Received Date : 23-Sep-2014

Revised Date : 09-Mar-2015

Accepted Date : 31-Mar-2015

Article type : Technical Advance

A library of synthetic Transcription Activator-Like Effector-activated promoters for coordinated orthogonal gene expression in plants

Kathleen Brückner[1,#], Petra Schäfer[1], Ernst Weber[2,‡], Ramona Grützner[1], Sylvestre Marillonnet[1,*] and Alain Tissier[1,*]

[1] Department of Cell and Metabolic Biology, Leibniz Institute of Plant Biochemistry, Weinberg 3, 06120 Halle (Saale), Germany

[2] Icon Genetics GmbH, Weinbergweg 22, 06120 Halle (Saale), Germany

* Corresponding authors:

Alain Tissier: Tel: +49 345 5582 1500; Fax: +49 345 5582 1009; Email: alain.tissier@ipb-halle.de

Sylvestre Marillonnet: Tel: +49 345 5582 1537; Fax: +49 345 5582 1009; Email : smarillo@ipb-halle.de

‡ Present Address: Bayer HealthCare, Protein Engineering and Assays, Nattermannallee 1, 50829 Cologne, Germany

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/tpj.12843

Abstract

A library of synthetic promoters containing the binding site of a single designer Transcription Activator-Like Effector (dTALE) was constructed. The promoters contain a constant sequence, consisting of an eighteen base long dTALE-binding site and a TATA box, flanked by degenerate sequences of forty-nine bases downstream and nineteen bases upstream. Forty-three of these promoters were sequenced and tested in transient assays in *Nicotiana benthamiana* using a GUS reporter gene. The strength of expression of the promoters ranged from around 5% to almost 100% of the viral 35S promoter activity. We then demonstrated the utility of these promoters for metabolic engineering by transiently expressing three genes for the production of a plant diterpenoid in *N. benthamiana*. The simplicity of the promoter structure shows great promise for the development of genetic circuits, with wide application potential in plant synthetic biology and metabolic engineering.

Keywords

synthetic promoters; transcription activator-like effectors; orthogonality; metabolic engineering , transient assays , Nicotiana benthamiana , terpenoid, synthetic biology; plants.

Running title

A library of dTALE-activated synthetic promoters

Introduction

One of the goals of synthetic biology is to establish artificial transcriptional networks for the expression of novel signalling or metabolic pathways. While this was successfully done in unicellular microorganisms (Brophy and Voigt, 2014), the development of such transcriptional networks in higher Eukaryotes, in particular in higher plants, presents

additional difficulties due to the lack of sufficient available data, multicellularity and chromatin structure. One prerequisite for the construction of such networks is the availability of a range of transcriptional promoters of varying strengths that can be activated by orthogonal transcription factors. Since such promoters are not available in nature, they have to be designed, synthesized, assembled in functional transcription units and tested.

Several approaches have been adopted for promoter engineering, including random mutagenesis, saturation mutagenesis of poorly conserved regions and the engineering of hybrid promoters (Blazeck and Alper, 2013). Expression levels achieved so far covered a ten-fold to over a thousand-fold range (Blazeck and Alper, 2013), indicating that such sets of promoters are adequate to finely tune the expression of individual genes for engineering of pathways. An additional desirable property of synthetic promoters is their activation by orthogonal transcription factors (TFs). This is particularly relevant for multicellular organisms such as higher plants, for example if the aim is to express a biosynthesis pathway exclusively in one tissue or under specific conditions, such as biotic or abiotic stresses. Most promoter libraries described so far in yeast are based on upstream activating sequences (UAS) of endogenous TFs (Blazeck and Alper, 2013). If the TF, on which the design of these synthetic promoters is based, is expressed constitutively, or only in specific tissues or conditions, this will considerably restrict their potential use. For constitutive TF, this will prevent the activation in specific tissues, and conversely for tissue-specific TFs, this will restrict the use of the promoters to a specific tissue, preventing their general application for other tissues or conditions. A set of promoters whose expression can be induced by an orthogonal TF would circumvent these problems. In this case, the localization and pattern of expression is determined by the orthogonal TF, which itself can be put under the control of a tissue-specific promoter for example.

The recently discovered Transcription Activator-Like Effectors (TALEs) provide the perfect raw material to develop such a tool (Boch and Bonas, 2010, Boch et al., 2009). The modular structure of their DNA binding site allows the design of truly orthogonal transcription factors by selecting a DNA sequence that does not occur in the genome, thus minimizing off-target

gene activation. Since their discovery, TALEs have been used both for transcription activation and genome editing, the latter by fusing them to nucleases (de Lange et al., 2014). While tools for the construction of custom-designed TALEs are available (Weber et al., 2011), there are, to the best of our knowledge, no described resources for libraries of synthetic promoters that can be activated by custom-designed TALEs. Such a resource would not only allow the coordinated expression of metabolic pathway genes in specific cell types or under given conditions, but also provide the building blocks to construct artificial transcriptional networks by introducing positive or negative feedback loops, or transcriptional cascades. To fill this gap and motivated by our interest in plant metabolic engineering, we undertook the design and construction of libraries of synthetic TALE-activated promoters (STAPs), which we describe here. These were assembled in Golden Gate compatible vectors to facilitate their downstream use (Engler et al., 2014). These STAPs were first evaluated with reporter genes (GUS and GFP) in transient assays in *Nicotiana benthamiana*, then, as proof of principle, tested in a metabolic engineering set-up for the production of a plant diterpene.

Results

Design and construction of a library of synthetic TALE-activated promoters

The structure of the STAPs is presented in Figure 1A. They contain an eighteen base-long DNA binding domain, named EBE002, flanked by a nineteen base long degenerate sequence immediately upstream, and a consensus TATA box downstream. The TATA box is followed by a forty-three base long degenerate sequence and the ATG start codon. This design is based on the observation that the DNA-binding sequence of TALEs from bacterial pathogens that activate genes *in planta* are typically located close to the transcription start site (less than 100 bp) and that insertion of a DNA binding sequence at position -55 or -40 was sufficient to confer TALE-mediated inducibility (Kay et al., 2009). The library was cloned

using the MoClo system based on Golden Gate cloning (Werner et al., 2012). The scheme for the construction of the library of STAPs is presented in Figure 1B. The library was made in a level -1 vector of the MoClo system, yielding pAGT559-L, and transferred in a level 0 vector, yielding pAGT582-L. Level 0 vectors contain parts (typically promoter, cDNA, terminator) which are brought together in level 1 vectors to assemble functional transcription units. Transferring the library from level -1 to level 0 ensures that only fragments containing correct overhangs are kept, thus limiting the number of false positives and increasing the efficiency of cloning in subsequent steps. Approximately 50 clones from the level 0 library were picked and checked for the presence of a fragment of the expected size. Fourty-three clones were thus selected and sequenced (see supplemental data). All promoter sequences contained the EBE002 DNA binding site.

Transient assays of the STAPs with a GUS reporter gene

Each STAP was cloned upstream of the GUS reporter gene and the ocs transcription terminator in a level 1 vector, which can also be used for *Agrobacterium tumefaciens* T-DNA transfer (Figure 2A). The transcriptional activity of these promoters was first evaluated in transient assays in *Nicotiana benthamiana* by measuring GUS activity in protein extracts. The dTALE is under control of the Act2 promoter (Figure 2A), which provides sufficient expression for the activation of transcription. For the quantification of the expression, leaves were infiltrated in triplicate with the STAP:GUS constructs, with or without the dTALE and with a 35S:GFP construct. GFP fluorescence was measured and used as a normalization factor. The results (Figure 2B) show a distribution of expression strengths ranging from less than 1% to over 90% of that of the 35S promoter. Since for some promoters there is significant residual background expression without the TALE, the useful range of expression in this transient assays is estimated to be between 5 to 90% of that of the 35S promoter. For most STAPs however, no significant expression was detected in the absence of the TALE,

indicating that the activity seen with the TALE does not stem from endogenous activators, establishing the orthogonality of the promoters.

Transient assays of selected STAPs with GFP

To confirm these results with another reporter gene, a selection of eight STAPs with different expression strengths (STAPs numbers 1, 2, 3, 5, 14, 17, 42 and 44) were cloned in front of the GFP gene and transiently expressed in *N. benthamiana*. The results (Figure 3) are consistent with those of the GUS assay, indicating that the strength of expression for GUS and GFP seems to be independent of the transcribed gene. For all the tested STAPs, no GFP signal can be seen in the absence of the TALE, confirming the high-specificity of the synthetic promoters. The strongest promoters (numbers 1, 3 and 5) show a signal intensity approaching that of the 35S promoter, while the weaker promoters (numbers 17 and 42) still produce a detectable signal.

Using the STAPs for diterpene metabolic engineering in N. benthamiana

We next sought to validate the utility of these promoters by using them for metabolic engineering of a plant diterpenoid. As output for our metabolic engineering assay we used the cembratrienol synthase (CBTS2a) from tobacco (Ennajdaoui et al., 2010, Wang and Wagner, 2003). This diterpene synthase produces a mix of two stereoisomers (α and β) of the macrocyclic cembratrienol (CBTol) (Figure 4). In plants, the terpenoid precursors isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) are produced by two distinct pathways, the plastidic 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway, and the cytosolic mevalonate (MEV) pathway (Vranova et al., 2012). Diterpenes are synthesized in the plastids and therefore predominantly use IPP and DMAPP from the MEP pathway. We previously showed that co-expression of 1-deoxyxylulose 5-phosphate synthase (DXS), the

first step of the MEP pathway, with a geranylgeranyl diphosphate synthase (GGPPS) leads to more than a doubling of diterpene production in *N. benthamiana* transient assays (Brückner and Tissier, 2013). These assays were carried out with the strong Cauliflower Mosaic Virus 35S promoter controlling the expression of all the genes. The *DXS2* (tomato), *GGPPS2* (tobacco) and the *CBTS2a* genes were cloned downstream of some of the strongest STAPs characterized above and transiently expressed together with the dTALE. The results show that both with the 35S promoter and the STAPs, the amount of CBTols produced is significantly higher when all genes (*CBTS2a*, *GGPPS2* and *DXS2*) are co-expressed compared to the other combinations (*CBTS2a* alone, *CBTS2a* + *DXS2* or *CBTS2a* + *GGPPS2*) (Figure 4C), confirming previous results with the 35S promoter (Brückner and Tissier, 2013). However, CBTols produced with the STAPs are around three-fold lower than with 35S. One explanation could be that gene expression is lower, leading to reduced CBTols levels. To check this, we measured transcript levels by quantitative RT-PCR from the same samples used for CBTol quantification. Surprisingly, the transcript levels were higher with the STAPs than with the 35S promoter (1.9 times for *CBTS2a*, 11.5 for *GGPPS2* and 1.5 for *DXS2*) (Figure 5). Thus, the lower CBTol levels cannot be caused by reduced transcript levels of the genes. Plotting the gene transcript levels versus the amount of CBTols produced indicates a negative correlation, regardless of the promoters used (STAPs or 35S) (supplemental Figure 1). Thus, this inverse effect of gene expression on metabolite levels does not seem to be specific to the STAPs.

These results show that the STAPs can be used for metabolic engineering applications and can deliver expression levels, which are comparable to those of the strongest known promoters.

Discussion

We report here on the design, construction and testing of a library of synthetic promoters activated by a designer TALE (dTALE) transcription factor in plants. Based on a remarkably simple architecture consisting of the dTALE-DNA binding site, a TATA box and degenerate sequences flanking these constant regions, these STAPs exhibit properties that make them ideal tools for the development of complex transcriptional networks or for metabolic engineering applications. First, based on the *GUS* reporter gene, they exhibit a usable range of expression levels in transient assays that varies from 5 to 90% of that of the 35S promoter, which is one of the strongest plant promoters available. It is however possible that transient assays in *N. benthamiana*, due to the high load of T-DNAs that are transferred, lead to background transcription activity. Whether this remains the case in stable transgenics remains to be seen. Large variations in expression levels of individual genes within a biosynthesis pathway have been observed in plants. For example, genes encoding the cytochrome P450 monooxygenases of the parthenolide pathway in feverfew (*Tanacetum parthenium*) can be expressed at levels several hundred fold higher than the gene encoding the germacrene A synthase, which represents the first committed step of the pathway (Liu et al., 2014). Thus, the range of expression levels delivered by the STAPs should prove particularly useful to tune the expression of individual pathway genes in metabolic engineering projects. This can serve to optimize pathway flux but also to avoid excess production of certain enzymes, which may cause metabolic overload of the host cells. Second, in the absence of the dTALE, no or background levels of expression were detected, demonstrating the orthogonal nature of the dTALE/STAP system.

We could also validate the utility of the STAPs for metabolic engineering in a transient assay. We could confirm that co-expression of GGPPS2 and DXS2 with the diterpene synthase CBTS2a leads to increased levels of CBTols (Figure 4C). However, the levels were lower than with the 35S promoter, which could have been due to lower gene expression levels. Unexpectedly, however, transcript levels measured by quantitative RT-PCR show

that the STAPs confer significantly higher expression than the 35S promoter, in particular up to more than tenfold higher for the *GGPPS2* gene. Several hypotheses can be proposed to account for this negative correlation between transcript levels and product yield. The high expression of the enzymes could lead to negative feedback on the MEP pathway and therefore reduce flux through this pathway. Another possibility is that CBTols are produced in larger amounts with the STAPs but are modified by endogenous *N. benthamiana* enzymes (e.g. acyl- or glycosyltransferases, glutathione transferases), as has been shown for mono- and sesquiterpenoids (Ting et al., 2013). Here, however, one would have to invoke that the more CBTols are produced, the more they are modified in proportion. Yet another possibility is that the higher transcript levels are not accompanied by higher translation rates, rather in the contrary this leads to reduced translation activity. Testing these different hypotheses will require extensive additional experiments which should be carried out in the future.

Eukaryotic promoters are currently described as containing two major domains, namely the core promoter element, which is defined as the region where the transcription machinery is recruited and initiates transcription, and the upstream elements, which are bound by transcription factors that can either activate or repress transcription (Hahn and Young, 2011). It is well established that binding to specific upstream elements by transcription factors triggers the recruitment of the core transcription machinery and thereby activates transcription (Hahn and Young, 2011). Hardly anything is known, however, on how the sequence of the core promoter element influences transcriptional activity (Lubliner et al., 2013). In an early study, the analysis of 95 yeast promoters lead to the identification of a so-called “locator” signal 30-10 bp upstream of the transcription start site (TSS), which is rich in Ts and poor in As, and is “stronger” in promoters with high activity (Maicas and Friesen, 1990). More recently, a survey of 859 yeast core promoters identified T-richness upstream of the TSS as a strong predictor of promoter activity, regardless of whether the promoters are constitutive or inducible (Lubliner et al., 2013). As far as we could tell from the available literature, nothing is known in plants. Because of their short sequence and the presence of

an identical DNA binding site, the STAPs constitute an appropriate material to study the influence of the base composition of the core promoter on transcription activity. However, the identification of significant correlations between expression level and sequence content in our dataset would require a larger number of sequences.

Transcript levels were measured by quantitative RT-PCR for three of the promoters (STAP-1, STAP-3, STAP-5; Figure 5). STAP-1 and STAP-3 promoters showed transcript levels that were similar to those of the 35S promoter, a result which is consistent with the GUS expression assays. Surprisingly STAP-5 showed transcript levels around ten times higher as with the 35S promoter, although GUS expression levels were similar to those with the 35S promoter. The base content between the transcription start site and the translation initiation codon (i.e. the 5'-UTR region) is known to have a dramatic impact on translation efficiency leading to up to 100-fold differences in translation efficiency (Rojas-Duran and Gilbert, 2012). It is therefore quite possible that in the case of STAP-5 the high transcript levels do not result in correspondingly high protein expression due to an unfavourable 5'-UTR sequence. One way to circumvent this issue would be to introduce a 5'-UTR sequence known to enhance translation efficiency such as the Ω–translation enhancer of the tobacco mosaic virus (Sleat et al., 1987).

Our design of the STAPs was based on reports that identify the presence of AvrBs3-binding sites within 100 bp of the transcription start site of genes activated by AvrBs3, a TAL-effector from the plant pathogenic bacterium *Xanthomonas campestris* (Kay et al., 2007, Kay et al., 2009, Romer et al., 2007). Furthermore, Kay and collaborators showed that a TATA box immediately downstream of the TALE binding site leads to greatly increased expression (Kay et al., 2009). Although our design gave satisfactory results, further variations could provide additional regulatory options. For example, adding multiple TALE-binding sites could lead to increased expression, as shown in mammalian cells (Perez-Pinera et al., 2013). However, since our strongest expressors are almost on a par with the strong viral 35S promoter this may be of limited interest. More relevant would be the insertion of TALE-

binding sites for positive or negative feedback regulation. Positive feedback can simply be done by adding a DNA-binding site for another dTALE. Negative feedback can be achieved in the same way but by replacing the activation domain of the second TALE by a repressor domain. Transforming dTALEs into transcriptional repressors was successfully done in various eukaryotic organisms including plants (Blount et al., 2012, Mahfouz et al., 2012). Thus, our STAPs provide the starting material to design complex regulatory circuits that could include transcriptional cascades, Boolean operators, as well as positive and negative feedback loops.

Given the ever-increasing number of genome sequences that become available, including from higher plants, it will be possible to design TALE-DNA binding sequences that do not or minimally interfere with endogenous gene expression and conversely that are not bound to by endogenous TFs. Thus, the dTALE/STAPs system can be considered a truly orthogonal system that neither depends nor interferes with endogenous transcription activation networks. The relevance of this system for metabolic engineering is obvious. Our transient expression assay for diterpene production in *N. benthamiana* demonstrated that several genes can be co-expressed by the dTALE/STAPs system at levels that are comparable to the strong CaMV 35S promoter. While using the same promoter to co-express multiple genes in a transient assay works properly, in stable transgenic plants this can lead to undesirable consequences such as gene silencing (Peremarti et al., 2010). Therefore, a particularly relevant application of these promoters will be in stable transgenic plants, where coordinated tissue-specific expression of pathway genes can be tested without requiring the identification of multiple endogenous promoters conferring the same expression profile and with different expression levels. Hence, by simply changing the promoter driving the expression of the activating dTALE and keeping the rest of the construct identical, it will be possible with minimum cloning effort to test different tissues in parallel for their performance in producing a compound of interest. For example, one could compare tissues such as the

seed endosperm, the epidermis, glandular trichomes, or root hairs, for which specific promoters are available (Potenza et al., 2004, Tissier, 2012).

Recently, the discovery of the bacterial immune system based on the CRISPR/Cas9 system has opened novel opportunities for genome engineering and synthetic biology (Fineran and Dy, 2014, Hsu et al., 2014, van der Oost et al., 2014). The CRISPR/Cas9 enzyme is a nuclease whose sequence specificity is determined by a guide RNA (gRNA) containing the target sequence (van der Oost et al., 2014). As such, CRISPR/Cas9 is perfectly suited for genome editing, with the possibility to simultaneously target several loci by co-expressing or even concatenating multiple gRNAs, which can be easily cloned due to their small size (Hsu et al., 2014). In this regard, CRISPR/Cas9 genome editing is simpler to implement than TALE-based editing, because for each locus to be targeted, two new dTALEs need to be designed and expressed (de Lange et al., 2014). The CRISPR/Cas system was also modified to generate TFs (CRISPR/Cas-TF), whose specificity is determined by gRNAs (Nissim et al., 2014). Typically, gRNAs are expressed from RNA pol III promoters which are constitutive, and therefore provide no flexibility for tissue or condition specific expression. One solution to this problem is to express the gRNAs by RNA pol II promoters and edit the RNA by the Csy4 enzyme to ensure processing of the gRNA in the nucleus (Qi et al., 2012), whose functionality in plants remained to be shown. In addition to this added complexity, the fact that the specificity of the CRISPR/Cas-TF is determined at the RNA level, and not at the protein level as with the TALEs, constitutes a limitation for the design of more complex regulatory networks with activating or repressing steps. Because all CRISPR/Cas proteins will recognize the gRNAs and their target regardless of whether there are modified to be transcriptional activators or repressors, co-expression of both types of CRISPR/Cas-TFs in a single cell would result in unpredictable expression of the target genes. Also, a single CRISPR/Cas-TF protein controlling the whole network cannot act simultaneously as a repressor and activator on different targets. Thus, to introduce a negative regulatory step in a CRISPR/Cas regulatory network, Nissim et al. had recourse to a miRNA (Nissim et al.,

2014). In contrast, distinct dTALES with either repressor or activating properties can be co-expressed, thus allowing much more flexibility and possibilities for the design of regulatory networks (de Lange et al., 2014).

These STAPs can be used for metabolic engineering of complex biosynthetic pathways requiring the simultaneous and tunable expression of multiple genes. This is particularly relevant in plants, where specific tissues or conditions can then be tested and compared for their performance in producing specific classes of compounds. Furthermore, these STAPs provide the basis for the design and construction of complex regulatory networks that contain transcriptional cascades, positive and negative feedback loops. Thus, their potential reaches beyond the field of metabolic engineering.

Materials and Methods

Construction of a library of synthetic TALE-activated promoters (STAPs)

To generate a library of promoter sequences as described in Figure 1, two degenerate primers, Talpro7 and Talpro8 (see sequences below), containing the TALE- DNA binding site (effector binding element 2 or EBE-002, described in (Weber et al., 2011) as the binding site for dTALE-2) and degenerate sequences on either side of the DNA binding site were made and annealed at a concentration of 0.2 µM each in a 50 µL reaction mix with the KOD polymerase (Takara). Primer extension was carried out by first denaturing for 5 min at 94°C, then annealing at 60°C for 30 s and extending at 72°C for 30 s. The product was then purified on a Qiaquick column (Qiagen) and inserted into vector pAGM1311 (universal level -1 cloning vector) (Engler et al., 2014) using a standard Golden Gate cloning procedure with BsaI and T4 DNA ligase (Werner et al., 2012). An aliquot (100 µL) of the transformed bacteria were spread over two Petri dishes for evaluation of the complexity of the library, and the remainder of the transformation inoculated in liquid culture for library preparation. DNA from the whole library was then prepared and further subcloned into level 0 vector

pICH41295 (Engler et al., 2014) using BpiI and T4 DNA ligase. The strategy of cloning the library in two steps (level -1 and 0) results in a higher amount of promoters with a correct structure compatible with Golden Gate cloning. In particular, the two successive cloning steps that use BsaI and then BpiI for cloning result in elimination of promoters that may have contained BpiI or BsaI sites in internal random sequences. The resulting level 0 constructs (pAGT582-n) were then picked individually, checked for the presence of single promoter sequences after digestion with BsaI, sequenced and used for further cloning. The sequences of the 43 promoters are provided as supplemental data.Talpro7: 5'-ttt ggtctc a acat GGAG $(n)_{19}$ tccccgcatagctgaacatc. Talpro8: 5'-ttt ggtctc a acaa CATT $(n)_{43}$ ttatata g atgttcagctatgcgggg (the capital letters indicate sequences that will become part of the BsaI cleavage sites in the final level 0 modules).

Reporter constructs

The STAPs coming from the pAGT582-n vectors were cloned either in front of a GUS reporter gene (plasmid pICH75111) (Engler et al., 2014) and the ocs terminator (pICH41432) (Engler et al., 2014) to give pAGT615-n vectors, or in front of the GFP reporter gene (pICH4153) (Engler et al., 2014) and the ocs terminator to give the pAGT917-n vectors. The constructs were made in T-DNA vector pICH75044 (Engler et al., 2014). All clonings were made using standard Golden Gate cloning procedure with the appropriate restriction enzymes (Engler et al., 2008).

GUS Transient assays

All T-DNA vectors were transformed into *Agrobacterium tumefaciens* strain GV3101(pMP90) (Koncz and Schell, 1986). *N. benthamiana* plants were grown in a greenhouse maintained at a constant temperature of 22°C with 82% humidity. For the GUS assays, the pAGT615-n constructs were co-infiltrated with the TALE containing the DNA binding domain specific for the EBE002 sequence (construct pICH74043, (Weber et al., 2011)) and a 35S:GFP construct as internal reference (construct pAGM4731, (Engler et al., 2014)). As control, the

pAGT615-n constructs were also infiltrated without the TALE. A construct with 35S:GUS (pICH75181, (Engler et al., 2014)) co-infiltrated with 35S:GFP was used as a positive control and external reference. We infiltrated the reporter constructs with and without the TALE and a positive control (35S:GUS) on different sectors of the same leaf to eliminate leaf-to-leaf variation. Agrobacterium strains were grown in LB medium, resuspended in infiltration medium (10 mM MES, 10 mM $MgSO_4$) at a final OD_{600} of 0.3 for infiltration. The plants were then placed back in the greenhouse for five days before analysis. For quantitative GUS assays, leaf discs of 0.9 cm diameter within the infiltrated areas were excised and transferred to microtubes containing three 0.2 mm steel beads and immediately frozen in liquid N_2. Frozen leaf tissues were then homogenized twice using a Retsch mixer mill MM400 at a frequency of 30 Hz for 1 min. The leaf powder was then resuspended in 1 mL phosphate buffered saline (pH 7.4), vortexed and filtered through a 0.2 µM PTFE filter on a 96-well plate (Chromafil, Macherey-Nagel, Germany). GUS activity was measured using the fluorogenic 4-methylumbelliferyl β-D-glucopyranoside (MUG) substrate using published procedures at 10, 30 and 60 min (Jefferson et al., 1987). In parallel, 200 µL of the extracts was used to measure GFP fluorescence. The GFP fluorescence of each sample was used as normalization factor. The strength of the individual promoters was then expressed as a percentage of the expression with the 35S promoter, based on the control from the same leaf.

GFP assays

Selected promoters from the pAGT582-n series were cloned into a level 1 vector (pICH47732) in front of GFP (vector pICH41531, (Engler et al., 2014)) and the ocs terminator (pICH41432). The constructs were introduced into Agrobacterium GV3101(pMP90) and expressed in *N. benthamiana* with or without the TALE-EBE002 (see above). The 35S:GFP construct (pAGM4731, see above) was used as positive control. The plants were incubated for three to five days and photos were taken under a UV lamp.

Diterpene metabolic engineering constructs

The coding sequences of the *SlDXS2* (pAGT345), *NtGGPPS2* (pAGT169) and *CBTS2a* (pAGT238) genes were those described in (Brückner and Tissier, 2013). They were cloned behind the promoters from pAGT582-3, pAGT582-5 and pAGT582-1 respectively. The terminators used were Sl vATPase1 (pICH 71431), Ocs (pICH 41432), and CBTS (pAGT168) respectively (Brückner and Tissier, 2013, Engler et al., 2014). These fragments were assembled into level 1 vectors pICH47751, pICH47742 and pICH47802 respectively. The vectors were transformed into *Agrobacterium tumefaciens* GV3101:pMP90 and used for *N. benthamiana* leaf infiltration as described above. The corresponding constructs with the 35S promoter were as described in (Brückner and Tissier, 2013). Metabolite extraction and diterpene GC-MS measurements were performed as described in (Brückner and Tissier, 2013) with the following modifications. The 1 mL hexane extracts from the 6 leaf discs per sample were not evaporated and 160µL of these extracts were combined with 40µL of internal standard (IS) (50µM Sclareol stock solution in hexane), mixed, centrifuged and transferred to a GC-vial. Samples were measured on a QP2010 Ultra GC-MS system (Shimadzu) in the following conditions. 1 µL was injected splitless at 250°C with a 1 min sampling time in a RXi-5Sil MS 30 m, 0.25 mm diameter column (Restek). Elution was done with a constant He flow of 0.9 mL/min. The oven temperature was raised from 50°C to 300°C (7°C/min), then to 320°C (50°C/min) and held for 2 min. The interface and the ion source temperatures were 300°C and 250°C respectively and the ionization voltage was set at 70 V.

Quantitative gene expression analysis in infiltrated leaf tissue

Total RNA was extracted from infiltrated leaf discs from which diterpenes were extracted by hexane wash before. Leaf discs were homogenized to a fine powder using a Mixer Mill MM-400 (Retsch) and RNA was extracted with the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich) according to the manufacturer's specifications. Any genomic contamination was removed before cDNA synthesis using the DNA-free Kit (Ambion) according to manufacturer instructions. The first strand of cDNA was

synthesized from 1 µg total RNA using Maxima H Minus First Strand cDNA Synthesis Kit (Thermo Scientific) and a mix of oligo(dT)18 and random hexamer primers. The cDNA was diluted 10-fold and 1 µl was used in a 10 µl reaction mix containing 1 µl (2 pmol/µl) of both forward and reverse primers and 2 µl my-Budget 5x EvaGreen QPCR Mix II (Bio&Sell). Quantitative real-time PCR was performed in 96-well plates using the CFX Connect real-time PCR detection system (Bio-Rad). The thermal profile of the reaction was as follows: an initial denaturation at 95°C for 15 min, followed by 40 cycles at 95°C for 15 seconds and 58°C for 30 seconds with fluorescence acquisition after each cycle. To verify primer specificity, a dissociation curve was finally generated by increasing the temperature from 65 to 95°C. qPCR primers for the amplification of *CBTS2a*, *DXS2* and *GGPPS2* genes were designed using the Primer3 program (Untergrasser et al., 2012). The *N. benthamiana* eukaryotic elongation factor 1a was selected as endogenous reference for normalization of transcript levels, and PCR amplified using primers as described in Liu et al., 2012. To check the specificity of the *CBTS2a*, *DXS2* and *GGPPS2* primers, qPCR was performed also on *N. benthamiana* cDNA and no amplification could be detected in any of the control samples (Figure 5). The sequence of the oligonucleotides used for real-time PCR are provided in the supplemental data. All the cDNA samples were amplified in three technical replicates from the same RNA preparation and the mean value was considered. Expression of transgenes in infiltrated leaf tissue was calculated using the $2^{(\text{delta-Ct})}$ method (Livak and Schmittgen, 2001).

Acknowledgements

KB is supported by DFG grant n° TI 800/2-1 to AT from the ERA-CAPS programme (http://www.eracaps.org/). This work was supported in part by the core funding of the Leibniz-Institute of Plant Biochemistry, Halle, Germany.

Conflict of interest

The authors declare that they have a conflict of interest. A patent application was filed. The promoters will be made available free of charge for non-commercial research purposes.

Short Legend of supporting information

Supplemental data 1. Sequences of the fourty-three STAPs.

Supplemental data 2. Sequences of the primers used for quantitative gene expression analysis

Supplemental Figure.1 Correlation between gene expression and CBTol levels

References

Blazeck, J. and Alper, H.S. (2013) Promoter engineering: recent advances in controlling transcription at the most fundamental level. *Biotechnol J*, **8**, 46-58.

Blount, B.A., Weenink, T., Vasylechko, S. and Ellis, T. (2012) Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. *PLoS One*, **7**, e33279.

Boch, J. and Bonas, U. (2010) Xanthomonas AvrBs3 Family-Type III Effectors: Discovery and Function. *Annu. Rev. Phytopathol.*, **48**, 419-436.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509-1512.

Brophy, J.A.N. and Voigt, C.A. (2014) Principles of genetic circuit design. *Nat. Methods*, **11**, 508-520.

Brückner, K. and Tissier, A. (2013) High-level diterpene production by transient expression in Nicotiana benthamiana. *Plant Methods*, **9**, 46.

de Lange, O., Binder, A. and Lahaye, T. (2014) From dead leaf, to new life: TAL effectors as tools for synthetic biology. *The Plant Journal*, **78**, 753-771.

Engler, C., Kandzia, R. and Marillonnet, S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One*, **3**, e3647.

Engler, C., Youles, M., Gruetzner, R., Ehnert, T.M., Werner, S., Jones, J.D., Patron, N.J. and Marillonnet, S. (2014) A Golden Gate Modular Cloning Toolbox for Plants. *ACS synthetic biology*, **Epub ahead of print**.

Ennajdaoui, H., Vachon, G., Giacalone, C., Besse, I., Sallaud, C., Herzog, M. and Tissier, A. (2010) Trichome specific expression of the tobacco (Nicotiana sylvestris) cembratrien-ol synthase genes is controlled by both activating and repressing cis-regions. *Plant Mol. Biol.*, **73**, 673-685.

Fineran, P.C. and Dy, R.L. (2014) Gene regulation by engineered CRISPR-Cas systems. *Curr. Opin. Microbiol.*, **18**, 83-89.

Hahn, S. and Young, E.T. (2011) Transcriptional Regulation in Saccharomyces cerevisiae: Transcription Factor Regulation and Function, Mechanisms of Initiation, and Roles of Activators and Coactivators. *Genetics*, **189**, 705-736.

Hsu, P.D., Lander, E.S. and Zhang, F. (2014) Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*, **157**, 1262-1278.

Jefferson, R.A., Kavanagh, T.A. and Bevan, M.W. (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *Embo j*, **6**, 3901-3907.

Kay, S., Hahn, S., Marois, E., Hause, G. and Bonas, U. (2007) A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science*, **318**, 648-651.

Kay, S., Hahn, S., Marois, E., Wieduwild, R. and Bonas, U. (2009) Detailed analysis of the DNA recognition motifs of the Xanthomonas type III effectors AvrBs3 and AvrBs3Δrep16. *The Plant Journal*, **59**, 859-871.

Koncz, C. and Schell, J. (1986) The Promoter of Tl-DNA Gene 5 Controls the Tissue-Specific Expression of Chimeric Genes Carried by a Novel Type of Agrobacterium Binary Vector. *Mol. Gen. Genet.*, **204**, 383-396.

Liu, D., Shi, L., Han, C., Yu, J., Li, D., Zhang, Y. (2012) Validation of reference genes for gene expression studies in virus-infected *Nicotiana benthamiana* using quantitative real-time PCR. *PLoS One*, **7**, e46451.

Liu, Q., Manzano, D., Tanic, N., Pesic, M., Bankovic, J., Pateraki, I., Ricard, L., Ferrer, A., de Vos, R., van de Krol, S. and Bouwmeester, H. (2014) Elucidation and in planta reconstitution of the parthenolide biosynthetic pathway. *Metab Eng*, **23**, 145-153.

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

Lubliner, S., Keren, L. and Segal, E. (2013) Sequence features of yeast and human core promoters that are predictive of maximal promoter activity. *Nucleic Acids Res.*, **41**, 5569-5581.

Mahfouz, M., Li, L., Piatek, M., Fang, X., Mansour, H., Bangarusamy, D. and Zhu, J.-K. (2012) Targeted transcriptional repression using a chimeric TALE-SRDX repressor protein. *Plant Mol. Biol.*, **78**, 311-321.

Maicas, E. and Friesen, J.D. (1990) A sequence pattern that occurs at the transcription initiation region of yeast RNA polymerase II Promoters. *Nucleic Acids Res.*, **18**, 3387-3393.

Nissim, L., Perli, S.D., Fridkin, A., Perez-Pinera, P. and Lu, T.K. (2014) Multiplexed and Programmable Regulation of Gene Networks with an Integrated RNA and CRISPR/Cas Toolkit in Human Cells. *Mol. Cell*, **54**, 698-710.

Peremarti, A., Twyman, R.M., Gomez-Galera, S., Naqvi, S., Farre, G., Sabalza, M., Miralpeix, B., Dashevskaya, S., Yuan, D.W., Ramessar, K., Christou, P., Zhu, C.F., Bassie, L. and Capell, T. (2010) Promoter diversity in multigene transformation. *Plant Mol. Biol.*, **73**, 363-378.

Perez-Pinera, P., Ousterout, D.G., Brunger, J.M., Farin, A.M., Glass, K.A., Guilak, F., Crawford, G.E., Hartemink, A.J. and Gersbach, C.A. (2013) Synergistic and tunable human gene activation by combinations of synthetic transcription factors. *Nat Meth*, **10**, 239-242.

Potenza, C., Aleman, L. and Sengupta-Gopalan, C. (2004) Targeting transgene expression in research, agricultural, and environmental applications: Promoters used in plant transformation. *In Vitro Cell Dev-Pl*, **40**, 1-22.

Qi, L., Haurwitz, R.E., Shao, W.J., Doudna, J.A. and Arkin, A.P. (2012) RNA processing enables predictable programming of gene expression. *Nat. Biotechnol.*, **30**, 1002-1006.

Rojas-Duran, M.F. and Gilbert, W.V. (2012) Alternative transcription start site selection leads to large differences in translation activity in yeast. *RNA*, **18**, 2299-2305.

Romer, P., Hahn, S., Jordan, T., Strauss, T., Bonas, U. and Lahaye, T. (2007) Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. *Science*, **318**, 645-648.

Sleat, D.E., Gallie, D.R., Jefferson, R.A., Bevan, M.W., Turner, P.C. and Wilson, T.M.A. (1987) Characterization of the 5'-leader sequence of tobacco mosaic virus RNA as a general enhancer of translation in vitro. *Gene*, **60**, 217-225.

Ting, H.M., Wang, B., Rydén, A.M., Woittiez, L., van Herpen, T., Verstappen, F.W., Ruyter-Spira, C., Beekwilder, J., Bouwmeester, H.J. and van der Krol, A. (2013) The metabolite chemotype of Nicotiana benthamiana transiently expressing artemisinin biosynthetic pathway genes is a function of CYP71AV1 type and relative gene dosage. *New Phytol.*, **199**, 352-366.

Tissier, A. (2012) Trichome specific expression: promoters and their applications. In *Transgenic Plants, advances and limitations* (Çiftçi, Y.O. ed: InTech, pp. 353-378.

Untergrasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G. (2012) Primer3 – new capabilities and interfaces. *Nucleic Acids Research*, **50**, e115.

van der Oost, J., Westra, E.R., Jackson, R.N. and Wiedenheft, B. (2014) Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nature Reviews Microbiology*, **12**, 479-492.

Vranova, E., Coman, D. and Gruissem, W. (2012) Structure and Dynamics of the Isoprenoid Pathway Network. *Mol. Plant.*, **5**, 318-333.

Wang, E. and Wagner, G.J. (2003) Elucidation of the functions of genes central to diterpene metabolism in tobacco trichomes using posttranscriptional gene silencing. *Planta*, **216**, 686-691.

Weber, E., Gruetzner, R., Werner, S., Engler, C. and Marillonnet, S. (2011) Assembly of Designer TAL Effectors by Golden Gate Cloning. *PLoS One*, **6**, e19722.

Werner, S., Engler, C., Weber, E., Gruetzner, R. and Marillonnet, S. (2012) Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. *Bioeng Bugs*, **3**, 38-43.

Figure legends

Figure 1. **Design and cloning procedure of the STAPs.** A. general sequence of the STAPS. The constant dTALE binding site (EBE002) is highlighted in blue and the TATA box in green. The atg triplet at the end of the promoter corresponds to the translation start codon. The degenerate sequences upstream and downstream of the constant region are indicated by $(n)_{19}$ and $(n)_{43}$ respectively. B. Cloning procedure of the STAPS. The library was synthesized using two degenerate oligonucleotides overlapping over the constant region and cloned into a level -1 Golden Gate vector. Subsequent cloning into a level 0 vector allowed elimination of aberrant products (for example dimers) and false positives. The level 0 vectors can then be used for assembly into transcription units.

Figure 2. Testing the STAPs in a transient assay with a GUS reporter gene. A. Overview of the constructs used for the transient assay. The dTALE, which binds to the constant EBE002 sequence of the STAPs, is under the control of the constitutive and moderate *Act2* promoter. The STAPs control the expression of a GUS transgene with introns (represented by white bars within the blue coding sequence) to prevent expression in Agrobacterium (which is used for the infiltration of *N. benthamiana* leaves). B. Distribution of the GUS activity values for the STAPs promoters with or without the dTALE, expressed in percentage of 35S:GUS expression.

Figure 3. Testing selected STAPs with GFP. Eight different STAPs displaying different levels of expression with the GUS reporter gene were tested with GFP. *N. benthamiana* leaves were infiltrated with Agrobacterium strains carrying different STAP:GFP constructs with (left side) or without (right side) the dTALE. Pictures of the leaves were taken in normal light (first and third column) or under UV illumination (second and fourth column) to visualize GFP.

Figure 4. Using STAPs for metabolic engineering of a plant diterpene. A. overview of the pathway for cembratrienols (CBTols) including the methylerythritol-phosphate (MEP) pathway, highlighting the genes that were over-expressed in *N. benthamiana*. Pyr, pyruvate; GA3P, glyceraldehyde-3-phosphate; DXP, 1-deoxyxylulose 5-phosphate; IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; GGPP, geranylgeranyl diphosphate; SlDXS2, tomato 1-deoxyxylulose 5-phosphate synthase; NtGGPPS2, tobacco GGPP synthase; CBTS2a, tobacco cembratrienol synthase. B. Overview of the constructs used for the overexpression of SlDXS2, NtGGPPS2 and CBTS2a using STAPs. The numbers of the STAPs correspond to the numbers in Figures 2 and 3. C. Comparison of TALE-driven expression and p35S-driven expression in *N. benthamiana* leaves agro-infiltrated with p19 and CBTS2a, DXS2 and GGPPS2 in different combinations. Two leaves in three individual plants were infiltrated for each construct combination (n=6). Mean CBTol amounts are expressed as equivalents of the internal standard (IS, sclareol) ± standard error. Groups indicated by different letters (a1/a2, and b1/b2) differ significantly from each other regarding their CBTol values ($P<0.05$, Student's t-test and two-way ANOVA). No CBTol was detected in leaves expressing the empty T-DNA vector as a control.

Figure 5. TALE- and 35S-driven expression of the CBTS2a, DXS2 and GGPPS2 genes in agro-infiltrated *N. benthamiana* leaves. Transcript levels were quantified by RT-PCR in three selected biological replicates from samples used in Figure 4 (rep1-3) from agro-infiltrated leaves co-expressing p19, CBTS2a, DXS2 and GGPPS2 driven by STAPS (TALE) or the 35S promoter (35S). Mean expression values of three technical replicates ± SD and the corresponding CBTol amounts for each sample are given.

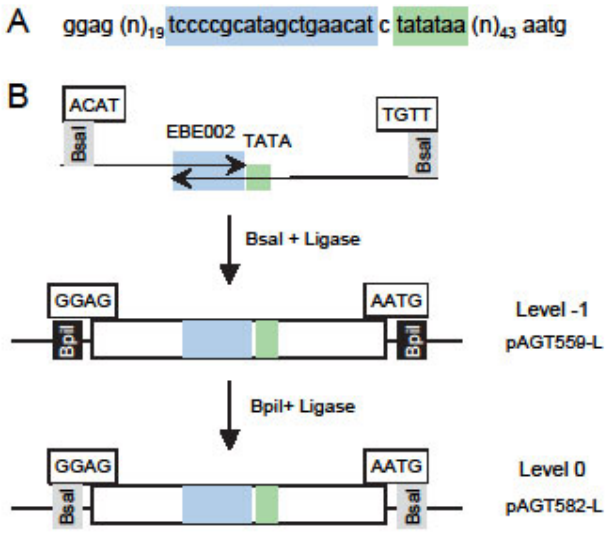

Figure 1. Design and cloning procedure of the STAPs. A. general sequence of the STAPS. The constant dTALE binding site (EBE002) is highlighted in blue and the TATA box in green. The atg triplet at the end of the promoter corresponds to the translation start codon. The degenerate sequences upstream and downstream of the constant region are indicated by $(n)_{19}$ and $(n)_{43}$ respectively. B. Cloning procedure of the STAPS. The library was synthesized using two degenerate oligonucleotides overlapping over the constant region and cloned into a level -1 Golden Gate vector. Subsequent cloning into a level 0 vector allowed elimination of aberrant products (for example dimers) and false positives. The level 0 vectors can then be used for assembly into transcription units.

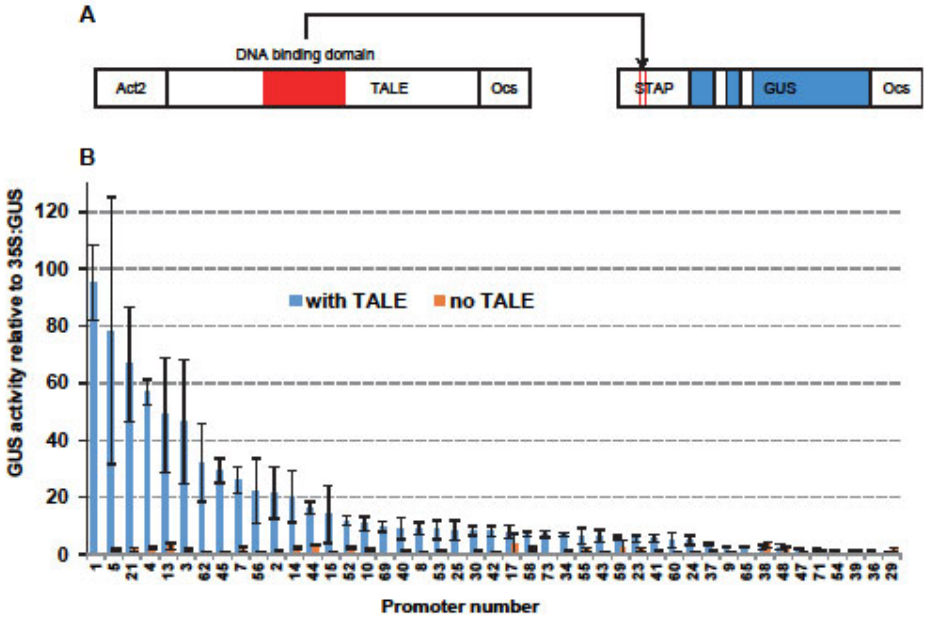

Figure 2. Testing the STAPs in a transient assay with a GUS reporter gene. A. Overview of the constructs used for the transient assay. The dTALE, which binds to the constant EBE002 sequence of the STAPs, is under the control of the constitutive and moderate Act2 promoter. The STAPs control the expression of a GUS transgene with introns (represented by white bars within the blue coding sequence) to prevent expression in Agrobacterium (which is used for the infiltration of *N. benthamiana* leaves). B. Distribution of GUS activity values for the STAPs promoters with or without the dTALE, expressed in percentage of the 35S:GUS expression.

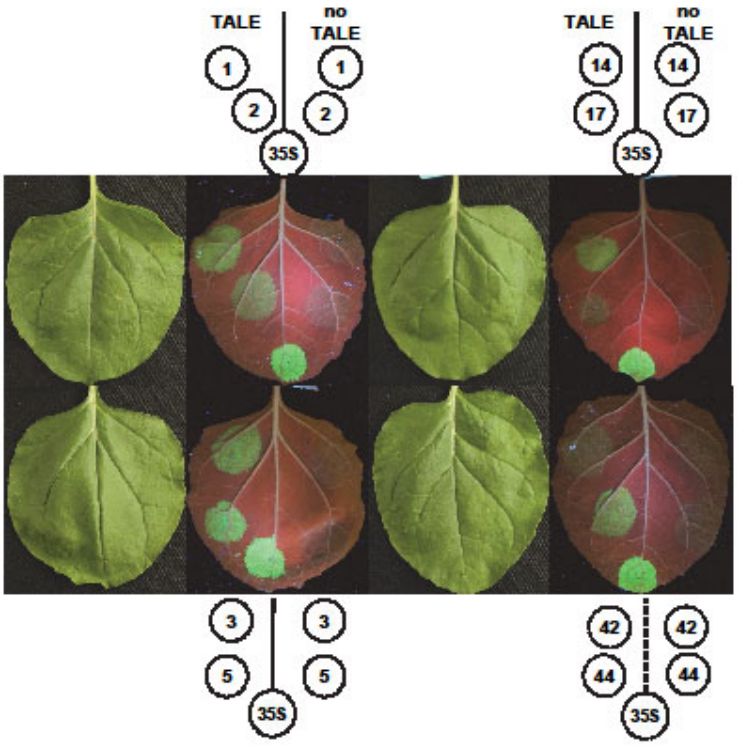

Figure 3. Testing selected STAPs with GFP. Eight different STAPs displaying different levels of expression with the GUS reporter gene were tested with GFP. *N. benthamiana* leaves were infiltrated with Agrobacterium strains carrying different STAP:GFP constructs with (left side) or without (right side) the dTALE. Pictures of the leaves were taken in normal light (first and third column) or under UV illumination (second and fourth column) to visualize GFP.

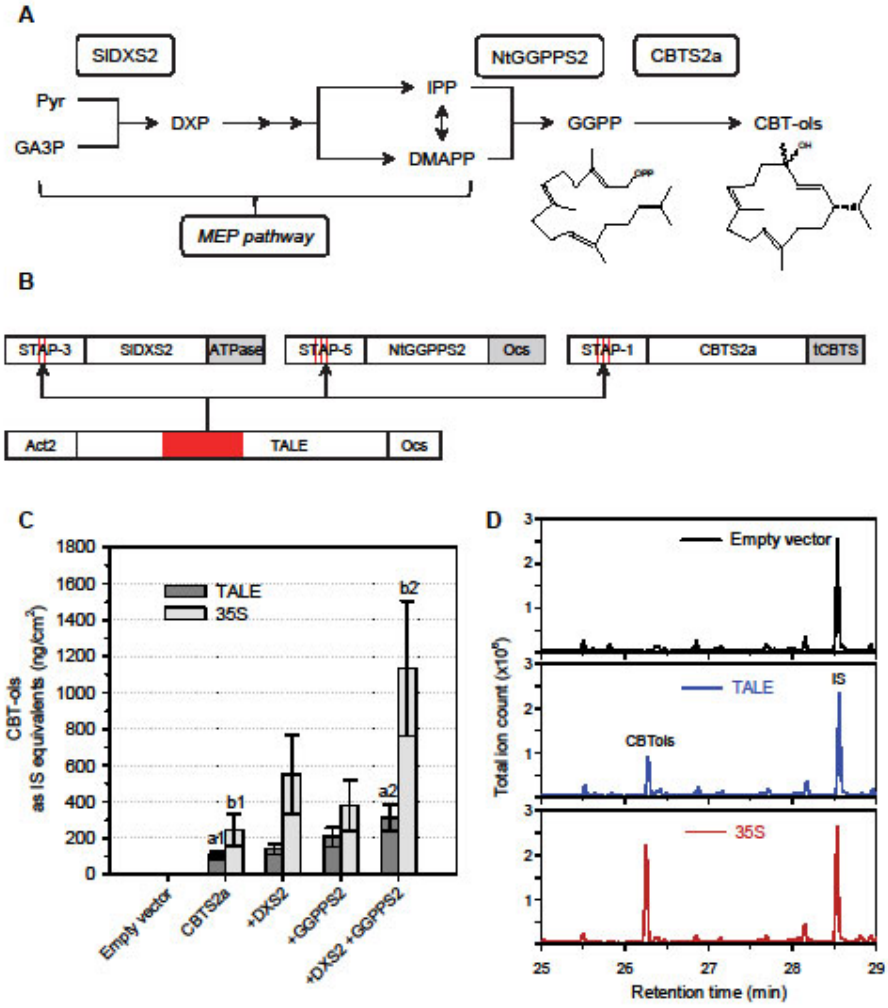

Figure 4. Using the STAPs for metabolic engineering of a plant diterpene. A. overview of the pathway for cembratrienols (CBTol) including the methylerythritol-phosphate (MEP) pathway, highlighting the genes that were over-expressed in *N. benthamiana*. Pyr, pyruvate; GA3P, glyceraldehyde-3-phosphate; DXP, 1-deoxyxylulose 5-phosphate; IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; GGPP, geranylgeranyl diphosphate; SlDXS2, tomato 1-deoxyxylulose 5-phosphate synthase; NtGGPPS2, tobacco GGPP synthase; CBTS2a, tobacco cembratrien-ol synthase. B. Overview of the constructs used for the overexpression of SlDXS2, NtGGPPS2 and CBTS2a using STAPs. The numbers of the STAPs correspond to the numbers in Figures 2 and 3. C. Comparison of TALE-driven expression (number of biological replicates= 5) and p35S driven expression (n=6) in *N. benthamiana* leaves agro-infiltrated with p19 and CBTS2a, DXS2 and GGPPS2 in different combinations. Mean CBTol amounts are expressed as equivalents of the internal standard (IS, sclareol) ± standard error. Groups indicated by different letters (a1/a2, and b1/b2) differ significantly from each other regarding their CBTol values ($P<0.05$, Student's t-test and two-way ANOVA). No CBTol was detected in leaves expressing the empty T-DNA vector as a control. D. Representative GC/MS chromatograms of surface extracts from leaves co-expressing all genes (*CBTS2a*, *GGPPS2*, and *DXS2*) activated by the TALE (blue), or under the 35S promoter (red) and an empty vector control (black).

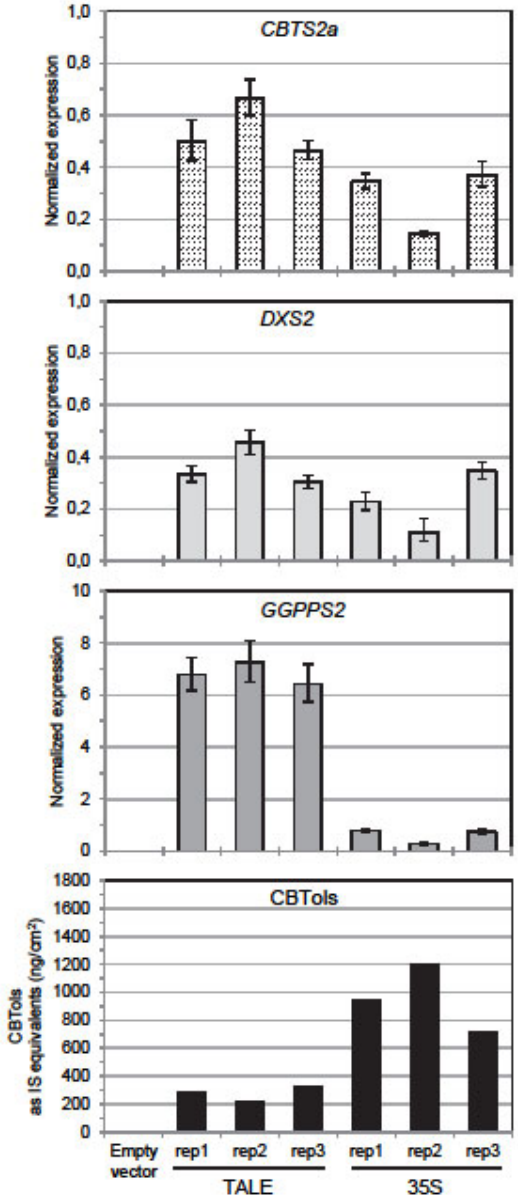

Figure 5. TALE- and 35S-driven expression of the *CBTS2a*, *DXS2* and *GGPPS2* genes in agro-infiltrated *N. benthamiana* leaves. Transcript levels were quantified by RT-PCR in three selected biological replicates from samples used in Figure 4 (rep1-3) from agro-infiltrated leaves co-expressing p19, CBTS2a, DXS2 and GGPPS2 driven by STAPS (TALE) or the 35S promoter (35S). Mean expression values of three technical replicates ± SD and the corresponding CBTol amounts for each sample are given.