the *acc2-1* mutant background and 4 events in 4 bombarded plates in the Sav-0 accession (Table 1-1). This transformation efficiency is comparable to the transformation efficiency obtained with current protocols in tobacco, 4 to 5 transplastomic events per bombardment [14].

This is a significant advance, as high frequency plastid transformation in Arabidopsis has been pursued since the publication of the original report in 1998 [8]. Since 2007, 26 plates of RLD and 5 plates of *Landsberg erecta* (Ler) leaf tissue were bombarded; none of which yielded a transplastomic event (Table 1-1). In contrast, 9 bombardments of leaves with the *acc2* null background yielded 12 transplastomic clones. Even though the technology significantly improved since 1998, no transplastomic clones were obtained until *ACC2* defective leaf tissue was used for bombardments (Table 1-1), providing overwhelming support for the absence of ACC2 activity being critical for high frequency plastid transformation in *Arabidopsis thaliana*.

Confocal Microscopy to Confirm Transplastomic events

GFP is encoded in the second ORF, thus GFP accumulation is expected only if the mRNA is translated in plastids on the prokaryotic type 70S ribosomes known to

translate polycistronic mRNAs [15]. Thus, GFP accumulation was anticipated only if the *gfp* gene is expressed in chloroplasts.

Putative transplastomic lines were identified by green cell cluster formation and were confirmed as transplastomic events by detecting localization of GFP to plastids by confocal microscopy (Fig. 1-4). Overlay of the GFP and chlorophyll channels indicates that the clones are heteroplastomic, carrying transformed and wild type plastids in the same cells. A good example for mixed plastids is shown in the overlay of GFP and chlorophyll channels in Col-0 *acc2-1* #3 in Fig. 1-4. The chloroplasts were not well developed in most tissue culture cells. Chlorophyll was detected in only a localized region of plastids in line with thylakoid biogenesis initiating from a localized center [16]. Good examples are overlays of Col-0 *acc2-1* #5 and Sav-0 #1 in Fig. 1-4.

The heteroplastomic state detected in the cells of the green clusters was not maintained, and eventually wild-type plastids (ptDNA) disappeared in the callus cells after continued cultivation on selective media. The homoplastomic state is confirmed by uniform accumulation of GFP in the leaves of a Sav-0 #6 plant shown in Fig. 1-4 and DNA gel blot analyses of calli shown in Fig. 1-2B.

Regeneration of Transplastomic Sav-0 Plants and Transmission of GFP to Seed Progeny

After bombardment of Col-0 and Sav-0 leaves, selection of transplastomic events was carried out according to the published RLD protocol [8]. However, when the

transplastomic clones were transferred to the RLD shoot-induction medium (ASI-N1B1 medium), the calli did not proliferate. Therefore, we transferred the transplastomic calli to media that were successfully used to regenerate plants from other accessions. We found that the two-step regeneration protocol described for shoot induction in the C24 background [17] triggered shoot regeneration in two surviving Sav-0 calli. Calli of Sav-0 transplastomic lines #3 and #6 were briefly (3 days) exposed to callus induction medium containing 0.5 mg/L 2,4-D and 0.05 mg/L kinetin, and then transferred to a shoot regeneration medium containing 0.15 mg/L IAA and 1.6 mg/L phenyl-adenine. Phenyl-adenine is a potent compound for shoot regeneration through inhibition of CYTOKININ OXIDASE/DEHYDROGENASE activity [17]. Shoots from the calli developed in 45 to 60 days; flowered and formed siliques in sterile culture in 250 ml Erlenmeyer flasks (Fig. 1-3E). The plants intensely glow when illuminated with UV light, indicating high-level GFP accumulation (Fig. 1-3F). Confocal microscopy suggests uniform transformation of plastid genomes in the leaves of regenerated Sav-0 #6 plants (Fig. 1-4) and was confirmed by molecular analyses (Fig. 1-2B).

The transplastomic shoots were transferred to larger 500 ml Erlenmeyer flasks containing ARM medium for seed set where they continued to grow. It is noteworthy that the shoots did not have any roots or rosette leaves, thus they could be best described as inflorescence cultures. The siliques harvested from the Sav-0 #6 plants were empty, while the Sav-0 #3 shoots produced six seeds. One

transplastomic Sav-0 #3 seed germinated on spectinomycin. The cotyledons of this seedling fluoresce under UV light, indicating GFP accumulation (Fig. 1-3G).

Molecular Analysis of Transplastomic Arabidopsis Clones

DNA and RNA gel blot analyses was carried out on the callus and shoots of Sav-0 transplastomic lines #3 and #6. Wild type plastids present in the cells of the green clusters were gradually lost by the time DNA gel blot analyses was carried out, confirming uniform transformation of the plastid genomes in both calli and shoots (Fig. 1-2B). RNA gel blot analyses indicate the presence of a 2.0-kb dicistronic transcript, detected by both the *aadA* and *gfp* probes (Fig. 1-2C).

Discussion

Development of a Plastid Transformation Protocol in Arabidopsis

We report here approximately 100-fold enhanced plastid transformation efficiency per bombardment in the *acc2* null background: 8 events in five bombarded samples in the Col-0 *acc2-1* line and 4 events in four bombarded samples in the Sav-0 background. The increase from 1 event per approximately 100 bombardments to 1 event per one bombardment is in part due to technological advances. However, the lack of success with the latest technology in a large number of bombarded samples (Table 1-1) provides overwhelming evidence that the key to success was the choice of Arabidopsis lines lacking ACC2 activity.

Identification of transplastomic events in the RLD ecotype took 5 to 12 weeks in 1998 [8]. Use of spectinomycin sensitive *acc2* knockout lines and the pATV1 dicistronic operon vector shortened the time period for identification of transplastomic events to 3 to 5 weeks. Use of the *acc2* knockout lines shortened scoring because proliferation of non-transformed cells was efficiently inhibited by spectinomycin, enabling identification of the spectinomycin resistant green cell clusters. Spectinomycin resistance may be due to integration of *aadA* in the plastid genome, integration of *aadA* in the nuclear genome and fortuitous expression from an upstream promoter, or spontaneous mutations in the *rrn16* gene [2]. GFP, encoded in the second ORF is expressed only in the chloroplasts, enabling rapid identification of transplastomic clones in a small number of heteroplastomic cells by confocal microscopy.

Once transplastomic clones are identified, the next major step is plant regeneration. There is diversity for shoot regeneration potential in Arabidopsis accessions. Columbia is well known for its recalcitrance to shoot regeneration from cultured cells. Therefore, no attempt was made to regenerate shoots from the Col-0 transplastomic callus tissue. There is no information about the tissue culture properties of the Sav-0 accession. Our first attempts at Sav-0 shoot regeneration from the transplastomic clones proved successful, yielding flowering shoots in culture (Fig. 1-3E). However, the seeds, with one exception, failed to germinate. Seed viability was apparently compromised by somaclonal variation, accumulated genetic changes due to the tissue remaining in culture for close to a year's time [18].

The first step towards obtaining a system that yields fertile transplastomic Arabidopsis will be obtaining *ACC2* null mutations in regenerable accessions. Shoot regeneration protocols have been worked out from root [19] and leaf explants [20] of the RLD accession; and from protoplasts [21], leaf explants [22] and inflorescence stem explants [23] of the Wassilewskya (Ws) accession. Thus, RLD and Ws will be our targets for *ACC2* mutagenesis.

Plastid Transformation in Arabidopsis Provides Template for Recalcitrant Crops

The recognition that the duplicated ACCase in Arabidopsis is an impediment to plastid transformation provides a rational template to implement plastid transformation in all Arabidopsis accessions, and in crops having a plastid-encoded *accD* gene and a plastid targeted ACC2 enzyme. The *Arabidopsis thaliana* ACC2 enzyme has an N-terminal extension compared to ACC1 (Fig. 1-5). The N-terminal extension is a plastid targeting sequence, as shown by subcellular localization of a GFP fusion protein [24]. The *ACC1* and *ACC2* genes are present in most Brassicaceae species, including *Arabidopsis lyrata*, *Camelina sativa*, *Camelina rubella*, *Brassica oleracea*, *Brassica napus* and *Brassica rapa*. The homomeric ACC2 enzyme in these species has an N-terminal extension compared to ACC1 (Fig. 1-5)[25]. Thus, a targeted mutation in the ACC2 N-terminal extension should create a spectinomycin hypersensitive variant. Plastid transformation has been achieved in cabbage (*Brassica oleracea* L. var. *capitata* L.). Thus, knockout of *ACC2* is apparently not

necessary to obtain transplastomic events in this crop, at least in the two cultivars tested [26,27]. Plastid transformation in cauliflower (*Brassica oleracea* var. *botrytis*) has been obtained at a very low frequency [28]. Plastid transformation in oilseed rape (*Brassica napus*) has also been obtained, but no homoplastomic plants could be obtained [29,30]. Plastid transformation in *Lesquerella fendleri*, another oilseed crop in the Brassicaceae, was feasible but inefficient [31]. Deletion of *ACC2* in the latter cases is expected to boost plastid transformation efficiency.

Materials and Methods

Tissue Culture Media

The tissue culture media were adopted from Sikdar *et al.* [8], originally described by Marton and Browse [19]. The culture media are based on MS salts [32]. Arabidopsis Revised Medium (ARM): MS salts, 3% sucrose, 0.8% agar (A7921; Sigma, St. Louis, MO), 200 mg myo-inositol, 0.1 mg biotin (1 mL of 0.1 mg/mL stock) and 1 mL vitamin solution (10 mg vitamin B1, 1 mg vitamin B6, 1 mg nicotinic acid, 1 mg glycine per mL) per liter, pH 5.8. ARM5 medium: ARM medium supplemented with 5% sucrose. ARMI medium: ARM medium containing 3 mg indolacetic acid (IAA), 0.6 mg benzyladenine (BA), 0.15 mg 2,4-dichlorophenoxyactetic acid and 0.3 mg isopentenyladenine (IPA) per liter. ARMIIr medium: ARM medium supplemented with 0.2 mg/L naphthaleneacetic acid and 0.4 mg isopentenyladenine per liter. The stocks of filter sterilized plant hormones and antibiotics (100 mg/L spectinomycin HCl) were added to media cooled to 45°C after autoclaving.

Shoot regeneration in the transplastomic Sav-0 clones was obtained on an ARM medium containing 2,4-D (0.5 mg/L), kinetin (0.05 mg/L) and spectinomycin (100 mg/L)(3 days) followed by incubation on an ARM medium containing IAA (0.15 mg/L), Phenyl-Adenine (1.6 mg/L) and spectinomycin (100 mg/L) [17]. Seed was obtained by growing shoots on MS salt medium containing 3% sucrose, 0.8 % agar (A7921; Sigma, St. Louis, MO), pH 5.8.

Plant Materials and Growth Conditions

The Sav-0 (CS28725) and Columbia homozygous *acc2-1* knockout line (SALK_148966C) seeds were obtained from The Arabidopsis Biological Resource Center, The Ohio State University. The Columbia (Col) seed was obtained from Prof. Juan Dong, Rutgers University. The RLD and Landsberg erecta (Ler) seed was purchased form Lehle Seeds, Round Rock, TX.

For surface sterilization, seeds (25 mg) were treated with 1.7% sodium hypochlorite (5x diluted 8.5% commercial bleach) in a 1.5 ml Eppendorf tube for 15 min. with occasional mixing (Vortex). The bleach was removed by pipetting and washed 3x with sterile distilled water. Seeds were germinated on 50 ml ARM5 medium in deep petri dishes (20 mm high, 10 cm in diameter). The plates were illuminated for 8 hours using cool-white fluorescent tubes (2,000 lux). The seeds germinated after 10-15 days of incubation at 24°C. To grow plants with larger leaves, seedlings were individually transferred to ARM5 plates (4 plants per deep petri dish). The plates were illuminated for 8 hours with cool-white fluorescent

bulbs (2000 lux) and incubated at 21°C during the day and 18°C during night. 1 to 2 cm long, dark green leaves were harvested for bombardment after incubation for an additional 5 to 6 weeks.

Plastid Transformation and Selection of Transplastomic Lines

The plastid transformation vector pATV1 reported here targets insertion in the inverted repeat region of the plastid genome upstream of the *trnV* gene (Fig. 1-2). The DNA sequence was deposited in GenBank under accession no. MF461355. The pAAK176 and pTT626 plastid transformation vectors share the plastid-targeting region with vector pATV1. The pAAK176 vector carries the Prrn:LrbcL:*aadA*:TpsbA marker gene present in vector pHK34 [33]. The *aadA* gene is between two *loxP* sites which facilitate the excision of the marker gene, leaving behind a *loxP* target site [34]. The pTT626 plastid transformation vector encodes *aadA-gfp* fusion protein [35] in PrrnLcry/TpsbA cassette [36].

Plastid transformation in Arabidopsis was carried out using our 1998 protocol, as shown in Fig. 1-3 [8]. The leaves (each 10 to 20 mm) were harvested from aseptically grown plants and covered the surface of agar-solidified ARMI medium in a 10 cm petri dish. We used 100 to 120 leaves to cover the surface of the plate. The leaves were cultured for 4 days on ARMI medium, then bombarded with pATV1 vector DNA. Transforming DNA was coated on the surface of microscopic (0.6 μm) gold particles, then introduced into chloroplasts by the biolistic process (1,100 psi) using

a helium-driven PDS1000/He biolistic gun equipped with the Hepta-adaptor [37]. The plates were placed on the shelf at the lowest position for bombardment.

Following bombardment, the leaves were incubated for two additional days on ARMI medium. After this time period, the leaves were stamped with a stack of 10 razor blades to create parallel incisions 1 mm apart. The stamped leaves were cut into smaller (1 cm^2) pieces and transferred onto the same medium (ARMI) containing 100 mg/L spectinomycin and incubated at 28°C illuminated for 16 hours with fluorescent tubes (CXL F025/741). After 8 to 10 days, the leaf strips were transferred onto selective ARMIIr medium containing 100 mg/L spectinomycin for the selection of spectinomycin resistant clones. The leaf strips were transferred to a fresh selective ARMIIr medium every two weeks until putative transplastomic clones were identified as resistant green calli.

Confocal Microscopy to Detect GFP in Plastids

Subcellular localization of GFP fluorescence was determined by a Leica TCS SP5II confocal microscope. To detect GFP and chlorophyll fluorescence, excitation wavelengths were at 488 nm and 568 nm, and the detection filters were set to 500-530 nm and 650-700 nm, respectively.

DNA and RNA Gel Blot Analyses

Total leaf DNA was prepared by the cetyltrimethylammonium bromide (CTAB) protocol [38]. DNA gel blot analyses was carried out as described [2]. Total cellular

DNA was digested with the *EcoR*I restriction enzyme. The DNA probe was the *ApaI-SphI* ptDNA fragment encoding the plastid *rrn16* gene (Fig. 2).

Total cellular RNA was isolated from leaves frozen in liquid nitrogen using TRIzol (Ambion/Life technologies), following the manufacturer's protocol. RNA gel blot analyses was carried out as described [33]. The probes were: *aadA*, 0.8-kb *Nco*I-*Xba*I fragment isolated from plasmid pHC1 [39]; *gfp*, fragment amplified from *gfp* coding region using primers *gfp*-forward p1(5'-TTTTCTGTCAGTGGAGAGGGTG-3') and *gfp*-reverse p2 (5'-CCCAGCAGCTGTTACAAACT-3'; Fig. 2).

Alignment of Homomeric ACCases

Alignment of homomeric ACCases in the Brassicaceae family was carried out with the MultAlin software [40].

Accession Numbers

DNA sequence of the pATV1 Arabidopsis plastid transformation vector is deposited in GenBank under accession number MF461355.
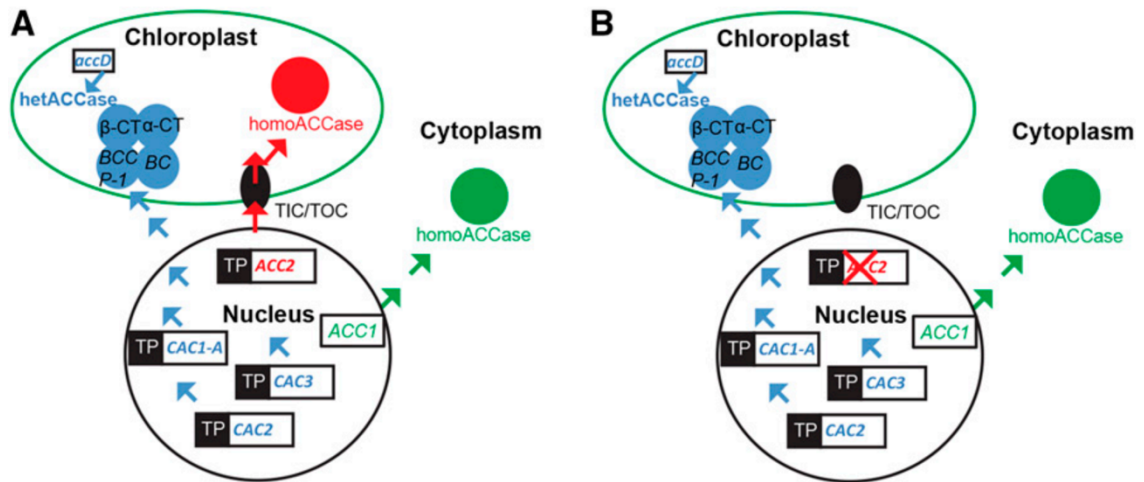
Figure 1-1. Elimination of ACC2 function makes plastid transformation efficient in Arabidopsis. A, Heteromeric ACCase (hetACC) localizes in the chloroplast and is encoded by nuclear genes *CAC1-A* (At5g16390; Biotin Carboxyl Carrier Protein 1 (BCCP-1)), *CAC1-B* (At5g15530; Biotin Carboxyl Carrier Protein 2 (BCCP-2)) (not depicted), *CAC2* (At5g35360; Biotin Carboxylase (BC)), *CAC3* (At2g38040, α subunit of Carboxyltransferase (α-CT) and the plastid encoded gene *accD* (AtCg00500; (β subunit of Carboxyltransferase (β-CT)). The homomeric ACC1 (At1g36160; homACCase) enzyme localizes in the cytoplasm and the ACC2 (At1g36180; homACCase) enzyme is imported into the chloroplast via the TIC/TOC membrane protein complex. If translation of the plastid *accD* mRNA is blocked by spectinomycin, the nuclear homomeric *ACC2* gene enables a limited amount of fatty acid biosynthesis, thereby reducing the impact of the absence of heteromeric enzyme in culture, making spectinomycin selection inefficient. B, In hypersensitive *acc2* mutants, the absence of the homomeric ACCase makes the cells

dependent on plastid translation to produce the heteromeric ACCase enzyme for fatty acid biosynthesis.

Figure 1-2. Molecular characterization of the Sav-0 transplastomic clones. A, Map of the plastid genome with the integrated *aadA-gfp* dicistronic operon. The *Nru*I-*Xba*I region is contained in the plastid transformation vector pATV1. P and T mark the position of PrrnLatpB promoter and TpsbA terminator in the dicistronic vector. The black box at the *aadA* N-terminus marks the *atpB* Downstream Box sequence [33]. The ribosome entry site is marked by black semi-ovals. Positions of the *rrn16* and *trnV* plastid genes and relevant restriction enzyme sites are marked. Thick black and red lines indicate probes used for DNA and RNA gel blot analysis, respectively.

B, DNA gel blot using the *rrn16* probe (Fig. 1-2A) indicates that the transplastomic

Sav-0 calli and leaves are homoplastomic, carrying only the 4.7-kb *Eco*RI fragment

and lacking the 2.7-kb wild type fragment. C, RNA gel blot analyses using both the

aadA and *gfp* probes (Fig. 1-2A) recognize the same 2 kb dicistronic mRNA.
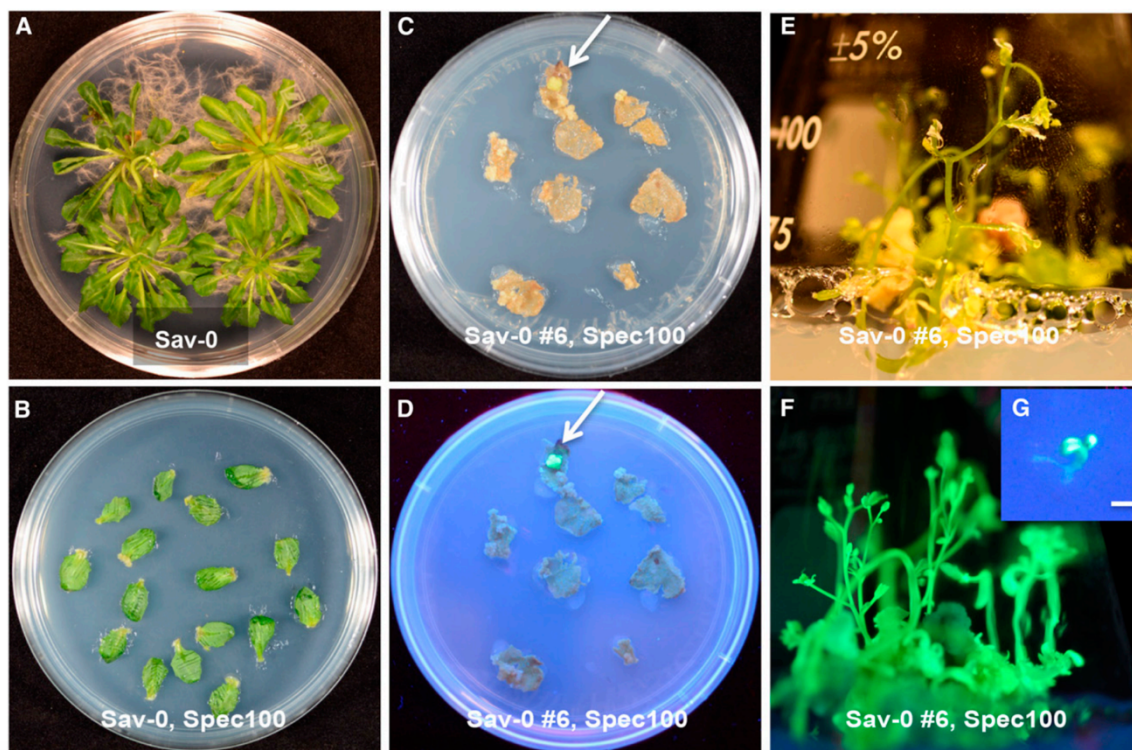
Figure 1-3. Identification of Arabidopsis transplastomic clones. A, Sterile Sav-0 plants grown in Petri dishes (diameter 10 cm) for six weeks. B, Two days after bombardment the Sav-0 leaves are incised and transferred to selective spectinomycin (100 mg/L) medium. C, Sav-0 leaves on selective medium one month after bombardment. Note scanty callus formation and green cell cluster (arrow). D, Culture shown in Fig. 1-3C, illuminated with UV light. Note green fluorescence indicating GFP accumulation in green cell cluster. E, Sav-0 plant regenerated from a transplastomic clone #6 in sterile culture. F, Culture shown in Fig. 1-3E, illuminated with UV light. G, Sav-0#3 seed progeny illuminated with UV light. The bar is 1 mm.
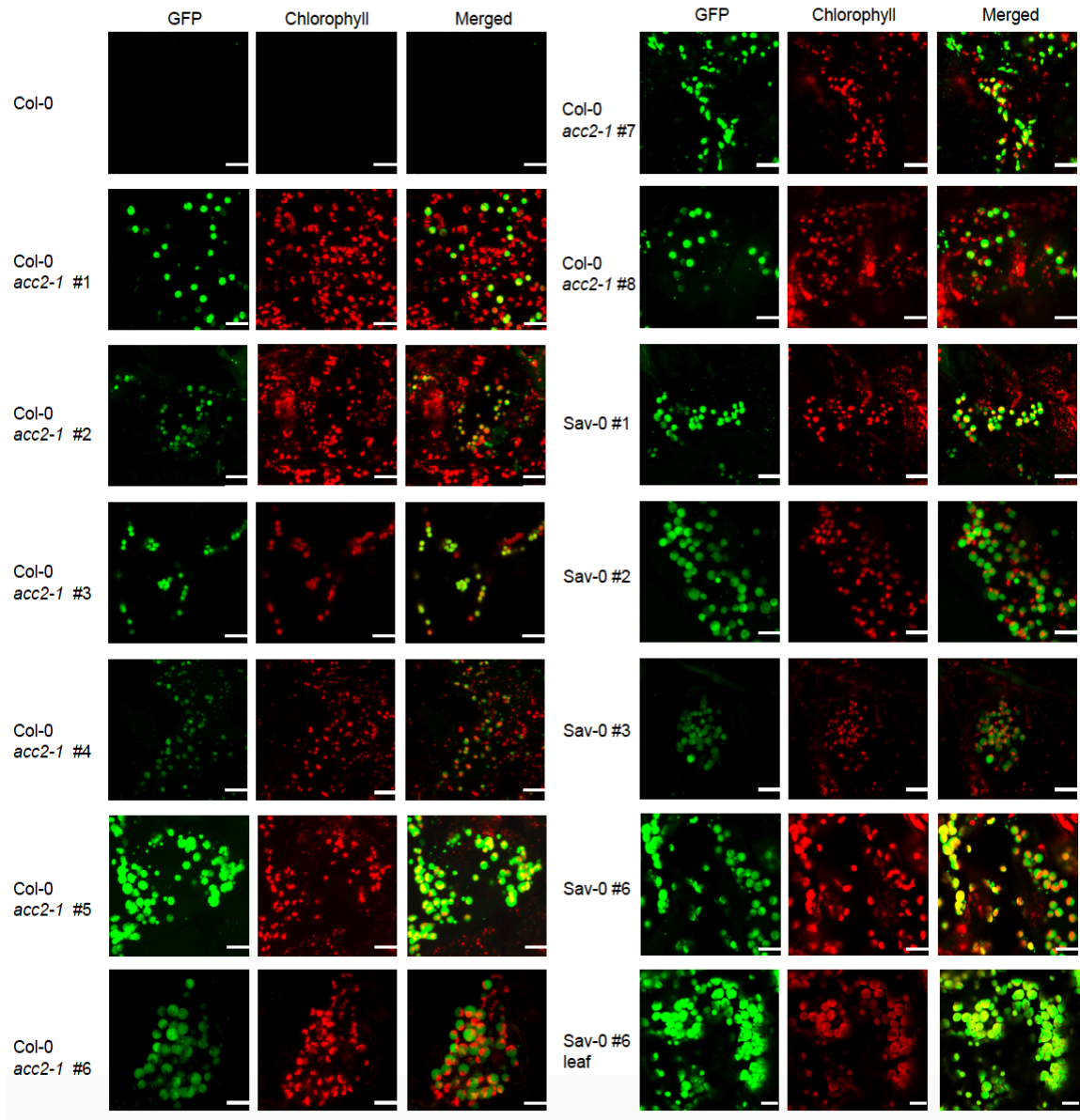
Figure 1-4. The green fluorescent protein (GFP) accumulates in chloroplasts.

Shown are confocal images collected in the GFP, chlorophyll, and merged channels

on a Leica TCS SP5II confocal microscope. Excitation wavelengths were at 488 nm

and 568 nm, detection at 500-530 nm and 650-700 nm, respectively. Note absence

of GFP and chlorophyll in the wild-type Col-0 callus cells and mixed GFP

expressing transgenic and wild type plastids in the Col-0-*acc2-1*#1 and Sav-0 #6 lines. Note the absence of wild type plastids in the leaves of Sav-O#6 plants. Yellow color in the merged images indicates co-localization of GFP to chlorophyll in plastids. Bar represents 10 μm. Note that cells in the small green cell clusters are heteroplastomic. The only exception are cells in Sav-0#6 leaves, which are homoplastomic due to prolonged selection in tissue culture.

A

```
             1        10        20        30        40        50        60        70        80        90       100
             |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
A.thalianaACC1
A.thalianaACC2 MEMRALGSSCSTGNGGSAPITLTNISPWITTVFPSTVKLRSSLRTFKGVSSRVRTFKGVSSTRVLSRTKQQFPLFCFLNPDPISFLENDVSEAERTVVLP
Consensus      ..........................................................................................................
```

```
             101      110       120       130       140       150       160       170       180       190       200
             |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
A.thalianaACC1     MAGSVNGNHSAVGPGINYETVSQVDEFCKALRGKRPIHSILIANNGMAAVKFIRSVRTWAYETFGTEKAILLVGMATPEDMRINAEHIRIADQFV
A.thalianaACC2 DGSVNGAGSVNGYHSDVVPGRN---VAEVNEFCKALGGKRPIHSILVATNGMAAVKFIRSVRTWAYETFGSEKAVKLVAMATPEDMRINAEHIRIADQFV
Consensus      .....gAGSVNGnHSaVgPGrN...Va#V#EFCKALrGKRPIHSIL!AnnGMAAVKFIRSVRTWAYETFGsEKA!lLVaMATPEDMRINAEHIRIADQFV
```

B

```
             1        10        20        30        40        50        60        70        80        90       100
             |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
A.thalianaACC1 MAGSVNGNHSAVGPGINYETVSQVDEFCKALRGKRPIHSILIANNGMAAVKFIRSVRTWAYETFGTEKAILLVGMATPEDMRINAEHIRIADQFVEVPGG
A.lyrataACC1   MAGSVNGYQSAIGPGINYETVSQVDEFCKALGGKRPIHSILIANNGMAAVKFIRSVRTWAYETFGTEKAILLVGMATPEDMRINAEHIRIADQFVEVPGG
C.sativaACC1-1 MAGSVNGYQSAVGPGINYESVSQVDEFCKALGGKRPIHSILIANNGMAAVKFIRSVRTWAYETFGTEKAILLVGMATPEDMRINAEHIRIADQFVEVPGG
C.rubellaACC1  MAGSVNGYQSSYGPGINYETVSQVDEFCKSLGGKRPIHSILIANNGMAAVKFIRSVRTWAYETFGTEKAILLVGMATPEDMRINAEHIRIADQFVEVPGG
B.oleraceaACC1 MAGSVNGYQT---PGRNHVSVSEVDDFCIALGGKRPIHSILIANNGMAAVKFIRSVRTWAYETFGTERAILLVGMATPEDMRINAEHIRIADQFVEVPGG
B.napusACC1-1  MAGSVNGYQT---PGRNHVSVSEVDDFCIALGGKRPIHSILIANNGMAAVKFIRSVRTWAYETFGTERAILLVGMATPEDMRINAEHIRIADQFVEVPGG
B.rapaACC1     MAGSVNGYQT---PGRNHVSVSEVDDFCIALGGKRPIHSILIANNGMAAVKFIRSVRTWAYETFGTERAILLVGMATPEDMRINAEHIRIADQFVEVPGG
C.sativaACC1-2 MAGSVNGYQSAVGPGINYETVSQVDEFCKALGGNRPIHSILIANNGMAAVKFIRSVRTWAYETFGTEKAILLVGMATPEDMRINAEHIRIADQFVEVPGG
B.napusACC1-2  MAGSVNGYQT---PGRNHVSVSEVDDFCIALGGKRPIHSILIANNGMAAVKFIRSVRTWAYETFGTERAILLVGMATPEDMRINAEHIRIADQFVEVPGG
Consensus      MAGSVNGYQs..gPGiNyesVS#VD#FCkALGGkRPIHSILIANNGMAAVKFIRSVRTWAYETFGTEkAILLVGMATPEDMRINAEHIRIADQFVEVPGG
```

```
             101      110       120       130       140       150       160       170       180       190       200
             |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
A.thalianaACC1 TNNNNYANVQLIVEMAEVTRVDAVWPGWGHASENPELPDALDAKGIIFLGPPASSMAALGDKIGSSLIAQAADVPTLPWSGSHVKIPPNSNLVTIPEEIY
A.lyrataACC1   TNNNNYANVQLIVEMAEVTRVDAVWPGWGHASENPELPDALDAKGIIFLGPPAASMAALGDKIGSSLIAQAADVPTLPWSGSHVKMPPNSNLVTIPEEIY
C.sativaACC1-1 TNNNNYANVQLIVEMAEVTRVDAVWPGWGHASENPELPDALDAKGIIFLGPPAASMGALGDKIGSSLIAQAADVPTLPWSGSHVKIPPNSNLVTIPEEIY
C.rubellaACC1  TNNNNYANVQLIVEMAEVTRVDAVWPGWGHASENPELPDALDAKGIIFLGPPAASMAALGDKIGSSLIAQAADVPTLPWSGSHVKIPPNSNLVTIPEEIY
B.oleraceaACC1 TNNNNYANVQLIVEMAEVTRVDAVWPGWGHASENPELPDALKAKGIIFLGPPAASMAALGDKIGSSLIAQAADVPTLPWSGSHVKIPPDSSLVTIPEEIY
B.napusACC1-1  TNNNNYANVQLIVEMAEVTRVDAVWPGWGHASENPELPDALKAKGIIFLGPPAASMAALGDKIGSSLIAQAADVPTLPWSGSHVKIPPDSSLVTIPEEIY
B.rapaACC1     TNNNNYANVQLIVEMAEVTRVDAVWPGWGHASENPELPDALKAKGIIFLGPPAASMAALGDKIGSSLIAQAADVPTLPWSGSHVKIPPDSSLVTIPEEIY
C.sativaACC1-2 TNNNNYANVQLIVEMAEVTRVDAVWPGWGHASENPELPDALDAKGIIFLGPPAASMGALGDKIGSSLIAQAADVPTLPWSGSHVKIPPNSNLVTIPEEIY
B.napusACC1-2  TNNNNYANVQLIVEMAEVTRVDAVWPGWGHASENPELPDALKAKGIIFLGPPAASMAALGDKIGSSLIAQAADVPTLPWSGSHVKIPPDSSLVTIPEEIY
Consensus      TNNNNYANVQLIVEMAEVTRVDAVWPGWGHASENPELPDALdAKGIIFLGPPAASMaALGDKIGSSLIAQAADVPTLPWSGSHVKIPP#SnLVTIPEEIY
```

C

```
             1        10        20        30        40        50        60        70        80        90       100
             |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
A.thalianaACC2 MEMRALGSSCSTGNGGSAPITLTNISPWITTV----------FPSTVKLRSS--LRTFKGVSSRVRTFKGVSSTRVLSRTKQQF-PLFCFL-----NPDP
A.lyrataACC2   MEMRALVSSCATGNGGSDPFSFTKVSPWITTVGG----KDRDFPTTVKLRTS--MRTFKGVSIRGRTFKGVS-TRVLSRNKQQF-PLFCFL-----NPDP
C.sativaACC2-1 MEMRALVSSYSTGNGGSDPISLTNGSPWITTVGGGASTMDREFPLTVKLGSS--MRAFKGVST----------TTVLSRTKQQF-PLVCLARNNANSTDP
C.sativaACC2-2 MEMRALVSSYSTGNGGSDPISLTNGSPWITTVGGGASTMDREFPATVKLGSS--MRAFKGVST----------ITVLSRTKQQF-PLVCLARNNGNSTDP
C.rubellaACC2  MEMRALVSSCSTGNGGSDPISLTNVSPWITTVGGGASSIDREFPTTVKLGSS--LRTFKGVSS----------TTVLSRTKQQF-PLVCLARNNANSTDP
B.oleraceaACC2 MEMRALVSCSAAGNGASDRFRLSNVSPWITSARGASGS---DSPATVKLRSSSMIRAFKGVSI----YKNKTRRNVLSQRNKQFRPMAYLGRKDLSSPDP
B.napusACC2-2  MEMRALVSCSAAGNGASDRFRLSNVSPWITSARGASGS---DSPATVKLGSSSMIRAFKGVSI----YKNKTRRNVLSQRNKQFRPMAYLGRKDLSSPDP
B.rapaACC2     MEMRALVSCSAAGNGASDRFRLSNVSPWITSARGASGS---DSPATVKLGSSSMIRAFKGVSI----YKNKSRRNVLSQRNKQFRPMAYLGRKDLSSPDP
B.napusACC2-1  MEMRALVSCSAAGNGASDRFRLSNVSPWITSARGASGS---DSPATVKLGSSSMIRAFKGVSI----YKNKTRRNVLSQRNKQFRPMAYLGRKDLSSPDP
Consensus      MEMRALVScsaaGNGaSDrfrLsNvSPWITsarGasgs...#sPaTVKLgSSsmiRaFKGVSi....yknktrrnVLSqrnkQFrP$aylgrkdlsspDP
```

```
             101      110       120       130       140       150       160       170       180       190       200
             |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
A.thalianaACC2 ISFLENDVSEAERTVVLPDGSVNGAGSVNGYHSDVVPGRN---VAEVNEFCKALGGKRPIHSILVATNGMAAVKFIRSVRTWAYETFGSEKAVKLVAMAT
A.lyrataACC2   TSFRONDISEAQRTVVLPGGSVNGY------HQSEVVPGRNDGTVAEVDEFCKALGGKRPIHSILVATNGMAAVKFIRSIRTWAYETFGTEKAVKLVAMAT
C.sativaACC2-1 TSFWENDISEVQRTV-LPAESING------DKSAVEPGRNDVTVSEVDEFCKALGGKRPIHSILVATNGHASVKFIRSIRTWAYQTFGSEKAISLVAMAT
C.sativaACC2-2 TSFWENDISETQRTV-LPAESING------DKSAVEPGRNDVTVSEVDEFCKALGGKRPIHRIHVATNGMAAVKFIRSIRTWAYQTFGSEKAISLVAMAT
C.rubellaACC2  TLFWENDISEAQSTV-LPSGS---------DESAVVPSGNDVKVSEVDEFCKALGGKRPIHSILVATNGMAAVKFIRSIRTWAYQTFGTEKAILLVAMAT
B.oleraceaACC2 TSFCONDISEPQGT-----GSINGN-----DHSAV-------RVSQVDEFCKAHGGKRPIHSILVATNGMAAVKLIRSYRAHSCQTFGSEKSISLVAMAT
B.napusACC2-2  TSFCONDISEPQGT-----GSINGN-----DHSAV-------RVSQVDEFCKAHGGKRPIHRILVATNGMAAVKFIRSYRAHSYQTFGSEKSISLVAMAT
B.rapaACC2     TSFCONDISEPQGT-----GSINGN-----DHSAV-------RVSQVDEFCKAHGGKRPIHSILVATNGMAAVKFIRSYRAHSYQTFGSEKSISLVAMAT
B.napusACC2-1  TSFCONDISEPQGT-----GSINGN-----DHSAV-------RVSQVDEFCKAHGGKRPIHSILVATNGMAAVKLIRSYRAHSYQTFGSEKSISLVAMAT
Consensus      TSFc#NDISEpQgT.....gS!NGn.....dhSaV.......rVs#VDEFCKAhGGKRPIHsILVATNGMAAVK1IRS!RaWsY#IFGsEKs!sLVAMAT
```

```
             201      210       220       230       240       250       260       270       280       290       300
             |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
A.thalianaACC2 PEDMRINAEHIRIADQFVEVPGGTNNNNYANVQLIVEMAEVTRVDAVWPGWGHASENPELPDALKEKGIIFLGPPADSMIALGDKIGSSLIAQAADVPTL
A.lyrataACC2   PEDMRINAEHIRIADQFVEVPGGTNNNNYANVQLIVEMAEVTRVDAVWPGWGHASENPELPDALKAKGIIFLGPPAASMIALGDKIGSSLIAQAADVPTL
C.sativaACC2-1 PEDMRINAEHIRIADQFVEVPGGTNNNNYANVQLIVEMAEATRVDAVWPGWGHASENPELPDALKAKGIIFLGPPATSMVALGDKIGSSLIAQAADVPTL
C.sativaACC2-2 PEDMRINAEHIRIADQFVEVPGGTNNNNYANVQLIVEMAEATRVDAVWPGWGHASENPELPDALKAKGIIFLGPPATSMVALGDKIGSSLIAQAADVPTL
C.rubellaACC2  PEDMRINAEHIRIADQFVEVPGGTNNHNYANVQLIVEMAEARSVDAVWPGWGHASENPELPDALKAKGIIFLGPSAASHVALGDKIGSSLIAQAADVPTL
B.oleraceaACC2 PEDMRINAEHIRIADQFMQVPGGTNNNNYANVHLIVEMAQATGVDAVWPGWGHASENPELPDALKAKGVIFLGPTAASHLALGDKIGSSLIAQAADVPTL
B.napusACC2-2  PEDMRINAEHIRIADQFMQVPGGTNNNNYANVHLIVEMAEATGVDAVWPGWGHASENPELPDALKAKGVIFLGPTAASHLALGDKIGSSLIAQAADVPTL
B.rapaACC2     PEDMRINAEHIRIADQFMQVPGGTNNNNYANVHLIVEMAEATGVDAVWPGWGHASENPELPDALKAKGVIFLGPTAASHLALGDKIGSSLIAQAADVPTL
B.napusACC2-1  PEDMRINAEHIRIADQFMQVPGGTNNNNYANVHLIVEMAQATGVDAVWPGWGHASENPELPDALKAKGVIFLGPTAASHLALGDKIGSSLIAQAADVPTL
Consensus      PEDMRINAEHIRIADQFm#VPGGTNNNNYANVhLIVEMA#aTgVDAVWPGWGHASENPELPDALKAKG!IFLGPtAaSH1ALGDKIGSSLIAQAADVPTL
```

Figure 1-5. Alignment of homomeric ACCases in the Brassicaceae family. A, Alignment of 200 the N-terminal amino acids of *Arabidopsis thaliana* ACC1 (At1g36160) and ACC2 (At1g36180) genes. B, Alignment of 200 the N-terminal amino acids of *Arabidopsis thaliana* ACC1: At1g36160; *Arabidopsis lyrata* ACC1: XM_002891166.1; *Camelina sativa* ACC1-1: LOC104777496; *Camelina sativa* ACC1-2: LOC104743830; *Capsella rubella* ACC1: CARUB_v10011872mg; *Brassica oleracea* ACC1: LOC106311006; *Brassica napus* ACC1-1: LOC106413885; *Brassica napus* ACC1-2: LOC106418889; *Brassica rapa* ACC1: LOC103833578. C, Alignment of 300 the N-terminal amino acids of *Arabidopsis thaliana* ACC2: At1g36180; *Arabidopsis lyrata* ACC2: XM_002891167.1; *Camelina sativa* ACC2-1: LOC104777495; *Camelina sativa* ACC2-2: LOC104742086; *Capsella rubella* ACC2: CARUB_v10008063mg; *Brassica oleracea* ACC2: LOC106301042; *Brassica napus* ACC2-1: Y10302; *Brassica napus* ACC2-2: X77576; *Brassica rapa* ACC2: LOC103871500.

Table 1-1. *Identification of transplastomic events in Arabidopsis*

Au, gold particles; Hepta, usin g the Biolistic gun Hepta adaptor instead of a single flying disk; Tu, tungsten particles.

Plasmid	Left/Right Arm	Marker Gene	Accession	Tissue	Gun	No. of Plates	No. of Transplastomic Events	Reference
	kb							
pGS31A	1.1/0.9	Prrn:LrbcL:aadA:TpsbA	RLD	Leaf	Single, Tu/1 μm	201	2	Sikdar et al. (1998)
pAAK176	1.7/0.8	Prrn:LrbcL:aadA:TpsbA	RLD	Leaf	Hepta, Au/0.6 μm	10	0	Reported here
			L*er*	Leaf	Hepta, Au/0.6 μm	4	0	Reported here
pTT626	1.7/0.8	Prrn:Lcry9:aadA-gfp:TpsbA	RLD	Leaf	Hepta, Au/0.6 μm	14	0	Reported here
pATV1	1.7/0.8	PrrnLatpB:aadA:Lcry9:gfp:TpsbA	RLD	Leaf	Hepta, Au/1 μm	2	0	Reported here
			L*er*	Leaf	Hepta, Au/1 μm	1	0	Reported here
			Col-0	Leaf	Hepta, Au/0.6 μm	4	0	Reported here
			Col-0 *acc2-1*	Leaf	Hepta, Au/0.6 μm	5	8	Reported here
			Sav-0	Leaf	Hepta, Au/0.6 μm	4	4	Reported here

References

1   Svab, Z., Hajdukiewicz, P. & Maliga, P. Stable transformation of plastids in higher plants. *Proc. Natl. Acad. Sci. USA* **87**, 8526-8530 (1990).

2   Svab, Z. & Maliga, P. High-frequency plastid transformation in tobacco by selection for a chimeric *aadA* gene. *Proc. Natl. Acad. Sci. USA* **90**, 913-917 (1993).

3   Ruf, S., Hermann, M., Berger, I. J., Carrer, H. & Bock, R. Stable genetic transformation of tomato plastids: foreign protein expression in fruit. *Nat. Biotechnol.* **19**, 870-875, doi:10.1038/nbt0901-870 (2001).

4   Valkov, V. T. *et al.* High efficiency plastid transformation in potato and regulation of transgene expression in leaves and tubers by alternative 5' and 3' regulatory sequences. *Transgenic Res.* **20**, 137-151, doi:10.1007/s11248-010-9402-9 (2011).

5   Kanamoto, H. *et al.* Efficient and stable transformation of *Lactuca sativa* L. cv. Cisco (lettuce) plastids. *Transgenic Res.* **15**, 205-217, doi:10.1007/s11248-005-3997-2 (2006).

6   Ruhlman, T., Verma, D., Samson, N. & Daniell, H. The role of heterologous chloroplast sequence elements in transgene integration and expression. *Plant Physiol.* **152**, 2088-2104, doi:10.1104/pp.109.152017 (2010).

7   Dufourmantel, N. *et al.* Generation of fertile transplastomic soybean. *Plant Mol. Biol.* **55**, 479-489, doi:10.1007/s11103-004-0192-4 (2004).

8   Sikdar, S. R., Serino, G., Chaudhuri, S. & Maliga, P. Plastid transformation in *Arabidopsis thaliana*. *Plant Cell Rep.* **18**, 20-24 (1998).

9   Wirmer, J. & Westhof, E. Molecular contacts between antibiotics and the 30S ribosomal particle. *Methods Enzymol.* **415**, 180-202 (2006).

10   Wilson, D. N. Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat. Rev. Microbiol.* **12**, 35-48, doi:10.1038/nrmicro3155 (2014).

11   Zubko, M. K. & Day, A. Stable albinism induced without mutagenesis: a model for ribosome-free plastid inheritance. *Plant J.* **15**, 265-271 (1998).

12   Parker, N., Wang, Y. & Meinke, D. Natural variation in sensitivity to a loss of chloroplast translation in Arabidopsis. *Plant Physiol.* **166**, 2013-2027, doi:10.1104/pp.114.249052 (2014).

13   Parker, N., Wang, Y. & Meinke, D. Analysis of Arabidopsis Accessions Hypersensitive to a Loss of Chloroplast Translation. *Plant Physiol.* **172**, 1862-1875 (2016).

14   Maliga, P. & Tungsuchat-Huang, T. in *Chloroplast Biotechnology: Methods and Protocols* Vol. 1132 *Methods in Molecular Biology* (ed P. Maliga) Ch. 8, 147-163 (Springer Science+Business Media, 2014).

15   Staub, J. M. & Maliga, P. Expression of a chimeric *uidA* gene indicates that polycistronic mRNAs are efficiently translated in tobacco plastids. *Plant J.* **7**, 845-848 (1995).

16   Schottkowski, M. *et al.* Biogenic membranes of the chloroplast in *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. USA* **109**, 19286-19291, doi:10.1073/pnas.1209860109 (2012).

17   Motte, H. *et al.* Phenyl-adenine, identified in a LIGHT-DEPENDENT SHORT HYPOCOTYLS4-assisted chemical screen, is a potent compound for shoot regeneration through the inhibition of CYTOKININ OXIDASE/DEHYDROGENASE activity. *Plant Physiol.* **161**, 1229-1241, doi:10.1104/pp.112.210716 (2013).

18   Bairu, M. W., Aremu, A. O. & Van Staden, J. Somaclonal variation in plants: causes and detection methods. *Plant Growth Regul.* **63**, 147-173, doi:10.1007/s10725-010-9554-x (2011).

19   Marton, L. & Browse, J. Facile transformation of *Arabidopsis*. *Plant Cell Rep.* **10**, 235-239 (1991).

20   Lutz, K. A., Martin, C., Khairzada, S. & Maliga, P. Steroid-inducible BABY BOOM system for development of fertile *Arabidopsis thaliana* plants after prolonged tissue culture. *Plant Cell Rep.* **34**, 1849-1856, doi:10.1007/s00299-015-1832-7 (2015).

21   Chupeau, M. C. *et al.* Characterization of the early events leading to totipotency in an Arabidopsis protoplast liquid culture by temporal transcript profiling. *Plant Cell* **25**, 2444-2463, doi:10.1105/tpc.113.109538 (2013).

22   Zhao, X., Liang, G., Li, X. & Zhang, X. Hormones regulate in vitro organ regeneration from leaf-derived explants in *Arabidopsis*. *Am. J. Plant Sci.* **5**, 3535-3550 (2014).

23   Zhao, X. Y. *et al.* Differences in capacities of in vitro organ regeneration between two *Arabidopsis* ecotypes Wassilewskija and Columbia. *Plant Cell Tissue Organ Cult.* **112**, 65-74 (2013).

24   Babiychuk, E. *et al.* Plastid gene expression and plant development require a plastidic protein of the mitochondrial transcription termination factor family. *Proc. Natl. Acad. Sci. USA* **108**, 6674-6679, doi:10.1073/pnas.1103442108 (2011).

25   Bryant, N., Lloyd, J., Sweeney, C., Myouga, F. & Meinke, D. Identification of nuclear genes encoding chloroplast-localized proteins required for embryo development in Arabidopsis. *Plant Physiol.* **155**, 1678-1689, doi:10.1104/pp.110.168120 (2011).

26   Liu, C. W., Lin, C. C., Chen, J. J. & Tseng, M. J. Stable chloroplast transformation in cabbage (*Brassica oleracea* L. var. *capitata* L.) by particle bombardment. *Plant Cell Rep.* **26**, 1733-1744, doi:10.1007/s00299-007-0374-z (2007).

27   Liu, C. W., Lin, C. C., Yiu, J. C., Chen, J. J. & Tseng, M. J. Expression of a *Bacillus thuringiensis* toxin (*cry1Ab*) gene in cabbage (*Brassica oleracea* L. var. *capitata* L.) chloroplasts confers high insecticidal efficacy against *Plutella xylostella*. *Theor. Appl. Genet.* **117**, 75-88, doi:10.1007/s00122-008-0754-y (2008).

28   Nugent, G. D., Coyne, S., Nguyen, T. T., Kavanagh, T. A. & Dix, P. J. Nuclear and plastid transformation of *Brassica oleracea* var. *botrytis* (cauliflower) using PEG-mediated uptake into protoplasts. *Plant Sci.* **170**, 135-142 (2006).

29   Hou, B. K. *et al.* Chloroplast transformation in oilseed rape. *Transgenic Res.* **12**, 111-114 (2003).

30   Cheng, L. *et al.* Chloroplast transformation of rapeseed (*Brassica napus*) by particle bombardment of cotyledons. *Plant Cell Rep.* **29**, 371-381, doi:10.1007/s00299-010-0828-6 (2010).

31   Skarjinskaia, M., Svab, Z. & Maliga, P. Plastid transformation in *Lesquerella fendleri*, an oilseed Brassicacea. *Transgenic Res.* **12**, 115-122 (2003).

32   Murashige, T. & Skoog, F. A revised medium for the growth and bioassay with tobacco tissue culture. *Physiol. Plant.* **15**, 473-497 (1962).

33   Kuroda, H. & Maliga, P. Sequences downstream of the translation initiation codon are important determinants of translation efficiency in chloroplasts. *Plant Physiol.* **125**, 430-436 (2001).

34   Lutz, K., Corneille, S., Azhagiri, A. K., Svab, Z. & Maliga, P. A novel approach to plastid transformation utilizes the phiC31 phage integrase. *Plant J.* **37**, 906-913 (2004).

35   Khan, M. S. & Maliga, P. Fluorescent antibiotic resistance marker to track plastid transformation in higher plants. *Nat. Biotechnol.* **17**, 910-915 (1999).

36   Chakrabarti, S. K., Lutz, K. A., Lertwirijawong, B., Svab, Z. & Maliga, P. Expression of the *cry9Aa2 B.t.* gene in the tobacco chlroplasts confers resistance to potato tuber moth. *Transgenic Res.* **15**, 481-488, doi:10.1007/s11248-006-0018-z (2006).

37   Lutz, K. A., Azhagiri, A. & Maliga, P. in *Chloroplast research in Arabidopsis* Vol. 774 (ed R.P. Jarvis) 133-147. (Springer Science+Business Media, LLC, 2011).

38   Tungsuchat-Huang, T. & Maliga, P. Visual marker and Agrobacterium-delivered recombinase enable the manipulation of the plastid genome in greenhouse-grown tobacco plants. *Plant J.* **70**, 717-725 (2012).

39   Carrer, H., Staub, J. M. & Maliga, P. Gentamycin resistance in *Nicotiana* conferred by AAC (3)-I, a narrow substrate specificity acetyl transferase. *Plant Mol. Biol.* **17**, 301-303 (1990).

40   Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16**, 10881-10890 (1988)

CHAPTER 2

New Tools for Engineering the Arabidopsis Plastid Genome

Abstract

Arabidopsis thaliana is the most characterized model plant and is used to study all aspects of basic science. The exception is that studies involving plastid genome engineering are carried out in *Nicotiana tabacum*, the only higher plant species where plastome engineering is routine [1]. Plastid transformation in Arabidopsis was reported in 1998 but only one transplastomic event was obtained per 100 bombarded samples; an efficiency 100-times lower than in tobacco [2]. A clue to why Arabidopsis plastid transformation was inefficient came years later from a study on nuclear genes essential for survival in the absence of chloroplast translation. Parker et al. [3] determined that most Arabidopsis ecotypes are tolerant to spectinomycin, the selective agent used in plastid transformation, due to a duplication of the ACCase enzyme. Spectinomycin is an inhibitor of plastid translation which curtails callus proliferation and greening in all species where it has been used to successfully recover transplastomic events [4]. In most Arabidopsis ecotypes, spectinomycin does not impair callus proliferation unless the ecotype carries a mutation in the nuclear *ACC2* gene. When the nuclear *ACC2* gene is not functional, the cells depend on the heteromeric ACCase enzyme for fatty acid biosynthesis, one subunit of which is encoded on the plastid genome [3]. We hypothesized that Arabidopsis plastid transformation inefficiency was due to tolerance of the RLD ecotype to

spectinomycin. The finding that plastid transformation in the spectinomycin hypersensitive Col-0 *acc2-1* and Sav-0 background proved 100-fold more efficient [5] corroborated this. However, the transplastomic plants regenerated from leaves failed to produce viable seed. Ruf et al. [6] confirmed that it is essential to carry out plastid transformation in an *ACC2* knockout background and showed that switching to root explants instead of leaves as the bombardment material in the C24 ecotype yields fertile plants. We report here the creation of spectinomycin hypersensitive lines in the Arabidopsis ecotypes RLD and Wassilewskija (Ws). We also extend the observation that it is essential to use the *ACC2* knockout background to transform the RLD and Ws ecotypes, indicating that the need to use a knockout recipient is probably true of all Arabidopsis ecotypes. We also report new Arabidopsis plastid transformation vectors that are suitable for the insertion of transgenes. In addition, we document significant differences in plastid transformation rates among different Arabidopsis ecotypes using a leaf system that enables rapid scoring of transplastomic callus events.

Results

To enable plastid transformation in additional Arabidopsis ecotypes, RLD and Wassilewskija (Ws) were chosen because efficient plant regeneration protocols are available for these ecotypes [7-10]. We designed a 20-bp sgRNA to target to the *ACC2* N-terminal chloroplast transit peptide, thereby avoiding mutagenesis of the cytoplasmic *ACC1* coding region (Fig. 2-1A, Fig. 2-3). The same sgRNA was used

to edit *ACC2* in the conserved region of RLD and Ws ecotypes (Fig. 2-1B). Putative mutants were identified by the T7 Endonuclease I assay that detects imperfectly matched DNA, and mutations were verified by direct sequencing of PCR products (Fig. 2-1C). The most readily identifiable mutations in *ACC2* are the 17-nt and 10-nt deletions in the RLD *acc2-3* and Ws *acc2-3* lines, respectively. The nucleotide insertions and deletions created a stop codon close to the target site which resulted in early translation termination in the *ACC2* reading frame (Fig. 2-3). Spectinomycin hypersensitivity of homozygous knockout lines was confirmed by germinating the seedlings on a selective spectinomycin medium, where wild type seedlings develop leaves and the knockout lines fail to develop any structure at the shoot apex (Fig. 2-1D) [3]. Under standard growth conditions, the knockout plants are fully fertile and produce seed.

Testing of plastid transformation efficiency was carried out with the pATV1 vector used in our earlier study [5] and the newly constructed pMEK14 vector (Fig. 2-2B). Vector pMEK14 is based on the pMEK18/pMEK19 backbone that carries a multiple cloning site to facilitate the insertion of passenger genes (Fig. 2-2A). The left targeting region of the pMEK vectors is truncated at the *Hin*dIII site in the plastid *rrn16* gene to enable utilization of the *Eco*RI-*Hin*dIII multiple cloning site. The multiple cloning site is adopted from the pUC18/pUC19 cloning vectors, which are routinely used for gene assembly in chloroplast engineering laboratories [11]. The pATV1 and pMEK14 vectors carry the same dicistronic *aadA-gfp* operon and target insertions between the plastid *trnV* gene and *3'-rps12/rps7* operon in the inverted

repeat region of the plastid genome (Fig. 2-2A). The first open reading frame (ORF) of the operon encodes aminoglycoside-3"-adenylyltransferase (AAD) that confers spectinomycin resistance to transformed cells. The second ORF encodes the green fluorescent protein (GFP) used for visual identification of transplastomic clones under UV light and confocal microscopy [5]. Truncation of the left targeting region in the pMEK vectors does not appear to have an impact on plastid transformation efficiency (Table 2-1).

Plastid transformation efficiency was evaluated by bombarding leaves, and culturing them on the one-step selective SPEED medium containing 100 mg/L spectinomycin for evaluation of transformation efficiency . Spectinomycin at this concentration severely inhibits callus proliferation in the *ACC2*-knockout leaves (Fig. 2-4). Spectinomycin resistant cells divide and form pale green and dark green clusters, but no shoots on the SPEED callus induction medium. Transplastomic cells are readily identified by GFP fluorescence five to six weeks after bombardment (Fig. 2-2C). Plastid transformation in these clones was confirmed by DNA gel blot analyses (Fig. 2-2D) and confocal microscopy (Fig. 2-2E). We note that calli no. 25, 27, 33 and 40 obtained with vector pATV1 and no. 86 obtained with vector pMEK14 were homoplastomic after one round of selection. This contrasts with tobacco, where typically multiple rounds of selection are required to obtain homoplasmy. Spontaneous mutations in the plastid 16S rRNA genes or nuclear insertions of *aadA* may also confer resistance to spectinomycin [12]. These cell clusters appear green due to chlorophyll accumulation but do not fluoresce under UV light (Fig. 2-2C).

A comparison of plastid transformation efficiency in the wild-type and knockout backgrounds for Col-0, RLD and Ws confirmed the absolute necessity for an *ACC2* defective background for the recovery of transplastomic events (Table 2-1). Only one transplastomic event was recovered in 23 samples with wild-type backgrounds, but many were obtained for all knockout backgrounds. Herein, transformation efficiencies varied significantly. The Col-0 *acc2* line yielded an average of 8 transplastomic events per bombarded sample. On the other extreme, Sav-0 yielded 1 event for every 2-4 bombarded samples. The poor transformation efficiency observed in Sav-0 may be due to the presence of a full length *ACC2* open reading frame with 15 missense mutations that may have maintained some residual enzymatic activity [13]. The reasons for higher plastid transformation efficiency in the Col-0 background are not understood and remain to be explored.

Discussion

Here we advance efforts to develop plastid transformation in Arabidopsis by developing a system for rapid evaluation of plastid transformation efficacy, complete with new transformation vectors, culture media and new RLD and Ws *ACC2*-knockout lines that expand the range of accessions available for experimentation. While the root system has proven its value by yielding fertile plants, leaf explants offer the advantage of recovering transplastomic events in 35 to 42 days (5 to 6 weeks), instead of the 120-240 days required in the root system[6]. Here we exploited the system to evaluate the impact of truncating the left

targeting region of the pMEK vectors. This rapid scoring system could also be used to evaluate new selective marker genes. However, endoreduplication in leaf tissues likely fostered the sterility that we encountered in plants regenerated from leaf tissue[5]. An alternative approach to overcome this issue may be to work with very young leaves or to explore the use of the BABY BOOM gene, which has the potential to preserve the diploid state in leaves[8]. Systematic research in the two systems will lead to reproducible protocols for plastid transformation in Arabidopsis to produce fertile transplastomic plants in an optimized timeframe. This will enable full exploitation of Arabidopsis as the model system.

Materials and Methods

Plant Materials and Growth Conditions

The *Arabidopsis thaliana* RLD seeds were purchased from Lehle Seeds, Round Rock, TX 78681, USA. The Col-0, Ws seeds were obtained from Prof. Juan Dong, Rutgers University. Sav-0 (CS28725) and Col-*acc2* (SALK_148966C) knockout seeds were ordered from the Arabidopsis Biological Resource Center. Seeds were surface sterilized with 5-fold diluted Clorox bleach (1.7% w/v sodium hypochlorite) in a 1.5 mL Eppendorf tube for 15 min. After removing the bleach, the seeds were washed three times with sterile distilled water. Seeds were germinated on ARM5 medium in deep petri dish (20 mm high and 10 cm in diameter) illuminated for 8 hours with cool-white fluorescent tubes (GE F34CW-RS-WM-Eco) at 21°C during the day and 18°C during the night. After 5 to 6 weeks, about 2-cm long green rosette

leaves were harvested for bombardment. Plants for nuclear transformation were cultivated in soil in a growth chamber illuminated Sylvania Octron 32W 4100K lights for 16 hours at 21ºC.

Deletion of *ACC2* by the CRISPR/Cas9 System

Genomic sequences of the *ACC2* gene (At1g36180) in the RLD and Ws ecotypes were gleaned from the Genome Browser-Arabidopsis 1001 (http://signal.salk.edu/atg1001/3.0/gebrowser.php) and aligned with the *ACC2* gene in Col-0 with MultAlin software [15]. The At1g36180 gene model (Fig. 2-1A) was annotated based on the Seqviewer website (https://seqviewer.arabidopsis.org).

Targeted mutations in *ACC2* were obtained using the CRISPR/Cas9 system described by Mao et al. [14]. Plasmid p998 contains a *cas9* gene driven by the Arabidopsis UBQ1 promoter. The expression of the sgRNA is under the control of the Arabidopsis RNA polymerase III U6 promoter. The sgRNA targeting *ACC2* was designed using website https://zlab.bio/guide-design-resources. After phosphorylation and annealing sgRNA-ACC2-F2 gattgCCTCACGAATATATCTCCA and sgRNA-ACC2-R2 aaacTGGAGATATATTCGTGAGGc, the forward and reverse oligonucleotides were ligated into the *Bbs*I-digested p998 vector. The CRISPR/Cas9 cassette was cloned into pCAMBIA-2300 vector, which was introduced into *Agrobacterium tumefaciens* GV3101 strain. The RLD and Ws plants were transformed with Agrobacterium using the standard floral dipping protocol [17]. Seedlings carrying the

CRISPR/Cas construct were identified by kanamycin resistance on a medium containing half-strength MS salts [18] and 50 mg/L kanamycin. Plants carrying mutations in the target gene were screened by the T7 endonuclease I assay that detects mismatched DNA and the mutations were identified by Sanger sequencing of PCR amplicons. Genomic DNA was prepared with a standard cetyltrimethylammonium (CTAB) method [19] and the amplicons for sequencing were obtained with forward primer 5'- TGGGTGAAGGAGTGTGTTGT-3' and reverse primer 5'-AGCCCATGTTCTGACACTACG-3'.

Plastid Transformation Vectors

Arabidopsis plastid transformation vectors pATV1, pMEK14, and pMEK18/pMEK19 are pBSKS+ plasmid derivatives. They are all derivatives of plasmid pGS6, a pBSKS+ plasmid carrying the *Kpn*I/*Sac*I plastid DNA fragment of RLD *Arabidopsis* ecotype (Lehle Farms, TX), encoding the *rrn16*, *trnV*, and *3'rps12/7* genes. Plasmid pBEN1 was obtained by ligation of the *Kpn*I site of pBSKS+ to the *Nru*I site in the plastid DNA, and the *Xba*I site in the plastid DNA to the *Sac*I site of the vector. pATV1 is a pBEN1 derivative, into which the dicistronic *aadA-gfp* gene (*Sac*I-*Hin*dIII fragment) was cloned at the *Hin*cII restriction site (GenBank Accession No. MF461355. The plastid DNA to the left of the multiple cloning site (left targeting region, Fig. 2-2A) in vectors pMEK18/pMEK19 has been truncated by ligating the *Bsp*QI site in the pBSKS+ backbone to the *Hin*dIII site in the ptDNA insert. The re-created *Hin*dIII site was

subsequently removed and a multiple cloning site inserted at the *Hin*cII site. Thus, the pMEK vectors have a shorter left flanking region relative to pATV1 (1055bp versus 1665bp). The *Eco*RI-*Hin*dIII multiple cloning site in the vectors is in the opposite relative orientation in pMEK18 (GenBank Accession No. MN326017) and pMEK19 (GenBank Accession No. MN326018) and is flanked by blunt sites for the insertion of marker genes. Vector pMEK14 is a construction intermediate, which carries the same *aadA-gfp* dicistronic marker as pATV1 and some additional restriction sites.

The pATV1 and pMEK14 vectors carry the same dicistronic *aadA-gfp* operon. The operon is expressed from the tobacco *rrn* operon promoter (P*rrn*, labeled as "P"); *aadA* has the leader sequence of the tobacco plastid *atpB* gene (L*atpB*); *gfp* has the *Bacillus thuringiensis Cry9Aa2* gene leader sequence (L*Cry9Aa2*); and the 3' UTR of plastid *psbA* gene (TpsbA, labeled as "T") [5].

Plastid Transformation

The one-step selective culture medium for the evaluation plastid transformation efficiency to obtain expedited detection (SPEED medium) is based on the ARM medium described for Arabidopsis [2,7]. ARM consists of MS salts, 3% (w/v) sucrose, 0.8% (w/v) agar (A7921; Sigma), 200 mg of myoinositol, 0.1 mg of biotin (1 mL of 0.1 mg/mL stock), and 1 mL of vitamin solution (10 mg of vitamin B1, 1 mg of vitamin B6, 1 mg of nicotinic acid, and 2 mg of glycine per mL) per liter, pH 5.8. The medium contains 0.5 mg/L a-naphthaleneacetic acid (NAA), 3.0 mg/L Dicamba

[20] and 1.0 mg/L 6-(3-hydroxybenzylamino) purine (*meta*-topolin) [9], and 100 mg/L spectinomycin.

Arabidopsis plastid transformation was carried out by the 1998 biolistic protocol [2], using the PDS-1000He biolistic gun equipped with a hepta adaptor [21]. Approximately 70 green leaves (10~20 mm long) were harvested from the aseptically grown plants in deep petri dishes on ARM5 medium and tightly arranged abaxial side up on SPEED medium in a 10 cm diameter petri dish. The leaves were bombarded with DNA-coated 0.6 μm gold particles and incubated for an additional two days in the absence of selection. Two days after bombardment the leaves were transferred, adaxial side up, to selective SPEED medium containing 100 mg/L spectinomycin. Every two weeks, the leaves were transferred to fresh selective SPEED medium. After five to six weeks the leaf cultures were screened for transplastomic events by illumination with a handheld UV light (Model UVL-56, manufactured by UVP, Upland, CA, 91786).

DNA Gel-Blot Analyses

DNA extraction and DNA gel blot hybridizations were carried out as previously described [23]. The only deviation from the published protocol was the use of a modified Church buffer (0.5M Na_2HPO_4, 7% SDS, 10mM EDTA, Ph 7.2) instead of the Rapid-hyb Buffer (GE Healthcare, Piscataway, NJ, USA). Briefly, total callus DNA was isolated by the CTAB protocol [21] and digested with the *Eco*RI restriction endonuclease. The digested fragments were separated in an agarose gel and

transferred to a nitrocellulose membrane. The DNA probe was the *Apa*I-*Sph*I ptDNA fragment of the 16S rRNA coding region (Fig. 2-2B) labelled with the Takara Bio. Inc. Random Primer DNA Labeling Kit.

Confocal Laser Scanning Microscopy

GFP fluorescence was detected by a Leica TCS SP5II confocal laser-scanning microscope. To detect GFP and chlorophyll fluorescence, the excitation wavelength were 488 nm and 568 nm, and the detection filters was set to 500-530 nm and 650-700 nm, respectively.
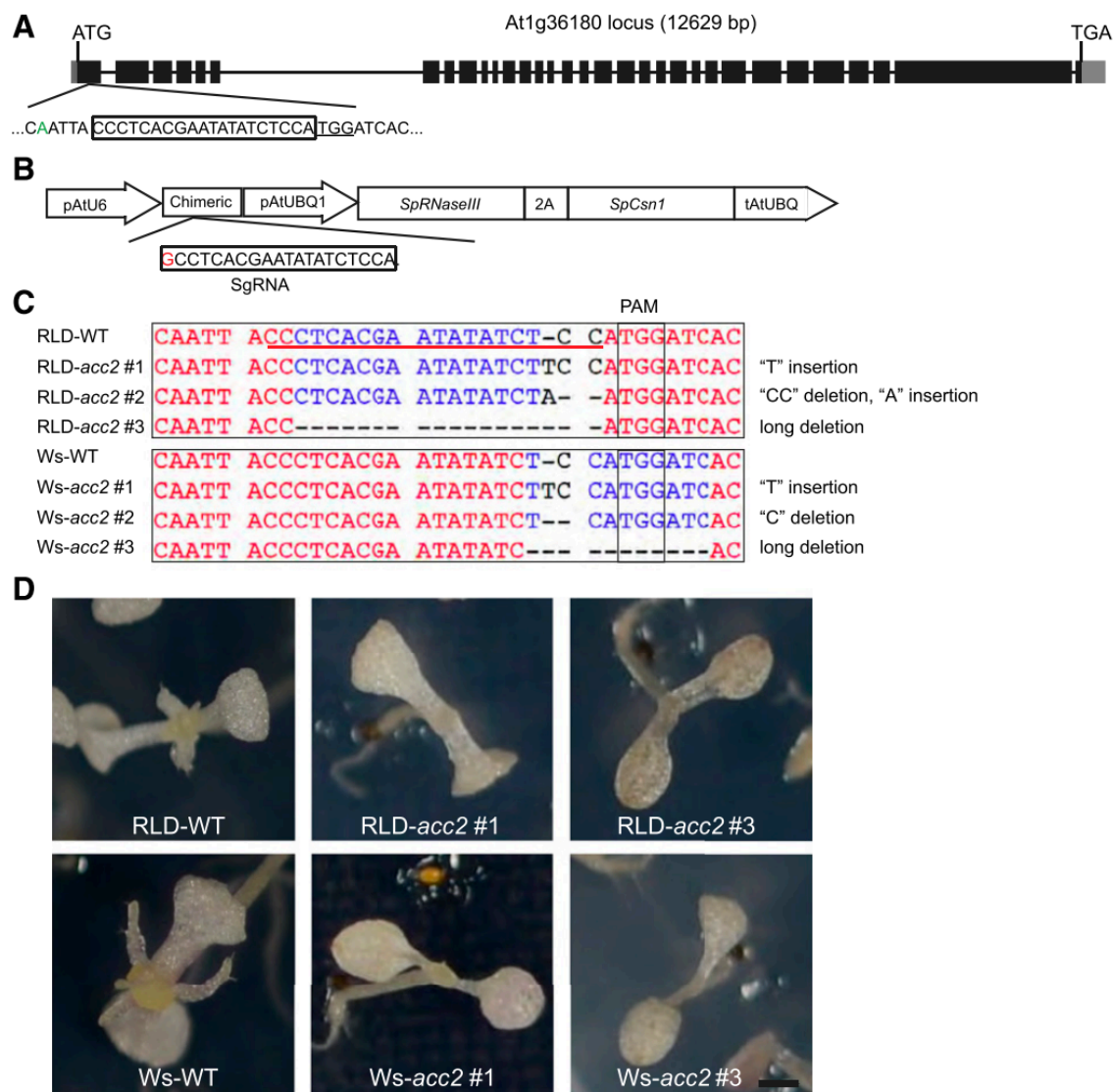
Figure 2-1. Generation of *ACC2* knockout lines using the CRISPR/Cas9 system. A,

Map of the *Arabidopsis thaliana ACC2* gene. The region targeted by sgRNA is

boxed. A single nucleotide polymorphism (SNP) in Col is highlighted in green. The

protospacer-adjacent motif (PAM) is underlined. B, Schematic diagram of the

sgRNA and Cas9 genes [14]. The 1st nucleotide of the sgRNA was changed from "C"

to "G" (in red) to facilitate transcription from the U6 promoter by RNA polymerase

III. C, Targeted mutations induced by CRISPR/Cas9 in the *ACC2* gene. We show

sequences in the targeted region aligned by the MultAlin software [15]. D, *ACC2* mutations make RLD and Ws hypersensitive to spectinomycin (100 mg/L). Note the arrested growth at the cotyledon stage and lack of shoot apex in the *acc2* lines, and outgrowth of the true leaves and development of shoot in wild type plants.
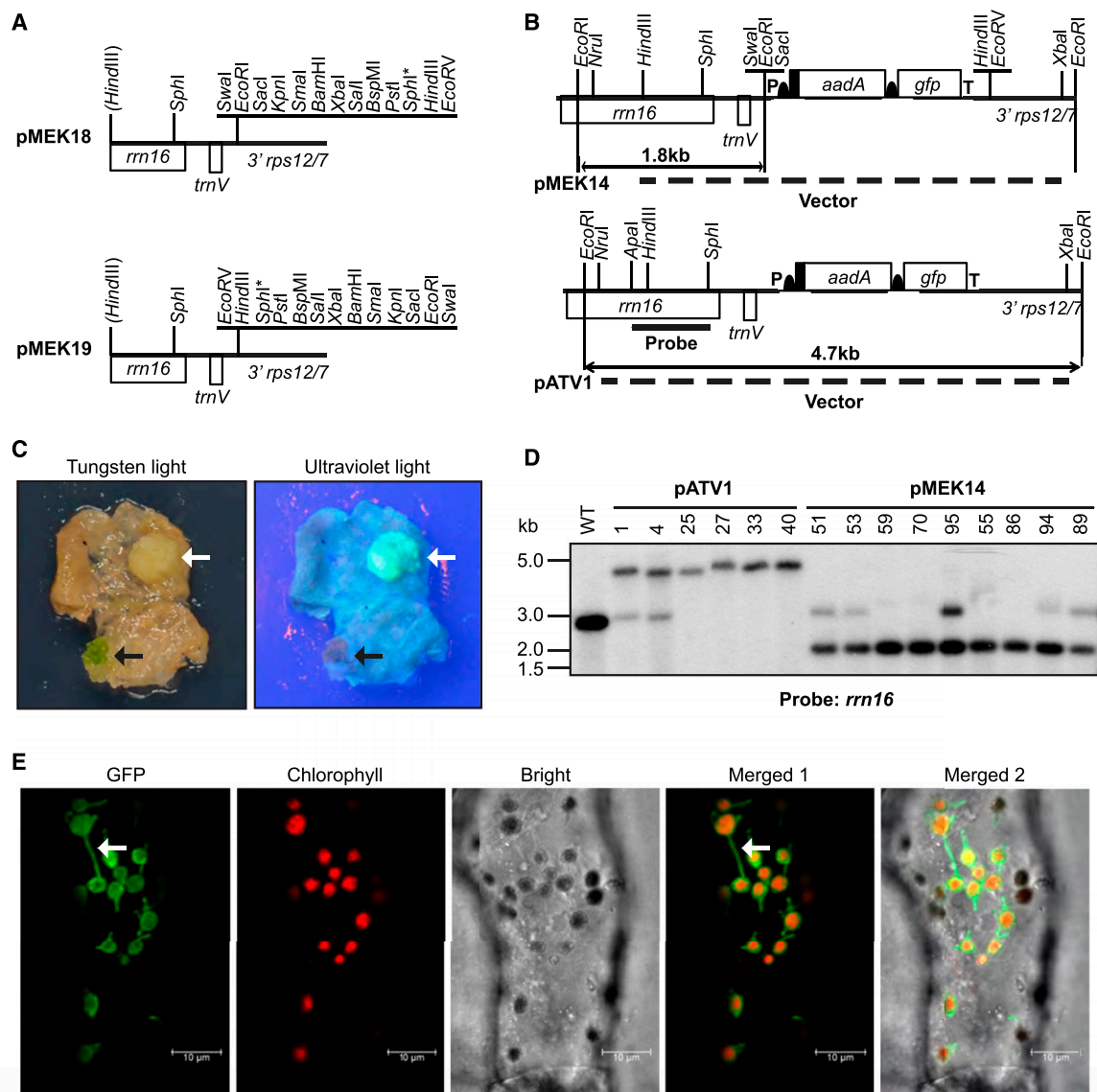
Figure 2-2. Plastid transformation in *Arabidopsis thaliana.* A, The plastid

targeting region of the pMEK18 (GenBank Accession No. MN326017) and

pMEK19 (MN326018) Arabidopsis plastid transformation vectors. The map

position of plastid *rrn16* and *trnV* genes and the promoter region of 3'-rps12/7

operon is shown as well as relevant restriction enzyme sites. B, Arabidopsis

plastid genome transformed with the pATV1 (MF461355) and pMEK14 vectors.

The vectors carry an identical dicistronic operon encoding an *aadA* and a *gfp* gene.

The position of the P*rrn*L*atpB* promoter (P), T*psbA* Terminator (T) and ribosome

binding site (semi-ovals) is indicated. The black box at the *aadA* N-terminus

marks the *atpB* Downstream Box sequence [16]. Vector pMEK14 has a shorter left

flanking region relative to pATV1 (1055bp versus 1665bp; targeting regions are

shown as dashed lines). The thick black line indicates the probe used for the DNA

gel blot analyses. C, Col-0 *acc2* transplastomic event (white arrow; fluoresces

under UV light) and spontaneous spectinomycin resistant mutant or *aadA* nuclear

insertion line (black arrow; green in normal light) one month after leaf

bombardment. D, DNA gel blot using the *rrn16* probe (Fig. 2-2B) confirms plastid

transformation in GFP-expressing Col-0 *acc2* clones. Plastid DNA transformed

with vector pATV1 and pMEK14 yields a 4.7 kb and 1.8 kb *Eco*RI fragment,

respectively. The 2.7 kb fragment is present in the wild-type and heteroplastomic

samples. E, Confocal microscopy to detect GFP accumulation in Col-0 *acc2*

chloroplasts. Arrow points to a stromule connecting two chloroplasts. Bar=10μm.

Merge 1 refers to the overlap of GFP and chlorophyll channels. Merge 2 refers to

the overlap of GFP, chlorophyll and bright fields.

>Col-0-WT
ATGGAGATGAGAGCTTTGGGTTCTTCGTGTTCTACTGGTAATGGAGGTTCTGCTCC
GATTACCCTCACGAATATATCTCCATGGATCACAACAGTTTTTCCGTCGACAGTGAA
GCTGAGAAGTAGTTTGAGAACCTTCAAAGGAGTTTCGTCAAGAGTGAGAACCTTTA
AAGGAGTTTCTTCGACAAGAGTTTTGTCTCGGACCAAACAACAGTTTCCTCTGTTTT
GTTTCCTAAACCCTGATCCGATCTCCTTCTTGGAAAATG
>Col-0-WT 1st exon amino acids
MEMRALGSSCSTGNGGSAPITLTNISPWITTVFPSTVKLRSSLRTFKGVSSRVRTFKGV
SSTRVLSRTKQQFPLFCFLNPDPISFLEN

>RLD-WT
ATGGAGATGAAAGCTTTGGGTTCTTCGTGTTCTACTGGTAATGGAGGTTCTGCTCC
AATTACCCTCACGAATATATCTCCATGGATCACAACAGTTTTTCCGTCGACAGTGAA
GCTGAAAAGTAGTTTGAGAACCTTCAAAGGAGTTTCGTCAAGAGTGAGAACCTTTAA
AGGAGTTTCTTCGACAAGAGTTTTGTCTCGGACCAAACAACAGTTTCCTCTGTTTTG
TTTCCTAAACCCTGATCCGATCTCCTTCTTGGATAATG
>RLD-WT. 1st exon amino acids
MEMKALGSSCSTGNGGSAPITLTNISPWITTVFPSTVKLKSSLRTFKGVSSRVRTFKGV
SSTRVLSRTKQQFPLFCFLNPDPISFLDN

>RLD-*acc2* #1
ATGGAGATGAAAGCTTTGGGTTCTTCGTGTTCTACTGGTAATGGAGGTTCTGCTCC
AATTACCCTCACGAATATATCTTCCATGGATCACAACAGTTTTTCCGTCGACAGTGA
AGCTGAAAAGTAGTTTGAGAACCTTCAAAGGAGTTTCGTCAAGAGTGAGAACCTTTA
AAGGAGTTTCTTCGACAAGAGTTTTGTCTCGGACCAAACAACAGTTTCCTCTGTTTT
GTTTCCTAAACCCTGATCCGATCTCCTTCTTGGATAATG
>RLD-*acc2* #1. 1st exon amino acids
MEMKALGSSCSTGNGGSAPITLTNISSMDHNSFSVDSEAEK*

>RLD-*acc2* #2
ATGGAGATGAAAGCTTTGGGTTCTTCGTGTTCTACTGGTAATGGAGGTTCTGCTCC
AATTACCCTCACGAATATATCTAATGGATCACAACAGTTTTTCCGTCGACAGTGAAG
CTGAAAAGTAGTTTGAGAACCTTCAAAGGAGTTTCGTCAAGAGTGAGAACCTTTAAA
GGAGTTTCTTCGACAAGAGTTTTGTCTCGGACCAAACAACAGTTTCCTCTGTTTTGT
TTCCTAAACCCTGATCCGATCTCCTTCTTGGATAATG
>RLD-*acc2* #2. 1st exon amino acids
MEMKALGSSCSTGNGGSAPITLTNISNGSQQFFRRQ*

>RLD-*acc2* #3
ATGGAGATGAAAGCTTTGGGTTCTTCGTGTTCTACTGGTAATGGAGGTTCTGCTCC
AATTACCATGGATCACAACAGTTTTTCCGTCGACAGTGAAGCTGAAAAGTAGTTTGA
GAACCTTCAAAGGAGTTTCGTCAAGAGTGAGAACCTTTAAAGGAGTTTCTTCGACAA
GAGTTTTGTCTCGGACCAAACAACAGTTTCCTCTGTTTTGTTTCCTAAACCCTGATC
CGATCTCCTTCTTGGATAATG
>RLD-*acc2* #3. 1st exon amino acids
MEMKALGSSCSTGNGGSAPITMDHNSFSVDSEAEK*

>Ws-WT
ATGGAGATGAAAGCTTTGGGTTCTTCGTGTTCTACTGGTAATGGAGGTTCTGCTCC

AATTACCCTCACGAATATATCTCCATGGATCACAACAGTTTTTCCGTCGACAGTGAA
GCTGAAAAGTAGTTTGAGAACCTTCAAAGGAGTTTCGTCAAGAGTGAGAACCTTTAA
AGGAGTTTCTTCGACAAGAGTTTTGTCTCGGACCAAACAACAGTTTCCTCTGTTTTG
TTTCCTAAACCCTGATCCGATCTCCTTCTTGGATAATG
>Ws-WT. 1st exon amino acids
MEMKALGSSCSTGNGGSAPITLTNISPWITTVFPSTVKLKSSLRTFKGVSSRVRTFKGV
SSTRVLSRTKQQFPLFCFLNPDPISFLDN

>Ws-*acc2* #1
ATGGAGATGAAAGCTTTGGGTTCTTCGTGTTCTACTGGTAATGGAGGTTCTGCTCC
AATTACCCTCACGAATATATCTTCCATGGATCACAACAGTTTTTCCGTCGACAGTGA
AGCTGAAAAGTAGTTTGAGAACCTTCAAAGGAGTTTCGTCAAGAGTGAGAACCTTTA
AAGGAGTTTCTTCGACAAGAGTTTTGTCTCGGACCAAACAACAGTTTCCTCTGTTTT
GTTTCCTAAACCCTGATCCGATCTCCTTCTTGGATAATG
>Ws-*acc2* #1. 1st exon amino acids
MEMKALGSSCSTGNGGSAPITLTNISSMDHNSFSVDSEAEK*

>Ws-*acc2* #2
ATGGAGATGAAAGCTTTGGGTTCTTCGTGTTCTACTGGTAATGGAGGTTCTGCTCC
AATTACCCTCACGAATATATCTCATGGATCACAACAGTTTTTCCGTCGACAGTGAAG
CTGAAAAGTAGTTTGAGAACCTTCAAAGGAGTTTCGTCAAGAGTGAGAACCTTTAAA
GGAGTTTCTTCGACAAGAGTTTTGTCTCGGACCAAACAACAGTTTCCTCTGTTTTGT
TTCCTAAACCCTGATCCGATCTCCTTCTTGGATAATG
>Ws-*acc2* #2. 1st exon amino acids
MEMKALGSSCSTGNGGSAPITLTNISHGSQQFFRRQ*

>Ws-*acc2* #3
ATGGAGATGAAAGCTTTGGGTTCTTCGTGTTCTACTGGTAATGGAGGTTCTGCTCC
AATTACCCTCACGAATATATCACAACAGTTTTTCCGTCGACAGTGAAGCTGAAAAGT
AGTTTGAGAACCTTCAAAGGAGTTTCGTCAAGAGTGAGAACCTTTAAAGGAGTTTCT
TCGACAAGAGTTTTGTCTCGGACCAAACAACAGTTTCCTCTGTTTTGTTTCCTAAAC
CCTGATCCGATCTCCTTCTTGGATAATG
>Ws-*acc2* #3. 1st exon amino acids
MEMKALGSSCSTGNGGSAPITLTNISQQFFRRQ*

Figure 2-3. DNA sequence and translation of the region targeted by the sgRNA in

exon 1 of the *ACC2 gene*. Amino acids that deviate from the wild type are in red.

Genomic sequences of *ACC2* (At1g36180) in RLD and Ws were gleaned from the

Genome Browser-Arabidopsis 1001

(http://signal.salk.edu/atg1001/3.0/gebrowser.php).

RLD-WT

RLD-*acc2* #1

Ws-WT

Ws-*acc2* #1

Col-0-WT

Col-0-*acc2*

Sav-0-WT
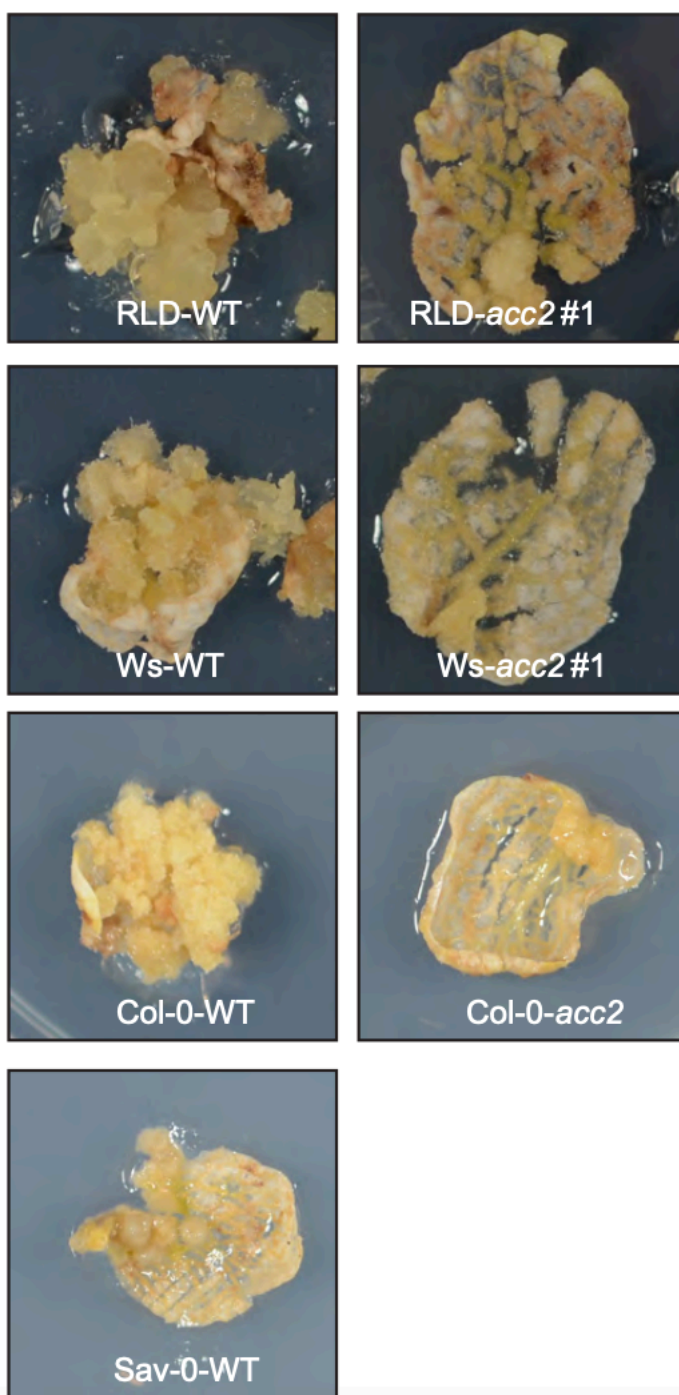
Figure 2-4. Callus phenotype of the wild-type and *acc2* knockout leaf explants.

Abundant callus forms on the wild-type leaf explants while spectinomycin (100

mg/L) effectively suppresses callus formation of *acc2* knockout explants on the

selective SPEED medium.

Table 2-1. The frequency of transplastomic events (TP) in Arabidopsis leaf

culture.

Accession	Vector	No. of Plates	No. of TP	No. of TP per Plate
Col-0	pATV1	6	0	0.0
	pMEK14	2	0	0.0
RLD	pATV1	6	0	0.0
Ws	pATV1	7	0	0.0
	pMEK14	2	1	0.5
Col-0 *acc2*	pATV1	9	76	8.4
	pMEK14	9	69	7.7
RLD acc2	pATV1	4	9	2.2
	pMEK14	2	6	3.0
Ws acc2	pATV1	4	11	2.8
	pMEK14	2	1	0.5
Sav-0	pATV1	10	4	0.4
	pMEK14	12	2	0.2

References

1   Bock, R. Engineering Plastid Genomes: Methods, Tools, and Applications in Basic Research and Biotechnology. *Annu. Rev. Plant. Biol.* **66**, 211-241, doi:10.1146/annurev-arplant-050213-040212 (2015).

2   Sikdar, S. R., Serino, G., Chaudhuri, S. & Maliga, P. Plastid transformation in *Arabidopsis thaliana*. *Plant Cell Rep.* **18**, 20-24 (1998).

3   Parker, N., Wang, Y. & Meinke, D. Natural variation in sensitivity to a loss of chloroplast translation in Arabidopsis. *Plant Physiol.* **166**, 2013-2027, doi:10.1104/pp.114.249052 (2014).

4   Maliga, P. in *Genomics of Chloroplasts and Mitochondria* Vol. 35 (eds R. Bock & V. Knoop) 393-414 (Springer, 2012).

5   Yu, Q., Lutz, K. A. & Maliga, P. Efficient plastid transformation in Arabidopsis. *Plant Physiol.* **175**, 186-193 (2017).

6   Ruf, S. *et al.* High-efficiency generation of fertile transplastomic Arabidopsis plants. *Nat Plants* **5**, 282-289, doi:10.1038/s41477-019-0359-2 (2019).

7   Marton, L. & Browse, J. Facile transformation of *Arabidopsis*. *Plant Cell Rep.* **10**, 235-239 (1991).

8   Lutz, K. A., Martin, C., Khairzada, S. & Maliga, P. Steroid-inducible BABY BOOM system for development of fertile *Arabidopsis thaliana* plants after prolonged tissue culture. *Plant Cell Rep.* **34**, 1849-1856, doi:10.1007/s00299-015-1832-7 (2015).

9   Chupeau, M. C. *et al.* Characterization of the early events leading to totipotency in an Arabidopsis protoplast liquid culture by temporal transcript profiling. *Plant Cell* **25**, 2444-2463, doi:10.1105/tpc.113.109538 (2013).

10   Zhao, X., Liang, G., Li, X. & Zhang, X. Hormones regulate in vitro organ regeneration from leaf-derived explants in *Arabidopsis*. *Am. J. Plant Sci.* **5**, 3535-3550 (2014).

11   Maliga, P. Engineering the plastid genome of higher plants. *Curr. Opin. Plant Biol.* **5**, 164-172 (2002).

12   Svab, Z. & Maliga, P. High-frequency plastid transformation in tobacco by selection for a chimeric *aadA* gene. *Proc. Natl. Acad. Sci. USA* **90**, 913-917 (1993).

13   Parker, N., Wang, Y. & Meinke, D. Analysis of Arabidopsis Accessions Hypersensitive to a Loss of Chloroplast Translation. *Plant Physiol.* **172**, 1862-1875 (2016).

14   Mao, Y. *et al.* Application of the CRISPR-Cas system for efficient genome engineering in plants. *Mol Plant* **6**, 2008-2011, doi:10.1093/mp/sst121 (2013).

15   Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16**, 10881-10890 (1988).

16   Kuroda, H. & Maliga, P. Sequences downstream of the translation initiation codon are important determinants of translation efficiency in chloroplasts. *Plant Physiol.* **125**, 430-436 (2001).

17   Clough, S. J. & Bent, A. F. Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735-743 (1998).

18   Murashige, T. & Skoog, F. A revised medium for the growth and bioassay with tobacco tissue culture. *Physiol. Plant.* **15**, 473-497 (1962).

19   Murray, M. G. & Thompson, W. F. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* **8**, 4321-4325 (1980).

20   Luo, Y. & Koop, H. U. Somatic embryogenesis in cultured immature zygotic embryos and leaf protoplasts of *Arabidopsis thaliana* ecotypes. *Planta* **202**, 387-396 (1997).

21   Lutz, K. A., Svab, Z. & Maliga, P. Construction of marker-free transplastomic tobacco using the Cre-*loxP* site-specific recombination system. *Nat. Protocols* **1**, 900-910, doi:10.1038/nprot.2006.118 (2006).

CHAPTER 3

Engineered PPR proteins as inducible switches to activate the expression of chloroplast transgenes

Abstract

The engineering of plant genomes presents exciting opportunities to modify agronomic traits and to produce high-value products in plants. Expression of foreign proteins from transgenes in the chloroplast genome offers advantages that include the capacity for prodigious protein output, the lack of transgene silencing, and the ability to express multicomponent pathways from polycistronic mRNA. However, there remains a need for robust methods to regulate plastid transgene expression. We designed orthogonal activators that boost expression of chloroplast transgenes harboring cognate *cis*-elements. Our system exploits the programmable RNA sequence-specificity of pentatricopeptide repeat proteins, and their native functions as activators of chloroplast gene expression. When expressed from nuclear transgenes, the engineered proteins stimulate expression of plastid transgenes up to ~40-fold, with maximal protein abundance approaching that of Rubisco. This strategy provides a means to regulate and optimize the expression of foreign genes in chloroplasts, and to avoid deleterious effects of foreign products on plant growth.

Introduction

The development of methodologies for manipulating chloroplast genomes [1, 2] presents exciting opportunities for use of the chloroplast as a biofactory to produce foreign products [3-5]. Chloroplast genes encode some of the most abundant proteins in nature, implying that the chloroplast gene expression system has the capacity for very high protein output. Indeed, chloroplast transgenes have been reported to yield as much as 70% of the soluble protein in tobacco seedlings [6], vastly exceeding outputs reported for nuclear transgenes. Chloroplast transgenes integrate at specified sites by homologous recombination and they are not subject to epigenetic silencing, ensuring stable expression from one generation to the next [3]. In addition, the fact that native chloroplast genes are expressed from polycistronic transcription units makes chloroplasts especially attractive for the incorporation of multigene pathways [7-15].

A challenge that stands in the way of fulfilling this promise is that expression of chloroplast transgenes often compromises growth of the host plant [16]. This may be due to toxicity of the foreign products or to diversion of resources from host functions. In either case, this limitation could be circumvented by maintaining low chloroplast transgene expression until shortly before harvest. Toward this end, several systems have been developed to induce chloroplast transgenes in tobacco, including methods based on the *lac* regulatory system [17], an inducible T7 RNA polymerase [14], and a modified sigma factor [18]. The most promising such system to date uses a theophylline-inducible riboswitch to activate translation of a plastid encoded T7 RNA polymerase, which in turn drives transgene transcription [19]. Still,

the dynamic range of this system (roughly 7-fold) and maximal protein yield (~1.75% of leaf protein) remain modest.

Recent advances in understanding mechanisms by which native chloroplast genes are activated by nucleus-encoded pentatricopeptide repeat (PPR) proteins offer new opportunities for the engineering of regulatory systems to control the expression of chloroplast transgenes. PPR proteins are helical repeat proteins that influence various RNA-mediated steps in gene expression in mitochondria and chloroplasts [20]. PPR proteins have tandem degenerate repeating units [21] that stack to form an elongated surface that binds single-stranded RNA via a modular 1 repeat-1 nucleotide recognition mode [20]. Nucleotide specificity is strongly influenced by the identity of amino acids at two positions in each repeat [22], which form a network of hydrogen bonds with the aligned nucleobase [23]. This binding mode is analogous to that by which TALE and PUF proteins bind specific DNA and RNA sequences, respectively [24]. The "PPR code" can be used to change the specificities of native PPR proteins in a predictable fashion [22, 25, 26], or to design synthetic proteins with desired sequence specificities [23, 27].

We demonstrate here that engineered PPR-encoding genes in conjunction with cognate binding sites upstream of plastid transgenes can be used to build a robust inducible switch for plastid transgenes. Our approach exploits the maize protein PPR10, which activates expression of the chloroplast *atpH* gene via well-defined mechanisms. PPR10 binds an RNA segment of ~20 nucleotides mapping a short distance upstream of the *atpH* ribosome binding site. PPR10 stimulates *atpH*

expression by blocking 5'→3' exoribonucleolytic decay and by increasing translational efficiency [28, 29]. Ribosome profiling analysis of *ppr10* mutants showed that the expression of *atpH* is reduced ~70-fold in the absence of PPR10 [30]. Furthermore, *atpH* is the most efficiently translated of any chloroplast mRNA in wild-type maize seedling leaves [31]. Site-directed mutations in PPR10 produced variants with altered sequence-specificity, as predicted by the PPR code [22, 32]. We reasoned, therefore, that these PPR10 variants expressed from the nuclear genome, in conjunction with a cognate binding site upstream of chloroplast transgenes, could be used to build orthogonal switches for transgene activation, with low basal expression and high activated protein output. These predictions were borne out by the results described below.

Results

Our general strategy is diagrammed in Fig. 3-1a. A PPR10 variant with amino acid changes that alter its sequence specificity (PPR10*) is expressed from the nuclear genome. This protein is targeted to the chloroplast, where it binds a cognate RNA sequence in the 5'-UTR of the transgene. The PPR10* binding sites have sequences that we predicted would interact poorly, if at all, with endogenous tobacco PPR10. Consequently, we anticipated that the plastid transgene would be expressed at low levels in the absence of PPR10*, and that PPR10* expression would activate the plastid transgene by stabilizing its RNA transcript and increasing translational efficiency (Fig. 3-1a).

The plastid reporter genes were provided with 5'-UTRs matching the 97-nucleotides upstream of the maize *atpH* start codon, excepting two nucleotides in the PPR10 binding site that were changed to correspond with the specificity of the PPR10 variants (Fig. 3-6a). We tested two PPR10 variants, which have amino acid substitutions in two adjacent PPR motifs that change their specificity from pyrimidines to either GG or AA (Fig. 3-1b). These variants are denoted PPR10^{GG} and PPR10^{AA}, respectively, and their binding sites are referred to below as the "GG site" and "AA site". Comprehensive analyses of the sequence specificities of wild-type PPR10 and these variants *in vitro* [22, 32] predict that the variants would bind negligibly to the endogenous *atpH* binding site in tobacco or to any other sequences in the tobacco plastid transcriptome, and that tobacco PPR10 would bind negligibly to these introduced sites (see binding site sequences in Fig. 3-1b). We tested the effects of the GG site in both a monocistronic and dicistronic reporter context (Fig. 3-1c), and we compared the efficacy of the CaMV 35S promoter, a soybean heat shock promoter [33] and an ethanol inducible promoter [34] to drive expression of the nucleus-encoded PPR10 variants (Fig. 3-6b). The chloroplast and nuclear transgenes were introduced into separate plants, whose progeny were then crossed together. Evidence for homoplasmy of the modified chloroplast genomes is shown in Fig. 3-6c.

Expression of the GFP Reporter Scales with the Abundance of the PPR10 Variant

An initial set of experiments used the soybean heat shock promoter to drive expression of PPR10^{GG} in the nucleus, with the intent of using heat shock to induce expression. However, we found that many transformed lines had high uninduced levels of PPR10^{GG} and that none of the lines exhibited robust heat shock induction. Nonetheless, we took advantage of the varying expression of PPR10^{GG} in different lines to address the quantitative relationship between PPR10^{GG} and expression of the chloroplast GFP reporters.

Immunoblot analyses of three dicistronic "GG" reporter lines expressing different levels of PPR10^{GG} are shown in Fig. 3-2a. The stained immunoblot shows a robust band corresponding to GFP in the high-expressing PPR10^{GG} line, where GFP abundance is similar to that of the large subunit of Rubisco (RbcL). The intensity of the stained GFP band correlates in a general sense with the abundance of PPR10^{GG}. Immunoblot analysis detected a low level of GFP in the absence of the PPR10^{GG} transgene (Fig. 3-2a); comparison of signal intensities among sample dilutions shows that this basal expression level is roughly 40-fold less than that in the high-expressing PPR10^{GG} line. PPR10^{GG} stimulated expression of the reporter genes in the monocistronic and dicistronic contexts to a similar extent (Fig. 3-2b). The highest expressing PPR10^{GG} line boosted GFP accumulation from the monocistronic reporter more than 40-fold, with respect to basal levels in the absence of the PPR10^{GG} transgene (Fig. 3-2c). GFP comprises approximately 25% of the soluble protein in leaf tissue from lines expressing the highest levels of PPR10^{GG} (Fig. 3-7a). This value is consistent with the relative abundance of Rubisco and GFP

in whole leaf lysates (see Ponceau stained filters in Fig. 3-2): the GFP and RbcL bands are of similar intensity, and Rubisco has been estimated to make up 30-50% of the soluble protein in C3 plants [35]. By extrapolation from the dilution series shown in Fig. 3-2c, we estimate basal GFP abundance (in the absence of PPR10^{GG}) to be roughly 0.6% of total soluble protein (i.e. 25%/40).

These results show that our system programs low basal expression of a plastid transgene regardless of whether it is found in a monocistronic context or as an internal ORF on a polycistronic RNA. Therefore, endogenous tobacco PPR10 has little, if any, cross-talk to the GG *cis*-element. Furthermore, the system has the capacity to achieve at least 40-fold induction if the expression of the nucleus-encoded activating factor can be suitably modulated.

Specificity of PPR10^{GG} and PPR10^{AA} for Their Cognate cis-elements

We next addressed the degree to which the PPR10 variants designed to bind either the "GG" or "AA" *cis*-elements exhibited crosstalk to the non-cognate binding site in the context of the dicistronic reporter (Fig. 3-3). The nuclear transgene encoding PPR10^{AA} was constitutively expressed from the CaMV 35S promoter. As observed for the PPR10^{GG} system, the AA *cis*-element programmed very low GFP expression in the absence of the PPR10^{AA} transgene, whereas GFP was more abundant than RbcL in its presence (Fig. 3-3a). PPR10^{AA} also increased expression from the GG site, albeit much less efficiently (Fig. 3-3a, last two lanes). By contrast, PPR10^{GG} had little if any impact on expression from the AA cis-element (Fig. 3-3b). GFP

fluorescence imaging of plants harboring these constructs supports these conclusions (Fig. 3-3c). These results show that PPR10^{GG} strongly discriminates between the "GG" and "AA" binding sites *in vivo*, as shown previously *in vitro* [22, 32]. PPR10^{AA} also acts with considerable specificity on its cognate site, but it exhibits some crosstalk to the GG site.

Ethanol-inducible PPR10^{GG} Drives Robust Inducible Expression of the GFP Reporter

We next tested the efficacy of an ethanol-inducible promoter as a means to induce expression of PPR10^{GG} and its cognate chloroplast GFP reporter. The same nuclear transformant was crossed with transplastomic plants harboring the dicistronic reporter construct with either the PPR10^{AA} or PPR10^{GG} binding site. Both lines showed negligible PPR10^{GG} expression in the absence of ethanol, and robust induction 4 days after ethanol application (Fig. 3-4a). PPR10^{GG} induction resulted in ~20-fold induction of the GFP reporter harboring the cognate PPR10^{GG} site (Fig. 3-4a, right panel). Ethanol was similarly effective at inducing GFP expression in 8- and 10-week old plants (Fig. 3-4, Fig. 3-7); GFP accumulated to roughly 15% of total soluble protein in both instances (Fig. 3-7a). By contrast, induction of PPR10^{GG} had no apparent effect on expression of the reporter with the AA *cis*-element (Fig. 3-4), consistent with the high specificity observed when PPR10^{GG} is expressed constitutively (Fig. 3-3). Therefore, induced expression of a PPR

activating protein in expanded leaves can induce a chloroplast transgene over a large

dynamic range.

Effects of PPR10^{GG} on RNA Stability and Processing

PPR10 activates expression of its native *atpH* target by stabilizing *atpH* mRNA and

increasing translational efficiency [28-30]. RNA stabilization results from the fact that

PPR10 blocks the progress of a 5'→3' exoribonuclease (see Fig. 3-1a). RNA gel

blot hybridizations showed that PPR10^{GG} promotes the accumulation of

monocistronic *gfp* mRNA from the dicistronic construct (Fig. 3-5a), analogous to

the stabilization of monocistronic *atpH* RNA by native PPR10; this is an expected

outcome of PPR10's barrier activity. PPR10^{GG} also caused a dramatic increase in

the abundance of RNA from the monocistronic reporter (Fig. 3-5a). We had not

anticipated such a strong effect on RNA accumulation from the monocistronic

reporter because the 5' triphosphate of primary transcripts prevents access by

bacterial Ribonuclease J [36], whose ortholog is proposed to mediate 5'→3' RNA

degradation in chloroplasts [28, 37, 38].

   To examine this phenomenon further, we mapped the RNA termini upstream

of each GFP reporter by primer extension (Fig. 3-5b). The 5'-end of the PPR10^{GG}-

dependent RNA isoform from the dicistronic construct mapped, as expected, to the

same position as the native processed PPR10-dependent *atpH* RNA. Unexpectedly,

however, PPR10^{GG} stabilized two transcript isoforms from the monocistronic

construct: one whose 5'-end mapped at the expected 5'-boundary of PPR10's

footprint (asterisk), and another whose 5'-end mapped to the transcription start site (diamond). The ability of a PPR protein to stabilize the 5'-end of a primary transcript mapping a considerable distance upstream from its binding site (~50 nucleotides in this instance) has not been described previously. This cannot result from the canonical exonuclease blockade mechanism. A hypothetical explanation is that the PPR10[GG] binding site sequence is sensitive to endonucleolytic cleavage; the binding of PPR10[GG] to this site might protect it from cleavage, similar to the action of PPR5 in the *trnG* group II intron [39].

Discussion

There is growing interest in exploiting the chloroplast gene expression system for biotechnological applications [4, 40-42]. However, improved methods to optimize, regulate, and tune expression of chloroplast transgenes are necessary to realize this promise. We show here that engineered PPR proteins can be used in conjunction with cognate *cis*-elements to activate plastid transgenes at the post-transcriptional level. Expression of the plastid transgene scales with that of the engineered PPR protein over a large dynamic range; this implies that the abundance of the PPR activator limits expression of the plastid transgene, and that the tunability of this system is limited primarily by the degree to which expression of the nuclear gene encoding the engineered PPR protein can be regulated. Experiments here were directed toward developing an inducible system, as a means to avoid negative impacts of recombinant proteins on plant vigor by separating the growth and

production phases. However, as discussed below, the results pave the way for tools to control the expression of plastid transgenes in other contexts.

Many chloroplast genes are activated by proteins whose mode of action is similar to that of PPR10 [20], and binding sites for several such proteins have been shown to stimulate transgene expression at the post-transcriptional level. This phenomenon was initially exemplified by the "intercistronic expression element" (IEE) [43] and its cognate binding protein HCF107 [44], and, more recently by several other native PPR binding sites [45]. Endogenous *cis*-elements like this provide tools for constitutively increasing transgene expression, but they can interfere with host functions by titrating the endogenous binding protein [45]. We selected PPR10 and its *atpH* target to develop an inducible system because *atpH* is expressed at very low levels in the absence of PPR10 [30], yet it is the most efficiently translated chloroplast gene in wild-type maize seedlings [31]. Furthermore, the sequence specificity of PPR10 and several PPR10 variants had been analyzed in depth [22, 32], facilitating the design of orthogonal activators. We predicted based on those data that the engineered PPR10/*atpH* system could achieve low basal expression (due to lack of crosstalk with tobacco PPR10) and high induced expression. These predictions were borne out by the results presented here.

Among the prior approaches for regulating plastid transgene expression, our method most closely resembles a system in the alga *C. reinhardtii* that exploits a native nucleus-encoded activating protein, NAC2, and its cognate chloroplast *cis*-element [46]. NAC2 is a helical repeat protein that is required specifically for the

expression of the chloroplast *psbD* gene. NAC2 acts via the *psbD* 5'UTR to stabilize *psbD* RNA and increase its translational efficiency [47, 48]. The *psbD* 5'UTR is sufficient to confer NAC2-dependent expression on any chloroplast ORF. In this system, a *NAC2* gene in the nucleus is put under the control of a vitamin-repressible riboswitch. This method has been used to silence specific chloroplast genes in order to address basic questions in chloroplast biology [46]. However, the dynamic range and maximal expression levels attainable with this system have not been reported. Furthermore, the use of the native *psbD* 5'-UTR results in the need to drive *psbD* expression with a different UTR.

In our highest PPR10*-expressing lines, the abundance of plastid-encoded GFP approached that of the large subunit of Rubisco, and was more than 40-fold higher than that in the absence of the cognate PPR protein. We estimate that GFP reached 25% of total soluble protein in our highest expressing material. This is a low estimate of the potential output of this system because our dose-response experiments did not reach an induction plateau. Our ethanol-inducible system yielded ~20-fold induction (~15% total soluble protein) after 4 days, despite the fact that we made little attempt to optimize the ethanol induction step. We are confident, therefore, that optimization of this method and advances in the design of regulatable nuclear promoters will translate to even greater dynamic range. But even without further optimization, these are substantial improvements over the ~1.75% total leaf protein and ~7-fold induction reported from the theophylline-inducible riboswitch system, which is the most promising system described previously [14, 17, 19]. In the

unusual situation in which a foreign protein compromises plant growth even at the low basal expression level reported here, expression could be scaled back by replacing the *rrn* promoter on the chloroplast transgene construct with a less active promoter.

These results lay a foundation for related approaches that could be used to control where, when, and to what extent plastid transgenes are expressed. For example, it may be possible to bypass the need for an inducer by driving PPR10* expression with a promoter that is activated at a suitably late stage of plant development. Tissue-specific expression of a nuclear PPR activator could be used to boost the expression of plastid transgenes in non-photosynthetic organs, an approach that we recently validated for potato tubers (Yu et al, accompanying manuscript). A more sophisticated goal is the tuning of relative expression levels of transgenes in a polycistronic transcription unit, a parameter that is important for the optimal performance of multienzyme pathways and the assembly of multicomponent structures [7, 12]. Such tuning may be achievable by small changes in *cis*-elements that modulate affinity for a single activating protein, or by use of multiple activating proteins with distinct *cis*-elements. The increasing repertoire of native PPR proteins with well-defined binding sites provides an expanding toolkit to draw upon for these and other applications that exploit the unique attributes of the chloroplast genetic system.

Materials and Methods

Plant material.

Nicotiana tabacum Petit Havana (tobacco) seeds were sown on agar plates containing Murashige and Skoog Basal Medium (Sigma), 0.3% Phytagel (Sigma), 2% sucrose. When working with transplastomic lines, the medium included 500 mg/ml Spectinomycin HCL to select for the *aadA* marker. Plants were grown for 2 weeks in a growth chamber at 25°C and a light intensity of ~150 μmol m^{-2} s^{-1} with 14-h light/10-h dark cycles. Seedlings were then transplanted to soil and grown under the same conditions prior to tissue sampling for protein and RNA analysis. Plants used for crossing were grown to flowering in a green house.

Chloroplast reporter constructs and transformation.

Plasmids pQY1 and pAI5 contain the dicistronic reporters Dicis^{AA} and Dics^{GG}, respectively, as diagrammed in Fig. 3-1 and Fig. 3-6. The dicistronic cassettes were inserted into the tobacco plastid transformation vector pPRV1 (Accession U12809) between the P$_{rrn}$L$_{atpB}$ promoter/5'UTR and the *psbA* 3'-UTR/terminator. The *aadA* gene is the same as that in plasmid pHK30 [49]. The *aadA* ORF is preceded by the *atpB* leader and starts with the first 14 codons from *atpB*. The *aadA* ORF is fused at its 3'-end with sequences encoding the c-myc 9E10 epitope EQKLISEEDL, as in the Arabidopsis plastid transformation vector pATV1 (GenBank Accession MF461355). The second ORF encodes GFP as in vector pATV1. The intergenic region consists of the tobacco *rbcL* 3'-UTR (T$_{rbcL}$) [50] followed by the 97 nucleotides

upstream of the maize chloroplast *atpH* gene (Genbank NC_001666), excepting the two nucleotide substitutions in the PPR10 binding site as shown in Fig. 3-1. The monocistronic[GG] reporter (construct pQY3) was inserted into the transformation vector pPRV111A (Accession No. U12812), with the *gfp* ORF preceded by the same modified *atpH* 5' UTR as for the dicistronic GG construct.

The reporter constructs were introduced into the tobacco chloroplast genome using the method described previously [51]. Transplastomic plants were selected by growth in the presence of 500 µg/ml Spectinomycin HCL, and all seedlings from lines used for the reported experiments remained green in the presence of Spectinomycin. Homoplasmy was confirmed by Southern blot hybridization (Fig. 3-6c).

Nuclear transgenes encoding PPR10 variants.

The nuclear *PPR10^AA* and *PPR10^GG* transgenes encode proteins that are identical to maize PPR10 except for the two amino acid substitutions shown in Fig. 3-1. The first 1177 nucleotides from the start codon were codon optimized for Arabidopsis. The remaining 1184 nucleotides to the stop codon used native maize *ppr10* codons. A Kozak consensus sequence was introduced just upstream of the start codon to optimize translation. The expression constructs for the PPR10 variants are diagrammed in Fig. 3-6b. PPR10[AA] was expressed from a CaMV 35S promoter in pEarlyGate100 [52]. PPR10[GG] was inserted into vector pCHSP6871 for expression from the soybean heat shock promoter, and into vector pMLBart AlcR for

expression from an ethanol inducible promoter. pEarlyGate100 was obtained from the Arabidopsis Biological Research Center. pCHSP6871 and pMLBart AlcR were generous gifts of Dr Steve Strauss (Oregon State University) and Detlef Weigel (Max Planck Institute for Developmental Biology), respectively. Transformations were performed at the Plant Transformation Core facilities at the University of Missouri and the University of California-Davis. Plants harboring the heat shock construct were selected by plating on agar plates (0.3% Phytagel) containing MS medium, 2% sucrose and kanamycin at 100 mg/liter. Plants harboring the CaMV 35S and ethanol-inducible constructs were selected by spraying 2-week old seedlings with 0.028% Finale and 0.1% Silwet 77; plants were sprayed two times, one-week apart.

Ethanol induction.

Induction was initiated by spraying plants with 4% ethanol until all leaves were wet (t=0). The sprayed plants were then incubated in ethanol vapor for 24 h as described in [53]. After an additional 48 h (t= 72 h), plants were sprayed again with 4% ethanol and incubated for an additional 24 h without vapor treatment. Samples and photos were collected 4 days after the initial spraying.

SDS-PAGE and immunoblot analysis.

Proteins from transplastomic samples were analyzed by SDS-PAGE using 4-20% polyacrylamide gels (Invitrogen). Immunoblotting was performed as described

previously, using affinity purified antibody raised against PPR10 in maize [28]. This antibody detects the maize PPR10 variants but does not cross-react with the endogenous tobacco PPR10 ortholog (see e.g. Fig. 3-2). GFP was detected with a monoclonal antibody for GFP (Clontech, cat #632380). To estimate the abundance of GFP as a fraction of the soluble protein in leaf extract, membranes were pelleted by centrifugation at 13000xg for 5 minutes, and the supernatants were resolved by SDS-PAGE. Gels were stained with Coomassie Blue R-250. The abundance of the GFP band as a fraction of the total was estimated by scanning the gels with a Typhoon Imager (model 2.0.0.6, Amersham) and quantification with ImageQuantTL.

RNA analyses.

RNA extractions and RNA gel blot hybridizations were performed as described previously [28], using a *gfp* probe generated by PCR with the primer pair: forward- 5' GGAGAAGAACTTTTCACTGGAGTTGTCCC and reverse 5' GATAATGGTCTGCTAGTTGAACGCTTCC. Primer extension was carried out as described previously [28], using 3 μg of leaf RNA and primer 5'GACAACTCCAGTGAAAAGTTCTTCTCC3'. Results were imaged with a Storm PhosphorImager (Molecular Dynamics).
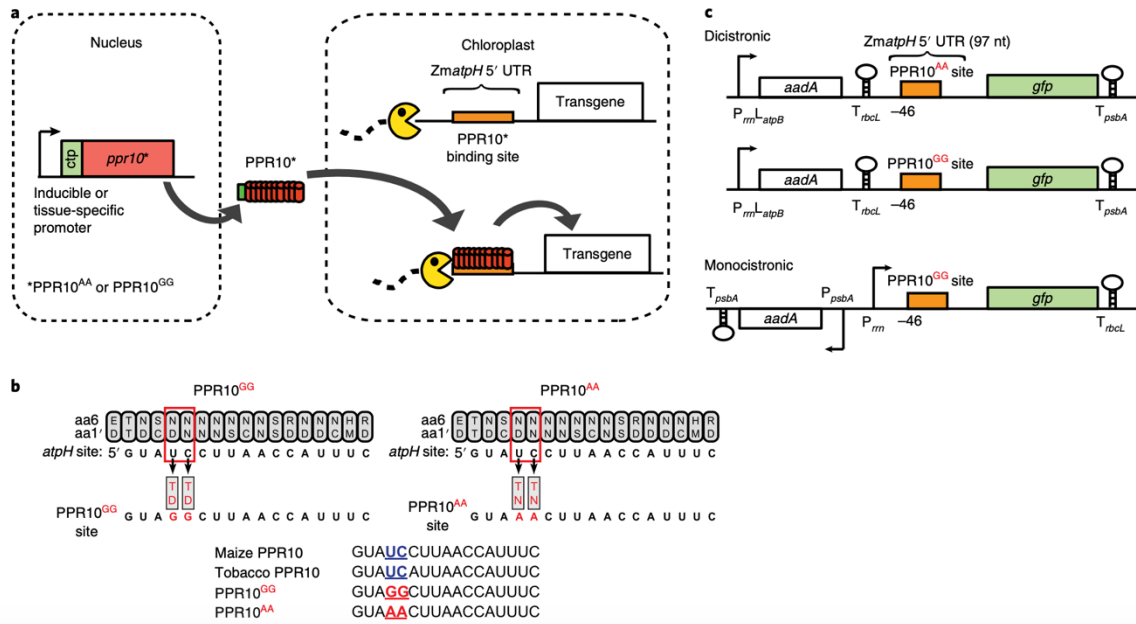
Figure 3-1. PPR10^{GG} dependent plastid transgene regulation. (a) Scheme overview. A chloroplast transgene is preceded by the maize *atpH* 5'-UTR with mutations in the PPR10 binding site to prevent its interaction with endogenous tobacco PPR10. PPR10 variants with a sequence-specificity matching the modified *cis*-element are expressed from a nuclear transgene, and targeted to the chloroplast with a chloroplast transit peptide (ctp), where they boost expression of the plastid transgene by blocking RNA degradation and increasing translational efficiency. Regulated expression of the PPR10 variants via a tissue-specific or inducible promoter is predicted to result in corresponding regulation of the plastid transgene.

(b) The PPR10^{GG} and PPR10^{AA} variants. PPR10's PPR motifs are diagrammed at top, with the amino acids at their specificity-determining positions (aa6 and 1') indicated [22]. The variants have mutations in two PPR motifs that change their nucleotide specificity as shown. The binding sites for maize PPR10, tobacco PPR10, and the two variants are shown below.

(c) Chloroplast reporter constructs. *aadA* is the selectable marker for chloroplast transformation. The *rbcL* and *psbA* 3'-UTRs (T$_{rbcL}$ and T$_{psbA}$) are illustrated as stem-loops. L*atpB* is the 5'-UTR and first few codons of the tobacco *atpB* mRNA and P$_{rrn}$ is the tobacco *rrn* promoter. The position of the 5'-end of the PPR10* binding site (-46) is indicated with respect to the GFP start codon. The transgenes were integrated between *trnV* and *rps12* in the inverted repeat of the chloroplast genome (see Fig. 3-6a). Features are not drawn to scale.
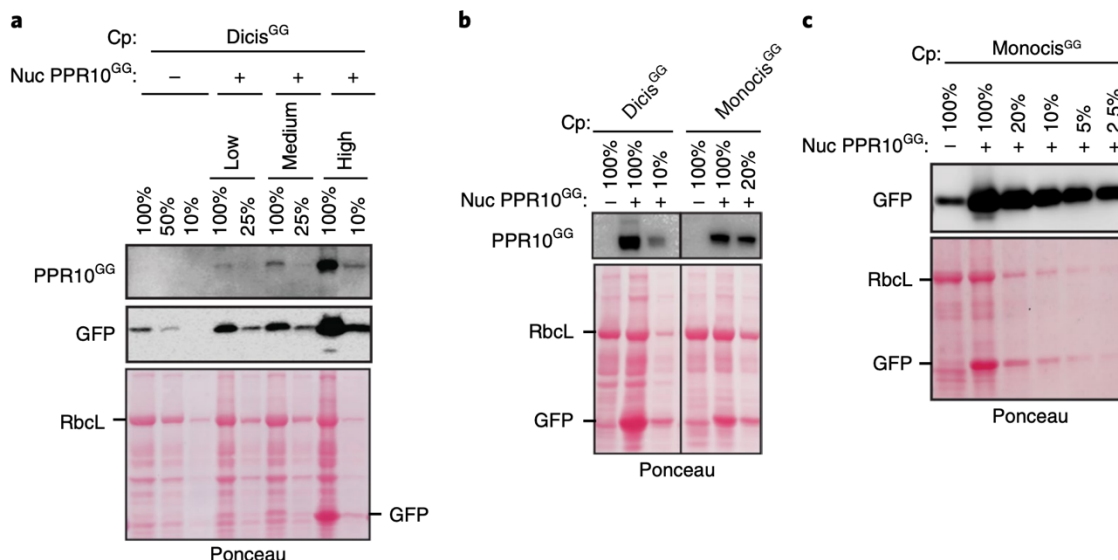
Figure 3-2. Quantitative relationship between PPR10^{GG} expression and expression of plastid GFP reporters with the GG *cis*-element. Images were cropped to show only the relevant part of each gel. Tissue was sampled from 4-week old plants.

(a) Immunoblot analysis of lines constitutively expressing different levels of PPR10^{GG}. Lanes contained equal quantities of total leaf protein or the indicated dilutions. The stained blot at bottom was probed consecutively to detect either PPR10^{GG} (top) or GFP (middle). RbcL is the large subunit of Rubisco.

(b) Comparison of the effects of PPR10^{GG} on the expression of a reporter in a monocistronic and dicistronic context. The experiment was performed as in panel (a). The vertical line separates non-contiguous lanes from the same gel and image.

(c) Dilution series illustrating >40- fold-induction from the monocistronic reporter in a constitutively high-expressing PPR10^{GG} line.
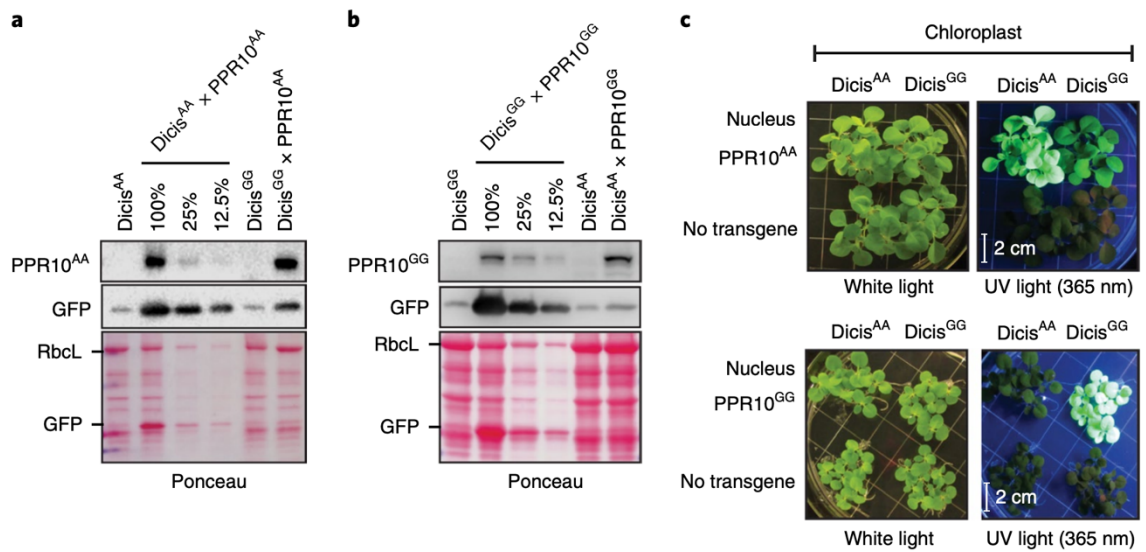
Figure 3-3. Specificity of PPR10^{GG} and PPR10^{AA} for plastid transgenes harboring the cognate binding site. Images were cropped to show only the relevant part of each gel.

Immunoblot analysis comparing effects of PPR10^{AA} (a) and PPR10^{GG} (b) on production of GFP from dicistronic constructs harboring the AA or GG *cis*-element. Samples come from plants with the indicated plastid reporter (Dicis^{AA} and Dicis^{GG}), in combination with the indicated nuclear transgene (PPR10^{AA}, PPR10^{GG}, or none). Leaf tissue was sampled four weeks after planting. See Fig. 3-2a for other details. (c) GFP fluorescence illustrating effects of each PPR10 variant on GFP output from each dicistronic reporter in young seedlings (18 days after planting).
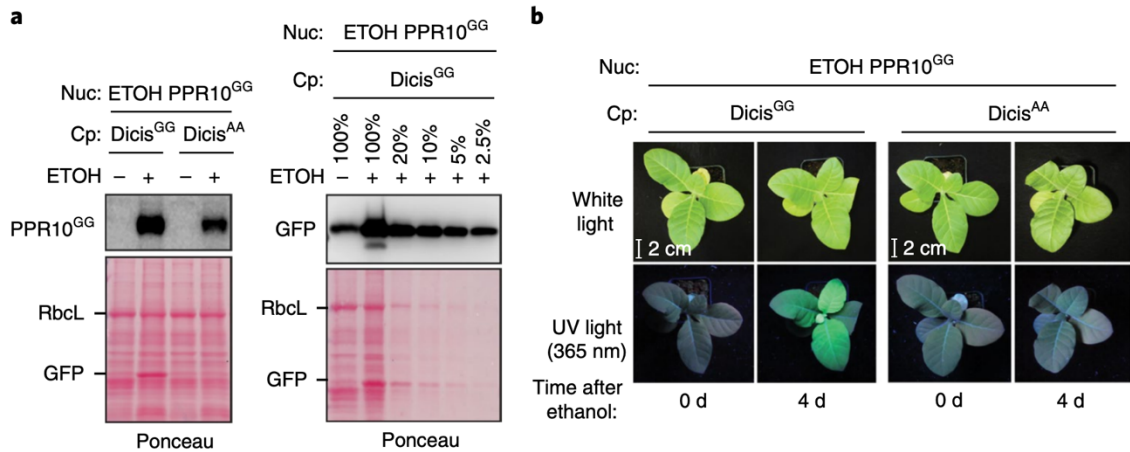
Figure 3-4. Ethanol-inducible expression of a plastid GFP reporter in a dicistronic context. Images were cropped to show only the relevant part of each gel. Leaf tissue was sampled eight weeks after planting. Analogous data for 10-week old plants are shown in Fig. 3-7b. (a) Immunoblot analysis demonstrating effects of ethanol on expression from an ethanol-inducible nuclear transgene encoding PPR10^{GG} and dicistronic chloroplast *gfp* reporters (Dicis^{GG} or Dicis^{AA}). Tissue samples were harvested from the same plants 4 days after ethanol treatment (+) or immediately prior to ethanol treatment (-). The dilution series to the right indicates that GFP expression from the Dicis^{GG} construct was induced approximately 20-fold.

(b) GFP fluorescence of plants used for the protein analysis shown in (a). Note that GFP fluorescence was muted by chlorophyll autofluorescence, which was not filtered out.
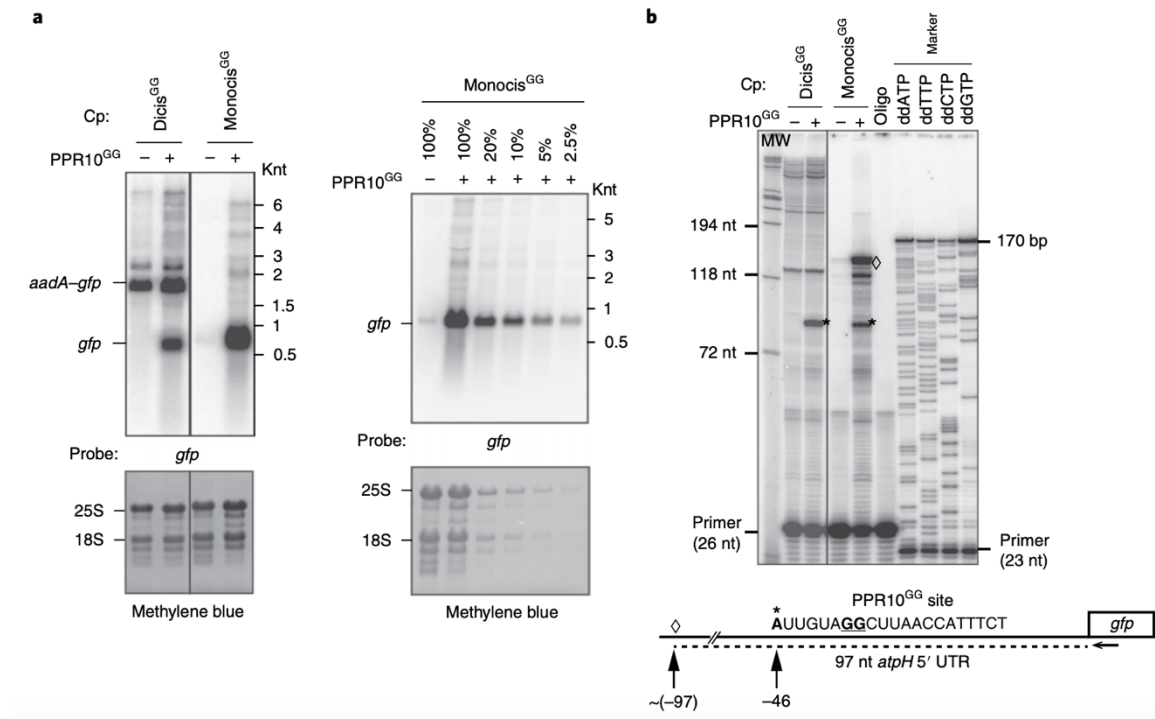
Figure 3-5. Effects of PPR10^{GG} on RNA transcripts from the monocistronic and dicistronic chloroplast reporters harboring the GG-*cis* element. Images were cropped to show only the relevant part of each gel. (a) RNA gel blot hybridizations showing *gfp* transcripts in the presence (+) and absence (-) of a constitutively-expressed PPR10^{GG} transgene. The RNA came from the "high-expressing" PPR10^{GG} line shown in Fig. 3-2a. Blots were hybridized with a probe specific for the *gfp* transgene. Images of the same blots stained with methylene blue are shown below to illustrate the abundance of rRNAs as a loading control. The line separates non-contiguous lanes from the same gel. (b) Primer extension mapping of transcript 5'-termini that increase in the presence of PPR10^{GG}. The sequencing ladders to the right come from a different gene; these provide markers to determine the sizes of the primer extension products at single-nucleotide resolution. The line separates

non-contiguous lanes from the same gel. The asterisk marks an RNA terminus at -46 nucleotides with respect to the start codon, which corresponds to the 5'-end of the PRP10 footprint. The diamond marks a 5' end ~100-nucleotides upstream of the start codon, the site of transcription initiation from the monocistronic reporter construct. The horizontal arrow in the map below denotes the primer.
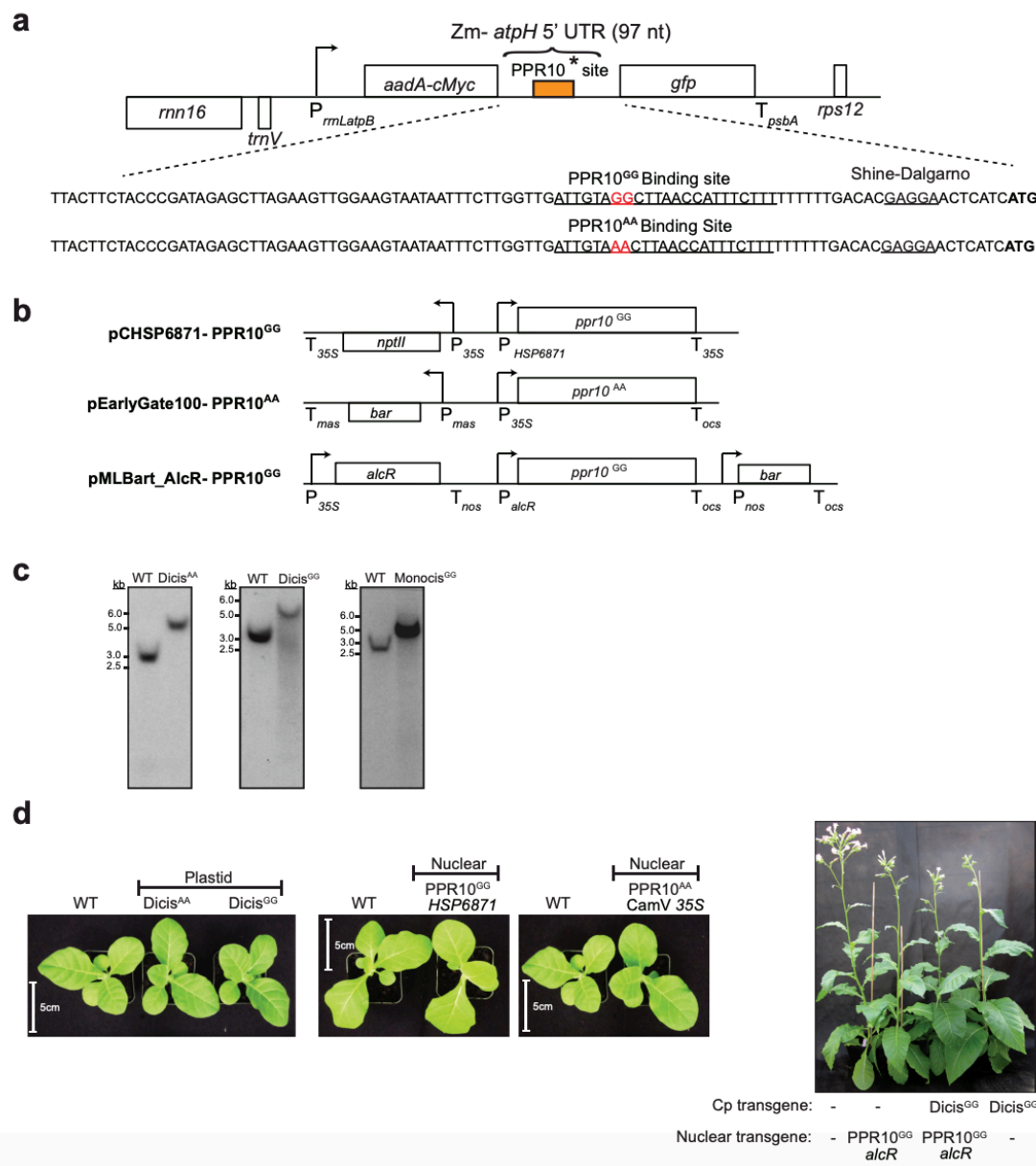
Figure 3-6. Plastid and nuclear transformation. (a) Context of dicistronic reporter constructs after insertion in the chloroplast genome. The monocistronic construct was inserted at this same position. (b) Constructs for expression of PPR10 variants from the nuclear genome. (c) DNA gel blot hybridizations documenting homoplasmy of the transplastomic lines. Leaf DNA was digested with BamHI and analyzed by hybridization to a radiolabeled probe specific for the tobacco *rrn16*

gene. The expected sizes of the WT and insertion alleles are 3.3 kb and 5.6 kb, respectively. (d) Phenotype of plants harboring the plastid reporter constructs or expressing PPR10^{GG} and PPR10^{AA} transgenes. The PPR10^{GG} HSP plant is from the high expressing line shown in Fig. 3-2a but lacking a chloroplast transgene. The PPR10^{AA} CamV plant is from the line shown in Fig. 3-3a but lacking the chloroplast transgene. The PPR10^{GG} plants to the right contain the ethanol inducible nuclear transgene (*alcR* promoter). The plants were grown for 6 weeks (left) or 13 weeks (right) in a greenhouse.
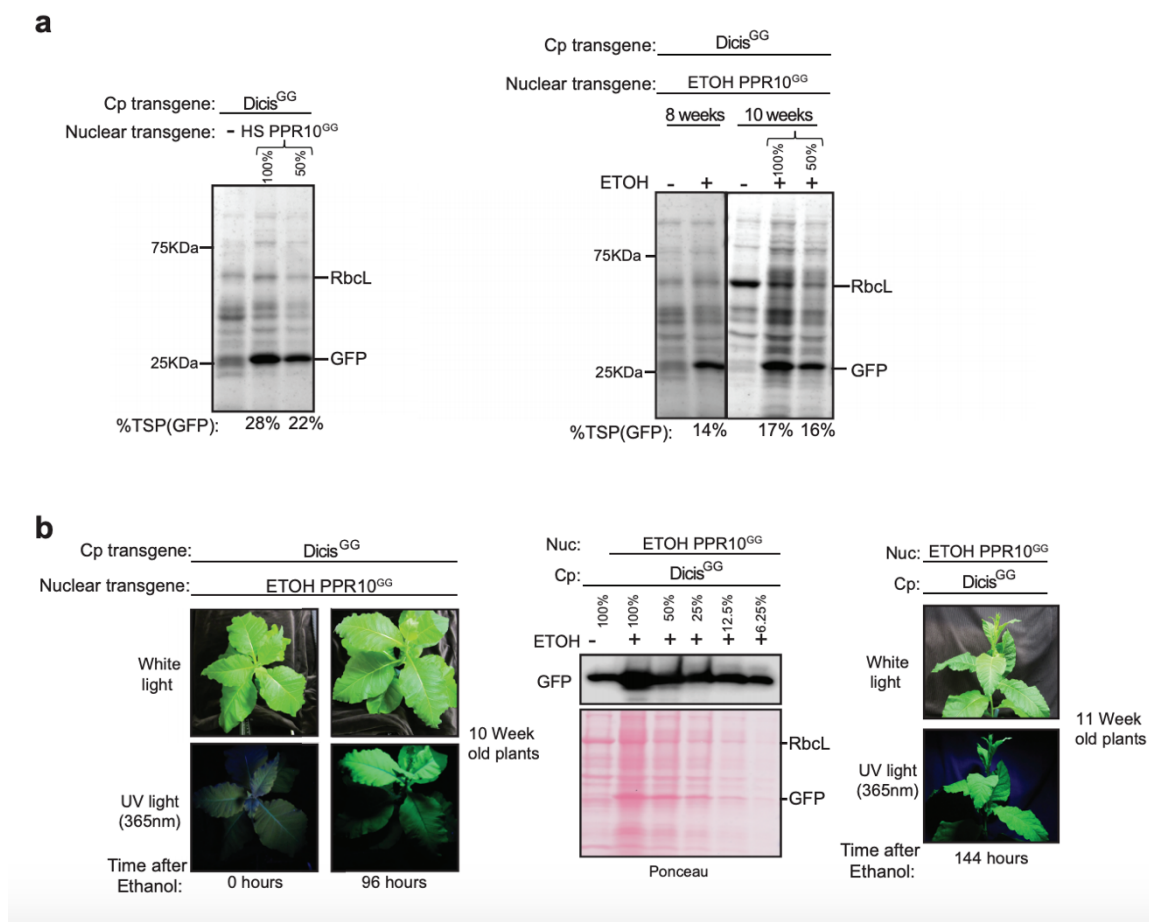
Figure 3-7. Abundance and ethanol induction of GFP from the dicistronic GG reporter in mature leaf tissue. (a) Estimate of GFP abundance as a fraction of Total Soluble Protein (TSP). Soluble extracts of the mature leaves of 4-week old plants (left panel), or 8 and 10-week old plants (right panel) were fractionated by SDS-PAGE and stained with Coomassie Brilliant Blue R-250. Dilutions of two samples (50%) were included to test for reproducibility. The stained gels were scanned with a Typhoon Imager (model 2.0.0.6, Amersham), and band intensities were quantified by ImageQuantTL. (b) Ethanol induction of GFP expression in 10-week old plants harboring the dicistronic GG reporter construct. Plants were imaged just prior to

ethanol treatment and 96 hours after treatment. The dilution series on the western

blot is shown to estimate fold-induction.

References

1.   Boynton, J.E. et al. Chloroplast transformation in *Chlamydomonas* with high velocity microprojectiles. *Science* **240**, 1534-1538 (1988).
2.   Svab, Z., Hajdukiewicz, P. & Maliga, P. Stable transformation of plastids in higher plants. *Proc Natl Acad Sci USA* **87**, 8526-8530 (1990).
3.   Bock, R. Engineering plastid genomes: methods, tools, and applications in basic research and biotechnology. *Annu Rev Plant Biol* **66**, 211-241 (2015).
4.   Ahmad, N., Michoux, F., Lossl, A.G. & Nixon, P.J. Challenges and perspectives in commercializing plastid transformation technology. *J Exp Bot* **67**, 5945-5960 (2016).
5.   Waheed, M.T., Ismail, H., Gottschamel, J., Mirza, B. & Lossl, A.G. Plastids: The Green Frontiers for Vaccine Production. *Front Plant Sci* **6**, 1005 (2015).
6.   Oey, M., Lohse, M., Kreikemeyer, B. & Bock, R. Exhaustion of the chloroplast protein synthesis capacity by massive expression of a highly stable protein antibiotic. *Plant J* **57**, 436-445 (2009).
7.   Fuentes, P. et al. A new synthetic biology approach allows transfer of an entire metabolic pathway from a medicinal plant to a biomass crop. *Elife* **5** (2016).
8.   Gnanasekaran, T. et al. Transfer of the cytochrome P450-dependent dhurrin pathway from Sorghum bicolor into Nicotiana tabacum chloroplasts for light-driven synthesis. *J Exp Bot* **67**, 2495-2506 (2016).
9.   Harada, H. et al. Construction of transplastomic lettuce (Lactuca sativa) dominantly producing astaxanthin fatty acid esters and detailed chemical analysis of generated carotenoids. *Transgenic Res* **23**, 303-315 (2014).
10.   Bohmert-Tatarev, K., McAvoy, S., Daughtry, S., Peoples, O.P. & Snell, K.D. High levels of bioplastic are produced in fertile transplastomic tobacco plants engineered with a synthetic operon for the production of polyhydroxybutyrate. *Plant Physiol* **155**, 1690-1708 (2011).
11.   Diretto, G. et al. Metabolic engineering of potato carotenoid content through tuber-specific overexpression of a bacterial mini-pathway. *PLoS One* **2**, e350 (2007).
12.   Hanson, M.R., Lin, M.T., Carmo-Silva, A.E. & Parry, M.A. Towards engineering carboxysomes into C3 plants. *Plant J* **87**, 38-50 (2016).
13.   Malhotra, K. et al. Compartmentalized Metabolic Engineering for Artemisinin Biosynthesis and Effective Malaria Treatment by Oral Delivery of Plant Cells. *Mol Plant* **9**, 1464-1477 (2016).
14.   Lossl, A. et al. Inducible trans-activation of plastid transgenes: expression of the R. eutropha phb operon in transplastomic tobacco. *Plant Cell Physiol* **46**, 1462-1471 (2005).
15.   Lu, Y., Rijzaani, H., Karcher, D., Ruf, S. & Bock, R. Efficient metabolic pathway engineering in transgenic tobacco and tomato plastids with synthetic multigene operons. *Proc Natl Acad Sci U S A* (2013).

16. Scotti, N. & Cardi, T. Transgene-induced pleiotropic effects in transplastomic plants. *Biotechnol Lett* **36**, 229-239 (2014).

17. Muhlbauer, S.K. & Koop, H.U. External control of transgene expression in tobacco plastids using the bacterial lac repressor. *Plant J* **43**, 941-946 (2005).

18. Buhot, L., Horvath, E., Medgyesy, P. & Lerbs-Mache, S. Hybrid transcription system for controlled plastid transgene expression. *Plant J* **46**, 700-707 (2006).

19. Emadpour, M., Karcher, D. & Bock, R. Boosting riboswitch efficiency by RNA amplification. *Nucleic Acids Res* **43**, e66 (2015).

20. Barkan, A. & Small, I. Pentatricopeptide Repeat Proteins in Plants. *Annu Rev Plant Biol* **65**, 415-442 (2014).

21. Small, I. & Peeters, N. The PPR motif - a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem Sci* **25**, 46-47 (2000).

22. Barkan, A. et al. A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet* **8**, e1002910 (2012).

23. Shen, C. et al. Structural basis for specific single-stranded RNA recognition by designer pentatricopeptide repeat proteins. *Nat Commun* **7**, 11285 (2016).

24. Filipovska, A. & Rackham, O. Modular recognition of nucleic acids by PUF, TALE and PPR proteins. *Mol Biosyst* **8**, 699-708 (2012).

25. Kindgren, P., Yap, A., Bond, C.S. & Small, I. Predictable alteration of sequence recognition by RNA editing factors from Arabidopsis. *Plant Cell* **27**, 403-416 (2015).

26. Colas des Francs-Small, C., Vincis Pereira Sanglard, L. & Small, I. Targeted cleavage of nad6 mRNA induced by a modified pentatricopeptide repeat protein in plant mitochondria. *Commun Biol* **1**, 166 (2018).

27. Miranda, R.G., McDermott, J.J. & Barkan, A. RNA-binding specificity landscapes of designer pentatricopeptide repeat proteins elucidate principles of PPR-RNA interactions. *Nucleic Acids Res* **46**, 2613-2623 (2018).

28. Pfalz, J., Bayraktar, O., Prikryl, J. & Barkan, A. Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts. *EMBO J* **28**, 2042-2052 (2009).

29. Prikryl, J., Rojas, M., Schuster, G. & Barkan, A. Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proc Natl Acad Sci USA* **108**, 415-420 (2011).

30. Zoschke, R., Watkins, K. & Barkan, A. A rapid microarray-based ribosome profiling method elucidates chloroplast ribosome behavior in vivo *Plant Cell* **25**, 2265-2275 (2013).

31. Chotewutmontri, P. & Barkan, A. Dynamics of Chloroplast Translation during Chloroplast Differentiation in Maize. *PLoS Genet* **12**, e1006106 (2016).

32. Miranda, R.G., Rojas, M., Montgomery, M.P., Gribbin, K.P. & Barkan, A. RNA-binding specificity landscape of the pentatricopeptide repeat protein PPR10. *RNA* **23**, 586-599 (2017).

33. Schoffl, F., Raschke, E. & Nagao, R.T. The DNA sequence analysis of soybean heat-shock genes and identification of possible regulatory promoter elements. *EMBO J* **3**, 2491-2497 (1984).

34. Roslan, H.A. et al. Characterization of the ethanol-inducible alc gene-expression system in Arabidopsis thaliana. *Plant J* **28**, 225-235 (2001).

35. Erb, T.J. & Zarzycki, J. A short history of RubisCO: the rise and fall (?) of Nature's predominant CO2 fixing enzyme. *Curr Opin Biotechnol* **49**, 100-107 (2018).

36. Hui, M.P., Foley, P.L. & Belasco, J.G. Messenger RNA degradation in bacterial cells. *Annu Rev Genet* **48**, 537-559 (2014).

37. Germain, A., Hotto, A.M., Barkan, A. & Stern, D.B. RNA processing and decay in plastids. *Wiley Interdiscip Rev RNA* **4**, 295-316 (2013).

38. Luro, S., Germain, A., Sharwood, R.E. & Stern, D.B. RNase J participates in a pentatricopeptide repeat protein-mediated 5' end maturation of chloroplast mRNAs. *Nucleic Acids Res* **41**, 9141-9151 (2013).

39. Beick, S., Schmitz-Linneweber, C., Williams-Carrier, R., Jensen, B. & Barkan, A. The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts. *Mol Cell Biol* **28**, 5337-5347 (2008).

40. Hanson, M.R., Gray, B.N. & Ahner, B.A. Chloroplast transformation for engineering of photosynthesis. *J Exp Bot* **64**, 731-742 (2013).

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

42. Daniell, H., Chan, H.T. & Pasoreck, E.K. Vaccination via Chloroplast Genetics: Affordable Protein Drugs for the Prevention and Treatment of Inherited or Infectious Human Diseases. *Annu Rev Genet* **50**, 595-618 (2016).

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

44. Hammani, K., Cook, W. & Barkan, A. RNA binding and RNA remodeling activities of the Half-a-Tetratricopeptide (HAT) protein HCF107 underlie its effects on gene expression. *PNAS* **109**, 5651-5656 (2012).

45. Legen, J. et al. Stabilization and translation of synthetic operon-derived mRNAs in chloroplasts by sequences representing PPR protein-binding sites. *Plant J* **94**, 8-21 (2018).

46. Ramundo, S. & Rochaix, J.D. Controlling expression of genes in the unicellular alga Chlamydomonas reinhardtii with a vitamin-repressible riboswitch. *Methods Enzymol* **550**, 267-281 (2015).

47. Boudreau, E., Nickelsen, J., Lemaire, S.D., Ossenbuhl, F. & Rochaix, J.D. The Nac2 gene of Chlamydomonas encodes a chloroplast TPR-like protein involved in psbD mRNA stability. *EMBO J* **19**, 3366-3376 (2000).

48. Kuchka, M.R., Goldschmidt-Clermont, M., van Dillewijn, J. & Rochaix, J.D. Mutation at the Chlamydomonas nuclear NAC2 locus specifically affects

stability of the chloroplast psbD transcript encoding polypeptide D2 of PS II. *Cell* **58**, 869-876 (1989).

49. Kuroda, H. & Maliga, P. Sequences downstream of the translation initiation codon are important determinants of translation efficiency in chloroplasts. *Plant Physiol* **125**, 430-436 (2001).

50. Shinozaki, K. & Sugiura, M. Sequence of the intercistronic region between the ribulose-1, 5-bisphosphate carboxylase/oxygenase large subunit and coupling factor beta subunit gene. *Nucleic Acids Res* **10**, 4923-4934 (1982).

51. Lutz, K.A., Svab, Z. & Maliga, P. Construction of marker-free transplastomic tobacco using the Cre-*loxP* site-specific recombination system. *Nature Protocols* **1**, 900-910 (2006).

52. Earley, K.W. et al. Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J* **45**, 616-629 (2006).

53. Werner, S., Breus, O., Symonenko, Y., Marillonnet, S. & Gleba, Y. High-level recombinant protein expression in transgenic plants by using a double-inducible viral vector. *Proc Natl Acad Sci U S A* **108**, 14061-14066 (2011).

CHAPTER 4

Engineered RNA-binding protein for transgene activation in non-green plastids

Abstract

Non-green plastids are desirable for the expression of recombinant proteins in edible plant parts to enhance the nutritional value of tubers or fruits or to deliver pharmaceuticals. However, plastid transgenes are expressed at extremely low levels in the amyloplasts of storage organs such as tubers[1-3]. Here we report a regulatory system consisting of a variant of the maize RNA binding protein PPR10 and a cognate binding site upstream of a plastid transgene encoding GFP. The binding site is not recognized by the resident potato PPR10 protein, restricting GFP protein accumulation to low levels in leaves. When the PPR10 variant is expressed from the tuber-specific patatin promoter, GFP accumulated up to 1.3% of total soluble protein, a 60-fold increase over 0.02%[2], the maximum protein yield achieved to date in potato tuber. This regulatory system enables an increase in transgene expression in non-photosynthetic plastids without interfering with chloroplast gene expression in leaves.

Introduction

The plastid genome of flowering plants contains about 110 genes, most of which encode components of the plastid transcription-translation machinery or genes required for photosynthesis. These plastid genes are complemented by ~ 3,000 nuclear genes that contribute to most functions required for photosynthesis, plastid gene expression, and various chloroplast-localized metabolic pathways [4]. Members of the plastid family are found in all plant cell types, but the plastid gene expression system is most active in the chloroplasts in leaves.

In contrast, chromoplasts and amyloplasts in fruits and tubers express only a few genes, such as *accD* required for fatty acid biosynthesis[5,6].

Chloroplast genome engineering holds great promise in many biotechnological applications, including production of vaccines and antibodies[7], conferring insect tolerance by expressing dsRNAs[8], implanting novel metabolic pathways[9,10] and improving photosynthetic efficiency[11-13]. All plastid types have identical genomes [14]. Non-green plastids are desirable hosts for the expression of recombinant proteins in edible plant parts to enhance the nutritional value of tubers or fruits. Seed-specific expression of vaccines may also be desirable when the seed is used as fodder. Attempts to express transgenes in non-green potato plastids (amyloplasts) thus far resulted in very low level recombinant protein, 0.02% of TSP[2] in tubers and 0.05% in microtubers[3]. We report here an expression system that stimulates protein expression in tuber amyloplasts with no significant impact on protein levels in leaves.

Results and Discussion

Our approach takes advantage of the maize protein PPR10, which activates expression of the chloroplast *atpH* gene by binding a cis-element in the *atpH* 5'UTR. This interaction stabilizes the RNA downstream and increases translational efficiency[15,16]. We placed the *gfp* plastid transgene downstream of the binding site for a PPR10 variant, denoted PPR10[GG], whose sequence differs from that of the endogenous potato PPR10 binding site (BS) at three positions: compare St-*atpH* 5' UTR and Zm-*atpH* 5' UTR (GG site) in Fig. 4-1b. The GG site is not recognized by maize PPR10[17], and we predicted the same would be true for endogenous potato PPR10[St]. The second component is a nuclear transgene encoding PPR10[GG,17] expressed from the tuber specific patatin promoter [18] (Fig. 4-4). PPR10[GG] has an N-terminal plastid targeting sequence that guides the protein to plastids (Fig. 4-1a).

We incorporated the GG site in the intergenic region of a dicistronic operon [19], between the first Open Reading Frame (ORF) encoding the spectinomycin resistance (*aadA*) gene and the second gene encoding Green Fluorescent Protein (GFP) (Fig. 4-1c). Plastids evolutionarily derive from photosynthetic bacteria and maintained the prokaryotic transcription-translation machinery, including expression of genes in operons. The dicistronic arrangement mimics the natural context where the PPR10 binding site is located in an intergenic region of a polycistronic transcription unit. We predicted that expression of PPR10[GG] would increase expression of the *gfp* reporter by stabilizing the RNA and increasing translational efficiency, analogous to the effects of wild-type PPR10 on *atpH*

expression. This approach is the same as that we used to achieve inducible expression of chloroplast transgenes[20] except that PPR10^{GG} is expressed from the *patatin* promoter to drive expression specifically in potato tubers. As a control, we analyzed the effects of the wild-type maize PPR10 binding site (Zm site) on expression of the same reporter system.

Transplastomic potato plants with the *gfp* reporter constructs were obtained by selection for the spectinomycin resistance conferred by the *aadA* gene [2]. The *aadA-gfp* construct integrated in the plastid genome (ptDNA) by two homologous recombination events via ptDNA flanking the dicistronic cassette. Uniform transformation of the plastid genomes was confirmed by DNA gel blot analyses (Fig. 4-5). We report here on two transplastomic potato (*Solanum tuberosum* cv. Desiree) lines: one having a wild-type (Zm site) and one having a mutated (GG site) maize binding site upstream of the *gfp* translation initiation codon (Fig. 4-1c). The *PPR10^{GG}* gene was introduced in the potato nucleus by Agrobacterium-mediated transformation. Because the transgenes insert at random sites, their expression levels and regulation vary among independent transformants. We chose two transgenic lines for further study by screening lines for intense fluorescence in tubers and the lack of fluorescence in leaves upon illumination by UV light.

The leaf phenotype of wild-type potato (WT) and transplastomic potato with the mutant BS^{GG} cis-element are indistinguishable and are very similar to that of transplastomic plants containing the *PPR10^{GG}* nuclear gene (Fig. 4-2). By contrast, use of the Zm site, which is expected to recognize the endogenous potato PPR10,

caused a dramatic decrease in plant vigor. These phenotypes correlated with the amount of GFP accumulation in leaves, as indicated by GFP fluorescence (Fig. 4-2 UV). GFP accumulation could be readily detected in the tubers of Dicis^{Zm} and Dicis^{GG}+PPR10^{GG} plants upon illumination with UV light (Fig. 4-2).

GFP in these materials was then quantified by SDS-PAGE and immunoblot analysis (Fig. 4-3). A stained GFP band was readily apparent in the leaf of the transplastomic reporter line harboring the wild-type maize binding site, whereas a stained GFP band was not visible in leaf material from lines with the GG site (Fig. 4-3a). We calculated that GFP in the Dicis^{Zm} line makes up 23.5% of the total soluble leaf protein (see Fig. 3 legend). However, high levels of GFP accumulation comes at a price. GFP accumulation in Dicis^{Zm} leaves is accompanied by a dramatic reduction in leaf size and smaller stature (Fig. 4-2). Poor growth could be due to high levels of GFP or the sequestration of native potato PPR10 protein. There is less GFP in the Dicis^{Zm} tuber (~1.2% of TSP) (Fig. 4-3a, b, e), reflecting the lower abundance of the mRNA (Fig. 4-3f).

In the Dicis^{GG} plants much less GFP accumulates in the leaves and tubers, 1.7% and 0.06% of TSP, respectively (Fig. 4-3a, b, e)., because the mutated maize GG binding site does not interact with native potato PPR10 protein. When the tuber specific PPR10^{GG} protein is expressed in the Dicis^{GG} plants, GFP levels in the leaf are unaffected, as no PPR10^{GG} protein is expressed in these tissues. However, PPR10^{GG} expressed in the Dicis^{GG} tuber binds to the mutant GG site and boosts GFP

accumulation to 1.0% or 1.3% of TSP depending on the genomic location of PPR10^{GG} (Fig. 4-3c-e).

RNA gel blot analyses using *aadA* and *gfp* probes revealed accumulation of dicistronic *aadA-gfp* and monocistronic *aadA* and *gfp* mRNAs. Monocistronic mRNAs are generated by processing of the dicistronic message. Monocistronic *gfp* mRNA is stabilized by PPR10 binding at the 5'-end and by the stem-loop structure in the *psbA* 3'UTR. Thus, accumulation of the monocistronic *gfp* message depends on protection of the *gfp* message at the 5' end by PPR10 binding, as reported for PPR10 function in the native context [15,16]. The Dicis^{Zm} plants accumulate monocistronic message in the leaf and tuber (Fig. 4-3f, g), presumably because the native potato PPR10 protein recognizes the wild type maize Zm site, despite the single nucleotide mismatch in the binding site (Fig. 4-1b). No monocistronic *gfp* message accumulates in the Dicis^{GG} plants, because the mutant GG site is not protected by the native potato PPR10St protein. However, when the engineered PPR10^{GG} is present, it binds to the mutated site resulting in the accumulation of monocistronic *gfp* mRNA. Monocistronic *gfp* mRNA in the Dicis^{GG} transplastomic plants accumulates only in tubers (Fig. 4-3f, g) due to tuber-specific expression of PPR10^{GG} from the patatin promoter.

The Dicis^{GG} + PPR10^{GG} regulatory system in potato tubers yielded as much GFP protein from the mutated GG site as from the wild-type Zm site. This finding suggests that GFP accumulation is subject to the same limiting factor, possibly the level of *gfp* mRNA. The 1.3% (TSP) protein is significantly higher than protein

levels obtained in potato amyloplasts in the past: 0.02% (TSP) in tubers[2] and 0.05% (TSP) in microtubers[3]. An increase in GFP levels in the tuber was obtained without an obvious deleterious impact on vegetative growth. GFP accumulated at the periphery of amyloplasts because the starch granules are filling up most of the amyloplasts, so that the stroma is pressed to the periphery (Fig. 4-6). Since this PPR10 regulatory system is significantly more efficient than the chimeric expression elements tested so far in tuber amyloplasts, we anticipate that protein accumulation in tomato chromoplasts can also be increased beyond what has already been achieved[21]. Further enhancement of protein expression in potato tubers may be obtained by increasing transcript levels (Fig. 4-3f) and reducing starch levels to allow accumulation of other storage products[22,23]. The orthogonal binding site/PPR10 control strategy described here proves the feasibility of exploiting modified natural PPR proteins for the manipulation of RNA metabolism[24].

Materials and Methods

Plant materials and growth conditions.

Potato (*Solanum tuberosum*) cv. Desire 2-24 plants obtained from Dr. Teruko Osumi (Simplot Plant Science). The plants were maintained aseptically on MS medium with 3% sucrose. The plants were incubated at 26 °C and illuminated for 16 hrs with fluorescent tubes. Potato plants were grown in Metromix 360 (Sungrow, Agawam, MA) in a growth chamber at 20 °C, under continuous illumination with Philips F72/T12 CW fluorescent bulbs at 40 µE m^{-2} s^{-1}. Under these conditions the plants never flower and are a reproducible source of leaves and tubers for biochemical analyses.

Construction of plastid and nuclear transformation vectors.

The *aadA* and *gfp* open reading frames in the pAI3 (GenBank Accession Number MK482729) and pAI5 (MK482730) dicistronic vectors are expressed in a P*rrn*L*atpB* promoter/T*psbA* cassette in tobacco plastid transformation vector pPRV1 (U12809). Transcription of the operon is from the tobacco rRNA operon promoter fused with the *atpB* gene leader and 14 amino acids of the *atpB* coding region N-terminus as described for plasmid pHK30 [25]. T*psbA* is the 3'-UTR of the plastid *psbA* gene. The *aadA* gene at its N-terminus is fused with the *atpB* gene N-terminal amino acids and at its C-terminus with the c-myc 9E10 epitope EQKLISEEDL, as in the Arabidopsis plastid transformation vector pATV1 (GenBank Accession Number MF461355). The second ORF encodes the Green

Fluorescent Protein (GFP) as in vector pATV1. Dicistronic vectors differ in sequence between the *aadA* and *gfp* ORFs: pAI3, T*rbcL*+Dicis[Zm]; pAI5, T*rbcL*+Dicis[GG]. T*rbcL* is the 3' UTR of the plastid *rbcL* gene [26]. The maize Zm and GG sites are the wild-type and mutant maize PPR10 binding sites [17].

The nuclear PPR10[GG] transgene encodes a protein that is similar to maize PPR10 except for two amino acid substitutions[20]. The nucleotide sequence encoding *ppr10[GG]* was codon optimized for potato and expressed from the potato patatin promoter[18]. The B33 patatin promoter/ocs terminator cassette was obtained from Dr. Frank Ludewig (University Erlangen-Nuremberg) in a pBIN19 Agrobacterium binary vector[27]. The PPR10[GG] coding sequence [20] was inserted in the B33 cassette and introduced into a pGV3101 (pMP90) Agrobacterium strain [28].

Transformation of the plastid and nuclear genomes.

Potato plastid transformation was carried out following the protocol of Valkov et al. [2], as modified by Zhang et al. [8]. For transformation of the nucleus, Agrobacterium carrying the binary vector was grown from frozen stock and resuspended in 50 mL liquid Minimal A medium with 0.2 M acetosyringone. Approximately 50 1 cm^2 potato leaf piece per construct were submerged in Agrobacterium solution for 15 minutes, then blotted briefly on sterile filter paper and left for 48 hours on antibiotic-free RMOS medium in the absence of selection. The RMOS medium contains modified MS potato medium (4.4 g /L, Phytotechnology Labs M516), vitamins (10 ml of 100x, Phytotechnology Labs M533), 30 g/L sucrose, thiamine (1 mg/L), NAA

(0.1 mg/L), BAP (1mg/L), myo-inositol 0.1 g/L, solidified with 0.7 % agar (7 g/L) and adjusted with KOH to pH5.8. Leaf pieces were then placed on RMOS medium containing 50 mg/L kanamycin and 500 mg/L carbenicillin. After 10 weeks, resistant callus was moved to MGS shoot formation medium with 50 mg/L kanamycin selection and 500 mg/L carbenicillin for 8 weeks. The MGS shoot induction medium is a modified MS potato medium (4.4 g /L, Phytotechnology Labs M516), vitamins (10 ml of 100x, Phytotechnology Labs M533), 16 g/L glucose, zeatin riboside (2.2 mg/L), NAA (0.02 mg/L), gibberellic acid (0.15 mg/L), solidified with 0.7 % agar (7 g/L) and adjusted with KOH to pH5.8. After 8 weeks, the shoots were excised and moved to RMS rooting medium with 50 mg/L kanamycin and 500 mg/L carbenicillin.

DNA gel blot analysis.

Total cellular DNA was extracted from leaf tissue using the cetyltrimethylammonium bromide (CTAB) protocol. DNA gel blot analysis was carried out as described[29]. Briefly, two microgram total leaf DNA was digested with the *Bam*HI restriction endonuclease, the DNA fragments were separated on 1% agarose gels, and transferred to Hybond-N membranes (Amersham Biosciences Corp) by capillary blotting. The probe was prepared by random-primed ^{32}P-labeling the 1.9 kb *Apa*I/*Bam*HI ptDNA DNA fragment encoding part of the 16S rRNA gene.

RNA gel blot analysis.

Total cellular RNA was isolated from the leaves and tubers of plants grown in the soil with the NucleoSpin RNA Plant kit (Macherey-Nagel, Germany) following the manufacturer's instruction. RNA gel blot analyses were carried out as described [8]. The RNA was electrophoresed on 1.5% agarose/formaldehyde gels, and then transferred to Hybond-N membranes (Amersham Biosciences Corp). The probe was prepared by random-primed ^{32}P-labeling. The probes were: *aadA*, a ~0.8kb NcoI-XbaI fragment isolated from plasmid pHC1[30]; and for *gfp*, a fragment amplified from gfp coding region using primers p1(5'-TTTTCTGTCAGTGGAGAGGGTG-3') and p2 (5'-CCCAGCAGCTGTTACAAACT-3';).

Western blot analysis.

Total leaf and tuber protein were extracted from plants grown in the soil using the protocol as described[31]. About 100 mg of powdered leaf or tuber tissue was extracted in 200 μl of sodium phosphate buffer (100 mM, pH 6.8) containing 1mM EDTA, 5mM DTT and a protease inhibitor cocktail (30 μl) specific for plant tissue (Sigma-Aldrich). The leaf and tuber supernatants were harvested after centrifugation at 14,000g for 25 min at 4 °C. The concentration of leaf and tuber total soluble protein was determined with the BCA protein Assay reagent kit. GFP was detected by immunoblot analysis using monoclonal antibody. ZmPPR10^{GG} was detected by immunoblot analysis using affinity purified PPR10 antibody [15]. GFP in the Dicis^{Zm} (pAI3) leaf extract was quantified in Coomassie blue stained protein gels using AlphaImager 2200. GFP in tissues containing less GFP was quantified

by Western analyses using Dicis^{Zm} leaf extracts as reference. Each value is an average of measurements of three tissue samples.
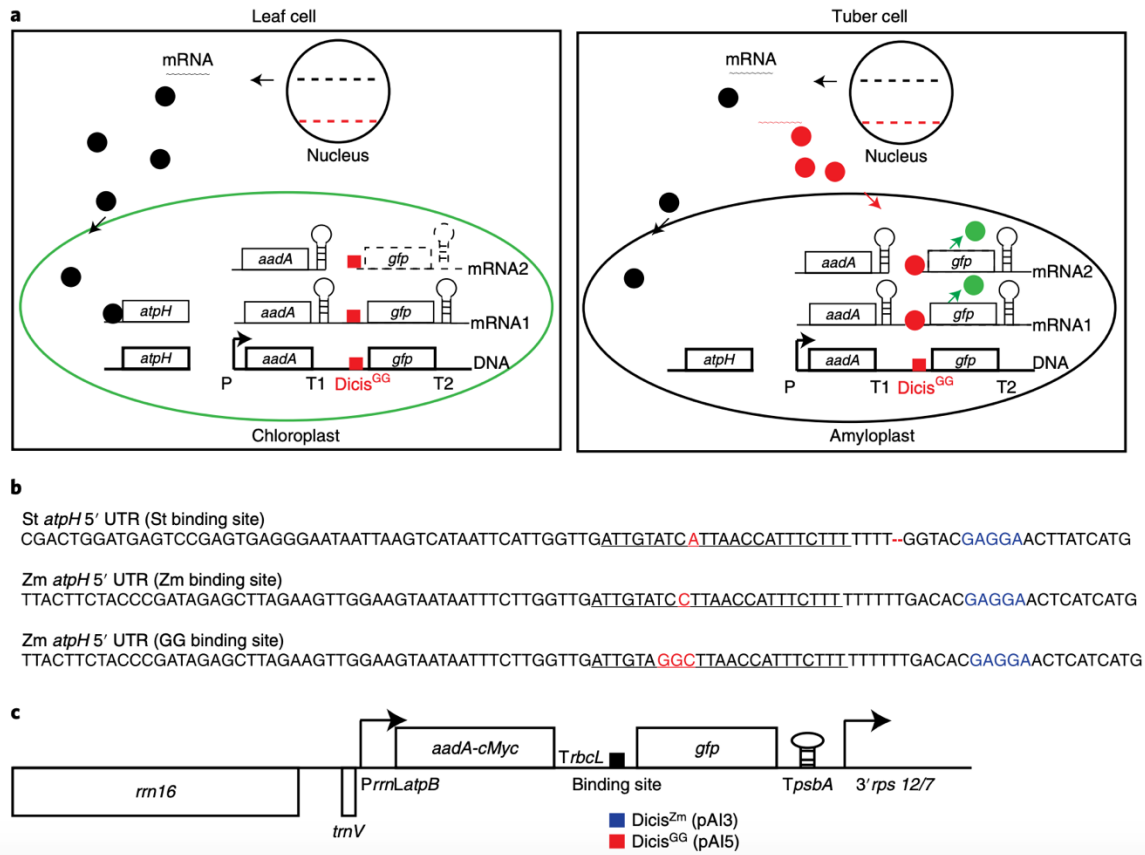
Figure 4-1. PPR10/binding site system for regulating gene expression in potato amyloplasts. a, Leaf cell with the nuclear and chloroplast compartments. Tuber cell with the nuclear and amyloplast compartments. Filled red circles are the engineered maize PPR10^{GG}; green circles are GFP. b, Alignment of sequence upstream of the *atpH* translation initiation codon (AUG) containing the wild type potato and the maize wild type (Zm) and mutant GG binding sites. The PPR10 Binding Sites are underlined. The ribosome binding site (RBS) is in blue. c, Schematic map of the pAI3 and pAI5 dicistronic plastid transformation vectors yielding transplastomic Dicis^{Zm} and Dicis^{GG} plants.
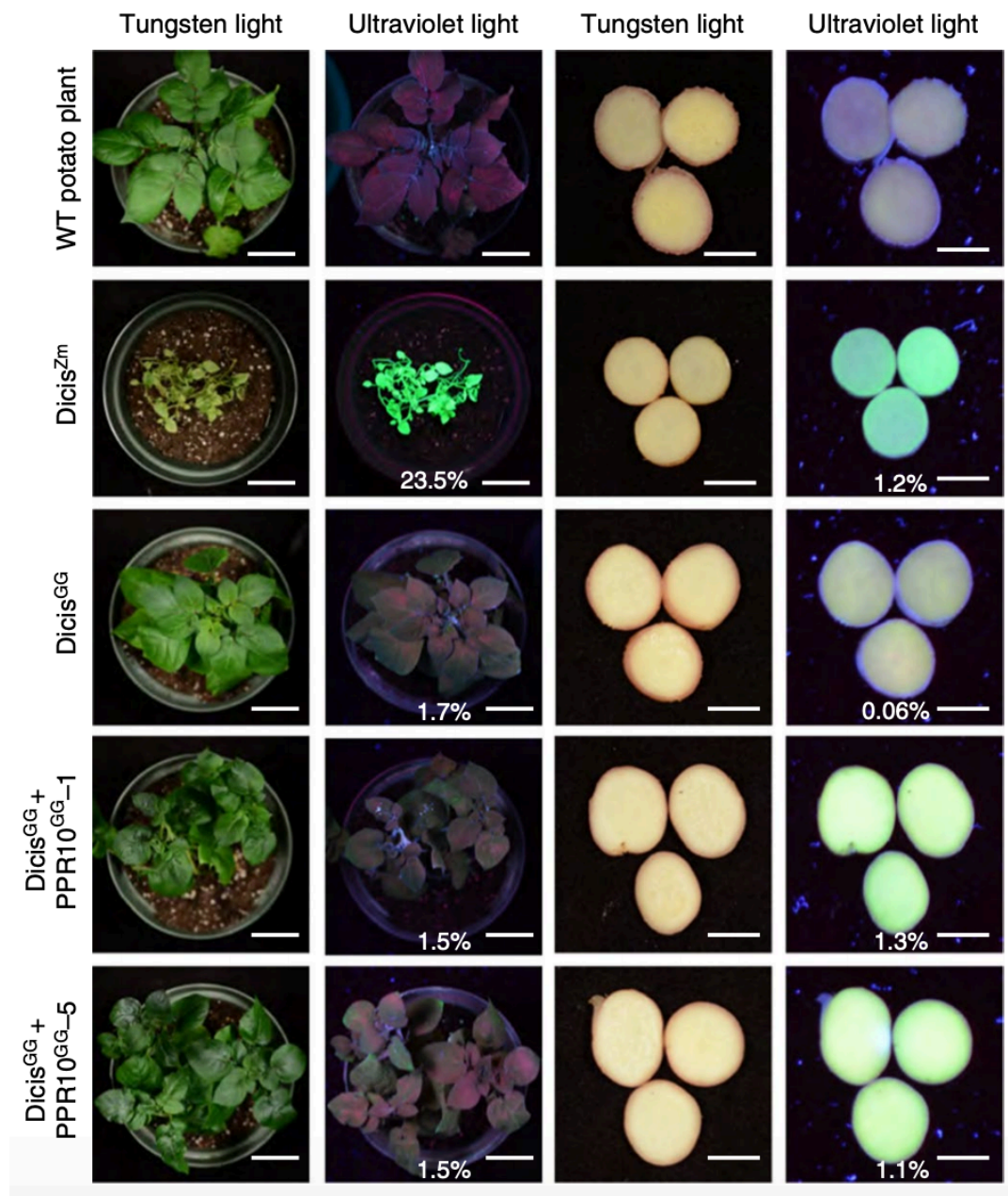
Figure 4-2. Leaf and tuber phenotype of potato plants illuminated with tungsten or UV light (366 nm) 4 months after planting *in vitro* shoots in soil. WT, wild type potato; Dicis^{Zm} and Dicis^{GG} transplastomic potato; Dicis^{GG}+PPR10^{GG} (transplastomic plants harboring the GG binding site and *patatin* promoter

PPR10^{GG}). -1 and -5 refer to two independent nuclear transgenic lines. Plant scale bar = 5 cm; tuber scale bar = 1 cm. Testing of phenotypes was carried out independently three times with similar results.
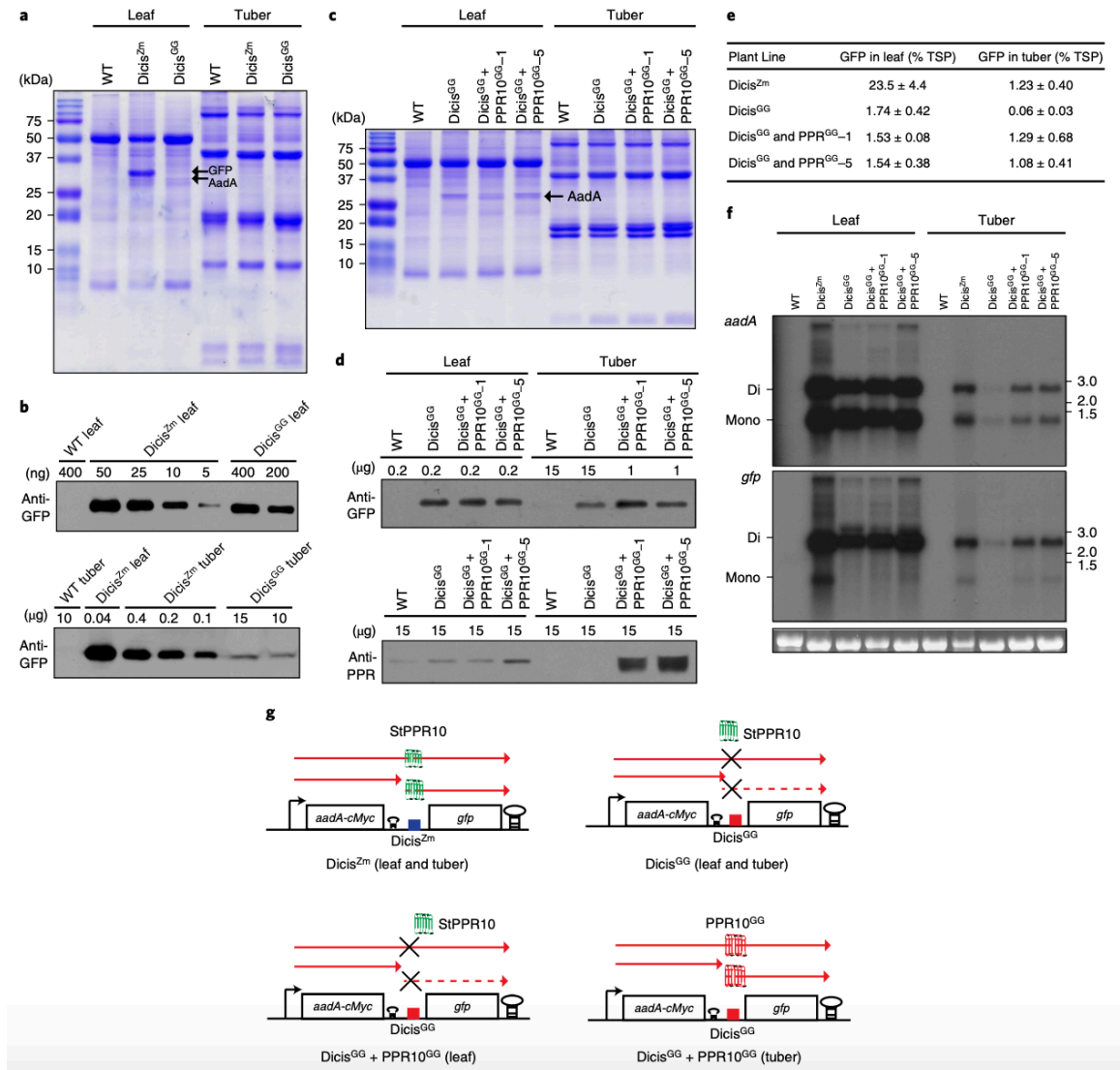
Figure 4-3. GFP expression in transplastomic potato and plants expressing the PPR10[GG] protein. a,c, Proteins separated in SDS-PAGE and stained with Coomassie brilliant blue. b,d, Western blot analyses probing with GFP and PPR10 antibody. Note signal in the wild-type and transplastomic Dicis[GG] lanes indicating that the maize PPR10 Ab weakly recognizes the native potato PPR10 protein. e, GFP levels based on three biological replicates (mean±SD., n=3). GFP concentration in the

highest expressing line Dicis^{Zm} was determined using 1D-Multi Lane Densitometry (Alphaimager 2200). Fold differences were quantified on Western blots with the ImageJ (version 2.0.0-rc-38/1.50b) software using Dicis^{Zm} leaf extracts as reference. f, RNA gel blot probed with the *aadA* and *gfp* probes. g, Schematic interpretation of mRNAs in transplastomic in Dicis^{Zm} and Dicis^{GG} plants. X indicates the absence of interaction between the PPR protein and binding site. Note that GFP in leaves accumulates at low level (1.5%) due to inefficiently translation from the Dicis^{GG} mRNA. Experiments shown in Fig. 3 a, b, c, d and f were repeated independently at least three times with similar results.
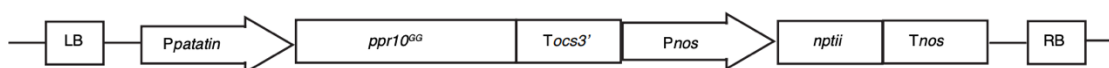
Figure 4-4. Schematic map of the T-DNA region of Agrobacterium binary vector pBIN19_B33_PPR10_mut10 plasmid, encoding PPR10^{GG}.
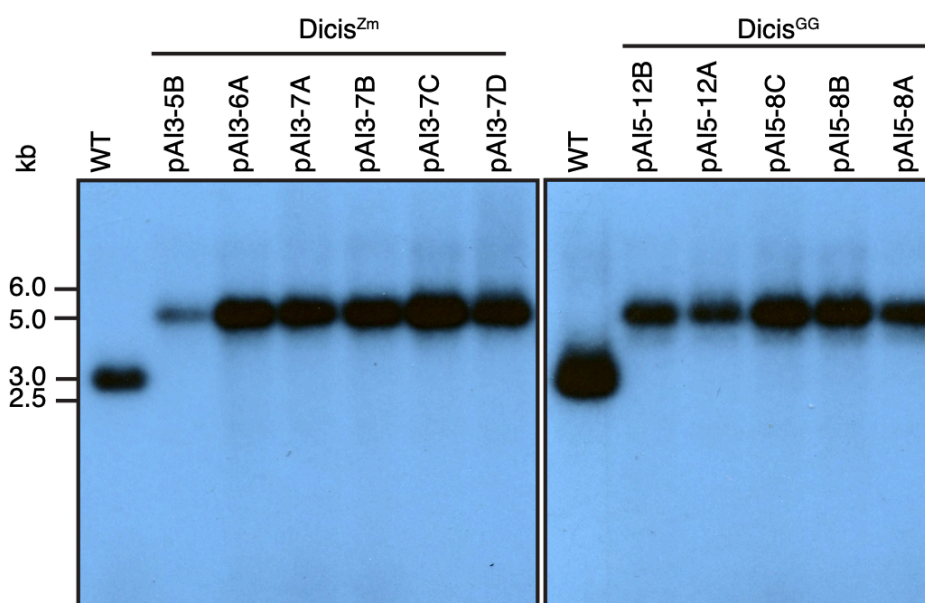
Figure 4-5. DNA gel blot analyses confirm uniform transformation of the plastid genome. Total cellular DNA was digested with the *Bam*HI restriction enzyme,and probed with an *Apa*I-*Bam*HI fragment containing the 16S rRNA coding sequence. Probe hybridizes with a 2.9 kb *Bam*HI fragment in the wild type and a 5.2 kb *Bam*HI fragment in transplastomic plants. Experiments were repeated independently two times with similar results.
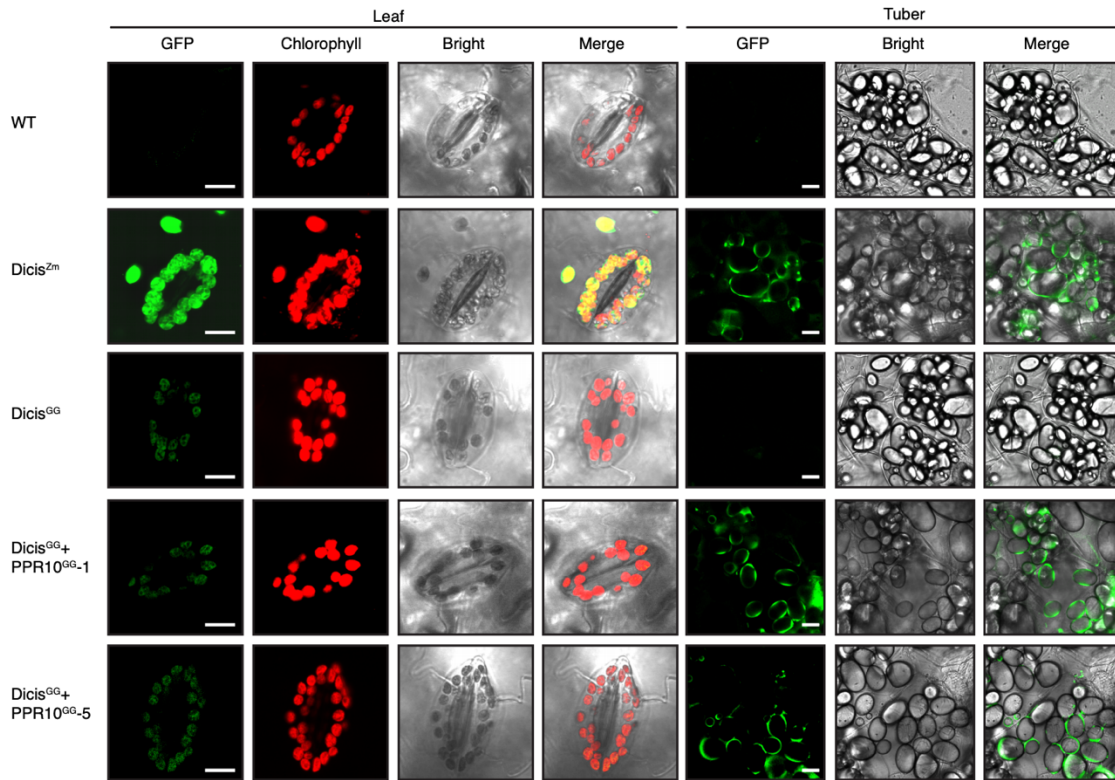
Figure 4-6. Confocal laser-scanning microscopy to detect GFP accumulation in potato leaf chloroplasts and tuber amyloplasts. Leaf bar = 10 μm; tuber bar = 20 μm. Experiments were repeated independently two times with similar results.

References

1   Zhang, J. *et al.* Identification of cis-elements conferring high levels of gene expression in non-green plastids. *The Plant Journal* 72, 115-128, doi:10.1111/j.1365-313X.2012.05065.x (2012).

2   Valkov, V. T. *et al.* High efficiency plastid transformation in potato and regulation of transgene expression in leaves and tubers by alternative 5' and 3' regulatory sequences. *Transgenic Res.* 20, 137-151, doi:10.1007/s11248-010-9402-9 (2010).

3   Sidorov, V. A. *et al.* Stable chloroplast transformation in potato: use of green fluorescent protein as a plastid marker. *Plant J.* 19, 209-216 (1999).

4   Martin, W. *et al.* 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 (2002).

5   Kahlau, S. & Bock, R. Plastid transcriptomics and translatomics of tomato fruit development and chloroplast-to-chromoplast differentiation: chromoplast gene expression largely serves the production of a single protein. *Plant Cell* 20, 856-874, doi:10.1105/tpc.107.055202 (2008).

6   Valkov, V. T. *et al.* Genome-wide analysis of plastid gene expression in potato leaf chloroplasts and tuber amyloplasts: transcriptional and posttranscriptional control. *Plant Physiol.* 150, 2030-2044, doi:10.1104/pp.109.140483 (2009).

7   Jin, S. & Daniell, H. The Engineered Chloroplast Genome Just Got Smarter. *Trends Plant Sci* 20, 622-640, doi:10.1016/j.tplants.2015.07.004 (2015).

8   Zhang, J. *et al.* Pest control. Full crop protection from an insect pest by expression of long double-stranded RNAs in plastids. *Science* 347, 991-994, doi:10.1126/science.1261680 (2015).

9   Boehm, C. R. & Bock, R. Recent advances and current challenges in synthetic biology of the plastid genetic system and metabolism. *Plant Physiol.*, doi:10.1104/pp.18.00767 (2018).

10   Maliga, P. & Bock, R. Plastid biotechnology: food, fuel and medicine for the 21st century. *Plant Physiol.* 155, 1501-1510, doi:pp.110.170969 [pii] 10.1104/pp.110.170969 (2011).

11   Hanson, M. R., Lin, M. T., Carmo-Silva, A. E. & Parry, M. A. Towards engineering carboxysomes into C3 plants. *Plant J.* 87, 38-50, doi:10.1111/tpj.13139 (2016).

12   Rae, B. D. *et al.* Progress and challenges of engineering a biophysical CO2-concentrating mechanism into higher plants. *J Exp Bot* 68, 3717-3737, doi:10.1093/jxb/erx133 (2017).

13   Long, B. M. *et al.* Carboxysome encapsulation of the CO2-fixing enzyme Rubisco in tobacco chloroplasts. *Nat Commun* 9, doi:ARTN 3570 10.1038/s41467-018-06044-0 (2018).

14   Jarvis, P. & Lopez-Juez, E. Biogenesis and homeostasis of chloroplasts and other plastids. *Nat Rev Mol Cell Bio* 14, 787-802, doi:10.1038/nrm3702 (2013).

15   Pfalz, J., Bayraktar, O. A., Prikryl, J. & Barkan, A. Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts. *EMBO J.* 28, 2042-2052, doi:10.1038/emboj.2009.121 (2009).

16   Prikryl, J., Rojas, M., Schuster, G. & Barkan, A. Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proc. Natl. Acad. Sci. USA* 108, 415-420, doi:10.1073/pnas.1012076108 (2011).

17   Barkan, A. *et al.* A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet* 8, e1002910, doi:10.1371/journal.pgen.1002910 (2012).

18   Diretto, G. *et al.* Metabolic Engineering of Potato Carotenoid Content through Tuber-Specific Overexpression of a Bacterial Mini-Pathway. *Plos One* 2, doi:ARTN e350 10.1371/journal.pone.0000350 (2007).

19   Yu, Q., Lutz, K. A. & Maliga, P. Efficient plastid transformation in Arabidopsis. *Plant Physiol.* 175, 186-193 (2017).

20   Rojas, M., Yu, Q., Williams-Carrier, R., Maliga, P. & Barkan, A. Engineered PPR proteins as inducible switches to activate the expression of chloroplast transgenes. *Nat Plants* in press (2019).

21   Caroca, R., Howell, K. A., Hasse, C., Ruf, S. & Bock, R. Design of chimeric expression elements that confer high-level gene activity in chromoplasts. *Plant J.* 73, 368-379, doi:10.1111/tpj.12031 (2013).

22   Slattery, C. J., Kavakli, I. H. & Okita, T. W. Engineering starch for increased quantity and quality. *Trends Plant Sci* 5, 291-298, doi:Doi 10.1016/S1360-1385(00)01657-5 (2000).

23   Liu, Q. *et al.* Genetic enhancement of oil content in potato tuber (*Solanum tuberosum* L.) through an integrated metabolic engineering strategy. *Plant Biotechnol J*, doi:10.1111/pbi.12590 (2016).

24   Miranda, R. G., Rojas, M., Montgomery, M. P., Gribbin, K. P. & Barkan, A. RNA-binding specificity landscape of the pentatricopeptide repeat protein PPR10. *RNA* 23, 586-599, doi:10.1261/rna.059568.116 (2017).

25   Kuroda, H. & Maliga, P. Sequences downstream of the translation initiation codon are important determinants of translation efficiency in chloroplasts. *Plant Physiol.* 125, 430-436 (2001).

26   Shinozaki, K. & Sugiura, M. Sequence of the intercistronic region between the ribulose-1, 5-bisphosphate carboxylase/oxygenase large subunit and coupling factor beta subunit gene. *Nucleic Acids Res.* 10, 4923-4934 (1982).

27   Bevan, M. Binary Agrobacterium vectors for plant transformation. *Nucleic Acids Res.* 12, 8711-8721 (1984).

28   Koncz, C. *et al.* in *Plant Molecular Biology Manual* (eds S. B. Gelvin & R. A. Schilperoort) Ch. B2, 1-22 (Kluver Academic, 1994).

29   Svab, Z. & Maliga, P. High-frequency plastid transformation in tobacco by selection for a chimeric *aadA* gene. *Proc. Natl. Acad. Sci. USA* 90, 913-917 (1993).

30   Carrer, H., Staub, J. M. & Maliga, P. Gentamycin resistance in *Nicotiana* conferred by AAC(3)-I, a narrow substrate specificity acetyl transferase. *Plant Mol. Biol.* 17, 301-303 (1991).

31   Zommick, D. H., Kumar, G. N., Knowles, L. O. & Knowles, N. R. Translucent tissue defect in potato (*Solanum tuberosum* L.) tubers is associated with oxidative stress accompanying an accelerated aging phenotype. *Planta* 238, 1125-1145, doi:10.1007/s00425-013-1951-8 (2013).

CHAPTER 5

Dicistronic operons as a novel marker system for chloroplast engineering and building blocks for plant synthetic biology

Abstract

We designed a dicistronic plastid marker system that relies on the plastid's ability to translate polycistronic mRNAs. The identification of transplastomic clones is based on selection for antibiotic resistance encoded in the first open reading frame (ORF) and accumulation of the reporter gene product in chloroplasts encoded in the second ORF. The antibiotic resistance gene may encode spectinomycin or kanamycin resistance based on the expression of *aadA* or *neo* genes, respectively. The reporter gene used in the study is the green fluorescent protein (GFP). Antibiotic resistance may be conferred by the expression of the prokaryotic-type marker in the plastid or be due to readthrough transcription in the nucleus from an upstream promoter, or a spontaneous mutation in the plastid 16S rRNA. However, if the dicistronic marker is integrated in the nucleus, GFP encoded in the second ORF will not be translated on the 80S cytoplasmic ribosomes. Indeed, when GFP accumulated in antibiotic resistant calli, plastid transformation has been confirmed in each case by DNA gel blot analyses and confocal microscopy. Therefore, high-level GFP accumulation from the second ORF in chloroplasts facilitates early identification of transplastomic events in callus or leaves by GFP fluorescence detected under UV light and by confocal microscopy. Because the dicistronic

mRNA is not processed, we could show that protein output from the second ORF is independent from the first ORF. Expression of multiple proteins from an unprocessed mRNA is an experimental design that enables predictable protein output from polycistronic mRNAs, expanding the toolkit of plant synthetic biology.

Introduction

The plastid genome encodes about 100 genes the products of which, in coordination with the about 3,000 nuclear gene products, carry out a wide range of vital metabolic processes including photosynthesis, fatty acid and amino acid biosynthesis [1]. As the consequence of endosymbiotic ancestry, plastids maintained prokaryotic features of the gene expression machinery [2,3]. The chloroplast's unique ability to selectively produce large quantities of proteins required for photosynthesis, such as RUBISCO, makes them useful hosts for the production of recombinant proteins [4-8]. Organization of plastid genes in operons makes the plastid genome an ideal target for the incorporation of operons encoding novel functions, such as the biosynthesis of astaxanthin, the red pigment that gives salmon its pink color [9]. Improving the efficiency of sunlight conversion into biomass is another process that depends on the availability of chloroplast engineering [10] [11,12] [13]. Expression of dsRNA in chloroplasts has also been successfully deployed against insects [14,15]. In addition, any trait imparted by chloroplast genome engineering will be exempt from dispersal by pollen because the plastid genome is inherited by the maternal parent only in most crops [16-18]

The plastid genome is highly polyploid and may be present in up to 10,000 genome copies per cell. Because initially only a few genome copies are transformed, sorting is protracted and may take several generations to obtain a uniform population of transformed plastid genomes in a cell [19,20]. Plastid sorting is accelerated by selection for an antibiotic resistance gene encoded in a vector. The most commonly used antibiotic is spectinomycin, an inhibitor of translation on the prokaryotic-type plastid ribosomes, that is inactivated by aminoglycoside-3''-adenylyltransferase (AAD or AadA), the product of the *aadA* gene [21]. Other inhibitors of translation on plastid ribosomes have also been successfully applied as markers for the selective enrichment of transformed plastid genomes, including kanamycin [22,23], chloramphenicol [24] and tobramycin [25]. Marker genes in the transformation vectors are expressed in promoter/terminator cassettes, which are inserted in the plastid genome by homologous recombination. Insertion, however, may also occur in the nuclear genome where antibiotic resistance may be expressed by readthrough transcription from an upstream promoter. A particularly high frequency of nuclear insertions was observed when kanamycin selection yielded only three transplastomic events in a pool of 99 kanamycin resistant clones [22]. No spontaneous kanamycin resistant mutants were recovered, because kanamycin resistant mutations in the 16S rRNA are probably not viable [26]. Therefore, we decided to develop a marker system that enables rapid identification of transplastomic events, without the need for PCR or DNA gel blot analyses to distinguish nuclear insertions from transplastomic events.

The dicistronic marker system consists of two ORFs: the first ORF encodes a selectable antibiotic resistance gene and the second ORF a visual marker. Rapid identification of transplastomic events exploits the plastid's ability to translate polycistronic mRNAs, a feature retained in plastids from the ancestral photosynthetic prokaryotes. This contrasts translation on 80S cytoplasmic ribosomes, which are based on the scanning mechanism of translation, that would be applicable to plastid gene constructs incorporated in the nucleus. The plastid antibiotic resistance gene encoded in the first ORF will be translated on the 80S ribosomes, but not the second ORF, enabling rapid identification of transplastomic events.

We report here dicistronic operons that are useful for the facile identification of transplastomic clones. High-level GFP accumulation from the second ORF is a reliable indicator of transplastomic clones, that can be readily identified by illumination of leaves or calli with a hand-held UV light. In our system only dicistronic mRNAs are detectable on RNA blots. That is why we can address translational coupling, the impact of protein accumulation from the first ORF on protein accumulation from the second ORF. We have found that, in the absence of processed mRNA, GFP accumulation from the second ORF is independent from protein accumulation from the first ORF, an experimental design that enables predictable protein output from polycistronic mRNAs.

Results

Design of synthetic dicistronic operons

Each dicistronic operon was expressed in a cassette consisting of the tobacco plastid

ribosomal RNA PEP promoter (P*rrn*) and the *psbA* plastid gene 3'-UTR (T*psbA*).

Prrn was fused with segments of plastid gene 5'-UTRs, referred to as leader

sequences (L) and the source gene name: the plastid *clpP*, *rbcL* and *atpB* genes and

the T7 phage gene 10 leader (L*T7g10*). The first ORF encodes a kanamycin

resistance (*neo*) or spectinomycin resistance (*aadA*) gene and the second ORF *gfp*,

with a *cry9Aa2* leader that encodes a ribosome binding site. The *cry9Aa2* leader was

chosen because is not processed when it is in an intergenic region [27] and is known

to yield high-level protein expression from the downstream ORF [28]. Among the *neo*

or *aadA* construct families the only variable is the 5'-UTR (Fig.5-1b).

To compare protein yields from monocistronic and dicistronic constructs, we

also included a monocistronic *gfp* with the *cry9Aa2* gene in the experiment (Fig. 5-

1c). Another control is a dicistronic *gfp* gene with the maize PPR10 binding site

between *aadA* and *gfp* that we used in the study of tissue-specific plastid gene

expression in potato [29] (PAI4, Fig. 5-1b).

Plastid transformation with the dicistronic vectors

Plastid transformation was carried in tobacco leaves with vectors pMRR13,

pMRR15 and pMRR18 which have spectinomycin resistance (*aadA*) as the selective

marker. Transplastomic clones were selected on the RMOP shoot regeneration

medium containing 500 mg/L spectinomycin. Bombarded leaf sections with

regenerating shoots are shown in Figure 2a. Chloroplast localization of GFP in the fluorescent sectors was confirmed by laser scanning confocal microscopy (Fig. 5-2b). GFP fluorescence as a reliable indicator of transplastomic state has been documented in detail elsewhere [30].

Plastid transformation efficiency was also tested using the pMRR20 and pMRR21 vector DNA and selection for kanamycin resistance on RMOP medium containing 50 mg/L kanamycin. Kanamycin selection after bombardment with vector pMRR20 yielded 61 T0 kanamycin resistant clones, of which 45 accumulated GFP under UV light. Out of the original 61 T0 clones, 52 maintained kanamycin resistance when transferred to a medium containing 100 mg/L kanamycin sulfate. 46 of these accumulated GFP at a level that it could detected under a hand-held UV light (Table 5-1). DNA gel blot analysis was carried out on 13 GFP fluorescent (Fig. 5-2c) and 6 non-fluorescent, kanamycin resistant clones (Fig. 5-2d). Probing of total cellular DNA from the fluorescent clones yielded a 5.7-kb transplastomic band, or a mix of 5.7-kb and 3.2-kb wild type band, confirming correlation of GFP fluorescence and transplastomic status. The non-fluorescent clones yielded only wild-type 3.2-kb fragments, and are probably nuclear gene transformants which carry wild-type ptDNA [22]. Analyses of kanamycin resistant clones obtained after bombardment with then pMRR21 vector yielded similar results: 15 kanamycin resistant clones, of which 13 showed GFP fluorescence in the original selection plate. 9 of the clones maintained kanamycin resistance on 100 mg/L kanamycin sulfate and the cells fluoresced under UV light (Table 5-1).

Protein accumulation in the leaves after transformation with the dicistronic operons

Early identification of transplastomic events is based on high-level GFP accumulation from the second ORF. The calli on the selective medium regenerate shoots, therefore measuring GFP accumulation in leaves is a good predictor whether or not the second ORF will express the protein at sufficiently high levels to allow identification of transplastomic clones under UV light. An addition question is whether or not protein accumulation from the 1st ORF will impact protein accumulation from the second ORF.

Many of our plastid transformation vectors have a spectinomycin resistance (*aadA*) marker gene. Therefore, we also tested GFP accumulation from the second ORF with *aadA* expressed from different leader sequences. The Prrn promoter was fused with 57 and 18 nucleotides of the *rbcL* plastid gene 5'-UTR (P*rrn*L57, PrrnL18) or the *atpB* gene 5'-UTR (P*rrn*L*atpB*) (Fig. 5-1). Vectors pMRR13, pMRR18 and pMRR15 yielded 6.6%, 2.9% and 0.5% AAD, while GFP from the second ORF yielded 42.2%, 12.7% and 29.4% GFP (Fig.5-4a, Table 5-2). The mRNA levels are the highest from the pMRR13 construct, with high-level GFP accumulation from the 2nd ORF, that is why we used this construct for the early identification of transplastomic clones in *Arabidopsis* [27,30].

P*rrn*L*clpP* is known to yield relatively low levels of NPTII (0.26%; pHK33 vector)[31], while PrrnLT7g10 yields high levels of NPTII (23%, pHK40) [32].

Kanamycin resistance encoded in the first ORF in vectors pMRR20 and pMRR21 are expressed from the L*clpP* and L*T7g10* leaders with predicted low- and high-levels of NPTII accumulation levels. Indeed, the transplastomic pMRR20 and pMRR21 leaves accumulated relatively low (1% TSP) and high-levels (9.1% TSP) of NPTII, as predicted, with comparably high levels of GFP accumulation from the second ORF, 24% and 31% TSP, respectively. This finding indicates that protein accumulation from the second ORF is independent from the 1st ORF (Table 5-2). The pMRR21 mRNA was 1.8-times more abundant then the pMMR20 mRNA indicating that the absolute levels of protein accumulation are dependent on mRNA abundance, as modified by mRNA stability.

GFP in the leaves of the plants in this study was >10% of TSP (Table 5-2). GFP was readily detectable in seedling cotyledons, or in the leaves of greenhouse plants upon illumination with UV light (Fig. 5-3). Plants transformed with vectors pMRR13 and pMRR17 are significantly smaller than then wild type. Both of these plants accumulate GFP in excess of 35% TSP (Fig. 5-4a, Table 5-2), what is the likely reason for the significantly reduced growth. Transgene expression in the young leaves of pMRR13, pMRR17, pMRR20 and pMRR21 plants cause a pigment deficient phenotypes (Fig. 5-3).

Expression of dicistronic operons in roots

We tested GFP accumulation in the roots of pMRR13 plants, which have the highest level GFP accumulation (42% TSP) in leaves. GFP in the roots accumulated to very

low levels, 0.14% TSP (Fig. 5-5). This level of GFP accumulation is similar to root expression of the pAI4 construct 0.13% TSP (Fig. 5-5), that accumulated 1.2% GFP in potato tubers [29].

Discussion

Dicistronic operons as plastid markers

GFP is the most commonly used reporter gene today, replacing GUS as the reporter gene used in the early studies on plastid gene expression [33,34]. The utility of GFP as a marker was first shown by transient expression of GFP in chloroplasts [35]. Stable integration of GFP in chloroplasts enabled visualized of stromules in transplastomic tobacco plants [30,36]. GFP was also deployed as a marker gene in *Chlamydomonas* chloroplasts where optimization of codon usage was necessary to obtain chloroplast expression [37]. A translational fusion of AAD, the *aadA* gene product with GFP yielded a fluorescent antibiotic resistance marker, that could be used to track transplastomic chloroplasts [38]. However, fluorescence of the fusion protein was dampened as compared to the non-fused GFP. That is why we started experimenting with dicistronic constructs that enabled GFP expression using the minimal number of promoters and terminators. Dicistronic GFP constructs were found useful to test regulated chloroplast gene expression [29,39]. Because polycistronic mRNAs are not expressed in the nucleus [40], we hypothesized that selection for a marker gene encoded in the first ORF, and GFP accumulation from the second ORF should indicate plastid transformation. This hypothesis was confirmed by the finding that

bombardment followed with spectinomycin selection yielded GFP expressing clones which were confirmed to be transplastomic by DNA gel blot analyses [30]. Here we report a similar finding, plastid transformation confirmed by DNA gel blot analyses, when the clones are selected for kanamycin resistance and are expressing GFP.

Transformation of the tobacco plastid genome with the *neo* (*AphA-2*) gene yielded only three transplastomic events in the pool of 99 kanamycin resistant clones [22]. The statistics in the present experiment are much more favorable. Bombardment with pMRR20 vector yielded mainly transplastomic events (45/61) and bombardment with the pMRR21 vector in most cases (13/15) yielded transplastomic clones. We attribute the reduced number of nuclear clones to inefficient translation of the plastid *atpB* and phage *T7G10* leader sequences when the plastid construct is inserted in the nucleus and transcribed from an upstream promoter. The potential problem of plastid gene expression in the nucleus is well documented. Frequent recovery of nuclear kanamycin resistant clones was reported after transformation with a kanamycin resistance gene encoded in a *psbA* plastid gene cassette [41]. Activation of a plastid *aadA* gene has also been characterized after transfer from the plastid genome to the nucleus [42]. No nuclear kanamycin resistant clones were obtained after transformation of chloroplasts with the *AphA-6* gene [23]. The absence of nuclear transformants is likely due to the choice of plastid expression signals, that were not translated in the cytoplasm when the *AphA-6* plastid gene was

inserted in the nucleus. The L*atpB* and L*T7g10* sequences tested in this study are equally useful.

Translational coupling in the chloroplast

Translational coupling is defined as the interdependence of translation efficiency of neighboring genes encoded by the same polycistronic mRNA. In bacteria the translation of overlapping ORFs is typically coupled, that is translation of the second ORF is dependent on the translation of the first ORF [43]. When two non-overlapping ORFs are encoded by the same mRNA, translational coupling depends on the distance between the stop codon and the downstream start codon. Translational coupling with enhancement level was found to vary up to 10-fold, dependent on the *cis*-elements of the upstream gene. The coupling effect in *E. coli* was attributed mainly to translational effect, and not to transcriptional modulation or mRNA stability [44].

Translation coupling has been reported in overlapping ORFs in tobacco chloroplasts. The *psbD-psbC* and *ndhC-ndhK* reading frames are encoded polycistronic mRNAs where the AUG translation initiation codon of the downstream ORF is upstream of stop codon of the upstream reading frame, and translation of the downstream ORF is dependent on the translation of the upstream ORF [45-48]. Translational coupling was also observed in an overlapping artificial dicistronic operon, where relatively high level (0.5%) GUS accumulation resulted from the translation of *gusA* mRNA in the *rbcL-gusA* operon [49]. The *atpB-atpE*

ORFs also overlap in a dicistronic operon, but in this case the downstream *atpE* cistron has its own cis-elements for translation, and is not dependent for translation of the upstream ORF [50,51].

To ensure that we obtain high level GFP accumulation from the second ORF, we addressed the impact of protein accumulation from the first ORF, that is the problem of translational coupling in chloroplasts. The ORFs of the dicistronic operons studied here are non-overlapping and have their own translation control sequences. All transgenes were expressed in the same promoter (P*rrn*)-terminator (T*psbA*) cassette, therefore we assume that the rate of transcription of each dicistronic operon is the same, and differences in mRNA accumulation are the consequence of differential mRNA stability mediated by the 5'-UTR (Fig. 5-1, Table 5-2). The simple assumption is that more mRNA of the same ORF should yield more protein. The level of mRNA accumulation among the *aadA* constructs (pMRR13, pMRR15, pMRR18) varied 10-fold, L*atpB* yielding the most stable and L*rbcL*57 the least stable mRNA (Fig. 5-4c, Table 5-2). GFP accumulation from the 2nd ORF in pMRR18 vs pMRR15 plants was proportional with the mRNA, about 3-times more mRNA yields 3-times more GFP, 12% as compared to 29% GFP plants. However, an additional 3-times more mRNA in the pMRR13 plants yielded only 42% GFP, indicating that further increase in mRNA levels yielded a diminishing return, probably due to exhausting the cells translation capacity.

We could also assess the impact of translation efficiency of the first ORF on the translation efficiency of the second ORF. NPTII protein output of the first ORF

in pMRR20 and pMRR21 plants differs 9x. When calculating translation efficiency (protein/unit mRNA), the difference is reduced 5-fold in favor pMRR21 plants. However, the pMRR20 plants produce as much or slightly more GFP from the second ORF than the pMRR21 plants per unit mRNA (24% as compared to 18%), indicating that there is no positive translational coupling between the first and second ORFs in the kanamycin resistant plants (Fig. 5-6). The lack of translational coupling between the first and second ORF is also confirmed, when comparing translation efficiency between the pMRR15 and pMRR18 spectinomycin resistant plants. Translation efficiency of the first ORF (*aadA*) differs 16-fold, while the second ORF accumulates similar amounts GFP, 10% and 13%, respectively. There is negative correlation between the translation efficiency of the first and second ORFs in case of the pMRR15 and pMRR13 plants: about 3-fold increase for the first ORF and 2x reduction from the 2nd ORF. However, reduced translation efficiency from the second ORF is likely to be the reflection of the limitations of translation capacity.

Because our transplastomic chloroplasts accumulate only dicistronic mRNAs, we had an ideal opportunity to quantify relative protein output from specific ribosome binding sites. The utility of information is complicated by differential mRNA stability, that will have to be considered when the information is used to design synthetic operons. While we should be able to predict relative protein output, absolute protein levels will be set by mRNA stability.

Materials and Methods

Plant material and growth conditions

Nicotiana tabacum L. cv. Petit Havana plants were grown in the greenhouse at 25 °C with illuminated with supplemental lighting for 16 hrs.

Plastid transformation vectors

The dicistronic operons were expressed in a cassette consisting of the tobacco plastid ribosomal RNA PEP promoter (P*rrn*)[52] and the *psbA* plastid gene 3'-UTR (T*psbA*) [53]. P*rrn* was fused with segments of plastid gene 5'-UTRs, referred to as leader sequences (L) and the source gene name: the plastid *clpP* [54] , *rbcL* [55] and *atpB* [56] genes and the T7 phage gene 10 leader (L*T7g10*) [57]. L*rbcL57* and L*rbcL18* refer to the number of nucleotides included from the *rbcL* leader. The first ORF encoded a kanamycin resistance (*neo*)[22,58] or spectinomycin resistance (*aadA*)[59] gene and the second ORF a GFP [60], with a *cry9Aa2* leader that encodes a ribosome binding site [28]. Among the *neo* or *aadA* construct families the only variable was the 5'-UTR. The schematic map of the vectors is shown in Figure 1. The vector parts and GenBank accession numbers are listed in Table 3. We also included a dicistronic *gfp* gene with maize PPR10 binding site between *aadA* and *gfp* (Fig. 5-5) , similar to vector pAI3, other than lacking TrbcL between the *aadA* and *gfp*[29] and a monocistronic *gfp* gene with a L*cry9aA2* leader (pMRR17) (Table 5-3).

Transformation of the tobacco plastid genome

Tobacco plastid transformation was carried out as described [61,62]. Briefly, 0.6 μm gold particles were coated with vector DNA, and tobacco leaves harvested from sterile plants were bombarded with the particles using the Bio-Rad PDS-1000 biolistic gun with the Hepta Adaptor. Wild type tobacco leaves form pigment deficient callus on selective RMOP shoot regeneration medium containing 500 mg/L spectinomycin or 50 mg/L kanamycin. Green shoots and calli formed in the bombarded leaf cultures indicating the expression of the selective marker genes. Uniform transformation of the plastid genomes was confirmed by DNA gel blot analyses (Figure 5-8).

DNA gel blot analysis

DNA gel blot analysis was carried out as described [59]. Briefly, total cellular DNA was extracted from leaf tissue using the cetyltrimethylammonium bromide (CTAB) protocol. Two microgram total leaf DNA was digested with the *Bam*HI restriction endonuclease, the DNA fragments were separated on 1% agarose gels, and then transferred to Hybond-N membranes (GE Healtcare RPNNL/02/10) by capillary blotting. The probe was prepared by random-primed ^{32}P-labeling using kit (Takara Bio. Inc., Cat. No. 6045). The probe was the 1.9 kb *Apa*I/*Bam*HI ptDNA fragment encoding part of the 16S rRNA gene (Fig. 5-1).

RNA gel blot analysis

RNA gel blot analyses was carried out as described [63]. Briefly, total cellular RNA was isolated from leaves frozen in liquid nitrogen using TRIzol (Thermo Fischer Scientific, Cat. No. 15596026). The RNA was separated in 1.5% formaldehyde agarose gels, and transferred to Hybond-N membranes by capillary blotting (GE Healthcare RPNNL/02/10). The probes were was prepared by random-primed ^{32}P-labeling using the Takara Bio. Inc., kit (Cat. No. 6045). The probes were: *aadA*, 0.8-kb *Nco*I-*Xba*I fragment isolated from plasmid pHC1 [64]; *gfp*, fragment amplified from *gfp* coding region with primers *gfp*-forward 5'-TTTTCTGTCAGTGGAGAGGGTG-3' and *gfp*-reverse 5'-CCCAGCAGCTGTTACAAACT-3' and *neo* amplified with primers *neo*-forward 5'-GGGAAGGG ACTGGCTGCTAT-3' and *neo*-reverse 5'-CGATACCGTAAAGCACGAGGA-3'.

Immunoblot analysis

Immunoblot analyses was carried out following the protocol described for potato tubers [65]. Briefly, total soluble protein was extracted from leaf and root samples frozen in liquid nitrogen and pulverized using stainless steel grinding balls (SPEX Sample Prep, Metuchen, NJ). About 100 mg of powdered tissue was extracted in 200 µl of sodium phosphate buffer (100 mM, pH 6.8) containing 1 mM EDTA, 5 mM DTT and a protease inhibitor cocktail (30 µl) specific for plant tissue (Sigma-Aldrich, Cat. No. P9599). The suspension was centrifuged at 14,000g for 25 min at 4 °C, then the supernatant containing the protein extract was removed and the

protein concentration was determined with the BCA protein assay reagent kit. The proteins were separated in 15% urea PAGE and transferred to Immun-Blot PVDF membrane (Bio-Rad, Cat. No. 1620177). GFP detection: the primary antibody was Monoclonal Antibody (JL-8) (Takara, 632381); Secondary antibody Anti-Mouse IgG Peroxidase antibody (Sigma-Aldrich, A4416-1 mL). NPTII detection: primary antibody was Anti-NPTII (Sigma-Aldrich, N6537-200 µl); secondary antibody Anti-Rabbit IgG Peroxidase antibody (Sigma-Aldrich, A0545-1 mL). Detecting the AAD c-Myc tag: the primary antibody was c-Myc Antibody (9E10) (Santa Cruz Biotechnology, sc-40); the secondary antibody was Anti-Mouse IgG Peroxidase antibody (Sigma-Aldrich, A4416-1 mL). Each value is an average of three biological replicate measurements.
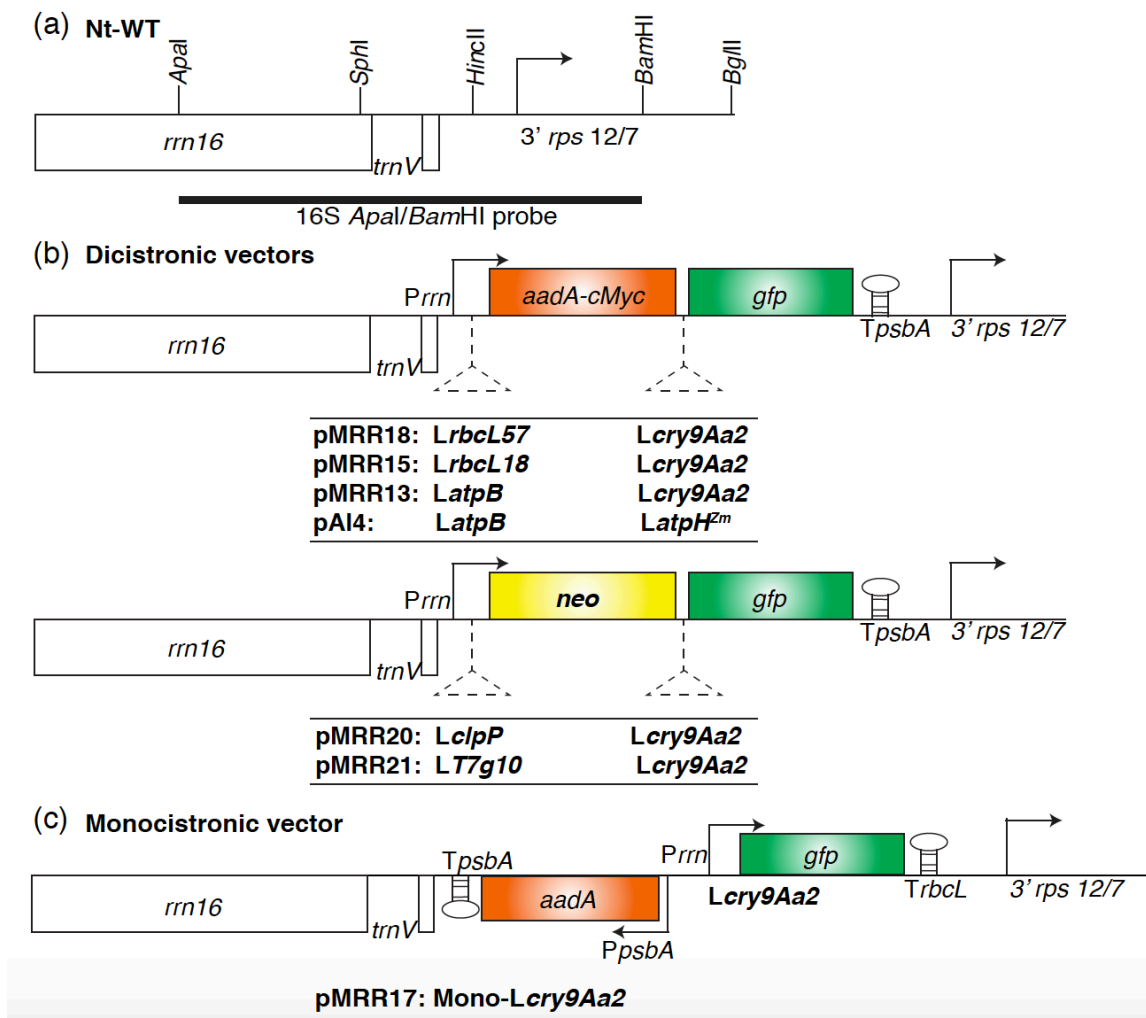
Figure 5-1. Plastid transformation vectors. (a) Map of the targeting region in the plastid genome. The transgenes are inserted into the *Hin*cII site. The position of *rrn16* and *trnV* genes and *3'rps12/7* promoter are shown [66]. (b) Dicistronic vectors with *aadA* or *neo* selectable marker gene encoded in the first ORF, and GFP in second ORF. Listed are: pMRR vectors with P*rrn* promoter and the source of 5'-UTR (L*rbcL*57, L*rbcL*18, L*atpB*, L*clpP*, L*T7g10*). Each vector has the *rry9Aa2* gene ribosome binding site (L*Cry9A2*) in the intergenic region of antibiotic resistance gene (*aadA*, spectinomycin resistance; *neo*, kanamycin resistance) and *gfp*. The operon has the *psbA* gene 3'-UTR (T*psbA*). (c) The pMRR17

monocistronic vector with *aadA* expressed in P*psbA*/T*psbA* cassette and *gfp* in P*rrn*L*cry9aA2*/T*rbcL* cassette.
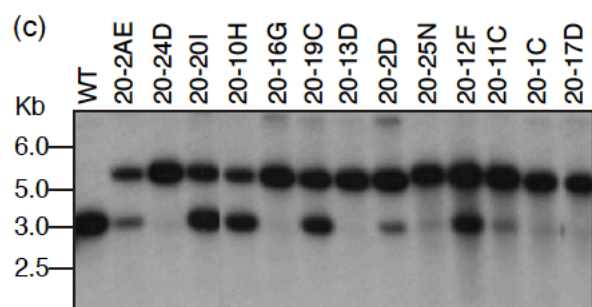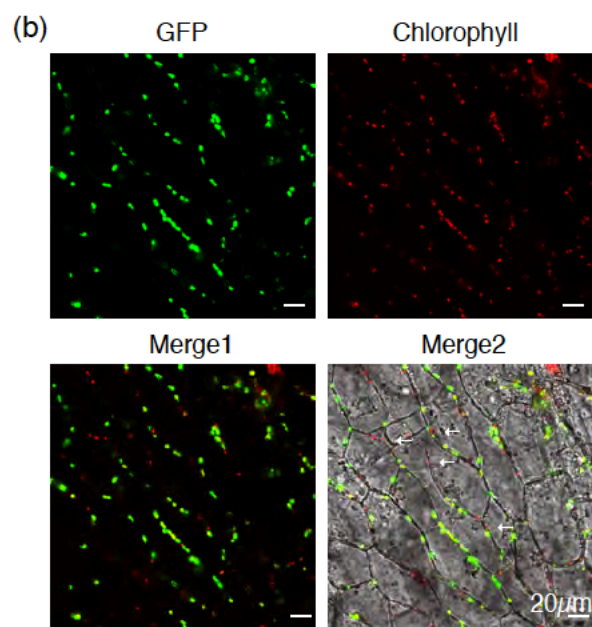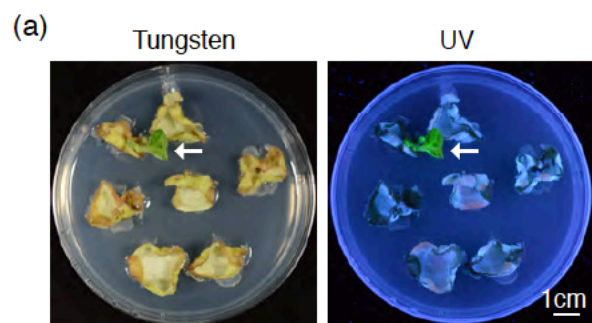
(a)

Tungsten   UV

1cm

(b)

GFP   Chlorophyll

Merge1   Merge2

20μm

(c)

Kb

6.0 —
5.0 —

3.0 —

2.5 —

WT | 20-2AE | 20-24D | 20-20I | 20-10H | 20-16G | 20-19C | 20-13D | 20-2D | 20-25N | 20-12F | 20-11C | 20-1C | 20-17D

(d)

Kb

6.0 —
5.0 —

3.0 —

2.5 —

WT | 20-7E | 20-14F | 20-18C | 20-3E | 20-8G | 20-7G | 20-20E | 20-14D
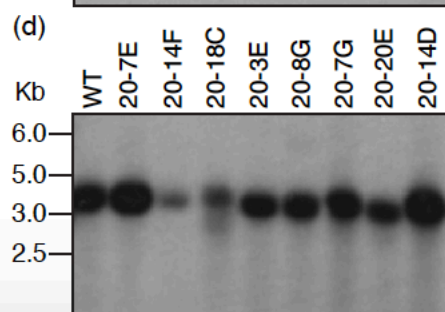
Figure 5-2. Identification of transplastomic clones using the dicistronic marker system. (a) Transplastomic shoot three weeks after the bombardment with vector pMRR13. Note GFP fluorescence under UV (366 nm) light. (b) Confocal laser-scanning microscopy to confirm GFP accumulation in chloroplast after bombardment with vector pMRR13. Shown are GFP channel (excitation 488 nm; detection 500-530 nm), chlorophyll chanel (excitation 568 nm; detection 650-700 nm), overlay of GFP and chlorophyll channels (Merge 1) and GFP/chlorophyll and bright field channels (Merge 2). The cells are heteroplastomic, with mixed transformed (GFP) and wild-type (red) plastids. (c) DNA gel blot hybridization confirms plastid transformation in GFP-expressing clones. Total cellular DNA was digested with the *Bam*HI restriction endonuclease and probed with the 1.9-kb *Apa*I-*Bam*HI probe (Figure 1a). Wild-type ptDNA yields a 2.9 kb fragment, transplastomic ptDNA a 5.7 kb fragment. Cells with mixed wild type and transformed ptDNA have both. (d) Kanamycin resistant clones which have no detectable GFP have 2.9 kb wild type ptDNA, that is indicative of nuclear insertion of the *neo* gene.
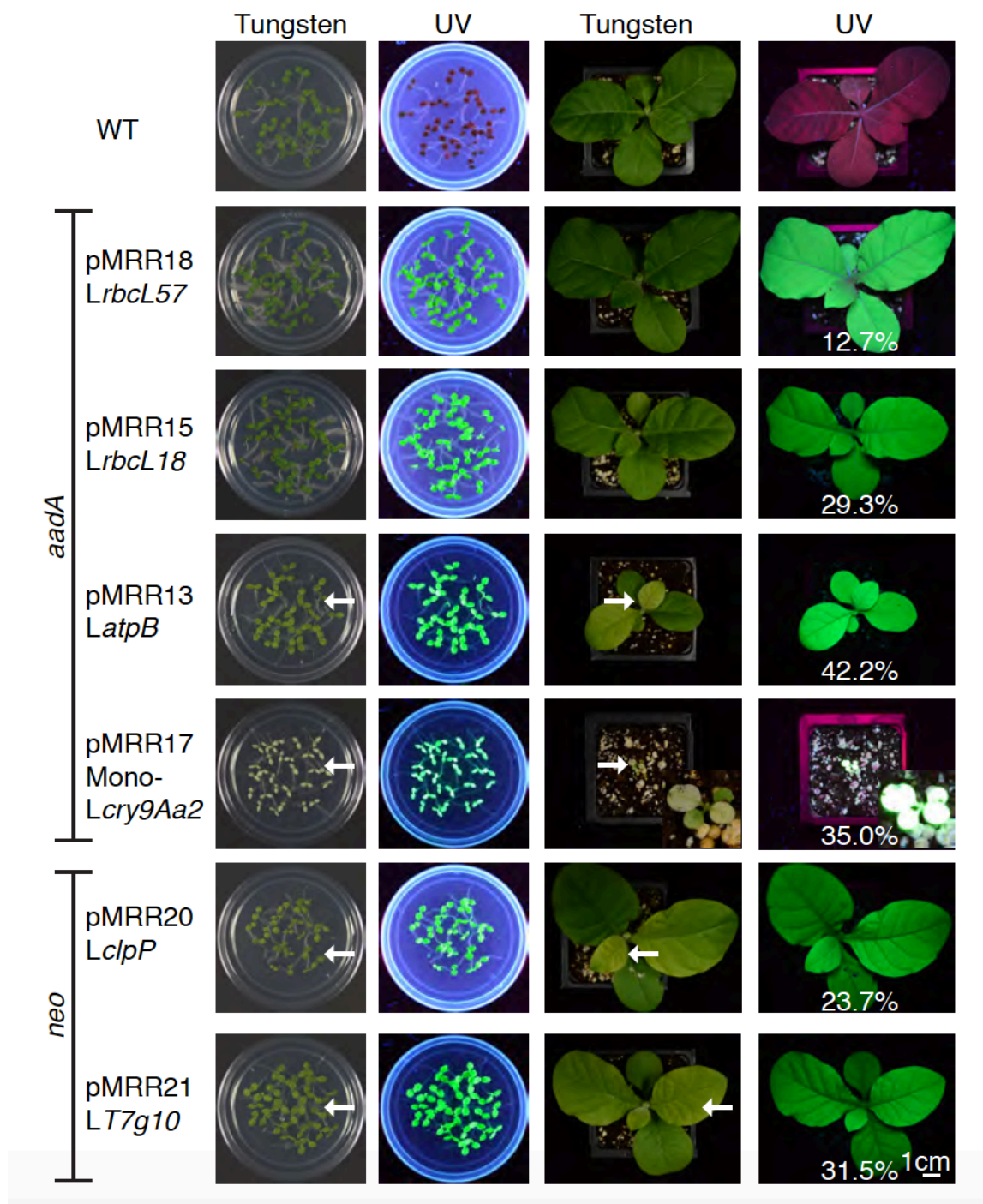
Figure 5-3. Phenotypes of transplastomic and wild type (WT) tobacco plants.

The lines are grouped by the selective marker encoded in the first ORF (*aadA*, *neo*)

and illuminated with tungsten or UV light to visualize GFP accumulation. One week

old transplastomic seedlings have been germinated in sterile culture on a selective medium (500 mg/L spectinomycin or 200 mg/L kanamycin). Greenhouse plants are four week old. Note stunted growth (pMRR13, pMRR17) and pigment deficiency (pMRR13, pMRR17, pMRR20, pMRR21; arrows).
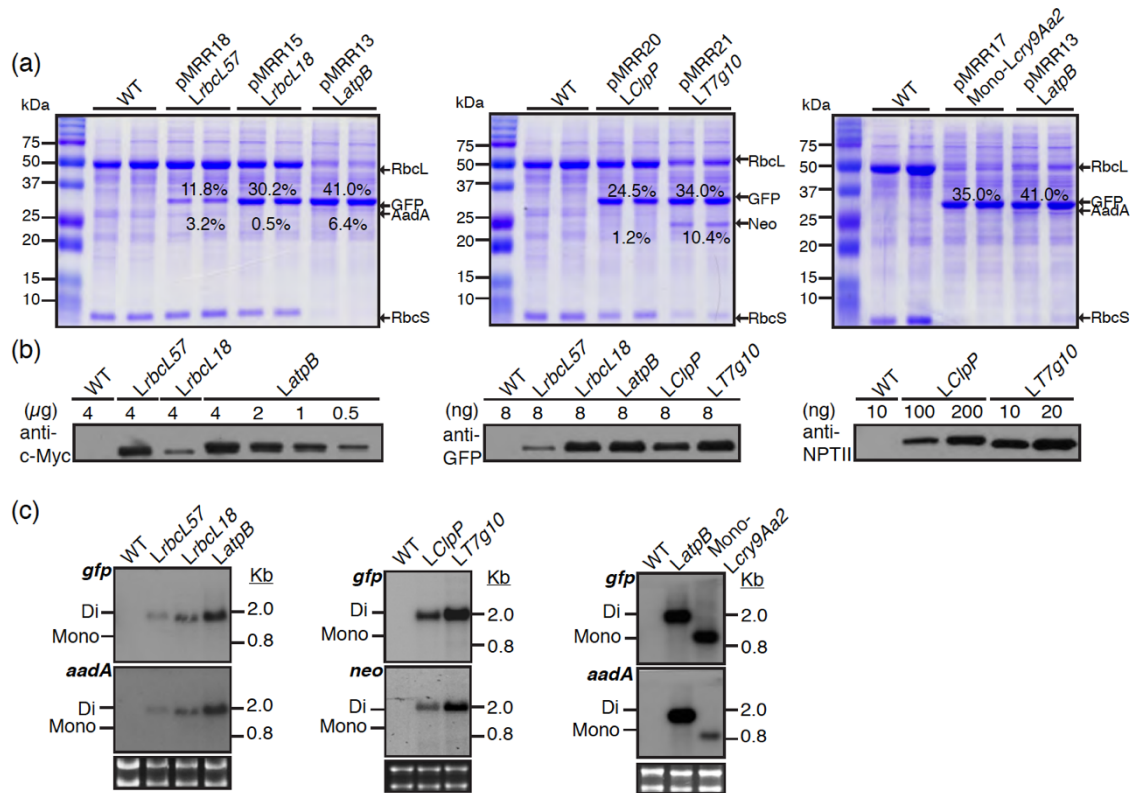
Figure 5-4. Expression of *gfp* and the selectable marker genes, *aadA* or *neo*, in transplastomic plants. (a) Protein (5 µg) separated in SDS-PAGE and stained with Coomassie brilliant blue. (b) Immunoblot analysis to quantify GFP, AadA and NPTII accumulation in transplastomic plants. (c) RNA gel blot hybridization to quantify mRNA abundance. The mRNA and protein levels are based on three biological replicates (mean±SD, n=3). Visible GFP, AAD and NPTII bands were quantified using 1D-Multi Lane Densitometry (Alphaimager 2200). Fold differences were quantified on Western blots with the ImageJ software (version 2.0.0-rc-38/1.50b) software. Northern blots were imaged with a Typhoon RGB scanner (GE Life Sciences) and quantified with the ImageQuant software.
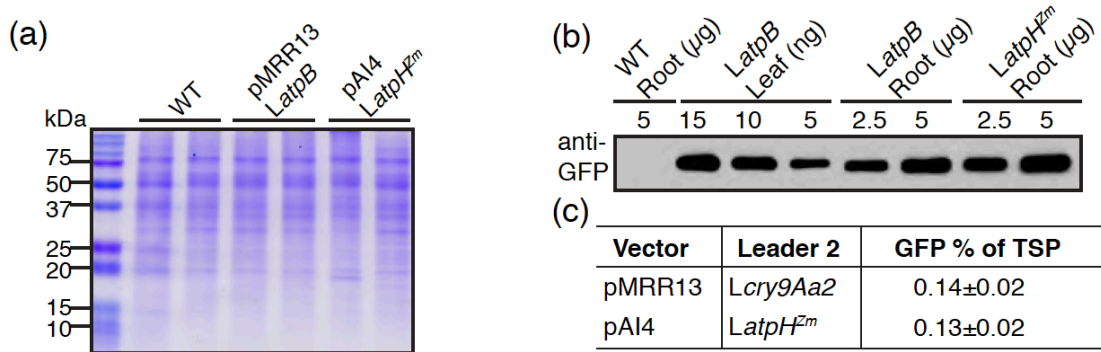
Figure 5-5. GFP accumulation from dicistronic operons in roots.

(a) Dicistronic vector pAI4 with *aadA* marker gene and *gfp* in the second ORF. The pAI4 vector is similar to pMRR13, other than the *aadA-gfp* intergenic region contains the maize wild-type *atpH* 5'-UTR, encoding the PPR10 protein binding site. (b) Tobacco root protein (5 µg) separated in SDS-PAGE and stained with Coomassie brilliant blue. (c) Western blot analyses probing with the GFP antibody using GFP from the leaf of pMRR13 as a reference. (d) Quantification of the GFP protein in the tobacco root amyloplast. GFP from pMRR13 leaf was included as reference. GFP was quantified on Western blots with the ImageJ (version 2.0.0-rc-38/1.50b) software. The calculation was based on three biological replicates (mean±SD, n=3).
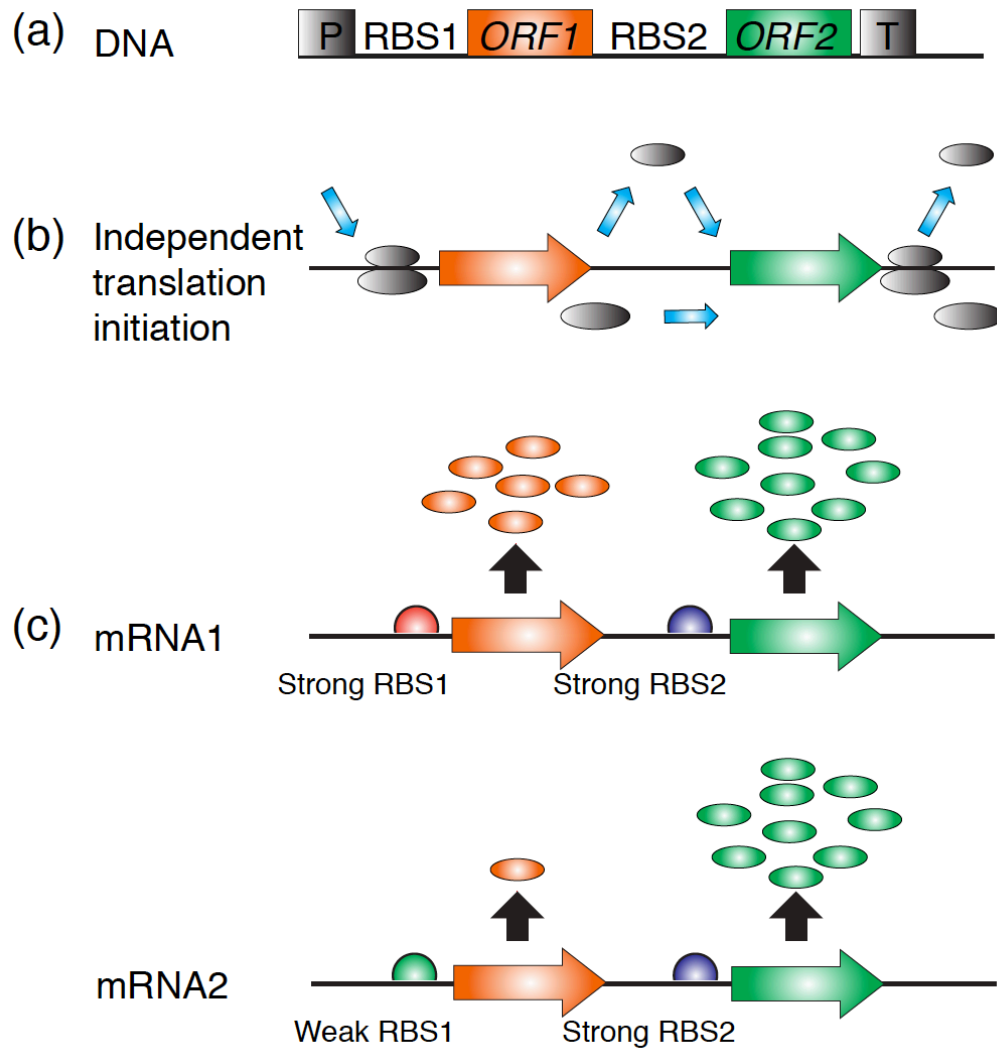
Figure 5-6. No translational coupling between the first and second ORF in the dicistronic markers. (a) Schematic map of dicistronic operons. ORF1 encodes a marker gene, the second ORF *gfp*. The operon has one promoter (P) and one 3'-UTR (T) for the stabilization of the mRNA. (b) Translation initiates independently from the first and second ORFs. Shown are the small and large ribosomal subunits entering the mRNA at the ribosome entry cite, and dissociating when translation of the 1st ORF is completed. (c) Protein accumulation from the first ORF has no

significant impact on protein accumulation from the 2nd ORF. Based on NPTII
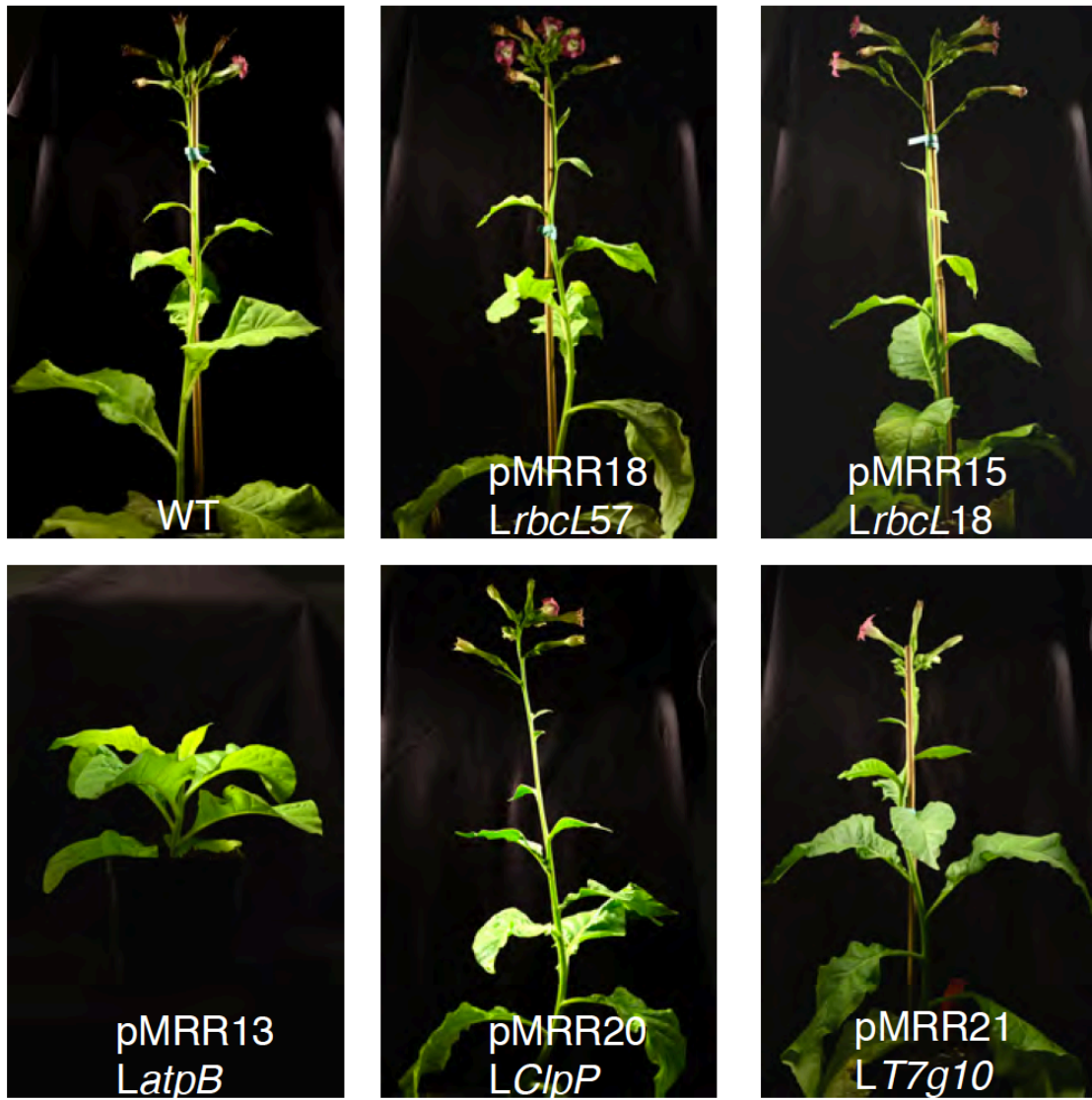
accumulation in the pMRR20 and pMRR21 plants (Table 2).

Figure 5-7. Plants in the greenhouse after eight weeks. Note delayed flowering of pMRR13 plant.
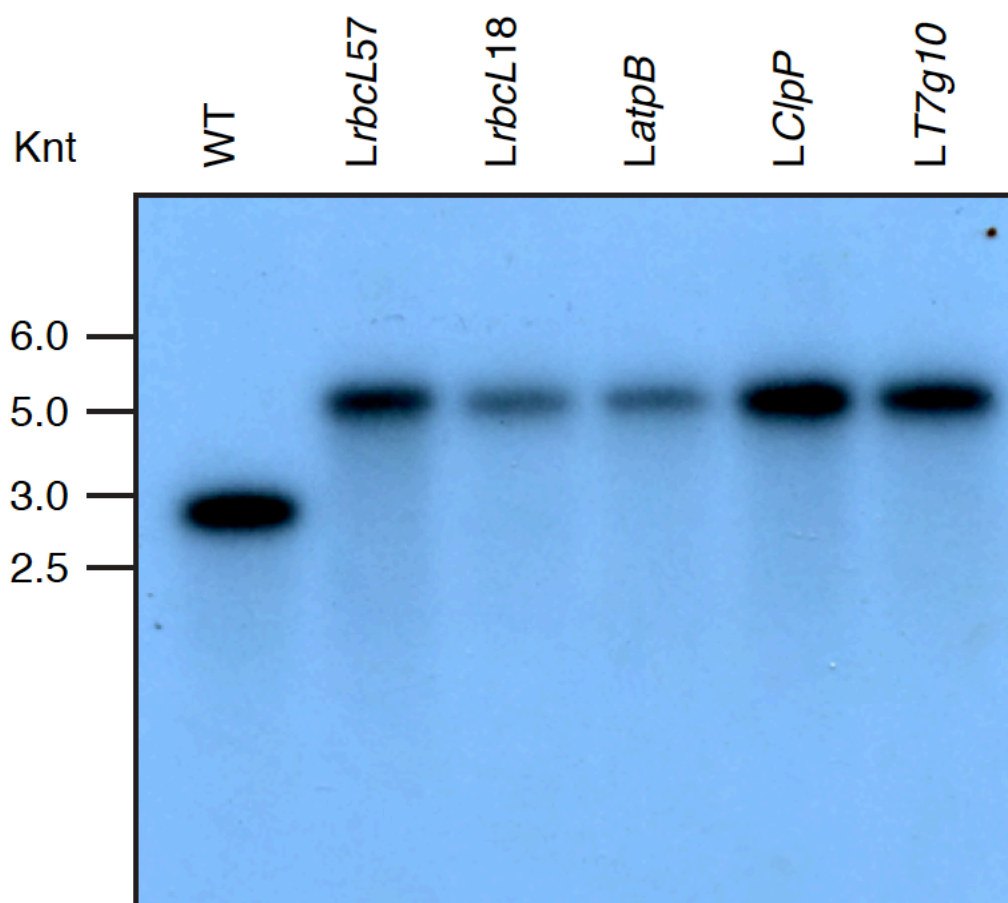
Figure 5-8. DNA gel blot confirms uniform transformation of the plastid genome. Total cellular DNA was digested with the *Bam*HI restriction enzyme and probed with the *Apa*I-*Bam*HI fragment (Figure 5-1a). Probe detects a 2.9 kb fragment in wild type DNA and the 5.7 kb fragment in transplastomic plants.

Table 5-1. Classification of kanamycin resistant clones by GFP expression

Vector	Leader 1	T0 – No. of events			1st Purification		
		KanR	GFP+	GFP-	KanR	GFP+	GFP-
pMRR20	L*ClpP*	61	45	16	52	46	6
pMRR21	L*T7g10*	15	13	2	9	9	0

Table 5-2. Quantification of GFP, AadA or NPTII in transplastomic leaves

Dicistronic vectors

Vector	Leader 1	1st ORF AAD/NPTII %	1st ORF Translation Efficiency	Leader 2	2nd ORF GFP %	2nd ORF Translation Efficiency	mRNA folds
pMRR18	L*rbcL57*	2.9±0.4	2.9±0.4	L*Cry9Aa2*	12.7±1.1	12.7±1.1	1.0
pMRR15	L*rbcL18*	0.5±0.1	0.18±0.06	L*Cry9Aa2*	29.3±0.6	9.9±3.0	3.2±1.2
pMRR13	L*atpB*	6.6±1.0	0.7±0.2	L*Cry9Aa2*	42.2±2.1	4.4±1.9	10.6±3.9
pMRR20	L*ClpP*	1.0±0.1	1.0±0.1	L*Cry9Aa2*	23.7±0.2	23.7±0.2	1.0
pMRR21	L*T7g10*	9.1±0.1	5.1±0.8	L*Cry9Aa2*	31.5±2.0	17.6±1.6	1.8±0.3

GFP accumulation from Monocistronic and Dicistronic LCry9Aa2 genes

Vector	*aadA* mRNA folds	AadA % of TSP	*gfp* mRNA folds	GFP %
pMRR13	1.0	6.6±1.0	1.0	42.2±2.1
pMRR17	0.1±0.02	<1	0.6±0.1	35.0±1.7

% GFP, AAD or NPTII of Total Soluble Protein (TSP).

Table 5-3. Plastid transformation vectors

Dicistronic vectors

Vector	P1	L1	ORF1	L2	ORF2	T	GenBank Accession No.
pMRR18	P*rrn*	L*rbcL57*	*aadA*	L*cry9Aa2*	*gfp*	T*psbA*	
pMRR15	P*rrn*	L*rbcL18*	*aadA*	L*cry9Aa2*	*gfp*	T*psbA*	
pMRR13	P*rrn*	L*atpB*	*aadA*	L*cry9Aa2*	*gfp*	T*psbA*	
pAI4	P*rrn*	L*atpB*	aadA	L*atpH*^{Zm}	*gfp*	T*psbA*	
pMRR20	P*rrn*	L*ClpP*	*neo*	L*cry9Aa2*	*gfp*	T*psbA*	
pMRR21	P*rrn*	L*T7g10*	*neo*	L*cry9Aa2*	*gfp*	T*psbA*	

Monocistronic vector

Vector	P1	L 1	ORF1	3'-UTR	P2	Leader 2	ORF2	T	GenBank Acc. No.
pMRR17	P*psbA*	L*psbA*	*aadA*	T*psbA*	P*rrn*	L*cvry9Aa2*	*gfp*	TrbcL	

References

1   Bock, R. Engineering Plastid Genomes: Methods, Tools, and Applications in Basic Research and Biotechnology. *Annu. Rev. Plant. Biol.* **66**, 211-241, doi:10.1146/annurev-arplant-050213-040212 (2015).

2   Zoschke, R. & Bock, R. Chloroplast Translation: Structural and Functional Organization, Operational Control, and Regulation. *Plant Cell* **30**, 745-770, doi:10.1105/tpc.18.00016 (2018).

3   Barkan, A. Expression of Plastid Genes: Organelle-Specific Elaborations on a Prokaryotic Scaffold. *Plant Physiol.* **155**, 1520-1532, doi:pp.110.171231 [pii]10.1104/pp.110.171231 (2011).

4   Waheed, M. T., Ismail, H., Gottschamel, J., Mirza, B. & Lossl, A. G. Plastids: The Green Frontiers for Vaccine Production. *Front Plant Sci* **6**, 1005, doi:10.3389/fpls.2015.01005 (2015).

5   Daniell, H., Lin, C. S., Yu, M. & Chang, W. J. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biology* **17**, doi:ARTN 134 10.1186/s13059-016-1004-2 (2016).

6   Ahmad, N., Michoux, F., Lossl, A. G. & Nixon, P. J. Challenges and perspectives in commercializing plastid transformation technology. *J Exp Bot* **67**, 5945-5960, doi:10.1093/jxb/erw360 (2016).

7   Liebers, M. *et al.* Regulatory Shifts in Plastid Transcription Play a Key Role in Morphological Conversions of Plastids during Plant Development. *Front Plant Sci* **8**, 23, doi:10.3389/fpls.2017.00023 (2017).

8   Stern, D. B., Goldschmidt-Clermont, M. & Hanson, M. R. Chloroplast RNA metabolism. *Annu. Rev. Plant. Biol.* **61**, 125-155, doi:10.1146/annurev-arplant-042809-112242 (2010).

9   Harada, H. *et al.* Construction of transplastomic lettuce (Lactuca sativa) dominantly producing astaxanthin fatty acid esters and detailed chemical analysis of generated carotenoids. *Transgenic Res.* **23**, 303-315, doi:10.1007/s11248-013-9750-3 (2014).

10   Ort, D. R. *et al.* Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proc. Natl. Acad. Sci. USA* **112**, 8529-8536, doi:10.1073/pnas.1424031112 (2015).

11   Lin, M. T. *et al.* beta-Carboxysomal proteins assemble into highly organized structures in *Nicotiana* chloroplasts. *Plant J.* **79**, 1-12, doi:10.1111/tpj.12536 (2014).

12   Lin, M. T., Occhialini, A., Andralojc, P. J., Parry, M. A. & Hanson, M. R. A faster Rubisco with potential to increase photosynthesis in crops. *Nature* **513**, 547-550, doi:10.1038/nature13776 (2014).

13   Parry, M. A. *et al.* Rubisco activity and regulation as targets for crop improvement. *J Exp Bot* **64**, 717-730, doi:10.1093/jxb/ers336 (2013).

14   Zhang, J. *et al.* Pest control. Full crop protection from an insect pest by expression of long double-stranded RNAs in plastids. *Science* **347**, 991-994, doi:10.1126/science.1261680 (2015).

15   Bally, J. *et al.* In-Plant Protection against *Helicoverpa armigera* by Production of Long hpRNA in Chloroplasts. *Front Plant Sci* **7**, doi:ARTN 145310.3389/fpls.2016.01453 (2016).

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

17   Azhagiri, A. & Maliga, P. Exceptional paternal inheritance of plastids in *Arabidopsis* suggests that low frequency leakage of plastid *via* pollen may be universal in plants. *Plant J.* **52**, 817-823, doi:10.1111/j.1365-313X.2007.03278.x (2007).

18   Ruf, S., Karcher, D. & Bock, R. Determining the transgene containment level provided by chloroplast transformation. *Proc. Natl. Acad. Sci. USA* **104**, 6998-7002 (2007).

19   Moll, B., Polsby, L. & Maliga, P. Streptomycin and lincomycin resistances are selective plastid markers in cultured *Nicotiana* cells. *Mol. Gen. Genet.* **221**, 245-250 (1990).

20   Lutz, K. A. & Maliga, P. Plastid genomes in a regenerating tobacco shoot derive from a small number of copies selected through a stochastic process. *Plant J.* **56**, 975-983, doi:10.1111/j.1365-313X.2008.03655.x (2008).

21   Maliga, P. in *Genomics of Chloroplasts and Mitochondria* Vol. 35 (eds R. Bock & V. Knoop) 393-414 (Springer, 2012).

22   Carrer, H., Hockenberry, T. N., Svab, Z. & Maliga, P. Kanamycin resistance as a selectable marker for plastid transformation in tobacco. *Mol. Gen. Genet.* **241**, 49-56 (1993).

23   Huang, F. C. *et al.* Efficient plastid transformation in tobacco using the *aphA-6* gene and kanamycin selection. *Mol. Genet. Genomics* **268**, 19-27, doi:10.1007/s00438-002-0738-6 (2002).

24   Li, W., Ruf, S. & Bock, R. Chloramphenicol acetyltransferase as selectable marker for plastid transformation. *Plant Mol. Biol.* **76**, 443-451, doi:10.1007/s11103-010-9678-4 (2011).

25   Tabatabaei, I., Ruf, S. & Bock, R. A bifunctional aminoglycoside acetyltransferase/phosphotransferase conferring tobramycin resistance provides an efficient selectable marker for plastid transformation. *Plant Mol. Biol.* **93**, 269-281, doi:10.1007/s11103-016-0560-x (2017).

26   Rosellini, D., LaFayette, P. R., Barone, P., Veronesi, F. & Parrott, W. A. Kanamycin-resistant alfalfa has a point mutation in the 16S plastid rRNA. *Plant Cell Rep.* **22**, 774-779, doi:DOI 10.1007/s00299-004-0757-3 (2004).

27   Yu, Q., Lutz, K. A. & Maliga, P. Efficient plastid transformation in Arabidopsis. *Plant Physiol.* **175**, 186-193 (2017).

28   Chakrabarti, S. K., Lutz, K. A., Lertwirijawong, B., Svab, Z. & Maliga, P. Expression of the *cry9Aa2 B.t.* gene in the tobacco chlroplasts confers resistance to potato tuber moth. *Transgenic Res.* **15**, 481-488, doi:10.1007/s11248-006-0018-z (2006).

29   Yu, Q., Barkan, A. & Maliga, P. Engineered RNA-binding protein for transgene activation in non-green plastids. *Nat Plants* **5**, 486-490, doi:10.1038/s41477-019-0413-0 (2019).

30   Yu, Q., LaManna, L., Kelly, M. E., Lutz, K. A. & Maliga, P. New Tools for Engineering the Arabidopsis Plastid Genome. *Plant Physiol.* **181**, 394-398 (2019).

31   Kuroda, H. & Maliga, P. Over-expression of the *clpP* 5'-UTR in a chimeric context causes a mutant phenotype suggesting competition for a *clpP*-specific RNA maturation factor in tobacco chloroplasts. *Plant Physiol.* **129**, 1600-1606, doi:10.1104/pp.004986 (2002).

32   Kuroda, H. & Maliga, P. Complementarity of the 16S rRNA penultimate stem with sequences downstream of the AUG destabilizes the plastid mRNAs. *Nucleic Acids Res.* **29**, 970-975 (2001).

33   Staub, J. M. & Maliga, P. Accumulation of D1 polypeptide in tobacco plastids is regulated via the untranslated region of the *psbA* mRNA. *EMBO J.* **12**, 601-606 (1993).

34   Staub, J. M. & Maliga, P. Translation of *psbA* mRNA is regulated by light *via* the 5'-untranslated region in tobacco plastids. *Plant J.* **6**, 547-553 (1994).

35   Hibberd, J. M., Linley, P. J., Khan, M. S. & Gray, J. C. Transient expression of green fluorescent protein in various plastid types following microprojectile bombardment. *Plant J.* **16**, 627-632. (1998).

36   Shiina, T., Hayashi, K., Ishii, N., Morikawa, K. & Toyoshima, Y. Chloroplast tubules visualized in transplastomic plants expressing green fluorescent protein. *Plant Cell Physiol* **41**, 367-371 (2000).

37   Franklin, S., Ngo, B., Efuet, E. & Mayfield, S. P. Development of a GFP reporter gene for *Chlamydomonas reinhardtii* chloroplast. *Plant J.* **30**, 733 (2002).

38   Khan, M. S. & Maliga, P. Fluorescent antibiotic resistance marker to track plastid transformation in higher plants. *Nat. Biotechnol.* **17**, 910-915 (1999).

39   Rojas, M., Yu, Q., Williams-Carrier, R., Maliga, P. & Barkan, A. Engineered PPR proteins as inducible switches to activate the expression of chloroplast transgenes. *Nat Plants* **5**, 505-511 (2019).

40   Kozak, M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **361**, 13-37, doi:10.1016/j.gene.2005.06.037 (2005).

41   Cornelissen, M. & Vandewiele, M. Nuclear transcriptional activity of the tobacco plastid *psbA* promoter. *Nucleic Acids Res.* **17**, 19-29 (1989).

42   Stegemann, S. & Bock, R. Experimental reconstruction of functional gene transfer from the tobacco plastid genome to the nucleus. *Plant Cell* **18**, 2869-2878, doi:10.1105/tpc.106.046466 (2006).

43   Huber, M. *et al.* Translational coupling via termination-reinitiation in archaea and bacteria. *Nat Commun* **10**, 4006, doi:10.1038/s41467-019-11999-9 (2019).

44   Levin-Karp, A. *et al.* Quantifying translational coupling in E. coli synthetic operons using RBS modulation and fluorescent reporters. *ACS synthetic biology* **2**, 327-336, doi:10.1021/sb400002n (2013).

45   Yukawa, M. & Sugiura, M. Termination codon-dependent translation of partially overlapping ndhC-ndhK transcripts in chloroplasts. *Proc. Natl. Acad. Sci. USA* **105**, 19550-19554, doi:10.1073/pnas.0809240105 (2008).

46   Yukawa, M. & Sugiura, M. Additional pathway to translate the downstream ndhK cistron in partially overlapping ndhC-ndhK mRNAs in chloroplasts. *Proc. Natl. Acad. Sci. USA* **110**, 5701-5706, doi:10.1073/pnas.1219914110 (2013).

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

48   Yao, W. B., Meng, B. Y., Tanaka, M. & Sugiura, M. An additional promoter within the protein-coding region of the psbD-psbC gene cluster in tobacco chloroplast DNA. *Nucleic Acids Res.* **17**, 9583-9591 (1989).

49   Avila, E. M., Gisby, M. F. & Day, A. Seamless editing of the chloroplast genome in plants. *Bmc Plant Biology* **16**, doi:ARTN 16810.1186/s12870-016-0857-6 (2016).

50   Suzuki, H., Kuroda, H., Yukawa, Y. & Sugiura, M. The downstream atpE cistron is efficiently translated via its own cis-element in partially overlapping atpB-atpE dicistronic mRNAs in chloroplasts. *Nucleic Acids Res.* **39**, 9405-9412, doi:10.1093/nar/gkr644 (2011).

51   Zoschke, R., Watkins, K. P. & Barkan, A. A Rapid Ribosome Profiling Method Elucidates Chloroplast Ribosome Behavior in Vivo. *Plant Cell* **25**, 2265-2275, doi:10.1105/tpc.113.111567 (2013).

52   Suzuki, J. Y., Sriraman, P., Svab, Z. & Maliga, P. Unique architecture of the plastid ribosomal RNA operon promoter recognized by the multisubunit RNA polymerase (PEP) in tobacco and other higher plants. *Plant Cell* **15**, 195-205 (2003).

53   Sugita, M. & Sugiura, M. Nucleotide sequence and transcription of the gene for the 32,000 dalton thylakoid membrane protein from *Nicotiana tabacum*. *Mol Gen Genet* **195**, 308-313 (1984).

54   Hajdukiewicz, P. T. J., Allison, L. A. & Maliga, P. 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 (1997).

55   Shinozaki, K. & Sugiura, M. The nucleotide sequence of the tobacco chloroplast gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase. *Gene* **20**, 91-102 (1982).

56   Orozco, E. M. J., Chen, L. J. & Eilers, R. J. The divergently transcribed *rbcL* and *atpB* genes of tobacco plastid DNA are separated by nineteen base pairs. *Curr. Genet.* **17**, 65-71 (1990).

57   Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60-89 (1990).

58   Beck, E., Ludwig, G., Auerswald, E. A., Reiss, B. & Schaller, H. Nucleotide sequence and exact localization of the neomycin phosphotransferase gene from transposon Tn5. *Gene* **19**, 327-336 (1982).

59   Svab, Z. & Maliga, P. High-frequency plastid transformation in tobacco by selection for a chimeric *aadA* gene. *Proc. Natl. Acad. Sci. USA* **90**, 913-917 (1993).

60   Davis, S. J. & Vierstra, R. D. Soluble, highly fluorescent variants of green fluorescent protein (GFP) for use in higher plants. *Plant Mol. Biol.* **36**, 521-528 (1998).

61   Lutz, K. A., Svab, Z. & Maliga, P. Construction of marker-free transplastomic tobacco using the Cre-*loxP* site-specific recombination system. *Nat. Protocols* **1**, 900-910, doi:10.1038/nprot.2006.118 (2006).

62   Maliga, P. & Tungsuchat-Huang, T. in *Chloroplast Biotechnology: Methods and Protocols* Vol. 1132 *Methods in Molecular Biology* (ed P. Maliga) Ch. 8, 147-163 (Springer Science+Business Media, 2014).

63   Kuroda, H. & Maliga, P. Sequences downstream of the translation initiation codon are important determinants of translation efficiency in chloroplasts. *Plant Physiol.* **125**, 430-436 (2001).

64   Carrer, H., Staub, J. M. & Maliga, P. Gentamycin resistance in *Nicotiana* conferred by AAC(3)-I, a narrow substrate specificity acetyl transferase. *Plant Mol. Biol.* **17**, 301-303 (1990).

65   Zommick, D. H., Kumar, G. N., Knowles, L. O. & Knowles, N. R. Translucent tissue defect in potato (*Solanum tuberosum* L.) tubers is associated with oxidative stress accompanying an accelerated aging phenotype. *Planta* **238**, 1125-1145, doi:10.1007/s00425-013-1951-8 (2013).

66   Shinozaki, K. *et al.* The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J.* **5**, 2043-2049 (1986).