



Universiteit
Leiden
The Netherlands

Phenotypic engineering of photosynthesis related traits in *Arabidopsis thaliana* using genome interrogation

Tol, Niels van

Citation

Tol, N. van. (2016, March 24). *Phenotypic engineering of photosynthesis related traits in Arabidopsis thaliana using genome interrogation*. Retrieved from <https://hdl.handle.net/1887/38624>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/38624>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/38624> holds various files of this Leiden University dissertation.

Author: Tol, Niels van

Title: Phenotypic engineering of photosynthesis related traits in *Arabidopsis thaliana* using genome interrogation1

Issue Date: 2016-03-24

**Phenotypic engineering of photosynthesis related traits in
Arabidopsis thaliana using genome interrogation**

Niels van Tol

Cover: The largest wild type Arabidopsis plant that I have grown. It grew for 88 days at short day and low light conditions without flowering and weighed approximately 14 grams upon harvesting.

ISBN: 978-94-6233-247-8

Printed by: Gildeprint, Enschede, The Netherlands

Phenotypic engineering of photosynthesis related traits in *Arabidopsis thaliana* using genome interrogation

Proefschrift

ter verkrijging van de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus Prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties te verdedigen op
donderdag 24 maart 2016 te klokke 13.45 uur

door

Niels van Tol

geboren te Leiden, Nederland
17 september 1988

Promotie commissie

Promotor: Prof. Dr. P.J.J. Hooykaas
Copromotor: Dr. Ir. E.J. van der Zaal
Overige leden: Dr. B.S. de Pater
Dr. E. van der Graaff
Prof. Dr. J. Memelink
Prof. Dr. H.P. Spaink
Prof. Dr. J.J.B. Keurentjes

Aan mijn vader, moeder en zus

Contents

Chapter 1	General introduction	9
Chapter 2	Artificial transcription factor-mediated regulation of gene expression	17
Chapter 3	Enhancement of Arabidopsis growth characteristics using genome interrogation	41
Chapter 4	Chloroplast genome interrogation in Arabidopsis seedlings	77
Chapter 5	An Arabidopsis mutant with a high operating efficiency of Photosystem II and low chlorophyll fluorescence	99
Chapter 6	Artificial transcription factor-induced salinity tolerance in Arabidopsis	127
Chapter 7	Dutch summary/Nederlandse samenvatting	161
Curriculum Vitae		165

Chapter 1

General Introduction

Improving the efficiency of photosynthesis

Niels van Tol

Photosynthesis is the process that harvests energy from light, and fixes it as chemical energy. It is performed by cyanobacteria, algae, and plants. In algae and plants photosynthesis is conducted by chloroplasts, which are highly specialized organelles of cyanobacterial origin. The stroma of chloroplasts harbors photosynthetic lipid membrane units named thylakoids, which are organized in stacks [1], and in which four large protein complexes are anchored: Photosystem II (PSII), Cytochrome b_6f , Photosystem I (PSI), and an ATP synthase complex. The light harvesting antennae that are associated with PSII are able to absorb and channel photons into PSII reaction centers, resulting in the photoexcitation of electrons obtained through the splitting of water. The tertiary electron acceptor to PSII, the cytochrome b_6f complex, functions as a proton pump, resulting in the buildup of a proton motive force across the thylakoid membrane. This proton motive force drives the synthesis of ATP by the ATP synthase complex. After passing the redox carrier protein plastocyanin, electrons are passed to PSI for another photoexcitation enabling the formation of the high energy electron carrier NADPH through the activity of the enzyme ferredoxin-NADP⁺ reductase. Both ATP and NADPH are used by the Calvin-Benson cycle in the chloroplast stroma, which fixes CO₂ into the three carbon molecule 3-phosphoglyceric acid (3-PGA) through the activity of the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo), ultimately resulting in the synthesis of carbohydrates. These carbohydrates are commonly stored in chloroplasts as starch during the day, and subsequently partitioned to different pathways and organs of the plant during the night [2].

The overall solar energy to biomass conversion efficiency of plant photosynthesis is widely considered to be very low. However, this does not necessarily imply that photosynthesis is an inefficient process for plant productivity. Presumptuous statements in that respect are often fueled by the thought that the future world population is soon to encounter a shortage of required plant derived biomass, either for food or practical applications. It is important to realize that the process of natural selection for plant genotypes has most likely not solely been guided by an intrinsic goal to maximize biomass production. Rather, depending on their life cycle, life strategy and environmental conditions, plants are much more likely to invest in defense against pathogen, coping with adverse conditions and to accumulate enough biomass to maximize offspring viability. Nonetheless, recent models have indicated that the human demand for plant biomass will exceed the current production capacity in the near future. Improving the efficiency of photosynthesis has since been designated as one of the primary targets for improving crop yield [3].

For plant species performing C3 photosynthesis, the theoretical maximum light energy to biomass conversion efficiency of photosynthesis is 4.6% [4], implying that for various reasons at least 95.4% of solar energy cannot be used for biomass production. One of the major losses is the fact that approximately 50% of solar radiation is not photosynthetically active and that significant fractions of the active radiation are either transmitted or reflected

by leaves [4]. In the case of crop plants growing in canopies, perception of light becomes more limiting due to shading within the leaf layers of individual plants and between different plants. Within the photosynthetically active portion of solar radiation, red photons are used more efficiently for photoexcitation than blue photons, meaning that an additional portion of solar energy is wasted due to the relatively inefficient use of blue photons [4]. Most of the losses of energy that occur either prior to or during light harvesting are intrinsic to the biophysical architecture of the thylakoid membrane proteins and are therefore very difficult to eliminate or circumvent using molecular genetics or plant breeding approaches. The most well studied biochemical problem of photosynthesis is the process of photorespiration, which results from the oxygenation reaction that is catalyzed by RuBisCo [5] especially when C3 crops are meeting hot and dry conditions. A more efficient photosynthetic pathway, designated C4 photosynthesis [6], has evolved over the course of plant evolution on several separate occasions and has allowed for the nearly complete elimination of photorespiration at the expense of a relatively small investment in terms of biomass. Plant species performing C4 photosynthesis therefore have a higher maximum light energy to biomass conversion efficiency, now reaching an efficiency of 6.0% [4]. Regardless, the photosynthetic efficiency of field-grown plants throughout the plant kingdom typically does not exceed 1-2% [7], clearly suggesting that for human purposes yield could be boosted substantially by increasing the overall efficiency of photosynthesis.

If one were to improve the overall efficiency of photosynthesis substantially, it would mean bringing *in planta* photosynthetic efficiencies closer to the theoretical maximal efficiency of plant photosynthesis in nature, which is the 6.0% of C4 photosynthesis. A great number of studies have discussed the processes that have been designated as the most important targets for enhancement of overall photosynthetic efficiency [8]. Most notably, the genetic modification of the enzyme RuBisCO [9-12] and the incorporation of either a part or the entire C4 photosynthesis pathway into C3 species [11, 13-17] have received a large amount of attention in literature. Other molecular genetic approaches have demonstrated that a wide variety of (heterologous) overexpression gene constructs can to a certain extent enhance photosynthesis and/or growth in the model species *Arabidopsis thaliana* and in other plants species (e.g. [18-21]). In addition, natural genetic variation and hybrid vigor have both been coined as potential sources of novel target genes for enhancement of photosynthetic efficiency [22, 23]. Regardless of the great number of suitable targets for improvement, no plant lines or cultivars with photosynthetic efficiencies and biomass conversion efficiencies that are high enough to meet the predicted demand have to my knowledge been documented. In addition, predictions have been made that current findings will most likely not be sufficient to meet the demands of the world population on the relatively short term that is required [4, 24, 25]. If more efficient overall photosynthesis could be achieved, it would also have to be translated into substantially more biomass accumulation, often under unfavorable environmental conditions that require stress tolerances. Therefore, the rather extreme demands that are posed on plant productivity

will ultimately require crop plants to integrate more efficient photosynthesis, more biomass accumulation and biotic and abiotic stress tolerances, all within a relatively short timeframe. Due to the extensive regulation of photosynthesis [26], the great influence of environmental conditions and the substantial number of genes involved, it seems to be very unlikely that a single change in the photosynthetic pathway will be sufficient to reach that goal. In order to acquire novel insights into the process of engineering more efficient photosynthesis I have investigated the use of zinc finger artificial transcription factor (ZF-ATF) mediated genome interrogation as a novel tool for the enhancement of photosynthesis-related traits, the results of which are described in this thesis.

Chapter 2 is a review of the use of artificial transcription factors (ATFs) as dominant and *in trans* regulators of endogenous gene expression in plants. Specifically, the use of ATFs with low complexity DNA binding domains to induce large scale changes in transcriptional activity is discussed, a technique which has been designated as ‘genome interrogation’. The merits of the three most commonly used types of DNA binding domains, i.e. zinc fingers (ZFs), TALEs, and complementary RNA molecules adopted from the CRISPR/Cas9 technology are discussed in relation to genome interrogation experiments. In particular, insights into zinc finger ATF (ZF-ATF) mediated genome interrogation are provided.

In **chapter 3**, the construction of two collections of Arabidopsis lines expressing fusions of three zinc fingers to the transcriptional repressor motif EAR (3F-EAR) and to the transcriptional activator VP16 (3F-VP16) is described, and the growth characteristics of populations of both types of plants is documented. Several lines with substantially enhanced growth were isolated, of which the performance compared to wild type plants was quantified in detail. New insights into the 3F-EAR and 3F-VP16 induced gene expression patterns are provided that could form the basis for the enhancement of growth.

In **chapter 4**, the design of a novel system to induce differential expression of the chloroplast genome is discussed, a technique which was designated ‘chloroplast genome interrogation’. The design of a nucleus-based expression cassette system that allows for the translocation of ZF-ATFs consisting of an array of two ZFs (2F) to bacterial transcriptional activators into Arabidopsis chloroplasts is described. Evidence is provided that ZF-ATFs can be targeted to the chloroplasts of Arabidopsis seedlings using this system and can induce phenotypic changes and changes in the operating light use efficiency of PSII. The data provided suggest that the engineering of chloroplasts might be a viable tool for the enhancement of plant photosynthesis.

In **chapter 5**, the isolation of a novel, recessive Arabidopsis mutant with a high operating light use efficiency of PSII and low chlorophyll fluorescence levels from a population of plants harboring 3F-VP16 encoding T-DNA constructs is described. The chlorophyll fluorescence and growth characteristics of this mutant were examined in detail, through which new insights are provided into the potential photosynthetic light use efficiency of Arabidopsis plants and its interconnection with biomass accumulation.

Finally, in **chapter 6** the quantitative phenotypic screening of the collection of Arabidopsis 3F-VP16 lines for salinity tolerance is described. Salinity greatly compromises photosynthesis and productivity in plants, making salinity tolerance a valuable trait in agriculture. Several different 3F-VP16 fusions were found to function as dominant and *in trans* triggers of salinity tolerance in Arabidopsis, through which new insights are provided into 3F-VP16 induced changes in gene expression levels that might form the basis of a novel salinity tolerance mechanism.

References

1. Jarvis P, Lopez-Juez E. (2013) Biogenesis and homeostasis of chloroplasts and other plastids. *Nat Rev Mol Cell Biol* 14(12):787-802.
2. Melis A. (2013) Carbon partitioning in photosynthesis. *Curr Opin Chem Biol* 17(3):453-6
3. Evans JR. (2013) Improving photosynthesis. *Plant Physiol* 162(4):1780-93.
4. Zhu XG, Long SP, Ort DR. (2010) Improving photosynthetic efficiency for greater yield. *Annu Rev Plant Biol* 61:235-61.
5. Maurino VG, Peterhansel C. (2010) Photorespiration: current status and approaches for metabolic engineering. *Curr Opin Plant Biol* 13(3):249-56.
6. Covshoff S, Burgess SJ, Knerova J, Kumpers BM. (2014) Getting the most out of natural variation in C4 photosynthesis. *Photosynth Res* 119(1-2):157-67.
7. Barber J. (2009) Photosynthetic energy conversion: natural and artificial. *Chem Soc Rev* 38(1):185-96.
8. Ort DR, Merchant SS, Alric J, Barkan A, Blankenship RE, Bock R, *et al.* (2015) Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proc Natl Acad Sci U S A* 112(28):8529-36.
9. Parry MA, Andralojc PJ, Mitchell RA, Madgwick PJ, Keys AJ. (2003) Manipulation of Rubisco: the amount, activity, function and regulation. *J Exp Bot* 54(386):1321-33.
10. John Andrews T, Whitney SM. (2003) Manipulating ribulose biphosphate carboxylase/oxygenase in the chloroplasts of higher plants. *Arch Biochem Biophys* 414(2):159-69.
11. Leegood RC. (2013) Strategies for engineering C(4) photosynthesis. *J Plant Physiol* 170(4):378-88.
12. Lin MT, Occhialini A, Andralojc PJ, Parry MA, Hanson MR. (2014) A faster Rubisco with potential to increase photosynthesis in crops. *Nature* 513(7519):547-50.
13. Ku MS, Cho D, Li X, Jiao DM, Pinto M, Miyao M, *et al.* (2001) Introduction of genes encoding C4 photosynthesis enzymes into rice plants: physiological consequences. Novartis Foundation symposium. 236:100-11; discussion 11-6.
14. Miyao M. (2003) Molecular evolution and genetic engineering of C4 photosynthetic enzymes. *J Exp Bot* 54(381):179-89.
15. Hibberd JM, Sheehy JE, Langdale JA. (2008) Using C4 photosynthesis to increase the yield of rice-rationale and feasibility. *Curr Opin Plant Biol* 11(2):228-31.
16. Sage TL, Sage RF. (2009) The functional anatomy of rice leaves: implications for refixation of photorespiratory CO₂ and efforts to engineer C4 photosynthesis into rice. *Plant Cell Physiol* 50(4):756-72.
17. Ruan CJ, Shao HB, Teixeira da Silva JA. (2012) A critical review on the improvement of photosynthetic carbon assimilation in C3 plants using genetic engineering. *Crit Rev Biotechnol* 32(1):1-21.
18. Zhang ZP, Yao QH, Wang LJ. (2011) Expression of yeast Hem1 controlled by Arabidopsis HemA1 promoter enhances leaf photosynthesis in transgenic tobacco. *Mol Biol Rep* 38(7):4369-79.
19. Hu F, Kang Z, Qiu S, Wang Y, Qin F, Yue C, *et al.* (2012) Overexpression of OsTLP27 in rice improves chloroplast function and photochemical efficiency. *Plant Sci* 195:125-34.
20. Uematsu K, Suzuki N, Iwamae T, Inui M, Yukawa H. (2012) Increased fructose 1,6-bisphosphate aldolase in plastids enhances growth and photosynthesis of tobacco plants. *J Exp Bot* 63(8):3001-9.
21. Tanaka Y, Sugano SS, Shimada T, Hara-Nishimura I. (2013) Enhancement of leaf photosynthetic capacity through increased stomatal density in Arabidopsis. *New Phytol* 198(3):757-64.

22. Flood PJ, Harbinson J, Aarts MG. (2011) Natural genetic variation in plant photosynthesis. *Trends Plant Sci* 16(6):327-35.
23. Fujimoto R, Taylor JM, Shirasawa S, Peacock WJ, Dennis ES. (2012) Heterosis of Arabidopsis hybrids between C24 and Col is associated with increased photosynthesis capacity. *Proc Natl Acad Sci U S A* 109(18):7109-14.
24. Long SP, Zhu XG, Naidu SL, Ort DR. (2006) Can improvement in photosynthesis increase crop yields? *Plant Cell Environ.* 29(3):315-30.
25. Murchie EH, Pinto M, Horton P. (2009) Agriculture and the new challenges for photosynthesis research. *New Phytol* 181(3):532-52.
26. Horton P. (2000) Prospects for crop improvement through the genetic manipulation of photosynthesis: morphological and biochemical aspects of light capture. *J Exp Bot* 51:475-85.

Chapter 2

Artificial transcription factor-mediated regulation of gene expression

Niels van Tol and Bert J. van der Zaal

Reprinted from
van Tol N, van der Zaal BJ. (2014)
Artificial transcription factor-mediated regulation of gene expression.
Plant Sci 225:58-67.
with permission from Elsevier

Abstract

The transcriptional regulation of endogenous genes with artificial transcription factors (TFs) can offer new tools for plant biotechnology. Three systems are available for mediating site-specific DNA recognition of artificial TFs: those based on zinc fingers, TALEs, and on the CRISPR/Cas9 technology. Artificial TFs require an effector domain that controls the frequency of transcription initiation at endogenous target genes. These effector domains can be transcriptional activators or repressors, but can also have enzymatic activities involved in chromatin remodeling or epigenetic regulation. Artificial TFs are able to regulate gene expression *in trans*, thus allowing them to evoke dominant mutant phenotypes. Large scale changes in transcriptional activity are induced when the DNA binding domain is deliberately designed to have lower binding specificity. This technique, known as genome interrogation, is a powerful tool for generating novel mutant phenotypes. Genome interrogation has clear mechanistic and practical advantages over activation tagging, which is the technique most closely resembling it. Most notably, genome interrogation can lead to the discovery of mutant phenotypes that are unlikely to be found when using more conventional single gene-based approaches.

1. Introduction

The phenotype of any given organism results from a complex interplay between its genome and the mechanisms that led to the expression of its genes. This interplay is characterized by intricate feedback loops that generate the essential robustness of the phenotype. The feedback loops must also allow for flexibility when endogenous or exogenous stimuli demand for specific phenotypic adaptations. The metaphor of Waddington's epigenetic landscape [1], a model describing the different developmental paths that an embryonic cell can take towards differentiation, is still very much relevant to modern developmental genetics. The stability of gene expression patterns controlled by established epigenetic cues enables cells to withstand most of the random biotic and abiotic noise. However, when a key determinant is able to induce a crucial epigenetic change, cells and organisms might be forced into a different state or developmental program. This epigenetic view of the regulation of gene expression complements the view where genetic variation is the source of phenotypic variation; genetic variation is futile when not expressed. The phenotype of a cell can be regarded as being the product of the epigenetic landscape, genome wide transcription patterns and variation at the sequence level at any given stage of development. Fundamental research on these processes has allowed us to gather knowledge on which genes or sets of genes are involved in phenotypes of interest. In this review, we address several means of placing phenotypes under artificial control by employing artificial transcription factors (TFs) as tools for regulating the expression of endogenous genes in plants.

1.1. Regulation of gene expression

The short sequence upstream of the transcription start site that in eukaryotic genes contains the binding sites for general transcription factors and RNA polymerase II [2] is often referred to as the "minimal promoter" of a gene. More gene-specific regulatory sequences can be found in the DNA sequence upstream of this minimal promoter. It has become common practice in the field of plant molecular biology to designate a rather arbitrary DNA fragment of one to a few kilobase (kb) pairs long and located upstream of the translational start site as the "promoter" of a gene. Plant molecular biologists are usually aware of the fact that many more regulatory sequences exist at greater distances at both the 5' and the 3' ends of a gene as well as within its coding sequence that contribute to the precise level of gene expression. Short statements regarding "promoter activity" usually refer to the contribution of at most a few kb of upstream DNA sequence on to the regulation of transcription levels. Within the context of artificial TF-mediated regulation of gene expression, it would be better to employ the term "gene control region" rather than "promoter". This control region is usually defined as the portion of a eukaryotic gene containing the core promoter as well as any other regulatory sequences that control or influence transcription of that gene. Within the control

region, the eukaryotic core promoter is defined as the region that can be bound by the general transcription factors required for RNA Polymerase II-dependent transcription initiation at the transcription start site, thus equaling the “minimal promoter” mentioned above. Apart from the core promoter, the control region contains enhancer and silencer sequences [3]. These regulatory sequences are potential docking sites for more specific transcription factors that can affect the number of transcription starts at the core promoter per unit of time. The regulatory sequences can be present *in cis* of the start site, within a distance of a few kb from the core promoter, or be located at much larger genomic distances where the term “*in cis*” gradually becomes practically irrelevant. In the latter cases, these regulatory elements are absent from the relatively short PCR-generated DNA sequences taken for the “promoter” in more pragmatic approaches. When discussing the effects of artificial TFs, it is much more appropriate to acknowledge all interactions that are formed within the larger gene control region.

The conserved Mediator complex is also required for successful initiation of RNA Polymerase II-dependent transcription at core promoters in eukaryotes. The Mediator complex functions as a highly complex co-activator of transcription, interacting with the protein domains of RNA polymerase II holoenzyme and general transcription factors. Mediator also interacts with the more specific transcription factors binding to sequences outside of the core promoter. Without the stimulatory contribution of the latter proteins, RNA polymerase II is unable to initiate gene transcription [4,5]. The Mediator complex can thus be thought of as a platform for integrating or relaying signals that can stimulate the initiation of transcription in the regulation of gene expression [4]. However, once the factors conducive for transcription are present and the expression of genes has been switched on in a stable manner, one could imagine that further information and activity is needed to subsequently decrease transcriptional activity or even switch off the expressed genes when this would be required, such as during developmental processes. Accumulating evidence connects the Mediator complex with epigenetic regulation, recruiting factors and enzymes that lead to the deposition of epigenetic molecular markers associated with gene silencing [6,7].

1.2. Chimeric transcription factors

Transcription factors contain a DNA binding domain and a domain that is able to affect transcriptional regulation. Such “effector” regulatory domains increase or decrease the number of transcriptional starts of a gene when bound to DNA at an appropriate position in the gene control region. The effector domain can be envisaged as directly interacting with one or more of the general transcription factors and/or RNA polymerase subunits at the transcription start site or indirectly by recruiting proteins that make these essential contacts.

The use of these effector domains has been reported in connection with natural transcription factors. Plant transcription factors equipped with signature DNA binding domains were fused to a small C-terminal peptide domain that inhibits gene expression [8,9]. This strategy is aimed at turning natural transcriptional regulators into dominant repressors of gene expression that specifically bind to the gene control region of their natural target genes. Changes in the phenotype are readily observed due to loss-of-function mutations resulting from the reduced expression of the genes that are under control of the transcription factors being experimentally manipulated. This strategy is termed Chimeric REpressor gene Silencing Technology (CRES-T) [10]. A system involving fusions with activating effector domains instead of repressing domains could also be envisaged, where an enhancing transcription factor would then affect transcription at its natural target loci in a positive manner.

In the CRES-T technology, as well as in its possible derivatives, DNA binding properties of natural TFs form the basis for the mode of action of these chimeric proteins. The artificial TFs discussed below allow for recognition of any target site of choice to affect the transcriptional activity of genes of interest at the control regions of their normal genomic position. However it is necessary to address relevant target sites within the control region to specifically regulate the expression of endogenous genes of interest. A technique that employs naturally occurring DNA binding domains is hardly an option. Even if a binding site for a known transcription factor would be present, such sites are usually of low complexity and occur at many positions within the genome. This could possibly affect the transcriptional regulation of a host of genes that are normally under control of this particular transcription factor. Custom made site-specific DNA binding domains are required to address unique sites within the genome. The molecular details of systems that allow for site-specific protein-DNA recognition have become understood during the last 15 years to such an extent that it has become possible to design and produce sequence specific DNA binding domains.

2. Artificial DNA binding domains

2.1. Zinc finger domains

From the late 1990s onwards, the DNA recognition code for Cys2His2 (C2H2)-type zinc finger domains has largely been elucidated. Each zinc finger domain (approximately 30 amino acids long) interacts with a triplet of consecutive bases on one strand of the DNA through one amino acid residue just before its alpha helix, and two amino acids within its alpha helix [11]. A fourth contact is made with a base on the opposite strand [11]. Changes in the amino acid composition of the alpha helix change the DNA binding specificity [11].

For site-specific recognition, complex DNA binding domains are required. With each ZF module recognizing a triplet of base pairs, it can easily be calculated that one will need fusions of at least five or six ZF domains to define a cognate contiguous 15 or 18 base pair sequence that is unique within a complex genome, encompassing a billion base pairs or more. The design of complex polydactyl zinc fingers and the consequences for their association or dissociation constants (K_a and K_d , respectively) have been studied [12-15]. The dissociation constants of the contiguous target sequences of six-fingered polydactyl zinc fingers are mostly in the low nanomolar range [13,16,17]. For the shorter three-fingered domains, the dissociation constants were in the range of 10 - 80 nM [17,18], but methods for determining their K_a/K_d values have never been standardized. Thus, the biochemical affinity of a polydactyl zinc finger domain for its target sequence increases with extra zinc finger-DNA triplet interactions.

Zinc finger-DNA interactions are not fully specific [19]. Zinc fingers cannot be expected to bind to their cognate DNA triplets with equal affinity. For several of the 64 possible DNA triplets, specific zinc finger domains are unlikely to become available as there seem to be insufficient structural means to establish the required protein-DNA contacts [20,21]. Moreover, the published triplet specificities of zinc fingers have been based on selection criteria where the particular zinc finger of interest was embedded within a larger structural context, such as being the middle one within a three finger framework aimed at binding a 9 bp target sequence. Altogether, established zinc finger-DNA recognition lexicons should be regarded as reflecting a preferred interaction rather than an exclusive one.

When aiming for highly specific, high affinity DNA binding, a positive bias towards selection of ZFs recognizing 5'-GNN-3' triplets (with N being any of the four DNA bases) [22] might be advisable, since the GNN-based ZF code provides for robust interactions with nearly all of its constituents [19]. A protocol such as Oligomerized Pool ENgineering (OPEN) provides for many details as how to generate arrays of zinc fingers [23]. Although the triplet recognition code might suggest otherwise, zinc finger-DNA interactions involve an extra contact with a fourth base in the opposite strand of DNA, as mentioned above. This phenomenon, known as "target site overlap", makes certain consecutive zinc finger combinations unfavorable for high affinity interactions. Despite all available data and having taken precautions regarding target site overlap, the precise affinity and specificity of polydactyl zinc finger domains for their intended target sites remains hard to predict [24]. The cognate 9 bp target sequence of three finger domains is likely to be more of a consensus sequence rather than an actual target sequence that is partly shared by all 9 bp sequences that can interact with the three finger module.

For the more complex six finger domains, it is tempting to assume that *in vivo* interactions only involve 18 bp contacts. However, when just looking at contiguous polydactyl zinc finger

domains and their target sites, each six finger domain provides for 2 five finger interactions, 3 four finger interactions, 4 three finger interactions, and 5 two finger interactions with 15, 12, 9, and 6 bp of DNA, respectively. The more complex interactions, involving three finger domains and more, could very well result in *in vivo* off- target interactions when expressed in target cells, as could combinations of two finger-two finger or two finger-three finger interactions within a single six finger domain. It seems unlikely that one or more mismatches with unrecognized DNA triplets will completely abolish DNA binding to subsets of an 18 bp target site, unless such mismatches will severely distort the binding at neighboring triplets as well. In any case, whether non-specific interactions occur *in vivo* depends on the affinity for the shorter and/or interrupted target sites as well as on the concentration of the polydactyl zinc finger proteins within a cell. Therefore, after assembly of a zinc finger-based DNA binding domain it remains to be seen whether it is able to bind at its intended genomic target site within the organism of interest.

Chromatin structure at or around the intended binding site of polydactyl zinc fingers is very likely to have a role in their interaction with DNA. The DNaseI sensitivity of chromatin embedded target sites positively correlates with the possibility to affect gene regulation by means of zinc finger artificial TFs [25,26], but correlations are not guarantees. The position of the target site within the gene control region is also crucial for any *in trans* regulation of target gene expression. Choosing target sites close to the core promoter, about 50-150 bp upstream of the transcription start site of the gene of interest, offers the best chances for artificial modulation of transcription. Still, the gene control region usually contains one or more binding sites for natural transcriptional regulators that are able to contribute to or repress the assembly of the general transcriptional machinery and the transcription initiation complex. Zinc finger artificial TFs targeting this region might therefore compete with or even replace the normal regulatory factors. The most practical application of the zinc finger technology is construction of site-specific zinc finger nucleases (ZFNs) [27].

2.2. TALEs

Some 10 years after the elucidation of the greater part of the DNA recognition code of ZFs, the DNA recognition code of transcription activator-like effectors (TALEs) became available [28,29]. TALEs are produced by the bacterial pathogen *Xanthomonas* and are injected into plant cells, where they bind to the regulatory regions of specific plant genes, activating their transcription [30]. The core DNA binding domains of TALEs consist of repeats of modules of 34 amino acids that each bind to 1 bp of DNA. Further insights and refinements of the TALE technology have led to the establishment of highly efficient protocols for the construction of designed TALE domains (dTALEs) [31,32]. As goes for ZF domains, TALE domains have been predominantly used for the construction of site-specific nucleases (TALENs). The “Golden

Gate” assembly allows for the relatively easy construction of dTALE domains with 15-31 repeats, thus in principle recognizing 15-31 bp targets [31]. The TALE technology presently seems to offer better options for making DNA binding domains with sufficient specificity, especially to investigators who do not have access to company-owned polydactyl zinc finger data.

Quantitative assessments of dTALE affinities for DNA are rare, in contrast to polydactyl zinc fingers. The ability to predict the *in vivo* performance of dTALEs is not as straightforward as might be expected. In an extensive analysis, the recognition of DNA by dTALEs was shown to still be surprisingly complex [33]. dTALEs constructed with 19 repeats had apparent K_d 's for their cognate 19 bp target sites ranging from (sub)nanomolar to low micromolar values, depending on the choice of alternative TALE repeats for the recognition of particular nucleotides. Furthermore, the N-terminal repeats contributed more to the DNA binding affinity than the C-terminal ones. An issue that might raise concern for some applications is that a DNA target sequence only providing for an interaction with 10 out of 19 bases could still interact with low nanomolar K_d . Such a phenomenon is likely to result in many off-target interactions within a complex genome [33].

2.3. CRISPR/Cas9

The most recently discovered DNA binding domains were found in the CRISPR (Clustered regularly Interspaced Short Palindromic Repeats)/Cas9 system, which is a defense system employed by a range of bacterial species aimed at the degradation of viral DNA [34]. In this system, specific guide RNAs direct the Cas9 endonuclease protein to their target DNA sequence, leading to subsequent cleavage of that sequence. Manipulation of this naturally occurring system demonstrated that the guide RNAs base pair with complementary DNA sequences at their 5' end, and interact with Cas9 through their 3' end [35]. The length of the homology-searching 5' RNA sequence is usually about 20 bases, but shorter sequences have recently been reported to have less off-target effects [36]. This might be due to the fact that relatively long RNA sequence can allow for more than one high affinity RNA-DNA interaction, while shorter ones can not. The CRISPR/Cas9 system has made an extremely rapid entry into biotechnology, predominantly for making site-specific double strand breaks and thereby targeted mutations within a genome, analogous to zinc finger nuclease and TALEN technology. For such genome engineering purposes, plasmids with genes encoding the guide RNA as well as Cas9 have to be introduced into target cells. The templates for guide RNAs can easily be edited by inserting oligonucleotide sequences complementary to the cognate target DNA region. Thus, the use of the CRISPR/Cas9 system for the induction of site-specific double strand breaks has gained tremendous attention during the last two years [37-39].

Derivatives of the Cas9 protein lacking nuclease activity (dCas9) can also be made amendable for generating artificial TFs. Induction of gene expression was achieved via dCas9 fusions to the powerful transcriptional activator VP64 [40,41]. Specific repression was observed by targeting just a dCas9 protein to potentially regulatory target sites [42]. The elegant and methodologically simple RNA-based targeting strategy of the CRISPR/Cas9 system might make CRISPR/Cas9-based ATFs very attractive as tools for the regulation of endogenous gene expression by offering further prospects for simultaneously controlling the regulation of sets of genes. This would require just the introduction of relatively simple guide RNA expression cassettes and one expression cassette encoding the dCas9 protein. It can be expected that further modifications of the dCas9 protein can provide better scaffolds for translational fusions to different types of effector domains.

3. Effector domains

We will not discuss all possible details about the types of effector domains that can be part of artificial TFs. Generally speaking, these domains can either have stimulatory or inhibitory effects on transcriptional regulation. Within the conceptual framework of gene regulation as mentioned above, both types of effects can be regarded as a consequence of the contribution of effector domain-mediated signals to the assembly of a transcription initiation complex at the gene control region. Effector domains can be chosen for their enzymatic activities, for instance the activities of proteins involved in establishing or removing epigenetic markings or in chromatin remodeling. Changes in epigenetic marks and/or chromatin structure can result in the recruitment of different kinds of transcriptional modulators to the gene control region [43,44]. Effector domains should be able to recruit accessory factors to a particular gene of interest if not active by themselves. Effectors such as (multimers of) the Herpes simplex virus-derived VP16 domain are thought to open chromatin structure by recruiting histone acetylating enzymes and enzymes with SWI/SNF chromatin remodeling activity [45,46]. The class of EAR (-like) repression domains offers the opportunity for gene silencing in plants. The classical LxLxL type EAR domains [47], as well as the more recently identified domain with an R/KLFGV motif [48], act by recruiting co-repressors such as TOPLESS (TPL) or the four TPL-related proteins. Expression of these proteins in *Arabidopsis* is largely constitutive, but enhanced during floral transition and flowering [49,50].

The choice of an effector domain largely depends on the investigator's imagination. The increasing knowledge about epigenetics and transcriptional regulation renders it easier to make an informed choice for a particular effector domain. If any further considerations need to be mentioned, they are rather generic and in fact rather obvious. When attempting

to rewrite epigenetic codes, the effector domain should preferably be active as a monomer, and possess the required enzymatic activity without being hampered by translational fusion with a DNA binding domain. When enzymatic partners need to be recruited, they and their substrates should be available in the cell types of interest. The enzymatic reactions often depend on preceding steps, such as di- and tri-methylation of lysine residue 36 in histone protein 3 of *Arabidopsis* requiring previous monomethylation [51]. Finally, the activity of all site-specific artificial TFs that harbor effector domains is only relevant at the intended target site. Interactions with partners at other cellular positions might lead to a phenomenon known as ‘squenching’, which is the sequestration of partners prohibiting the intended effect at the site of interest.

4. Applications of artificial transcription factors

The scope of present and future applications is in principle the same for any type of artificial TF, whether it is zinc finger-, TALE-, or CRISPR/Cas9-based. Some issues regarding the ease of construction and possible off-target effects have briefly been mentioned above and have recently been reviewed [27]. It seems to be too early to make the clear statement that one type of DNA binding domain, whether it is protein or RNA-based, is to be preferred over any of the others. Several studies have demonstrated the *in vivo* activity of zinc finger-based artificial TFs [52-54]. Designing new zinc finger-based domains for specific target site recognition currently has its problems, certainly for academic researchers with limited funds for outsourcing the synthesis of polydactyl zinc fingers to parties that have exclusive access to company-protected information. If more information will become publicly available and/or the prices of outsourcing drop, the zinc finger technology might very well regain the key position that it had prior to the rapid entry of dTALE domains into this field. The original guidelines for the design of dTALEs might have been an oversimplification [33], but TALE-based artificial TFs have also been the subject of several studies [55-57]. The current pitfalls and problems in dTALE design might be accounted for in the future. The possibilities are very promising for CRISPR/Cas9-based ATFs. At the time of writing this review however, only very few papers have demonstrated their *in vivo* activity [40,41]. A simplified overview of the artificial TFs based on the different DNA binding domain technologies and their mode of action is presented in Fig. 1.

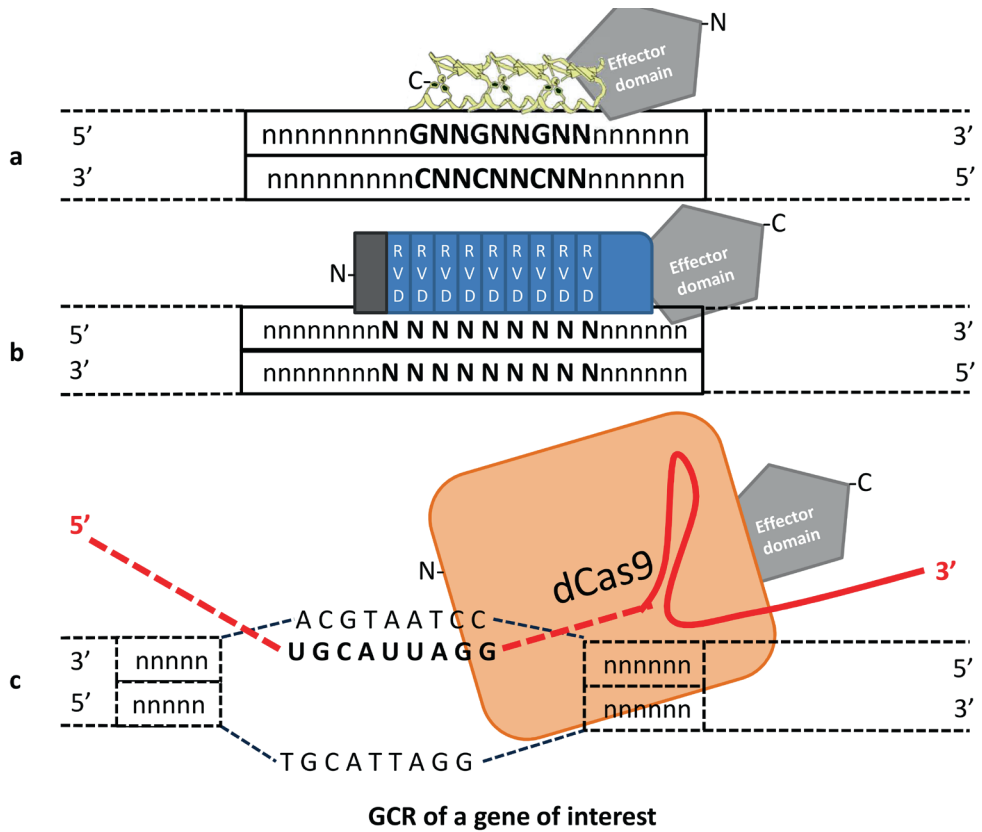


Fig. 1 A schematic overview of the three major classes of DNA binding domains available for the engineering of artificial transcription factors. A simplified model of the respective interaction with 9 bp of DNA of (a) GNN binding three zinc finger-based, (b) TALE-based and (c) CRISPR/dCas9-based artificial transcription factors is presented here. Note that –N and –C refer to the N and C terminal ends of the fusion proteins represented in the figure, but not necessarily to an optimal position for the effector domain.

4.1. Regulation of the expression of specific endogenous genes

Dominant regulatory control of the expression of selected endogenous genes became possible with the emergence of zinc finger artificial TFs. The type of DNA binding domain used for the design of artificial TFs is a practical choice rather than a fundamental one. Thus, it is not surprising that the ideas that were put forward for zinc finger artificial TFs more than 10 years ago [58] also hold for dTALE- or CRIPR/Cas9-based artificial TFs. Several studies with zinc finger- and dTALE-based artificial TFs in the plant field are mentioned in a recent review about advanced genetic tools for plant biotechnology [59]. They all demonstrate the possibility of changing the expression pattern of a plant gene within its native context. Of course, as described above, there are various reasons why other attempts to do so could have failed. Lack of success might be due to the failure of generating a site-specific DNA binding

domain or to an inability of the respective DNA binding domain to access its intended target site. In a yeast model system, only the more complex five and six fingered artificial TFs were able to transactivate a glucose repressed reporter gene [15]. The position of the target site within the gene control region is very important. Nevertheless, with the choice of tools presently available, there is a reasonably high probability that it will be possible to generate an artificial TF influencing the expression of a gene of interest.

4.2. Genome interrogation

Instead of aiming at gene-specific regulation via complex DNA binding domains intended to interact with high affinity at a single cognate target site, artificial TFs with more simple DNA binding domains also allow for interesting applications. In a seminal paper, novel mutant phenotypes could be evoked in yeast cells as well as in mammalian cells when such cells expressed members of a library of about 100,000 zinc finger artificial transcription factors with different three finger and four finger domains [60]. The gist of this method was that single gene constructs encoding a particular three finger or four finger artificial TF were introduced and expressed in transformed cells, as well in their clonal offspring. Three and four finger domains, might allow for multiple genome-wide interactions due to their relatively short 9 bp or 12 bp cognate target sequences. For instance, each randomly chosen 9 bp sequence will on average occur about once every 130,000 bp of double stranded DNA, and a 12 bp sequence about once per 8.3 million bp. Supposing that a target sequence is accessible and that it is located at a suitable position within a gene control region, binding of the zinc finger artificial TF is likely to affect the expression pattern of the gene. In this manner, a relatively simple zinc finger artificial TF can act as an ectopic master switch that modulates the regulation of a multitude of target genes. These primary effects probably result in a plethora of changes in transcription patterns of genes further downstream. These changes in the transcriptome will obviously be accompanied by changes in the proteome and the metabolome.

The ability to cause massive changes in gene expression by just expressing a single artificial TF with lower target site-specificity enables researchers to seek for novel phenotypes, possibly never observed when down regulating or overexpressing single genes. Mutagenesis by means of introducing a library of gene constructs each encoding a unique artificial TF with a rather simple DNA binding domain has been termed 'genome interrogation'. Since the target cells contain the same genome (except for different artificial TF-encoding transgenes), an overview of the potential of a genome is achieved in this way by expressing it differently and/or differentially.

Thus far, genome interrogation has only been accomplished using libraries of zinc finger artificial transcription factors. It is relatively easy to have essentially all members of a large

library of different artificial TF-encoding gene constructs represented in single celled organisms or in cell cultures [60-64]. When the libraries consist of tens of thousands or even millions of different artificial TFs, it is simply not realistic to study the full spectrum of mutant phenotypes that are induced. However, when the introduction of artificial TFs is combined with a single or multiple rounds of selective screening for phenotypes of interest, particular phenotypes might easily be found, even when the causal artificial TFs are relatively rare. Successful isolation of mutants with a phenotype of interest and the subsequent delivery of the proof of principle that the respective isolated zinc finger artificial TF-encoding construct induces this phenotype has led to the notion that genome interrogation is indeed a powerful tool for generating novel phenotypes [60-65]. Interestingly, genome interrogation has previously yielded bacterial phenotypes that could not be mimicked by overexpression of single or two combined candidate genes that were identified as the most likely target genes of the respective zinc finger artificial TF. This can be regarded as a strong indication that the causal artificial TF was able to induce the phenotype by affecting the expression of a combination of genes, precisely as the principle of genome interrogation predicts [62]. Obviously, genome interrogation is not restricted to the zinc finger technology. Both dTALE- and CRISPR/Cas9-based technology should also allow for the generation of libraries of artificial TFs that are suitable for genome interrogation. These alternative DNA binding domains seem not to have been applied in the context of genome interrogation. The first proofs of principle will probably be published soon or have been so at the time of publication of this review.

Applying genome interrogation at the level of multicellular organisms requires generating and culturing transgenic organisms in sufficient quantities to obtain meaningful experimental depth. The first proof of principle of genome interrogation in multicellular organisms was delivered for the model plant species *Arabidopsis* by using a relatively small collection of about 4000 GNN based three fingered artificial TF-encoding genes. In this study, a specific artificial TF inducing very high levels of somatic homologous DNA recombination [65] was identified. Further experimental evidence indicated that this three finger ATF acts as an ectopic master switch orchestrating the timely expression of a set of endogenous genes. This then leads to enhanced somatic recombination. The resulting enhancement is much greater than the one that is accomplished by the overexpression of each of the individual genes [66].

4.3. Comparing genome interrogation to activation tagging

Thus far, the use of genome interrogation for finding novel phenotypes of interest in plants has not yet been reported by other groups. This could be due to various reasons, the most prosaic one being that the method is relatively new and unknown. However, more fundamental questions might relate to its performance and utility in comparison to other methods. Since genome interrogation involves raising transgenic individuals with dominant phenotypes, it

is very different from chemical mutagenesis or radiation mutagenesis which yields mostly recessive mutations in a dose-dependent manner. Moreover, these more traditional mutagenic techniques are based on introducing genetic variation at the DNA sequence level. The phenotypic changes induced by genome interrogation are in principle due to changes in the expression patterns of particular sets of genes. A comparison to activation tagging, however, is appropriate. This method and its derivatives are also based on the introduction of foreign DNA and are also aimed at recovering dominant mutations [67,68]. The essential difference between the two techniques is that genome interrogation is based upon the *in trans* activity of an artificial TF-encoding gene (with its precise location being irrelevant), while activation tagging is based upon the integration of a transcriptional enhancer sequence that can only affect the expression of a gene nearby (*in cis*), mostly within a few kb of the insertion site [68].

It is important to have at least some idea about the size of the mutant population that is required for finding a particular phenotype of interest when considering large scale forward mutagenesis as a tool for genetics. A typical three finger domain might, on average, find about 1000 loci that contain its cognate 9 bp recognition sequence in Arabidopsis, with a haploid genome size of approximately 135 Mbp. Hence, screening a library of only a few thousand transgenic Arabidopsis plants can already be regarded as a potent brute force approach to discover genes or sets of genes where differential expression results in novel and dominant traits. About 750,000 nucleosomes and their cognate 180 bp DNA stretches exist in Arabidopsis when simply thinking of chromatin-embedded DNA as a string of nucleosome structures occurring every 180 bp of DNA. Suppose that the expression level of a particular three finger artificial TF and accessibility of its cognate 9 bp recognition sites are not limiting factors, then one out of each 750 nucleosomes might already be targeted by this 3F-ATF. Having a total of a few thousand different zinc finger artificial TFs represented in about 3000 transgenic plants would in theory be enough to probe every nucleosome structure on average about four times. A large collection of transgenic plants is required to find an *in cis* effect of activation tagging on a particular gene. Up to 100,000 insertion mutants would be required for Arabidopsis to be able to state that each endogenous gene has had a fair chance of being affected by the introduced activation tagging construct. The number of transgenic plants required increases in a linear fashion to perform a saturating screen for a given phenotype of interest in plant species with larger genomes. The number of genome interrogation mutant plants will stay the same; larger genomes contain correspondingly more binding sites for any given artificial TF that can all be targeted *in trans*, provided that the expression level of the artificial TF encoding transgene is sufficiently high. Therefore, screening a relatively small mutant population enables the discovery of novel dominant phenotypes. However, this very strength of genome interrogation could also be considered a weakness when one is interested in elucidating the gene expression patterns that are induced by the artificial TF

and induce the phenotype of interest. Each ATF is a potential trigger of changes in genome wide transcription patterns. Each of the changes in primary target gene expression can have their subsequent pleiotropic effects, apart from the primary effects at the genomic loci having an accessible recognition site for any given artificial TF. Only after having gathered sufficient experimental evidence that reintroduction of a particular artificial TF indeed induces the phenotype, it can be concluded that the particular artificial TF is the master switch triggering all of these primary and pleiotropic effects. However, finding the initial artificial TF-induced changes among the plethora of changes that abide in plants exhibiting a phenotype of interest might seem like an impossible task, especially when considering that any given artificial TF might bind at hundreds of primary recognition sites. In that respect, activation tagging has a clear advantage over genome interrogation. Simply identifying the genomic locus of integration of the activation tagging construct, for instance by TAIL-PCR [69], is in fact a direct lead to the primary candidate gene(s) in the vicinity of the insertion site. This is provided that the genome is adequately annotated. Linking a single gene with a phenotype of interest can then proceed by more or less standard protocols, as it will just require the rather straightforward experimental proof of causality that the ectopic overexpression of the candidate gene also induces the phenotype of interest. An overview of this comparison between genome interrogation and activation tagging is presented in Fig. 2.

4.4. Practical considerations for the application of genome interrogation

Genome interrogation in plants has thus far only been performed with zinc finger artificial TFs [65,66]. When using three finger domains, it is to be expected that a considerable fraction of the zinc finger artificial TF library can only bind to DNA with low affinity. The potential target sites can at best only transiently be occupied by the artificial TF, particularly when the expression levels of the artificial TF-encoding transgene are low. The chances of a given target site being occupied at a given time will increase when expressing artificial TFs with low binding affinity at higher levels. However, one could wonder whether inherently weak DNA-binding affinities will allow for building a molecular platform affecting transcriptional regulation. Polydactyl zinc fingers must have a very high affinity for their target sites to access DNA in a region with an actively repressed chromatin state, and require at least 5F domains for transactivation of gene expression [13]. Surprisingly, access to DNA within typical heterochromatic regions was already possible using 3F domains [70].

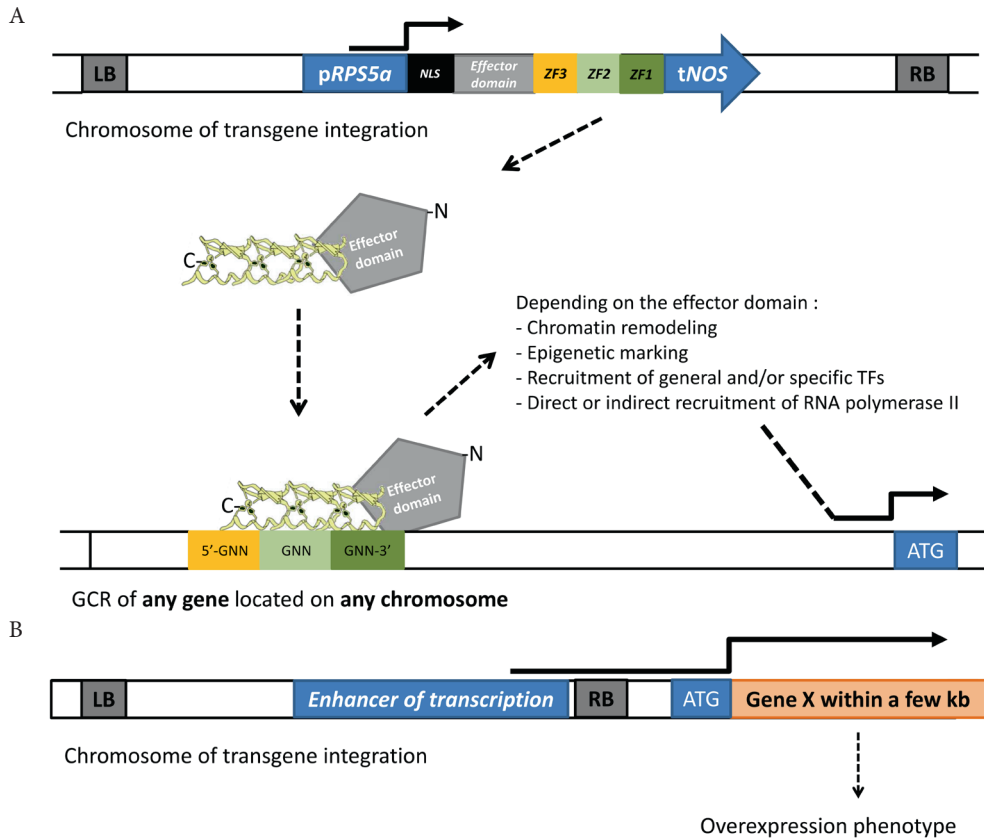


Fig. 2 A comparison of genome interrogation (a) and activation tagging (b). For this example, expression cassettes are assumed to have been introduced as T-DNAs with a left and a right border sequence (LB and RB, respectively) via *Agrobacterium tumefaciens*-mediated transformation. The genome interrogation construct in (a) encodes an effector domain (ED) fused to an array of three different zinc fingers (ZFs). Expression of this gene is controlled by the promoter of the *RPS5a* gene from *Arabidopsis* and the *NOS* terminator sequence. The N-terminus of the protein contains a Nuclear Localization Signal (NLS) to send it to the nucleus. Note that –N and –C refer to the N and C terminal ends of the fusion proteins represented in the figure, but not necessarily to an optimal position for the effector domain. The activation tagging construct in (b) harbors just an element that is able to enhance transcription.

When good transformation or transfection protocols exist, such as for unicellular organisms or (plant) cell cultures, very large numbers of different artificial TFs can be expressed and assessed for their ability to induce a phenotype of interest. If needed, one or more additional rounds of selective screening can be employed to further enrich for the cells that harbor an artificial TF-encoding gene construct inducing such a phenotype. After their recovery, candidate gene constructs can then be tested individually for their phenotype-inducing properties. Obviously, generating very large numbers of genome interrogation mutants at the plant level can never be a practical option. However, as further described below, a wealth of

different phenotypes can already be observed in small collections of plants, representing at most a few thousand artificial TF-encoding constructs. In our opinion, the most attractive application of artificial TF-encoding genes therefore lies in their use as factors for the induction of novel, complex phenotypes of interest.

4.5. Preliminary results with genome interrogation in plants

We have established collections of seeds of the primary transformants obtained through floral dip transformation of *Arabidopsis* plants with libraries of zinc finger artificial TF constructs. These seed collections are valuable tools for the discovery of (novel) mutant phenotypes. Amongst others, plant lines with phenotypes of interest such as increased vigor, increased tolerance to salinity, and changes in flowering time were easily isolated from a seed pool representing a maximum of approximately 3500 VP16-based three finger artificial TFs (Van Tol and Van der Zaal, unpublished data). An example of the variety of phenotypes observed among primary transformants expressing a fusion of an EAR repressor domain to the C-terminus of three fingers is shown in Fig 3. These plants were transferred to soil in a random fashion from medium containing kanamycin for selection for the presence of the transgene. It is evident that this particular set of plants is very rich in different leaf phenotypes, but one can not immediately attribute all phenotypes in primary transformants to expression of artificial TFs. In our experience however, the most conspicuous phenotypes will reappear as dominant traits in the next generation. For a large fraction of the phenotypes that were experimentally addressed, a zinc finger artificial TF-encoding construct was recovered that could reestablish the respective phenotype when introduced into wild type plants. Obviously, a particular artificial TF-induced phenotype will be linked to transgene activity and might disappear when the gene gets inactivated. Up till now, we have no evidence that silencing occurs. The artificial TF-induced somatic recombination phenotype [65] has been stable over the next six generations that have been studied thus far. We are now particularly interested in screening for phenotypes associated with more complex traits, such as growth and photosynthesis. Orchestrating novel gene expression patterns by means of genome interrogation may prove valuable in aiding breakthroughs in these fields of research.



Fig. 3 An example of the variety of leaf phenotypes observed in randomly picked primary transformant plants expressing 3F-EAR fusion proteins.

One could very well wonder whether artificial TFs in genome interrogation experiments should best be expressed in a constitutive, inducible, or in a tissue specific manner. We have used the promoter of the *RPS5A* gene (*pRPS5a*) from *Arabidopsis* for driving the expression of zinc finger artificial TFs [65]. We were motivated by the fact that this promoter is highly active in zygotes, early embryos, and meristematic tissue [71]. Hence, all cells in an *Arabidopsis* rosette have expressed the zinc finger artificial TF-encoding transgene during their juvenile stages, but are not experiencing unnecessary transcription factor-induced stress when they have matured and/or differentiated. It might be expected, however, that *pRPS5A* controlled expression of artificial TFs in the early embryo shortly after floral dip transformation induces lethal phenotypes. Expression under control of *pRPS5a* in dividing cells might hamper cell division in tissue culture-based transformation protocols, thereby abolishing the formation of any regenerated plant material. Hence, unless early lethality is eliminated by a particular choice of expression cassette, it might result in a final mutant population that is biased for expressing only those artificial TFs that induce relatively mild phenotypes. Also, drastic changes in gene expression patterns invoked by artificial TFs might themselves be lethal in early stages of development independently of the expression cassette that is chosen. Indeed, most of the primary transformants harboring *pRPS5A* expression cassettes appeared to have rather mild phenotypes. Sterility (either male or female) was often found, strongly suggesting that expression of the transgene in gametes can have a substantial effect on fertility. Remarkably,

the majority of surviving primary transformants also lacked an obvious phenotype when using the constitutive CaMV 35S promoter instead of pRPS5A. We did not observe a very conspicuous drop in the efficiency of floral dip transformation. Thus, we tend to believe that indeed most artificial TF-induced phenotypes in our mutant libraries are relatively mild. This does not mean, however, that transcription patterns are also only marginally affected. In plants that were macroscopically normal there was differential expression of more than 1500 genes [66]. Apparently, plants are able to thrive whilst experiencing drastic changes in genome wide transcription patterns.

Different artificial TFs should be found that trigger a similar phenotype when working with high complexity artificial TF libraries and when also being able to generate the required number of transgenic individuals. This approach has led to the identification of gene expression signatures related to the development of drug resistance in human cancer cell lines [64]. However, it might be a problem to find truly comparable phenotypes in relatively small mutant populations when screening at the plant level. A way out of this dilemma is to make a dedicated collection of artificial TFs that are very similar to the one that triggers a uniquely found phenotype of interest. Such an approach was feasible for three finger artificial TFs. By keeping two zinc finger domains constant and alternating the third one, alternative zinc finger artificial TFs were obtained that triggered the same phenotype of interest. Comparative transcriptome analyses were performed between plants expressing phenotype-inducing three finger artificial TFs (Nearest Active Neighbors) and plants expressing highly similar three finger artificial TFs that did not induce the phenotype (Nearest Inactive Neighbors). A core set of genes was identified that specifically contributes to the phenotype of interest [66]. An alternative approach might be to use samples taken from pools of plants expressing related zinc finger artificial TFs that do not induce a phenotype of interest as negative controls for RNA sequencing experiments. When it is of particular interest to identify the genes that cause to an artificial TF-induced phenotype, several approaches could be used. One could use an inducible artificial TF expression system and investigate gene expression changes soon after induction. Target gene identification could then further be corroborated when cognate artificial TF binding sites can be found in their respective gene control regions, if these have functionally been delineated. However, considering the lack of knowledge on the true extent of the gene control region, and the possibly low binding specificity, we do not think that an *in silico* screening for biologically relevant artificial TF target sites among the large number of target sites that exist in a given genome can yield more than just tentative clues. This is especially true in the case of genomes that have not completely or incorrectly been annotated.

5. Concluding remarks

The availability of modular DNA binding domains allows for a wealth of possibilities for artificial regulation of gene expression. Still, alternative methods for gene regulation are available as well. When looking for the specific merits of artificial TF technology for plant biotechnology, we think its niche should be to help us assess which novel traits and phenotypes can be generated by differentially expressing essentially the same genome. This might pave the way for finding novel phenotypes that have not yet been or cannot be made available by single gene-based methods. In this way, it might enable exploring the potential of a genome in terms of the phenotypes it can produce, and to get to know the hidden properties of a selected plant line or cultivar. In this way, artificial TF-mediated genome interrogation will allow us to place these traits under experimental control. It should be realized however, that the genome interrogation technology greatly depends on genetic modification. It should be noted that when there would be the requirement of introducing complex, multilocus properties into a selected plant line or cultivar, genetic modification is almost certainly going to be a part of such a procedure. In any case, genome interrogation allows for an intriguing complementary approach to methods based on learning from extant genetic variation in plant species.

Acknowledgements

This work was carried out within the research programme of BioSolar Cells, co-financed by the Dutch Ministry of Economic Affairs, Agriculture and Innovation, and was cofinanced by the Consortium for Improving Plant Yield (CIPY) which is part of the Netherlands Genomics Initiative / Netherlands Organisation for Scientific Research. We are indebted to Dr. Martijn Rolloos for experimental data on genome interrogation with ZF-EAR constructs, and to Prof. Dr. P.J.J. Hooykaas and Dr. B.S. de Pater for critically reading the manuscript.

References

1. J.E. Ferrell, Bistability, bifurcations, and Waddington's epigenetic landscape, *Current Biol.* 22 (2012) 458-466.
2. I. Moore, M. Samalova, Smita Currup, Transactivated and chemically inducible gene expression in plants, *Plant J.* 45 (2006) 651-683.
3. J.J. Riethoven, Regulatory regions in DNA: promoters, enhancers, silencers and insulators, *Methods Mol. Biol.* 674 (2010) 33-42.
4. Z.C. Poss, C.C. Ebmeier, D.J. Taatjes, The Mediator Complex and transcription regulation, *Crit. Rev. Biochem. Mol. Biol.* 48 (2013) 575-608.
5. J. Yin, and G. Wang, The Mediator complex: a master coordinator of transcription and cell lineage development, *Development* 141 (2014) 977-987.
6. N. Ding *et al.*, Mediator links epigenetic silencing of neuronal gene expression with x-linked mental retardation, *Mol. Cell* 31 (2008) 347-359
7. T. Tsutsui *et al.*, Mediator complex recruits epigenetic regulators via its two cyclin-dependent kinase subunits to repress transcription of immune response genes, *J. Biol. Chem* 288.(2013) 20955-20965.
8. K. Hiratsu, K. Matsui, T. Koyama, and M. Ohme-Takagi, Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in Arabidopsis, *Plant J.* 34 (2003) 733-739
9. K. Matsui, K. Hirahatsu, K. Koyama, H. Tanakana, M. Ohme-Takagi, A chimeric AtMYB23 repressor induces hairy roots, elongation of leaves and stems, and inhibition of the deposition of mucilage on seed coats in Arabidopsis, *Plant Cell Physiol.* 46 (2005) 147-155.
10. N. Mitsuda *et al.*, CREST-T, an effective gene silencing system utilizing chimeric repressors, *Meth. Mol. Biol.* 754 (2011) 87-105.
11. S.A. Wolfe, L. Nekudova and C.O. Pabo, DNA recognition by Cys2His2 zinc finger proteins, *Annu. Rev. Biophys. Biomol. Struct.* 29 (2000) 183-212.
12. J.S. Kim, and C.O. Pabo, Getting a handhold on DNA: design of polyzinc finger proteins with femtomolar dissociation constants, *Proc. Natl. Acad. Sci. USA* 95 (1998) 2812-2817.
13. M. Moore, A. Klug, Y. Choo, Improved DNA binding specificity from polyzinc finger peptides by using strings of two finger units, *Proc. Natl. Acad. Sci. USA* 98 (2001) 1437-1441
14. P. Blancafort, D.J. Segal, C.F. Barbas III, Designing transcription factor architectures for drug discovery, *Mol. Pharmacol.* 66 (2004), 1361-1371
15. L.W. Neuteboom *et al.*, Effects of different zinc finger transcription factors on genomic targets, *Biochem. Biophys. Res. Commun.* 339 (2006) 263-270.
16. Q. Liu, D.J. Segal, J.B. Ghiara, C.F. Barbas III, Design of polydactyl zinc-finger proteins for unique addressing within complex genomes, *Proc. Natl. Acad. Sci. USA* 94 (1997) 5525-5530.
17. R.R. Beerli, D.J. Segal, B. Dreier, C.F. Barbas III, Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks, *Proc. Natl. Acad. Sci. USA* 95 (1998) 14628-14633.
18. D.J. Segal, B. Dreier, R.R. Beerli, C.F. Barbas III, 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 (1999) 2758-2763.
19. D. Carroll, J.J. Mortom, K.J. Beumer, D.J. Segal, Design, construction and *in vitro* testing of zinc finger nucleases, *Nat. Protoc.* 1 (2006) 1329-1341.
20. H. Kono *et al.*, Rational design of DNA sequence-specific zinc fingers, *FEBS Lett.* 586 (2012) 918-923.

21. A.V. Persikov, and M. Singh, De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins, *Nucleic Acids Res.* 42 (2014) 97-108.
22. D.J. Segal *et al.*, 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 (1999) 2758-2763.
23. M.L. Maeder *et al.*, Oligomerized pool engineering (OPEN): an 'open-source' protocol for making customized zinc-finger arrays, *Nat. Protoc.* 4 (2009) 1471-1501.
24. A.V. Persikov and M. Singh, An expanded binding model for Cys2His2 zinc finger protein-DNA interfaces, *Phys. Biol.* 8 (2011) 035010.
25. P.Q. Liu *et al.*, Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A, *J. Biol. Chem.* 276 (2001) 11323-11334.
26. L. Magnenat, P. Blancafort and C.F. Barbas III, In vivo selection of combinatorial libraries and designed affinity maturation of polydactyl zinc finger transcription factors for ICAM-1 provides new insights into gene regulation, *J. Mol. Biol.* 341 (2004) 635-649.
27. T.L. Gaj, C.A. Gersbach, and C.F. Barbas III, ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering, *Trends Biotechnol.* 31 (2013) 397-405.
28. M.J. Moscou and A.J. Bogdanove, A simple cipher governs DNA recognition by TAL effectors, *Science* 326 (2009) 1501.
29. J. Boch *et al.*, Breaking the code of DNA binding specificity of TAL-type III effectors, *Science* 326 (2009) 1509-1512.
30. J. Boch and U. Bonas, *Xanthomonas* AvrBs3 family-type III effectors: discovery and function, *Annu. Rev. Phytopathol.* 48 (2010) 419-436.
31. T.I. Cermak *et al.*, Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting, *Nucleic Acids Res.* 39 (2011) e82.
32. D. Reyon *et al.*, Engineering customized TALE nucleases (TALENs) and TALE transcription factors by fast ligation-based automatable solid-phase high-throughput (FLASH) assembly, *Curr. Protoc. Mol. Biol.* 12 (2013) doi: 10.1002/0471142727.mb1216s103.
33. J.F. Meckler *et al.*, Quantitative analysis of TALE-DNA interactions suggest polarity effects, *Nucleic Acids Res.* 41 (2013) 4118-4128.
34. R. Barrangou *et al.*, CRISPR provides acquired resistance against viruses in prokaryotes, *Science* 315 (2007) 1709-1712.
35. E. Deltcheva *et al.*, CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III, *Nature* 471 (2011) 602-607.
36. Y.I. Fu *et al.*, Improving CRISPR-Cas nuclease specificity using truncated guide RNAs, *Nat. Biotech.* 32 (2014) 279-286.
37. C. Chen, L.A. Fenk, M. de Bono, Efficient genome editing in *Caenorhabditis elegans* by CRISPR-targeted homologous recombination, *Nucleic Acids Res.* 41 (2013) e193.
38. B. Gennequin, D.M. Otte and A. Zimmer, CRISPR/Cas-induced double-strand breaks boost the frequency of gene replacements for humanizing the mouse *Cnr2* gene, *Biochem. Biophys. Res. Commun.* 441 (2013) 815-819.
39. A. Bassett, and J.L. Liu, CRISPR/Cas9 mediated genome engineering in *Drosophila*, *Methods* (2014), <http://dx.doi.org/10.1016/j.ymeth.2014.02.019>.
40. M.L. Maeder *et al.*, CRISPR RNA-guided activation of endogenous human genes, *Nat. Methods* 10 (2013) 977-979.
41. F. Farzadfard, S.D. Perli, and T.K. Lu, Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas, *ACS Synth. Biol.* 2(2013) 604-613.

42. L.S.I. Qi, *et al.*, Repurposing CRISPR as an RNA-guided platform for sequence-specific control of genes expression, *Cell* 152 (2013) 1173-1183.
43. D. Jiang *et al.*, Arabidopsis COMPASS-like complexes mediate histone H3 lysine-4 trimethylation to control floral transition and development, *Plos Genetics* 7 (2011) e1001330.
44. J.-H. Jeong *et al.*, Repression of FLOWERING LOCUS T chromatin by functionally redundant histone H3 lysine 4 demethylases in Arabidopsis, *PLoS One* 4 (2009) e8033.
45. W.J.L. Cummings, D.W. Bednarski, N. Maizels, Genetic variation by epigenetic modification, *PLoS One* 3 (2008) e4075.
46. S. Memedula, and A. Belmont, Sequential recruitment of HAT and SWI/SNF components to condensed chromatin by VP16, *Curr. Biol.* 13 (2003) 241-246.
47. M. Ohta *et al.*, Repression domains of class II ERF transcriptional repressors share an essential motif for active repression, *Plant Cell* 13 (2001) 1959-1968.
48. M. Ikeda, and M. Ohme-Takagi, A novel group of transcriptional repressors in Arabidopsis, *Plant Cell Physiol.* 50 (2009) 970-975.
49. L. Pauwels, and A. Goossens, The JAZ proteins a crucial interface in the jasmonate signaling cascade, *Plant Cell* 23 (2012) 3089-3100.
50. B. Causier *et al.*, The TOPLESS interactome; a framework for gene repression in Arabidopsis, *Plant Physiol.* 158 (2012) 423-438.
51. L.I. Xu *et al.*, Di- and tri- but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in flowering time regulation and other processes in Arabidopsis thaliana, *Mol. Cell Biol.* 28 (2008) 1348-1360.
52. J.T. Stege *et al.*, Controlling gene expression in plants using synthetic zinc finger transcription factors, *Plant J.* 32 (2002) 1077-1086.
53. J.P. Sanchez *et al.*, Regulation of *Arabidopsis thaliana* 4-coumarate:coenzyme-A ligase-1 expression by artificial zinc finger chimeras, *Plant Biotech. J.* 4 (2006) 103-114.
54. C. Passananti *et al.*, Transgenic mice expressing an artificial zinc finger regulator targeting an endogenous gene. *Methods Mol. Biol.* 649 (2010) 183-206.
55. J. Streubel *et al.*, Five phylogenetically close rice SWEET genes confer TAL effector-mediated susceptibility to *Xanthomonas oryzae* pv. *oryzae*, *New Phytol* 200 (2013) 808-819.
56. M.C. Politz, M.F. Copeland, and B.F. Pfeleger, Artificial repressors for controlling gene expression in bacteria. *Chem. Commun. (Camb).* 14 (2013) 4325-4327.
57. M.L. Maeder *et al.*, Robust, synergistic regulation of human gene expression using TALE activators. *Nat. Methods* 10 (2013) 243-245.
58. R.R.I. Beerli, and C.F. Barbas III, Engineering polydactyl zinc-finger transcription factors, *Nat. Biotechnol.* 20 (2002) 135-141.
59. W.L. Liu, J.S. Yuan, and C.N. Stewart Jr., Advanced genetic tools for plant biotechnology, *Nat. Rev. Genet.* 14 (2013) 781-793.
60. K.S. Park *et al.*, Phenotypic alteration of eukaryotic cells using randomized libraries of artificial transcription factors, *Nat. Biotechnol.* 21 (2003) 1208-1214.
61. A.L. Beltran *et al.*, Interrogating genomes with combinatorial artificial transcription factor libraries: asking zinc finger questions, *Assay Drug Dev.* 4 (2006) 317-331.
62. J.Y. Lee *et al.*, Phenotypic engineering by reprogramming gene transcription using novel artificial transcription factors in *Escherichia coli*, *Nucleic Acids Res.* 36 (2008) e102.
63. J.Y. Lee, Engineering butanol-tolerance in *Escherichia coli* with artificial transcription factor libraries, *Biotech. Bioeng.* 108 (2011) 742-749.
64. J.L. Lee *et al.*, Induction of stable drug resistance in human breast cancer cells using a combinatorial zinc finger transcription factor library, *PLoS One* 6 (2011) e21112.

65. B.I. Lindhout *et al.*, Employing libraries of zinc finger artificial transcription factors to screen for homologous recombination mutants in Arabidopsis, *Plant J.* 48 (2006) 475-483.
66. Q. Jia *et al.*, 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 (2013) 1069-1079.
67. D.L. Weigel *et al.*, Activation tagging in Arabidopsis, *Plant Physiol.* 122 (2000) 1003-1013.
68. T.L. Waki *et al.*, A GAL4-based targeted activation tagging system in Arabidopsis thaliana, *Plant J.* 73 (2013) 357-367.
69. Y.G.L. Liu *et al.*, Efficient isolation and mapping of Arabidopsis thaliana T-DNA insert junctions by thermal asymmetric interlaced PCR, *Plant J.* 8 (1995) 457-463.
70. B.I. Lindhout *et al.*, Live cell imaging of repetitive DNA sequences via GFP-tagged polydactyl zinc finger proteins, *Nucleic Acids Res.* 35 (2007) e107.
71. D. Weijers *et al.*, An Arabidopsis Minute-like phenotype caused by a semi-dominant mutation in a RIBOSOMAL PROTEIN S5 gene, *Development* 128 (2001) 4289-4299.

Chapter 3

Enhancement of Arabidopsis growth characteristics using genome interrogation

Niels van Tol¹, Martijn Rolloos¹, Johan Pinas¹, Christiaan V. Henkel¹,
Dieuwertje Augustijn², Paul J.J. Hooykaas¹ and Bert J. van der Zaal¹

¹Institute of Biology Leiden, Faculty of Science, Leiden University,
Sylviusweg 72, 2333 BE, Leiden, The Netherlands

²Leiden Institute of Chemistry, Faculty of Science, Leiden University,
Einsteinweg 55, 2333 CC, Leiden, The Netherlands

Abstract

The rapidly growing world population has a greatly increasing demand for plant biomass, thus creating a great interest in the development of methods to enhance the growth and biomass accumulation of crop species. In this study, we used zinc finger artificial transcription factor (ZF-ATF)-mediated genome interrogation to successfully manipulate the growth characteristics and biomass of *Arabidopsis*. We describe the construction of two collections of ZF-ATF *Arabidopsis* lines expressing fusions of three zinc fingers (3F) to the transcriptional repressor motif EAR (3F-EAR) or the transcriptional activator VP16 (3F-VP16), and the phenotypic characterization of their growth characteristics. In total, six different 3F-ATF lines with a consistent increase in rosette surface area (RSA) of up to 55% were isolated. For two lines we demonstrated that 3F-ATF constructs function as dominant *in trans* acting causative agents for an increase in RSA and biomass, and for five lines we addressed the 3F-ATF induced transcriptomic changes in relation to the enhancement of RSA. Our results indicate that genome interrogation can be used as a powerful tool for the manipulation of plant growth and biomass and that it might supply novel cues for the discovery of genes and pathways involved in these properties.

Introduction

The world population is growing at a tremendous rate and requires an increasingly large amount of plant biomass for a variety of different industries. The high demand for plant biomass also comes at a time of the rapid urbanization and salinization of agricultural land, which greatly compromise the overall agricultural performances of crops. Substantial increases in the yield of many important crop species have been realized over the last couple of decades [1], but for a number of different crop species there has been a stagnation in the improvement of yield since the late 1990s [1-3], indicating that breeders have to a large extent exhausted the existing tools for yield improvement. Currently, photosynthetic efficiency is regarded as one of the primary targets for the further enhancement of crop yield [1, 4]. As the overall solar energy to biomass conversion efficiency of photosynthesis is generally regarded as low [1, 5], the improvement of photosynthetic efficiency is expected to allow for substantial yield increases. Regardless, the projections for the conventional means of crop improvement are that the overall biomass production will at some point in the future not be sufficient to supply the world population [6]. Therefore, there is great interest in the development of novel tools and breeding techniques to enhance the biomass production of crops.

The plant species *Arabidopsis thaliana* has been used as a model system for the discovery of genes that are involved in growth and the accumulation of biomass [7], which was greatly facilitated by its well established genetics. The mechanisms of rosette development and growth have been well documented in *Arabidopsis* [8], and the ectopic overexpression or mutation of a substantial number of so-called Intrinsic Yield Genes (IYGs) has been demonstrated to enhance the growth of *Arabidopsis* through various mechanisms involving the rates of cell expansion and cell division [7]. In some cases IYGs were also able to confer an increase in the growth of other plant species [9-11]. There is also variation in growth rate among natural *Arabidopsis* accessions, many of them being much larger than the most commonly studied wild type accession Columbia-0 (Col-0). Through the discovery of hybrid vigor among crosses between accessions, (epi)genetic factors leading to growth and biomass increases are being established [12-15]. As is the case for *Arabidopsis*, enhancement of crop yield has mostly been achieved in single (mutant) genotypes or crosses between genotypes. These classical strategies have had a great impact on yield, but are being exhausted and will most likely not be sufficient to meet the future demands as described above. When plant productivity is considered to be part of the so-called phenotypic space, the predicted impending failure to meet the demand on productivity thus indicates that the phenotypic space has already been stretched to such an extent that the boundaries of the normal genetic and epigenetic variation are close to being met. Therefore, means are required to introduce novel genetic and epigenetic variation in plants.

Using *Arabidopsis* as a model system, we have investigated whether artificial distortion of gene expression patterns can trigger plants to develop rosettes with a larger surface area and to accumulate more biomass. A promising method to do so has been coined ‘genome interrogation’ and is originally based on the use of zinc finger artificial transcription factors (ZF-ATFs) to drastically change genome-wide transcription patterns and to induce novel phenotypes of interest [16]. While initially successfully applied in unicellular organisms or cell cultures [17-19], studies in our lab have demonstrated that genome interrogation is also possible in plants [20, 21]. In our setup, ZF-ATFs consist of an array of three zinc fingers (3F) which each recognize a cognate 3 base pair (bp) DNA consensus sequence of 5'-GNN-3' [22], with N being any of the four bases. From the sixteen 5'-GNN-3' binding ZFs, 4096 3F combinations can be generated that each function as DNA binding domains recognizing a 9 bp target sequence, which on average occur once in approximately 130,000 bp of genomic DNA. When fused to an effector protein domain that is known to influence transcription of genes, a plethora of genomic loci can simultaneously be addressed for differential expression leading to a phenotype of interest in a multicellular organism. Even in the relatively small genome of *Arabidopsis*, each particular 9 bp sequence will on average occur approximately 1000 times. The level and potential tissue specificity of ZF-ATF expression in the organism of interest depends on the choice of the promoter controlling the transgene encoding the protein. In *Arabidopsis* the promoter of the *Arabidopsis RPS5a* gene, which is predominantly active in embryonic and meristematic tissue [23], proved to be suitable [20].

In a previous study conducted in our lab [20], we have successfully used an effector domain originating from the VP16 protein of the herpes simplex virus and acts as a potent transcriptional activator [24]. As an interesting addition to this we have opted to explore the possibilities for genome interrogation in plants using a transcriptional repressor domain, such as the ERF-associated Amphiphilic Repression (EAR) motif, which originates from *Arabidopsis* [25, 26] and has been reported to confer dominant repressor activity on transcription factors [27]. Here, we describe the construction and phenotypic characterization of two collections of *Arabidopsis* genome interrogation lines harboring 3F-EAR encoding T-DNA constructs (~700 lines) and 3F-VP16 encoding T-DNA constructs (~4200 lines). A multitude of different phenotypes was found in the 3F-EAR collection, and some of the plants were substantially larger than the wild type Col-0. Although the increase in growth of these lines is very likely to be correlated with the presence of a 3F-EAR encoding T-DNA construct, transformation of Col-0 plants with reconstituted T-DNA constructs encoding the same 3F-EAR fusions did not prove a causal connection with enhanced plant growth. In the larger 3F-VP16 collection however, we found lines with a strong increase in surface area and biomass compared to Col-0 for which we were also able to prove that the 3F-VP16 expression constructs are indeed the causative agents. Overall, we show that 3F-ATFs are able to introduce substantial phenotypic, developmental and growth related changes in *Arabidopsis*, and we describe 3F-ATF induced

transcriptomic changes that might be indicative for increased growth and biomass in Arabidopsis. Our results not only suggest that genome interrogation is a powerful tool for the manipulation of the growth and the transcriptome of Arabidopsis, but also that it can be regarded as a novel source of genes and pathways that could be targeted for the improvement of biomass yield in plants.

Results

Construction of the collections of 3F-ATF expressing Arabidopsis lines

In order to obtain collections of plant lines expressing different ZF-ATFs a previously generated library of 3F constructs was used [20]. In brief, this library was composed of 15 non-overlapping subpools, each of which was named after one of the sixteen 5'-GNN-3' binding ZFs that was first cloned to establish a theoretical maximum of 256 3F constructs. A 16th pool (with the 5'-GAA-3' binding ZF as founder) proved to be unstable for unknown reasons [20], and was therefore not constructed. The 3F encoding sequences were translationally fused with the VP16 transcriptional activation domain or the EAR transcriptional repression domain by cloning them into the pRF-VP16-Kana binary vector [20] or a similar vector designated pRF-EAR-Kana, respectively.

The collection of Arabidopsis lines expressing 3F-VP16 fusions was generated by transformation of Col-0 plants with each of the 15 subpools separately. It should be noted that each primary transformant (T1 generation) could thus harbor any of the maximally 256 different 3F-ATFs, or more than one in the case of simultaneous transformation with another construct. Over an extended period of time, a total of approximately 5400 viable primary transformants were obtained, more or less evenly distributed over the 15 different 3F-VP16 subpools. The vast majority of the 3F-VP16 plants did not exhibit conspicuous growth phenotypes at standard growth conditions. The handling complexity of the collection was reduced by cultivating five primary transformants originating from the same subpool together in a single pot and collecting the T2 progeny of these plants in seed bags named 'five-bags'. Due to losses of plants during cultivation and infertility in about 4% of the plants, these five bags in total contained the offspring of 4278 primary transformants, meaning that a fraction of the bags had less complicated seed mixtures originating from 3 or 4 plants rather than 5. The collection of seeds of 3F-VP16 lines was finally represented in a total of 1034 five-bags, and each line was assigned a code name consisting of three parts: effector domain – 3F pool number – five-bag of origin. Assuming that primary transformants are heterozygous for the T-DNA insertion, it is important to note that its T2 offspring will exhibit Mendelian segregation for the 3F-VP16 encoding transgene.

For the collection of 3F-EAR expressing plant lines, we decided to generate a smaller collection representing just 7 different 3F subpools and to aim for approximately 100 primary transformants for each subpool. As the maximal complexity of each subpool is 256 constructs, we could in this way ensure that most primary transformants express different 3F-EAR fusions. Therefore, a total of just over 700 viable primary transformants were cultivated and their growth was documented as described below. Subsequently the seeds (T2 generation) of the individual plants were harvested.

Selection of 3F-ATF lines with increased rosette surface area

Among the primary transformants harboring 3F-EAR encoding constructs there was great variation in rosette size, phenotype, leaf morphology and flowering time (Fig. 1), clearly indicating that 3F-EAR fusions induce phenotypic variation in *Arabidopsis*. To quantitatively assess the growth of 3F-EAR plants, we firstly quantified the rosette surface area (RSA) of all 700 primary transformants. RSA was used as a measure for growth, because it is a non-destructive proxy for biomass in *Arabidopsis* [28]. Because the primary transformants were obtained through kanamycin selection, their growth could not be compared to Col-0, as this cannot be selected for with kanamycin. Also, not all plants could be raised and assessed simultaneously. Therefore, we decided to grow and compare plants to each other in batches of 28 individuals. In this way, approximately 100 primary transformants with at least 80% larger RSA than the batch average were selected for further analysis. Subsequently, the RSA of a segregating population of their T2 progeny was quantified throughout development. Again, a substantial amount of variation was observed in the RSA of the T2 progeny (Fig. 2). For approximately one third of the 100 selected lines the mean RSA values were also larger than Col-0 in the T2 generation. Four lines that displayed the most reproducible and substantial increases in RSA were selected for further analysis as described below.

Overall, there was much less phenotypic variation among 3F-VP16 plants compared to 3F-EAR plants. Approximately 5% of the 3F-VP16 primary transformants exhibited clearly visible phenotypic differences with Col-0 in for instance leaf morphology, leaf color and flowering time. When further investigated in the T2 generation, such conspicuous phenotypes often segregated in a 3:1 ratio, demonstrating that the phenotypes are indeed induced in a dominant manner as expected when induced by ZF-ATFs. Since we did not collect RSA data for the seven-fold larger collection of 3F-VP16 plants and also could not store offspring of each primary transformant individually, we decided to screen for 3F-VP16 plants with large RSA in a pragmatic fashion in the T2 generation, just aiming for plants with the most extreme increases in RSA, which were noted in the lines VP16-02-003 and VP16-05-014. We therefore focused on further analyzing the T3 offspring of these lines, as described below.



Fig. 1 An overview of the variation in rosette phenotypes and growth characteristics among a population of primary transformants (T1) harboring 3F-EAR encoding T-DNA constructs. The presented individuals are representative of the extent of variation. The plants were first grown on selection medium containing kanamycin, and were transferred to soil after approximately 2 weeks. The presented individuals are approximately 1 month old. The size of the individual in the right top corner is representative of a wild type Col-0 plant at this stage of development.

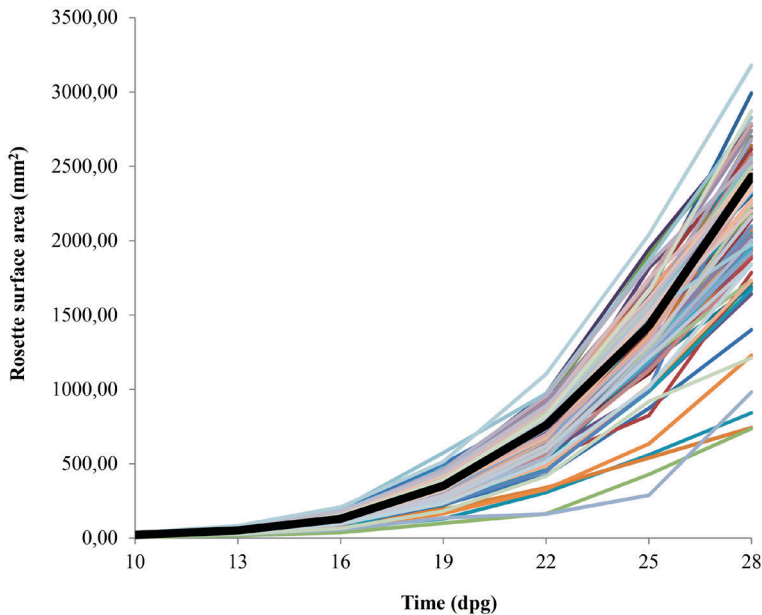


Fig. 2 Growth curves of the T2 progeny (segregating) of the 100 largest 3F-EAR primary transformants in terms of RSA ($n=100$ for Col-0; $n=7$ for the transgenic lines). The growth curve of the wild type Col-0 is presented in black.

Quantification of RSA and PCR analysis of selected 3F-ATF lines

Altogether four 3F-EAR lines (T2 generation) and two 3F-VP16 lines (T3 generation) with consistently larger RSA than Col-0 were selected from the 3F-ATF libraries. Subsequently, a comparative analysis of RSA was performed to verify this selection and to assess the performance of each of these 3F-ATF lines compared to Col-0 in detail. The overall increase in RSA of the lines compared to Col-0 ranged from approximately 20% to 55% (Fig. 3A). It should be noted that measuring RSA by this non-destructive imaging of plants viewed from the top unavoidably leads to an increasing underestimation of RSA along with the gradual overlapping of leaves. The ZF-ATF encoding T-DNA fragments were amplified from the genomic DNA of plants of the six selected lines by PCR. Full length 3F encoding sequences were recovered from each line and all encoded 3Fs with different predicted cognate DNA recognition sites (Fig. 3B). For all 3F-EAR lines and for VP16-05-014, the increases in RSA were attributable to an increase in the surface area of the individual leaves (Fig. 3A). For VP16-02-003 plants, the RSA increase was mostly reflected by an increase in the number of leaves, which were also slightly serrated and lanceolate in shape compared to Col-0 (Fig. 3B).

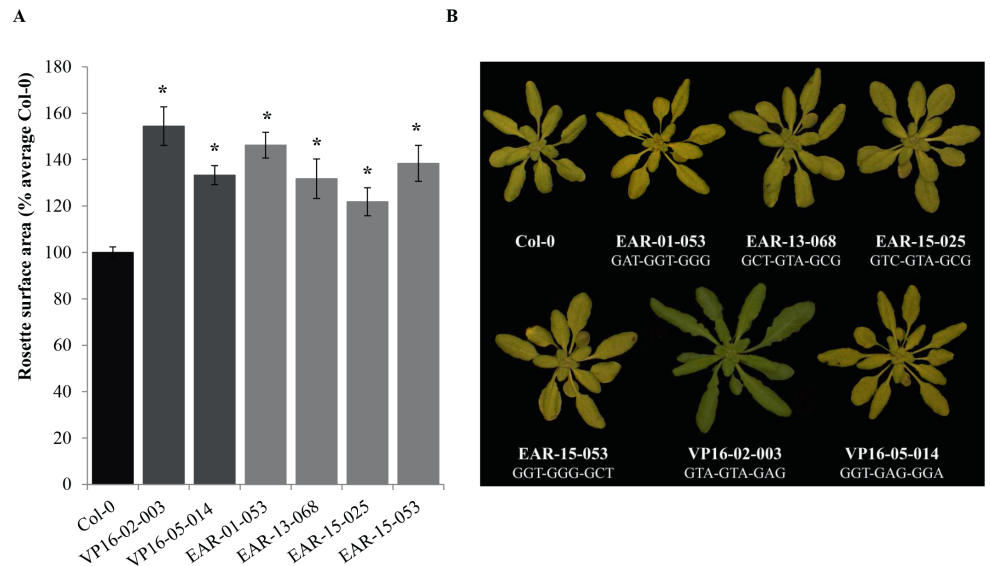


Fig. 3 Overview of the properties of selected 3F-EAR lines (T2; segregating) and 3F-VP16 lines (T3; only VP16-05-014 segregating) with a consistent increase in rosette surface area (RSA) compared to the wild type Col-0. **A**) Quantification of the relative RSA of the selected 3F-ATF lines compared to Col-0. The RSA of each plant was calculated in terms of percentage of the average of Col-0. Error bars represent SEM values (n=182 for Col-0, n=16-18 for the transgenic lines). Significant differences with the Col-0 are indicated by an * ($p < 0.05$). **B**) Overview of the rosette phenotypes or the selected 3F-ATF lines (25 dp), and of the 9 bp DNA recognition sequences of the 3Fs that were isolated from them. The presented individuals had RSA values closest to the average of their genotypes.

T3 populations of VP16-02-003 plants also had a remarkably uniform appearance, with all individuals being substantially larger than Col-0 throughout development (Fig. 4). As we also did not observe segregation for kanamycin resistance among T3 VP16-02-003 plants, this means that the original VP16-02-003 plant that was isolated in the T2 generation must have already been homozygous for the T-DNA insertion involved in enhanced RSA. VP16-02-003 plants started to flower approximately one week earlier than Col-0 plants at a 12 h photoperiod and developed substantially larger inflorescences which produced only few viable seeds (Fig. S1). Although VP16-02-003 plants had a much higher yield in terms of RSA, this was therefore not reflected by seed yield.

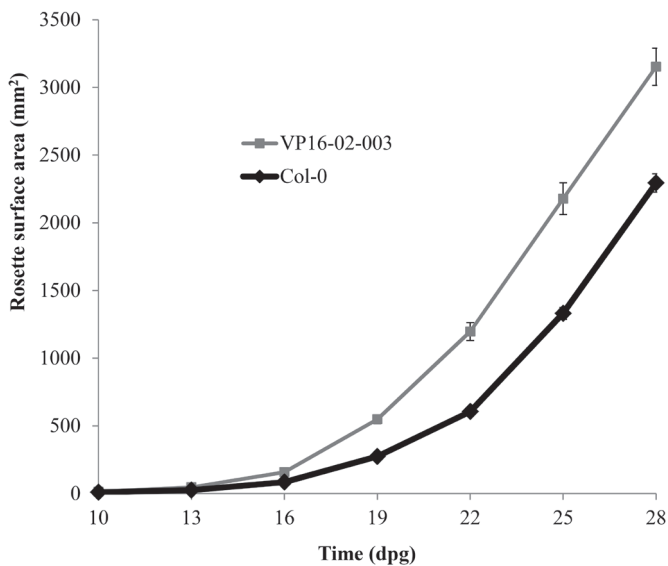


Fig. 4 Growth curves of the wild type Col-0 and VP16-02-003 (T3). Error bars represent SEM values ($n=66$ for Col-0, $n=17$ for VP16-02-003), and are visible only when exceeding the data points. From approximately 22 dpg onwards, overlapping of leaves started to occur, causing an increasing underestimation of RSA for larger plants for which corrections could not be made.

The effect of 3F-ATF encoding T-DNA constructs on the growth of *Arabidopsis*

To investigate whether the 3F-ATF encoding genes from the six selected lines (Fig. 3) were indeed the causative agents for the increase in RSA, reconstituted T-DNA constructs were generated harboring the 3F sequences that had been isolated from these lines. Col-0 plants were transformed with these reconstituted T-DNA constructs and RSA of a segregating population of T2 retransformant plants was analyzed throughout development. The T2 retransformant lines harboring reconstituted 3F-EAR constructs did not exhibit any clear increases in RSA compared to Col-0, indicating that they were not the direct causative agents for increased RSA, or that they induced RSA in a rather intricate manner possibly dependent on the genomic T-DNA integration locus.

Five out of the eight retransformant lines reconstituted from the original VP16-02-003 line displayed a consistently larger RSA than Col-0 throughout development (Fig. 5). At 28 dpv, these differences were in all cases significant at $p < 0.05$ (Fig. 5), except for retransformant line 3, which had a significantly larger RSA at $p < 0.07$. For two retransformant lines we observed an overall decrease in RSA (Fig. 5), and one line displayed wild type growth (Fig. 5). With the majority of the retransformant lines having a larger RSA than Col-0, and with the increases being comparable to or even higher than the increase noted for the original VP16-02-003 line, we concluded that the expression of the 3F-VP16 encoding T-DNA construct from VP16-02-003 induced an increase in RSA in Arabidopsis. Moreover, as these differences in RSA were already detectable from a relatively small population of plants which was still segregating for the transgene, the 3F-VP16 construct from VP16-02-003 most likely induced the increases in RSA in a dominant manner. Among the T2 retransformant plants there was variation in leaf morphology, shape and number throughout development which started to become clearly visible between 13 and 16 dpv (Fig. S2), and which can likely be attributed to segregation of the transgene. In contrast to the original VP16-02-003 line, all retransformant plants had wild type levels of fertility. The retransformant lines that were significantly larger than Col-0 exhibited increases in RSA ranging from 20 to 80% when quantified at 28 dpv, and had an increased number of leaves which were also slightly lanceolate in shape (Fig. S3A). Among the population of T2 retransformant plants there were individuals that were substantially larger than the largest Col-0 individual (Fig. S3B), indicating that the 3F-VP16 construct from VP16-02-003 introduced variation in RSA reaching beyond the variation that can be produced by the Col-0 (background) genotype itself. Five out of the eight retransformant lines reconstituted from the original VP16-02-003 line also had a significantly higher relative fresh weight (Fig. 6A) and dry weight (Fig. 6B) than Col-0 at 28 dpv, thereby confirming the correlation between RSA and biomass. The average relative increases in fresh weight (Fig. 6A) and dry weight (Fig. 6B) of these T2 retransformant lines varied from approximately 30 to 110%, with the increases for the original VP16-02-003 line being 65% and 74%, respectively.

Through retransformation with a 3F-VP16 construct reconstituted from the original VP16-05-014 line (T3 generation) we found that two out of the five T2 retransformant lines constructed in this way displayed a significantly larger RSA than Col-0 throughout development, and a third one was larger at 28 dpv, albeit not significantly so (Fig S4). Therefore, these results indicated that the 3F-VP16 construct from VP16-05-014 is also capable of inducing an increase in RSA *in trans* and in a dominant manner. Analogous observations could be made in terms of fresh weight (Fig. S5A) and dry weight (Fig. S5B), in which cases the overall increases varied from approximately 4 to 30 %, respectively. Altogether, our data demonstrate that we have isolated two 3F-VP16 encoding T-DNA constructs that can induce a significant increase in both RSA and biomass in Arabidopsis.

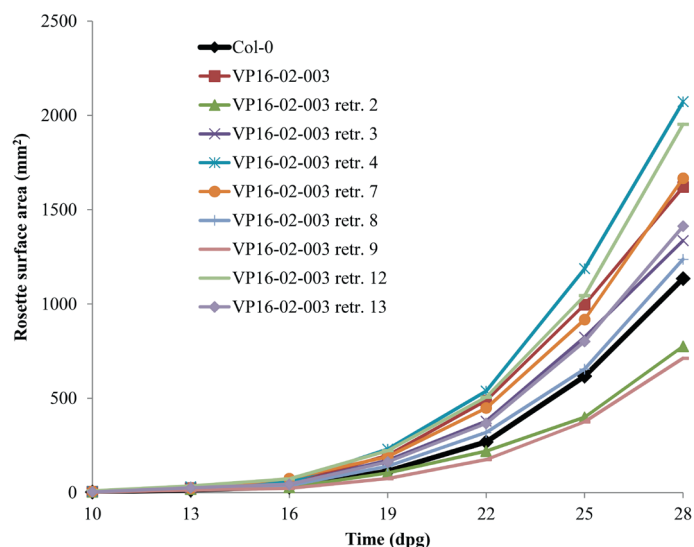


Fig. 5 Growth curves of the wild type Col-0, VP16-02-003 (T3) and retransformants reconstituted from VP16-02-003 (T2; segregating) ($n=48$ for Col-0, $n=15-18$ for the transgenic lines). From approximately 22 dpg onwards, overlapping of leaves started to occur, causing an increasing underestimation of RSA for larger plants for which corrections could not be made. Significant increases compared to Col-0 at 28 dpg are indicated by black asterisks (*) ($p < 0.05$), and significant decreases are indicated by red asterisks (*) ($p < 0.05$).

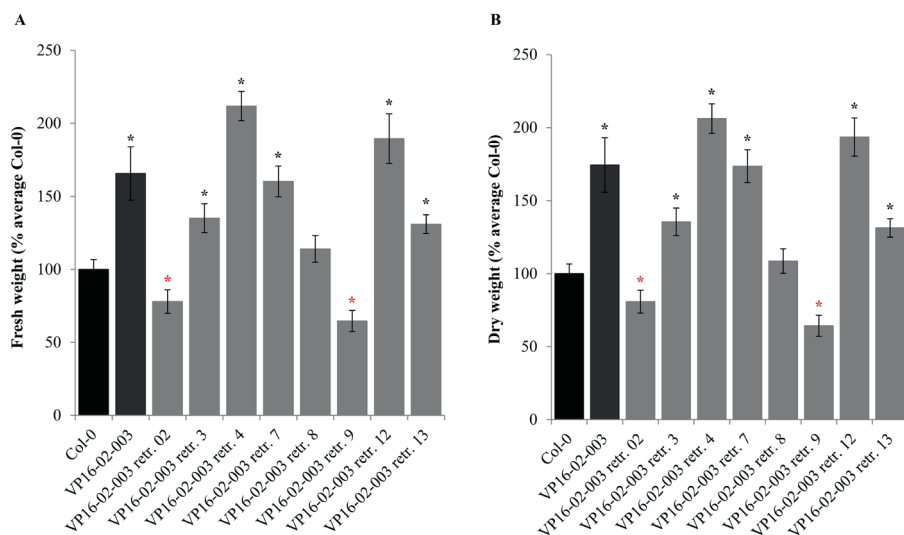


Fig. 6 Quantification of the relative fresh weight (A) and the relative dry weight (B) of VP16-02-003 plants (T3, non-segregating) and retransformant plants reconstituted from VP16-02-003 (T2; segregating) compared to the wild type Col-0 (28 dpg). The fresh and dry weights of each plant were calculated in terms of percentage of the average of Col-0. Error bars represent SEM values ($n=48$ for Col-0, $n=15-18$ for the transgenic lines). Significant increases compared to Col-0 are indicated by black asterisks (*) ($p < 0.05$), and significant decreases are indicated by red asterisks (*) ($p < 0.05$).

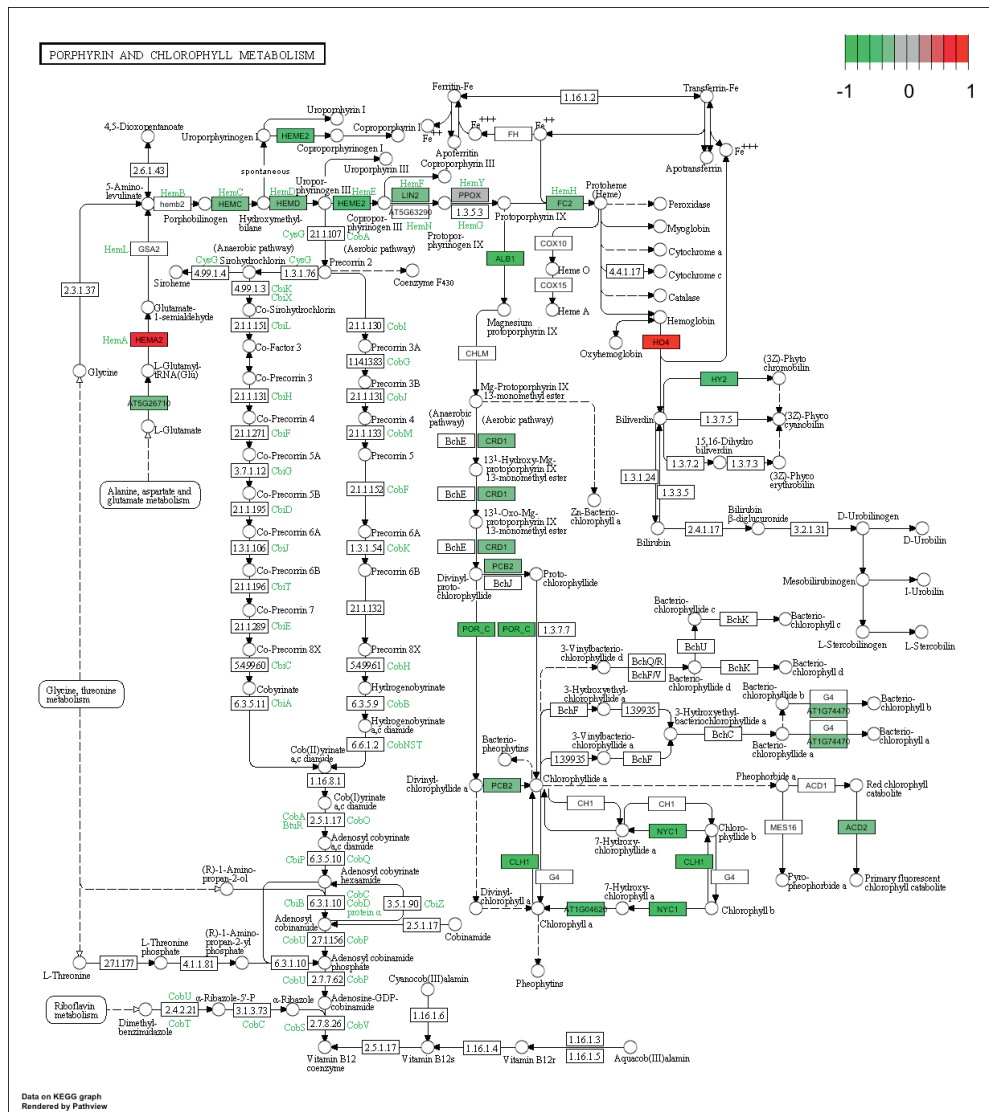


Fig. 7 Overview of the porphyrin and chlorophyll metabolism KEGG pathway of VP16-02-003. Boxes represent gene products (either RNA or protein). Colored boxes indicate differential expression compared to the wild type Col-0. The color legend in the top right corner represents the $2\log(\text{fold change})$ values of gene expression.

The effect of 3F-VP16 expression driven by the CaMV 35S promoter

Both the VP16-02-003 and VP16-05-014 lines harbor 3F-VP16 encoding T-DNA constructs of which the expression is driven by the promoter of the Arabidopsis *RPS5a* gene, which is primarily active in meristematic tissue. As these constructs already induced substantial

Transcriptome analysis of the selected 3F-ATF lines

To gather insight into the transcriptomic basis of 3F-ATF induced enhancement of growth and biomass, RNA sequencing was performed on total RNA extracted from the shoots of Col-0 plants and plants of the selected original 3F-ATF lines (Fig. 3). For budgetary reasons, EAR-01-053 could not be included. For every selected line we also included a complex mixture of plants that express 3F fusions originating from the same 3F-ATF subpools but that were not found to induce an increase in RSA as controls for the transcriptomic changes that are 3F-ATF specific, but are not causative for increased RSA and biomass. These controls theoretically also allow for the discrimination between the 3F-specific and effector domain-specific transcriptomic changes leading to increased RSA, and are further referred to as 'background'. The plant tissue was harvested at 15 dpv, which is around the time point at which the increase in RSA of the 3F-ATF plants became visible. RNA sequencing was performed on three biological replicates consisting of different plants (therefore, $n=3$). Principal component (PCA) analysis showed that for unknown reasons one of the three replicates of VP16-05-014 had such a large sample-sample distance to the other two replicates that it could be disregarded for the data analysis.

Transcriptomes of the three 3F-EAR lines that were assessed exhibited a large number of differentially expressed genes (DEGs) compared to Col-0 (Table S1), which could be matched with significant changes in gene expression for a number of metabolic pathways (Table S2). In the case of line EAR-13-068 for instance, 2598 genes were differentially expressed compared to Col-0 and there was significant enrichment for genes involved in 17 different KEGG pathways among them (Table S2). When compared to the corresponding background samples however, only 1346 genes were differentially expressed (Table S1) and 12 of the significant enrichments for genes in the pathways were lost (Table S2). Similarly, for the other two lines (EAR-15-025 and EAR-15-053) the great majority or even all of the significant enrichments for genes were lost in this manner, respectively (Table S2). These observations indicate that most of the significant transcriptomic changes in these lines occur irrespective of the presence of a specific 3F domain, but are rather triggered by the EAR domain in a more unspecific manner. We cannot exclude that the more subtle changes in gene expression in the 3F-EAR lines form the basis of RSA enhancement, but these were not included in our analysis due to the applied DEG selection criteria.

In case of the 3F-VP16 lines, extensive differential gene expression (Table S3) was also found along with significant enrichments for DEGs in a substantial number of KEGG pathways compared to Col-0 (Table S4). However, as opposed to the 3F-EAR lines, these differences were still very significant when the background was subtracted, clearly indicating that the transcriptomic changes in the 3F-VP16 lines are predominantly 3F-specific and more likely to be correlated with increased RSA, biomass or any of the other phenotypic characteristics. To illustrate this, 4889 genes were differentially expressed in line VP16-02-003 compared to Col-0 ($p < 0.05$), 2437 of which were upregulated, and 2452 of which were down-

regulated. When compared to the corresponding background, 4535 genes were differentially expressed ($p < 0.05$), among which there was a significant enrichment for genes involved in a variety of KEGG pathways (Table S4), suggesting that the enhanced growth of VP16-02-003 is achieved through the combined differential regulation of several metabolic pathways. For instance, a significant down-regulation of a number of steps in the porphyrin and chlorophyll biosynthesis pathway was found (Fig. 7), indicating that the biosynthesis of these metabolites is compromised in VP16-02-003. In addition, several steps in the sulfur metabolism pathway along with the basic steps of the sulfate assimilation were down-regulated at the transcriptional level (Fig. 8), which might have a multitude of downstream effects on various aspects of sulfur metabolism. Other DEGs were indicative of down-regulation of genes involved in plant-pathogen interactions, glucosinolate biosynthesis and alpha-linoleic acid metabolism, suggesting that VP16-02-003 plants might have compromised defense mechanisms as well. Altogether, these observations might indicate that VP16-02-003 plants achieve an increase in RSA through a transcriptional trade-off between several metabolic pathways and growth.

In VP16-05-014, 1753 genes were differentially expressed compared to Col-0 ($p < 0.05$), 815 of which were upregulated, and 938 of which were down-regulated. When compared to the background, there was still differential expression of 1633 genes, again clearly suggesting that in this case the majority of the DEGs are induced in a 3F-specific manner and are causal for the enhancement of RSA. Among the DEGs compared to the background, there was a significant enrichment for genes involved in three KEGG pathways (Table S4), among which the most substantial transcriptional changes were found in the signal transduction pathways of several hormones that are involved in plant growth (Fig. 9), such as those of auxin, cytokinin and gibberellin, suggesting that the enhancement of growth in VP16-05-014 is mediated by the combined differential regulation of hormonal responses that are already known to be involved in enhanced growth of plants. The overall transcriptomic footprint of VP16-05-014 was very different and less complex than that of VP16-02-003, demonstrating that we have found at least two novel transcriptional mechanisms through which an increase in RSA can be achieved in *Arabidopsis*.

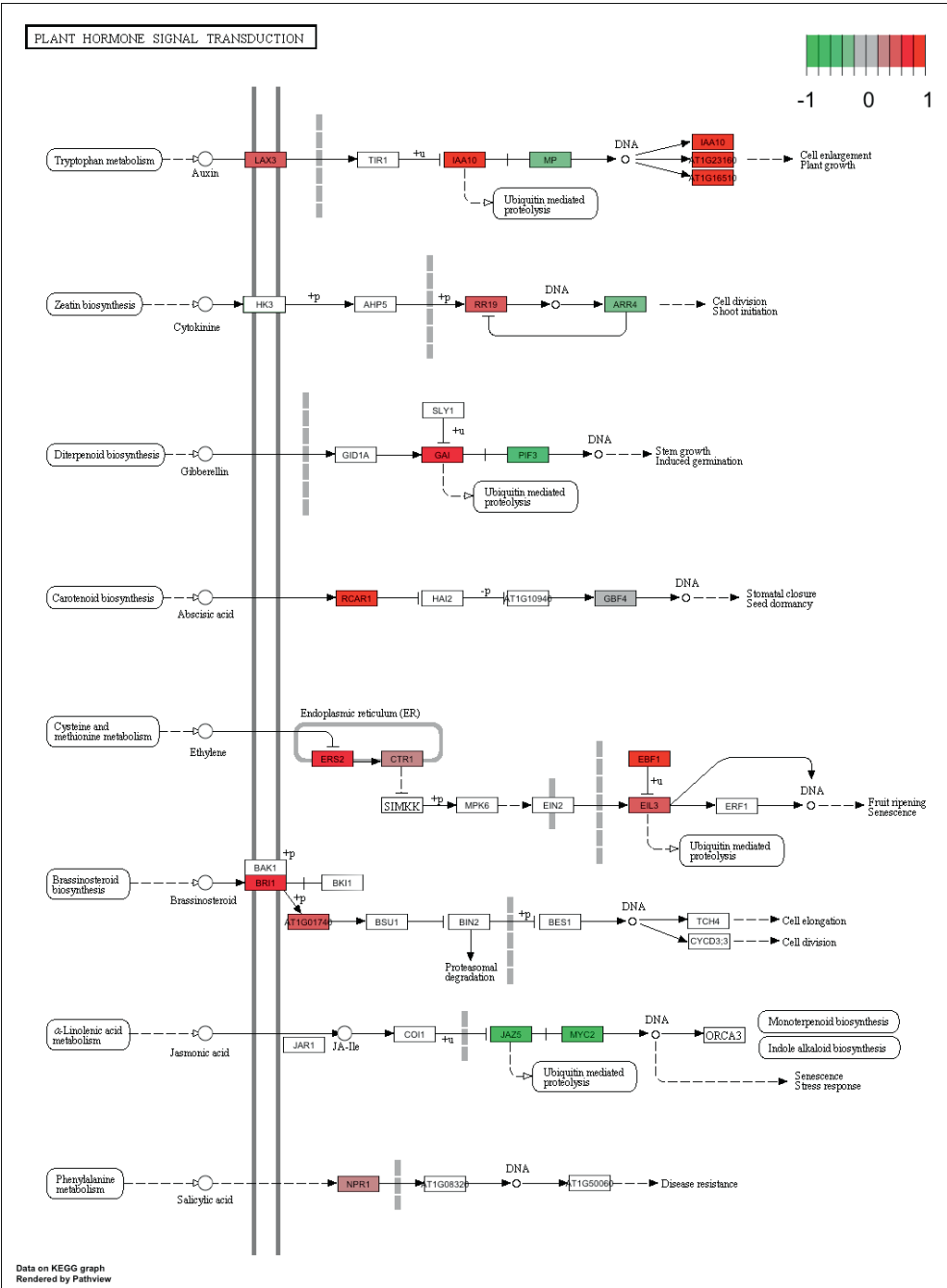


Fig. 9 Overview of several hormone signal transduction pathways of VP16-05-014. Boxes represent gene products (either RNA or protein). Colored boxes indicate differential expression compared to the wild type Col-0. The color legend in the top right corner represents the 2log(fold change) values of gene expression.

Discussion

In this study, we have described the construction and phenotypic characterization of two collections of *Arabidopsis* genome interrogation lines harboring 3F-EAR and 3F-VP16 artificial transcription factor encoding genes. There was extensive variation in the phenotype and RSA of 3F-EAR plants which was accompanied by a multitude of transcriptomic changes, but we were unable to demonstrate a causality between the presence of the 3F-EAR constructs and an increase in RSA. However, in the case of two 3F-VP16 lines we could show that the 3F-VP16 constructs were causal for a significant increase in RSA and biomass through highly 3F-specific transcriptomic changes.

In some studies leaf area was used as a measure of the growth of *Arabidopsis* (e.g. [29, 30]), but this parameter is more difficult to quantify in a high throughput manner and throughout plant development. Leaf area was also found not to be linearly correlated with biomass accumulation [31], making RSA a more suitable measure for growth. RSA is much more easily determined and still provides a reasonable correlation with plant biomass, as also found in our study. Despite that RSA measurements generally provide an underestimate of plant sizes when leaves gradually start overlapping during the later stages of development, it proved to be possible to demonstrate RSA increases in 3F-ATF plant lines, even when analyzing a relatively small number of plants of genetically segregating lines. For all six selected 3F-ATF lines, we observed RSA increases to occur over the course of more than one generation (T2 and T3 for the 3F-EAR lines; T2, T3 and T4 for the 3F-VP16 lines), indicating that they are stably inherited. While almost all RSA increases were highly reproducible, it could sometimes not be demonstrated for VP16-05-014 segregating lines (Fig. S4) and was never found in a VP16-05-014 T3 line that did not segregate for T-DNA presence. This suggests that the enhancement of RSA and biomass induced by the 3F-VP16 construct is gene dosage dependent, which is in accordance with the negative effect on growth when its expression is driven by the 35S promoter (Fig. S7).

We observed extensive variation in rosette phenotype, flowering time and RSA among the 700 primary transformants with 3F-EAR constructs under control of the *RPS5A* promoter (Fig. 1), indicating that 3F-EAR fusions can be dominant triggers of phenotypic variation in *Arabidopsis* and that their expression during early embryogenesis and in meristematic tissue [23] is sufficient for this. Expression of 3F-EAR fusions under control of the constitutive CaMV 35S promoter did not induce as much phenotypic variation in *Arabidopsis* [16], suggesting that the timing rather than the level of transgene expression is most important for the extent of phenotypic variation. Alternatively, the use of a strong and constitutive promoter like 35S might lead to infertility and/or death of plants when driving the expression of very active 3F-ATFs, thus directly or indirectly resulting in a population of plants that is enriched for 3F-ATF fusions with lower activities or limited expression levels due to the

low transcriptional activity of the genomic insertion locus. The use of pathway, cell type or tissue specific promoters for genome interrogation might very well lead to the discovery of novel mutants affected in specific processes of interest. Therefore, the choice of the promoter driving ATF expression for genome interrogation experiments greatly depends on the research question, the organizational level of interest (cell, tissue, organ, organism) and the organism of interest [16].

Our screening strategy for the isolation of 3F-EAR lines with large RSA consisted of first selecting 100 primary transformants which displayed an 80% or higher increase compared to the batch average, followed by a secondary comparison of the RSA of their T2 progeny with Col-0. By choosing this strategy, we aimed to combine the benefits of having RSA data of all 700 primary transformants and making a fair and accurate comparison to Col-0. Of the 100 3F-EAR lines that were at least 80% larger than the batch average, only one out of three displayed an increase in RSA compared to Col-0 in the T2 generation (Fig. 2), albeit mostly lower than 80% (the initial selection criterion) and not statistically significant. The screening of the 3F-VP16 library for lines with RSA phenotypes was more pragmatic and aimed at isolating the largest plants. While doing so we observed relatively little phenotypic and RSA variation among the plants of this library in comparison to the 3F-EAR library. In this case, it might also very well be that the T2 population of 3F-VP16 plants was enriched for 3F-ATF constructs with rather mild effects on the phenotype due to death or infertility of primary transformant (T1) plants harboring 3F-VP16 with lethal activities, thus making the T2 population more uniform. Among the six selected 3F-ATF lines with consistently larger RSA than Col-0 however, the 3F-VP16 lines displayed by far the largest and most conspicuous increases in RSA accompanied by drastic and highly 3F-specific transcriptional changes, demonstrating that the 3F-VP16 library still harbored very interesting lines.

The majority of the retransformant lines harboring T-DNA constructs that were reconstituted from the selected 3F-VP16 lines displayed an increase in RSA compared to Col-0, whereas the increases found for the 3F-EAR lines could not be reproduced by means of retransformation experiments. This might be explained by the fact that 3F-EAR constructs in general tend to decrease the average RSA of Arabidopsis, which is a negative effect on growth that needs to be overcome before any possible increase can be noticed. In that manner, the appearance of 3F-EAR mediated RSA increases would be more challenging and more likely to be dependent on subtle differences in transgene expression levels than for 3F-VP16 mediated increases. In that respect, it was interesting to note that any 3F-EAR expression already triggered a multitude of shared transcriptomic changes, while 3F-VP16 expression in general only had a relatively mild effect. Therefore, the few 3F specific DEGs found in the 3F-EAR expressing lines with increased RSA would then need to compensate for the gross transcriptional changes that lead to a decreased plant size, thereby making the RSA increase rather intricate and difficult to reproduce. In the case of 3F-VP16 lines however, we were able

to clearly provide the proof of principle of genome interrogation through retransformation, and we found a multitude of highly 3F specific transcriptomic changes with relatively little background. In contrast to 3F-EAR constructs most 3F-VP16 plants also do not have growth phenotypes, and the 3F-VP16 fusions therefore seem to induce a highly 3F specific response that can directly increase RSA in a robust and easily reproducible manner. Although it might now be tempting to state that 3F-VP16 fusions are more useful for genome interrogation than 3F-EAR fusions, this is not necessarily the case and really depends on the research question to be answered. In case of an interest in phenotypic variation the 3F-EAR library could be most suitable, whereas the 3F-VP16 lines would be more useful for the elucidation of 3F-ATF induced transcriptomic changes that are causative for a given phenotype.

In VP16-02-003 plants we found differential expression of approximately one quarter of all *Arabidopsis* genes, clearly illustrating the drastic effect that genome interrogation can have on transcription patterns. Among the DEGs there was a significant enrichment for genes involved in a number of metabolic pathways, suggesting that VP16-02-003 achieves an increase in RSA and biomass through the drastic differential regulation of metabolism at the transcriptomic level. For instance, overall down regulation of porphyrin and chlorophyll metabolism was found (Fig. 7), suggesting that VP16-02-003 plants have impaired pigmentation. In support of this hypothesis, we have observed that VP16-02-003 plants can be mildly pale, and we have some preliminary evidence that VP16-02-003 plants have a significantly skewed chlorophyll a/b ratio. Among the DEGs there was also a significant enrichment of genes mediating the essential steps of sulfur assimilation (Fig. 8), such as the downregulation of sulfite reductase (*SiR*), which is an essential factor in assimilatory sulfate reduction [32], thereby possibly compromising sulfur metabolism and leading to sulfur deficiency. Sulfur deficiency has previously been connected to reduced seed viability in plants [33, 34], which we have also observed in VP16-02-003 plants (Fig. S1). Changes in sulfur metabolism have recently been connected to the evolution of C4 photosynthesis [35], which might be positively correlated with biomass accumulation. We have some preliminary evidence that VP16-02-003 plants have a higher operating light use efficiency of Photosystem II (ϕ PSII) at around 15 dpv, and therefore that the increase in RSA and biomass might be driven by more efficient overall photosynthesis [36]. We also observed overall downregulation of glucosinolate biosynthesis genes, which might be a consequence of sulfur deficiency, as glucosinolate degradation is regulated through sulfur deficiency responsive transcription factors [37].

VP16-05-014 plants also displayed substantial differential gene expression, which was enriched for genes involved in the signaling pathways of the hormones auxin, cytokinin and gibberellic acid, all known regulators of plant growth. The auxin signaling pathway of VP16-05-014 might possibly be enhanced through the transcriptional downregulation of *IAA10* (Fig. 9), which is a known repressor of the ARF-mediated auxin response [38]. The cytokinin response is seemingly positively regulated through an enhanced negative feedback loop

between ARR19 and ARR4 (Fig. 9), and thereby possibly promoting cell divisions. Finally, we noticed upregulation of the expression of the DELLA gene *GAI* (Fig. 9), which is a known repressor of the gibberellic acid signaling pathway [39] which, however, would lead to a reduction of growth rather than an increase. Altogether, it seems that VP16-05-014 plants achieve a net increase in RSA and biomass through a transcriptomic interplay between several hormone signaling pathways. In the case of both 3F-VP16 lines however, the phenotypes are very likely to be orchestrated through a complex combination of all of the transcriptomic changes, as we have previously observed in 3F-ATF lines [21] and is accordance with the concept of genome interrogation [16].

Overall, we have isolated Arabidopsis lines with enhanced growth characteristics from both 3F-ATF libraries at relatively high frequencies. The use of 3F-ATFs has thus enabled us to introduce variation in the growth and transcriptome of Arabidopsis, and has provided us with a novel tool to break through the figurative yield barrier of Arabidopsis. It should be noted that we have used RSA and rosette biomass as measures for yield, but other plant properties could also be used as selection criteria for phenotypic screens of the two collections of 3F-ATF lines, such as for instance seed yield. In other screens that we have performed, we have readily found lines with other traits of interest, such as a high rate of homologous DNA recombination [20], early flowering, salinity tolerance, and early flowering combined with increased biomass accumulation at a short photoperiod, indicating that genome interrogation is a potent novel concept for the enhancement of a variety of Arabidopsis traits of interest. In the future our novel knowledge on the pathways that are involved in the growth of Arabidopsis could possibly be used in crop breeding for enhanced yield.

Materials and methods

Growth conditions and plant material

All plants were grown on soil in a climate-controlled growth chamber at a constant temperature of 20 °C, 70% relative humidity, at a light intensity of approximately 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$ of photosynthetically active radiation (PAR), and at a 12 h photoperiod. The plants were watered every 3-4 days. The Arabidopsis accession Columbia-0 (Col-0) was used as wild type and as the background genotype for all transformations described below. The 3F-EAR and 3F-VP16 libraries were constructed as described below.

Construction of the binary vector pRF-EAR-Kana and the collection of 3F-EAR expressing Arabidopsis lines

The amino acid sequence of the EAR (SRDX) domain, LDLDLELRIGFA, was derived from Hiratsu *et al.* [27]. A double stranded oligonucleotide encoding this amino acid sequence

was made (5'-GGTACCGAGGCCAGGCGGCCTCGAGAACTAGTGGCCAGGCC GGCCAATTGGATTGGATTGGAATTGAGATTGGGATTTGCTTAGGAGCTC-3'; codon-optimized for Arabidopsis), preceded by *Sfi*I sites for cloning of 3F domains that were generated in a previous study [20]. The oligo was first cloned into the pJET Blunt cloning vector using the CloneJet PCR Cloning Kit (Thermo Scientific), and after Sanger sequencing (Macrogen Europe, Amsterdam, The Netherlands) the plasmid was digested with *Kpn*I and *Sac*I. The insert was subsequently cloned into the similarly digested plasmid pSDM3835 [40]. The resulting binary vector construct, designated pRF-EAR-Kana, allows for expression of 3F domains preceded by an N-terminal FLAG-tag and an SV40 derived nuclear localization signal (NLS), fused to the C-terminal EAR domain under control of the promoter of the Arabidopsis *RPS5a* gene. The same procedure was followed for plasmid pSDM3838 [40], resulting in an identical coding region under control of the constitutive CaMV 35S promoter, and was designated p35S-EAR-Kana. 3F encoding sequences from subpools 1, 5, 7, 9, 11, 13, and 15 [20] were used for 3F-EAR vector library construction with more than 200 clones per pool in *E. coli*, apart from pools 7 and 13, which only contained about 100 clones. For p35S-EAR-Kana, all pools contained more than 200 clones. Each subpool of pRF-EAR-Kana plasmids encoding 3F-EAR fusions was mobilized to the Agrobacterium strain Agl1 through triparental mating [20]. Arabidopsis Col-0 plants were subsequently transformed separately with each of the subpools of 3F-EAR constructs using the floral dip method [41]. Primary transformants (T1 generation) were selected for on MA medium containing 35 µg/mL kanamycin, 100 µg/mL nystatin and 100 µg/mL Timentin and were transferred to soil after approximately 2 weeks. Plasmid sequences are available upon request.

Rosette surface area analysis of the collection of 3F-EAR lines

Approximately 700 primary transformants harboring 3F-EAR T-DNA constructs were generated as described above (at least 100 primary transformants per ZF subpool). The plants were transferred to soil in 67 x 67 x 65 mm pots (Pöppelmann, Lohne, Germany) after 2 weeks of growth on selection medium. Approximately 28 pots corresponding to the same ZF pool were placed together on 50 x 32.5 x 2.5 cm trays (Bachmann). At one fixed time point in development every tray was photographed from the top with a digital camera (Canon EOS 1100D). Due to the large number of plants that had to be generated, transferred to soil and analyzed, this time point varied per tray. Using an ImageJ plugin, the intensity of the green channel of the resulting RGB images was multiplied by two, both the red and blue channels were subtracted and the image was converted to a binary image using the ImageJ 'Intermodes' Threshold Method. The binary images were manually inspected to ensure that all the leaves of a plant were connected to each other. In cases where leaves were not connected, they were connected manually to the rest of the rosette by a black line of two pixels in width. The rosette surface area (RSA) in pixel² of every plant was subsequently calculated using the 'Analyze

Particles' function of ImageJ. As no fair comparison could be made with wild type Col-0 plants (cannot be selected for with kanamycin), the largest plants were selected by comparing the RSA of each individual to the average RSA of its respective tray. The individuals that had $\geq 80\%$ larger RSA than the tray average were selected for further characterization of growth, which narrowed the number of plant lines for subsequent analysis down to approximately 100. These individuals were allowed to complete their life cycle and their seeds (T2 progeny) were harvested.

For RSA quantification throughout development of the selected 100 lines, approximately 50 T2 seeds of each genotype and 200 for Col-0 were sown on soil in separate pots with a diameter of 15.7 cm and height of 6.5 cm (Soparco, Condé-sur-Huisne, France) and stratified for 3 days at 4 °C. At 7 dpv, the largest individuals of each 3F-EAR line ($n=7$) and Col-0 ($n=100$) were transferred to soil in individual 67 x 67 x 65 mm pots (Pöppelmann, Lohne, Germany). In this manner, it was ensured that the growth of possibly larger 3F-EAR plants was always compared to the growth of the largest Col-0 individuals. From 10 dpv onwards and every 3 days, photos were taken from the top with a digital camera (Canon EOS 1100D). The resulting RGB images were processed into binary images as described above, and from pixel² converted to mm² using the 'Analyze Particles' function of ImageJ by multiplying with the mm²/pixel² ratio of every RGB image, respectively. This ratio was calculated from the dimensions of the pots that the plants were growing in, as this value is constant throughout the experiment. From the 100 3F-EAR lines for which this analysis was performed, the lines with $\geq 20\%$ larger RSA than Col-0 at 25 dpv were selected and subjected to a secondary quantification of RSA throughout development as is described below.

Construction of the collection of 3F-VP16 lines

A library of approximately 3500 3F-VP16 fusion encoding T-DNA constructs was constructed in the binary vector pRF-VP16-Kana, and was mobilized to the *Agrobacterium* strain Agl1 through triparental mating, as described previously [20]. Col-0 plants were subsequently transformed with the 3F-VP16 constructs using the floral dip method [41]. Primary transformants (T1) were selected for on MA medium containing 35 µg/mL kanamycin, 100 µg/mL nystatin and 100 µg/mL Timentin and were transferred to soil after approximately 2 weeks. Five primary transformants originating from the same subpool were placed together in a pot, and their seeds were combinedly harvested and stored in seed bags named 'five-bags'. A fraction of the five-bags contained less complicated seed mixtures (originating from 3 or 4 primary transformant rather than 5) due to losses of plants during cultivation and occasional infertility.

Screening of the 3F-VP16 library for plants with surface area phenotypes

Approximately 20 seeds from each five-bag of the 3F-VP16 seed library were sown together on soil in a single square 67 x 67 x 65 mm pot (Pöppelmann, Lohne, Germany). The pots were divided over 50 x 32.5 x 2.5 cm trays (Bachmann), with every tray receiving 23 pots with transgenic seeds and one pot with approximately 20 wild type seeds. The seeds were stratified on soil for 3 days at 4 °C, and every pot was inspected by eye at 14 dpv to identify plants with large rosettes compared to Col-0. These individuals were isolated, transferred to fresh soil and allowed to set seeds, which were harvested.

Transformation with reconstituted T-DNA constructs

The 3F-EAR or 3F-VP16 encoding DNA fragments were PCR amplified from the genomic DNA of the selected 3F-ATF lines with increased RSA using a forward primer within the *RPS5a* promoter sequence (5'-GCCCAAACCCTAAATTTCTCATC-3') and a reverse primer within the *NOS* terminator sequence of the T-DNA (5'-CAAGACCGGCAACAGGAT-3'). Reconstituted pRF-EAR-Kana and pRF-VP16-Kana binary vectors were generated with the 3F fragments from the PCR products [20]. The binary vector p35S-VP16-Kana was obtained by introduction of the CaMV 35S promoter sequence as a *XmaI*-*SacI* fragment into similarly digested pRF-VP16-Kana. Reconstituted p35S-VP16-Kana binary vectors were generated with the 3F fragments from the PCR products mentioned above [20]. Arabidopsis Col-0 plants were transformed with the reconstituted T-DNA constructs using the floral dip method [41]. Primary retransformants were selected by plating sterilized seeds on MA medium containing 35 µg/mL kanamycin, 100 µg/mL nystatin and 100 µg/mL Timentin, transferred to soil after 2-3 weeks and allowed to set seeds, which were harvested.

RSA and biomass quantification of selected 3F-ATF lines with increased RSA and retransformant lines harboring reconstituted 3F-ATF constructs

Approximately 50 seeds of the selected 3F-ATF lines (T2 for 3F-EAR and T3 for 3F-VP16; segregating except for VP16-02-003), the retransformant lines reconstituted from these lines (T2; segregating) and Col-0 were sown on soil in pots with a diameter of 15.7 cm and height of 6.5 cm (Soparco, Condé-sur-Huisne, France) and stratified for 3 days at 4 °C. The largest Col-0, 3F-ATF and retransformant individuals were selected at 7 dpv and were each transferred to fresh soil in separate 67 x 67 x 65 mm pots that were placed on 50 x 32.5 x 2.5 cm trays, again ensuring that the growth of possibly larger transgenic individuals was always compared to the growth of the largest Col-0 individuals. Every tray contained 18 transgenic individuals and 6 wild type individuals. As every individual plant in such a grid of 24 pots is growing in the vicinity of either two (~ 17%), three (50%) or four (~33%) neighboring individuals, Col-0 plants were positioned in such a manner that these conditions were correspondingly represented among the Col-0 population. If every row of pots were assigned a letter (A-D),

and every column a number (1-6), wild type individuals were grown at positions A3, B1, B5, C3, D1 and D5, respectively. From 10 dpg onwards and every 3 days, photos were taken from the top and RSA was calculated in mm² from RGB images using the ImageJ protocol described above. The relative RSA compared to Col-0 was calculated by dividing the RSA of every individual by the average RSA of Col-0 and multiplying by 100%. For the quantification of biomass, the shoots of each plant were harvested at 28 dpg and fresh weight was determined. Dry weights were determined after 2 days of incubation at 60 °C. The relative fresh and dry weights compared to Col-0 were calculated by dividing the fresh and dry weights of every individual by the average of Col-0 and multiplying by 100%. All relative surface area, fresh weight and dry weight data were statistically analyzed for significant differences with Col-0 using the heteroscedastic T-Test function of Microsoft Excel 2010 (assuming unequal variance between samples). A *p*-value of 0.05 was used as a threshold for significance.

RNA extraction and sequencing

Approximately 100 seeds of Col-0, the selected 3F-ATF lines with larger RSA than Col-0, and a mixture of lines originating from the same 3F pools as those lines, but which do not have noticeably increased RSA (referred to as 'background') were sown on soil in separate pots with a 15.7 cm diameter and a height of 65 mm (Soparco, Condé-sur-Huisne, France) and stratified for 3 days at 4 °C. At 7 dpg, the largest 36 individuals of each of each genotype were selected and were transferred to individual 67 x 67 x 65 mm pots (Pöppelmann, Lohne, Germany) in three replicates with 12 individuals per replicate, ensuring that the transcriptomes of possibly larger transgenic individuals were compared to the transcriptomes of the largest wild type individuals. All pots were randomly distributed over trays in the greenhouse to mitigate the effect of local variation in growth conditions on the transcriptome data. The above ground parts of each replicate of 12 plants were combined and grinded to powder in liquid nitrogen with pistil and mortar at 15 dpg. Total RNA was extracted from approximately 50 mg of tissue powder of each replicate using the RNA Plant Mini Kit (Qiagen), and sequenced with Illumina HiSeq (50 bp single reads).

RNA sequencing data analysis

A sequencing library was constructed and sequenced by Illumina sequencing (50 cycles; single read). Genomic reference sequences and annotations were obtained from TAIR (version TAIR10) and supplemented with fragments mapping to VP16 from pRF-VP16-Kana and EAR from pRF-EAR-Kana. The splicing-aware aligner TopHat (version 2.0.10, [42]) was used to map reads, using the 'very-sensitive' and 'coverage-search' options, and allowing for a maximum intron size of 15000 bp. Secondary alignments were removed from the BAM files using SAMtools (version 0.1.18, [43]) and Perl. Reads aligning to annotated exons were summarized at the level of TAIR genes using HTSeq (version 0.5.3p9, [44]) using

the ‘intersection-strict’ setting. At least 91% of raw sequencing reads could be uniquely assigned to a gene. Read counts were processed in R (version 3.0.2) using the edgeR package (version 3.4.2, [45]). Normalized expression values per gene (excluding mitochondrial and chloroplast sequences) were obtained by scaling using a robust estimate of the library size [46] and dividing by the mean length of the annotated transcripts in kbp. Differentially expressed genes (DEGs) between Col-0, the background, and the 3F-ATF lines were determined using the Bioconductor R package DESeq2 [47]. Genes were considered DEGs if their *p*-value adjusted by the Benjamini-Hochberg method was lower than 0.05. Active gene expression was distinguished from background noise using the method of Hart *et al.* [48]. The resulting list of DEGs was used as input for KEGG pathway enrichment analysis using the Bioconductor R package KEGGprofile [49]. The KEGG pathways were visualized with the R package Pathview [50].

Acknowledgements

We would like to thank Prof. Dr. H.J.M. de Groot and Dr. A. Alia Matysik (Leiden Institute of Chemistry, Leiden, the Netherlands) for stimulating discussions. This work was carried out within the research programme of BioSolar Cells, co-financed by the Dutch Ministry of Economic Affairs, and was cofinanced by the Consortium for Improving Plant Yield (CIPY), which is part of the Netherlands Genomics Initiative/ Netherlands Organization for Scientific Research. A separate CIPY grant for RNA sequencing was received for an Enabling Technology Platform Hotel project. We would like to thank Dr. Gabino Sanchez-Perez (Bioinformatics, Wageningen UR, The Netherlands) for advice on the set up of the RNA sequencing experiment, and Dr. Eric van der Graaff (Section of Crop Science, Department of Plant and Environmental Sciences (PLEN), Copenhagen University, Denmark) for helpful discussions and advice regarding the RNA sequencing data analysis.

References

1. Zhu XG, Long SP, Ort DR. (2010) Improving photosynthetic efficiency for greater yield. *Annu Rev Plant Biol* 61:235-61.
2. Wiebe K. (2008) The State of Food and Agriculture. Biofuels: Prospects, Risks and Opportunities. *Food and Agriculture Organization of the United Nations, Rome, Italy*.
3. Peng SB, Tang QY, Zou YB. (2009) Current Status and Challenges of Rice Production in China. *Plant Prod Sci* 12(1):3-8.
4. Evans JR. (2013) Improving photosynthesis. *Plant Physiol* 162(4):1780-93.
5. Barber J. (2009) Photosynthetic energy conversion: natural and artificial. *Chem Soc Rev* 38(1):185-96.
6. Long SP, Ainsworth EA, Leakey AD, Nosberger J, Ort DR. (2006) Food for thought: lower-than-expected crop yield stimulation with rising CO₂ concentrations. *Science* 312(5782):1918-21.
7. Gonzalez N, Beemster GT, Inze D. (2009) David and Goliath: what can the tiny weed *Arabidopsis* teach us to improve biomass production in crops? *Curr Opin Plant Biol* 12(2):157-64.
8. Gonzalez N, Vanhaeren H, Inze D. (2012) Leaf size control: complex coordination of cell division and expansion. *Trend Plant Sci* 17(6):332-40.
9. Choe S, Fujioka S, Noguchi T, Takatsuto S, Yoshida S, Feldmann KA. (2001) Overexpression of DWARF4 in the brassinosteroid biosynthetic pathway results in increased vegetative growth and seed yield in *Arabidopsis*. *Plant J* 26(6):573-82.
10. Horvath BM, Magyar Z, Zhang YX, Hamburger AW, Bako L, Visser RG (2006) EBP1 regulates organ size through cell growth and proliferation in plants. *Embo J* 25(20):4909-20.
11. Century K, Reuber TL, Ratcliffe OJ. (2008) Regulating the regulators: The future prospects for transcription-factor-based agricultural biotechnology products. *Plant Physiol* 147(1):20-9.
12. Meyer RC, Torjek O, Becher M, Altmann T. (2004) Heterosis of biomass production in *Arabidopsis*. Establishment during early development. *Plant Physiol* 134(4):1813-23.
13. Grossmann M, Gonzalez-Bayon R, Greaves IK, Wang L, Huen AK, Peacock WJ (2014) Intraspecific *Arabidopsis* hybrids show different patterns of heterosis despite the close relatedness of the parental genomes. *Plant Physiol* 166(1):265-80.
14. Andorf S, Meyer RC, Selbig J, Altmann T, Repsilber D. (2012) Integration of a systems biological network analysis and QTL results for biomass heterosis in *Arabidopsis thaliana*. *PloS One* 7(11):e49951.
15. Barth S, Busimi AK, Friedrich Utz H, Melchinger AE. (2003) Heterosis for biomass yield and related traits in five hybrids of *Arabidopsis thaliana* L. Heynh. *Heredity* 91(1):36-42.
16. van Tol N, van der Zaal BJ. (2014) Artificial transcription factor-mediated regulation of gene expression. *Plant Sci* 225:58-67.
17. Lee JY, Sung BH, Yu BJ, Lee JH, Lee SH, Kim MS (2008) Phenotypic engineering by reprogramming gene transcription using novel artificial transcription factors in *Escherichia coli*. *Nucleic Acids Res* 36(16):e102.
18. Park KS, Jang YS, Lee H, Kim JS. (2005) Phenotypic alteration and target gene identification using combinatorial libraries of zinc finger proteins in prokaryotic cells. *J Bacteriol* 187(15):5496-9.
19. Park KS, Lee DK, Lee H, Lee Y, Jang YS, Kim YH (2003) Phenotypic alteration of eukaryotic cells using randomized libraries of artificial transcription factors. *Nat Biotechnol* 21(10):1208-14.
20. Lindhout BI, Pinas JE, Hooykaas PJ, van der Zaal BJ. (2006) Employing libraries of zinc finger artificial transcription factors to screen for homologous recombination mutants in *Arabidopsis*. *Plant J* 48(3):475-83.

21. Jia Q, van Verk MC, Pinas JE, Lindhout BI, Hooykaas PJ, van der Zaal BJ. (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 Biotech J* 11(9):1069-79.
22. Segal DJ, Dreier B, Beerli RR, Barbas CF, 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(6):2758-63.
23. Weijers D, Franke-van Dijk M, Vencken RJ, Quint A, Hooykaas P, Offringa R. (2001) An Arabidopsis Minute-like phenotype caused by a semi-dominant mutation in a RIBOSOMAL PROTEIN S5 gene. *Development* 128(21):4289-99.
24. Sadowski I, Ma J, Triezenberg S, Ptashne M. (1988) GAL4-VP16 is an unusually potent transcriptional activator. *Nature* 335(6190):563-4.
25. Ohta M, Matsui K, Hiratsu K, Shinshi H, Ohme-Takagi M. (2001) Repression domains of class II ERF transcriptional repressors share an essential motif for active repression. *Plant Cell* 13(8):1959-68.
26. Hiratsu K, Ohta M, Matsui K, Ohme-Takagi M. (2002) The SUPERMAN protein is an active repressor whose carboxy-terminal repression domain is required for the development of normal flowers. *FEBS Lett* 514(2-3):351-4.
27. Hiratsu K, Matsui K, Koyama T, 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(5):733-9.
28. Leister D, Varotto C, Pesaresi P, Niwergall A, Salamini F. (1999) Large-scale evaluation of plant growth in Arabidopsis thaliana by non-invasive image analysis. *Plant Physiol Bioch* 37(9):671-8.
29. Gonzalez N, De Bodt S, Sulpice R, Jikumaru Y, Chae E, Dhondt S, *et al.* (2010) Increased leaf size: different means to an end. *Plant Physiol* 153(3):1261-79.
30. Gonzalez N, Pauwels L, Baekelandt A, De Milde L, Van Leene J, Besbrugge N, *et al.* (2015) A Repressor Protein Complex Regulates Leaf Growth in Arabidopsis. *Plant Cell* 27(8):2273-87.
31. Weraduwege SM, Chen J, Anozie FC, Morales A, Weise SE, Sharkey TD. (2015) The relationship between leaf area growth and biomass accumulation in Arabidopsis thaliana. *Front Plant Sci* 6:167.
32. Khan MS, Haas FH, Samami AA, Gholami AM, Bauer A, Fellenberg K, *et al.* (2010) Sulfite reductase defines a newly discovered bottleneck for assimilatory sulfate reduction and is essential for growth and development in Arabidopsis thaliana. *Plant Cell* 22(4):1216-31.
33. Zuber H, Poignavent G, Le Signor C, Aime D, Vieren E, Tadla C, *et al.* (2013) Legume adaptation to sulfur deficiency revealed by comparing nutrient allocation and seed traits in Medicago truncatula. *Plant J* 76(6):982-96.
34. Sorin E, Etienne P, Maillard A, Zamarreno AM, Garcia-Mina JM, Arkoun M, *et al.* (2015) Effect of sulphur deprivation on osmotic potential components and nitrogen metabolism in oilseed rape leaves: identification of a new early indicator. *J Exp Bot* 66(20):6175-89.
35. Weckopp SC, Kopriva S. (2014) Are changes in sulfate assimilation pathway needed for evolution of C4 photosynthesis? *Front Plant Sci* 5:773.
36. Baker NR. (2008) Chlorophyll fluorescence: a probe of photosynthesis in vivo. *Annu Rev Plant Biol* 59:89-113.
37. Wawrzynska A, Sirko A. (2014) To control and to be controlled: understanding the Arabidopsis SLIM1 function in sulfur deficiency through comprehensive investigation of the EIL protein family. *Front Plant Sci* 5:575.
38. Rademacher EH, Lokerse AS, Schlereth A, Llavata-Peris CI, Bayer M, Kientz M, *et al.* (2012) Different auxin response machineries control distinct cell fates in the early plant embryo. *Dev Cell* 22(1):211-22.

39. Claeys H, De Bodt S, Inze D. (2014) Gibberellins and DELLAs: central nodes in growth regulatory networks. *Trends Plant Sci* 19(4):231-9.
40. de Pater S, Neuteboom LW, Pinas JE, Hooykaas PJ, van der Zaal BJ. (2009) ZFN-induced mutagenesis and gene-targeting in Arabidopsis through Agrobacterium-mediated floral dip transformation. *Plant Biotech J* 7(8):821-35.
41. Clough SJ, Bent AF. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J* 16(6):735-43.
42. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14(4).
43. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078-9.
44. Anders S, Pyl PT, Huber W. (2015) HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2):166-9.
45. Robinson MD, McCarthy DJ, Smyth GK. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 26(1):139-40.
46. Anders S, Huber W. (2010) Differential expression analysis for sequence count data. *Genome Biol* 11(10). doi: ARTN R106
47. Love MI, Huber W, Anders S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):550.
48. Hart T, Komori HK, LaMere S, Podshivalova K, Salomon DR. (2013) Finding the active genes in deep RNA-seq gene expression studies. *BMC Genom* 14:778.
49. Zhao S SY. (2014) KEGGprofile: an Annotation and Visualization Package for Multi-Groups Expression Data in KEGG Pathways.
50. Luo W, Brouwer C. (2013) Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* 29(14):1830-1.

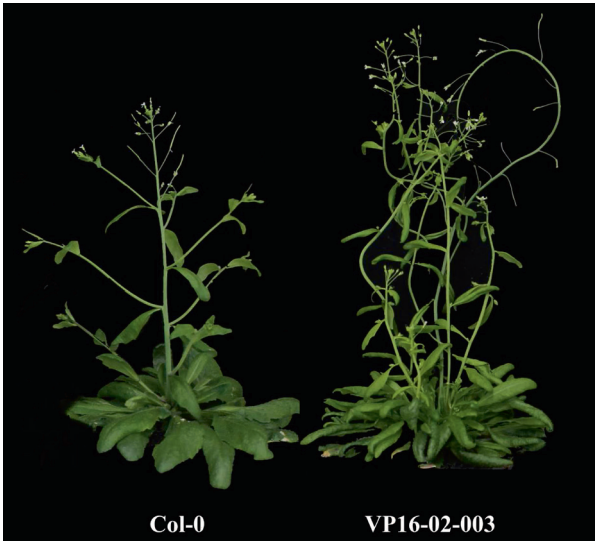


Fig. S1 An overview of the inflorescence phenotypes of the wild type Col-0 and of VP16-02-003 (T3 generation) at 65 dpv.

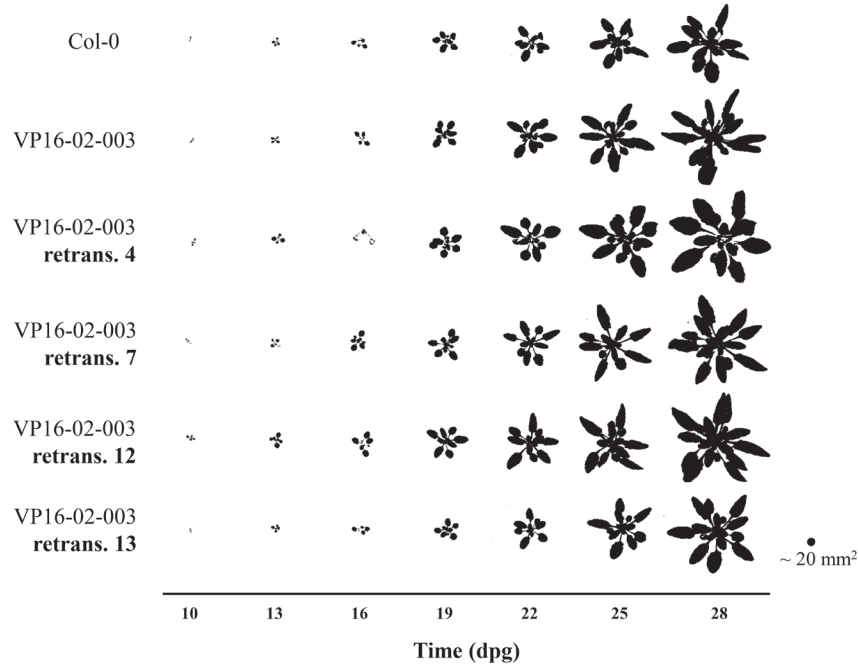


Fig. S2 An overview of the development of wild type Col-0 plants, VP16-02-003 plants (T3) and plants of the retransformant lines reconstituted from VP16-02-003 (T2) that have significantly larger RSA than Col-0. The presented individuals had RSA values closest to the average of that particular genotype at 28 dpv, thereby representing the average development of their respective genotypes in terms of RSA.

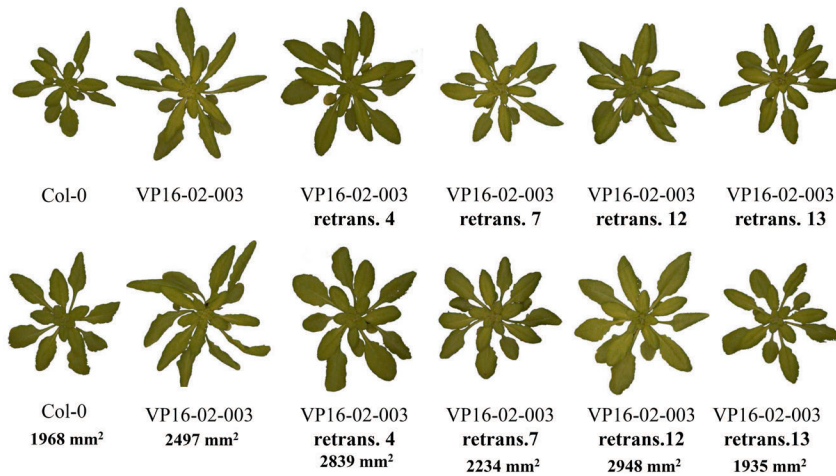


Fig. S3 A to scale overview of the phenotypes and sizes of wild type Col-0 plants, VP16-02-003 plants (T3) and plants of retransformant lines reconstituted from VP16-02-003 that have significantly larger RSA than the wild type Col-0 (T2; segregating). **A**) Representative individuals of Col-0 (out of 48 plants), VP16-02-003 and retransformant lines harboring a T-DNA construct reconstituted from VP16-02-003. The transgenics that are larger than Col-0 have slightly lanceolate shaped leaves and exhibit an increase in the number of leaves (28 dpv). **B**) The largest individuals of the indicated genotypes in terms of RSA among the analyzed population of plants at 28 dpv.

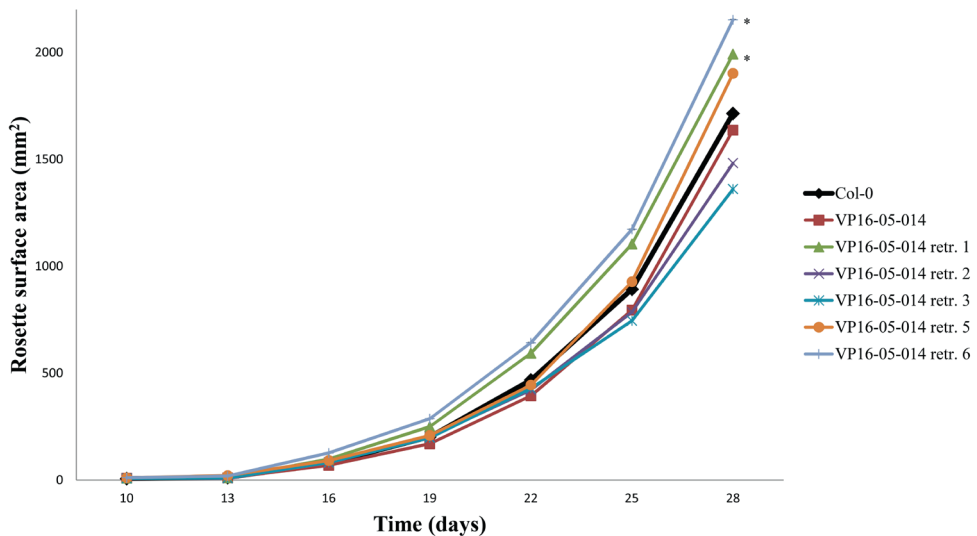


Fig. S4 Growth curves of the wild type Col-0, VP16-05-014 (T3; segregating) and retransformants reconstituted from VP16-05-014 (T2; segregating) (n=36 for Col-0, n=18 for the other genotypes). From approximately 22 dpv onwards, overlapping of leaves started to occur, causing an increasing underestimation of RSA for larger plants for which corrections could not be made. Significant differences with Col-0 at 28 dpv are indicated by an * ($p < 0.05$). In this experiment we were not able to reproduce the increase in RSA of VP16-05-014.

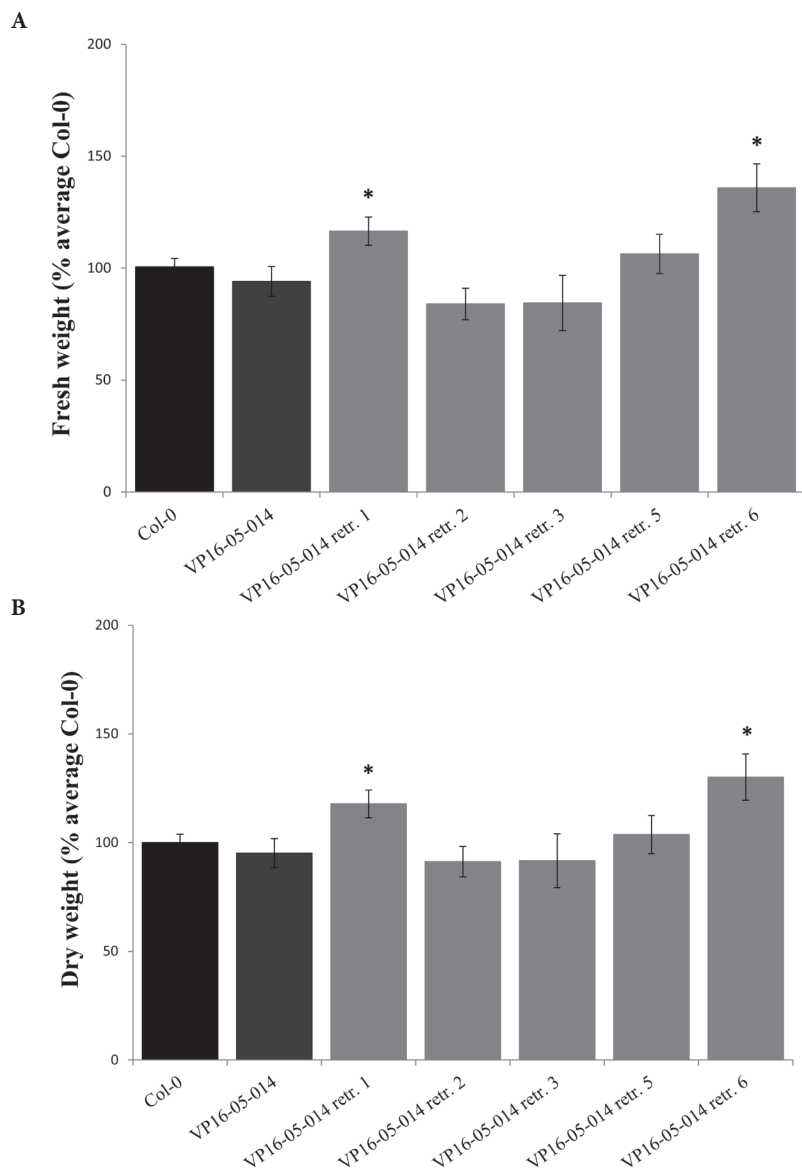


Fig. S5 Quantification of the relative fresh weight (**A**) and dry weight (**B**) of wild type Col-0 plants, VP16-05-014 plants (T3; segregating) and retransformant plants reconstituted from VP16-05-014 (T2; segregating) compared to the wild type Col-0 (28 dpv). The fresh and dry weights of plants of the indicated genotypes was calculated in terms of percentage of the average of Col-0. Error bars represent SEM values ($n=36$ for Col-0, $n=18$ for the other genotypes). Significant differences with Col-0 are indicated by an * ($p < 0.05$). In this experiment we were not able to reproduce the increase in biomass of VP16-05-014.

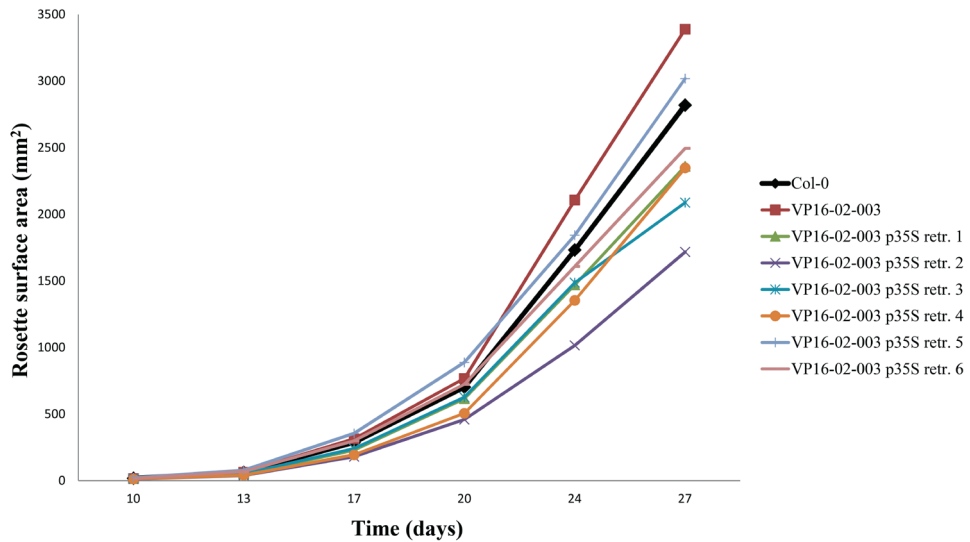


Fig. S6 Growth curves of the wild type Col-0, VP16-02-003 (T3) and retransformants reconstituted from VP16-02-003 in the binary vector construct p35S-VP16-Kana (T2; segregating) (n=84 for Col-0, n=11-18 for the transgenic lines). From approximately 22 dpg onwards, overlapping of leaves started to occur, causing an increasing underestimation of RSA for larger plants for which corrections could not be made.

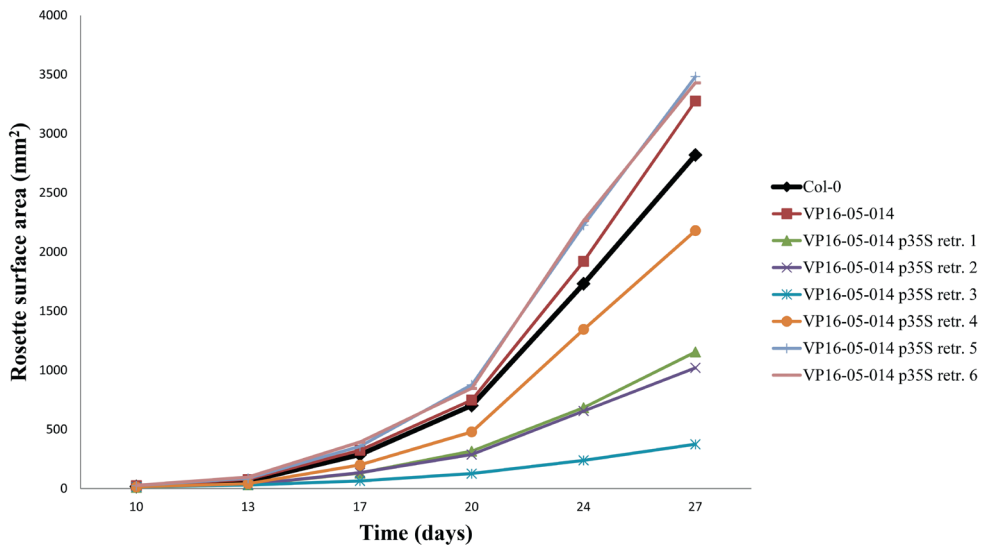


Fig. S7 Growth curves of the wild type Col-0, VP16-05-014 (T3; segregating) and retransformants reconstituted from VP16-05-014 in the binary vector construct p35S-VP16-Kana (T2; segregating) (n=84 for Col-0, n=10-18 for the transgenic lines). From approximately 22 dpg onwards, overlapping of leaves started to occur, causing an increasing underestimation of RSA for larger plants for which corrections could not be made.

Table S1 Overview of differentially expressed genes (DEGs) compared to the wild type Col-0 in the RNA sequencing data sets of the indicated 3F-EAR transgenic lines ($p < 0.05$). ‘Background’ refers to RNA expression data derived from the pool of lines expressing 3F-EAR fusions similar to the specific 3F-EAR fusion expressed in the selected lines, but without a noticeable increase in RSA.

Genotype	DEGs <u>without</u> subtraction of background			DEGs <u>with</u> subtraction of background		
	Total	Down-regulated	Up-regulated	Total	Down-regulated	Up-regulated
EAR-13-068	2598	1346	1252	1346	613	733
EAR-15-025	1371	732	639	557	251	306
EAR-15-053	3007	1580	1427	1543	782	761
EAR pool 13 background	2365	1304	1061	-	-	-
EAR pool 15 background	2165	1166	999	-	-	-

Table S2 Overview of significantly enriched KEGG pathways among differentially expressed genes (DEGs) of the indicated 3F-EAR lines compared to the wild type Col-0 ($p < 0.05$). ‘Background’ refers to RNA expression data derived from the pool of lines expressing 3F-EAR fusions similar to the specific 3F-EAR fusion expressed in the selected lines, but without a noticeable increase in RSA.

Genotype	KEGG pathway ID	KEGG pathway name	Number of genes in pathway	DEGs in pathway (% of genes in pathway)	
				<u>Without</u> subtraction of background	<u>With</u> subtraction of background
EAR-13-068	04712	Circadian rhythm	29	14 (48%)	
	04146	Peroxisome	61	15 (25%)	
	01110	Biosynthesis of secondary metabolites	790	146 (18%)	70 (9%)
	01100	Metabolic pathways	1554	257 (17%)	129 (8%)
	00941	Flavonoid biosynthesis	19	10 (53%)	5 (26%)
	00910	Nitrogen metabolism	43	16 (37%)	
	00906	Carotenoid biosynthesis	24	11 (46%)	
	00760	Nicotinate and nicotamide metabolism	11	5 (45%)	
	00710	Carbon fixation	78	18 (23%)	12 (15%)
	00592	Alpha-linolenic acid metabolism	27	10 (37%)	
	00480	Glutathione metabolism	62	15 (24%)	
	00400	Phenylalanine, tyrosine and tryptophan biosynthesis	43	15 (35%)	8 (19%)
	00350	Tyrosine metabolism	26		
	00340	Histidine metabolism	19	8 (42%)	
	00260	Glycine, serine and threonine metabolism	47	14 (30%)	
	00250	Alanine, aspartate and glutamate metabolism	47	13 (28%)	
	00130	Ubiquinone and other terpenoid-quinone biosynthesis	24	11 (46%)	

EAR-15-025	04712	Circadian rhythm	29	9 (31%)	
	01110	Biosynthesis of secondary metabolites	790	77 (10%)	
	01100	Metabolic pathways	1554	145 (9%)	63 (4%)
	00910	Nitrogen metabolism	43	10 (23%)	5 (12%)
	00906	Carotenoid biosynthesis	24	7 (29%)	
	00860	Porphyrin and chlorophyll metabolism	40	9 (23%)	
	00592	Alpha-linolenic acid metabolism	27	6 (22%)	
	00350	Tyrosine metabolism	26	7 (27%)	
	00340	Histidine metabolism	19	5 (26%)	
	00260	Glycine, serine and threonine metabolism	47	10 (21%)	6 (13%)
EAR-15-053	01110	Biosynthesis of secondary metabolites	70	134 (17%)	
	01100	Metabolic pathways	1554	237 (15%)	
	00130	Ubiquinone and other terpenoid-quinone biosynthesis	24	10 (42%)	

Table S3 Overview of differentially expressed genes (DEGs) compared to the wild type Col-0 in the RNA sequencing data sets of the indicated 3F-VP16 lines ($p < 0.05$). ‘Background’ refers to RNA expression data derived from the pool of lines expressing 3F-VP16 fusions similar to the specific 3F-VP16 fusion expressed in the selected lines, but without a noticeable increase in RSA.

Genotype	DEGs <u>without</u> subtraction of background			DEGs <u>with</u> subtraction of background		
	Total	Down-regulated	Up-regulated	Total	Down-regulated	Up-regulated
VP16-02-003	4889	2437	2452	4535	2258	2277
VP16-05-014	1753	815	938	1633	769	864
VP16 pool 2 background	688	259	429	-	-	-
VP16 pool 5 background	387	267	120	-	-	-

Table S4 Overview of significantly enriched KEGG pathways among differentially expressed genes (DEGs) of the indicated 3F-VP16 lines compared to the wild type Col-0 ($p < 0.05$). 'Background' refers to RNA expression data derived from the pool of lines expressing 3F-VP16 fusions similar to the specific 3F-VP16 fusion expressed in the selected lines, but without a noticeable increase in RSA.

	KEGG pathway ID	KEGG pathway name	Number of genes in pathway	DEGs in pathway (% of genes in pathway)	
				<u>Without</u> subtraction of background	<u>With</u> subtraction of background
Genotype					
VP16-02-003	04712	Circadian rhythm	29	10 (34%)	
	04626	Plant-pathogen interaction	148	47 (32%)	43 (29%)
	01110	Biosynthesis of secondary metabolites	790	208 (26%)	205 (26%)
	01100	Metabolic pathways	1554	334 (21%)	325 (21%)
	00966	Glucosinolate biosynthesis	18	9 (50%)	9 (50%)
	00941	Flavonoid biosynthesis	19	8 (42%)	8 (42%)
	00920	Sulfur metabolism	34	15 (44%)	15 (44%)
	00860	Porphyrin and chlorophyll metabolism	40	18 (45%)	18 (45%)
	00592	Alpha-linolenic acid metabolism	27	12 (44%)	
	00480	Glutathione metabolism	62	21 (34%)	21 (34%)
	00400	Phenylalanine, tyrosine, tryptophan biosynthesis	43	14 (33%)	14 (33%)
	00380	Tryptophan metabolism	38	13 (34%)	13 (34%)
	00340	Histidine metabolism	19	8 (42%)	8 (42%)
	00290	Valine, leucine and isoleucine biosynthesis	36	14 (39%)	14 (39%)
	00071	Fatty acid metabolism	41	13 (32%)	13 (32%)
VP16-05-014	04712	Circadian rhythm	29	5 (17%)	5 (17%)
	04075	Plant hormone signal transduction	232	10 (4%)	10 (4%)
	01100	Metabolic pathways	1554	32 (2%)	30 (2%)
	00592	Alpha-linolenic acid metabolism	27	5 (19%)	

Chapter 4

Chloroplast genome interrogation in Arabidopsis seedlings

Niels van Tol, Vera Veltkamp, Johan Pinas,
Paul J.J. Hooykaas and Bert J. van der Zaal

Abstract

The large majority of core photosynthesis proteins in plants are encoded by nuclear genes, but a small portion has been retained in the plastid genome. Here, we report about the use of nuclear encoded, chloroplast targeted zinc finger artificial transcription factors (ZF-ATFs) to modulate the transcription patterns of chloroplast genes, a technique designated chloroplast genome interrogation. Using this system, we obtained evidence that ZF-ATFs can be translocated to chloroplasts, can induce phenotypic changes and can influence the operating light use efficiency of Photosystem II. Our data suggest that the distortion of chloroplast gene expression patterns might be a feasible approach to manipulate the efficiency of plant photosynthesis.

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 core of the photosynthetic apparatus. Sunlight is absorbed by chlorophyll molecules that are associated with Photosystems I and II (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 Cytochrome 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 CO₂ 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 number of 4500 bacterial genes has been incorporated into the nuclear genomes of plants [1]. These genes have acquired eukaryotic gene expression signals and in many cases sequences encoding N-terminal signal peptides known as chloroplast transit peptides (CTPs), which mediate chloroplast import [2]. Remarkably, a small but significant portion of the bacterial genes has 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. Chloroplastic 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 [3]. The chloroplast genome of the model plant species *Arabidopsis thaliana* encodes 54 structural thylakoid membrane proteins, 31 proteins involved in the regulation of plastid gene expression and contains 45 tRNA and rRNA encoding genes [4].

The engineering of chloroplast genes has been designated as one of the targets for increasing the efficiency of photosynthesis in plants [5]. 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 [6] to introduce gene constructs into target chloroplast genomic loci through homologous DNA recombination, but it is typically very tedious to generate homoplasmic plants with stably transgenic chloroplasts, especially in tissue in which the chloroplast genome is maintained in very high copy numbers. Another pitfall of this approach is that mutation of the core thylakoid membrane components usually results in impairment rather than gain of thylakoid membrane function. To our knowledge, no chloroplast mutants with enhanced photochemistry have been reported. We therefore

hypothesized that changing the stoichiometry rather than the structure of chloroplast encoded proteins might be an alternative and more promising approach to influence thylakoid membrane function. In order to change the stoichiometry of chloroplast encoded thylakoid membrane proteins, we considered artificial transcription factor (ATF)-mediated genome interrogation a suitable method.

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 [7]. In our lab, we have successfully used ATFs with zinc fingers (ZFs) as DNA binding domains (ZF-ATFs) for genome interrogation experiments in *Arabidopsis* [8-10]. In our setup, the ZF-ATFs contained an array of three of the 16 different ZFs that can recognize a cognate 3 base pair (bp) consensus DNA sequence of 5'-GNN-3' [11], 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 [12-16]. 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 1000 times within the 130 Mbp *Arabidopsis* genome. In this way, the expression of nearby genes might be distorted by 3F-ATFs *in trans* and in a dominant manner, potentially leading to the differential expression of a large number of genes, which in turn might trigger novel phenotypes to arise.

In the present study we explored the use of ZF-ATFs in chloroplasts. To change chloroplast gene expression patterns it had to be taken into account that chloroplasts have also retained their own transcriptional and regulatory machinery [17], 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. The process of plastid gene expression is also tightly regulated through anterograde and retrograde signaling with the nucleus [18, 19]. In view of these considerations, we had to redesign our previously established genome interrogation setup in such a way that ZF-ATFs can function in a 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 2F DNA binding domains. We obtained evidence that a very small collection of ZF-ATFs already contained constructs that induced variation in the phenotype and operating light use efficiency of PSII reaction centers of *Arabidopsis* seedlings, indicating 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 we had no guarantee that established modulators of eukaryotic gene expression could also function as such within a prokaryotic context, we decided to look for prokaryotic protein modules that could be direct activators of chloroplast gene expression. Firstly, we selected the *E.coli* Cyclic AMP Receptor Protein (CRP), which has been shown to activate *lac* gene expression in *E.coli* through a direct interaction with RNA polymerase [20]. CRP has also previously been used for genome interrogation experiments in *E.coli* [21]. For the present study we opted for the use of the C-terminal part of CRP consisting of amino acids 134-190 (designated CRPD2) [21], which lacks the cAMP binding domain and which is a more potent transcriptional activator than the full length CRP protein [21]. As a second option we selected the *Aliivibrio fischeri* protein LuxR, which is a regulator of *lux* gene promoters [22]. 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* [23, 24]. Importantly, LuxRΔN was also shown to possess transcriptional transactivation activity in *E. coli* [25]. Based on the published characteristics of CRPD2 and LuxRΔN we thus hypothesized that both could be suitable modulators of PEP activity in chloroplasts without the requirement of other regulatory proteins.

As the concept of genome interrogation relies on generating relatively large changes in gene expression patterns we did not consider the use of 3Fs, as the 155 kbp chloroplast genome on average contains only one 3F binding site. Instead, we opted to make use of 2F domains, which have 6 bp DNA recognition sites that occur approximately 75 times in a typical chloroplast genome. Provided that the affinity of 2F domains for DNA is still high enough to allow for a preferential presence at these target sites, the activity of CRPD2 or LuxRΔN could lead to differential gene expression at many chloroplast genomic loci. In support of this idea, we have previously found that expression of different nuclear targeted 2F-ATFs can lead to transcriptional changes in *Arabidopsis* [9, 10]. More recently, we have found that salinity tolerance can be induced by a 2F-VP16 fusion (van Tol *et al.*, manuscript in preparation). For chloroplast genome interrogation experiments, we thus decided to randomly select eight different 2Fs for ZF-ATF construction. These 2Fs were denoted as 2F1 through 8 (Table 2).

To avoid any inhibitory effects on the activity of the effector domains when translationally fused to 2F domains, we selected a flexible linker peptide optimized for LuxR activity [25] 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 [25], and provides a distance of 150 Å between the 2Fs and the effector domains, which was determined to be optimal for LuxRΔN activity [25]. In order achieve the translocation of ZF-ATFs into the stroma of chloroplasts, we chose to use the N-terminal chloroplast targeting peptide (CTP) of the FedA protein of Arabidopsis [26, 27]. This CTP has been shown to mediate the translocation of heterologous proteins into chloroplasts [28, 29]. We simultaneously chose to make use of the promoter of the *FEDA* gene to drive ZF-ATF expression.

An overview of the expression cassettes that were designed based on the considerations described is presented in Figure 1, and an overview of the amino acid sequences of the translational fusions encoded by the effector constructs is provided in Figure 2. CTP-mediated chloroplast translocation of ZF-ATFs was verified by confocal microscopy on the mesophyll tissue of Col-0 plants and second generation transformant (T2) plants harboring the construct pCTP-2F1-GFP. As expected, GFP signal could be detected in the chloroplasts of pCTP-2F1-GFP plants (Fig 3).

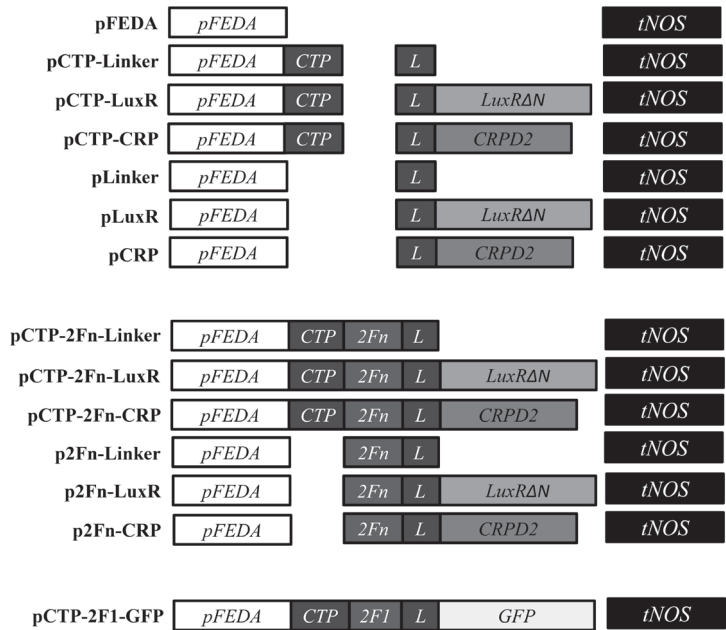


Fig. 1 Overview of expression cassettes that were generated for chloroplast genome interrogation experiments. Expression of all fusion constructs is under control of the promoter sequence of *AtFEDA* (*pFEDA*) and the *NOS* terminator sequence (*tNOS*). The names of the constructs are indicated on the left side of the panels. **A)** Control constructs encoding fusions of effectors (*LuxRΔN* or *CRPD2*) without DNA binding domains to an N-terminal linker (*L*) and with or without N-terminal chloroplast transit peptide (*CTP*). **B)** ZF-ATF encoding constructs consisting of fusions of eight different 2Fs (2Fn; n=1-8) to either *LuxRΔN* or *CRPD2* through the linker sequence, either with or without *CTP*. **C)** Construct encoding a fusion of 2F1 and GFP through the linker sequence with an N-terminal *CTP*.

Characterization of primary transformants to establish the experimental setup

To investigate whether chloroplast genome interrogation can lead to phenotypic differences and altered photosynthetic performance, Arabidopsis Col-0 plants were transformed with T-DNA constructs encoding CTP-LuxRΔN and CTP-CRPD2 fusions without (Fig. 1A) or with 2Fs as DNA binding domains (Fig. 1B). A few dozen primary transformants were readily obtained for all constructs at approximately equal transformation efficiencies, indicating that none of the constructs had effects that are detrimental to embryonic development. We observed some variation in seedling phenotype, but this was randomly distributed and is thus likely to be due to kanamycin selection. When further cultivated on soil the primary transformants did not gain other conspicuous phenotypes that could be attributed to the expression constructs, nor to the presence of a CTP or of 2F domains. Based on these observations in the primary transformant stage we concluded that the expression of chloroplast targeted ZF-ATFs did not have any marked negative effects on chloroplast function.



Fig. 2 Overview of the amino acid (aa) 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 Figure 1A. The linker, LuxRΔN and CRPD2 peptides are presented in green, blue and red, respectively. The insertion sites of the 2Fs are labelled ‘2F insert’.

To avoid the interference of phenotypic aberrations due to kanamycin selection with further quantitative experiments, we decided to assess the T2 progeny of the primary transformants without applying any antibiotic selection. To obtain a first indication of whether any of the expression cassettes had effects on chloroplast performance, we chose to harvest the seeds of 12 randomly chosen primary transformants representing a particular expression construct and to analyze a large number of seedlings consisting of an equal mixture of these 12 T2 lines. These mixtures were designated ‘T2-pools’. As there are no practical options available to assess chloroplast genome wide transcription patterns we opted to use growth and photosynthesis as indicators of ZF-ATF induced changes in chloroplast performance.

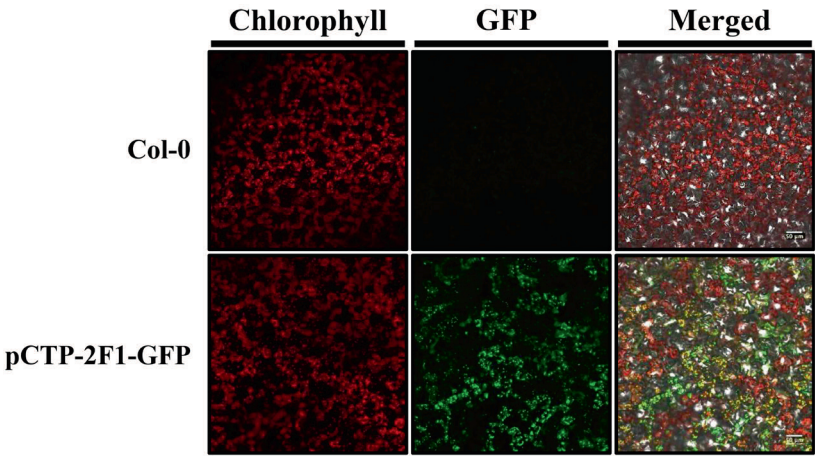


Fig. 3 Confocal microscopy images of abaxial mesophyll tissue of the sixth leaf of Col-0 and pCTP-2F1-GFP (T2) plants at 25 dpv (20x magnification). Merged images are an overlay between 633 nm, GFP and bright field images.

Phenotypic and photosynthetic properties of mixed T2 populations

When sown on soil wild type Col-0 seedlings and T2-pool seedlings harboring the pCTP-Linker construct had germinated homogenously at 4 days post germination (dpg) and did not display conspicuous phenotypic variation (Fig. S1A). However, T2-pool seedlings harboring the constructs pCTP-CRP and pCTP-LuxR exhibited noticeable variation in size (Fig. S1A), suggesting that chloroplast targeting of CRPD2 and LuxRΔN without DNA binding domains can already influence seedling development. Among the seedlings harboring the constructs pCTP-2Fn-CRP and pCTP-2Fn-LuxR (Fig. S1B and Fig. S1C, respectively) the variation in growth and pigmentation was visibly more substantial than the variation induced by pCTP-CRP and pCTP-LuxR alone, indicating that ZF-ATFs can trigger phenotypic variation in Arabidopsis seedlings when targeted to chloroplasts.

To investigate whether LuxRΔN and CRPD2 can have an effect on photosynthesis, we quantified the operating light use efficiency of PSII (ϕ PSII) of populations of T2-pool seedlings by CF imaging. We used ϕ PSII as a measure for photosynthesis because it can be quantified in a high throughput manner for large numbers of seedlings, and because there is a strong correlation between ϕ PSII, linear electron transport rate and the rate of CO₂ fixation in plants [30]. As expected, seedlings harboring the construct pFEDA, which lacks an open reading frame, did not display changes in ϕ PSII compared to Col-0 (Fig. 4). However, T2-pool seedlings harboring the constructs pCTP-LuxR and pCTP-CRP did display significant increases in ϕ PSII compared to Col-0 (Fig. 4). These ϕ PSII increases were not found for seedlings harboring the constructs pLuxR and pCRP (Fig. 4), indicating that LuxRΔN and CRPD2 can only trigger ϕ PSII increases when translocated to chloroplasts. For unknown reasons significant ϕ PSII increases were also noted for seedlings harboring the constructs pCTP-Linker and pLinker, suggesting that the linker domain has an effect ϕ PSII regardless of the presence of a CTP.

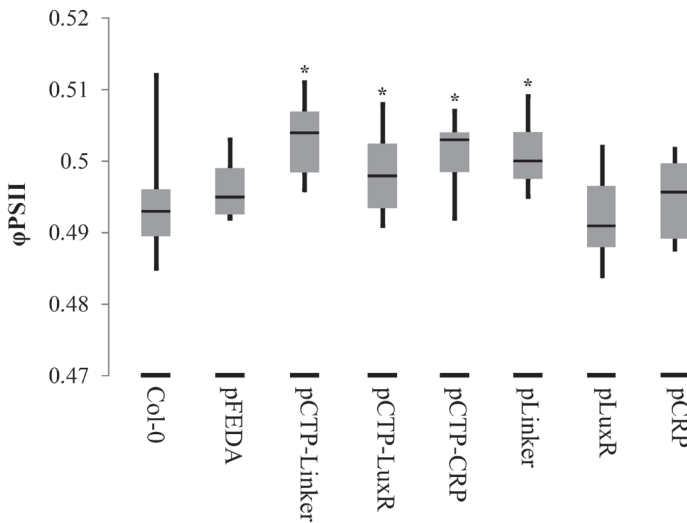


Fig. 4 Quantification of the operating light use efficiency of Photosystem II (ϕ PSII) at 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$ of actinic light of Col-0 seedlings and seedlings harboring control constructs without putative chloroplast genome interrogation activity (7 dp). The presented boxplots were generated from quadrant data obtained from approximately 250 seedlings ($n=9$). Asterisks (*) indicate significant differences with Col-0 ($p < 0.05$) determined by one-way ANOVA analysis.

To investigate whether ZF-ATF activity in chloroplasts can have an effect on photosynthesis, we quantified ϕ PSII of T2-pool seedlings harboring the two series of ZF-ATF encoding constructs pCTP-2Fn-LuxR and pCTP-2Fn-CRP, respectively. There were significant increases in the ϕ PSII of seedlings harboring the constructs pCTP-2F3-LuxR, pCTP-2F4-

CRP and pCTP-2F7-CRP compared to the respective empty vector control constructs lacking 2Fs as DNA binding domains (Fig. 5A and 5B). There was also a noticeable upward shift in the distribution of the ϕ PSII data of T2-pool seedlings harboring the construct pCTP-2F1-LuxR which was not statistically significant, but which we considered to be of interest regardless. Altogether these data suggested that the activity of chloroplast targeted 2F1-LuxR Δ N, 2F3-LuxR Δ N, 2F4-CRPD2 and 2F7-CRPD2 fusions can lead to enhancement of ϕ PSII.

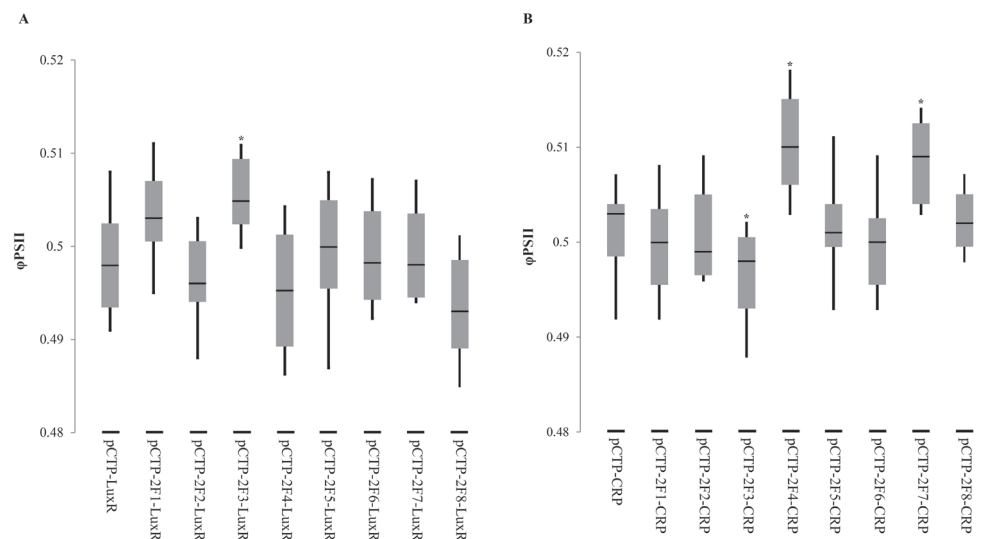


Fig. 5 Quantification of the operating light use efficiency of Photosystem II (ϕ PSII) at 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$ of actinic light of seedlings harboring T-DNA constructs encoding chloroplast targeted 2F-LuxR Δ N (**A**) or 2F-CRPD2 (**B**). Seedlings harboring constructs without 2Fs serve as negative controls. The presented boxplots were generated from quadrant data obtained from approximately 250 seedlings ($n=9$). Asterisks (*) indicate significant differences with the negative controls pLuxR and pCRP ($p < 0.05$) determined by one-way ANOVA analysis.

Identification of individual 2F-LuxR lines with CTP-dependent ϕ PSII increases

To corroborate the indications that ZF-ATFs can modulate ϕ PSII and do so in a 2F dependent manner, we investigated the contribution of the individual lines that were used to generate T2-pools to the ϕ PSII increases. In this way it could be confirmed that five out of eight independent pCTP-2F1-LuxR lines and three out of 12 independent pCTP-2F3-LuxR lines displayed significantly higher ϕ PSII than the empty vector control pCTP-LuxR (Fig. 6). To examine whether the induction of these ϕ PSII increases is dependent on chloroplast translocation of the ZF-ATFs, Col-0 plants were transformed with the constructs p2F1-LuxR and p2F3-LuxR, which encode the same ZF-ATFs but lack the CTP. Subsequently, the ϕ PSII of populations of the T2 progeny of five randomly chosen primary transformants with these constructs was quantified. Seedlings harboring the constructs p2F1-LuxR and p2F3-LuxR displayed ϕ PSII

values which were either similar to or significantly lower than those of seedlings harboring the empty vector control construct pLuxR (Fig. 7A and B), indicating that 2F1-LuxRΔN and 2F3-LuxRΔN fusions can only induce increases in ϕ PSII when translocated to chloroplasts.

We attempted to perform a similar analysis as described above for lines harboring the constructs pCTP-2F4-CRP and pCTP-2F7-CRP, but for unknown reasons the seed germination percentage of these lines declined steeply within a few months, making unbiased analysis of the seedling impossible.

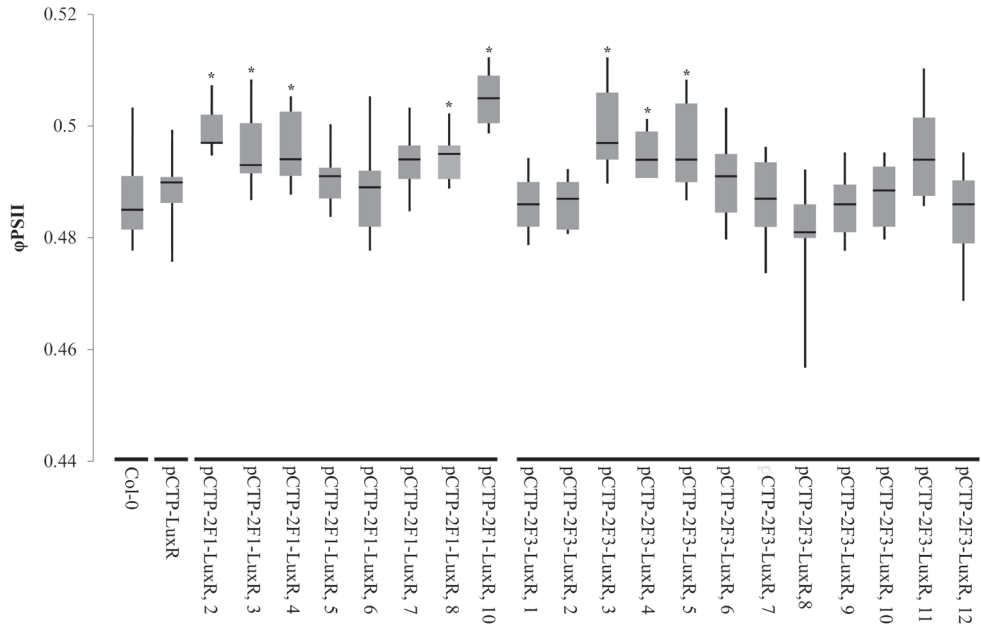


Fig. 6 Determination of the contribution of individual T2 pCTP-2F-LuxR and pCTP-2F3-LuxR lines to the overall increase in ϕ PSII of the complex mixture of these lines presented in Figure 5. The operating light use efficiency of Photosystem II (ϕ PSII) was quantified at $200 \mu\text{mol m}^{-2} \text{s}^{-1}$ of actinic light. The presented boxplots were generated from quadrant data obtained from approximately 250 seedlings ($n=9$). Asterisks (*) indicate significant differences with pCTP-LuxR ($p < 0.05$) determined with a heteroscedastic T-test assuming unequal variance to account for transgene segregation.

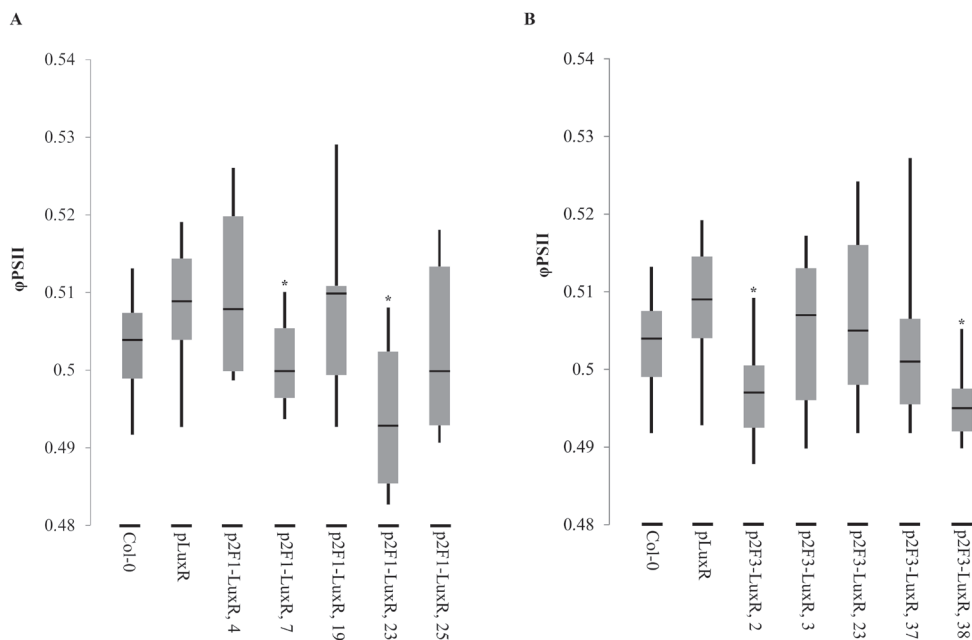


Fig. 7 Quantification of the operating light use efficiency of Photosystem II (ϕ_{PSII}) at 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$ of actinic light of seedlings (11 dpv) harboring constructs encoding 2F1-LuxR (A) and 2F3-LuxR (B) fusions lacking a CTP (five independent lines each). Seedlings harboring the construct pLuxR serve as a negative control. The presented boxplots were generated from quadrant data obtained from approximately 250 seedlings ($n=9$). Asterisks (*) indicate significant differences with pLuxR ($p < 0.05$) determined with a heteroscedastic T-test assuming unequal variance to account for transgene segregation.

Plants harboring the constructs pCTP-2F1-LuxR and pCTP-2F3-LuxR display changes in growth and experience leaf damage at the later stages of development

All of the ϕ_{PSII} measurements described above were performed on young Arabidopsis seedlings. To examine the effect of the expression of chloroplast targeted 2F1-LuxR Δ N and 2F3-LuxR Δ N fusions on the growth of Arabidopsis plants in later stages of development, we quantified the rosette surface area (RSA) of segregation populations of T2 plants harboring the constructs pCTP-2F1-LuxR (lines 2, 8 and 10) and pCTP-2F3-LuxR (lines 1, 3 and 11). RSA was used as a non-destructive proxy for growth, because there is a strong correlation between RSA and biomass in Arabidopsis [31]. For both constructs, two out of three lines displayed a reduction in RSA throughout development compared to Col-0 (Fig. S2A), and one out of three displayed wild type growth, respectively (Fig. S2A). At 31 dpv, a substantial fraction of the T2 plants of five out the six lines displayed mild to severe leaf necrosis, ranging from small spots to partial or even complete death of mature leaves (Fig. S2B). Regardless, these observations did not indicate that 2F1-LuxR and 2F-LuxR have marked effects in the later stages of development.

Discussion

In this study, we have described the design of a novel system to perform chloroplast genome interrogation in *Arabidopsis* seedlings. This system was tested using two types of chloroplast targeted ZF-ATFs consisting of fusions of the bacterial transcriptional activators CRPD2 and LuxRΔN to arrays of 2Fs as DNA binding domains. Using a relatively small number of 2Fs we have found evidence that both types of chloroplast targeted ZF-ATFs can induce phenotypic variation of *Arabidopsis* seedlings and can modulate their ϕ PSII.

Although chloroplast genome engineering is considered an option for manipulating plant photosynthesis [5], the transformation of chloroplasts and subsequent selection of homoplasmic plant lines is an inherently tedious procedure and has not yet been reported for any commercially important plant species. However, the chloroplast genome interrogation system that we have investigated in this study should allow for the *in trans* manipulation of gene expression patterns in all chloroplasts and all copies of the chloroplast genome simultaneously through the integration of a single artificial gene in the nuclear genome. Since nuclear transformation protocols have become available for numerous plant species, the chloroplast genome interrogation system could in principle readily be applied to commercially interesting plant species without the requirement of detailed *a priori* knowledge regarding their plastid biology.

Using ϕ PSII as read-out we have gathered evidence that several 2F and effector domain combinations from a rather limited pool can already have effects on chloroplast performance. Even though 2Fs are expected to bind to DNA with much lower specificity and lower affinity than 3Fs [8] and recognize just 6 or 7 bp of DNA (the latter due to target site overlap [32]), our data indicate that they still have sufficiently distinct binding specificity and affinity for DNA to interact with the chloroplast genome *in planta*. The use of the promoter of the *FedA* gene combined with the sequence encoding its CTP is therefore likely to have led to a ZF-ATF protein concentration in the chloroplast stroma that is sufficiently high to allow for 2F-specific DNA interactions. In the present study we did not yet attempt to directly investigate ZF-ATF induced transcriptional changes in chloroplasts because this has a number of practical limitations. Our data therefore do not yet allow for any conclusive statements regarding the activities of the LuxRΔN or the CRPD2 domains. Our evidence in terms of ϕ PSII indicates that they do modulate chloroplast transcription. As mentioned above, there was a marked decrease in the viability of seeds harboring pCTP-CRP constructs, forcing us to refrain from further analysis of these seedlings. Even though CRPD2 is seemingly detrimental to long term seed survival, two out of the eight T2-pools harboring pCTP-2Fn-CRP constructs still exhibited an increase in ϕ PSII, which was comparable to the two out of eight T2-pools harboring pCTP-2Fn-LuxR constructs that were further analyzed. It also has to be taken into account that LuxRΔN still possesses some putative DNA binding activity [25], meaning

that it could potentially bind to DNA regardless of the presence of a 2F domain. This might also explain why the seedlings harboring the construct pCTP-LuxR have significantly higher ϕ PSII values than Col-0 seedlings.

The initial ϕ PSII quantifications were performed on complex populations consisting of T2 seedlings from multiple lines (T2-pools) that were segregating for the transgenes. As the observed increases in ϕ PSII were relatively small, this could have been attributable to a single or small number of lines with very pronounced increases in ϕ PSII. We also noted that the ZF-ATF encoding constructs on average had a substantial effect on seedling growth and pigment composition (Fig. S1B and C), which could to some extent have masked otherwise higher average ϕ PSII values. To be able to attribute the ϕ PSII increases specifically to ZF-ATFs, further verification of the phenotype in individual T2 seedling populations showed to be essential. As described, the majority of 5 out of 8 independent pCTP-2F1-LuxR T2 lines exhibited a significantly higher ϕ PSII levels (Fig. 5A), and 3 out of 12 independent pCTP-2F3-LuxR lines did so as well (Fig. 5B). With these rather large fractions of lines exhibiting the phenotype, we consider the conclusion that both pCTP-2F1-LuxR and pCTP-2F3-LuxR can enhance ϕ PSII justified. Not all transgenic lines did necessarily display the phenotype, but this might very well be dependent on individual differences in transgene expression levels due to differences in copy number and/or insertion site(s).

Even though our data suggest that ZF-ATFs can induce increases in the ϕ PSII of Arabidopsis seedlings, we have not observed that this is translated into more growth (Fig. S2), possibly suggesting that the ϕ PSII increases do not lead to more efficient overall photosynthesis. We think that this is most likely explained by the fact that ϕ PSII increases induced at the level of chloroplast gene expression also require differential retrograde signalling to the nucleus [4, 18], as both growth and photosynthesis are performed and regulated at the cellular and tissue levels. We therefore consider it likely that the translation of chloroplast performance to more growth and more efficient overall photosynthesis requires additional differential expression of nuclear genes. There is also debate about whether or not photosynthetic efficiency is at all positively correlated with biomass accumulation [33]. Quite recently, a transplastomic tobacco mutant expressing a hybrid RuBisCo with a spectacularly high rate of CO₂ fixation was generated [34], but these plants are rather small and albino, further suggesting that is difficult to synchronize chloroplast photosynthetic performance with overall plant performance.

In conclusion, based on basic knowledge of chloroplast biology and without requiring further *a priori* knowledge of the chloroplast genome we have successfully designed a novel chloroplast genome interrogation system. Using a relatively small setup we have already found evidence that ZF-ATF mediated chloroplast genome interrogation can induce small but significant changes in the photosynthetic performance of chloroplasts. Altogether our work suggests that it would be worthwhile to further investigate chloroplast genome interrogation as a novel tool to enhance the photosynthetic performance of plants.

Materials and methods

Growth conditions and plant material

The *Arabidopsis* accession Columbia-0 (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 days at 4 °C prior to 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 $\mu\text{mol m}^{-2} \text{s}^{-1}$ of photosynthetically active radiation (PAR), and at a 12 h photoperiod (referred to as ‘standard growth conditions’). Primary floral dip transformants were first grown on selection medium in a climate controlled tissue culture chamber at a constant temperature of 20 °C, 50% relative humidity, a PAR light intensity of approximately 50 $\mu\text{mol m}^{-2} \text{s}^{-1}$, and at a 16 h photoperiod, subsequently transferred to soil after approximately 3 weeks and were further cultivated at standard growth conditions.

Construction of *Arabidopsis* plant lines expressing ZF-ATFs

A library of plasmids containing DNA fragments encoding all 256 different 2Fs was previously constructed [35]. Eight different 2Fs consisting of two different ZFs were randomly selected from this library (Table 1). The DNA sequence of CRP was derived from NCBI (REFSEQ accession NC_000913.2). The amino acid (aa) sequence of CRPD2 was derived from *Lee et al.*, 2008 [21]. The aa residues 134-136 (NLA) were also included, as these were reported to constitute a flexible hinge [36]. The DNA sequence of LuxR was derived from NCBI (accession M25752, version 1). The aa sequences of LuxR Δ N and the flexible linker were derived from Volzing *et al.*, 2011 [25]. The promoter of *AtFEDA* (At1g60950) including the 5'-UTR sequence and the sequences encoding the CTP and the first 8 amino acids of FedA was amplified by PCR from the genomic DNA of Col-0 using the forward primer pFEDA FW (Table 2) and the reverse primers pFEDA REV1 and pFEDA REV2 (Table 2), yielding a 2029 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, Amsterdam). The binary vector plasmid pRF [9] was used as the backbone for all cloning steps.

The *RPS5A* promoter sequence and the 3F-VP16 ORF [9] were removed, and the *pFEDA* fragment was subsequently ligated in, yielding the plasmid pFEDA. A 700 bp oligo DNA fragment (Fig. S3) encoding the flexible linker, LuxR Δ N and CRPD2 (codon optimized for *Arabidopsis*) was synthesized by the company ShineGene (Shanghai, China), 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, pCTP-LuxR and pCTP-Linker as *Sfi*I fragments, yielding the plasmids which were designated pCTP-2Fn-

CRP, pCTP-2Fn-LuxR and pCTP-2Fn-Linker, respectively. Using pFEDA as a template, a PCR product lacking the CTP was generated using the primer combination pFEDA FW and MASTAL REV (Table 2), and was ligated into pCTP-CRP, pCTP-LuxR and pCTP-Linker, yielding the plasmids which were named pCRP, pLuxR and pLinker, respectively. The DNA sequence encoding eGFP was amplified by PCR using the forward primer GFP FW and reverse primer GFP RV (Table 2), and ligated into pCTP-2F1-Linker, yielding the plasmid designated pCTP-2F1-GFP. Plasmid sequences are available upon request. Col-0 plants were transformed with each of the generated constructs separately using the floral dip method [37] as described previously [9].

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

Table 2. 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	<u>GTTCGACT</u> GCCTTTTACGGAAAGATTTCGATTG (Sall)
pFEDA REV1	<u>CCTCTCGAGG</u> ATGAACCTTGACCTTGATGTAGC (XhoI)
pFEDA REV2	<u>GGAGCTCAGG</u> CCTCTCGAGGATGAACCTTGACCTTG (SacI)
MASTAL REV	<u>GCTCGAGAG</u> CAGTGGAAGCCATTTTTTTTG (XhoI)
GFP FW	<u>GACTAGTGT</u> GAGCAAGGGCGAGGAGCTGTTCACCG (SpeI)
GFP RV	<u>GGAGCTCT</u> TACTTGTACAGCTCGTCCATGCCG (SacI)
FEDA FW	CACGCCATTTCACAAGC

Confocal microscopy

Confocal microscopy was performed on the abaxial tissue of the sixth leaf of Col-0 plants (25 dpv) and T2 plants harboring the construct pCTP-2F1-GFP using a Zeiss L5M5 Exciter (Zeiss, Jena, Germany) at 20x magnification. Excitation of the tissue was performed with a 488 nm laser. All leaves were imaged at the same laser power. The emission of chlorophyll fluorescence was collected with a 560 nm long pass filter and GFP fluorescence was collected with a 505 to 530 nm band pass filter. Merges between chlorophyll fluorescence, GFP fluorescence and bright field images were generated using ImageJ.

ϕ PSII quantification of Arabidopsis seedlings

Approximately 250 T2 seeds from either T2-pools (as explained in the results section) or from individual T1 parents were sown on soil in pots with a diameter of 15.7 cm and height of 65 mm (Soparco, Condé-sur-Huisne, France). At one time point in development (varying between experiments depending on seedling growth rate), the operating light use efficiency of Photosystem II (F_q'/F_m' [30]; referred to as ϕ PSII) was quantified using a CF Imager (Technologica, Essex, United Kingdom). The seedlings were exposed to 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$ actinic light in the CF imaging chamber (the same light intensity as in the growth chamber) for 1 min, after which F' and F_m' images were generated by exposure to a 6226 $\mu\text{mol m}^{-2} \text{s}^{-1}$ saturating actinic light pulse. F_q'/F_m' images were constructed from the F' and F_m' images. Local differences in seedling density along the surface of individual pots led to variation in F_q'/F_m' images. To account for this variation, images containing data of 250 seedlings were subdivided into 9 equal and non-overlapping quadrants, each representing the average ϕ PSII of approximately 28 seedlings. As the quadrants did not overlap and the independent seedlings grew in different parts of large pots, the quadrants were considered biological replicates ($n=9$). For the first comparative measurements of T2-pools (Fig. 4 and 5) the quadrant data were statistically analyzed by one-way ANOVA. For comparisons of individual T2 lines (Fig. 6 and 7) the heteroscedastic T-Test function of Microsoft Excel was used, assuming unequal variance between samples due to possible differences in transgene segregation. In all cases a p -value of 0.05 was used as a threshold for significance.

Growth analysis of pCTP-2F1-LuxR and pCTP-2F3-LuxR lines

For rosette surface area quantification, approximately 50 seeds of Col-0 and of independent pCTP-2F1-LuxR and pCTP-2F3-LuxR lines were sown in pots with a diameter of 15.7 cm and height of 65 mm (Soparco, Condé-sur-Huisne, France), and stratified at 4 °C for 3 days. Seedlings were transferred to 67 x 67 x 65 mm pots (Pöppelmann, Lohne, Germany) at 7dpg. Photos were taken of all trays from the top with a fixed digital camera (Canon EOS 1100D) from 10 dpg onwards and every two to five days. The surface area of each rosette was subsequently calculated in pixel² using the 'Analyze Particles' function of ImageJ, and then converted to mm² by multiplying this value by the mm²/pixel² ratio of each RGB image.

Acknowledgements

We would like to thank Paul de Mooij for his help in the initial stages of the experimental design of this study. We would like to thank Prof. Dr. H.J.M. de Groot and Dr. A. Alia Matysik (Leiden Institute of Chemistry, Leiden University, the Netherlands) for stimulating discussions. This work was carried out within the research programme of BioSolar Cells, co-financed by the Dutch Ministry of Economic Affairs.

References

1. 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(19):12246-51.
2. Bruce BD. (2000) Chloroplast transit peptides: structure, function and evolution. *Trends Cell Biol* 10(10):440-7.
3. Flores-Perez U, Jarvis P. (2013) Molecular chaperone involvement in chloroplast protein import. *BBA Mol Cell Res* 1833(2):332-40.
4. Jarvis P, Lopez-Juez E. (2013) Biogenesis and homeostasis of chloroplasts and other plastids. *Nat Rev Mol Cell Biol* 14(12):787-802.
5. Ort DR, Merchant SS, Alric J, Barkan A, Blankenship RE, Bock R, *et al.* (2015) Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proc Natl Acad Sci USA* 112(28):8529-36.
6. Bock R. (2015) Engineering plastid genomes: methods, tools, and applications in basic research and biotechnology. *Annu Rev Plant Biol* 66:211-41.
7. Beltran A, Liu YZ, Parikh S, Temple B, Blancafort P. (2006) Interrogating genomes with combinatorial artificial transcription factor libraries: Asking zinc finger questions. *Assay Drug Dev Techn* 4(3):317-31.
8. van Tol N, van der Zaal BJ. (2014) Artificial transcription factor-mediated regulation of gene expression. *Plant Sci* 225:58-67.
9. Lindhout BI, Pinas JE, Hooykaas PJ, van der Zaal BJ. (2006) Employing libraries of zinc finger artificial transcription factors to screen for homologous recombination mutants in Arabidopsis. *Plant J* 48(3):475-83.
10. Jia Q, van Verk MC, Pinas JE, Lindhout BI, Hooykaas PJ, van der Zaal BJ. (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 Biotech J* 11(9):1069-79.
11. Segal DJ, Dreier B, Beerli RR, Barbas CF, 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(6):2758-63.
12. Sadowski I, Ma J, Triezenberg S, Ptashne M. (1988) GAL4-VP16 is an unusually potent transcriptional activator. *Nature* 335(6190):563-4.
13. Lindhout I, Markus M, Hoogendijk T, Borst S, Maingay R, Spinhoven P, *et al.* (2006) Childrearing style of anxiety-disordered parents. *Child Psychiatry Hum Dev* 37(1):89-102.
14. Park KS, Seol W, Yang HY, Lee SI, Kim SK, Kwon RJ, *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(3):664-70.
15. Hiratsu K, Matsui K, Koyama T, 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(5):733-9.
16. Mito T, Seki M, Shinozaki K, Ohme-Takagi M, Matsui K. (2011) Generation of chimeric repressors that confer salt tolerance in Arabidopsis and rice. *Plant Biotech J* 9(7):736-46.
17. Borner T, Aleynikova AY, Zubo YO, Kusnetsov VV. (2015) Chloroplast RNA polymerases: Role in chloroplast biogenesis. *Biochim Biophys Acta* 1847(9):761-9.
18. Woodson JD, Chory J. (2008) Coordination of gene expression between organellar and nuclear genomes. *Nat Rev Genet* 9(5):383-95.
19. Leister D. (2005) Genomics-based dissection of the cross-talk of chloroplasts with the nucleus and mitochondria in Arabidopsis. *Gene* 354:110-6.

20. Borukhov S, Lee J. (2005) RNA polymerase structure and function at lac operon. *C R Biol* 328(6):576-87.
21. Lee JY, Sung BH, Yu BJ, Lee JH, Lee SH, Kim MS, *et al.* (2008) Phenotypic engineering by reprogramming gene transcription using novel artificial transcription factors in *Escherichia coli*. *Nucleic Acids Res* 36(16):e102.
22. Dunlap P. (2014) Biochemistry and genetics of bacterial bioluminescence. *Adv Biochem Eng Biotechnol* 144:37-64.
23. Schu DJ, Carlier AL, Jamison KP, von Bodman S, Stevens AM. (2009) Structure/function analysis of the *Pantoea stewartii* quorum-sensing regulator EsaR as an activator of transcription. *J Bacteriol* 191(24):7402-9.
24. Choi SH, Greenberg EP. (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(24):11115-9.
25. Volzing K, Biliouris K, Kaznessis YN. (2011) proTeOn and proTeOff, new protein devices that inducibly activate bacterial gene expression. *ACS Chem Biol* 6(10):1107-16.
26. Somers DE, Caspar T, Quail PH. (1990) Isolation and Characterization of a Ferredoxin Gene from *Arabidopsis thaliana*. *Plant Physiol* 93(2):572-7.
27. Smeekens S, Geerts D, Bauerle C, Weisbeek P. (1989) Essential Function in Chloroplast Recognition of the Ferredoxin Transit Peptide Processing Region. *Mol Gen Genet* 216(1):178-82.
28. Smeekens S, van Steeg H, Bauerle C, Bettenbroek H, Keegstra K, Weisbeek P. (1987) Import into chloroplasts of a yeast mitochondrial protein directed by ferredoxin and plastocyanin transit peptides. *Plant Mol Biol* 9(4):377-88.
29. Jin R, Richter S, Zhong R, Lamppa GK. (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(4):493-507.
30. Baker NR. (2008) Chlorophyll fluorescence: a probe of photosynthesis in vivo. *Annu Rev Plant Biol* 59:89-113.
31. Leister D, Varotto C, Pesaresi P, Niwergall A, Salamini F. (1999) Large-scale evaluation of plant growth in *Arabidopsis thaliana* by non-invasive image analysis. *Plant Physiol Bioch* 37(9):671-8.
32. Wolfe SA, Neklodova L, Pabo CO. (2000) DNA recognition by Cys2His2 zinc finger proteins. *Annu Rev Biophys Biomol Struct* 29:183-212.
33. Kirschbaum MU. (2011) Does enhanced photosynthesis enhance growth? Lessons learned from CO₂ enrichment studies. *Plant Physiol* 155(1):117-24.
34. Lin MT, Occhialini A, Andralojc PJ, Parry MA, Hanson MR. (2014) A faster Rubisco with potential to increase photosynthesis in crops. *Nature* 513(7519):547-50.
35. Neuteboom LW, Lindhout BI, Saman IL, Hooykaas PJ, van der Zaal BJ. (2006) Effects of different zinc finger transcription factors on genomic targets. *Biochem Biophys Res Commun* 339(1):263-70.
36. Baker CH, Tomlinson SR, Garcia AE, Harman JG. (2001) Amino acid substitution at position 99 affects the rate of CRP subunit exchange. *Biochemistry* 40(41):12329-38.
37. Clough SJ, Bent AF. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16(6):735-43.

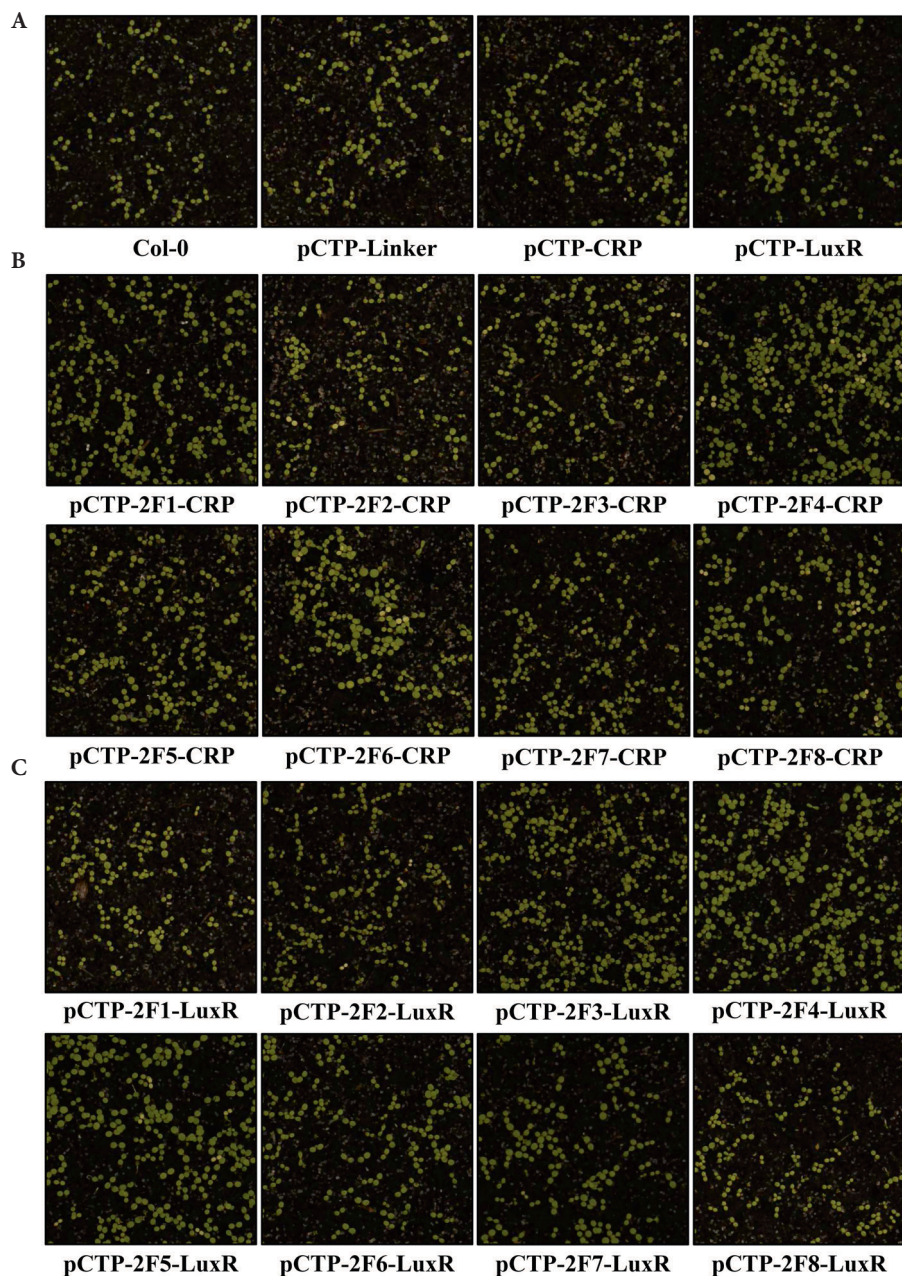


Fig. S1 An overview of the phenotypes of Arabidopsis seedlings (4 dpg) harboring empty vector control constructs (A), 2F-CRPD2 encoding constructs (B) and 2F-LuxR encoding constructs (C). An overview of the composition of the constructs is provided in Figure 1. Presented are populations of T2 seedlings that are segregating for the constructs, and have germinated from a complex mixture of seeds originating from 10-12 independent primary transformants (T1).

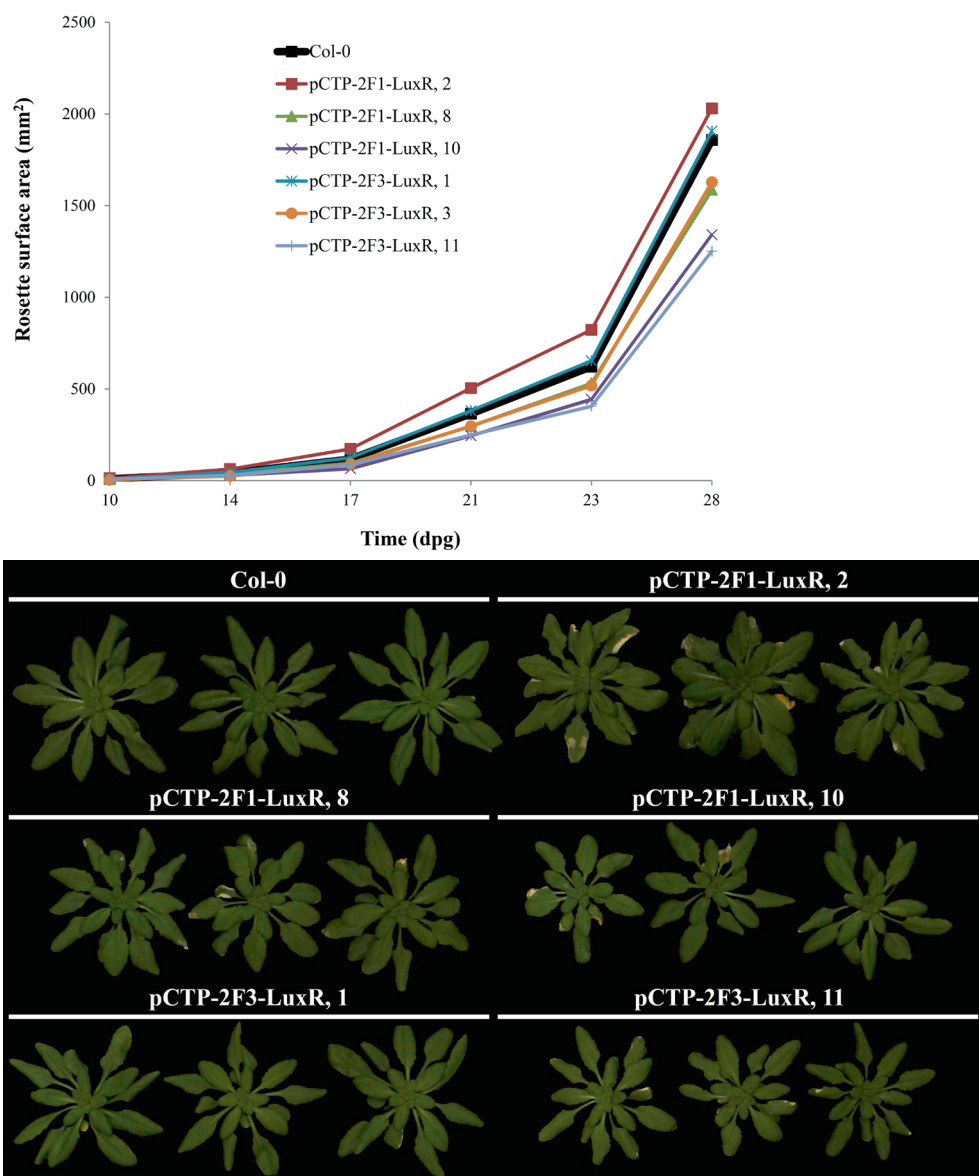


Fig. S2 Overview of growth of Col-0 plants, pCTP-2F1-LuxR plants (lines 2, 8 and 10; T2 generation; segregating for the expression construct) and pCTP-2F3-LuxR plants (lines 1,3 and 11; T2 generation; segregating for the expression construct). **A)** Quantification of rosette surface area throughout development (n=33 for Col-0, n=14-18 for the transgenic lines). Error bars were not included for the sake of clarity of the figure. **B)** The rosette phenotypes of three representative individuals for each genotype at 31 dpg. The transgenic plants all display some degree of leaf damage or death, except for pCTP-2F3-LuxR,3 which was therefore not included.

Spacer

CTCGAGCCCAGGCGGCCTTAATAATCGTGCACTCGGCCAGGCCGGCCAATTAGTGGTGGTGGTGC
XhoI *SfiI* *SfiI*

TAGGACACAATATTCAGAAAGTATGGGTGCTAGAACCCAATACTCGGAAAGCATGGGAGCTAGAACCC
 AATACTCTGAAAGCATGGGAGCTAGAACTCAATACTCCGAGTCTATGGGAGCCAGAACACAATATAGT
 GAGTCAATGGGAGGGGGAACTAGTAGTACTCCGTCCCTCGTTGATAACTATAGGAAAGATTAATATCGCA
SpeI *ScaI*

AATAATAAGTCCAATAACGATCTGACGAAGAGGGAGAAAGAATGCTTGGCTTGGGCTTGTGAGGGAAA
 AAGCTCATGGGACATCTCTAAAATTTGGGCTGTTCAAAGCGTACAGTTACATTTCATCTGACCAACGC
 ACAGATGAAATTGAATACGACGAACAGATGCCAGAGCATTTCTAAGGCAATTTGACAGGAGCTATCG
 ACTGCCCATACTTCAAGTCATAATCTAGAAACCGGTTTAGCATTCTTGATGTGACCGGCAGAATTGCAC
XbaI *AgeI*

AAACACTCCTCAACCTTGCCAAGCAGCCAGATGCAATGACTCACCCAGACGGTATGCAGATTAAGATC
 ACACGTCAGGAGATTGGCCAGATTGTAGGGTGTTGCGGTGAAACAGTAGGGCGGATACTCAAGATGTT
 STOP STOP
 GTAACCCGGGTGAGCTC
SmaI *SacI*

Fig. S3 Overview of the 700 bp oligo DNA sequence that was synthesized for the construction of chloroplast genome interrogation expression cassettes. Sequences encoding the linker, LuxRΔN and CRP are presented in green, blue and red font, respectively. Restriction sites are underlined. Stop codons are labelled 'STOP'.

Chapter 5

An Arabidopsis mutant with high operating efficiency of Photosystem II and low chlorophyll fluorescence

Niels van Tol, Johan Pinas, Paul J.J. Hooykaas and Bert J. van der Zaal

Abstract

Photosynthesis is the process by which light energy is converted to chemical energy, and is the driving force behind biomass accumulation in plants. However, the overall efficiency of light energy to biomass conversion by photosynthesis is generally regarded as low. The improvement of photosynthetic efficiency has become a primary target for enhancement of crop yield. Here, we describe the isolation and characterization of an *Arabidopsis* mutant with the novel phenotype of a high operating efficiency of Photosystem II (ϕ PSII) and low chlorophyll fluorescence. This mutant was denoted as Low Chlorophyll Fluorescence 1 (LCF1). The high ϕ PSII of LCF1 was due to an increase in photochemical quenching, and was insensitive to stressful light conditions. The increase in ϕ PSII of LCF1 did not result in the accumulation of more biomass. Rather, at standard light conditions LCF1 had a growth defect. LCF1 plants could be rescued from this defect by growing them at low light intensity and short day conditions, or by inducing senescence.

Introduction

Photosynthesis is the process that harvests energy from sunlight and fixes it as chemical energy. The light is absorbed by chlorophyll molecules that are associated with the antenna complexes of Photosystems I and II (PSI and PSII) in the thylakoid membranes of chloroplasts, which results in the photoexcitation and subsequent charge separation of the reaction center chlorophylls P700 and P680, respectively. The water-plastoquinone A oxidoreductase activity of PSII, referred to as photochemistry, initiates the linear flow of photoexcited electrons from PSII to the final electron acceptor NADP, thus generating NADPH, and drives the synthesis of ATP by chemiosmotic coupling. NADPH and ATP can be used in the Calvin-Benson cycle for CO₂ fixation by the enzyme complex RuBisCo, where after the resulting carbohydrate product is partitioned to different parts and processes of the plant. Photosynthesis is therefore considered the driving force behind plant productivity. However, the overall light energy to biomass conversion efficiency of photosynthesis is generally regarded as remarkably low [1]. Photosynthesis has therefore become a primary target for the improvement of crop yield [2], and the genetic modification of photosynthesis has received a considerable amount of attention [3-6].

Apart from photochemistry, photoexcitation events in PSII reaction centers can lead to the emission of chlorophyll fluorescence (CF), and to the dissipation of excess excitation energy through various processes that are together named non-photochemical quenching (NPQ) [7]. Upon illumination of a plant with actinic light, CF can relatively easily be observed and quantified with a fluorescence camera. Photochemistry and NPQ are both quenchers of CF, and changes in the rates of photochemistry and NPQ are therefore proportionally reflected by changes in CF levels. CF imaging with Pulse-Amplitude-Modulation (PAM) equipment therefore allows for the relatively accurate quantification of photochemistry, the rate of linear electron transport and NPQ. Importantly also, CF imaging allows for the quantification of the operating light use efficiency of Photosystem II (ϕ PSII) [8], which is a measure for how efficiently absorbed light is used by PSII reaction centers for the reduction of plastoquinone A. The ϕ PSII of a plant is also considered a measure for the efficiency of overall photosynthesis, because there is a strong correlation between ϕ PSII, linear electron transport rate and the rate of CO₂ fixation in both C3 and C4 plant species [8]. CF imaging thus allows for the relatively high throughput and simple estimation of the overall photosynthetic performance of plants, as opposed to for instance gas exchange measurements. With the emergence of CF imaging, the isolation of photosynthesis mutants from populations of plants has become feasible.

Several published and ongoing studies have used ϕ PSII as a measure for the assessment of variation in photosynthesis among natural *Arabidopsis* accessions. Among the three most widely used accessions, Columbia-0 (Col-0) and Landsberg erecta-0 (Ler-0) have very similar ϕ PSII values, whereas that of Wassilewskija-4 (Ws-4) is lower [9]. Overall, very little

natural variation for ϕ PSII has been found in Arabidopsis, with the most extreme variation being due to impairment rather than gain of PSII function [10]. No natural accessions with substantially higher ϕ PSII than Col-0 have been documented [10]. Surprisingly, no large scale mutant screens have been performed with ϕ PSII as a selection criterion, or with any other CF parameter for that matter. The most extensively studied class of CF mutants are the recessive high CF (*hcf*) mutants. Although interesting from a scientific point of view, these mutants have impaired rather than enhanced PSII function, and often completely lack the activity of one of the major thylakoid membrane protein complexes. The *hcf* mutants display very high levels of CF upon illumination, and can therefore relatively easily be isolated from populations of plants using a fluorescence camera. In this way, a substantial number of *HCF* genes has been documented [11, 12]. To our knowledge, no phenotypic screens for PSII gain-of-function mutants have been documented. Also, no mutants with low rather than high CF have been described. Screening a population of mutants for changes in ϕ PSII might allow for the isolation of such gain of function mutants.

In our lab, we have established zinc finger artificial transcription factor (ZF-ATF) mediated genome interrogation [13] as a tool for generating novel Arabidopsis mutants with enhanced traits of interest [14, 15]. In our setup, ZF-ATFs consist of an array of three zinc fingers (3F) fused to the transcriptional activator protein VP16 [14]. Each ZF recognizes a cognate 3 base pair (bp) 5'-GNN-3' consensus DNA sequence ('N' can be any of the DNA bases) [16]. There are 16 possible 5'-GNN-3' binding ZFs, which can theoretically be assembled into 256 two finger (2F) and 4096 3F combinations. Arabidopsis plants are transformed with the artificial 3F-VP16 encoding gene construct under control of the promoter of the Arabidopsis *RPS5a* gene [17], which is predominantly active in embryonic and meristematic tissue. Due to the relatively short DNA recognition sequence, each 3F-VP16 fusion has on average approximately 1000 binding sites in the 130 Mbp Arabidopsis genome, where it can influence the expression of nearby genes *in trans* and in a dominant manner through the activity of VP16. Introducing a single artificial gene in an otherwise wild type genome thus allows for the drastic perturbation of genome-wide gene expression patterns, and the potential induction of (novel) traits of interest.

Here, we describe the isolation and phenotypic characterization of a novel, recessive Arabidopsis mutant with a high ϕ PSII and low CF from a population of genome interrogation Arabidopsis plant lines harboring 3F-VP16 fusion encoding T-DNA constructs. This mutant was named Low Chlorophyll Fluorescence 1 (LCF1). Interestingly, LCF1 has a light quality dependent growth defect at standard light conditions, which can be rescued by short day and low light conditions, and by dark induced senescence.

Results

Genome interrogation library construction and isolation of LCF1

In order to investigate whether genome interrogation can be used to generate novel plant lines with photosynthesis phenotypes, a library of Arabidopsis plant lines harboring genome interrogation T-DNA constructs was screened for significant changes in ϕ PSII compared to the wild type Col-0 using CF imaging. Briefly, the genome interrogation library was constructed by floral dip transformation of Col-0 plants with a library of pRF-VP16-Kana binary vector constructs encoding 3F-VP16 fusions [14]. The genome interrogation library consisted of seeds that were harvested from primary transformants with these constructs (T1), and are therefore referred to as the second generation of transformants (T2). Similarly, the progeny of the secondary transformants is referred to as the tertiary generation of transformants (T3), and so on. The total complexity of the library was approximately 3500 individual genome interrogation constructs represented as T-DNAs in a total of 4278 lines. A population of genome interrogation T2 plants was screened for ϕ PSII values at 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$ of actinic that deviated strongly from Col-0. One mutant individual with high ϕ PSII and low CF was isolated, which was named Low Chlorophyll Fluorescence 1 (LCF1).

Genetic analysis of LCF1

By means of Southern blotting, only one copy of the 3F-VP16 encoding T-DNA construct was detected in the genome of LCF1 (Fig. 1). Through PCR analysis and sequencing with T-DNA specific primers, the T-DNA insert in LCF1 was demonstrated to encode a ZF-ATF with a 3F supposedly binding to the 9 bp sequence GCG-GTG-GCG. However, newly generated Col-0 plants that were transformed with a reconstituted 3F-VP16 encoding T-DNA construct that contained the identified 3F sequence did not display the high ϕ PSII phenotype, suggesting that the 3F-VP16 T-DNA construct from LCF1 does not induce the phenotype *in trans*. Moreover, when LCF1 (σ) was crossed with Col-0 (ρ) one quarter of the F2 offspring displayed the high ϕ PSII phenotype (Fig. S1), indicating that the LCF1 phenotype is inherited in a recessive manner. By means of Thermal Asymmetric Interlaced PCR (TAIL-PCR) the integration site of the T-DNA construct was demonstrated to be the 6th intron of the gene At4g36280 (chromosomal position 17165331-17169375 bp). With a combination of T-DNA and At4g36280 specific primers, it was confirmed that this locus indeed contained a full length T-DNA construct encoding the particular 3F-VP16 fusion. Remarkably, among the progeny of 15 available SALK At4g36280T-DNA knock-out lines of the Nottingham Arabidopsis Stock Center we did not find any plants that displayed the high ϕ PSII phenotype. To investigate whether the phenotypes of LCF1 are genetically linked to the insertion of the T-DNA in At4g36280, LCF1 was crossed with the accession Ler-0 to establish a mapping population of F2 hybrid plants. From this population, 110 individuals with high ϕ PSII and low

chlorophyll fluorescence were genotyped by PCR analysis using primers for the indel marker 4-AL022141-9227, which is located close to At4g3680 at the chromosomal position 17148559 bp. Only two out of the 220 chromosomes were found to be recombinant, indicating that the genetic element linked to the phenotypes of LCF1 is located within a genetic distance of approximately 1 cM from this marker. As none of the F2 individuals contained a wild type At4g3680 locus the phenotype of LCF1 must somehow be caused by the insertion of the 3F-ATF encoding T-DNA into the At4g36280 locus or by an adjacent mutation very close by. Regardless, LCF1 has a unique phenotype that has not yet been documented in literature, and therefore will be described in detail below.

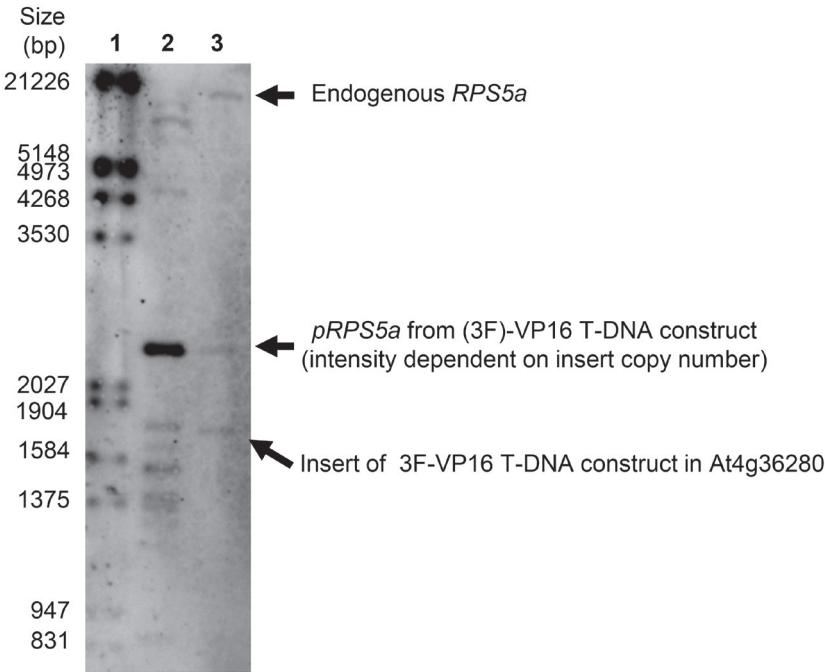


Fig. 1 Southern blot on *Nco*I digested genomic DNA (gDNA) of an empty vector control (Col-0 transformed with pRF-VP-Kana not containing a 3F fragment; lane 2) and LCF1 (lane 3). Lane 1 contains the DIG-labelled molecular weight marker. The DIG-labelled probe is expected to hybridize with fragments of the promoter of the endogenous *RPS5a* gene (10695 bp fragment), the *RPS5a* promoter (p*RPS5a*) from the pRF-VP16-Kana T-DNA construct (2590 bp fragment with intensity depending on the copy number of the insert), and the (3F)-VP16 fusion encoding part of each T-DNA insert in the genome (fragments of $\geq \sim 1500$ bp for LCF1 and $\geq \sim 800$ bp for the empty vector control without 3F). The empty vector control contains at least eight T-DNA inserts. LCF1 only contains one insert.

LCF1 has high ϕ PSII and low chlorophyll fluorescence

The ϕ PSII of LCF1 was typically 10-15% higher than that of Col-0 at $200 \mu\text{mol m}^{-2} \text{s}^{-1}$ of actinic light (Fig. 2A), which was observed at every time point in rosette development that

was analyzed. Based on preliminary gas exchange experiments we do not have evidence to conclude that this increase in ϕPSII is reflected by an increase in the rate of CO_2 fixation and therefore, the overall rate of photosynthesis. As our initial CF measurements were only performed at standard light conditions ($200 \mu\text{mol m}^{-2} \text{s}^{-1}$), we were also interested in the response of LCF1 to other light intensities. To this end, ϕPSII of Col-0 and LCF1 plants (T3) was quantified at a range of different light intensities with long adaptation steps to the actinic light, to ensure that the physiological responses of the plants to the changes in light intensity were reflected by ϕPSII . It was found that LCF1 plants had a significantly higher ϕPSII than Col-0 at every light intensity (Fig. 2B). As we had observed low steady state CF levels in LCF1, we were also interested in the dynamics of CF induction (the Kautsky effect; [18]). To this end, the induction of CF was examined by quantifying the fluorescence of dark adapted Col-0 and LCF1 plants (T3) upon illumination with $50 \mu\text{mol m}^{-2} \text{s}^{-1}$ actinic light. Although the time resolution of our CF imaging setup did not allow for quantification of the Kautsky effect on a micro to millisecond time scale, the fluorescence induction kinetics of LCF1 were clearly different from those of Col-0 (Fig. 3). The CF of LCF1 was consistently lower over the whole course of both the fast and slow transients of the Kautsky effect (referred to in literature as the Origin, Peak, Steady state (OPS) transient and the Steady State, Maximum, Terminal state (SMT) transient, respectively [19]), which can both be observed at a timescale of seconds. This possibly suggested that LCF1 plants have higher photochemical quenching activity, resulting in a substantial decrease in CF from the Origin of induction onwards.

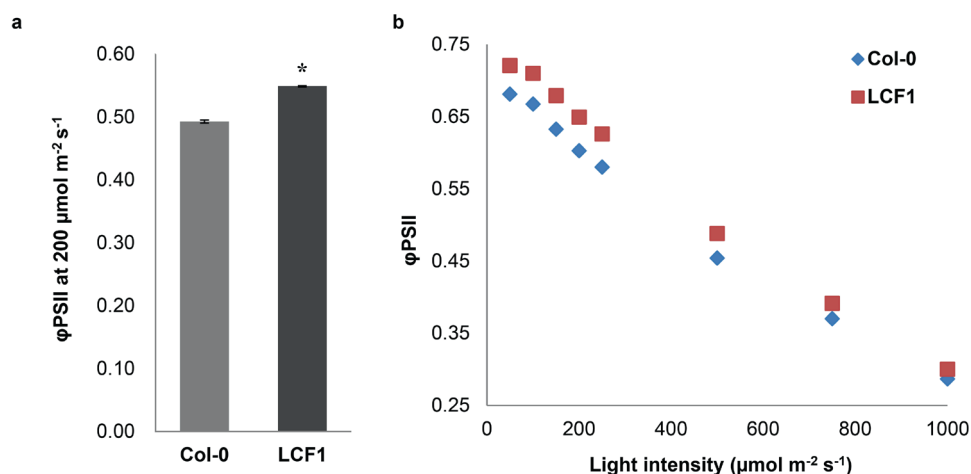


Fig. 2 Quantification of the operating efficiency of Photosystem II (ϕPSII) of LCF1 and wild type Col-0 plants using CF imaging. (A) Quantification of ϕPSII at standard light conditions of $200 \mu\text{mol m}^{-2} \text{s}^{-1}$ (14 dp). (B) Quantification of ϕPSII at a range of different light intensities (19 dp). Plants were adapted to every light intensity for 15 min. Asterisks (*) represent a significant difference with the wild type ($p < 0.05$). Error bars represent SEM values ($n=6$ per genotype) and are visible only when exceeding the data points.

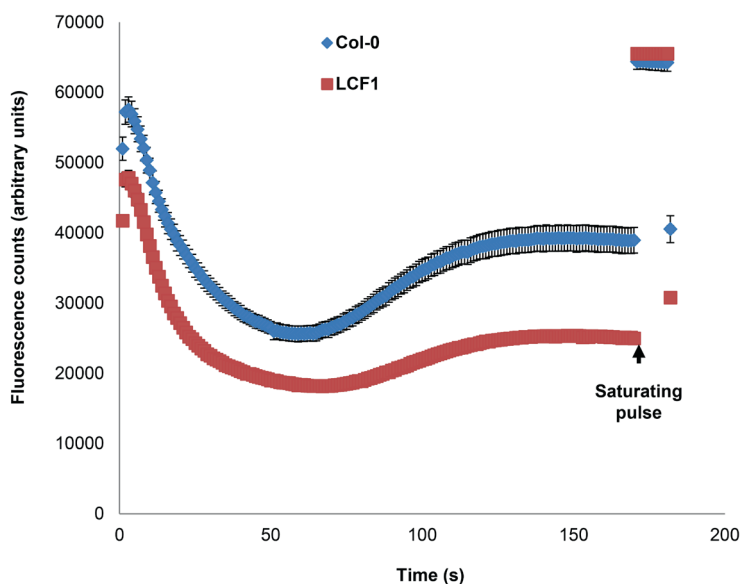


Fig. 3 Chlorophyll fluorescence induction traces in wild type Col-0 and LCF1 plants. Plants were dark adapted for ≥ 30 min and placed in the imaging chamber. Subsequently, actinic light of $50 \mu\text{mol m}^{-2} \text{s}^{-1}$ was switched on and the counts of fluorescence were quantified for 180 s, followed by the quantification of ϕPSII through a saturating pulse. Error bars represent SEM values ($n=12$ per genotype) and are visible only when exceeding the data points.

The high ϕPSII of LCF1 is due to an increase in photochemical quenching

To further examine the role of photochemical quenching in the high ϕPSII and low CF phenotypes of LCF1, light response curves were made for both the maximum efficiency of PSII in the light (F_v'/F_m') and the PSII efficiency factor (F_q'/F_v'), which is a measure for the fraction of all PSII reaction centers that are open and thereby relates the maximum efficiency to the operating efficiency ϕPSII [8]. A significant increase in the maximum efficiency of PSII in the light might mean that LCF1 has intrinsically more efficient PSII reaction centers, whereas a significant increase in photochemical quenching might mean that LCF1 has a larger fraction of PSII reaction centers performing photochemistry. There were no significant differences in the F_v'/F_m' values of LCF1 compared to Col-0 (Fig. 5A). However, F_q'/F_v' values of LCF1 were significantly higher than those of Col-0 (Fig. 5B). There were no significant changes in non-photochemical quenching (NPQ). Altogether, these results clearly suggest that the increase in ϕPSII of LCF1 is due to an increase in photochemical quenching.

The ϕPSII phenotype of LCF1 is robust and insensitive to light stress

All of the CF measurements described above were performed on LCF1 plants that were performing steady state photosynthesis at non-stressful light conditions. However, these conditions are not representative of field conditions, where plants are consistently exposed to

fluctuating, low or high light intensities. As the high ϕ PSII phenotype of LCF1 might also have agricultural value, we were interested in to what extent the phenotype would be sensitive to stressful light conditions. To examine this, Col-0 and LCF1 plants (T3) were exposed to either low ($20 \mu\text{mol m}^{-2} \text{s}^{-1}$), high ($1000 \mu\text{mol m}^{-2} \text{s}^{-1}$), very high ($2000 \mu\text{mol m}^{-2} \text{s}^{-1}$) or fluctuating ($50 \mu\text{mol m}^{-2} \text{s}^{-1}$ for 5 min, followed by $500 \mu\text{mol m}^{-2} \text{s}^{-1}$ for 1 min) actinic light intensities for a period of 1 h, before, during and after which ϕ PSII was quantified every 15 min (Fig. S2). From the resulting response curves it was concluded that, regardless of the light intensity, ϕ PSII was consistently higher in LCF1 than in Col-0. The relative increase in ϕ PSII during light stress treatment compared to Col-0 varied from approximately 6% to 10% (Fig. S2). LCF1 plants were also able to either partially or almost fully recover from every light stress treatment when they were again exposed to $200 \mu\text{mol m}^{-2} \text{s}^{-1}$ of actinic light (Fig. S2), clearly suggesting that the high ϕ PSII phenotype of LCF1 is robust and insensitive to stressful light conditions.

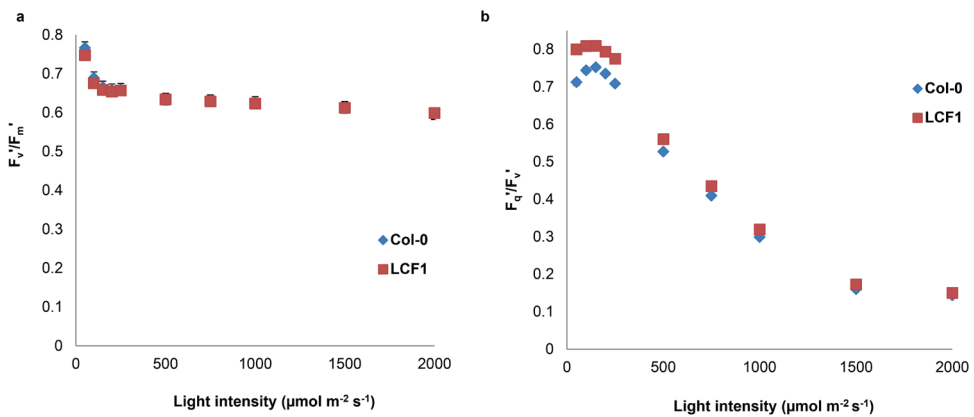


Fig. 4 Quantification of the maximum efficiency of Photosystem II in the light (F_v'/F_m') and the coefficient of photochemical quenching (F_q'/F_v') of LCF1 and wild type Col-0 plants (19 dpv) at different light intensities. Response curves were made by adapting the plants to every light intensity for 30s, followed by a saturating pulse. (A) Response curves of F_v'/F_m' to different light intensities. (B) Response curves of F_q'/F_v' to different light intensities. Error bars represent SEM values (n=12 per genotype) and are visible only when exceeding the data points.

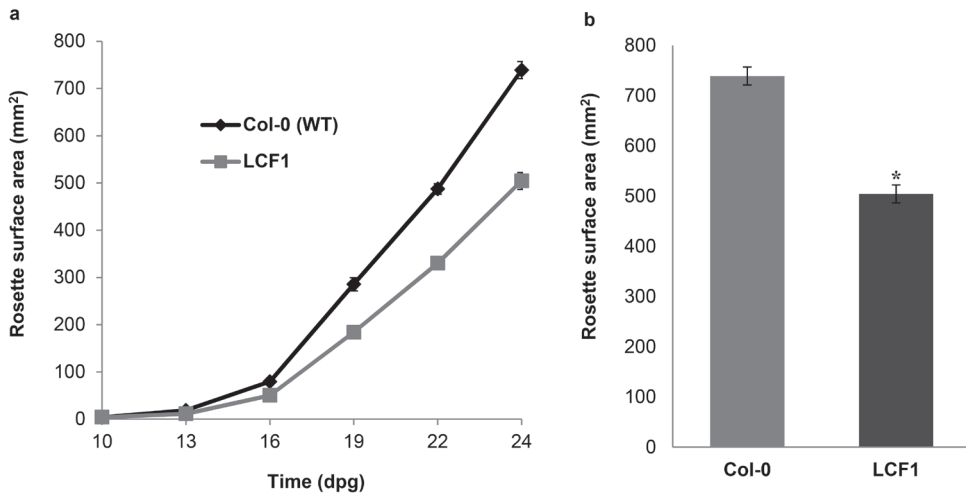


Fig. 5 Quantification of the rosette surface area throughout development of the wild type Col-0 and the mutant LCF1. **(A)** Growth curves of the wild type and mutant LCF1. **(B)** Rosette surface area at 24 dpg. Asterisks (*) represent a significant difference with the wild type ($p < 0.05$). Error bars represent SEM values ($n=36$ for Col-0, $n=99$ for LCF1), and are only visible when exceeding the data points.

The phenotype of LCF1 does not compare to natural variation in ϕ PSII

To our knowledge, no *Arabidopsis* mutants with a phenotype similar to that of LCF1 have been described in literature. To put the phenotype into some perspective, we performed a comparative analysis of ϕ PSII for plants of Col-0, LCF1 (T3) and eight natural *Arabidopsis* accessions with either significantly higher or lower ϕ PSII than Col-0 ([20]; personal communications with Pádraic Flood and Dr. Mark Aarts, Laboratory of Genetics, Wageningen UR, The Netherlands). The expected changes compared to Col-0 could be confirmed for LCF1 as well as for all of the natural accessions (Fig. S3A). Most notably, the increase in ϕ PSII of LCF1 compared to Col-0 was much higher than any of the more efficient natural accessions (Fig. S3A). The accession Lisse, which was the most efficient among the natural accessions, had an approximately 4% higher ϕ PSII than Col-0, whereas the ϕ PSII of LCF1 was approximately 15% higher (Fig. S3A). In these experiments LCF1 again displayed a reduction in rosette surface area as noted before (Fig. S3B). It was of interest to note that the natural accessions that we examined also did not display a positive correlation between ϕ PSII and growth (Fig. S3B).

LCF1 plants have a growth defect that is mediated by light quality dependent changes in photosynthetic activity

As LCF1 has a high ϕ PSII, one might expect it to accumulate more biomass. To examine whether LCF1 indeed accumulates significantly more biomass than Col-0, we quantified its rosette surface area (RSA) throughout development. We used RSA as a proxy for biomass, because these two parameters are strongly correlated in *Arabidopsis* [21]. Surprisingly, rather

than having a significantly larger RSA than Col-0, we found LCF1 to be consistently smaller throughout development (Fig. 5A). At 24 dpv, LCF1 had an approximately 32% smaller RSA than Col-0 (Fig. 5B). LCF1 plants did not exhibit any other conspicuous growth phenotypes, except for sometimes flowering later than Col-0. To examine whether this growth defect also occurs at light conditions other than standard (12 h photoperiod at 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$ PAR light; mercury lamps), Col-0 and LCF1 plants (T3) were first grown at standard light conditions for 14 days and were then transferred to short day and low light conditions (8 h photoperiod at 50 $\mu\text{mol m}^{-2} \text{s}^{-1}$ PAR light; fluorescent tubes), thereby reducing both the duration and the absolute amount of irradiance on the plants. There were no significant differences between LCF1 and Col-0 in terms of rosette surface area (Fig. S4A,B) and biomass (Fig. S4C,D) at these conditions. When these short day and low light conditions were supplemented with 2000 ppm of CO_2 , there was again a reduction in growth over time (Fig. 6). Since an increase in the ambient CO_2 concentration is stimulatory to photosynthesis [22, 23], these data altogether suggested that the growth defect of LCF1 at standard light conditions is caused by light intensity and/or quality dependent changes in photosynthesis. To further delineate the effect of light intensity and quality on the growth of LCF1, Col-0 and LCF1 plants (T3) were first grown at standard light conditions for 14 days and were then transferred to either high intensity red light (500 $\mu\text{mol m}^{-2} \text{s}^{-1}$ PAR in total; LEDs) or similarly intense blue light (180 $\mu\text{mol m}^{-2} \text{s}^{-1}$ PAR in total; LEDs) at otherwise unchanged conditions. Surprisingly, there was now an increase in the biomass of LCF1 in the case of red light (Fig. S5A) and a decrease in the case of blue light (Fig. S5B) compared to Col-0. Although these changes were not statistically significant, these results clearly suggest that the growth defect of LCF1 is mediated by light quality rather than intensity. Overall we conclude from this that the growth defect of LCF1 at standard light conditions is due to light quality-dependent changes in photosynthetic activity.

The growth of LCF1 can be rescued and increased by dark induced senescence

As LCF1 plants have low CF and a light quality-dependent growth phenotype, we hypothesized that both might be due to differences in the pigment composition (chlorophylls, β -carotene, xanthophylls) of PSII antennae. At standard growth conditions, we found no significant differences between LCF1 and Col-0 in terms of chlorophyll A and B content, and the chlorophyll A to chlorophyll B ratio. To examine the effect of chlorophyll degradation and pigment synthesis on ϕPSII non-destructively and *in vivo* [24], senescence was induced in Col-0 and LCF1 plants (T3) by adaptation to complete darkness from 21 dpv onwards at otherwise unchanged conditions. LCF1 plants had a significantly higher ϕPSII than Col-0 plants regardless of the number of days of dark incubation (DDI) (Fig. 7A), strongly suggesting that LCF1 has a higher ϕPSII irrespective of its (absolute) chlorophyll and/or pigment content. The plants were transferred back to standard light conditions at 14 DDI and after 14 days of recovery in the light, both Col-0 and LCF1 plants had regained ϕPSII values from before dark-induced senescence (Fig. 7B). Surprisingly, LCF1 plants now had a significantly larger

surface area than Col-0 (Fig. 7C), strongly suggesting that chlorophyll content and/or the synthesis of other pigments mediate the growth defect of LCF1.

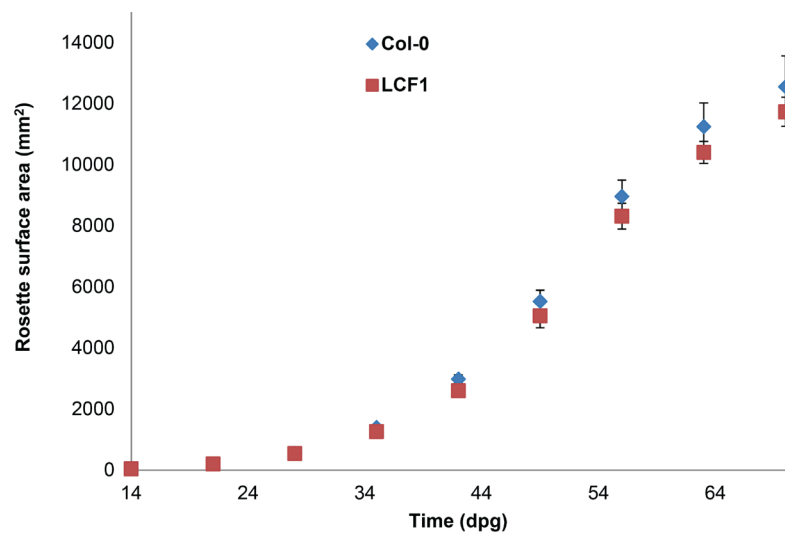


Fig. 6 Quantification of rosette surface area of wild type Col-0 and LCF1 plants grown at an 8 h photoperiod, a light intensity of 50 $\mu\text{mol m}^{-2} \text{s}^{-1}$ and 2000 ppm CO_2 . Plants were first grown at standard light conditions for 14 days and were then transferred to the indicated conditions. Error bars represent SEM values ($n=6$ per genotype), and are only visible when exceeding the data points.

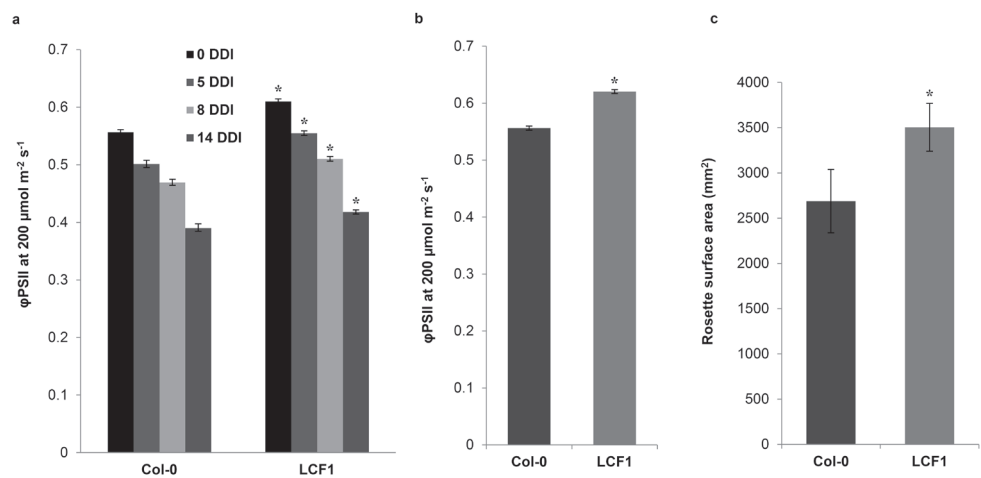


Fig. 7 Dark induced senescence of wild type Col-0 and LCF1 plants. Plants were initially grown at standard light conditions, and were transferred to complete darkness at otherwise unchanged conditions at 21 dpg. (A) ϕPSII measurements after 0, 5, 8 and 14 days of dark incubation (DDI), respectively. (B) Quantification of ϕPSII after 14 days of recovery in the light. (C) Quantification of rosette surface area after 14 days of recovery in the light. Asterisks (*) represent a significant difference with the wild type ($p < 0.05$ for ϕPSII values and $p < 0.1$ for rosette surface area values). Error bars represent SEM values ($n=7$ per genotype).

Discussion

Through screening a population of Arabidopsis plant lines harboring genome interrogation T-DNA constructs for changes in ϕ PSII, the recessive mutant LCF1 was isolated. LCF1 has a 10-15% higher ϕ PSII than Col-0, and has low CF levels. The high ϕ PSII of LCF1 is reflected by a significant increase in the coefficient of photochemical quenching, and is relatively insensitive to stressful light conditions. Rather than exhibiting an increase in growth as one might expect, LCF1 plants have a light quality dependent growth defect at standard light conditions compared to Col-0 plants, which can be alleviated by growing the plants at short day and low light conditions, or by inducing senescence through dark incubation.

To our knowledge, no other mutant screens in Arabidopsis for changes in ϕ PSII have been reported in literature. There are, however, a large number of publications available on screens for CF mutants. All of these screens have yielded mutants with high CF (HCF) phenotypes [11, 12, 25-27]. The most extensive screen for Arabidopsis mutants with HCF phenotypes was described by Meurer *et al.* in 1996, and yielded 34 recessive *hcf* mutants from a population of 7700 ethyl methane sulfonate (EMS)-mutagenized Arabidopsis plants [11]. Remarkably, we have not encountered any conspicuous HCF mutants in our population of 4278 plant lines. To our knowledge, no low CF mutants with an apparent gain in PSII function like LCF1 have previously been isolated in Arabidopsis, or other species for that matter. A (saturating) mutant screen for ϕ PSII mutants has most likely never been performed in Arabidopsis, because quantifying ϕ PSII can at best be made semi-high throughput and screening a large population of thousands of mutant plants is very laborious and requires sophisticated equipment.

The genetic evidence that we have gathered showed that the phenotypes of LCF1 are linked to the insertion of a 3F-VP16 encoding T-DNA construct in At4g36280 (*AtMORC2*). However, none of the available SALK T-DNA insertion lines of *AtMORC2* displayed the high ϕ PSII phenotype and retarded growth observed in LCF1. The phenotypes of LCF1 are not due to a combination of the knock-out of *AtMORC2* and the activity of the 3F-VP16 encoded by the transgene insert, as transformation of the SALK T-DNA insertion lines for *AtMORC2* with the same 3F-VP16 construct did not result in transformants that displayed the phenotypes of LCF1 (data not shown). We have also attempted to investigate whether the phenotypes of LCF1 are due to the combination of a knock-out of *AtMORC2* and highly insertion locus specific 3F-VP16 expression by crossing several of the SALK insertion lines with LCF1, but these crosses did not yield any viable F1 seeds, indicating that these genotypes cannot produce viable progeny. It is therefore most likely that we have isolated a novel and recessive gain of function *AtMORC2* insertion allele. *AtMORC2* encodes an ATPase of the Microrchidia (MORC) family that in plants and animals is involved in the large scale epigenetic transcriptional silencing of transposons through the condensation of heterochromatin [28],

possibly suggesting that the phenotypes of LCF1 are due to novel epigenetic regulation of AtMORC2 target loci.

The most striking feature of LCF1 is its low CF phenotype. The CF induction kinetics of LCF1 are very different from Col-0 (Fig. 2), and CF of LCF1 is much lower than Col-0 over the course of the whole Kautsky effect from the Origin onwards (Fig. 3). Although LCF1 performs more steady state photochemical quenching than Col-0 (Fig. 5B), photochemistry, NPQ and the Calvin cycle in principle require time to reach steady state activity on the time scale of CF induction. Altogether, these observations could suggest that CF in LCF1 is quenched by an alternative electron sink downstream of PSII that is more rapidly light inducible than photochemistry and NPQ, or not induced by light at all. At this point, we do not have enough evidence to pinpoint which particular sink would function as this electron acceptor. The water-water cycle (WWC) could be considered as a candidate, as it has been reported to act as a major sink for linearly transported electrons [29]. Changes in WWC activity are known to be reflected by changes in ϕ PSII [30]. Therefore, the high ϕ PSII phenotype of LCF1 could be due to enhanced WWC activity. Alternatively, the PSII/PSI ratio of LCF1 might be different from Col-0. A lower PSII/PSI ratio would theoretically allow for the more efficient quenching of PSII fluorescence that we observe during the Kautsky effect of LCF1. In support of this hypothesis, the high ϕ PSII phenotype of LCF1 was found to be robust and insensitive to light stress, which clearly suggests that the thylakoid membranes of LCF1 perform intrinsically more efficient photochemistry.

As LCF1 has substantially higher ϕ PSII than Col-0, one might expect it to have a correspondingly higher overall rate of photosynthesis and, therefore, to accumulate more biomass. However, LCF1 has stunted rather than enhanced growth compared to Col-0 at standard growth conditions. There is some debate about whether or not more efficient photosynthesis will result in an increase in growth and yield [23], but the Arabidopsis accessions that we have analyzed do not display any correlation between ϕ PSII and growth. There are both relatively efficient and inefficient accessions that have a large surface area compared to Col-0. Of the three most commonly used background accessions, Ws-4, has the lowest ϕ PSII, but a larger surface area than both Col-0 and Ler-0 [9]. The ϕ PSII of LCF1 was much higher than any of the natural accession that we have studied, but it had by far the smallest surface area, thereby clearly suggesting that there is no positive correlation between ϕ PSII and biomass accumulation. We have, however, been able to rescue LCF1 plants from this growth defect by growing them at short day and low light conditions, at intense red light conditions, and growth could even be enhanced compared to Col-0 through dark induced senescence. This demonstrates that conditions can be found at which the high ϕ PSII of LCF1 is beneficial to its growth.

In conclusion, LCF1 has provided us with the novel high ϕ PSII and low chlorophyll fluorescence phenotype that has never before been documented in literature. Therefore, LCF1

might be a valuable tool for exploring how more efficient photosynthesis can be achieved in *Arabidopsis* at the level of PSII. As mentioned, LCF1 plants exhibit retarded growth, thereby questioning the positive correlation between photosynthesis and growth that is often postulated. We have demonstrated that the growth of LCF1 can be controlled by the manipulation of growth conditions, and can even be enhanced through the induction of dark induced senescence, which demonstrates that the high ϕ PSII of LCF1 can be profitable for biomass accumulation in *Arabidopsis*. Through further research, we will hopefully be able to learn how to convert the 10 to 15% increase in ϕ PSII of LCF1 into substantially more efficient overall photosynthesis and biomass accumulation in plant species that are of commercial interest.

Materials and Methods

Growth conditions

All plants in this study were grown on soil in a climate-controlled growth chamber at a constant temperature of 20 °C, 70% relative humidity, and at a light intensity of approximately 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$ of photosynthetically active radiation (PAR) from mercury lamps, unless otherwise specified. These light conditions are referred to as 'standard light conditions'. All plants were grown at a 12 h photoperiod, unless otherwise specified. Plants grown at a low light intensity were firstly grown at standard light conditions, and were transferred to a growth cabinet with fluorescent tubes at a PAR light intensity of approximately 50 $\mu\text{mol m}^{-2} \text{s}^{-1}$ at 14 dpg. Plants grown at high light intensities were firstly grown at standard light conditions, and at 14 dpg were transferred to either red (625-680 nm) LED light supplemented with LED daylight (480 and 20 $\mu\text{mol m}^{-2} \text{s}^{-1}$ PAR, respectively) or blue (450-480 nm) LED light supplemented with LED daylight (160 and 20 $\mu\text{mol m}^{-2} \text{s}^{-1}$ PAR, respectively), at otherwise unchanged conditions.

Plant material and genome interrogation library construction

The *Arabidopsis* accession Columba-0 (Col-0) was used as the wild type in this study. Seeds of the other accessions were kindly provided by Dr. Mark Aarts (Laboratory of Genetics, Wageningen UR, The Netherlands). SALK T-DNA insertion knock-out lines for the genomic locus At4g36280 (Table 1) were obtained from the Nottingham *Arabidopsis* Stock Centre. A library of T-DNA constructs encoding fusions of three zinc fingers (3Fs) to the transcriptional activator VP16 was constructed in the binary vector pRF-VP16-Kana [14]. These constructs were transformed to Col-0 plants using the floral dip method [31]. Primary transformants (T1) were selected on MA medium containing 35 $\mu\text{g/mL}$ kanamycin, and were transferred to soil after approximately 2 weeks, with five plants per pot. The primary transformants were then raised, and their seeds were harvested. Due to plant loss and infertility of a fraction of the

plants, not all plants produced progeny. Seeds of five or less primary transformants (T2 seeds) originating from the same 3F pool were combined and stored in a single seed bag. These bags were named ‘five-bags’. The genome interrogation library consisted of the T2 seeds of 4278 plant lines in total representing approximately 3500 3F-VP16 encoding constructs, and were stored in a total of 1034 five-bags.

Table 1. The SALK T-DNA insertion lines for the genomic locus At4g36280 that were used in this study.

NASC Code	Accession name
N521267	SALK_021267
N572456	SALK_072456
N572774	SALK_072774
N585402	SALK_085402
N585404	SALK_085404
N589534	SALK_089534
N623179	SALK_123179
N635895	SALK_135895
N636332	SALK_136332
N636502	SALK_136502
N636598	SALK_136598
N636599	SALK_136599
N645794	SALK_145794
N669832	SALK_072774C
N674058	SALK_021267C

Library screening for ϕ PSII mutants

Approximately 20 seeds from every five-bag were sown on soil in 67 x 67 x 65 mm pots (Pöppelmann, Lohne, Germany). Every five-bag was assigned to its own pot. The seeds were stratified on soil for 3-5 days at 4 °C. The ϕ PSII of the plants was quantified at 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$ of actinic light (the same PAR light intensity as in the growth chamber) at 14 dpv, using a Technologica CF Imager (Technologica, Colchester, United Kingdom). The ϕ PSII (F_q'/F_m' ; [8]) images were generated after 30 s of exposure to the actinic light with a 800 ms saturating pulse of 6226 $\mu\text{mol m}^{-2} \text{s}^{-1}$ (maximal intensity).

Plant retransformation with reconstituted T-DNA constructs

The 3F-VP16 encoding fragment was isolated from LCF1 by PCR using the T-DNA specific primer combination RPS5a FW and tNOS REV (Table 2). Using the 3F encoding sequence from this PCR product a reconstituted T-DNA construct was generated in the binary vector pRF-VP16-Kana [14]. Col-0 plants were transformed with this T-DNA construct using the floral dip method [31]. Primary retransformants were selected by plating sterilized seeds on MA medium containing 35 $\mu\text{g/mL}$ kanamycin, and transferred to soil after 2-3 weeks.

Table 2. Primers used for the genetic analysis of LCF1. Nucleotide codes: ‘N’ is any of the four bases; ‘W’ is either A or T; ‘S’ is either G or C.

Primer name	Sequence (5’-3’)
NOS1 FW	GATTGAATCCTGTTGCCGGTCTT
NOS2 FW	GCATGACGTTATTTATGAGATGG
NOS3 FW	CGCAAACCTAGGATAAAATTATCGC
AD1	NTCGASTWTSGWGTT
AD2	NGTCGASWGANAWGAA
AD3	WGTGNAGWANCANAGA
At4g36280 FWa	ACGGAGTAGTAGGAGGAAGAGA
At4g36280 FWb	TTGGAAGGCGGGAGATTACT
At4g36280 REVa	CTTTTTC AACCTCGCCTCCA
At4g36280 REVb	TGTAGGTTGTGGGTTGAGCTG
RPS5a FW	GCCCAAACCCTAAATTTCTCATC
tNOS REV	CAAGACCGGCAACAGGAT
171 FW	ACTTGTTTGGTATTTGTCTC
171 REV	AATTCTACGGATAAGTTCAG
880 p1 FW	GGTTTCTCTCCTTTCTT
1325 p1 REV	TCCTGACCAGTTTTTCT

Chlorophyll fluorescence analysis

All chlorophyll fluorescence measurements were performed on Col-0 and LCF1 plants (T3) at either 14, 21, or 28 dpv. The quantification of ϕ PSII was performed as described above, with adaptation steps to the actinic light lasting 30 to 60 s, depending on the number of plants that had to be analyzed. For the quantification of ϕ PSII at different light intensities (‘slow’ light response curves), plants were adapted to every light intensity for 15 min instead of 30 s as described above. The maximum quantum efficiency of PSII photochemistry (F_v/F_m) was also quantified using the Technologica CF Imager (Technologica, Essex, United Kingdom) on dark adapted (≥ 30 min) plants. F_m images were generated with a 800 ms saturating pulse of $6226 \mu\text{mol m}^{-2} \text{s}^{-1}$ (maximal intensity).

‘Fast’ light response curves of ϕ PSII were made on dark adapted (≥ 30 min) plants after F_v/F_m quantification, thereby also allowing for the calculation of NPQ, F_q'/F_v' and F_v'/F_m' images, respectively. For the quantification of the Kautsky effect, dark adapted (≥ 30 min) plants were exposed to $50 \mu\text{mol m}^{-2} \text{s}^{-1}$ of actinic light, and the absolute counts of fluorescence were quantified for 3 min using the Technologica CF Imager, followed by quantification of ϕ PSII as described above. For the quantification of the sensitivity of ϕ PSII to light stress in LCF1, ϕ PSII was quantified every 15 min for 1 h at $200 \mu\text{mol m}^{-2} \text{s}^{-1}$ of actinic light as described above, then conditions were switched to either low ($20 \mu\text{mol m}^{-2} \text{s}^{-1}$), fluctuating ($50 \mu\text{mol m}^{-2} \text{s}^{-1}$ for 5 min, followed by $500 \mu\text{mol m}^{-2} \text{s}^{-1}$ for 1 min), high ($1000 \mu\text{mol m}^{-2} \text{s}^{-1}$) or very high ($2000 \mu\text{mol m}^{-2} \text{s}^{-1}$) actinic light. ϕ PSII was then quantified again every 15 min for 1 h (except for fluctuating light conditions, where ϕ PSII was quantified after every 5 min of $50 \mu\text{mol m}^{-2} \text{s}^{-1}$ light and every 1 min of $500 \mu\text{mol m}^{-2} \text{s}^{-1}$), followed by a quantification of ϕ PSII

every 15 min for 6 h at $200 \mu\text{mol m}^{-2} \text{s}^{-1}$. The calculation of the CF parameters was performed with the average of all pixels in every rosette. All CF data were statistically analyzed using the heteroscedastic T-Test function of Microsoft Excel 2010 (assuming unequal variance between samples). A p -value of 0.05 was used as a threshold for significance.

Rosette surface area and biomass quantification

For rosette surface area quantification, approximately 50 seeds of Col-0 and LCF1 (T3 generation) were sown in pots with a diameter of 15.7 cm and height of 65 mm (Soparco, Condé-sur-Huisne, France), and stratified for 3 days at 4 °C. Seedlings were transferred to 67 x 67 x 65 mm pots (Pöppelmann, Lohne, Germany) at 7dpg. Photos were taken of all trays from the top with a fixed digital camera (Canon EOS 1100D) from 10 dpg onwards and every three days. Using an ImageJ plugin, the intensity of the green channel of these RGB images was multiplied by two, both the red and blue channels were subtracted and the image was converted to a binary image using the ImageJ 'Intermodes' Threshold Method. The binary images were manually inspected to make sure that all the leaves of each individual plant were connected to each other as a rosette. In cases where leaves were not properly connected, they were connected manually to the rest of the rosette by a black line of two pixels in width. The rosette surface area of each rosette was subsequently calculated in pixel^2 using the 'Analyze Particles' function of ImageJ, and then converted to mm^2 by multiplying this value by the $\text{mm}^2/\text{pixel}^2$ ratio of every RGB image separately. This ratio was calculated from the dimensions of the pots that the plants were growing in, as this value is constant for every individual throughout the experiment. All rosette surface area data were statistically analyzed using the heteroscedastic T-Test function of Microsoft Excel 2010 (assuming unequal variance between samples). A p -value of 0.05 was used as a threshold for significance.

Dark induced senescence

Seeds of Col-0 and mutant LCF1 (T3 generation) were sown in 67 x 67 x 65 mm pots (Pöppelmann, Lohne, Germany) and stratified for 3 days at 4 °C. At 21 dpg, ϕPSII was quantified at $200 \mu\text{mol m}^{-2} \text{s}^{-1}$ of actinic light as described above and plants were subsequently transferred to complete darkness at otherwise unchanged conditions. At 5 days of dark incubation (DDI), 8 DDI and 14 DDI, respectively, plants were temporarily adapted to $200 \mu\text{mol m}^{-2} \text{s}^{-1}$ of PAR light in the greenhouse for approximately 3 h and ϕPSII was quantified at $200 \mu\text{mol m}^{-2} \text{s}^{-1}$ of actinic light, as described above. At 14 DDI, plants were transferred back to standard light conditions. After 14 days of incubation in the light (now 49 dpg), ϕPSII was quantified as described above. Rosette surface area data for all individuals were also obtained from the fluorescence images (in mm^2). Data were statistically analyzed using the heteroscedastic T-Test function of Microsoft Excel 2010 (assuming unequal variance between samples). A p -value of 0.1 was used as a threshold for significance.

Genetic analysis of LCF1

For segregation analysis, LCF1 (♂) was crossed with Col-0 (♀). F1 individuals were selected for presence of the 3F-VP16 T-DNA construct on half strength MS medium containing 50 µg/mL of kanamycin. After 2-3 weeks on selection medium, resistant individuals were transferred to soil. The F2 seeds were harvested and ϕ PSII of a population of F2 individuals was quantified as described above. The T-DNA copy number in the genome of LCF1 was examined using the Southern blot method [32] on *Nco*I predigested genomic DNA isolated with the CTAB extraction protocol [33]. As a probe, DIG-labelled PCR product was generated with the primer combination RPS5a FW and tNOS REV (Table 2) from LCF1 genomic DNA, using PCR DIG Labelling Mix (Roche). Detection of hybridization signal was performed with the DIG Wash and Block Buffer Set (Roche), according to the instructions provided by the manufacturer. The insertion site of the 3F-VP16 encoding T-DNA construct in the genome of LCF1 was mapped via TAIL-PCR [34], using forward primers NOS1 FW, NOS2 FW and NOS3 FW (Table 2), respectively, for the three consecutive PCR reactions each with one of degenerative primers AD1, AD2 and AD3 (Table 2), respectively. TAIL-PCR products were cloned into the pJET Blunt cloning vector using the CloneJet PCR Cloning Kit (Thermo Scientific), and sequenced by Sanger sequencing (Macrogen Europe, Amsterdam, The Netherlands). The 3' regions of the PCR products were BLASTed (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) for identification of the insertion locus. For the confirmation of insertion locus At4g36280, PCRs were performed using combinations of At4g36280 specific primers (FWa, FWb, REVa, REVb; Table 2) and combinations of At4g36280 and T-DNA specific primers (RPS5a FW, tNOS REV; Table 2).

Genetic mapping

To establish a population of plants for genetic mapping LCF1 (♂) was crossed with the *Arabidopsis* accession Landsberg erecta (Ler-0; ♀). Hybrid F1 plants were allowed to set seeds (F2) and ϕ PSII of a population of F2 seedlings was quantified at 200 µmol m⁻² s⁻¹ actinic light at 7 dpg as described above. Genomic DNA was isolated from a single young leaf of individuals exhibiting the high ϕ PSII phenotype using a quick DNA extraction protocol [35]. The individuals were then genotyped by PCR analysis using primers 171 FW and REV (Table 2) for the indel marker 4-AL022141-9227 (<http://amp.genomics.org.cn>, [36]), which yields a 157 bp product for a Col-0 sequence and a 134 bp product for a Ler-0 sequence. The presence of wild type At4g36280 loci or loci with T-DNA inserts was determined by PCR analysis with the primer combination 880 FW and 1325 p1 REV (Table 2).

Acknowledgements

We would like to thank the laboratory of Dr. Mark Aarts (Laboratory of Genetics, Wageningen UR, the Netherlands) for kindly providing seeds of the natural *Arabidopsis* accessions used in this study. We would like to thank Prof. Dr. H.J.M. de Groot and Dr. A. Alia Matysik (Leiden Institute of Chemistry, Leiden, the Netherlands) for stimulating discussions, and Dr. Silvia Lang for critically reading the manuscript. This project was carried out within the research programme of BioSolar Cells, co-financed by the Dutch Ministry of Economic Affairs.

References

1. Barber J. (2009) Photosynthetic energy conversion: natural and artificial. *Chem Soc Rev* 38(1):185-96.
2. Evans JR. (2013) Improving photosynthesis. *Plant Physiol* 162(4):1780-93.
3. Zhu XG, Long SP, Ort DR. (2010) Improving photosynthetic efficiency for greater yield. *Annu Rev Plant Biol* 61:235-61.
4. Peterhansel C, Maurino VG. (2011) Photorespiration redesigned. *Plant Physiol* 155(1):49-55.
5. Miyao M. (2003) Molecular evolution and genetic engineering of C4 photosynthetic enzymes. *J Exp Bot* 54(381):179-89.
6. Parry MA, Andralojc PJ, Mitchell RA, Madgwick PJ, Keys AJ. (2003) Manipulation of Rubisco: the amount, activity, function and regulation. *J Exp Bot* 54(386):1321-33.
7. Muller P, Li XP, Niyogi KK. (2001) Non-photochemical quenching. A response to excess light energy. *Plant Physiol* 125(4):1558-66.
8. Baker NR. (2008) Chlorophyll fluorescence: a probe of photosynthesis in vivo. *Annu Rev Plant Biol* 59:89-113.
9. Yin L, Fristedt R, Herdean A, Solymosi K, Bertrand M, Andersson MX, *et al.* (2012) Photosystem II function and dynamics in three widely used Arabidopsis thaliana accessions. *PloS One* 7(9):e46206.
10. El-Lithy ME, Rodrigues GC, van Rensen JJ, Snel JF, Dassen HJ, Koornneef M, *et al.* (2005) Altered photosynthetic performance of a natural Arabidopsis accession is associated with atrazine resistance. *J Exp Bot* 56(416):1625-34.
11. Meurer J, Meierhoff K, Westhoff P. (1996) Isolation of high-chlorophyll-fluorescence mutants of Arabidopsis thaliana and their characterisation by spectroscopy, immunoblotting and northern hybridisation. *Planta* 198(3):385-96.
12. Shikanai T, Munekage Y, Shimizu K, Endo T, Hashimoto T. (1999) Identification and characterization of Arabidopsis mutants with reduced quenching of chlorophyll fluorescence. *Plant Cell Physiol* 40(11):1134-42.
13. van Tol N, van der Zaal BJ. (2014) Artificial transcription factor-mediated regulation of gene expression. *Plant Sci* 225:58-67.
14. Lindhout BI, Pinas JE, Hooykaas PJ, van der Zaal BJ. (2006) Employing libraries of zinc finger artificial transcription factors to screen for homologous recombination mutants in Arabidopsis. *Plant J* 48(3):475-83.
15. Jia Q, van Verk MC, Pinas JE, Lindhout BI, Hooykaas PJ, van der Zaal BJ. (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 Biotech J* 11(9):1069-79.
16. Segal DJ, Dreier B, Beerli RR, Barbas CF, 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(6):2758-63.
17. Weijers D, Franke-van Dijk M, Vencken RJ, Quint A, Hooykaas P, Offringa R. (2001) An Arabidopsis Minute-like phenotype caused by a semi-dominant mutation in a RIBOSOMAL PROTEIN S5 gene. *Development* 128(21):4289-99.
18. Stirbet A, Govindjee. (2011) On the relation between the Kautsky effect (chlorophyll a fluorescence induction) and Photosystem II: basics and applications of the OJIP fluorescence transient. *J Photochem Photobiol B* 104(1-2):236-57.
19. Papageorgiou GC, Tsimilli-Michael M, Stamatakis K. (2007) The fast and slow kinetics of chlorophyll a fluorescence induction in plants, algae and cyanobacteria: a viewpoint. *Photosynth Res* 94(2-3):275-90.

20. van Rooijen R, Aarts MG, Harbinson J. (2015) Natural genetic variation for acclimation of photosynthetic light use efficiency to growth irradiance in *Arabidopsis*. *Plant Physiol* 167(4):1412-29.
21. Leister D, Varotto C, Pesaresi P, Niwergall A, Salamini F. (1999) Large-scale evaluation of plant growth in *Arabidopsis thaliana* by non-invasive image analysis. *Plant Physiol Bioch* 37(9):671-8.
22. Campbell WJ, Allen LH, Bowes G. (1988) Effects of CO₂ Concentration on Rubisco Activity, Amount, and Photosynthesis in Soybean Leaves. *Plant Physiol* 88(4):1310-6.
23. Kirschbaum MU. (2011) Does enhanced photosynthesis enhance growth? Lessons learned from CO₂ enrichment studies. *Plant Physiol* 155(1):117-24.
24. Sakuraba Y, Park SY, Kim YS, Wang SH, Yoo SC, Hortensteiner S, *et al.* (2014) *Arabidopsis* STAY-GREEN2 is a negative regulator of chlorophyll degradation during leaf senescence. *Mol Plant* 7(8):1288-302.
25. Dinkins RD, Bandaranayake H, Baeza L, Griffiths AJ, Green BR. (1997) *hcf5*, a nuclear photosynthetic electron transport mutant of *Arabidopsis thaliana* with a pleiotropic effect on chloroplast gene expression. *Plant Physiol* 113(4):1023-31.
26. Lennartz K, Plucken H, Seidler A, Westhoff P, Bechtold N, Meierhoff K. (2001) HCF164 encodes a thioredoxin-like protein involved in the biogenesis of the cytochrome b(6)f complex in *Arabidopsis*. *Plant Cell* 13(11):2539-51.
27. Schmitz J, Schottler MA, Krueger S, Geimer S, Schneider A, Kleine T, *et al.* (2012) Defects in leaf carbohydrate metabolism compromise acclimation to high light and lead to a high chlorophyll fluorescence phenotype in *Arabidopsis thaliana*. *BMC Plant Biol* 12:8. doi: 10.1186/1471-2229-12-8.
28. Moissiard G, Bischof S, Husmann D, Pastor WA, Hale CJ, Yen L, *et al.* (2014) Transcriptional gene silencing by *Arabidopsis* microorchidia homologues involves the formation of heteromers. *Proc Natl Acad Sci USA* 111(20):7474-9.
29. Miyake C. (2010) Alternative electron flows (water-water cycle and cyclic electron flow around PSI) in photosynthesis: molecular mechanisms and physiological functions. *Plant Cell Physiol* 51(12):1951-63.
30. Miyake C, Shinzaki Y, Nishioka M, Horiguchi S, Tomizawa K. (2006) Photoinactivation of ascorbate peroxidase in isolated tobacco chloroplasts: *Galdieria partita* APX maintains the electron flux through the water-water cycle in transplastomic tobacco plants. *Plant Cell Physiol* 47(2):200-10.
31. Clough SJ, Bent AF. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16(6):735-43.
32. Southern E. (2006) Southern blotting. *Nat Protoc* 1(2):518-25.
33. Richards E, Reichardt M, Rogers S. (2001) Preparation of genomic DNA from plant tissue. *Curr Protoc Mol Biol* chapter 2, unit 2.
34. Liu YG, Mitsukawa N, Oosumi T, Whittier RE. (1995) Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J* 8(3):457-63.
35. Edwards K, Johnstone C, Thompson C. (1991) A Simple and Rapid Method for the Preparation of Plant Genomic DNA for PCR Analysis. *Nucleic Acids Res* 19(6):1349.
36. Hou X, Li L, Peng Z, Wei B, Tang S, Ding M, *et al.* (2010) A platform of high-density INDEL/CAPS markers for map-based cloning in *Arabidopsis*. *Plant J* 63(5):880-8.

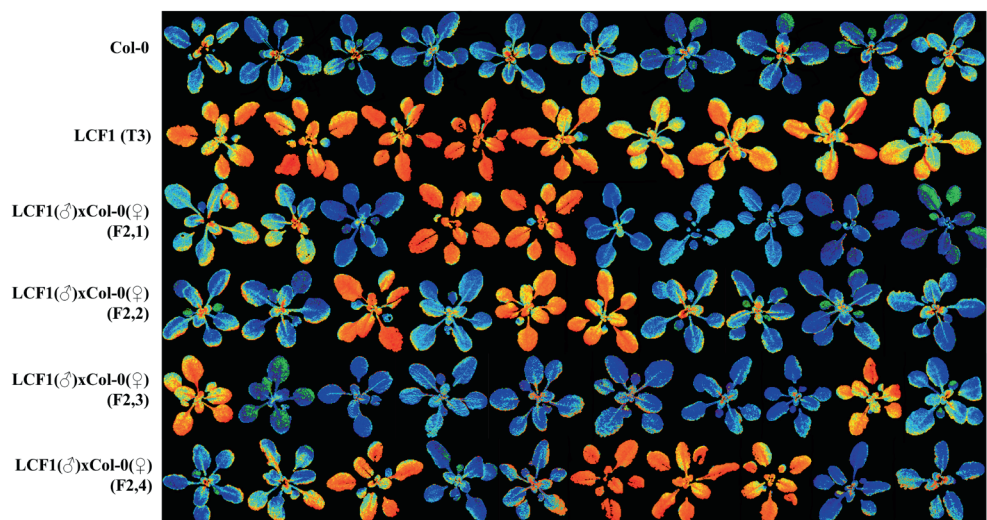


Fig. S1 False color ϕ PSII images of 10 individuals each of the wild type, LCF1 (T3 generation), and four independent F2 progeny lines (F2,1-4) of a cross between LCF1 (pollen donor, ♂) and the wild type (♀). Plants were grown at standard light conditions, and ϕ PSII of every individual was quantified at $200 \mu\text{mol m}^{-2} \text{s}^{-1}$ of actinic light. Orange colored individuals have the high ϕ PSII phenotype; blue colored individuals have wild type ϕ PSII values. Images do not provide a quantitative measure of ϕ PSII.

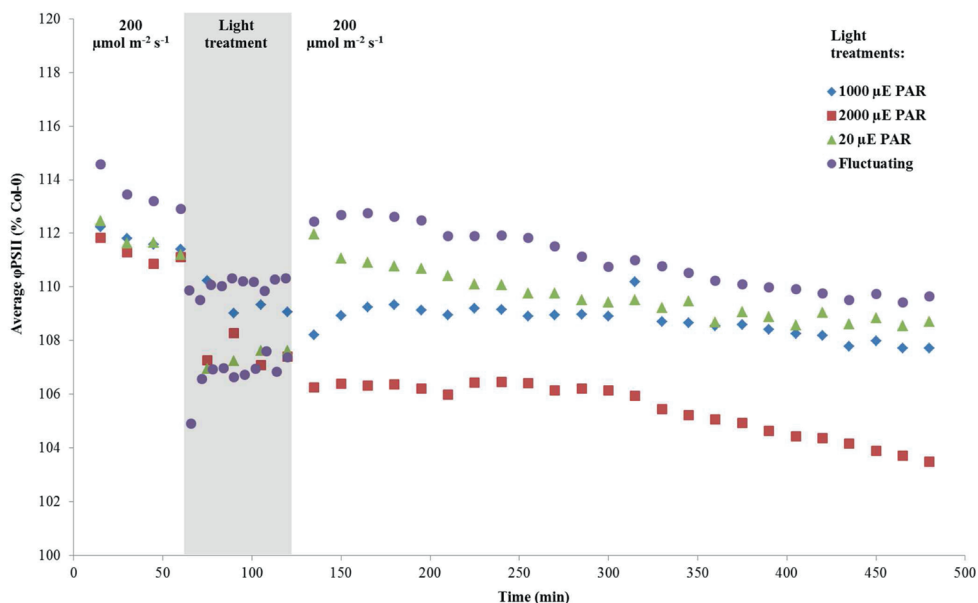


Fig. S2 The relative response of the operating efficiency of Photosystem II (ϕ PSII) of LCF1 plants to light stress treatments compared to wild type Col-0 plants. The ϕ PSII of LCF1 and Col-0 plants that were grown at standard light conditions of $200 \mu\text{mol m}^{-2} \text{s}^{-1}$ was quantified every 15 min for 1 h at $200 \mu\text{mol m}^{-2} \text{s}^{-1}$ of actinic light. Subsequently, as a light stress treatment, plants were exposed to actinic light of high intensity (1000 or $2000 \mu\text{mol m}^{-2} \text{s}^{-1}$; 14 and 16 dpg respectively), low intensity ($20 \mu\text{mol m}^{-2} \text{s}^{-1}$; 21 dpg) or fluctuating intensity ($50 \mu\text{mol m}^{-2} \text{s}^{-1}$ for 5 min, followed by $500 \mu\text{mol m}^{-2} \text{s}^{-1}$ for 1 min; 24 dpg), and ϕ PSII was again quantified every 15 min for 1h (except for fluctuating light conditions, were ϕ PSII was quantified after every 5 min of $50 \mu\text{mol m}^{-2} \text{s}^{-1}$ light and after every 1 min of $500 \mu\text{mol m}^{-2} \text{s}^{-1}$). After the light stress treatment, standard light conditions were switched back on and ϕ PSII was quantified every 15 min for 6 h. The response curves represent the average ϕ PSII of LCF1 relative to that of the wild type ($n \geq 4$ each).

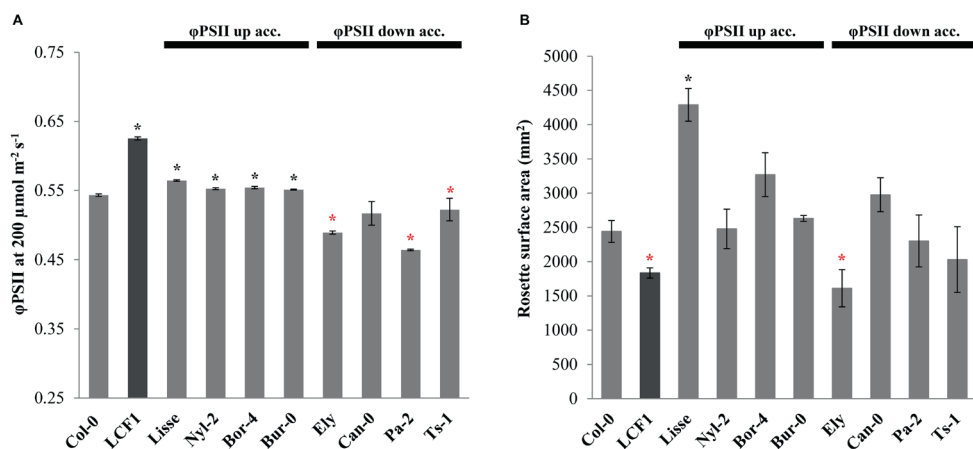


Fig. S3 A comparison of the operating efficiency of Photosystem II (ϕPSII) at $200 \mu\text{mol m}^{-2} \text{s}^{-1}$ of actinic light (**A**) and rosette surface area (**B**) between the wild type Col-0, LCF1 and eight natural *Arabidopsis* accessions with either high ϕPSII (ϕPSII up acc.) or low ϕPSII (ϕPSII down acc.) compared to Col-0 (28 dp). Asterisks (*) represent a significant difference with Col-0 ($p < 0.05$). Black asterisks represent significant increases; red asterisks represent significant decreases. Error bars represent SEM values ($n=6$ per genotype for ϕPSII ; $n=4$ per genotype for rosette surface area).

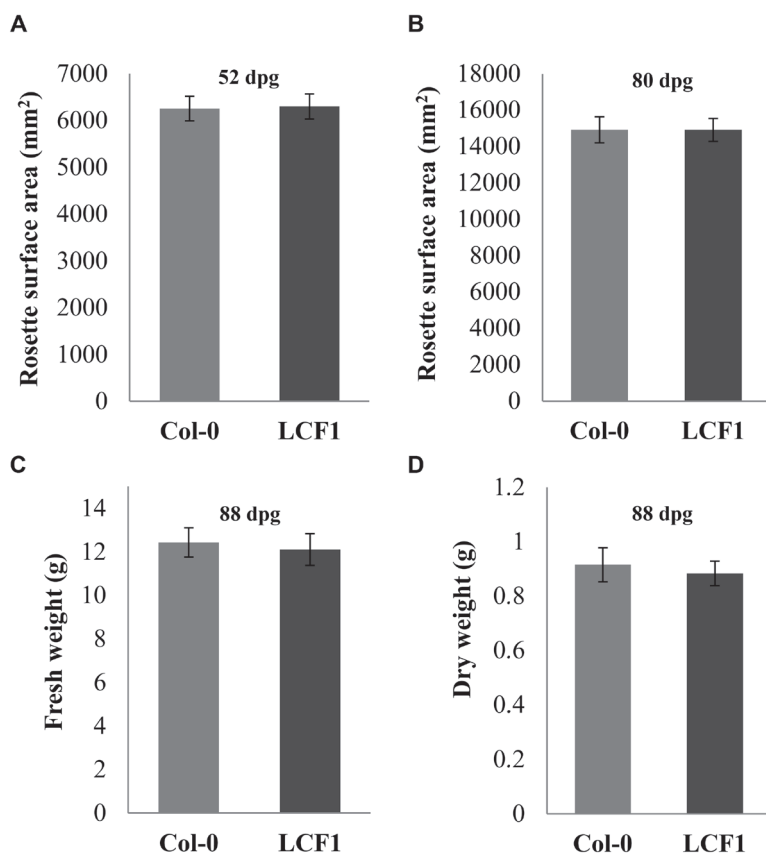


Fig. S4 Rosette surface area and biomass quantifications of wild type Col-0 and LCF1 plants grown at an 8 h photoperiod and a light intensity of 50 $\mu\text{mol m}^{-2} \text{s}^{-1}$. Plants were first grown at standard light conditions, and were transferred to the indicated conditions at 14 dpg. (A) Rosette surface area at 52 dpg. (B) Rosette surface area at 80 dpg. (C) Fresh weight of the above ground parts of Col-0 and LCF1 plants at 88 dpg. (D) Dry weight at 88 dpg, determined after drying the material for 3 days at 60 °C. Error bars represent SEM values (n=6 per genotype).

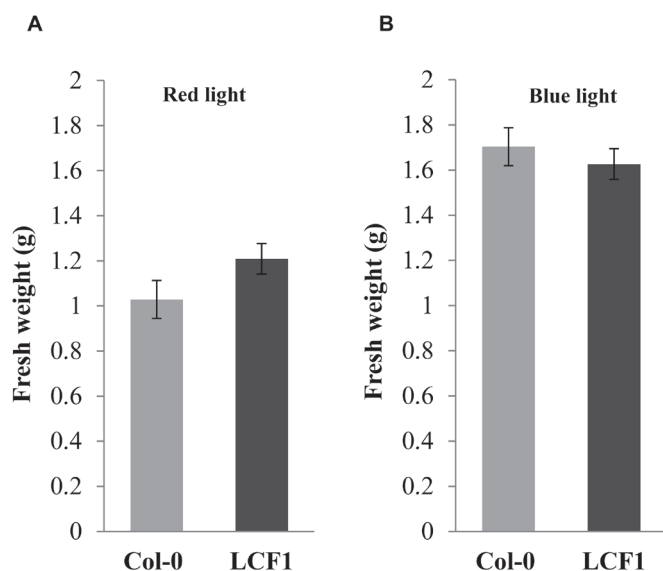


Fig. S5 Biomass quantification of wild type Col-0 plants and LCF1 plants grown at high intensities of red and blue LED light, respectively. Plants were initially grown at standard light conditions, and were then transferred to either $480 \mu\text{mol m}^{-2} \text{s}^{-1}$ red light, or $140 \mu\text{mol m}^{-2} \text{s}^{-1}$ blue light, both supplemented with $20 \mu\text{mol m}^{-2} \text{s}^{-1}$ daylight. (A) Fresh weight of above ground parts of plants grown for a further 11 days at high intensity red light (25 dpg in total). (B) Fresh weight of above ground parts of plants grown at high intensity blue light for a further 10 days (32 dpg in total). Error bars represent SEM values (n=10 per genotype).

Chapter 6

Artificial transcription factor-induced salinity tolerance in Arabidopsis

Niels van Tol, Johan Pinas, Paul J.J. Hooykaas and Bert J. van der Zaal

Abstract

Soil salinity is becoming an increasingly large problem in agriculture. Here, we describe the use of zinc finger artificial transcription factor (ZF-ATF)-mediated genome interrogation as a novel tool for the induction of salinity tolerance. We report the isolation of 41 *Arabidopsis* lines that are tolerant to 100 mM NaCl from a library of 4278 genome interrogation lines collectively harboring approximately 3800 different ZF-ATF encoding T-DNA constructs. The most strongly salinity tolerant lines were used to demonstrate that the ZF-ATF encoding gene constructs are the causative agents for salinity tolerance. For two lines expressing different ZF-ATFs we demonstrated that the constructs also induce tolerance to saline soil in terms of rosette surface area and biomass. Transcriptomic analysis of these two lines revealed that their salinity tolerance was accompanied by highly similar changes in gene expression, even though their ZF-ATFs harbored different DNA recognition sequences. Altogether, our data provide strong evidence that salinity tolerance can be evoked relatively easily and at high frequencies by ZF-ATF mediated genome interrogation.

Introduction

Salinity tolerance has become a valued trait in agriculture due to the gradually increasing salinity of agricultural soil [1, 2]. It has been estimated that more than 800 million hectares of agricultural land worldwide are affected by salts [3], with the most predominant being sodium chloride (NaCl). The engineering of salinity tolerance in crops has therefore received a considerable amount of attention [4-7]. Soil is considered saline when it has salt concentrations equivalent to or higher than 40 mM NaCl, which generates osmotic pressures of at least approximately 0.2 MPa [8]. The accumulation of salts in agricultural soils has mostly been attributed to irrigation, with approximately 20% of the irrigated land being affected [3]. This is especially worrisome because irrigated land can in principle be highly productive [2].

Plants experience osmotic stress to the root system immediately upon exposure to salts, which greatly impairs water uptake from the soil and results in a loss of turgor pressure. Despite that leaf cells are able to relatively quickly regain turgor pressure through osmotic adjustment by solute accumulation, there still is an overall reduction in cell expansion upon osmotic stress [9]. Osmotolerance of the root system can to some extent be mediated by the extrusion or efflux of ions from the roots [2]. On the long term salt ions can penetrate into the plant tissue, where the resulting ionic stress causes further problems such as genome instability [10] and oxidative stress that inhibits photosynthesis [2]. Plants respond to ionic stress by compartmentalization of ions in the vacuole, and by controlling the overall flux of ions through the apoplast [2]. In the non-growing tissues of mature leaves salt ions are not diluted by cell divisions and expansion, thereby resulting in cell death and a diminishing of the overall photosynthetic capacity on a leaf area basis [2].

There is a large extent of natural variation in salinity tolerance among plant species, with some species being relatively tolerant to NaCl, such as bread wheat (*Triticum aestivum*) and barley (*Hordeum vulgare*), and some being particularly sensitive, such as rice (*Oryza sativa*) and the model plant species *Arabidopsis thaliana* [2]. The growth of *Arabidopsis* is completely arrested in the presence of approximately 100 mM NaCl [2], and it has therefore been used extensively as a model system for salinity tolerance and stress responses [11, 12]. The physiological responses of *Arabidopsis* to salinity have been well documented [13], with the core of the salinity tolerance response pathway having been discovered through a genetic screen for salt overly sensitive (*sos*) mutants [14-16]. In this way, the proteins SOS1, SOS2 and SOS3 were identified that together form the SOS signaling pathway for salinity tolerance. *SOS1* encodes a plasma membrane Na^+/H^+ antiporter that has an important role in the extrusion of Na^+ ions [13]. Overexpression of *SOS1* has been demonstrated to induce salinity tolerance in *Arabidopsis* [17]. The activity of SOS1 is regulated through phosphorylation by the kinase SOS2, of which the activity is positively regulated by the Ca^{2+} binding protein SOS3 upon exposure to high Na^+ concentrations, thereby triggering Na^+ efflux [13]. Forward genetic

screens for *sos* suppressor mutants resulted in the identification of the Na⁺ transporter HKT1, which is another important component in the response to salinity stress [18]. HKT1 plays an important role in controlling the influx of Na⁺ into the roots and is putatively involved in the circulation of Na⁺ between the root system and the shoot [19], which is supported by the fact that natural variation in Na⁺ accumulation among natural Arabidopsis accessions has been mapped to *HKT1* [20]. The salinity stress signaling responses and their subsequent impact on growth are regulated by the phytohormones abscisic acid (ABA) [21] and gibberellic acid [22] and are modulated by transcription factors from the NAC [23, 24], DREB/CBF (Dehydration-Responsive Element Binding/C-repeat Binding Factor) [25] and MYB families [12].

The physiological adaptations of plants to adverse conditions such as salinity are part of the dynamic phenotypic space. Failure to adapt to adverse conditions and the resulting stress indicates that the phenotypic space cannot be stretched far enough to cope. In those cases, the normal epigenetic adaptations have been exhausted and novel (epi)genetic changes (e.g. spontaneous mutations) are required for adaptation. The correlation between gradual increases in soil salinity and the emergence of salinity tolerance among different Arabidopsis populations [20] might very well be considered as a natural example of stretching the phenotypic space towards salinity tolerance. A range of different studies have described artificial attempts to do so by the construction of various types of transgenic plant lines that are tolerant to salinity (e.g. [26-30]). These approaches mostly involve the use of only one or just a few (heterologous) genes, and are heavily based on *a priori* knowledge.

In this study, we have investigated whether the more or less random and large scale distortion of gene expression patterns can trigger plants to develop a higher level of salinity tolerance than normal for a given genotype. To this end, we have used a method coined ‘genome interrogation’ [31], which employs zinc finger artificial transcription factors (ZF-ATFs) to modify the phenotypic space of Arabidopsis [32, 33]. In the setup described in this study, ZF-ATFs consist of an array of three zinc fingers (3F) as a DNA binding domain, fused to the transcriptional activator VP16 as an effector domain [32]. Each individual ZF recognizes a cognate 3 base pair (bp) DNA sequence of 5'-GNN-3' [34], with N being any of the four DNA bases. There are 16 possible 5'-GNN-3' binding ZFs, which can theoretically be assembled into 256 two finger (2F) and 4096 three finger (3F) combinations. The artificial 3F-VP16 encoding genes are expressed under control of the promoter of the Arabidopsis *RPS5a* gene [35], which is predominantly active in embryonic and meristematic cells. Each 3F-VP16 fusion protein can potentially interact with on average approximately 1000 randomly distributed recognition sites in the 130 Mbp Arabidopsis genome. Even when only a fraction of these interactions would result in differential expression of one or more genes located *in cis* of a 3F recognition site, the expression of a multitude of genes is affected in a dominant manner, causing the perturbation of genome-wide gene expression patterns and the potential induction of traits of interest.

Here, we describe the isolation of 41 salinity tolerant *Arabidopsis* lines from a library of 3F-VP16 genome interrogation lines at the overall high frequency of approximately one out of every 80 3F-VP16 lines. The nine most tolerant lines were subjected to more extensive analysis and for four we were able to demonstrate that their 3F-VP16 encoding genes are the causative agents for salinity tolerance. Transcriptomic changes induced by two different 3F-VP16 encoding genes are discussed in relation to enhanced salinity tolerance.

Results

Construction of the 3F-VP16 genome interrogation *Arabidopsis* seed library and screening for salinity tolerant lines

For the construction of the *Arabidopsis* 3F-VP16 genome interrogation seed library, we used a library of pRF-VP16-Kana binary vector constructs encoding 3F-VP16 fusions described previously [32]. Briefly, this library was composed of 15 non-overlapping subpools, each of which was named after one of the sixteen 5'-GNN-3' binding ZFs that was first cloned to establish a theoretical maximum of 256 3F constructs. The subpool with the 5'-GAA-3' binding ZF as founder (subpool 3) was unstable for unknown reasons [32], and was therefore not constructed. Hence, the maximal complexity of the library of 3F-VP16 constructs was limited to a theoretical maximum of 3840 different gene constructs (15 subpools of at most 256 different fusions). Wild type Col-0 plants were transformed with each of the 15 subpools separately and primary transformants (T1 generation) were allowed to set seeds (T2 generation). In total, 4278 viable primary transformants were raised, more or less evenly distributed over the different subpools (Table 1). Approximately 5% of the primary transformants exhibited clearly visible phenotypic differences with Col-0 in terms of for instance leaf morphology, pigmentation and flowering time. Such conspicuous phenotypes often segregated in a 3:1 ratio in the T2 generation, demonstrating that they were indeed induced in a dominant manner as expected. The majority of the 3F-VP16 plants did not exhibit conspicuous phenotypes at standard growth conditions. To reduce the handling complexity of the library, seeds of five primary transformants originating from the same subpool were combined and stored in seed bags named 'five-bags'. Some five-bags contained less complex seed mixtures (originating from three or four primary transformants rather than five) due to occasional losses of plants during cultivation and infertility of a small fraction of the plants. Equal aliquots of seeds of ten five-bags from the same 3F-VP16 subpool were subsequently combined into 'fifty-bags' (Table 1).

Table 1. The composition of the 3F-VP16 Arabidopsis genome interrogation seed library.

Subpool number	Binding site of founding ZF (5'-3')	Number of primary transformants	Number of five-bags	Number of fifty-bags
1	GGG	298	69	6
2	GGA	234	67	5
3	GGT	210	66	4
4	GGC	280	67	6
5	GAG	343	80	7
7	GAT	323	90	7
8	GAC	279	78	6
9	GTG	259	60	5
10	GTA	396	90	8
11	GTT	238	56	5
12	GTC	240	56	5
13	GCG	234	50	5
14	GCA	302	62	6
15	GCT	378	84	6
16	GCC	264	59	5
Total		4278	1034	86

To investigate whether genome interrogation can be used to induce salinity tolerance in Arabidopsis, the 3F-VP16 library was screened for salinity tolerant lines by plating approximately 200 seeds of each fifty-bag on salinity stress induction medium containing 100 mM NaCl, a concentration that was found to be a sufficiently high to arrest the growth of wild type plantlets in the cotyledon stage without having a marked negative effect on the germination percentage (Fig. S1). There was a clear reduction in overall germination along with increasing NaCl concentration and, as expected, germination was completely abolished in the presence of 250 or 500 mM NaCl (Fig. S1). In total, 46 salinity tolerant individuals were isolated (Table 2), that were all able to develop green and in most cases fully expanded leaves in the presence of 100 mM NaCl. These individuals were denoted S1 through S46 and were isolated from all library subpools except for subpools 9 and 14 (Table 2), suggesting that there is great potential for the induction of salinity tolerance for a range of different 3F-VP16 fusions. The 46 salinity tolerant individuals were transferred to soil, further cultivated and allowed to set seeds (T3 generation).

Table 2. The total output of screening the 3F-VP16 library for salinity tolerant lines.

Library pool number	Number of salinity tolerant lines found (T2 generation)	Number for which salinity tolerance could be confirmed (T3 generation)
1	2	2
2	1	1
3	2	2
4	7	7
5	4	4
7	6	5
8	4	3
10	7	7
11	1	1
12	1	1
13	1	0
15	9	7
16	1	1
Total	46	41

Salinity tolerance confirmation and PCR analysis

To confirm the salinity tolerance of the isolated individuals, their T3 seed progeny was plated on salinity stress induction medium. In this way, salinity tolerance could be confirmed for 41 out of the 46 lines (Table 2), indicating that the salinity tolerance phenotype is stable over the course of at least two generations. As a result of the use of 50-bags we estimate that the seeds of approximately 80% of the 4278 3F-VP16 lines were represented in the salinity tolerance screen. Based on this estimate, the 41 salinity tolerant lines were isolated from in total approximately 3400 lines, and therefore at the overall frequency of approximately one out of every 80. Assuming that all 41 tolerant lines harbor different 3F-VP16 fusions, this means that approximately one out of every 75 3F-VP16 fusions induces salinity tolerance in Arabidopsis.

From the 41 tolerant lines we selected nine that displayed the strongest tolerance to 100 mM NaCl in terms of vigor for further analysis (Fig. 1). The 3F encoding T-DNA fragments were isolated from the genomic DNA of these lines by PCR, and were sequenced. All of the selected lines harbored full length 3F encoding sequences except for S27, which for unknown reasons harbored a 2F encoding sequence (Table 3). Lines S6 and S7 originated from the same subpool and turned out to also contain the same 3F sequence (Table 3). In total, 3Fs originating from five different subpools were recovered (Table 3) and no clear consensus or overlap was found among their DNA recognition sequences (Table 3), corroborating the idea that salinity tolerance can be induced through different 3F recognition sequences.

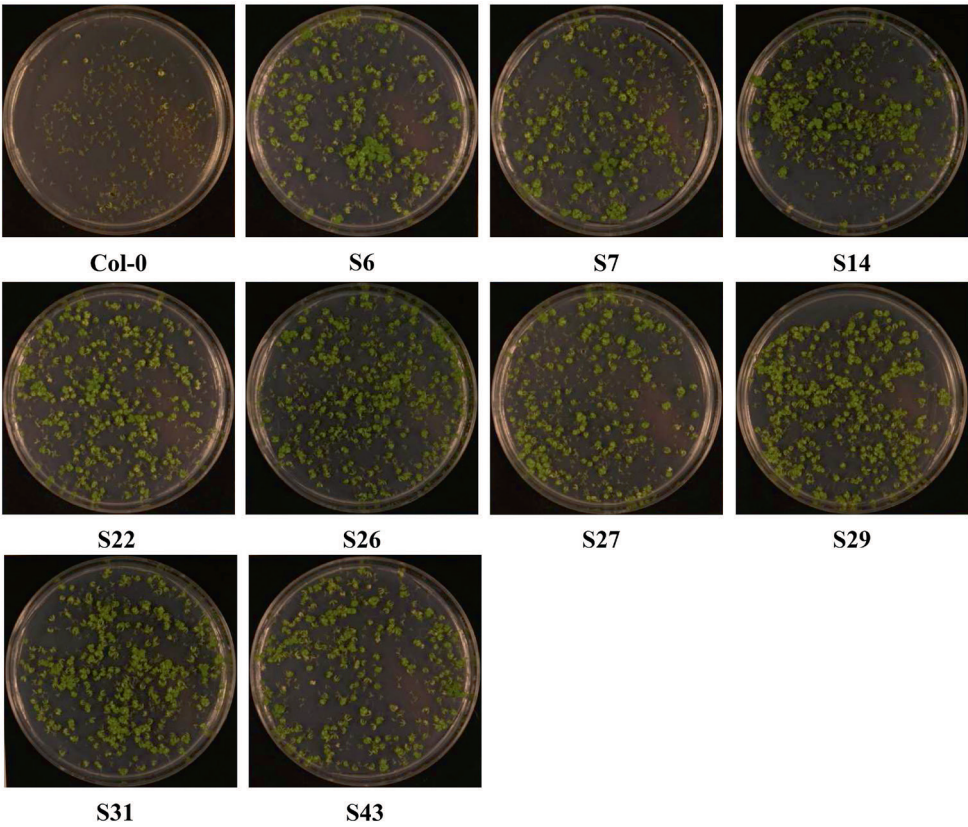


Fig. 1 Seedlings of the salinity tolerant lines that were selected for further analysis (T3 generation) growing on salinity stress induction medium containing 100 mM NaCl (21 dpv). The wild type Col-0 serves as a negative control.

Table 3. Salinity tolerant lines selected for further analysis.

Line name	Isolated from subpool	DNA recognition site of isolated 3F (5'-3')
S6	4	GAG-GGA-GGC
S7	4	GAG-GGA-GGC
S14	5	GAC-GTT-GAG
S22	7	GGT-GAG-GTA
S26	8	GAT-GCG-GAC
S27 [§]	10	GGT-GTA
S29	10	GAT-GCC-GTA
S31	10	GTA-GTA-GAG
S43	15	GTA-GGC-GCT

[§] Contained a two zinc finger (2F) encoding sequence

The 3F-VP16 constructs from four different lines are the causative agents of salinity tolerance

To determine whether in the nine selected lines the 3F-VP16 encoding T-DNA constructs are causal for salinity tolerance, wild type plants were transformed with 3F-VP16 constructs that were reconstituted with the 3F sequences that were recovered from the lines (Table 3). The presence of the correct constructs in the primary retransformants (T1 generation) was verified by PCR analysis and sequencing of the PCR product. The seed progeny of four to five independent T2 retransformant lines (segregating for the T-DNA constructs) was plated on salinity stress induction medium containing an additional 10 mM of NaCl (110 mM NaCl in total) to eliminate the background tolerance that we occasionally observed among Col-0 plants. The majority of the retransformant lines that were reconstituted from S27 and S29, and all of the retransformant lines that were reconstituted from S31 and S43 displayed salinity tolerance (Fig. 2), demonstrating that we have isolated four different 3F-VP16 encoding genes that can induce salinity tolerance in *Arabidopsis* in a dominant manner and seemingly independent of the copy number and/or the insertion loci of the 3F-VP16 encoding T-DNA construct. In the cases of lines S6, S7, S14, S22 and S26 only a minority of the retransformant lines displayed salinity tolerance (Table 3), indicating that the 3F-VP16 encoding genes from these lines are not as potent at inducing salinity tolerance as the constructs from S27, S29, S31, and S43. Overall, 24 out of the 43 retransformant lines that were investigated displayed some degree of salinity tolerance (Table S1).

Retransformant lines reconstituted from S31 and S43 accumulate more biomass at salinity stress conditions

Although all retransformant lines reconstituted from S31 and S43 displayed strong tolerance to salinity stress induction medium (Fig. 2), such *in vitro* tolerance does not necessarily translate into tolerance to the agriculturally relevant soil salinity. Therefore, Col-0 plants and T2 retransformant plants harboring T-DNA constructs that were reconstituted from S31 and S43 (three and four independent lines, respectively; segregating for the T-DNA constructs) were grown on soil that was watered with a solution of 100 mM NaCl from 17 dpv onwards, and rosette surface area (RSA) of the plants was quantified both before and during this salt treatment. RSA was used as a non-destructive proxy for growth and biomass during development, because these parameters are strongly correlated in *Arabidopsis* [36]. During this experiment, all of the retransformant lines displayed a consistent increase in the mean RSA compared to Col-0, both before and during the salt treatment (Fig. 3A). Five out of the seven retransformant lines had also accumulated significantly more dry weight (Fig. 3B) at 28 dpv, with the average increase in biomass varying from approximately 20 to 80 %. To ascertain that these increases in RSA were not only attributable to enhanced growth in the absence of salt, we again quantified the RSA of Col-0 plants and T2 retransformant plants of

lines T31-5 and T43-11 throughout development, but now also without salt treatment. There turned out to be a significant increase in RSA of the retransformant plants compared to Col-0 both with and without salt treatment at 28 dp_g, but the effect was more substantial with salt treatment (Fig. 4), thus showing that the 3F-VP16 constructs reconstituted from S31 and S43 induce an increase in RSA at both standard and salinity stress conditions. During our experiments on soil, the maximum quantum efficiency of PSII photochemistry (F_v/F_m) [37] was highly similar for all lines both at standard and salt stress conditions. Hence, the observed differences in RSA could not directly be attributed to changes in photosynthetic efficiency as a result of salt stress.

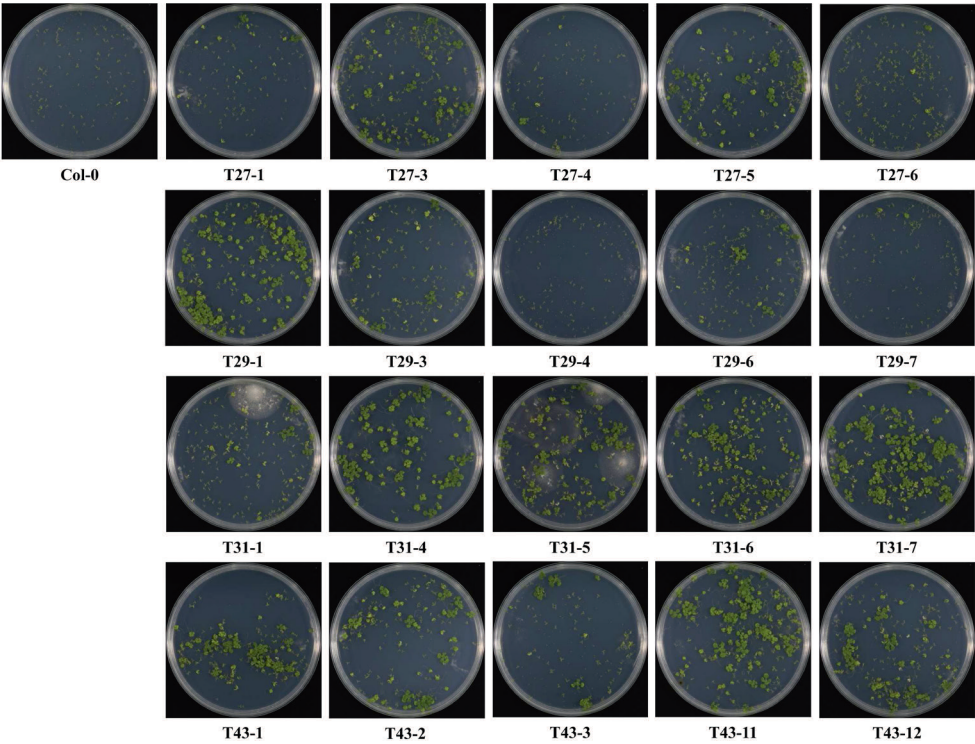


Fig. 2 Seedlings of retransformant lines harboring 3F-VP16 encoding T-DNA constructs that were reconstituted from the original salinity tolerant lines S27, S29, S31 and S43 (T2 generation; segregating for the T-DNA constructs) growing on salinity stress induction medium containing 110 mM NaCl (21 dp_g). The wild type Col-0 serves as a negative control.

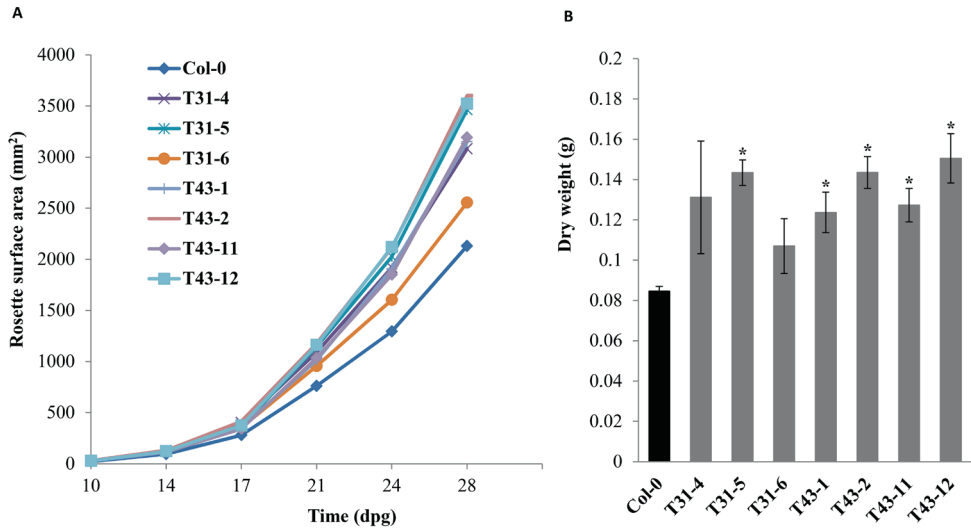


Fig. 3 Overview of the growth characteristics during salinity stress of wild type Col-0 plants and T2 retransformant plants harboring T-DNA constructs that were reconstituted from salinity tolerant 3F-VP16 lines S31 and S43 (segregating; three and four independent lines, respectively). **A)** Growth curves of plants (n=6 per genotype) that were watered every 3-4 days with demi water, and from 17 dpg onwards were watered every 3-4 days with a solution of 100 mM NaCl in demi water. RSA was quantified at the indicated time points. Errors bars were not included for the sake of clarity of the figure. **B)** Dry weight after 12 days of salt treatment (28 dpg in total) and subsequent incubation of the shoots at 60 °C for two days. Error bars represent SEM values (n=6 per genotype). Asterisks (*) indicate significant differences with Col-0 ($p < 0.05$).

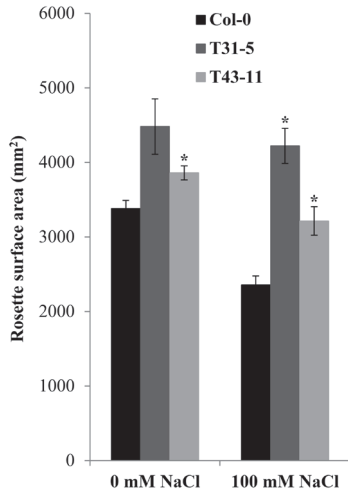


Fig. 4 Quantification of RSA of soil grown wild type Col-0 plants and plants from retransformant lines T31-5 and T43-11 (T2 generation; segregating for the T-DNA constructs). From 17 dpg onwards, the plants were watered every 3-4 days with either demi water (0 mM NaCl) or a solution of 100 mM NaCl for another 12 days (28 dpg in total). Error bars represent SEM values (n=6 per genotype). Asterisks (*) indicate significant differences with Col-0 within treatments ($p < 0.05$).

Transcriptome analyses

To gather some insight into the transcriptomic basis of 3F-VP16 induced salinity tolerance, RNA sequencing was performed on total RNA extracted from Col-0 plants, T2 plants harboring empty vector control T-DNA constructs (a mixture of five independent lines expressing VP16 without 3Fs; segregating for the T-DNA construct) and T2 retransformant plants of lines T31-5 and T43-11 (segregating for the T-DNA constructs). We adopted an experimental set up without biological replicates which did not allow for the precise quantification of gene expression, but should give qualitative clues about the transcriptomic changes in the 3F-VP16 lines compared to the controls. The plants were all grown on half strength MS medium containing either 0 mM NaCl (referred to as 'control') or 75 mM NaCl (referred to as 'salt treated'). The NaCl concentration of 75 mM was chosen because it allowed wild type plants to grow whilst experiencing salt stress, whereas 100 mM NaCl would have been lethal. The empty vector control lines were included to investigate the changes brought about by the expression of free VP16 as compared to a complete 3F-VP16 transcription factor.

All plants grew normally on control medium. As expected, salt treated Col-0 and empty vector control plants displayed stunted growth, whereas salt treated T31-5 and T43-11 plants grew much better in comparison. On average, the RNA samples yielded about 20 million mapped sequencing reads (Table S2). The expression levels of the (3F-)VP16 encoding T-DNA constructs were found to be rather similar to the reads per kilobase per million mapped reads (RKPM) values for the endogenous *RPS5a* gene (At3g11940) from which the promoter driving (3F-)VP16 expression was derived, indicating that the constructs were well expressed *in planta*. To avoid overestimation of differentially expressed genes (DEGs) due to random fluctuations in low levels of gene expression, we skipped the analysis of genes that exhibited an RKPM value below 5 as the mean for all transcriptomes. By doing so, we had to accept the fact that only the data for about half of the annotated Arabidopsis genes were further analyzed (13374 out of the 27416 annotated genes in the TAIR10 release of the Arabidopsis genome), and that we had to refrain from considering the behavior of the majority of the usually lowly expressed transcriptional regulators. Genes were marked as DEGs between treatments and/or genotypes when exhibiting changes in RKPM values that were larger than 2-fold up or down.

Salt treatment led to considerable transcriptomic differences with the controls for all of the transgenic lines with in each case approximately half of the DEGs being upregulated and half being downregulated. In Col-0, a total of 907 genes out of the 13374 present in our list (6.8%) were DEGs after salt treatment. For the analysis described below, these genes were denoted as 'salt regulated', with 459 being 'salt upregulated' (Table S3A) and 448 being 'salt downregulated' (Table S3B), thus at the ratio of almost exactly 1:1. As expected, a considerable fraction of the salt regulated DEGs were annotated as salt or osmotic stress responsive in the TAIR 10 release of the annotated Arabidopsis genome. Despite the fact that we deliberately did not analyze lowly expressed genes, about 45% of the genes reported as differentially

regulated during salt acclimation (48 hours on 50 mM NaCl containing medium, [38]) were also present in our list of salt regulated genes. Salt treatment reduced the expression levels of most of the highly expressed genes (mean expression values higher than 1500 RKPM) known to be involved in photosynthesis by approximately 30% in all genotypes.

3F-VP16 specific transcriptomic changes upon salt treatment

In order to relate 3F-VP16 induced transcriptomic changes to salinity tolerance, we first determined the number of DEGs between transcriptomes of the salt treated transgenic lines and salt treated Col-0, and subsequently analyzed their overlap (Fig. 5A). This procedure clearly demonstrated that the large majority of the resulting DEGs ($\geq 90\%$) were shared between the empty vector control and salinity tolerant 3F-VP16 lines. For instance, the salt treated empty vector control and T43-11 had 188 and 267 upregulated DEGs compared to salt treated Col-0, respectively, of which 167 were shared (Fig. 5A). To delineate which of the DEGs in the 3F-VP16 mutant lines were 3F specific, we again performed a sequential comparison. Firstly, we identified the DEGs of the salt treated T31-5 and T43-11 transcriptomes compared to the salt treated empty vector control, and subsequently looked for the overlap between these two. In this way, 12 and 9 annotated genes were found to be 3F-VP16 specifically upregulated and downregulated, respectively (Fig. 5B, and Table S5 for further details). Among the gene products of these 21 3F-VP16 specific DEGs, there were three (putative) peroxidases (At1g05250; At1g25820; At4g26010), two oxygenases (At3g12900; At4g31940) and a glutathione S-transferase (At1g49860) (Table 4), indicating that the 3F-VP16 induced salinity tolerance mechanisms have an oxidative stress related component. Furthermore, there were two extensin(-like) proteins (At3g54590; At5g35190), two glucan modifying enzymes (At4g25820; At3g57260) and an arabinogalactan protein (At4g40090), suggesting that the 3F-VP16 induced salinity tolerance mechanisms also involved modulation of the cell wall. Five of the 21 DEGs encode gene products of presently unknown function (Table 4). For two genes, At3g57260 and At3g01345, we found such seemingly random variations in the expression pattern that their presence in the list might better be ignored. Although not necessarily causal for acquiring salinity tolerance, several of these 21 DEGs might be useful transcriptomic markers for salinity tolerance.

Table 4. Overview of differentially expressed genes (DEGs; average normalized expression values over different treatments > 5 RPKM) compared to the empty vector control (pRF-VP16-Kan^r[-]) that are shared between salinity tolerant 3F-VP16 lines T31-5 and T43-11, all treated with 75 mM. Upregulated DEGs have > 2 fold normalized expression values, downregulated DEGs have < 0.5 fold normalized expression values.

Upregulated (12)		Downregulated (9)	
Gene ID	Gene name; encodes; description	Gene ID	Gene name; encodes; description
^a At1g05240	Putative peroxidase	"At1g13609	Defensin-like (DEFL) family protein.
^a At1g05250	PRX2 ; Peroxidase 2	"At1g47395	Unknown protein
At1g49860	GSTF14 ; putative Glutathione S-transferase	"At1g47400	Unknown protein
At3g20380	Meprin and TRAF homology (MATH) domain-containing protein	"At2g30766	Unknown protein
At3g43850	Unknown protein	At3g01345 [*]	Unknown protein; putative hydrolase
At3g54590	HRGPI ; proline-rich extensin-like family protein	"At3g12900	Oxidoreductase; 2OG-Fe(II) oxygenase
"At3g57260 [*]	BGL2 ; Beta-1,3-Glucanase 2	At3g09922	IPSI
^a At4g02270	RHS13 ; Pollen Ole e 1 allergen and extensin family protein; contains Pfam domain	At4g14690	ELIP2 ; chlorophyll A-B binding family protein
^a At4g25820	XTH14 ; XTR9 ; xyloglucan endotransglycosylase	At4g31940	CYP82C4 ; Putative cytochrome P450 monooxygenase
^a At4g26010	Putative peroxidase		
At4g40090	AGP3 ; Arabinogalactan Protein 3		
^a At5g35190	EXT13 ; Extensin 13; Proline-rich extensin-like family protein		

^{*}Meet the indicated DEG selection criteria, but expression values are highly variable. Might better be disregarded.

^aSalt downregulated in Col-0.

^u Salt upregulated in Col-0.

Grey highlights indicate genes that are downregulated by salt treatment in wild type plants, but upregulated in the 3F-VP16 line, or were very close to meeting these criteria. At3g12900 was very close to meeting the opposite of these criteria.

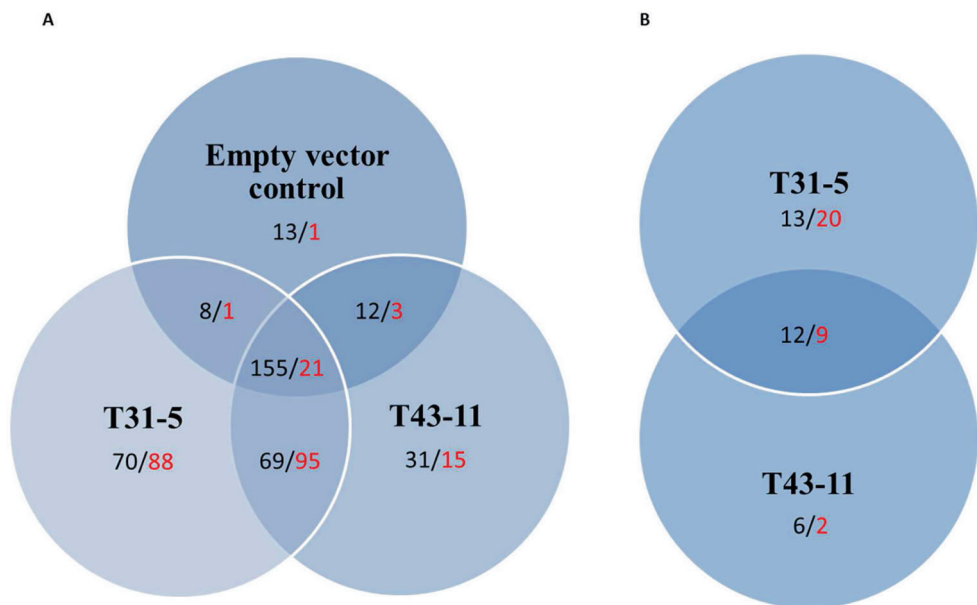


Fig. 5 Overview of the numbers of differentially expressed genes (DEGs; average normalized expression values over different treatments > 5 RKPM) in salt treated samples (75 mM NaCl) of the indicated transgenic lines compared to **A**) the salt treated wild type Col-0 and **B**) the salt treated empty vector control. Upregulated DEGs (> 2 fold higher normalized expression values) are presented in black font; downregulated DEGs (< 0.5 fold normalized expression values) are presented in red font.

Transcriptomic changes in relation to the salt stress response of wild type plants

When looking at (3F-)VP16 induced DEGs in control plants, we noticed that more than 50% of these could be found in the list of 907 genes that were salt regulated in wild type plants, as defined above. Expression of (3F-)VP16 proteins thus already affected potentially salt regulated genes even prior to any salt treatment. To further investigate the mechanism underlying 3F-VP16 induced salinity tolerance in relation to a wild type salt response, we therefore systematically determined whether selected groups of (3F-)VP16 specific DEGs were salt regulated and, if so, whether they were up- or downregulated. To this end, we performed Chi-square tests with an expectancy of 6.8% for salt regulation and an expectancy of 50% for then being up- or downregulated by salt. Out of the total of 581 DEGs found by comparing salt treated (3F-)VP16 lines to salt treated Col-0 (Fig. 5A), 229 genes were in this manner defined as salt regulated (39.4%), which is a highly significant enrichment ($p < 0.001$). Out of these 229 DEGs, 111 were salt upregulated and 118 were salt downregulated, which almost perfectly corresponds to the 50% expectancy for a gene to be salt up- or downregulated. When we specified for DEGs that were upregulated in the (3F-)VP16 lines in comparison to Col-0, 111 out of the 357 DEGs (Fig. 5A) were salt regulated (31%), but now with a skewed salt upregulated to salt downregulated ratio of 22:89. Similarly, out of

the 224 genes that were downregulated in the (3F-)VP16 lines compared to Col-0 (Fig. 5A), 118 were salt regulated (52.6%) with a salt upregulated to salt downregulated ratio of 103:15. Subtracting the 14 DEGs that were specific to the empty vector control (Fig. 5A) did not make a difference in this respect. A nearly identical pattern was found for the DEGs shared by the salt treated transcriptomes of T31-5 and T43-11 compared to salt treated Col-0 (Fig. 5A). Out of the total of 340 DEGs, 141 were salt regulated (41%), with 71 up- and 70 downregulated by salt. When specifying for up- and downregulated DEGs, 72 of the 224 upregulated DEGs were salt regulated (32.1%; 10 salt upregulated, 62 salt downregulated) and 69 of the 116 downregulated DEGs (59.5%; 61 salt upregulated, 8 salt downregulated). The enrichments for salt regulated genes and also the skewed distributions were highly significant ($p < 0.001$). The patterns observed after analyzing the DEGs shared by T31-5 and T43-11 but not by the empty vector control (Fig. 5A) were highly similar (Table S4), and yielded a (nearly) significant enrichment for genes that we denominated as being salt regulated and always a significant negative correlation between the direction of gene regulation with respect to untreated control and the direction of regulation by salt. These observations further corroborate the idea that (3F-)VP16 fusions induce salinity tolerance through the differential regulation of salt responsive genes.

Salt responsiveness of 3F-VP16 specific DEGs

To investigate the salt responsiveness of the 3F specific DEGs, we examined whether the DEG categories depicted in Figure 5B were salt regulated. Out of these 62 genes (Table S5), 29 were salt regulated (46.8%). For all 41 genes unique to either T31-5 or T43-11, 16 were salt regulated (significant 39% enrichment). Among the DEGs that are shared between T31-5 and T43-11, 13 out of 21 are salt regulated (significant 61.9% enrichment). Remarkably, bearing in mind that there is a highly significant enrichment for salt regulated DEGs in the salt treated (3F-)VP16 transcriptomes (Table S5), we have found no evidence for a significant enrichment for salt regulated genes among the in total 19 genes that are classified as specifically upregulated by either T31-5 or T43-11 (Table S5, columns 1 and 3). Among both the upregulated and downregulated DEGs compared to the salt treated empty vector control that are shared between T31-5 and T43-11, a large fraction is salt regulated (Table S4). Interestingly, when inspecting the gene expression patterns more closely, we found that several of the shared and 3F-VP16 specific salt regulated genes were oppositely regulated in the salt treated Col-0 samples (highlighted in Table S5). These observations demonstrate that there is also a significant enrichment of salt regulated genes among the DEGs that are 3F specific, and therefore also mediate salinity tolerance through the differential regulation of salt responsive genes.

Discussion

In this study we described the construction of a library of Arabidopsis lines expressing 3F-VP16 fusions, the screening of that library for salinity tolerant lines, and their further characterization. Four different 3F-VP16 encoding T-DNA constructs were shown to be dominant *in trans* causative agents for salinity tolerance. A tentative analysis of the effect of two of these constructs in relation to salinity tolerance indicated that they prime Arabidopsis plants for salinity tolerance through generic VP16-induced changes that form a background for the emergence of salinity tolerance through 3F specific changes.

In order to evaluate the method of genome interrogation as a forward genetics tool in relation to salinity tolerance, some comparison to other mutagenic treatments needs to be made. Our library consisted of 4278 plant lines, representing approximately 3800 different 3F-VP16 encoding T-DNA constructs, and our screening procedure yielded 41 stably salinity tolerant lines, corresponding to a salinity tolerance frequency of approximately 1 out of every 80 3F-VP16 mutant lines. This frequency appears to be exceptionally high when compared to published data. For instance, screening Arabidopsis populations for germination in the presence of 250 mM NaCl has yielded EMS, fast neutron and T-DNA insertion mutant lines at much lower frequencies (5 out of 12000, 34 out of 45452 and 12 out of 6480, respectively) [39]. The salt hypersensitive mutant line *sos1* was isolated along with three other stable hypersensitive mutants from a pool of approximately 50,000 EMS mutagenized seeds [14], which is also a much lower frequency. Of course these studies cannot directly be compared to each other, as differences in the salt concentration that is used as selection criterion might have resulted in different screening stringencies. It should also be noted that traits induced by genome interrogation are in principle inherited in a dominant manner, while the more classical methods predominantly yield recessive mutations. Therefore, the frequency of occurrence of a given trait in the progeny of genome interrogation lines will anyhow be 3-fold higher than the progeny of mutants generated with more classical forward genetics techniques. In addition to triggering phenotypes of interest, genome interrogation can therefore also be considered a powerful forward genetics tool. When compared to T-DNA tagging, which is an *in cis* type of mutagenesis used for the isolation of dominant types of ectopic gene expression and to some extent resembles genome interrogation mechanistically [31], genome interrogation might theoretically reach a similar level of saturation for ectopic expression of a particular gene with up to 1000-fold less plants having to be screened. Therefore, our library of just a few thousand 3F-VP16 lines has in principle allowed us to address the role of every genetic locus in salinity tolerance [31]. In addition, genome interrogation will theoretically trigger differential expression of a multitude of genes *in trans*, whereas T-DNA tagging will in principle at best address only a few loci *in cis*. Altogether, our screening for salinity tolerance using genome interrogation has illustrated that it might be sufficient to screen relatively few

plant lines to achieve a saturated screening for a particular phenotype. It also proved to be possible to isolate four 3F-VP16 encoding genes that were able to induce salinity tolerance *in trans* after retransformation, demonstrating that genome interrogation provides the additional benefit of placing a trait under experimental control. Salinity tolerance has already been placed under control of (heterologous) artificial transcription factors (e.g. [40]), but these approaches are mostly based on *a priori* knowledge of known regulators of the salinity stress responses of plants, whereas genome interrogation has allowed us to induce salinity tolerance by the introduction of a single artificial gene and without *a priori* knowledge of the salinity stress response.

It became evident from our transcriptomics data that the expression of 3F-VP16 fusions but also of free VP16 led to considerable changes in gene expression patterns, particularly in the case of salt treatment. The expression of (3F-)VP16 encoding genes thus preferentially triggered differential regulation of a considerable fraction of salt regulated genes, which might very likely be indicative of a more generic stress response. Although originally derived from the herpes simplex virus [41], the VP16 domain is known to interact with a variety of transcriptional regulators involved in the activation of gene expression in large variety of eukaryotic organisms, most likely via the recruitment of proteins that can open up the local chromatin structure [42, 43]. It could thus be envisaged that an abundance of VP16 in the nuclei of meristematic plant cells easily results in the distortion of normal gene expression patterns, regardless of the presence of a DNA binding domain. The fact that VP16 responsive DEGs are enriched for salt regulated genes might therefore indicate that these genes are readily affected by transcriptional modulation, and are easily controlled by direct or indirect regulation via endogenous transcriptional regulators with VP16-like activity. Furthermore, there was an enrichment for salt regulated genes in the (3F-)VP16 lines both at control conditions and upon salt treatment, but in the case of salt treatment there was a much larger number of genes involved, suggesting that (3F-)VP16 fusions specifically modulate the salt response. Together with the skewed overall salt up- to salt down ratios this seems to indicate that (3F-)VP16 mediated up- or downregulation of a given gene is negatively correlated with salt regulation. Mechanistically, this might be explained by the fact that (3F-)VP16 fusions sequester the proteins that are needed for a salt response of particular genes, a phenomenon which has been designated as ‘squenching’[44], and that salt regulated genes are less likely to be affected by this.

In order to isolate the salt responsive 3F specific transcriptional changes from the VP16 specific ones, it proved to be very helpful to discriminate against the transcriptome of salt treated empty vector control (Fig. 6B). We found relatively few 3F-VP16 specific DEGs in this way, but they were still highly enriched for salt regulated genes. This further illustrates the specific modulation of the salt responsive transcriptome by 3F-VP16 fusions. Remarkably, while the total number of genes in the different categories of DEGs exhibited a highly

significant enrichment for salt regulated genes, no such enrichment was found among the DEGs truly specific to either T31-5 or T43-11, which are thus apparently not involved in differential regulation of the salt response. Interestingly, for the genes that were regulated by both 3F-VP16 fusions there often was an opposite direction of salt regulation than in Col-0, further corroborating the idea that the salinity tolerance of T31-5 and T43-11 is brought about by a highly similar adaptation or modulation of the wild type response to salinity. Although it might be tempting to state that only the shared 3F-VP16 DEGs are truly contributing to the salinity tolerance phenotype, it seems much more likely that the broader VP16 mediated modulation of the salt responsive transcriptome already provides a background that somehow supports the development of salinity tolerance through these shared 3F-specific DEGs. This might explain the high frequency of salinity tolerant lines in our genome interrogation library. It might be surprising to note that so many of the 3F specific DEGs are shared between T31-5 and T43-11, because they expectedly recognize very different DNA sequences. It should be realized however that there might very well be binding sites for different 3Fs within the regulatory regions of key regulatory genes of the salinity stress response through which highly similar cascades of events could be triggered. Based on our transcriptome analysis the roles of regulatory genes in the 3F-VP16 mediated salinity stress remains elusive because they are typically expressed at low levels. As already indicated above, it might however be expected that 3F-VP16 fusions orchestrate a novel salinity tolerance mechanism through the differential regulation of a rather complex set of genes, part of which is already involved in the wild type salinity stress response. This is in accordance with the principle of genome interrogation, where an artificial transcription factor is thought to function as an exogenous master switch that evokes a particular phenotype through the induction of drastic differential gene expression [31]. In a previous study concerning another Arabidopsis trait of interest, we have also found that the phenotype was only fully induced by a combination of all differentially expressed genes, and could not or only mildly be triggered by single DEGs [32, 33].

In conclusion, we have successfully used genome interrogation to generate and isolate salinity tolerant Arabidopsis lines at very high frequencies, and it has thus enabled us to place salinity tolerance under the control of artificial genes. In principle, genome interrogation could be applied to any plant species of choice regardless of *a priori* knowledge and the availability and/or size of the genome sequence [31]. In this study, we have used 100 and 110 mM NaCl as selection criteria, but the 3F-VP16 library could in principle also be screened for any salinity responsive trait that one would be interested in, and lines with phenotypes or resistances of interest might also be found at high frequencies. Our transcriptomics data suggest 3F-VP16 mediated salinity tolerance is triggered by the transcriptional modulation of the normal salinity stress response, and based on predictions of gene function it involves novel oxidative stress and cell wall modification related genes in Arabidopsis. Our findings through genome interrogation also indicate that the development of salinity tolerance might

be relatively easy in natural populations once conditions arise that cause destabilization of normal gene expression patterns. In the future, we hope to be able to exploit this knowledge further, and to eventually use it for the engineering of salinity tolerant crop plants.

Material and methods

Growth conditions

All *in vitro* salinity tolerance experiments were performed in a climate controlled tissue culture chamber at 20 °C, 50% relative humidity, approximately 50 $\mu\text{mol m}^{-2} \text{s}^{-1}$ of photosynthetically active radiation (PAR), and a 16 h photoperiod (referred to as tissue culture conditions). For salinity stress experiments on soil, plants were grown in a climate controlled growth chamber at 20 °C, 70% relative humidity, a light intensity of approximately 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$ PAR, and a 12 h photoperiod (referred to as growth chamber conditions). Primary floral dip transformants were first grown on selection medium at tissue culture conditions, subsequently transferred to soil after approximately 3 weeks and were further cultivated at growth chamber conditions.

Plant material and construction of the Arabidopsis 3F-VP16 genome interrogation seed library

The Arabidopsis accession Columbia-0 (Col-0) was used as the wild type and background genotype for all transformations. For construction of the 3F-VP16 genome interrogation seed library, a library consisting of 15 subpools of T-DNA constructs encoding fusions of 3Fs to the transcriptional activator VP16 was generated in the binary vector construct pRF-VP16-Kana in *E.coli*, and mobilized to the *Agrobacterium tumefaciens* strain Agl1 through triparental mating [32]. Col-0 plants were transformed with each subpool of 3F-VP16 constructs separately using the floral dip method [45], and primary transformants were selected by plating the resulting seed pools on MA medium [46] containing 35 $\mu\text{g/mL}$ kanamycin, 100 $\mu\text{g/mL}$ nystatin and 100 $\mu\text{g/mL}$ Timentin. Kanamycin resistant plants were transferred from selection medium to soil after 2-3 weeks (five plants originating from the same subpool per pot), further cultivated and allowed to set seeds (T2 generation). The seeds of all plants in a pot were harvested together and stored in seed bags that were denoted 'five-bags'. Due to losses of plants during cultivation, some five-bags contained seeds of three or four primary transformants rather than five. Equal aliquots of seeds from ten five-bags from the same subpool were combined into 'fifty-bags'.

Library screening for salinity tolerant 3F-VP16 lines

Approximately 200 seeds of each fifty-bag were sterilized by washing with 70% ethanol for 5 min and subsequent incubation for 15 min in a solution of 1% active chlorine containing

aliquots of Tween-20. The seeds were then washed 4 times with sterile demineralized water and suspended in 10 mL of 0.1% (w/v) agarose. The seed suspensions were stratified for 3 days at 4 °C, and were subsequently plated on salinity stress induction medium (half strength MS medium [47] without sucrose, containing 100 mM NaCl, 100 µg/mL nystatin and 100 µg/mL Timentin). Plates were incubated for 2-3 weeks at tissue culture conditions, and tolerant individuals were subsequently transferred to soil, further cultivated and allowed to set seeds (T3 generation). For the confirmation of salinity tolerance, these seeds were collected and approximately 200 were sterilized, stratified and plated on salinity stress induction medium as described above. Plates were incubated for 2-3 weeks at tissue culture conditions and then evaluated for salinity tolerance.

Retransformation with reconstituted T-DNA constructs

From the selected salinity tolerant lines the 3F-VP16 encoding fragments were isolated by PCR using the T-DNA specific primer combination pRF Uni FW (5'-GAAGCGTAAGGTCGAGC-3') and 2pol REV (5'-CTCGATGCATTCGCGAG-3'). Reconstituted T-DNA constructs were generated in pRF-VP16-Kana using the 3F encoding sequences from these PCR products [32], and were transformed to Col-0 plants using the floral dip method [45]. Primary retransformants were selected by plating sterilized and stratified seeds (as described above) on MA medium [46] containing 35 µg/mL kanamycin, 100 µg/mL nystatin, and 100 µg/mL Timentin. The primary retransformants were transferred from selection medium to soil after 2-3 weeks, further cultivated and allowed to set seeds, which were harvested (T2 generation).

Rosette surface area quantification and chlorophyll fluorescence imaging during growth on saline soil

Seeds of Col-0 and the retransformant lines harboring T-DNA constructs that were reconstituted from the original salinity tolerant lines S31 and S43 were sown on soil in separate pots (n=6 per genotype) and stratified for 3 days at 4 °C. The plants were watered every 3-4 days with demineralized water. From 17 dpg onwards, the plants were watered every 3-4 days either with demineralized water, or with a solution of 100 mM NaCl in demineralized water. From 10 dpg onwards and every 3-4 days, photos of the plants were taken from the top with a fixed digital camera (Canon EOS 1100D). Using an ImageJ plugin, the intensity of the green channel of these RGB images was multiplied by two, both the red and blue channels were subtracted and the image was converted to a binary image using the ImageJ 'Intermodes' Threshold Method. The binary images were manually inspected to ensure that all the leaves of each individual plant were connected. When leaves were not connected to each other, they were connected manually with a black line of two pixels in width. The surface area of each rosette was subsequently calculated in pixel² using the 'Analyze Particles' function of ImageJ, and converted to mm² by multiplication with the mm²/pixel² ratio of

every RGB image separately. The $\text{mm}^2/\text{pixel}^2$ ratio was calculated from the dimensions of the pots the plants were growing in, as this value is constant for every individual throughout the experiment. At 28 dpv, individual shoots were harvested and fresh weight was determined. Dry weight was determined after a further 2 days of incubation at 60 °C. The maximum efficiency of Photosystem II photochemistry (F_v/F_m) was quantified at 17 dpv (before salt treatment) and 28 dpv (after 12 days of salt treatment), respectively, as a measure of the effect of salinity on photosynthetic performance. F_0 and F_m images were captured using a CF Imager (Technologica, Essex, United Kingdom) by exposing dark adapted (> 30 min) plants to a maximal intensity $6226 \mu\text{mol m}^{-2} \text{s}^{-1}$ saturating actinic light pulse. The rosette surface area, biomass and chlorophyll fluorescence data were statistically analyzed using the heteroscedastic T-Test function of Microsoft Excel 2010 (assuming unequal variance between samples). A p -value of 0.05 was used as a threshold for significance.

Tissue sampling and RNA extraction

Seeds of Col-0, a mixture of five independent empty vector control lines in Col-0 background (transformed with a pRF-VP16-Kana T-DNA construct without 3F sequence), and of retransformant lines T31-5 and T43-11 were sterilized and stratified as described above. Seeds were then spotted on 245 mm x 245 mm Square BioAssay Dishes (Corning, Acton, MA, USA) containing salinity stress induction medium, and which were divided into quadrants. The seeds were spotted onto three independent dishes in different and separate quadrants for each condition (3 plates with 0 mM; 3 plates with 75 mM). At 21 dpv, the seedlings of each genotype from every quadrant were combinedly harvested in a 1:1:1 mass ratio, and grinded to powder in liquid nitrogen with pestil and mortar. In this way, two samples were obtained for each genotype consisting of tissue from three different dishes and three different quadrants (one for 0 and one for 75 mM). Total RNA was extracted from approximately 100 mg of each tissue sample with the RNeasy Plant Mini Kit (Qiagen). The quality of the RNA samples (RIN value > 8) was assessed with a Bioanalyzer (Agilent Technologies, Inc.) by the company BaseClear (Leiden, The Netherlands).

RNA sequencing and data analysis

A sequencing library was constructed from the RNA samples by the company BaseClear (Leiden, the Netherlands) [48, 49]. The library was sequenced by Illumina sequencing (all eight samples in one sequencing lane; 50 cycles; single read) by BaseClear. The quality of the samples was examined using FASTQC (version 0.10.1, www.bioinformatics.babraham.ac.uk/projects/fastqc). Genomic reference sequences and annotations were obtained from TAIR (version TAIR10) and supplemented with a fragment mapping to VP16 from pRF-VP16-Kana. The splicing-aware aligner TopHat (version 2.0.10, [50]) was used to map reads, using the 'very-sensitive' and 'coverage-search' options, and allowing for a maximum intron

size of 15000 bp. Secondary alignments were removed from the BAM files using SAMtools (version 0.1.18, [51]) and Perl. Reads aligning to annotated exons were summarized at the level of TAIR genes using HTSeq (version 0.5.3p9, [52]) with the ‘intersection-strict’ setting. At least 96% of raw sequencing reads could be uniquely assigned to a gene. Read counts were processed in R (version 3.0.2) using the edgeR package (version 3.4.2, [53]). Normalized expression values per gene (excluding mitochondrial and chloroplast sequences) were obtained by scaling using a robust estimate of the library size [54] and dividing by the mean length of the annotated transcripts in kilobasepairs. To acquire a more robust data set, genes with average mean normalized expression values lower than 5 over the different treatments and genotypes were not considered for the assessment of differential gene expression. Genes were considered differentially expressed between samples when their normalized expression values were > 2 fold higher or lower. The web tools “Venn Selector” and “Duplicate Remover” (<http://bar.utoronto.ca/welcome.htm>) were used for additional analysis of lists of gene accession numbers.

Acknowledgements

We would like to thank Dr. Christiaan Henkel (Institute of Biology Leiden, Leiden University, the Netherlands) for his help with transcriptome assembly and data analysis. We would like to thank Prof. Dr. H.J.M. de Groot and Dr. A. Alia Matysik (Leiden Institute of Chemistry, Leiden University, the Netherlands) for stimulating discussions. This work was carried out within the research programme of BioSolar Cells, co-financed by the Dutch Ministry of Economic Affairs.

References

1. Roy SJ, Negrao S, Tester M. (2014) Salt resistant crop plants. *Curr Opin Biotech* 26:115-24.
2. Munns R, Tester M. (2008) Mechanisms of salinity tolerance. *Ann Rev Plant Biol* 59:651-81.
3. FAO and Plant Nutrition Management Service. <http://www.fao.org/ag/agl/agll/spush2008>.
4. Zhang X, Lu G, Long W, Zou X, Li F, Nishio T. (2014) Recent progress in drought and salt tolerance studies in Brassica crops. *Breeding Sci* 64(1):60-73.
5. Deshmukh R, Sonah H, Patil G, Chen W, Prince S, Mutava R, *et al.* (2014) Integrating omic approaches for abiotic stress tolerance in soybean. *Front Plant Sci* 5:244.
6. Sahi C, Singh A, Kumar K, Blumwald E, Grover A. (2006) Salt stress response in rice: genetics, molecular biology, and comparative genomics. *Funct Integr Genomics* 6(4):263-84.
7. Yu S, Wang W, Wang B. (2012) Recent progress of salinity tolerance research in plants. *Genetika* 48(5):590-8.
8. USDA-ARS. Research Databases. Bibliography on Salt Tolerance (2008). *George E. Brown, Jr. Salinity Lab. US Dep. Agric., Agric. Res. Serv. Riverside, CA.* <http://www.ars.usda.gov/Services/docs.htm?docid=8908>
9. Cramer GR. (2002) Response of abscisic acid mutants of Arabidopsis to salinity. *Funct Plant Biol* 29(5):561-7.
10. Boyko A, Golubov A, Bilichak A, Kovalchuk I. (2010) Chlorine ions but not sodium ions alter genome stability of Arabidopsis thaliana. *Plant Cell Physiol* 51(6):1066-78.
11. Gupta B, Huang B. (2014) Mechanism of salinity tolerance in plants: physiological, biochemical, and molecular characterization. *Int J Genomics*. 2014:701596.
12. Goldack D, Li C, Mohan H, Probst N. (2014) Tolerance to drought and salt stress in plants: Unraveling the signaling networks. *Front Plant Sci* 5:151.
13. Zhang JL, Shi HZ. (2013) Physiological and molecular mechanisms of plant salt tolerance. *Photosynth Res* 115(1):1-22.
14. Wu SJ, Ding L, Zhu JK. (1996) SOS1, a genetic locus essential for salt tolerance and potassium acquisition. *Plant Cell* 8(4):617-27.
15. Liu JP, Zhu JK. (1998) A calcium sensor homolog required for plant salt tolerance. *Science* 280(5371):1943-5.
16. Zhu JK, Liu JP, Xiong LM. (1998) Genetic analysis of salt tolerance in Arabidopsis: Evidence for a critical role of potassium nutrition. *Plant Cell* 10(7):1181-91.
17. Shi HZ, Lee BH, Wu SJ, Zhu JK. (2003) Overexpression of a plasma membrane Na⁺/H⁺ antiporter gene improves salt tolerance in Arabidopsis thaliana. *Nat Biotech* 21(1):81-5.
18. Rus A, Yokoi S, Sharkhuu A, Reddy M, Lee BH, Matsumoto TK, *et al.* (2001) AtHKT1 is a salt tolerance determinant that controls Na⁺ entry into plant roots. *Proc Natl Acad Sci USA* 98(24):14150-5.
19. Berthomieu P, Conejero G, Nublat A, Brackenbury WJ, Lambert C, Savio C, *et al.* (2003) Functional analysis of AtHKT1 in Arabidopsis shows that Na⁺ recirculation by the phloem is crucial for salt tolerance. *Embo J* 22(9):2004-14.
20. Baxter I, Brazelton JN, Yu D, Huang YS, Lahner B, Yakubova E, *et al.* (2010) A coastal cline in sodium accumulation in Arabidopsis thaliana is driven by natural variation of the sodium transporter AtHKT1;1. *PLoS Genetics* 6(11):e1001193.
21. Cutler SR, Rodriguez PL, Finkelstein RR, Abrams SR. (2010) Abscisic acid: emergence of a core signaling network. *Annu Rev Plant Biol* 61:651-79.
22. Colebrook EH, Thomas SG, Phillips AL, Hedden P. (2014) The role of gibberellin signalling in plant responses to abiotic stress. *J Exp Biol* 217(Pt 1):67-75.

23. Nakashima K, Takasaki H, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. (2012) NAC transcription factors in plant abiotic stress responses. *Biochim Biophys Acta* 1819(2):97-103.
24. Tran LS, Nishiyama R, Yamaguchi-Shinozaki K, Shinozaki K. (2010) Potential utilization of NAC transcription factors to enhance abiotic stress tolerance in plants by biotechnological approach. *GM Crops* 1(1):32-9.
25. Akhtar M, Jaiswal A, Taj G, Jaiswal JP, Qureshi MI, Singh NK. (2012) DREB1/CBF transcription factors: their structure, function and role in abiotic stress tolerance in plants. *J Genet* 91(3):385-95.
26. Hichri I, Muhovski Y, Clippe A, Zizkova E, Dobrev PI, Motyka V, *et al.* (2016) SIDREB2, a tomato DEHYDRATION RESPONSIVE ELEMENT BINDING 2 transcription factor, mediates salt stress tolerance in tomato and Arabidopsis. *Plant Cell Environ* 39(1):62-79.
27. Zhang T, Zhang D, Liu Y, Luo C, Zhou Y, Zhang L. (2015) Overexpression of a NF-YB3 transcription factor from *Picea wilsonii* confers tolerance to salinity and drought stress in transformed Arabidopsis thaliana. *Plant Physiol Biochem* 94:153-64.
28. Saibi W, Feki K, Ben Mahmoud R, Brini F. (2015) Durum wheat dehydrin (DHN-5) confers salinity tolerance to transgenic Arabidopsis plants through the regulation of proline metabolism and ROS scavenging system. *Planta* 242(5):1187-94.
29. Qin Y, Tian Y, Liu X. (2015) A wheat salinity-induced WRKY transcription factor TaWRKY93 confers multiple abiotic stress tolerance in Arabidopsis thaliana. *Biochem Biophys Res Commun* 464(2):428-33.
30. Tiwari LD, Mittal D, Chandra Mishra R, Grover A. (2015) Constitutive over-expression of rice chymotrypsin protease inhibitor gene OCPI2 results in enhanced growth, salinity and osmotic stress tolerance of the transgenic Arabidopsis plants. *Plant Physiol Biochem* 92:48-55.
31. van Tol N, van der Zaal BJ. (2014) Artificial transcription factor-mediated regulation of gene expression. *Plant Sci* 225:58-67.
32. Lindhout BI, Pinas JE, Hooykaas PJ, van der Zaal BJ. (2006) Employing libraries of zinc finger artificial transcription factors to screen for homologous recombination mutants in Arabidopsis. *Plant J* 48(3):475-83.
33. Jia Q, van Verk MC, Pinas JE, Lindhout BI, Hooykaas PJ, van der Zaal BJ. (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 Biotech J* 11(9):1069-79.
34. Segal DJ, Dreier B, Beerli RR, Barbas CF, 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(6):2758-63.
35. Weijers D, Franke-van Dijk M, Vencken RJ, Quint A, Hooykaas P, Offringa R. (2001) An Arabidopsis Minute-like phenotype caused by a semi-dominant mutation in a RIBOSOMAL PROTEIN S5 gene. *Development* 128(21):4289-99.
36. Leister D, Varotto C, Pesaresi P, Niwergall A, Salamini F. (1999) Large-scale evaluation of plant growth in Arabidopsis thaliana by non-invasive image analysis. *Plant Physiol Bioch* 37(9):671-8.
37. Baker NR. (2008) Chlorophyll fluorescence: a probe of photosynthesis in vivo. *Annu Rev Plant Biol* 59:89-113.
38. Shen X, Wang Z, Song X, Xu J, Jiang C, Zhao Y, *et al.* (2014) Transcriptomic profiling revealed an important role of cell wall remodeling and ethylene signaling pathway during salt acclimation in Arabidopsis. *Plant Mol Biol* 86(3):303-17.
39. Quesada V, Ponce MR, Micol JL. (2000) Genetic analysis of salt-tolerant mutants in Arabidopsis thaliana. *Genetics* 154(1):421-36.
40. Mito T, Seki M, Shinozaki K, Ohme-Takagi M, Matsui K. (2011) Generation of chimeric repressors that confer salt tolerance in Arabidopsis and rice. *Plant Biotech J* 9(7):736-46.

41. Triezenberg SJ, Kingsbury RC, McKnight SL. (1988) Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. *Genes Dev* 2(6):718-29.
42. Cummings WJ, Bednarski DW, Maizels N. (2008) Genetic variation stimulated by epigenetic modification. *PLoS One* 3(12):e4075.
43. Memedula S, Belmont AS. (2003) Sequential recruitment of HAT and SWI/SNF components to condensed chromatin by VP16. *Curr Biol* 13(3):241-6.
44. Gill G, Ptashne M. (1988) Negative Effect of the Transcriptional Activator Gal4. *Nature* 334(6184):721-4.
45. Clough SJ, Bent AF. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16(6):735-43.
46. Masson J, Paszkowski J. (1992) The Culture Response of *Arabidopsis thaliana* Protoplasts Is Determined by the Growth-Conditions of Donor Plants. *Plant J* 2(5):829-33.
47. Murashige T, Skoog F. (1962) A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiol Plantarum* 15(3):473-97.
48. Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobisch S, *et al.* (2009) Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res* 37(18):e123.
49. Levin JZ, Yassour M, Adiconis X, Nusbaum C, Thompson DA, Friedman N, *et al.* (2010) Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat Methods* 7(9):709-15.
50. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14(4):R36
51. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078-9.
52. Anders S, Pyl PT, Huber W. (2015) HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2):166-9.
53. Robinson MD, McCarthy DJ, Smyth GK. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139-40.
54. Anders S, Huber W. (2010) Differential expression analysis for sequence count data. *Genome Biol* 11(10): R106

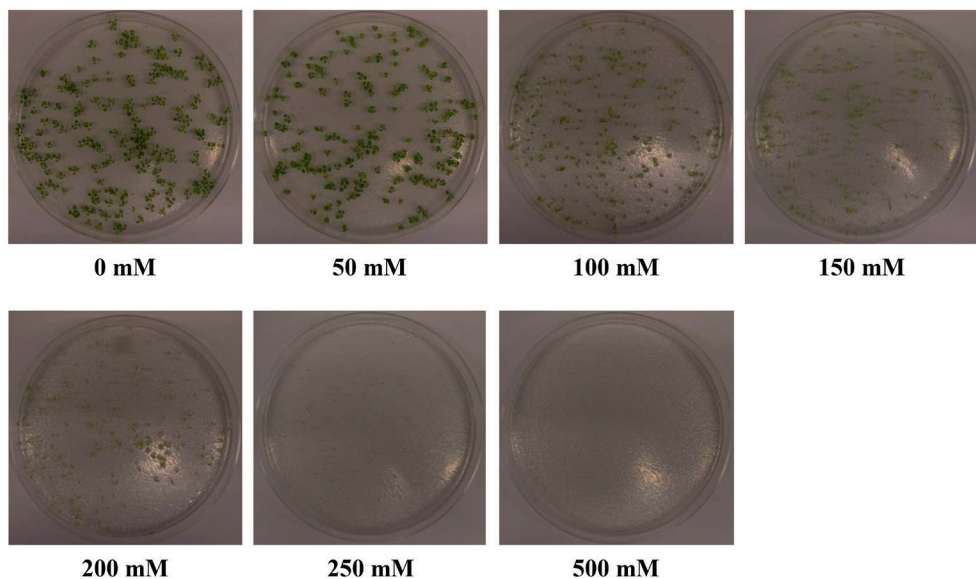


Fig. S1 Wild type Col-0 seeds plated on salinity stress induction medium containing the indicated concentrations of NaCl (14 dpv).

Table S1. Overview of the frequency of salinity tolerance among the retransformant lines harboring T-DNA constructs that were reconstituted from the original salinity tolerant 3F-VP16 lines (T2 generation; segregating for the T-DNA construct).

Retransformation with	Number of independent retransformant lines assayed	Number of salinity tolerant lines
pRF-VP16-Kana[S6 3F]	5	2
pRF-VP16-Kana[S7 3F]	5	2
pRF-VP16-Kana[S14 3F]	4	1
pRF-VP16-Kana[S22 3F]	4	1
pRF-VP16-Kana[S26 3F]	5	2
pRF-VP16-Kana[S27 2F]	5	3
pRF-VP16-Kana[S29 3F]	5	3
pRF-VP16-Kana[S31 3F]	5	5
pRF-VP16-Kana[S43 3F]	5	5
Total	43	24

Table S2. Overview of genotypes and treatments used for RNA sequencing, and the total RNA sequencing output. The accession Col-0 was used as the wild type; ‘empty vector control’ represents a mixture of 5 independent lines harboring T-DNA constructs with VP16, but without 3Fs.

Genotype	Treatment	Mapped reads
Col-0	0 mM NaCl	21,454,455
Col-0	75 mM NaCl	12,949,249
Empty vector control (pRF-VP16-Kana[-])	0 mM NaCl	18,588,293
Empty vector control (pRF-VP16-Kana[-])	75 mM NaCl	21,693,999
T31-5	0 mM NaCl	21,634,528
T31-5	75 mM NaCl	22,714,114
T43-11	0 mM NaCl	27,323,273
T43-11	75 mM NaCl	21,065,890

Table S3A. Overview of upregulated (> 2 fold) differentially expressed genes (DEGs; average normalized expression values over different treatments > 5 RPKM) in salt treated wild type Col-0 plants compared to control Col-0 plants. These genes are referred to as 'salt upregulated'.

Salt upregulated(459)										
AtCg00670	Atlg22370	Atlg60730	At2g04160	At2g32487	At3g12520	At3g49620	At4g12480	At4g30660	At5g14470	At5g42800
Atlg01470	Atlg22570	Atlg60740	At2g05520	At2g33150	At3g12700	At3g50260	At4g12490	At4g31050	At5g14920	At5g43440
Atlg02205	Atlg22890	Atlg61810	At2g14247	At2g33380	At3g12900	At3g50480	At4g12550	At4g31800	At5g15960	At5g43570
Atlg02300	Atlg22800	Atlg62710	At2g14560	At2g34300	At3g13450	At3g50770	At4g12720	At4g32650	At5g15970	At5g43580
Atlg02660	Atlg25275	Atlg62790	At2g14610	At2g34930	At3g13790	At3g51430	At4g14020	At4g33150	At5g16360	At5g44005
Atlg02850	Atlg25400	Atlg64660	At2g14900	At2g37330	At3g14060	At3g51660	At4g14060	At4g33720	At5g16450	At5g44568
Atlg02920	Atlg25530	Atlg65500	At2g15880	At2g38290	At3g14310	At3g51860	At4g14365	At4g34030	At5g16550	At5g44572
Atlg02930	Atlg26450	Atlg65510	At2g15890	At2g38390	At3g15356	At3g52180	At4g15233	At4g34180	At5g16910	At5g44585
Atlg04990	Atlg26770	Atlg65970	At2g15960	At2g38470	At3g15500	At3g52400	At4g15393	At4g34950	At5g17520	At5g45070
Atlg05340	Atlg27020	Atlg66160	At2g15970	At2g38530	At3g18080	At3g53990	At4g15610	At4g36430	At5g17760	At5g45750
Atlg07135	Atlg27730	Atlg67265	At2g16660	At2g38870	At3g18250	At3g54150	At4g16260	At4g36670	At5g18490	At5g46050
Atlg08315	Atlg29395	Atlg67300	At2g16700	At2g39030	At3g18290	At3g55005	At4g16370	At4g36760	At5g20230	At5g46180
Atlg08890	Atlg30720	Atlg67660	At2g16710	At2g39050	At3g18830	At3g55970	At4g17140	At4g36900	At5g20950	At5g48380
Atlg08980	Atlg30730	Atlg69410	At2g16750	At2g39330	At3g20470	At3g56170	At4g17230	At4g37220	At5g22500	At5g49520
Atlg09500	Atlg32170	Atlg69870	At2g17280	At2g39800	At3g21090	At3g56360	At4g17490	At4g37370	At5g23210	At5g50915
Atlg09932	Atlg32940	Atlg69880	At2g17840	At2g41260	At3g21230	At3g56980	At4g17500	At4g37450	At5g23575	At5g51390
Atlg10070	Atlg33790	Atlg69890	At2g18150	At2g41410	At3g21720	At3g57260	At4g18170	At4g37520	At5g23660	At5g51570
Atlg10550	Atlg33960	Atlg71140	At2g18690	At2g41730	At3g21770	At3g57780	At4g18280	At4g37530	At5g24030	At5g52310
Atlg12780	Atlg33970	Atlg72060	At2g22470	At2g42530	At3g22840	At3g58570	At4g18930	At4g38400	At5g24160	At5g52450
Atlg13609	Atlg35140	Atlg72070	At2g22880	At2g42540	At3g23550	At3g58750	At4g19200	At4g39090	At5g24210	At5g52810
Atlg13990	Atlg35320	Atlg72416	At2g23000	At2g42790	At3g23570	At3g59820	At4g19390	At4g39210	At5g24530	At5g53120
Atlg14250	Atlg35710	Atlg72680	At2g23010	At2g43510	At3g25770	At3g59930	At4g19410	At4g39955	At5g25190	At5g53450
Atlg14880	Atlg35720	Atlg72790	At2g23030	At2g43530	At3g25780	At3g60130	At4g19810	At5g03350	At5g25210	At5g54190
Atlg15010	Atlg36060	Atlg72900	At2g23120	At2g43570	At3g25900	At3g60140	At4g21680	At5g05250	At5g25240	At5g54810
Atlg15790	Atlg43160	Atlg73260	At2g24850	At2g43620	At3g26170	At3g60180	At4g21830	At5g05600	At5g25250	At5g57220
Atlg17020	Atlg47395	Atlg75190	At2g25200	At2g45220	At3g26740	At3g60420	At4g22470	At5g06320	At5g25610	At5g57550
Atlg17745	Atlg47400	Atlg75750	At2g25730	At2g46620	At3g26840	At3g62550	At4g22505	At5g06690	At5g25900	At5g57560
Atlg18200	Atlg49570	Atlg75900	At2g25770	At2g48090	At3g27210	At3g62860	At4g22753	At5g07010	At5g26340	At5g57900
Atlg18570	Atlg50040	Atlg76180	At2g25910	At3g03470	At3g27870	At4g01610	At4g23010	At5g07830	At5g27280	At5g59320

Table S3A. Continued

Salt upregulated(459)											
At1g19020	At1g51140	At1g76930	At2g26440	At3g03640	At3g28180	At4g02330	At4g23600	At5g08000	At5g27420	At5g59820	
At1g19180	At1g51940	At1g76960	At2g26560	At3g04000	At3g29320	At4g02370	At4g24570	At5g08380	At5g27520	At5g61520	
At1g19380	At1g52000	At1g77210	At2g26870	At3g04070	At3g29575	At4g02520	At4g25110	At5g10380	At5g28830	At5g63810	
At1g19610	At1g52040	At1g77640	At2g27080	At3g04640	At3g43270	At4g03820	At4g25810	At5g10760	At5g33290	At5g64120	
At1g20440	At1g52120	At1g78830	At2g28900	A3g04720	At3g44300	At4g04745	At4g26670	At5g11150	At5g35735	At5g64310	
At1g20450	At1g52400	At1g80160	At2g29350	At3g05290	At3g44720	At4g05330	At4g27440	At5g12420	At5g37540	At5g64560	
At1g20620	At1g52710	At1g80840	At2g29460	At3g05640	At3g44860	At4g06746	At4g27730	At5g13080	At5g39520	At5g64860	
At1g21250	At1g54040	At2g01023	At2g29670	At3g08860	At3g45140	At4g08040	At4g29010	At5g13200	At5g39580	At5g65390	
At1g21520	At1g54100	At2g01170	At2g30250	At3g09390	At3g45970	At4g09020	At4g29780	At5g13220	At5g40240	At5g67245	
At1g21650	At1g54575	At2g02850	At2g30550	At3g09940	At3g46370	At4g09030	At4g30270	At5g13330	At5g40450	At5g67480	
At1g21910	At1g55510	At2g02930	At2g30766	At3g10300	At3g46970	At4g11650	At4g30280	At5g13740	At5g40780		
At1g22065	At1g57990	At2g03760	At2g31360	At3g10410	At3g47800	At4g12290	At4g30290	At5g14180	At5g41910		
At1g22360	At1g59870	At2g04080	At2g31880	At3g11230	At3g49120	At4g12470	At4g30650	At5g14450	At5g42050		

Table S3B. Overview of downregulated (> 2 fold) differentially expressed genes (DEGs; average normalized expression values over different treatments > 5 RKPM) in salt treated wild type Col-0 plants compared to control Col-0 plants. These genes are referred to as 'salt downregulated'.

Salt downregulated(448)										
AtCG00020	Atlg19050	Atlg55960	Atlg76240	At2g35960	At3gl13980	At3g53720	At4g16880	At4g38960	At5g22580	At5g49730
AtCG00080	Atlg19150	Atlg56210	Atlg77490	At2g36690	At3gl16670	At3g53830	At4g17090	At4g38970	At5g22620	At5g49740
AtCG00280	Atlg21050	Atlg57770	Atlg77760	At2g36885	At3gl16690	At3g54500	At4g17460	At4g39235	At5g23360	At5g51460
AtCG00470	Atlg21060	Atlg58520	Atlg79160	At2g38300	At3gl17160	At3g54510	At4g17970	At4g39710	At5g23730	At5g51720
AtCG00480	Atlg21350	Atlg60270	Atlg79270	At2g39510	At3gl17330	At3g54830	At4g18220	At4g39770	At5g23750	At5g52070
Atlg01060	Atlg21460	Atlg60950	Atlg79410	At2g39850	At3gl17609	At3g55230	At4g18340	At5g01015	At5g24000	At5g52570
Atlg01790	Atlg22590	Atlg61580	Atlg80270	At2g40100	At3gl18500	At3g55800	At4g19170	At5g01075	At5g24120	At5g52640
Atlg02820	Atlg22650	Atlg62180	Atlg80760	At2g40260	At3gl19030	At3g56160	At4g19830	At5g01610	At5g24380	At5g52780
Atlg03055	Atlg22690	Atlg62250	At2g01520	At2g40460	At3g20395	At3g56290	At4g21990	At5g01740	At5g24420	At5g52900
Atlg03440	Atlg24148	Atlg62500	At2g04039	At2g40900	At3g21250	At3g57040	At4g24120	At5g02020	At5g24580	At5g54585
Atlg04240	Atlg26945	Atlg62560	At2g04570	At2g41040	At3g21390	At3g58850	At4g24510	At5g02120	At5g25840	At5g55570
Atlg04250	Atlg27370	Atlg62750	At2g05185	At2g41050	At3g22231	At3g58990	At4g24540	At5g02180	At5g26200	At5g55620
Atlg04520	Atlg29430	Atlg64500	At2g10940	At2g41310	At3g22420	At3g60390	At4g24670	At5g02540	At5g35190	At5g56860
Atlg04550	Atlg29440	Atlg64770	At2g14285	At2g41880	At3g23000	At3g62040	At4g24700	At5g02760	At5g35480	At5g57780
Atlg04800	Atlg29450	Atlg64780	At2g14660	At2g43140	At3g23050	At3g62930	At4g24972	At5g03170	At5g35490	At5g58390
Atlg05240	Atlg29460	Atlg65370	At2g15020	At2g43520	At3g23510	At4g00050	At4g25290	At5g04190	At5g36910	At5g58770
Atlg05250	Atlg29465	Atlg65390	At2g15620	At2g44670	At3g23805	At4g02270	At4g25400	At5g06530	At5g37020	At5g60530
Atlg06690	Atlg29500	Atlg65490	At2g16500	At2g45470	At3g24460	At4g04610	At4g25820	At5g07000	At5g37260	At5g61420
Atlg07180	Atlg29510	Atlg66100	At2g17300	At2g46650	At3g24500	At4g04770	At4g25830	At5g07460	At5g37990	At5g61440
Atlg07400	Atlg29720	Atlg66840	At2g18328	At2g46830	At3g25820	At4g04840	At4g26010	At5g07580	At5g38020	At5g62280
Atlg07570	Atlg30870	Atlg66940	At2g20560	At2g47240	At3g25830	At4g09890	At4g26150	At5g08350	At5g38200	At5g64770
Atlg08250	Atlg31335	Atlg67865	At2g21080	At2g47370	At3g26932	At4g10120	At4g26530	At5g08640	At5g38410	At5g64850
Atlg08810	Atlg31350	Atlg67870	At2g21320	At2g47560	At3g27170	At4g11211	At4g26850	At5g08650	At5g38420	At5g65140
Atlg09240	Atlg32080	Atlg68010	At2g21330	At3g01060	At3g27690	At4g11460	At4g26860	At5g09220	At5g38430	At5g66400
Atlg09340	Atlg32190	Atlg68238	At2g21650	At3g01440	At3g27750	At4g11830	At4g27030	At5g09660	At5g38520	At5g67350
Atlg10060	Atlg32450	Atlg68740	At2g23670	At3g01500	At3g28740	At4g12400	At4g28780	At5g12050	At5g40610	
Atlg10960	Atlg32470	Atlg68780	At2g26080	At3g01550	At3g29970	At4g12830	At4g30610	At5g15310	At5g40850	
Atlg11080	Atlg32900	Atlg68810	At2g26500	At3g02380	At3g43850	At4g12970	At4g31910	At5g15440	At5g41070	
Atlg11850	Atlg32928	Atlg68840	At2g26695	At3g03190	At3g45160	At4g12980	At4g33010	At5g15520	At5g41080	

Table S3B. Continued

Salt downregulated(448)										
Atlg12020	Atlg33811	Atlg69490	At2g26710	At3g06130	At3g46650	At4g13235	At4g33960	At5g16980	At5g42070	
Atlg12110	Atlg43800	Atlg69523	At2g27402	At3g06145	At3g46900	At4g13540	At4g34610	At5g17170	At5g42250	
Atlg12900	Atlg44000	Atlg70000	At2g28350	At3g07273	At3g47070	At4g13570	At4g34740	At5g17300	At5g42760	
Atlg13650	Atlg46768	Atlg70940	At2g28410	At3g09600	At3g47430	At4g13572	At4g36540	At5g18020	At5g44060	
Atlg14150	Atlg47600	Atlg70985	At2g28780	At3g10040	At3g47675	At4g13575	At4g36570	At5g18030	At5g44130	
Atlg14280	Atlg48480	Atlg71880	At2g29500	At3g10120	At3g48100	At4g14480	At4g37095	At5g18670	At5g44260	
Atlg15250	Atlg51270	Atlg72130	At2g30100	At3g11550	At3g48460	At4g14560	At4g37540	At5g19530	At5g46240	
Atlg15260	Atlg51470	Atlg73830	At2g30520	At3g11750	At3g48800	At4g14680	At4g37590	At5g19730	At5g46800	
Atlg17050	Atlg51805	Atlg73870	At2g31380	At3g12320	At3g49160	At4g15430	At4g37610	At5g20220	At5g47370	
Atlg17180	Atlg52060	Atlg74310	At2g32010	At3g12580	At3g51150	At4g15550	At4g37980	At5g20630	At5g47610	
Atlg18060	Atlg52070	Atlg74880	At2g32765	At3g12920	At3g52290	At4g15990	At4g38100	At5g21100	At5g48485	
Atlg18460	Atlg54050	Atlg75250	At2g34060	At3g13000	At3g52360	At4g16515	At4g38420	At5g22300	At5g48570	
Atlg18810	Atlg55910	Atlg75820	At2g35130	At3g13730	At3g52720	At4g16780	At4g38950	At5g22390	At5g49480	

Table S4. Overview of differentially expressed genes (DEGs; average normalized expression values over different treatments > 5 RKPM) in salinity tolerant retransformant lines T31-5 and T43-11 compared to the wild type Col-0, all treated with 75 mM NaCl, which are not DEGs in the empty vector control (expressing VP16 without 3Fs). Upregulated DEGs have > 2 fold normalized expression values, downregulated DEGs have < 0.5 fold normalized expression values.

Upregulated (69)			Downregulated (95)			
Salt regulated (13)	Not salt regulated (56)		Salt regulated (55)	Not salt regulated (40)		
"At1g02205	At1g13590	At3g28345	^d At1g02820	"At3g55970	At1g22400	At4g15680
^d At1g05240	At1g19530	At3g54040	"At1g19610	"At3g56360	At1g23205	At4g16563
"At1g29395	At1g20190	At3g59220	"At1g25400	"At3g60140	At1g23020	At4g18010
^d At1g29460	At1g21310	At3g62270	"At1g27020	"At4g02330	At1g24530	At4g20860
^d At2g15020	At1g22440	At4g11320	"At1g35140	^d At4g10120	At1g28660	At4g21870
^d At3g07273	At1g23080	At4g12510	"At1g36060	"At4g15233	At2g04040	At4g31870
^d At3g11550	At1g33800	At4g12520	^d At1g65390	"At4g16370	At2g05380	At5g05440
^d At3g48100	At1g51830	At4g15160	"At1g66160	"At4g17500	At2g25000	At5g08150
^d At3g49160	At1g51840	At4g30170	"At1g72900	"At4g18170	At2g25735	At5g20150
^d At4g09890	At1g62770	At4g30670	"At1g75750	"At4g21680	At2g28630	At5g20240
^d At4g25400	At2g04170	At4g33610	"At1g76930	"At4g22470	At2g32530	At5g44575
^d At5g23750	At2g05510	At4g38080	"At2g14247	"At4g25810	At2g32550	At5g65010
^d At5g54585	At2g18980	At4g38860	"At1g21910	"At4g27730	At2g34070	
	At2g28760	At4g39675	"At2g24850	"At4g37450	At2g38310	
	At2g32100	At5g03150	"At2g25770	"At5g05250	At2g40330	
	At2g38380	At5g11420	"At2g26440	"At5g05600	At2g43920	
	At2g38380	At5g15230	"At2g27080	"At5g13220	At3g02140	
	At2g43590	At5g17330	"At2g30766	"At5g14180	At3g09922	
	At3g05890	At5g18860	"At2g39030	^d At5g16980	At3g16150	
	At3g07010	At5g25460	"At2g43510	"At5g24160	At3g21530	
	At3g10940	At5g25475	"At3g12900	"At5g39580	At3g22060	
	At3g12710	At5g47100	"At3g15356	"At5g43570	At3g23730	
	At3g13435	At5g50760	"At3g21090	^d At5g44130	At3g25760	
	At3g16920	At5g53250	"At3g22840	"At5g44572	At3g26320	
	At3g20380	At5g56540	^d At3g28740	"At5g46050	At3g30775	
	At3g23180	At5g57090	"At3g44860	"At5g53450	At3g50560	
	At3g23480	At5g58000	"At3g50480	"At5g57220	At3g50570	
	At3g25930	At5g59520	^d At3g54830		At4g14690	

^d Salt downregulated in Col-0.

" Salt upregulated in Col-0.

Table S5. Overview of differentially expressed genes (DEGs; average normalized expression values over different treatments > 5 RKPM) in salinity tolerant retransformant lines T31-5 and T43-11 compared to the empty vector control (expressing VP16 without 3Fs), all treated with 75 mM NaCl. Upregulated DEGs have > 2 fold normalized expression values, downregulated DEGs have < 0.5 fold normalized expression values.

Specific for T31-5		Specific for T43-11		Shared between T31-5 and T43-11	
Upregulated (13)	Downregulated (20)	Upregulated (6)	Downregulated (2)	Upregulated (12)	Downregulated (9)
At1g27565	^d At1g02820	At4g07820	At2g14610	^d At1g05240	"At1g13609
At1g51830	"At1g52120	At4g12510	At4g11460	^d At1g05250	"At1g47395
At1g51840	"At1g69880	At4g12520		At1g49860	"At1g47400
At1g80240	"At2g14247	At5g14330		At3g20380	"At2g30766
At2g01530	At2g19800	At5g43500		^d At3g43850	At3g09922
At2g43590	"At2g25770	At5g53250		At3g54590	At3g01345*
At3g48410	"At2g39030			"At3g57260"	"At3g12900
At3g53980	At3g16150			^d At4g02270	At4g14690
^d At4g13570	"At3g22840			^d At4g25820	At4g31940
At4g25580	"At3g55970			^d At4g26010	
At4g30670	"At3g60140			At4g40090	
At5g23980	"At4g12490			^d At5g35190	
At5g38940	At4g15680				
	At4g26260				
	"At5g05250				
	^d At5g16980				
	At5g20240				
	"At5g43570				
	AtCg00140				
	AtCg00220				

*Meet the indicated DEG selection criteria, but expression values are highly variable; might better be disregarded.

^d Salt downregulated in Col-0.

" Salt upregulated in Col-0.

Grey highlights indicate genes that are downregulated by salt treatment in wild type plants, but upregulated in the 3F-VP16 line, or were very close to meeting these criteria. At3g12900 was very close to meeting the opposite of these criteria.

Chapter 7

Nederlandse samenvatting

Niels van Tol

Fotosynthese is het proces dat planten gebruiken om energie te oogsten uit licht en vast te leggen in de vorm van chemische verbindingen en wordt daarom beschouwd als de drijvende kracht achter de productiviteit van planten. Ondanks het feit dat fotosynthese vormgegeven is door vele miljoenen jaren van evolutie lijkt de theoretisch gezien maximaal haalbare efficiëntie waarmee lichtenergie wordt omgezet tot biomassa betrekkelijk laag: 4,6% voor C3 planten en 6,0% en voor C4 planten. Onder natuurlijke omstandigheden zijn de condities waaronder planten groeien meestal suboptimaal en daarom zijn de *in planta* omzettingsefficiënties vaak niet hoger zijn dan 1%, wat ook geldt voor belangrijke gewasplanten. Er lijkt dus ruimte te zijn voor het verhogen van de efficiëntie van fotosynthese van gewasplanten tot de theoretisch mogelijke 6,0%. De verhoging van fotosynthetische efficiëntie - op wat voor manier dan ook - is essentieel om de snel groeiende wereldbevolking van voldoende plantaardige producten te kunnen voorzien. Hoewel er een aantal verschillende componenten van fotosynthese zijn benoemd als de voornaamste limitaties in termen van efficiëntie, is het tot op heden echter nog niet gelukt om met de genetische modificatie van planten werkelijk verbeteringen te realiseren die voldoende zijn om de wereldbevolking in haar toekomstige behoefte te kunnen voorzien.

In **Hoofdstuk 1** van dit proefschrift worden de verschillende limitaties van fotosynthese besproken en wordt in meer detail toegelicht waarom de reeds gekozen wetenschappelijke strategieën om fotosynthese te verbeteren niet adequaat lijken te zijn. Het is daarom van groot belang dat er nieuwe technieken ontwikkeld worden om de extreme variatie in fotosynthetische efficiëntie, stresstolerantie en productiviteit van planten te genereren die vereist zijn om werkelijk vooruitgang te kunnen boeken. In dit proefschrift zijn de resultaten beschreven van mijn onderzoek naar het gebruik van artificiële transcriptiefactoren in de modelplant *Arabidopsis thaliana* om nieuwe fotosynthese-gerelateerde fenotypen te genereren. Deze relatief nieuwe techniek, genaamd 'genome interrogation' wordt in detail uitgelegd in **Hoofdstuk 2** en biedt de mogelijkheid tot verandering van de expressie van een groot aantal genomische loci tegelijkertijd door artificiële transcriptiefactoren te introduceren die binden aan relatief korte DNA sequenties.

Omdat de verhoging van fotosynthetische efficiëntie in principe een middel is om een verbeterde productiviteit van planten te bereiken, is in **Hoofdstuk 3** van dit proefschrift ten eerste onderzocht of 'genome interrogation' gebruikt kan worden om *Arabidopsis* planten meer biomassa te laten accumuleren. Hiertoe zijn er twee collecties van transgene *Arabidopsis* planten gegenereerd die artificiële genconstructen bevatten die coderen voor een groot aantal verschillende 'zinc finger' artificiële transcriptie factoren (ZF-ATFs). Van deze transgene planten zijn in detail de groei-eigenschappen gedocumenteerd. Op deze manier zijn er ZF-ATF genconstructen gevonden waarvan expressie in *Arabidopsis* planten resulteert in een significante groei- en biomassa toename.

Naast genen in de celkern van planten zijn ook een relatief groot aantal genen in het chloroplastgenoom betrokken bij fotosynthese. Veel van deze chloroplastgenen coderen voor structurele componenten van de fotosynthetische eiwitcomplexen in chloroplasten. Differentiële expressie van deze genen zou mogelijkwerwijs kunnen leiden tot veranderingen in fotochemische activiteit. In **Hoofdstuk 4** is het ontwerp beschreven van een nieuw systeem om 'genome interrogation' toe te passen op het chloroplastgenoom van Arabidopsis door middel van genconstructen die integreren in het kerngenoom en coderen voor ZF-ATFs met bacteriële transcriptionele activatordomeinen. Door middel van deze aanpak zijn sterke aanwijzingen gevonden dat ZF-ATF geïnduceerde differentiële expressie van chloroplastgenen een goede aanpak zou kunnen zijn om fotosynthetische efficiëntie te verhogen.

De voornaamste techniek om fotosynthetische activiteit te fenotyperen in grote populaties van transgene planten is de kwantificatie van chlorofyl fluorescentie (CF) parameters. De efficiëntie waarmee Fotosysteem II in de thylakoidmembranen van chloroplasten lichtenergie gebruikt voor de reductie van plastoquinon A wordt ϕ PSII genoemd en heeft grote voorspellende waarde voor de algehele fotosynthetische activiteit van planten. Verassend genoeg is nog nooit een grootschalige screen voor Arabidopsis mutanten met significant verhoogde ϕ PSII waarden gepubliceerd. In **Hoofdstuk 5** is de isolatie beschreven van een nieuwe, recessieve Arabidopsis mutant met zeer hoge ϕ PSII waarden en lage CF door middel van CF analyse van een grote populatie van Arabidopsis planten met ZF-ATF constructen. Dit hoofdstuk beschrijft verder in detail alle groei- en CF eigenschappen van deze mutant.

Omdat de meest efficiënte fotosynthese alleen plaats kan vinden onder zeer gunstige omgevingscondities is het belangrijk om planten te kunnen generen die naast hoge productiviteit ook stresstolerant zijn. In de hedendaagse agricultuur is de ophoping van zouten in de bodem één van de meest relevante abiotische stressfactoren voor planten. In **Hoofdstuk 6** is de screening beschreven van een collectie van Arabidopsis lijnen getransformeerd met ZF-ATF constructen voor tolerantie tegen 100 mM NaCl, een concentratie die normaal gesproken lethaal is voor Arabidopsis planten. Op deze manier is een relatief groot aantal nieuwe mutanten gevonden die zoutstress goed kunnen verdragen. Door analyse van deze mutanten zijn een aantal ZF-ATFs gevonden die zouttolerantie via een niet eerder gedocumenteerd mechanisme kunnen induceren in Arabidopsis.

Curriculum Vitae

Curriculum Vitae

Niels van Tol was born on the 17th of September, 1988, in the city of Leiden, the Netherlands. He grew up in Leiden, and received his high school diploma from the Adelbert College located in the nearby village of Wassenaar in 2006. In September 2006, he started his BSc studies of Biology at Leiden University, receiving his BSc diploma in July 2009. In September 2009, he continued his scientific education by entering the 'Molecular and Cellular Biosciences' MSc program at Leiden University. After having received his MSc degree in June 2011, he started his PhD research in July 2011 in the research group of Molecular and Developmental Genetics of the Institute Biology Leiden under the supervision of Dr. Bert J. van der Zaal and Prof. Dr. Paul J.J. Hooykaas, the results of which are described in this thesis. Currently, he is working as a postdoc in the same research group on the STW funded project 'Novel strategies to control meiotic recombination', which is a collaboration between Leiden University and the seed company Rijk Zwaan.

