PCP
PLANT & CELL PHYSIOLOGY

Zinc Finger Artificial Transcription Factor-Mediated Chloroplast Genome Interrogation in *Arabidopsis thaliana*

Niels van Tol, Gema Flores Andaluz, Hendrika A.C.F. Leeggangers, M. Reza Roushan, Paul J.J. Hooykaas and Bert J. van der Zaal*

Institute of Biology Leiden, Faculty of Science, Leiden University, Sylviusweg 72, Leiden 2333 BE, The Netherlands
*Corresponding author: E-mail, b.j.v.d.zaal@biology.leidennuniv.nl.
Subject areas: Regulation of gene expression; photosynthesis, respiration and bioenergetics
(Received July 19, 2018; Accepted November 1, 2018)

The large majority of core photosynthesis proteins in plants are encoded by nuclear genes, but a small portion have been retained in the plastid genome. These plastid-encoded chloroplast proteins fulfill essential roles in the process of photochemistry. Here, we report the use of nuclear-encoded, chloroplast-targeted zinc finger artificial transcription factors (ZF-ATFs) with effector domains of prokaryotic origin to modulate the expression of chloroplast genes, and to enhance the photochemical activity and growth characteristics of *Arabidopsis thaliana* plants. This technique was named chloroplast genome interrogation. Using this novel approach, we obtained evidence that ZF-ATFs can indeed be translocated to chloroplasts of Arabidopsis plants, can modulate their growth and operating light use efficiency of PSII, and finally can induce statistically significant changes in the expression levels of several chloroplast genes. Our data suggest that the distortion of chloroplast gene expression might be a feasible approach to manipulate the efficiency of photosynthesis in plants.

Keywords: Arabidopsis • Artificial transcription factors • Chloroplasts • Photosynthesis • PSII • Zinc fingers.

Introduction

Photosynthesis is the process that fixes solar energy as chemical energy. In green plant tissues, it is conducted by specialized plastid organelles named chloroplasts, which harbor the photosynthetic apparatus. Light is absorbed by Chl molecules that are associated with PSI and PSII that are anchored in the thylakoid membranes of chloroplasts, and catalyze the photoexcitation of electrons. The resulting linear electron transport leads to the photoreduction of $NADP^+$, and indirectly drives the synthesis of ATP through a pH gradient that is generated by the Cyt b_6f proton pump through chemiosmotic coupling. In the light, the energy-rich compounds NADPH and ATP are used in the Calvin–Benson cycle for CO_2 fixation by the enzyme complex RuBisCo to yield a carbohydrate product that can be partitioned to different plant organs and used for various metabolic processes supporting plant growth and development.

During the domestication of photosynthetic bacteria as chloroplasts, an estimated 4,500 bacterial genes have been lost, or incorporated into the nuclear genomes of plants (Martin et al. 2002). These genes have acquired eukaryotic gene expression signals and in many cases the encoded proteins contain N-terminal signal peptides known as chloroplast transit peptides (CTPs) which mediate chloroplast import (Bruce 2000). Importantly, a small fraction of the bacterial genes have been retained within chloroplasts. In higher plants, these genes now reside on a single circular chromosome of 120–170 kb that is maintained in high copy numbers in the chloroplast stroma. Chloroplast DNA therefore accounts for a very significant portion of the total cellular DNA, with up to 50 copies per chloroplast and up to 100 chloroplasts per cell in mature photosynthetic tissue (Flores-Perez and Jarvis 2013). The chloroplast genome of the model plant species *Arabidopsis thaliana* encodes 54 thylakoid membrane proteins, 31 plastid ribosomal proteins and also contains 45 tRNA- and rRNA-encoding genes (Jarvis and Lopez-Juez 2013). The relatively high gene density of the chloroplast genome is largely due to the organization of clusters of genes into operons, thus placing expression of a polycistronic mRNA encoding multiple gene products under control of a single promoter and its accessory regulatory elements (Stoppel and Meurer 2013, Borner et al. 2015a).

The engineering of chloroplast genes has been designated as one of the targets for increasing the efficiency of photosynthesis in plants (Ort et al. 2015). As many chloroplast-encoded proteins have structural or catalytic roles in chloroplast function, the introduction of mutant alleles or orthologous genes from other photosynthetic organisms might result in more efficient thylakoid membrane function. For a limited number of plant species, there are plastid transformation protocols available to introduce gene constructs (e.g. encoding antibiotic resistance genes) into target chloroplast genomic loci through homologous DNA recombination (Day and Goldschmidt-Clermont 2011, Bock 2015). It is, however, typically very tedious to generate homoplasmic plants, i.e. plants with cells containing non-segregating chloroplasts with the genetically modified copies of the chloroplast genome. Achieving homoplasmy typically requires multiple rounds of selection for the presence of the marker gene and subsequent regeneration of resistant tissue into plants (Day and Goldschmidt-Clermont 2011). This procedure is particularly difficult to complete for tissues in which

Plant Cell Physiol. 0(0): 1–14 (2018) doi:10.1093/pcp/pcy216, Advance Access publication on 6 November 2018, available online at www.pcp.oxfordjournals.org

the chloroplast genome is maintained in very high copy numbers. Another important pitfall of reverse genetic approaches to mutate or differentially express thylakoid membrane components is that they usually result in impairment rather than gain of thylakoid membrane function. In this study, we have explored the use of artificial transcription factor (ATF)-mediated genome interrogation to modulate chloroplast gene expression levels as a novel method to induce changes in the photochemical performance of Arabidopsis plants.

The key principle of genome interrogation is based on the introduction of ATFs with low complexity DNA-binding domains to induce large-scale changes in gene expression patterns that might lead to different phenotypes of interest (Beltran et al. 2006). In our lab, we have successfully used ATFs with zinc fingers (ZFs) as DNA-binding domains (ZF-ATFs) for genome interrogation experiments in Arabidopsis (Lindhout et al. 2006, Jia et al. 2013, van Tol and van der Zaal 2014, van Tol et al. 2016, van Tol et al. 2017b). In our set-up, the ZF-ATFs contained an array of three of the 16 different ZFs that can each recognize a cognate 3 bp consensus DNA sequence of 5′-GNN-3′ (Segal et al. 1999), with 'N' being any of the four bases. The ZF domains were fused to protein moieties that can either stimulate or repress transcription, such as the transcriptional activation domain of the herpes simplex VP16 protein or the EAR transcriptional repressor motif from Arabidopsis itself (Sadowski et al. 1988, Hiratsu et al. 2003, Park et al. 2005, Lindhout et al. 2006, Mito et al. 2011). Gene constructs encoding these 3F-ATFs can be introduced into the nuclear plant genome through *Agrobacterium tumefaciens*-mediated floral dip transformation to obtain transgenic plants. The cognate 9 bp target site of each 3F-ATF will on average occur once in every 130,000 bp of double-stranded DNA, and thus approximately 1,000 times within the 130 Mbp Arabidopsis genome. In this way, by expression of a single 3F-ATF, a large number of genes near the cognate target sites might be differentially expressed *in trans* and in a dominant manner, which could trigger novel phenotypes to arise.

In the present study, we have explored the use of ZF-ATFs in chloroplasts. To modulate chloroplast gene expression, it had to be taken into account that chloroplasts have also retained their own transcriptional and regulatory machinery, consisting of the phage-type nuclear-encoded RNA polymerase (NEP), which mostly transcribes plastid housekeeping genes, and the bacterial type plastid-encoded RNA polymerase (PEP), which mostly transcribes photosynthesis genes (Borner et al. 2015b). The process of plastid gene expression is also tightly regulated through anterograde and retrograde signaling with the nucleus (Leister 2005, Woodson and Chory 2008). In view of these considerations, we had to redesign our previously established genome interrogation set-up in such a way that ZF-ATFs can function in an essentially prokaryotic environment.

Here, we describe the construction of ZF-ATF expression cassettes that can be introduced into the nuclear plant genome using standard methods, and can result in ZF-ATF activity in chloroplasts. This system was tested by expressing chloroplast-targeted fusions of the bacterial transcriptional activators CRP and LuxR to low complexity DNA-binding domains consisting of two zinc fingers (2Fs). We obtained evidence that a relatively small collection of 2F-ATFs already contained constructs that induced variation in the growth and operating light use efficiency of PSII reaction centers of Arabidopsis plants, through which we provide the first potential tool for organellar genome interrogation. Our data indicate that manipulation of chloroplast gene expression patterns could further be explored as an option for the enhancement of plant photosynthesis.

Results

Design of the chloroplast genome interrogation system

Gene constructs encoding ZF-ATFs with novel features had to be designed for genome interrogation experiments in chloroplasts. Foremost, as described above, the expression of chloroplast genes is mediated by a system of polymerases and regulatory proteins that are of bacterial origin. As there is no evidence at present that established modulators of eukaryotic gene expression (e.g. VP16) can also function as such within a prokaryotic context, we prospected for prokaryotic protein modules that are direct activators of gene expression, and could therefore also activate PEP-mediated gene expression. Firstly, we selected the *Escherichia coli* cAMP receptor protein (CRP), which has been shown to activate *lac* gene expression in *E. coli* through a direct interaction with RNA polymerase (Borukhov and Lee 2005). CRP has also previously been used for genome interrogation experiments in *E. coli* (Lee et al. 2008). For the present study, we opted for the use of the C-terminal part of CRP (truncation CRPD2) consisting of amino acids 134–190, which lacks the cAMP-binding domain and is a more potent transcriptional activator than the full-length CRP protein (Lee et al. 2008). As a second option we selected the *Aliivibrio fischeri* protein LuxR, which is a regulator of *lux* gene promoters (Dunlap 2014). The C-terminal part of LuxR lacking the N-terminal amino acids 2–162 (designated LuxRΔN) was reported to contain the most critical amino acids for the interaction with RNA polymerase and to lead to inducer-independent transcriptional activation activity in *A. fischeri* (Choi and Greenberg 1991, Schu et al. 2009). Importantly, LuxRΔN was also shown to possess transcriptional transactivation activity in *E. coli* (Volzing et al. 2011). Based on the published characteristics of CRPD2 and LuxRΔN, we thus hypothesized that both might be suitable modulators of PEP-mediated transcriptional activity in chloroplasts without the requirement for other regulatory proteins. For the sake of clarity, CRPD2 and LuxRΔN are referred to herein as CRP and LuxR, respectively.

As the concept of genome interrogation relies on generating genome-wide changes in gene expression patterns, we did not consider the use of 3Fs, as the 155,000 bp chloroplast genome on average contains only one 9 bp 3F-binding site. Instead, we opted to make use of 2F domains, which have 6 bp DNA recognition sites that each can be expected to occur approximately 75 times in a typically sized chloroplast genome.

The activity of translationally fused CRP or LuxR domains could then lead to differential gene expression at many chloroplast genomic loci, provided that the affinity of 2F domains for DNA is still high enough to allow for a preferential presence at these target sites. In support of this idea, we have previously found that expression of different nuclear-targeted 2F-ATFs can lead to transcriptional changes in Arabidopsis (Lindhout et al. 2006, Jia et al. 2013). More recently, we have found that salinity tolerance can be induced by a 2F–VP16 fusion (van Tol et al. 2016). For chloroplast genome interrogation experiments, we thus decided to select eight different 2Fs randomly for ZF-ATF construction. These 2Fs were denoted as 2F1–2F8 (**Table 1**). To avoid any inhibitory effects on the activity of the effector domains when translationally fused to 2F domains, we selected a previously published flexible linker peptide optimized for LuxR activity (Volzing et al. 2011) to function as a spacer between the 2F and effector domain modules. This linker consists of five repeats of the peptide ARTQYSESM (each separated by the amino acid G) (Volzing et al. 2011), and provides for a distance of 150 Å between the 2Fs and the effector domains, which was determined to be optimal for LuxRΔN activity (Volzing et al. 2011). In order to achieve the translocation of ZF-ATFs into the stroma of chloroplasts, we chose to use the N-terminal CTP of the FedA protein of Arabidopsis (Smeekens et al. 1989, Somers et al. 1990). This CTP has been shown to mediate the translocation of heterologous proteins into chloroplasts (Smeekens et al. 1987, Jin et al. 2003). We simultaneously opted to make use of the promoter of the *FEDA* gene to drive ZF-ATF expression. A hypothetical model of chloroplast genome interrogation activity using the presented components is provided in **Fig. 1**, and an overview of the expression cassettes that were designed is presented in **Fig. 2**. Finally, an overview of the amino acid sequences of the translational fusions encoded by the effector constructs is provided in **Fig. 3**.

To confirm that the CTP of FedA also allows for import of 2F fusions into chloroplasts, and to assess the effect of the presence of 2Fs on translocation of proteins into the chloroplasts, leaves from primary transformants harboring the constructs pGFP, pCTP-GFP and pCTP-2F1-GFP were analyzed by confocal microscopy. Confocal images were captured using leaf number 5 or 6 (Telfer et al. 1997) from approximately five randomly chosen primary transformants for each construct and were analyzed for co-localization of Chl fluorescence signal with green fluorescent protein (GFP) signal, represented by yellow/orange pixels in merged images. As expected, GFP signal could be detected in the case of all three constructs (**Fig. 4A**). Primary transformants harboring pGFP displayed GFP signal which did not overlap with Chl fluorescence (**Fig. 4A**), indicating that these transformants expressed GFP in the cytoplasm and/or the nucleus as expected, but not in the chloroplast stroma. In the merged channel of the images taken from primary transformants harboring pCTP-GFP and pCTP-2F1-GFP, there was a clear and similar overlap of the GFP and Chl fluorescence signal (**Fig. 4A**), indicating that the CTP of FedA indeed allowed for the translocation of GFP as well as 2F1–GFP into chloroplasts and that the addition of two ZFs to the GFP thus did not disrupt chloroplast import. To corroborate co-localization between GFP signal and chloroplasts, we examined leaf protoplasts of kanamycin-resistant T_2 plants harboring the constructs pGFP, pCTP-GFP and pCTP-2F1-GFP, which were transfected with the plastidial marker plasmid plastid-mCherry by confocal microscopy (**Fig. 4B**). Expression of plastid-mCherry results in chloroplast targeting of the red fluorescent protein mCherry, thereby allowing for the fluorescent labeling of chloroplasts. While protoplasts harboring pGFP displayed GFP signal spread throughout the cells without any apparent co-localization, protoplasts harboring pCTP-GFP and pCTP-2F1-GFP clearly displayed GFP signal which co-localized with plastid-mCherry signal (**Fig. 4B**), thereby corroborating that the CTP of FedA allows for chloroplast translocation which was not hampered by fusion to two ZFs.

Table 1 2Fs that were randomly assembled for chloroplast genome interrogation binary vector construction

Name	5′–3′ DNA recognition sequence
2F1	GTC-GGG
2F2	GGG-GGA
2F3	GGA-GAG
2F4	GAG-GAT
2F5	GGG-GTA
2F6	GAT-GTC
2F7	GCC-GCT
2F8	GGA-GCC

Assessment of transformation efficiency and primary transformant viability

To investigate whether ZF-ATF-mediated chloroplast genome interrogation can be used to introduce variation in the growth and photosynthetic performance of Arabidopsis plants, we first transformed Arabidopsis Columbia-0 (Col-0) plants with T-DNA constructs encoding CTP–LuxR and CTP–CRP fusions with or without 2Fs as DNA-binding domains (**Fig. 2**) through *A. tumefaciens*-mediated floral dip transformation (Clough and Bent 1998). For all constructs, we could obtain kanamycin-resistant primary transformants (T_1 generation) at equal transformation frequencies of approximately 1% (10 transformants per 1,000 seeds screened), except—and repeatedly so—for constructs pCTP-2F1-CRP and pCTP-2F5-CRP, which did not yield enough primary transformants for quantitative analysis, indicating that expression of these particular 2F–CRP fusions resulted in embryo or seedling lethality. Except for these cases, there was therefore no obvious correlation between transformation efficiency and the composition of the constructs used for transformation. However, all the other constructs allowed for recovery of viable primary transformants. We did observe some variation in the growth of primary transformant seedlings, but this was randomly distributed and thus most probably due to differences in antibiotic resistance during the selection on medium containing kanamycin. When further cultivated on soil the primary transformants did not exhibit any conspicuous phenotypes that could be attributed to the expression constructs, or to the presence of a CTP or of 2F domains. When

studied in more detail, however, there was variation in rosette size and photochemical efficiency, which will be discussed further below.

Chloroplast genome interrogation with 2F–CRP but not with 2F–LuxR fusions induces significant variation in growth and photosynthetic performance

To investigate whether chloroplast genome interrogation can be used to trigger variation in the photosynthetic properties of Arabidopsis plants, we quantified projected rosette surface area (RSA) and the operating light use efficiency of PSII (φPSII) of primary transformants harboring chloroplast genome interrogation constructs. RSA was quantified as a non-destructive proxy for biomass accumulation (Leister et al. 1999) and φPSII as a non-destructive proxy for overall photosynthetic rate (Baker 2008), both still allowing for harvesting seeds of the primary transformants (T_2 seeds). These two data sets were subsequently combined to generate two-dimensional plots for the comparison of the overall performances of the primary transformants in terms of growth and photochemical efficiency. First, we assessed the variation for these two parameters among a population of wild-type Col-0 plants (Fig. 5). Both RSA and φPSII were normally distributed for Col-0, and there was clear clustering of the data points (Fig. 5), demonstrating that Col-0 plants displayed relatively little variation in growth and photochemical efficiency. Given the clustering of the data points for Col-0, we

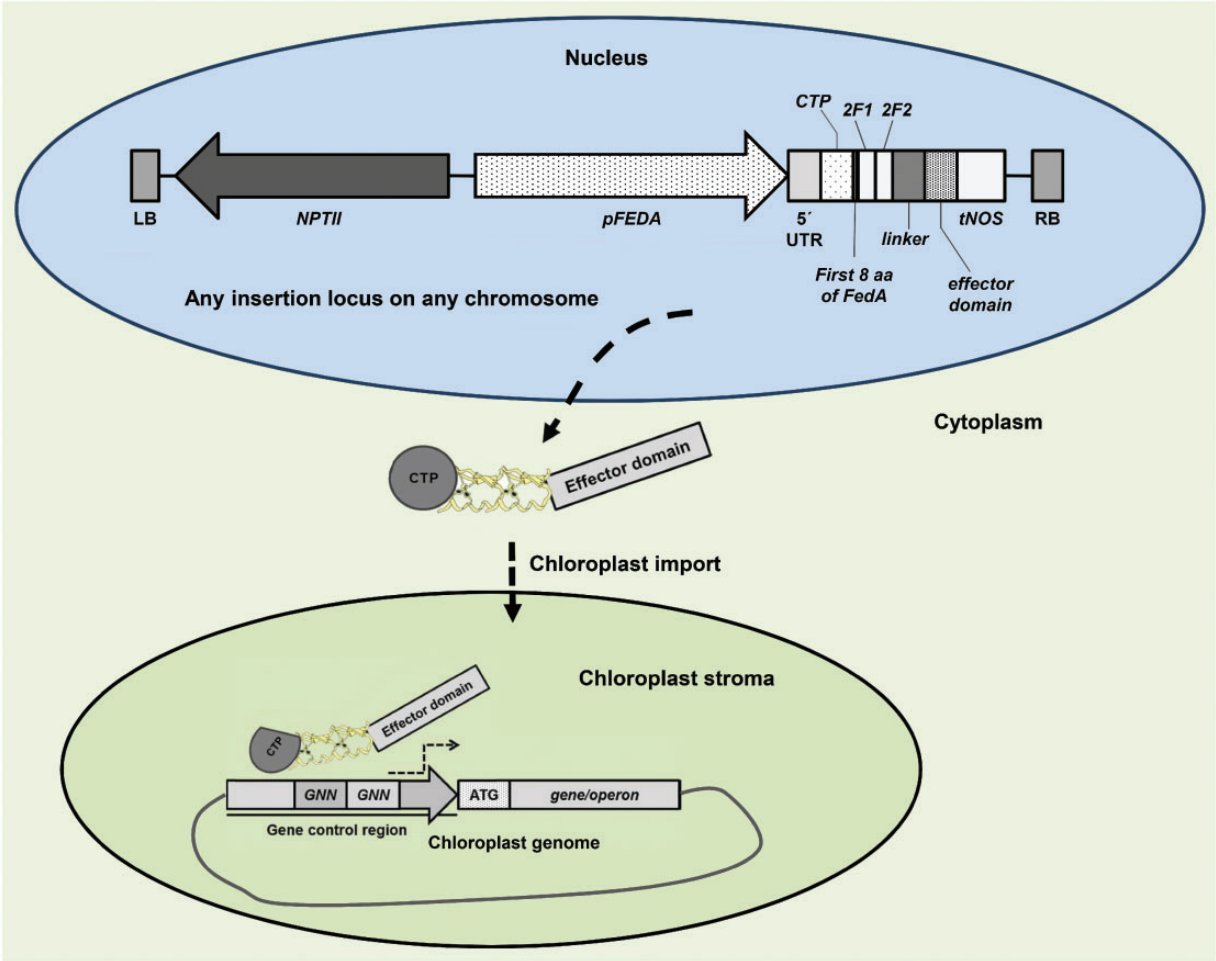

Fig. 1 A hypothetical model for genome interrogation in chloroplasts through nuclear-encoded ZF-ATFs. Constructs encoding ZF-ATFs are introduced into the nuclear genome of Arabidopsis as T-DNAs with left and right border sequences (LB and RB, respectively) through *Agrobacterium tumefaciens*-mediated transformation. The ZF-ATF constructs encode fusion proteins consisting of two zinc fingers modules (*ZF1* and *ZF2*) fused to an effector domain with transcriptional activation activity (*effector domain*) through a flexible linker peptide (*linker*), and tagged with the chloroplast transit peptide (*CTP*) of FedA and the first eight amino acids of the mature FedA protein (*First 8 aa of FedA*) at the N-terminus. Expression of the fusion construct is under control of the promoter of the Arabidopsis *FEDA* gene (*pFEDA*) including the 5′-untranslated region and the NOS terminator (*tNOS*). Presence of the T-DNA in the genome of primary transformants is determined by selection for kanamycin resistance due to expression of the *NPTII* gene, which is also on the T-DNA. When the ZF-ATF-encoding transgene is expressed in the nucleus, the resulting ZF-ATF protein is imported into the chloroplast stroma through recognition and cleavage of the CTP. In the chloroplast stroma, transient interactions between the ZF-ATF and 6 bp GNNGNN recognition sites of the two zinc fingers (with N being any of the four bases depending on the particular ZF) in the gene control region of any chloroplast gene or operon can result in modulation of transcription.

considered analyzing the combination of RSA and φPSII as a suitable method to test for introduction of alterations in growth and photosynthetic performance by chloroplast genome interrogation. We then analyzed RSA and φPSII of primary transformants harboring either 2F-ATF chloroplast genome interrogation constructs (pCTP-2Fn-CRP or pCTP-2Fn-LuxR), or control constructs without 2Fs as DNA-binding domains, which translocated either only the linker peptide (pCTP-Linker) or only the effector domains (pCTP-CRP and pCTP-LuxR). The resulting plots for primary transformants expressing CRP fusions are presented in **Fig. 5** and for primary transformants expressing LuxR fusions in Supplementary Fig. S1.

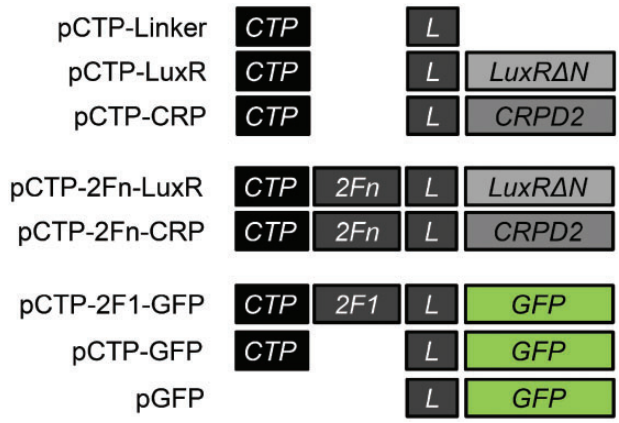

Fig. 2 The expression cassettes that were generated for chloroplast genome interrogation in this study. Control constructs (pCTP-Linker, pCTP-LuxR and pCTP-CRP) encode fusions of effectors (*LuxRΔN* or *CRPD2*) without DNA-binding domains to an N-terminal linker (*L*) and an N-terminal chloroplast transit peptide (*CTP*). Chloroplast genome interrogation constructs (pCTP-2Fn-LuxR and pCTP-2Fn-CRP) encode fusions consisting of eight different and randomly selected 2Fs (2F*n*; $n = 1–8$), respectively, to either LuxRΔN or CRPD2, spaced by the linker sequence and also with an N-terminal CTP. Constructs for confocal microscopy encode either a fusion of GFP to 2F1 spaced by the linker and with an N-terminal CTP (pCTP-2F1-GFP), GFP without a DNA-binding domain and with a CTP (pCTP-GFP) or GFP without a DNA-binding domain and without a CTP (pGFP). Expression of all fusion constructs is under control of the promoter of *AtFEDA* (*pFEDA*) and the *NOS* terminator (*tNOS*).

Just as observed for wild-type Col-0 plants (**Fig. 5**), primary transformants harboring control constructs pCTP-Linker and pCTP-CRP displayed strong clustering of RSA and φPSII data (**Fig. 5**). Surprisingly, despite the data points clustering strongly together, we noted that in the case of pCTP-Linker, but not pCTP-CRP, there seemed to be an enhancement of φPSII compared with Col-0 (**Fig. 5**). Indeed, 50% (14/28) of φPSII data points of the primary transformants harboring pCTP-Linker compared with approximately 3.5% (1/28) of the transformants harboring pCTP-CRP were outside of the 0.05 and 0.95 confidence intervals of Col-0. These observations showed that, for unknown reasons, translocation of the linker peptide outside of the context of a fusion could affect φPSII of Arabidopsis plants. This notion is corroborated further below.

Clustering was very much distorted when primary transformants harbored pCTP-2Fn-CRP constructs compared with pCTP-CRP (**Fig. 5**), demonstrating that an additional presence of 2F domains within the translocated 2F–CRP fusions indeed triggered variation in plant performance, probably due to an interaction of the 2Fs with chloroplast DNA. Statistical analysis at 35 d after germination showed that loss of clustering due to the presence of the 2F domains 2F3, 2F4, 2F6, 2F7 and 2F8 was primarily due to an increase in the variation in φPSII, leading to lower average values. There was only a small variation in RSA values (**Fig 5**), with only expression of 2F7–CRP triggering a

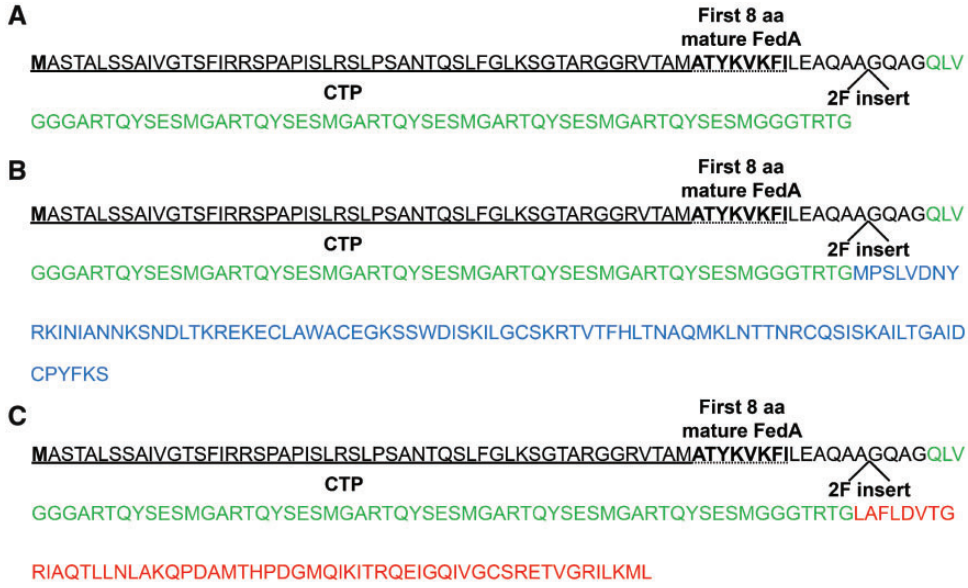

Fig. 3 The amino acid sequences encoded by the ORFs of pCTP-Linker (A), pCTP-LuxR (B) and pCTP-CRP (C). An overview of the composition of these constructs is presented in **Fig. 2**. The linker, LuxRΔN and CRPD2 peptides are presented in green, blue and red font, respectively. The insertion sites of the 2Fs are indicated by '2F insert'.

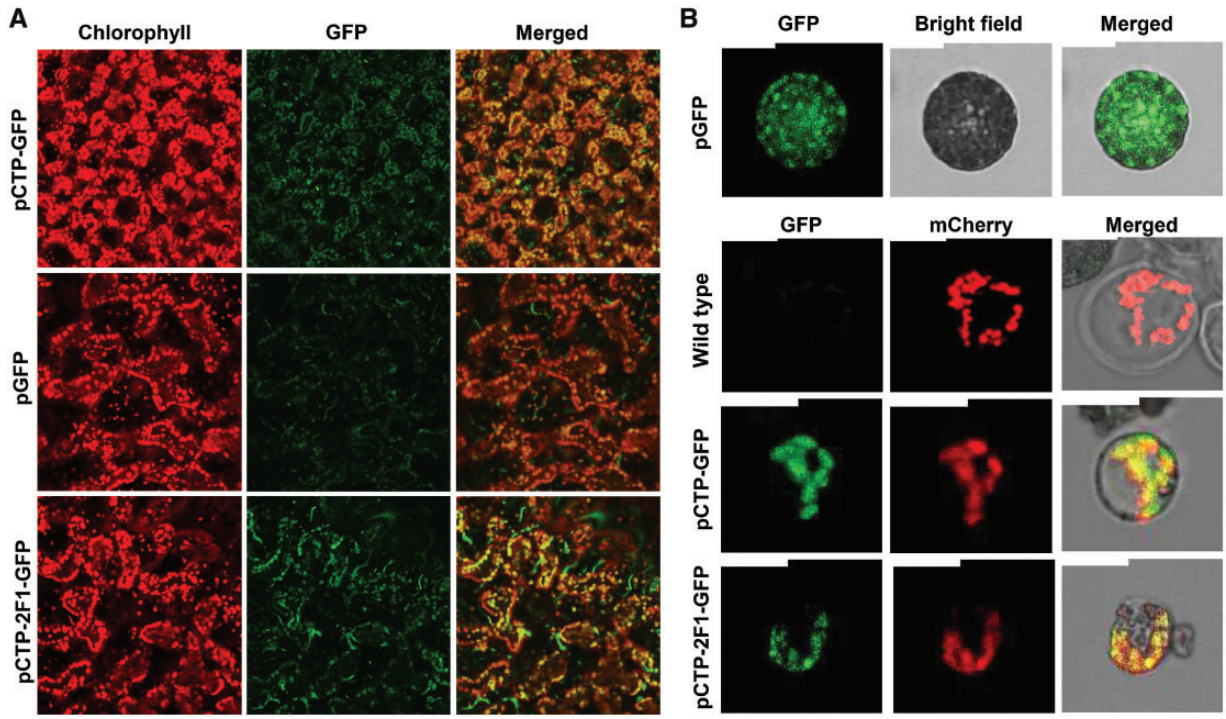

Fig. 4 (A) Confocal microscopy images of primary transformants harboring constructs pCTP-GFP, pGFP and pCTP-2F1-GFP at 35 d after germination (abaxial sides of leaf 5 or 6 at ×20 magnification). 'Chlorophyll' images represent Chl autofluorescence collected with a 560 nm filter. Merged images represent an overlay of the 'Chlorophyll' and GFP channels. (B) Confocal microscopy images of leaf protoplasts of kanamycin-selected T_2 plants (13 d after germination) harboring the constructs pGFP, pCTP-GFP and pCTP-2F1-GFP, respectively, and of the wild-type Col-0 (not kanamycin selected), transfected with the plastidial marker plasmid plastid-mCherry to visualize chloroplasts by fluorescence microscopy. Representative images for three independent lines per construct are shown. Scale bars represent 15 μm.

significant decrease in RSA (**Fig. 5**). Clustering of combined RSA and φPSII data was also observed for pCTP-LuxR-expressing plants, but in this case addition of 2Fs as DNA-binding domains did not significantly alter the mean values (Supplementary Fig. S1), suggesting that translocation of 2F–LuxR fusions only had a minor impact on chloroplast performance. As also discussed below, the LuxRΔN protein could still possess residual DNA binding activity (Choi and Greenberg 1991) that might prohibit 2F-mediated DNA binding, thus, in combination with our observations, making LuxRΔN fusion proteins likely to be less favorable tools for chloroplast genome interrogation. We therefore considered CRP-containing constructs most suitable to investigate 2F-ATF expression further as a tool for chloroplast genome interrogation.

The effect of 2F–CRP-induced variation in growth and photosynthetic performance is stably inherited

To investigate whether effects of 2F–CRP-encoding constructs on growth and photosynthetic performance of primary transformants were stably inherited into the next generation (T_2), we chose to analyze the progeny of four or five primary transformants obtained after transformation with each of the different 2F–CRP-encoding constructs. To address the question of stable 2F–CRP-induced inheritance most effectively, we analyzed the progeny of the primary transformants for which RSA and φPSII values were outside of the 95% confidence interval of the values obtained for primary transformants harboring the control construct pCTP-CRP. In this way, we could most clearly attribute any effects to the additional presence of the DNA-binding 2F domains. We thus generated two-dimensional φPSII and RSA plots for kanamycin-resistant T_2 plants of the independent lines (**Fig. 6**) in a similar fashion as was done for the primary transformants. Kanamycin selection ensured that only transgenic plants expressing the transgene (heterozygous or homozygous) were included in the data analysis, and excluded the wild-type segregant. While there was again a clear clustering of φPSII and RSA values found for T_2 plants of independent lines harboring pCTP-CRP, this clustering was largely lost for the independent lines harboring constructs encoding CRP fusions with 2Fs as DNA-binding domains (**Fig. 6**). These observations therefore indicated that expression of 2F–CRP proteins induced variation in φPSII and RSA values. This was again corroborated with a statistical analysis, showing that in the cases of the constructs pCTP-2F3-CRP, pCTP-2F4-CRP and pCTP-2F8-CRP, all independent lines were significantly different from control lines harboring pCTP-CRP in terms of φPSII (**Table 2**). In the cases of pCTP-2F6-CRP and pCTP-2F7-CRP, only three out of five and one out of four lines, respectively, showed significant differences from pCTP-CRP (**Table 2**), possibly indicating that the originally detected changes in φPSII values observed for these constructs at the primary transformant stage were partially due to variation in antibiotic resistance. In terms of RSA values, we only observed consistent

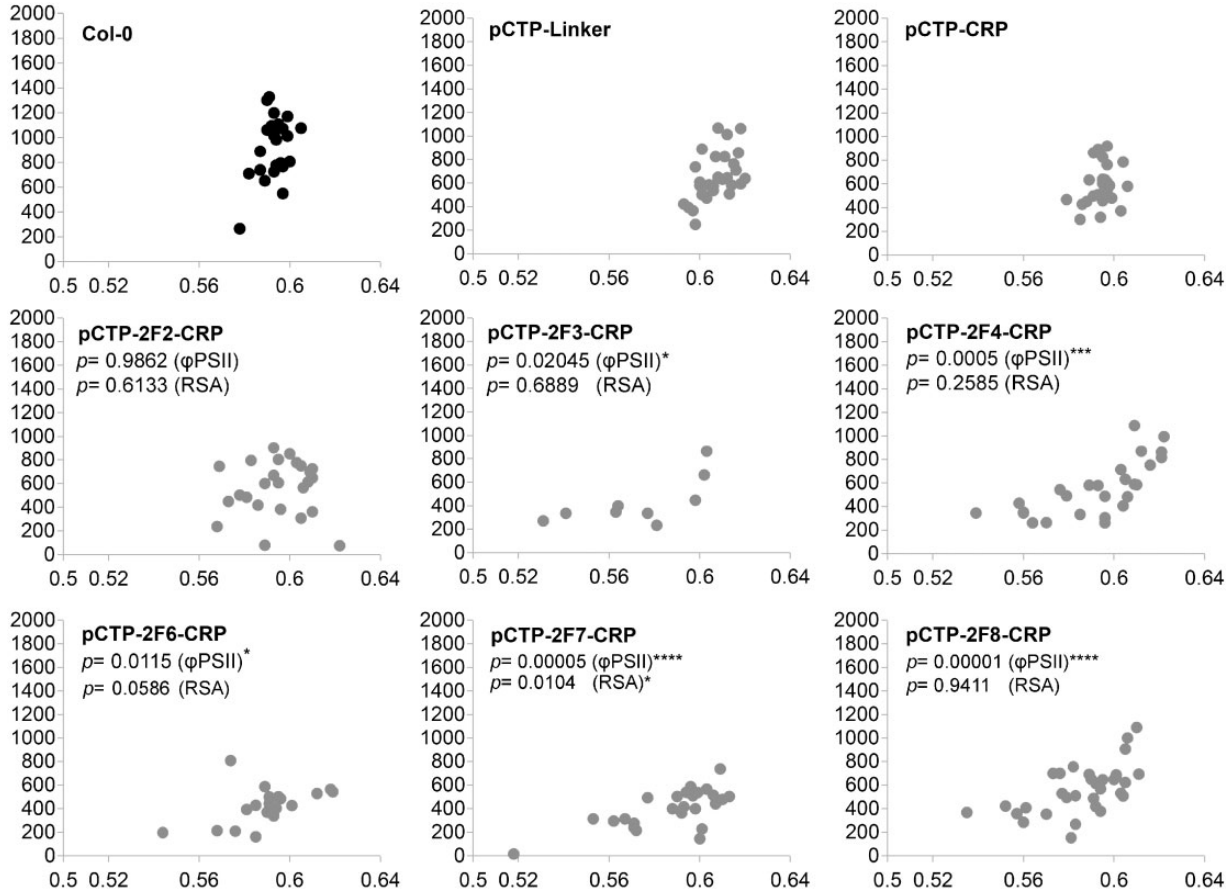

Fig. 5 Analysis of rosette surface area (RSA) and the operating light use efficiency of PSII (φPSII) of wild-type Col-0 plants ($n = 24$ at 28 d after germination; not kanamycin selected) and kanamycin-selected primary transformants (T_1 generation; 28 d after germination) harboring chloroplast genome interrogation constructs encoding fusions of CRPD2 to 2Fs as DNA-binding domains with an N-terminal CTP (pCTP-2Fn-CRP). Primary transformants harboring the constructs pCTP-Linker (importing only the flexible linker peptide) and pCTP-CRP (importing only CRPD2 fused to the flexible linker, without a DNA-binding domain) served as negative controls. Axis labels are not shown for the sake of clarity of the figure. The φPSII values are plotted on the *x*-axes, RSA (mm^2) is plotted on the *y*-axes. Data points represent individual primary transformants. Calculated *P*-values compared with primary transformants harboring pCTP-CRP at 35 d after germination are provided for φPSII and RSA separately, and were calculated compared with the control pCTP-CRP using *t*-tests assuming unequal variance. Asterisks represent the following: $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$. For constructs containing 2F1 and 2F5 only, insufficient numbers of viable primary transformants could be obtained (see the Results).

significant differences compared with pCTP-CRP among the independent lines harboring pCTP-2F4-CRP and pCTP-2F8-CRP, indicating that these constructs also introduced variation in growth in addition to photochemical performance.

As translocation of the unfused linker peptide at the primary transformant stage was observed, for unknown reasons, to affect φPSII, we also assessed the T_2 progeny of independent lines for which at the primary transformant stage such an effect on φPSII was observed. This analysis showed that in two out of four lines this effect was heritable (**Fig. 6**), demonstrating that translocation of the linker peptide could indeed, for unknown reasons and in some cases stably, affect φPSII. Similar to the primary transformant stage, however, independent T_2 lines harboring pCTP-CRP clustered strongly together like Col-0 (**Fig. 6**), corroborating that the presence of the linker by itself does not affect φPSII, and that the effect is only observed when the linker is translocated outside the context of a fusion.

As the independent lines harboring pCTP-2F3-CRP, pCTP-2F4-CRP and pCTP-2F8-CRP displayed different effects on φPSII and RSA (**Fig. 6**), especially so in the case of pCTP-2F8-CRP (**Fig. 6**), we hypothesized that this might be due to differences in nuclear transgene expression between the independent lines. To assess this, we performed quantitative reverse transcription–PCR (RT–qPCR) analysis with primers for an amplicon encoding part of the linker peptide to determine transgene expression levels in kanamycin-selected T_2 plants of these lines, with wild-type Col-0 plants and lines harboring pCTP-CRP as controls. We found that the transgenes were indeed expressed, but also that there was a large amount of variation in nuclear transgene expression levels (Supplementary Fig. S2). We did not, however, find a positive correlation between transgene expression levels and the extent to which or the direction in which φPSII and RSA were affected, indicating that variation in the transgene expression level is not directly causative for changes in these parameters.

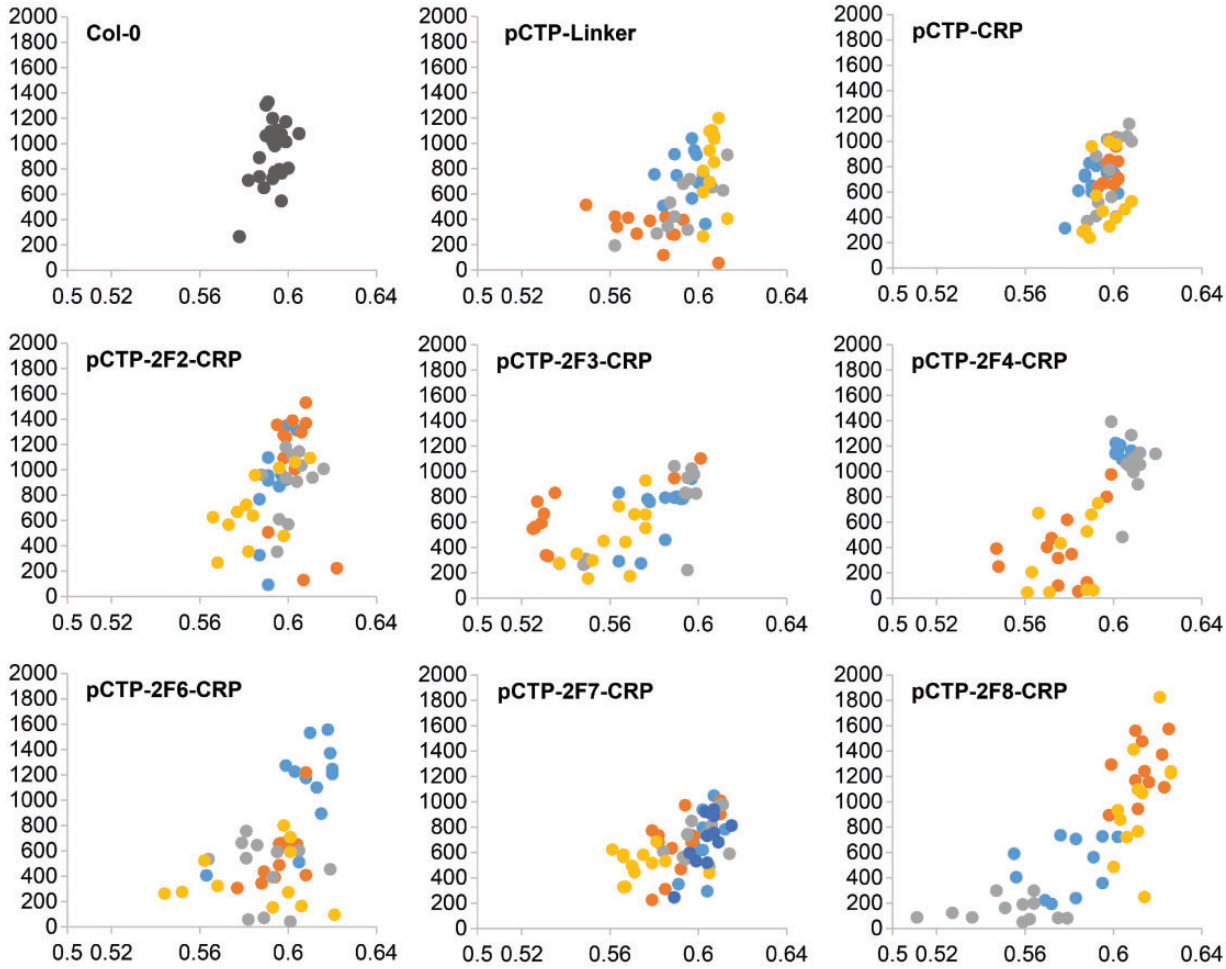

Fig. 6 Analysis of rosette surface area (RSA) and the operating light use efficiency of PSII (φPSII) of kanamycin-selected T_2 plants (28 d after germination) harboring chloroplast genome interrogation constructs encoding fusions of CRPD2 to 2Fs as DNA-binding domains with an N-terminal CTP (pCTP-2F*n*-CRP). Wild-type Col-0 and lines harboring the construct pCTP-CRP (importing only CRPD2 fused to the flexible linker, without a DNA-binding domain) and pCTP-Linker (importing only the flexible linker peptide) are presented as controls. RSA (mm^2) is plotted on the *y*-axes, φPSII is plotted on the *x*-axes. Axis labels are not shown in the figure for the sake of clarity. Data points represent individual plants. Colors represent independent lines.

Chloroplast gene expression analysis of candidate chloroplast genome interrogation lines expressing 2F–CRP fusions

To corroborate that translocated 2F–CRP fusions can activate chloroplast gene expression, we analyzed the expression levels of the chloroplast genes *clpP* [chloroplast-encoded ClpP catalytic subunit of the plastid Clp protease system (Olinares et al. 2011)], *psbA* [encoding the PSII reaction center protein D1 (Shen 2015)] and *rbcL* [encoding the large subunit of RuBisCo (Bracher et al. 2017)]. We selected these genes because they all have well-known and characterized functions in chloroplasts and are important for overall plant development. In addition, these genes are spaced relatively far apart on the chloroplast genome (Sato et al. 1999), thereby excluding possible co-activation of their transcription due to close proximity of the loci. To analyze expression of these genes, we selected independent lines harboring the constructs pCTP-2F3-CRP, pCTP-2F4-CRP and pCTP-2F8-CRP, which displayed significant differences in φPSII compared with the control pCTP-CRP both at the primary transformant stage and in the T_2 generation (**Figs. 5**, **6**; **Table 1**). Independent lines harboring the constructs pCTP-Linker and pCTP-CRP were included as controls, along with the wild-type Col-0. Expression data of *clpP*, *psbA* and *rbcL* determined by RT–qPCR are provided in **Fig. 7**. Analysis of *clpP* expression of kanamycin-resistant T_2 plants showed that lines harboring pCTP-2F3-CRP, but not pCTP-2F4-CRP and pCTP-2F8-CRP, had significantly up-regulated *clpP* expression (~32%) compared with the wild-type Col-0, which was not the case for control plants harboring the constructs pCTP-Linker or pCTP-CRP (**Fig. 7**). Together, these observations demonstrated that 2F3–CRP fusions can activate *clpP* expression in chloroplasts in a 2F3-dependent manner. Analysis of *psbA* expression showed that lines harboring pCTP-2F4-CRP, but in this case not pCTP-2F3-CRP and pCTP-2F8-CRP, had significantly up-regulated *psbA* expression (~60%) compared with Col-0 (**Fig. 7**). However, the control lines harboring pCTP-CRP in this case also displayed a small but nonetheless

Table 2 Statistical analysis of operating light use efficiency of PSII (φPSII) and rosette surface area (RSA) values of kanamycin-selected T_2 progeny plants from selected primary transformants harboring 2F–CRP constructs

		φPSII		RSA	
Construct name	**Line**	***P*-value**	**Significance**	***P*-value**	**Significance**
pCTP-2F2-CRP	1	0.0226	*	0.9118	NS
	8	0.0269	*	0.0242	*
	16	0.0130	*	0.0331	*
	23	0.2882	NS	0.1656	NS
pCTP-2F3-CRP	4	1.2×10^{-6}	****	0.0104	*
	6	0.00005	****	0.6451	NS
	8	0.0346	*	0.6825	NS
	9	0.0012	**	0.9645	NS
pCTP-2F4-CRP	1	1.4×10^{-7}	****	0.0001	***
	2	0.0001	***	2.2×10^{-16}	****
	9	0.0018	**	0.0016	**
	17	0.0014	**	0.0047	**
pCTP-2F6-CRP	4	0.2178	NS	0.0012	**
	5	0.9279	NS	0.1357	NS
	8	0.1564	NS	0.0070	**
	19	0.0265	*	0.0016	**
pCTP-2F7-CRP	7	0.1803	NS	0.806	NS
	9	0.0022	**	0.8268	NS
	18	0.1140	NS	0.6807	NS
	20	4.7×10^{-5}	****	0.0003	***
	22	0.0102	*	0.8492	NS
pCTP-2F8-CRP	2	5.2×10^{-5}	****	2.5×10^{-6}	****
	3	0.0065	**	0.0167	*
	9	9.4×10^{-6}	****	2.2×10^{-16}	****
	14	3.9×10^{-5}	****	0.0373	*

The *P*-values for statistically significant differences ($P < 0.05$) compared with the control pCTP-CRP are presented, determined by *t*-tests assuming unequal variance between samples ($n = 11$ or 12 per individual line)
NS, not significant; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$.

statistically significant up-regulation of *psbA* expression (∼11%; **Fig. 7**), indicating that the *psbA* gene can be activated to a small extent by CRP lacking a DNA-binding domain. This effect was much stronger in the case of lines harboring pCTP-2F4-CRP, suggesting that transcriptional activation of *psbA* by CRP is significantly enhanced in a 2F4-dependent manner. Finally, expression analysis of *rbcL* showed that lines harboring pCTP-2F8-CRP, but not pCTP-2F3-CRP and pCTP-2F4-CRP, had significantly reduced *rbcL* expression (∼19%) compared with Col-0 (**Fig. 7**), which, however, was also the case for the control lines harboring pCTP-CRP, and much more so for pCTP-Linker (∼58%; **Fig. 7**). As the pCTP-CRP construct also included the artificial linker peptide sequence, these results indicated that for unknown reasons the mere presence of the linker sequence inhibited *rbcL* expression. This effect was alleviated by fusion to CRP, and even more so by fusion to the 2F3 and 2F4 domains, as the pCTP-2F3-CRP and pCTP-2F4-CRP constructs not longer led to significant differences in *rbcL* expression compared with the levels found in Col-0 plants (**Fig. 7**).

Discussion

In this study, we have described the design of a novel system for chloroplast genome interrogation in Arabidopsis aimed to introduce changes in growth and photosynthetic properties of plants. This system was tested using two types of chloroplast-targeted ZF-ATFs consisting of fusions of the bacterial transcriptional activators CRPD2 or LuxRΔN to arrays of 2Fs as DNA-binding domains. Using only a small number of different 2Fs, we already found evidence that ZF-ATFs can induce variation in the growth of Arabidopsis plants, can modulate their φPSII and can activate chloroplast gene expression.

Modifying the composition of the chloroplast genome of plants can be regarded as an option for manipulating plant photosynthesis (Ort et al. 2015). Despite the fact that there are at present efficient protocols available for plastid transformation of Arabidopsis (Yu et al. 2017), this is not the case for all model plant species and species of commercial interest. In addition, generating homoplasmic plant lines can be a challenging procedure (Day and Goldschmidt-Clermont 2011). In contrast,

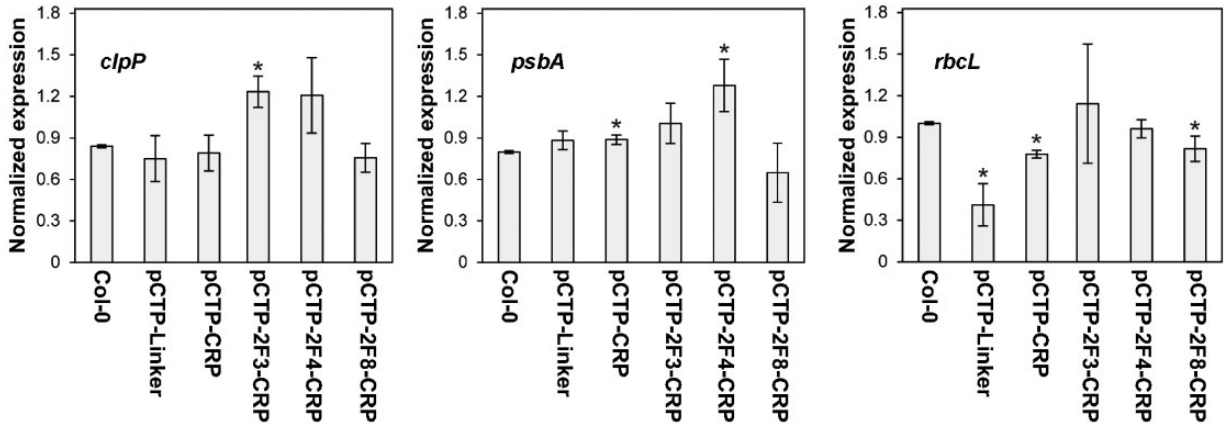

Fig. 7 Expression levels of the chloroplast genes *clpP*, *psbA* and *rbcL* in wild-type Col-0 plants and in kanamycin-selected T_2 plants harboring the control constructs pCTP-Linker and pCTP-CRP or the chloroplast genome interrogation constructs encoding fusions of 2F3, 2F4 or 2F8 to CRP. Expression was normalized to the expression level of the nuclear reference gene *ATG6*. Error bars represent the SD of 3–4 lines ($n = 3–4$). Asterisks (*) represent significant differences from Col-0 ($P < 0.05$).

the chloroplast genome interrogation system that we have investigated in this study should simply allow for *in trans* manipulation of gene expression patterns in all chloroplasts and all copies of the chloroplast genome by just integrating a single artificial gene in the nuclear genome via standard methods. Since nuclear transformation protocols have become available for most plant species, the chloroplast genome interrogation system could thus in principle readily be applied to commercially interesting plant species without the requirement for detailed a priori knowledge regarding their plastid biology and the availability of a plastid transformation protocol.

We have explored the use of the bacterial transcriptional activators CRP (Lee et al. 2008) and LuxR (Choi and Greenberg 1991) as effector domains of the ATFs to achieve chloroplast genome interrogation. It has to be noted that published data regarding prokaryotic transcriptional regulators that directly affect RNA polymerase activity without requiring other regulatory proteins are extremely scarce; to our knowledge, as described in more detail above, the bacterial transcriptional activators CRP and LuxR are the only protein domains for which data have been published that support these requirements. Still, it has to be taken into account that the LuxR protein contains residual DNA binding activity (Choi and Greenberg 1991). Compared with chloroplast-targeted transcriptional regulators containing the CRP domain, which were shown in this study to induce changes in φPSII and RSA (**Figs. 5**, **6**), constructs encoding 2F–LuxR fusion proteins were much less effective (Supplementary Fig. S1). As already indicated above, the residual DNA binding activity of the LuxR domain might obstruct the intended use of 2F domains as variable 6 bp targeting DNA-binding moieties. Theoretically, this would mean that higher affinity LuxR-mediated DNA binding would direct the 2F–LuxR fusions away from their cognate 2F target sites, thereby abrogating the basic principle of genome interrogation. Given the cyanobacterial origin of chloroplasts, cyanobacterial transcriptional activators might be explored as candidate effector domains in future studies. In addition, bacteriophage or cyanophage transcription factors acting on sigma factor-mediated transcription could also be useful candidates (e.g. Liu et al. 2014).

In the present study, we have designed a novel system for chloroplast genome interrogation in Arabidopsis with CRP-based transcriptional regulators. Surprisingly, this was already achieved by using only eight randomly selected 2Fs out of the 256 possible 2F combinations of ZFs with consensus GNN recognition sites, thereby only scratching the surface of the possibilities of 2F-mediated chloroplast genome interrogation. The low-complexity and low-affinity 2F domains can thus be instrumental in evoking different responses to artificial chloroplast-targeted transcriptional regulators at the phenotypic (**Figs. 5**, **6**) and transcriptional levels (**Fig. 7**). This corroborates previous observations for 2F-based nuclear-targeted transcription factors found to induce salinity tolerance in Arabidopsis (van Tol et al. 2016).

The goal for employing chloroplast genome interrogation was to enhance photosynthetic performance and growth of plants. However, at the primary transformant stage, we observed that the overall performance of the plants was negatively impacted by translocation of 2F–CRP fusions (**Fig. 5**), suggesting that artificial transcriptional modulation of chloroplast genes disrupts rather than enhances photochemical activity and growth. This suggests that interference with the normal regulation of photochemical activity generally leads to loss of photochemical capacity. Although mechanistically not comparable, this is in line with the fact that mutations in thylakoid membrane proteins typically negatively impact thylakoid membrane function (e.g. Meurer et al. 1996, Varotto et al. 2002, Ihnatowicz et al. 2004). In the T_2 generation, however, we also identified 2F–CRP-expressing lines which outperformed the control lines expressing only CRP (**Fig. 6**), which suggests that variations in levels of nuclear transgene expression—often related to the different positions of the loci where T-DNA integration occurred—might greatly influence the extent to which and direction in which photosynthetic performance is

Table 3 Primers that were used for the construction of the library of chloroplast genome interrogation binary vector constructs

Name	5′–3′ DNA sequence (restriction site underlined)
pFEDA FW	GGTCGACTGCCTTTTACGGAAAGATTCGATTTGG (*Sal*I)
pFEDA REV1	CCTCTCGAGGATGAACTTGACCTTGTATGTAGC (*Xho*I)
pFEDA REV2	GGAGCTCAGGCCTCTCGAGGATGAACTTGACCTTG (*Sac*I)
GFP FW	GACTAGTGTGAGCAAGGGCGAGGAGCTGTTCACCG (*Spe*I)
GFP RV	GGAGCTCTTACTTGTACAGCTCGTCCATGCCG (*Sac*I)
MASTAL REV	GCTCGAGAGCAGTGGAAGCCATTTTTTTTTG (*Xho*I)

Table 4 NCBI accession numbers of constructs generated in this study

Sequence name	Sequence label	NCBI accession number
pFEDA	Seq1	MK078636
tNOS	Seq2	MK078637
CDS 2F1	Seq3	MK078638
CDS 2F2	Seq4	MK078639
CDS 2F3	Seq5	MK078640
CDS 2F4	Seq6	MK078641
CDS 2F5	Seq7	MK078642
CDS 2F6	Seq8	MK078643
CDS 2F7	Seq9	MK078644
CDS 2F8	Seq10	MK078645
CDS pCTP-Linker	Seq11	MK078646
CDS pCTP-CRP	Seq12	MK078647
CDS pCTP-LuxR	Seq13	MK078648
CDS pCTP-2F1-CRP	Seq14	MK078649
CDS pCTP-2F1-LuxR	Seq15	MK078650
CDS pCTP-GFP	Seq16	MK078651
CDS pCTP-2F1-GFP	Seq17	MK078652
CDS pGFP	Seq18	MK078653

affected. We indeed found that different plant lines that contained the same 2F–CRP-encoding construct varied greatly in photochemical performance and growth (**Fig. 6**), and exhibited a large variation in the levels of nuclear transgene expression (Supplementary Fig. S2), possibly suggesting that nuclear transgene expression levels might affect the extent of variation. For future chloroplast genome interrogation studies, nuclear 2F–CRP-encoding transgenes with inducible promoters could also be considered to have more control over the timing and strength of the expression and subsequent translocation.

In conclusion, just guided by basic knowledge of chloroplast biology, we have designed a novel and prospective system for chloroplast genome interrogation that can be used without requiring any further a priori knowledge of the chloroplast genome. Using a relatively limited experimental set-up, we have already found evidence that ZF-ATF-mediated chloroplast genome interrogation can induce significant changes in the photosynthetic performance of chloroplasts. Taken together, our work suggests that it would be worthwhile to investigate chloroplast genome interrogation further as a potential tool to enhance the photosynthetic performance of plants.

Materials and Methods

Plant material and growth conditions

The Arabidopsis accession Col-0 was used as the wild type and as the background genotype for all transformations described below. All seeds were stratified for 3–4 d at 4°C prior to the start of the experiments. Soil-grown seedlings and plants were cultivated in a climate-controlled growth chamber at a constant temperature of 20°C, 70% relative humidity, a light intensity of approximately 200 μmol m^{-2} s^{-1} of photosynthetically active radiation (PAR) and a 12 h photoperiod (referred to as 'standard growth conditions'). Primary transformants were first grown on MA medium (Masson and Paskowski, 1992) containing 35 μg ml^{-1} kanamycin, in a climate-controlled tissue culture chamber at a constant temperature of 20°C, 50% relative humidity, a PAR light intensity of approximately 50 μmol m^{-2} s^{-1} and a 16 h photoperiod, subsequently transferred to soil after approximately 10 d, and were further cultivated under standard growth conditions. Kanamycin-selected T_2 plants received the same treatment as primary transformants, except that they were grown on a different selection medium (half-strength MS medium containing 50 μg ml^{-1} kanamycin). Col-0 plants also received this treatment, but were grown on medium (half-strength MS) without kanamycin.

Construction of Arabidopsis plant lines expressing ZF-ATFs

A library of plasmids containing DNA fragments encoding all 256 different 2Fs was previously constructed (Neuteboom et al. 2006). Eight different 2Fs each consisting of two different ZFs were randomly selected from this library (**Table 1**). The DNA sequence of CRP was derived from the NCBI (REFSEQ accession NC_000913.2). The amino acid sequence of CRPD2 was derived from Lee et al. (2008). The amino acid residues 134–136 (NLA) were also included, as these were reported to constitute a flexible hinge (Baker et al. 2001). The DNA sequence of LuxR was derived from the NCBI (accession M25752, version 1). The amino acid sequences of LuxRΔN and the flexible linker were derived from Volzing et al. (2011). The promoter of *AtFEDA* (At1g60950; including the 5′-untranslated region sequence, the sequence encoding the CTP and the sequence encoding the first eight amino acids of the mature FedA peptide) was amplified by PCR from the genomic DNA of Col-0 using the forward primer pFEDA FW (**Table 3**) and the reverse primers pFEDA REV1 and pFEDA REV2 (**Table 3**), yielding a 2,029 bp *pFEDA* fragment. The sequences of all DNA fragments obtained by the insertion of oligonucleotides or PCR products were verified by Sanger sequencing (Macrogen Europe). The binary vector pRF (Lindhout et al. 2006) was used as the backbone for all cloning steps described below.

The *RPS5A* promoter sequence and the 3F–VP16-encoding open reading frame (ORF) were removed from the plasmid pRF (Lindhout et al. 2006), and the *FEDA* promoter fragment was subsequently ligated in, yielding the plasmid pFEDA. A 700 bp oligo DNA fragment (Supplementary Fig. S3) encoding the flexible linker, LuxRΔN and CRPD2 (both codon optimized for Arabidopsis) was synthesized by the company ShineGene, and was ligated into pFEDA. Either one or both of the two effector-encoding modules were subsequently removed, yielding the plasmids pCTP-CRP, pCTP-LuxR and pCTP-Linker, respectively. The eight randomly selected 2F fragments were each ligated into pCTP-CRP and pCTP-LuxR as *Sfi*I fragments, yielding the plasmids which were designated pCTP-2Fn-CRP and pCTP-2Fn-LuxR, respectively. The DNA sequence encoding enhanced GFP (eGFP) was amplified by PCR using the forward primer GFP FW and reverse primer GFP RV (**Table 3**), and ligated into pCTP-Linker, yielding the plasmid pCTP-GFP, and into pCTP-2F1-Linker, yielding the plasmid which was named pCTP-2F1-GFP. Using pFEDA as a template, a PCR product of the *FEDA* promoter lacking the CTP was generated using the primer combination pFEDA FW and MASTAL REV (**Table 2**), and was ligated into *Sal*I- and *Xho*I-digested pCTP-Linker, yielding the plasmid which was named pLinker. The PCR fragment of eGFP was subsequently also ligated into pLinker, yielding the plasmid pGFP.

Table 5 Primers that were used for RT–qPCR analysis of chloroplast gene expression

Name	5'–3' oligo DNA sequence	Reference
clpP FW	TATGCAATTTGTGCGACCC	Weihe (2014)
clpP REV	TTGGTAATTGCTCCTCCGACT	Weihe (2014)
psbA FW	ATACAACGGCGGTCCTTATG	Tadini et al. (2012)
psbA REV	CGGCCAAAATAACCGTGAGC	Tadini et al. (2012)
rbcL FW	CGTTGGAGAGACCGTTTCTT	Tadini et al. (2012)
rbcL REV	CAAAGCCCAAAGTTGACTCC	Tadini et al. (2012)
ATG6 FW	AGACACAGGTTGAACAGCCA	van Tol et al. (2017a, 2017b)
ATG6 REV	GTATGCTTCCACGTCCCTCG	van Tol et al. (2017a, 2017b)
Linker FW	GGTGGTGGTGCTAGGACA	This study
Linker REV	CCCCTCCCATTGACTCACTA	This study

Col-0 plants were transformed with each of the generated constructs separately using the floral dip method (Clough and Bent 1998). The DNA sequences of the constructs listed in **Fig. 2** have been deposited in an NCBI BankIt database. Database accession numbers are listed in Table 4. For the sequences encoding 2F–CRP or 2F–LuxR fusions (pCTP-2Fn-CRP and pCTP-2Fn-LuxR), we have provided pCTP-2F1-CRP and pCTP-2F1-LuxR as samples. Other in-frame 2F fusions (2F2–2F8) can be constructed by simply removing the coding sequence of 2F1 and replacing it by the coding sequences of the other 2Fs, which have also been deposited there.

Confocal microscopy

Five to six primary transformants harboring the constructs pGFP, pCTP-GFP and pCTP-2F1-GFP, respectively, were randomly selected from a population of primary transformants using Research Randomizer (http://www.randomizer.org) at 34–35 d after germination for confocal microscopy. For each randomly selected individual, five confocal images were taken from the abaxial side of either leaf 5 or 6 (Telfer et al. 1997) using a Zeiss LSM5 exciter with Illuminator HXP 120 V. Excitation of the tissue was performed at a wavelength of 488 nm. Chl fluorescence emission was collected with a 560 nm long pass filter and GFP fluorescence with a 505–530 nm band pass filter. The settings were first calibrated for the primary transformants harboring pCTP-GFP. The rest of the transformants were subsequently imaged using the same settings and at the same laser power. The images most representative for the variation among the individuals were selected manually and are presented in **Fig. 4A**.

Leaf protoplast transfection

Leaves of kanamycin-resistant T_2 plants harboring pGFP, pCTP-GFP and pCTP-2F1-GFP (three independent lines each), and leaves of Col-0 plants (not kanamycin selected) were harvested at 13 d after germination and protoplasted as described previously (Doelling and Pikaard 1993, Doelling et al. 1993). Leaf protoplasts were subsequently transfected with the plastidial marker plasmid plastid-mCherry (CD3-1000) (Nelson et al. 2007) and examined for co-localization of plastid-mCherry and GFP signal by confocal microscopy as described above. Representative images for three independent lines are shown in **Fig. 4B**.

Quantification of rosette surface area

From 21 d after germination onwards and every 3–4 d, photos were taken of primary transformants and of kanamycin-selected T_2 progeny from candidate primary transformants using a fixed digital camera set-up (Canon EOS 1100D). These RGB images were converted to binary images, and projected RSA was subsequently calculated in square millimeters as described previously (van Tol et al. 2017a). The distribution of the data was determined with quantile–quantile plots and Shapiro–Wilk normality tests (Ghasemi and Zahediasl 2012), both using RStudio. These tests showed that the data points were collected from a normally distributed population with unequal variance. The data were then statistically analyzed with *t*-tests assuming unequal variance compared with the respective control constructs, with $P < 0.05$ as a threshold for significance.

Chl fluorescence imaging

Operating light use efficiency of PII (φPSII; F_q'/F_m') (Baker 2008) was measured using a Chl fluorescence imager (Technologica Ltd.). The plants were placed in the imaging chamber and exposed to 200 μmol m^{-2} s^{-1} actinic light (the same light intensity as in the growth chamber) for 30 s, after which a saturating pulse of 6,626 μmol m^{-2} s^{-1} actinic light (maximal intensity) was given for 800 ms. Background fluorescence was manually removed from the resulting F_q'/F_m' images using the FluorImager software package. The F_q'/F_m' data were statistically analyzed as described above for the quantification of RSA.

RT–qPCR analysis

For 3–4 independent plant lines per construct (therefore $n = 3–4$), we collected pools of 10–12 kanamycin-selected T_2 plants (progeny of candidate primary transformants harboring chloroplast genome interrogation constructs outside of the 95% confidence intervals for controls, or harboring control constructs without 2Fs) which were ground to powder in liquid nitrogen with pestles and mortars at 12 d after germination. Col-0 plants (not kanamycin selected) were harvested to assess wild-type expression. Total RNA was extracted from 50–100 mg of tissue powder of each individual using the RNeasy Plant Mini Kit (QIAGEN). To remove residual traces of DNA, a DNase treatment was performed with DNase I (Ambion), followed by a phenol/chloroform (1:1) extraction and ethanol precipitation. First-strand cDNA synthesis was performed with 500 ng of RNA as a template using the Script cDNA Synthesis Kit (Jena Bioscience), and reactions were carried out in a Biometra TAdvanced (Analytik Jena). RT–qPCRs were prepared using the SYBR Green PCR Master Mix (Applied Biosystems). The RT–qPCRs were performed using the CFX96 Touch™ Real-Time PCR Detection System (BIORAD), with previously published primer combinations for chloroplast genes (**Table 5**), and for the reference gene *ATG6* (van Tol et al. 2017a, van Tol et al. 2017b). Relative expression levels were calculated by the ΔΔCt method (Livak and Schmittgen 2001), with *ATG6* as a reference gene. Independent lines were examined as biological replicates. Expression of the nuclear transgene was verified with a primer combination amplifying the linker sequence (*pLinker* FW and REV; **Table 5**), which was designed using the Primer3Web tool (http://primer3.ut.ee). At the end of the PCRs, a melt curve determination was performed to check for single amplification products for each of the primer combinations.

Supplementary Data

Supplementary data are available at PCP online.

Funding

This work was supported by the research program of BioSolar Cells, co-financed by the Dutch Ministry of Economic Affairs [grant No. 10TBSC23].

Acknowledgments

We would like to thank Paul de Mooij for his help in the initial stages of the experimental design of this study, and Johan Pinas, Nick Surtel and Vera Veltkamp for technical assistance and help with carrying out the experiments.

Disclosures

The authors have no conflicts of interest to declare.

References

Baker, C.H., Tomlinson, S.R., Garcia, A.E. and Harman, J.G. (2001) Amino acid substitution at position 99 affects the rate of CRP subunit exchange. *Biochemistry* 40: 12329–12338.

Baker, N.R. (2008) Chlorophyll fluorescence: a probe of photosynthesis in vivo. *Annu. Rev. Plant Biol.* 59: 89–113.

Beltran, A., Liu, Y.Z., Parikh, S., Temple, B. and Blancafort, P. (2006) Interrogating genomes with combinatorial artificial transcription factor libraries: asking zinc finger questions. *Assay Drug Dev. Technol.* 4: 317–331.

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

Borner, T., Aleynikova, A.Y., Zubo, Y.O. and Kusnetsov, V.V. (2015) Chloroplast RNA polymerases: role in chloroplast biogenesis. *Biochim. Biophys. Acta* 1847: 761–769.

Borukhov, S. and Lee, J. (2005) RNA polymerase structure and function at lac operon. *C.R. Biol.* 328: 576–587.

Bracher, A., Whitney, S.M., Hartl, F.U. and Hayer-Hartl, M. (2017) Biogenesis and metabolic maintenance of Rubisco. *Annu. Rev. Plant Biol.* 68: 29–60.

Bruce, B.D. (2000) Chloroplast transit peptides: structure, function and evolution. *Trends Cell Biol.* 10: 440–447.

Choi, S.H. and Greenberg, E.P. (1991) The C-terminal region of the *Vibrio fischeri* LuxR protein contains an inducer-independent *lux* gene activating domain. *Proc. Natl. Acad. Sci. USA* 88: 11115–11119.

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

Day, A. and Goldschmidt-Clermont, M. (2011) The chloroplast transformation toolbox: selectable markers and marker removal. *Plant Biotechnol. J.* 9: 540–553.

Doelling, J.H., Gaudino, R.J. and Pikaard, C.S. (1993) Functional analysis of *Arabidopsis thaliana* rRNA gene and spacer promoters in vivo and by transient expression. *Proc. Natl. Acad. Sci. USA* 90: 7528–7532.

Doelling, J.H. and Pikaard, C.S. (1993) Transient expression in *Arabidopsis thaliana* protoplasts derived from rapidly established cell suspension cultures. *Plant Cell Rep.* 12: 241–244.

Dunlap, P. (2014) Biochemistry and genetics of bacterial bioluminescence. *Adv. Biochem. Eng. Biotechnol.* 144: 37–64.

Flores-Perez, U. and Jarvis, P. (2013) Molecular chaperone involvement in chloroplast protein import. *Biochim. Biophys. Acta* 1833: 332–340.

Ghasemi, A. and Zahediasl, S. (2012) Normality tests for statistical analysis: a guide for non-statisticians. *Int. J. Endocrinol. Metab.* 10: 486–489.

Hiratsu, K., Matsui, K., Koyama, T. and Ohme-Takagi, M. (2003) Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in Arabidopsis. *Plant J.* 34: 733–739.

Ihnatowicz, A., Pesaresi, P., Varotto, C., Richly, E., Schneider, A., Jahns, P., et al. (2004) Mutants for photosystem I subunit D of *Arabidopsis thaliana*: effects on photosynthesis, photosystem I stability and expression of nuclear genes for chloroplast functions. *Plant J.* 37: 839–852.

Jarvis, P. and Lopez-Juez, E. (2013) Biogenesis and homeostasis of chloroplasts and other plastids. *Nat. Rev. Mol. Cell Biol.* 14: 787–802.

Jia, Q., van Verk, M.C., Pinas, J.E., Lindhout, B.I., Hooykaas, P.J.J. and van der Zaal, B.J. (2013) Zinc finger artificial transcription factor-based nearest inactive analogue/nearest active analogue strategy used for the identification of plant genes controlling homologous recombination. *Plant Biotechnol. J.* 11: 1069–1079.

Jin, R., Richter, S., Zhong, R. and Lamppa, G.K. (2003) Expression and import of an active cellulase from a thermophilic bacterium into the chloroplast both in vitro and in vivo. *Plant Mol Biol.* 51: 493–507.

Lee, J.Y., Sung, B.H., Yu, B.J., Lee, J.H., Lee, S.H., Kim, M.S., et al. (2008) Phenotypic engineering by reprogramming gene transcription using novel artificial transcription factors in *Escherichia coli*. *Nucleic Acids Res.* 36: e102.

Leister, D. (2005) Genomics-based dissection of the cross-talk of chloroplasts with the nucleus and mitochondria in Arabidopsis. *Gene* 354: 110–116.

Leister, D., Varotto, C., Pesaresi, P., Niwergall, A. and Salamini, F. (1999) Large-scale evaluation of plant growth in *Arabidopsis thaliana* by non-invasive image analysis. *Plant Physiol. Biochem.* 37: 671–678.

Lindhout, B.I., Pinas, J.E., Hooykaas, P.J. and van der Zaal, B.J. (2006) Employing libraries of zinc finger artificial transcription factors to screen for homologous recombination mutants in Arabidopsis. *Plant J.* 48: 475–483.

Liu, B., Shadrin, A., Sheppard, C., Mekler, V., Xu, Y., Severinov, K., et al. (2014) A bacteriophage transcription regulator inhibits bacterial transcription initiation by sigma-factor displacement. *Nucleic Acids Res.* 42: 4294–4305.

Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25: 402–408.

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

Masson, J. and Paszkowski, J. (1992) The culture response of Arabidopsis thaliana protoplasts is determined by the growth conditions of donor plants. *Plant J.* 2: 829–833.

Meurer, J., Meierhoff, K. and Westhoff, P. (1996) Isolation of high-chlorophyll-fluorescence mutants of *Arabidopsis thaliana* and their characterisation by spectroscopy, immunoblotting and northern hybridisation. *Planta* 198: 385–396.

Mito, T., Seki, M., Shinozaki, K., Ohme-Takagi, M. and Matsui, K. (2011) Generation of chimeric repressors that confer salt tolerance in Arabidopsis and rice. *Plant Biotechnol. J.* 9: 736–746.

Nelson, B.K., Cai, X. and Nebenfuhr, A. (2007) A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. *Plant J.* 51: 1126–1136.

Neuteboom, L.W., Lindhout, B.I., Saman, I.L., Hooykaas, P.J. and van der Zaal, B.J. (2006) Effects of different zinc finger transcription factors on genomic targets. *Biochem. Biophys. Res. Commun.* 339: 263–270.

Olinares, P.D., Kim, J. and van Wijk, K.J. (2011) The Clp protease system; a central component of the chloroplast protease network. *Biochim. Biophys. Acta* 1807: 999–1011.

Ort, D.R., Merchant, S.S., Alric, J., Barkan, A., Blankenship, R.E., Bock, R., et al. (2015) Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proc. Natl. Acad. Sci. USA* 112: 8529–8536.

Park, K.S., Seol, W., Yang, H.Y., Lee, S.I., Kim, S.K., Kwon, R.J., et al. (2005) Identification and use of zinc finger transcription factors that increase production of recombinant proteins in yeast and mammalian cells. *Biotechnol. Prog.* 21: 664–670.

Sadowski, I., Ma, J., Triezenberg, S. and Ptashne, M. (1988) GAL4-VP16 is an unusually potent transcriptional activator. *Nature* 335: 563–564.

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

Schu, D.J., Carlier, A.L., Jamison, K.P., von Bodman, S. and Stevens, A.M. (2009) Structure/function analysis of the *Pantoea stewartii* quorum-sensing regulator EsaR as an activator of transcription. *J. Bacteriol.* 191: 7402–7409.

Segal, D.J., Dreier, B., Beerli, R.R. and Barbas, C.F., 3rd (1999) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5′-GNN-3′ DNA target sequences. *Proc. Natl. Acad. Sci. USA* 96: 2758–2763.

Shen, J.R. (2015) The structure of photosystem II and the mechanism of water oxidation in photosynthesis. *Annu. Rev. Plant Biol.* 66: 23–48.

Smeekens, S., Geerts, D., Bauerle, C. and Weisbeek, P. (1989) Essential function in chloroplast recognition of the ferredoxin transit peptide processing region. *Mol. Gen. Genet.* 216: 178–182.

Smeekens, S., van Steeg, H. Bauerle, C., Bettenbroek, H., Keegstra, K. and Weisbeek, P. (1987) Import into chloroplasts of a yeast mitochondrial protein directed by ferredoxin and plastocyanin transit peptides. *Plant Mol Biol.* 9:377–388.

Somers, D.E., Caspar, T. and Quail, P.H. (1990) Isolation and characterization of a ferredoxin gene from *Arabidopsis thaliana*. *Plant Physiol.* 93: 572–577.

Stoppel, R. and Meurer, J. (2013) Complex RNA metabolism in the chloroplast: an update on the psbB operon. *Planta* 237: 441–449.

Tadini, L., Romani, I., Pribil, M., Jahns, P., Leister, D. and Pesaresi, P. (2012) Thylakoid redox signals are integrated into organellar-gene-expression-dependent retrograde signaling in the prors1-1 mutant. *Front. Plant Sci.* 3: 282.

Telfer, A., Bollman, K.M. and Poethig, R.S. (1997) Phase change and the regulation of trichome distribution in *Arabidopsis thaliana*. *Development* 124: 645–654.

van Tol, N., Pinas, J., Schat, H., Hooykaas, P.J. and van der Zaal, B.J. (2016) Genome interrogation for novel salinity tolerant Arabidopsis mutants. *Plant Cell Environ.* 39: 2650–2662.

van Tol, N., Rolloos, M., Augustijn, D., Alia, A., de Groot, H.J., Hooykaas, P.J.J., et al. (2017a) An Arabidopsis mutant with high operating efficiency of Photosystem II and low chlorophyll fluorescence. *Sci. Rep.* 7: 3314.

van Tol, N., Rolloos, M., Pinas, J.E., Henkel, C.V., Augustijn, D., Hooykaas, P.J., et al. (2017b) Enhancement of Arabidopsis growth characteristics using genome interrogation with artificial transcription factors. *PLoS One* 12: e0174236.

van Tol, N. and van der Zaal, B.J. (2014) Artificial transcription factor-mediated regulation of gene expression. *Plant Sci.* 225: 58–67.

Varotto, C., Pesaresi, P., Jahns, P., Lessnick, A., Tizzano, M., Schiavon, F., et al. (2002) Single and double knockouts of the genes for photosystem I subunits G, K, and H of Arabidopsis. Effects on photosystem I composition, photosynthetic electron flow, and state transitions. *Plant Physiol.* 129: 616–624.

Volzing, K., Biliouris, K. and Kaznessis, Y.N. (2011) proTeOn and proTeOff, new protein devices that inducibly activate bacterial gene expression. *ACS Chem. Biol.* 6: 1107–1116.

Weihe, A. (2014) Quantification of organellar DNA and RNA using real-time PCR. *Methods Mol. Biol.* 1132: 235–243.

Woodson, J.D. and Chory, J. (2008) Coordination of gene expression between organellar and nuclear genomes. *Nat. Rev. Genet.* 9: 383–395.

Yu, Q., Lutz, K.A. and Maliga, P. (2017) Efficient plastid transformation in Arabidopsis. *Plant Physiol.* 175: 186–193.