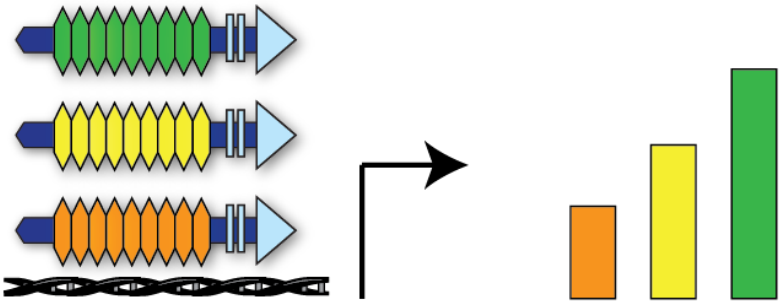
Tunable synthetic transcription factors assembled from TALE-like repeat elements

Authors: Orlando de Lange*, Niklas Schandry*, Markus Wunderlich, Kenneth Berendzen & Thomas Lahaye

*Joint first authors

Graphical abstract:



Abstract

A protein-DNA interface can be engineered in ways that an RNA-DNA interface cannot. Programmable TAL effector derived proteins bind DNA with a series of repeats. Each repeat binds a single DNA base. Repeats can be treated as modules, with a pair of residues in the center defining the target base. A set of 31-33 residues flanking the base specifying residues affect other DNA binding parameters. Efforts at engineering TAL effector DNA binding repeats have till now focused on the base specifying residues only. Here we show that natural repeat sequence diversity can be used to alter DNA binding strength. We generated sets of chimeric repeat arrays through a random assembly approach. These sequence diverse repeat proteins activate or repress promoters with a range of activities. Our design leaves the choice of base binding residues for each repeat open. This allows users to tune binding strength without altering sequence preference.

Introduction

(Limit ~750 words; currently only 500 words so plenty of extra room if needed)

Synthetic transcription factors (TFs) are tools to control gene expression. They are composed of a DNA binding domain combined with a transcriptional regulation domain. Fixed sequence DNA binding domains (DBDs), such as Gal4 and LexA, have been used extensively, but lack flexibility. The recent advent of programmable DNA binding proteins is now allowing researchers to create DBDs for any sequence of interest. Freed from the constraint of a fixed binding sequence can easily target synthetic TFs to endogenous promoters or regulate diverse promoter sequences in the context of synthetic genetic circuits. Synthetic transcription factors with programmable DBDs are being rapidly adopted as tools in fundamental and applied molecular biology[1].

Two commomly used forms of programmable DNA binding proteins are CRISPR-Cas9 and TALEs (CRISPR, Clustered Regularly Interspersed repeats; TALE, Transcription activator like effector). Both are bacterial in origin, and both are involved in disease, though whilst natural CRISPR-Cas9 systems form an adaptive immune system to defend bacteria against viruses[2], while natural TALEs are bacterial weapons injected into host plant cells to alter transcription and promote pathogen growth[3]. The CRISPR-Cas9 system is bipartite, with a fixed protein chassis (Cas9) guided to specific DNA sequences by a guide RNA molecule. DNA binding of CRISPR-Cas systems is based on RNA-DNA interactions. Thus base preference and binding energy of each pair are innately linked, limiting the range of DNA binding parameters that can be engineered. The same principle and thus the same limitation applies to other nucleic acid guided DNA-binding proteins Cpf1[4] and NgAgo[5].

TALE proteins bind DNA directly via arrays of 33-35 amino-acid repeats, which collectively form a super-helical structure. Positions 12 and 13 of each repeat, termed the Repeat Variable Diresidue (RVD[6]) make base-specific interactions. Other repeat positions are involved in stabilizing the structure of the array or non-base-specific DNA binding[7,8]. Multiple cloning approaches have been developed to assemble TALE repeat arrays with user defined RVD compositions and therefore user-defined target sequences[9,10]. When fused to a functional domain of interest such artificial TALE repeat arrays are referred to as dTALEs (designer-TALEs), and are used within the field of synthetic biology[11] as well as fundamental research[12] and have recently been demonstrated to have advantages over CRISPR-Cas9 based transcriptional regulators for the generation of genetic logic gates[13] in eukaryotic cells.

The protein-DNA interface of TALE-repeats provides opportunities to modulate their base affinity and/or specificity. However, the engineering potential of TALE repeats has been little explored beyond the RVD positions. The dTALE design approach views the RVDs as modular units within an otherwise fixed scaffold, and in practice dTALEs are assembled from a limited RVD-pool of the four most commonly occurring RVDs in natural TALEs (NI, HD,NG,NN[14]). There has been considerable interested in characterizing uncommon or completely novel RVDs, culminating in the publication of the activities of dTALE repeats bearing the complete set of 400 possible RVDs[15]. This study takes a different approach, exploring the potential inherent in TALE repeat sequence diversity beyond the RVDs. We believe this diversity can be used to create tunable, but synonymous programmable TFs. Synonymous, in this context, means that they recognize the same DNA sequences. Tunable means that the DNA-binding properties of each TF can be modified to achieve a range of activities at the target promoter. We have created a set of Variable Sequence TALEs (VarSeTALEs) with a conserved RVD composition but considerable diversity at non-RVD positions and demonstrated that they execute a range of activation or repression levels at promoters in a set of reporter assay.

We drew on natural sequence diversity to create the VarSeTALEs in this study.

In nature, TALE-like proteins are found in at least three bacterial genera[16] and polymorphisms are found across repeats at every position (Figure S1). We used the natural pool of sequence-diverse TALE repeats to assemble a novel set of synonymous TALE-TFs with different activities.

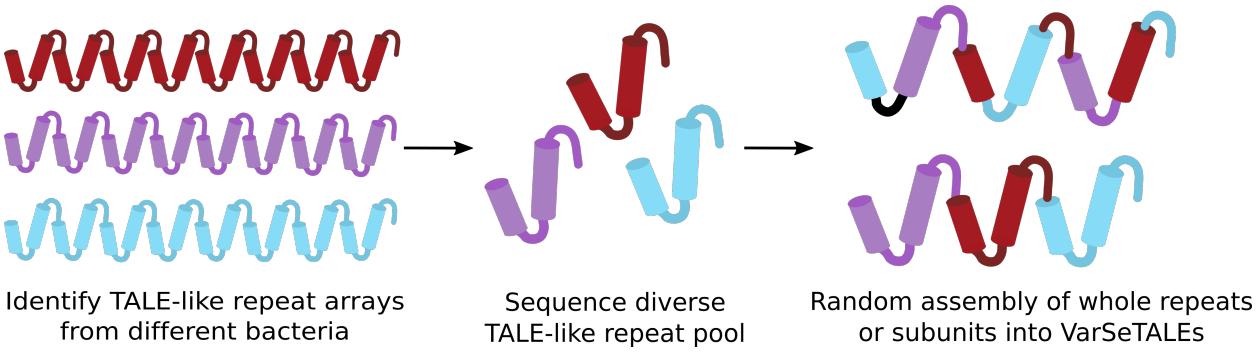
A core assumption of this approach is that polymorphisms in non-RVD positions have only a very minor impact on base preference, but do modify DNA binding strength, as measured through promoter activity. This is supported by previous work carried out by authors of this study [16–18].

This is the first report on the use of natural TALE-like sequence diversity to tune activities of synonymous TALE repeat arrays.

Results & discussion

For this work we used sequences from previously characterized TALE-like proteins encoded in the genomes of bacterial clades *Ralstonia solanacearum*[19] and *Burkholderia rhizoxinica*[18]. TALE-likes all share a common DNA binding code, with the same code linking RVDs to target bases. However, TALE-like repeats differ considerably at the sequence level (Figure S1). We also drew on the full diversity of *Xanthomonas* TALEs, which is generally not utilized (most dTALEs previously published are derived from two TALEs: AvrBs3[9] and Hax3[20]). Together the repeat sequences from across the *Xanthomonas* TALEs and the TALE-likes were used to assemble our chimeras, termed VarSeTALEs (sequences are listed in Table S1). The workflow that we followed is outlined in figure 1.

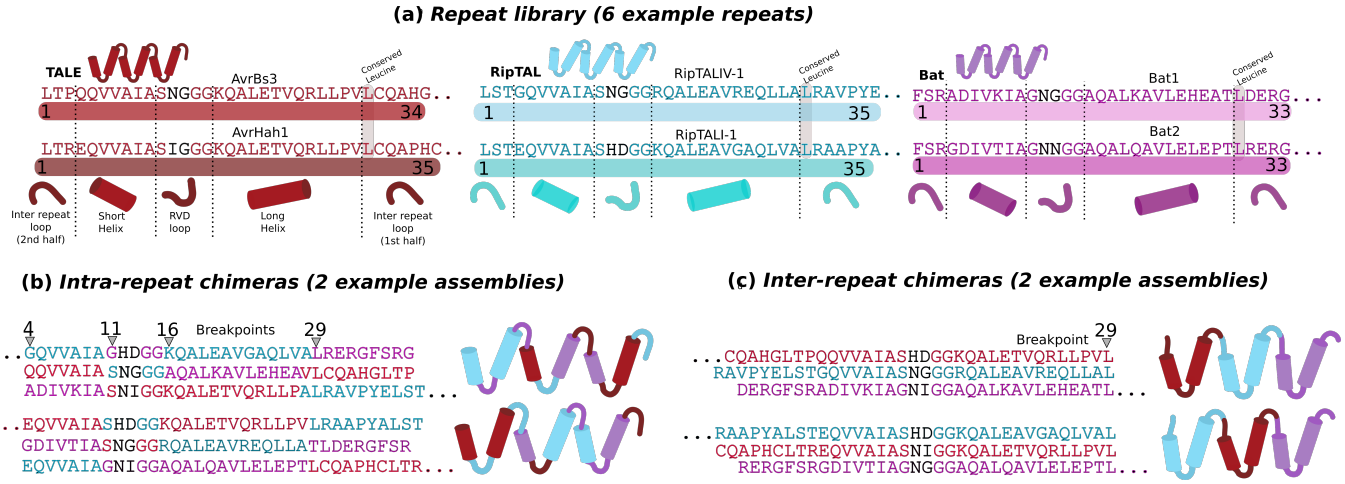
Figure 1: Workflow pursued in this study



The first step was to assemble sequences of TALE-likes from *Xanthomonas*, *Ralstonia* and *Burkholderia* bacteria (red, blue and magenta cylinders and loops respecitively; cylinders and loops reflect the paired helix-loop structure of TALE repeats). Sequences of individual repeats and sub-repeat elements were then compiled. From these we used a random assembly approach to create sequences of novel TALE-like chimeras termed VarSeTALEs. In the 'Intra-repeat' approach (top right) these secondary structural units were used as the basic sequence units for assembly. Whole repeats were the basis of 'inter-repeat' chimeras (bottom-right).

We explored two alternative approaches for VarSeTALE design: Intra-repeat and inter-repeat chimeras. That is to say either replacing whole TALE repeats (inter-repeat) or repeat subunits (intra-repeats). Repeat subunits used in our intra-repeat chimeras correspond to secondary structural elements (short-helix, RVD loop, long-helix and inter-repeat loop). For inter-repeat chimeras the highly conserved leucine residue at position 29 was used as the breakpoint between repeats of different origins. Figure 2 illustrates the two design approaches using example sequences.

Figure 2: VarSeTALE design



The starting material is an *in-silico* repeat library of non-identical repeat sequences (a). We stored sequences based on bacterial origin (TALEs of *Xanthomonas*, RipTALs of *Ralstonia* and Bats of *Burkholderia*). Colour coding reflects these groupings throughout this figure. The different shades within each colour group reflect the fact that individual repeats within each group do differ from one another at the sequence level. Numbers indicate residue positions within each repeat, as classically defined. Throughout this figure RVD residues are left uncoloured because RVDs are left open to user definition in this design approach.

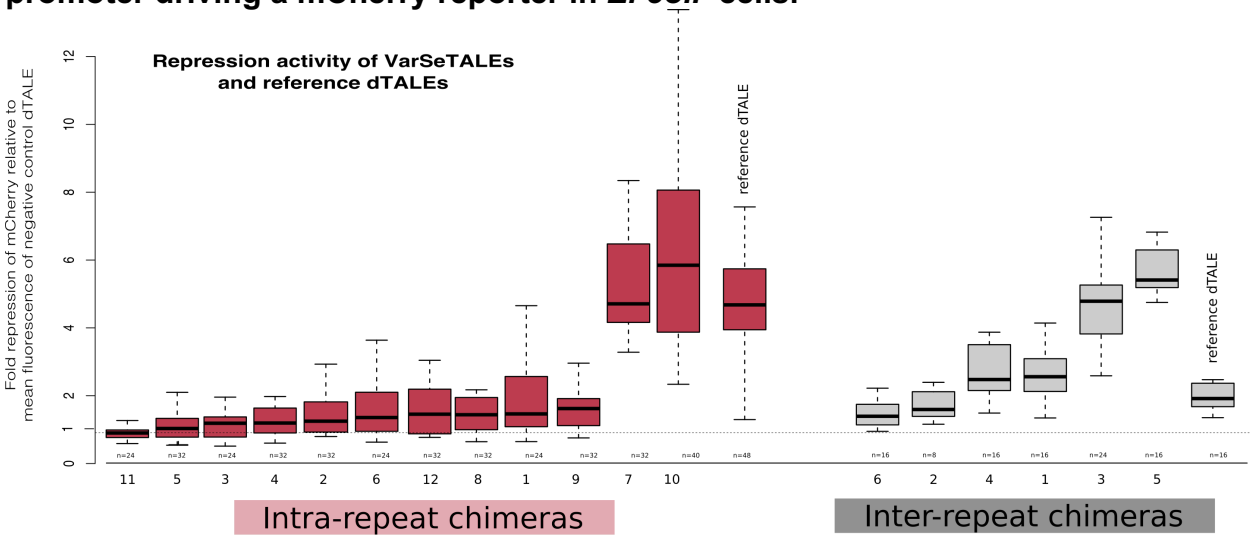
To facilitate intra-repeat chimera design, we used the known TALE repeat structure to divide up repeats into predicted secondary structural elements. To facilitate inter-repeat chimera design, we searched for a repeat position that is conserved across all TALE-like repeats within our library. This is Leucine 29. We then randomly shuffled subunits (b) or whole repeats (c) to design sequences endoding blocks of chimeric repeats. These repeats were synthesized and then cloned into otherwise non-chimeric dTALE repeat arrays for functional testing.

For both intra- and inter-repeat VarSeTALEs only a subset of repeats is chimeric within the full array. A set of chimeric repeats are embedded within an AvrBs3-derived dTALE. In the case of Intra-repeat chimeras only 3-4 repeats per array are chimeric, whereas 5-10 are chimeric in Inter-repeat chimeras. In all cases chimeric repeats were placed within the first ten repeats of the dTALE repeat array. The first ten repeats of a dTALE array have been shown to make a greater contribution to TALE-DNA interactions than subsequent repeats[21]. Please refer to figure S2 for further details.

The RVD compositions of all dTALEs and VarSeTALEs in this study were chosen to recognize the sequence of the natural AvrBs3 target box from *Capsicum annuum* gene *Bs3*. Multiple RVDs are known to recognize Adenine bases[15], and for this reason the RVD composition differs slightly between VarSeTALEs. Specifically the RVDs of repeats 1 and 3 differ between Inter and Intra repeat chimeras and thus separate reference dTALEs are provided for each. Please refer to figure S2 for further details.

The first approach we used to compare activities of VarSeTALEs and reference TALEs was a repression assay in *E. coli* , based on a TALE-repressor system[23 16]. In this assay a TALE binds to a modified *Trc* promoter driving constitutive *mCherry* expression in E.coli. Strong binding of the TALE is assumed to impair promoter activity by occluding the RNA polymerase complex. We were previously able to demonstrate that strong versus weak repression correlates to higher and lower DNA-binding affinity *in vitro*[16]. VarSeTALE and reporter plasmids were co-transformed into *E. coli* and resulting colonies were used to inoculate separate cultures in wells of a 96-well plate and then after 3.5 hours growth mCherry expression and cell density (OD 600) were measured in a plate reader. Results are shown in Figure 3.

Figure 3: Ability of a set of VarSeTALEs and reference dTALEs to repress a promoter driving a mCherry reporter in *E. coli* cells.

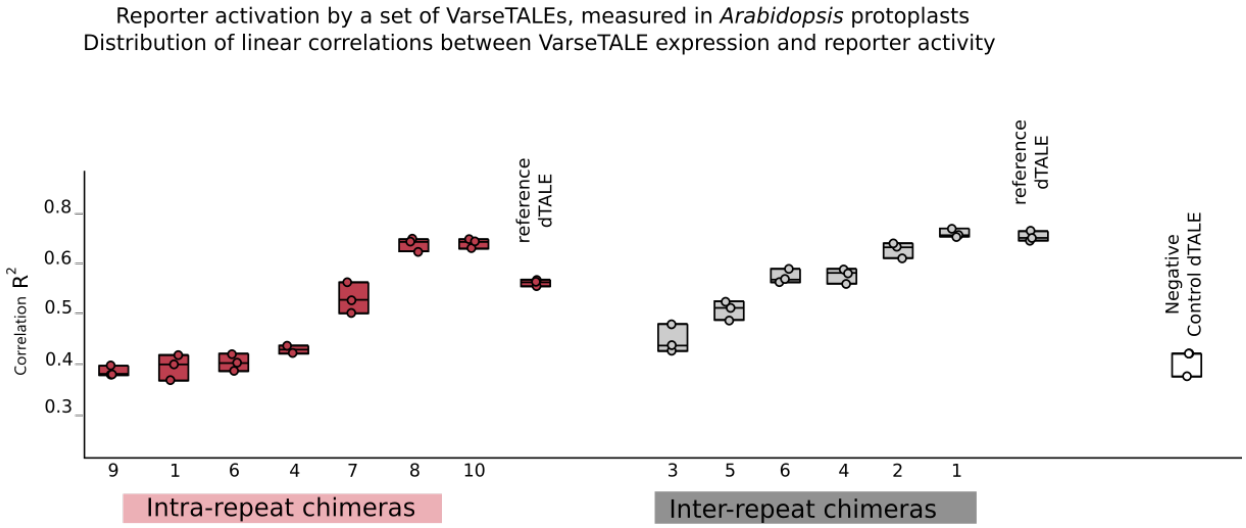


13 Intra-repeat (red) and 6 Inter-repeat (grey) chimeras were assembled based on a pool of diverse TALE repeat sequences. All chimeras and control dTALEs were tested for their ability to repress transcription from a bacterial promoter containing a cognate binding element, in a promoter driving expression of an mCherry reporter gene. Boxplots of fold mCherrys repression, relative to an unrelated, negative control, dTALE are shown along with the number of biological replicates tested given underneath in each case. A dotted line at one indicates basal reporter activity without repression. Each biological replicate corresponds to a single colony picked into a 96-well assay plate. Reference dTALEs were assembled entirely from AvrBs3-derived repeats[9]. VarseTALEs are ordered within their groups based on increasing repression strength, with numbers (X) below each boxplot giving the identifier of each VarSeTALE (IntraX or InterX in Table S1).

Our design goal was to create VarSeTALEs that mediate a range of reporter activities. No prediction was made as to the activities of individual VarSeTALE sequences, only that due to the spread of sequence polymorphisms that the set of VarSeTALEs would capture a range of reporter repression levels. That is indeed what we observed (Figure 3). For both the intra- and inter-repeat chimeras the range of repression strengths ranged from barely detectable to above the upper reference, as inferred from comparison of sample medians. However, there was a striking difference between the performance of the two design approaches. Only two out of 13 intra-repeat chimeras (Figure 3, red) were able to achieve at least 2-fold reporter repression, compared to four out of six inter-repeat chimeras.

The repression assay in *E. coli* is a simple system to measure VarSeTALE-DNA binding strength using reporter output as a proxy. However, we assume that it relies on simple stoichiometric repression. No additional interactions between the VarSeTALE and the cellular machinery are required. When dTALEs are used in eukaryotes for regulation of synthetic genetic circuits, they are fused to activation or repression domains[24]. Such domains rely on interactions with specific components of the cellular machinery, which differ between prokaryotes and eukaryotes. So we next tested the same constructs in a transcriptional activation assay in eukaryotic cells. We chose to work in a plant cell system to exploit the natural C-terminal domain of AvrBs3, which encodes a strong *in planta* transactivation domain[25]. Specifically, we used Arabidopsis root cell culture protoplasts. Full length C-terminal domains from AvrBs3, including the natural transactivation domain where included in each construct. Each VarSeTALE was also fused C-terminally to GFP to allow us to control for expression variability between transfected protoplasts. In this case, the reporter gene used was an *mCherry* coding sequence located 3' of the 360bp *Bs3* promoter fragment, which has previously shown to be transcriptionally silent *in planta* but inducible by a cognate dTALE [22]. Activation strengths are shown in Figure 4, for all inter-repeat chimeras and a subset of VarSeTALEs capturing the full range of activities measured for the repressor reporter (Figure 1).

Figure 4: Reporter activation by a set of VarSeTALEs, measured in Arabidopsis protoplasts.



7 intra- (red) and 6 inter- (grey) repeat chimeras, tagged with GFP, were transfected into A. thaliana protoplasts alongside an mCherry reporter gene containing a cognate binding site in its promoter (*pBs3*). Fluorescence levels were measured using a MoFlo XDP flow cytometer. The linear correlation value linking GFP to mCherry for each population of cells was calculated and correlations for each of three replicate populations are plotted for each construct. Reference dTALEs were assembled entirely from AvrBs3-derived repeats[9], the negative control dTALE bears a repeat array with no cognate target site in the Bs3 promoter. VarSeTALEs are ordered within their groups based on increasing repression strength, with numbers (X) below each boxplot giving the identifier of each VarSeTALE (IntraX or InterX in Table S1).

As for the repressor assay (Figure 1) we found that VarSeTALEs spanned a range of activation strengths (Figure 3). Activation strength is approximated with the correlation between GFP (VarSeTALE) and mCherry (reporter) within each population. Although VarSeTALEs were all expressed from the same, constitutive promoter, there is a natural heterogeneity in dTALE/VarSeTALE expression among cells in each measured population. Since we could measure dTALE/VarSeTALE expression through GFP fluorescence this was correlated to reporter activation, as inferred from mCherry fluorescence. Correlation statistics are thus used here as measures of activation strength for each dTALE/VarSeTALE .

The results for particular VarSeTALEs agree in several cases between the two assay systems, though in most cases they do not. Examples of agreement include intra7 and intra10. These were the only two intra-repeat chimeras to mediate more than two-fold median repression(Figure 3) and also performed well in the activation assay (Figure 4). In addition Intra-repeat chimeras 1, 6 and 9 all fall in the lower range of both repression and activation activities (Figures 2 and 3).

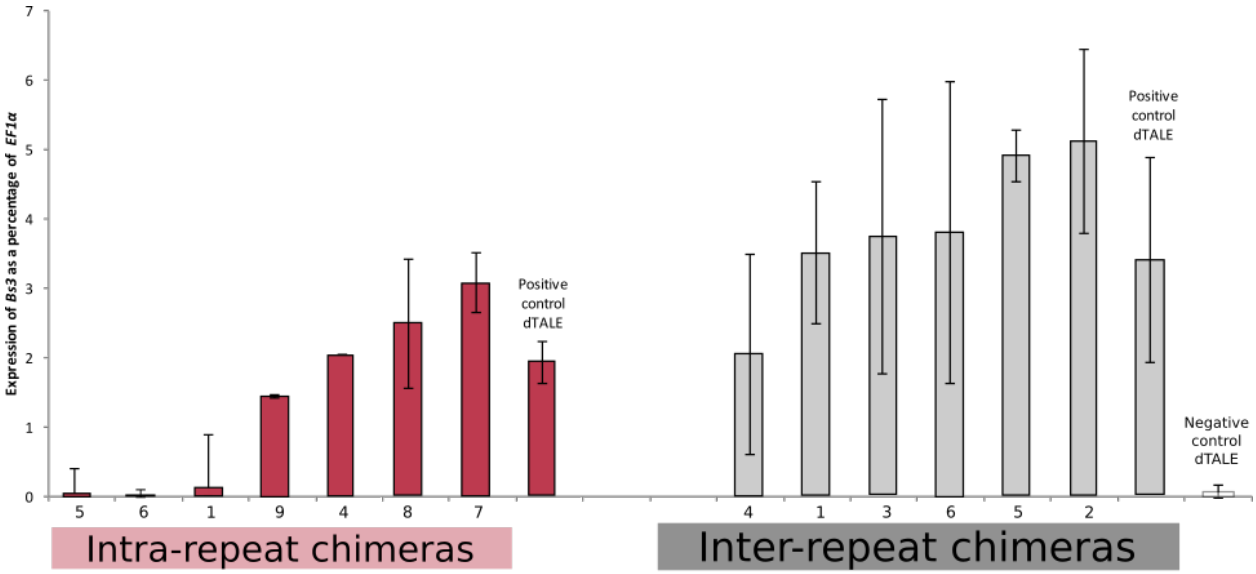
For some constructs the results for the two assay systems differ considerably. Inter-repeat chimeras 3 and 5 were the top performers in the repressor assay but the weakest activators (Figures 2 and 3). This may relate to the difference between stoichiometric repression, where higher binding affinity is correlated to higher repression strength[16]. In the case of activation, binding must be accompanied by recruitment of the

transcriptional machinery and unwinding of the double helix downstream to allow transcription. In such a scenario a high affinity, particularly a low K_{off} may be disadvantageous. A study that derived K_Ds as well as fold-activations for a set of 20 dTALEs, differing in RVD composition, found an overall positive correlation between the two measures, but that this correlation disappeared for very those with very low K_Ds (high affinity)[21]. This is supported by the relative performances of the intra- and inter-repeat chimera reference dTALEs, which differ from each other by a single RVD (Figure S2). The stronger activator is the weaker repressor and vice versa (Figures 2 and 3). Thus some of the observed discrepancies are a consequence of the chosen assays and are not necessarily contradictory.

The key specification we were hoping to achieve from our designs is that together they cover a range of activation levels, and in this they achieved their aim. We therefore next tested whether this property was preserved in the activation of a chromosomally embedded gene, a common application of dTALEs[12,26]. The *Bs3* gene of bell pepper (*Capsicum annuum* ECW30-R) contains a target site for AvrBs3 in its promoter[27]. All VarSeTALEs in this study were made with an RVD composition matching the AvrBs3 target box in the *Bs3* promoter. We introduced VarSeTALE genes into bell pepper leaves via *Agrobacterium tumefaciens* transient transformation and quantified *Bs3* transcript levels via qPCR which provides a proxy for promoter activation levels.

We hoped to see a range of activation levels of the usually transcriptionally silent *Bs3* gene. This is indeed what we observed (Figure 5), with VarSeTALEs of both design types. This demonstrates that VarSeTALEs can be used to achieve a range of endogene activation levels. However, a greater range of endogene activation levels, compared to the relevant reference dTALE, was captured by the intra-repeat chimeras (red) than the inter-repeat chimeras (grey). Absolute values for these two groups should not be directly compared because the assays were performed on separate days on separate plants.

Figure 5: A range of gene expression levels achieved for a *Capsicum annuum* endogene by using a set of VarSeTALEs.



Leaves of *C. annuum* plants containing the Bs3 gene, a natural target of TALE AvrBs3 and therefore of the VarSeTALEs in this study, were infiltrated with *A. tumefaciens* strains to deliver VarSeTALEs and reference dTALEs. Expression of the Bs3 gene was quantified with qPCR and compared to housekeeping gene EF1a (Elongation factor 1 alpha). Data shown are means with error bars showing plus and minus one standard deviation. The RVD composition of the negative control dTALE shares no identity with the VarSeTALEs and is thus unable to activate the *Bs3* promoter. VarseTALE are ordered within their groups based on increasing repression strength, with numbers (X) below each boxplot giving the identifier of each VarSeTALE (IntraX or InterX in Table S1).

Conclusion

(not actually a separate section in the ACS format)

We report here a general method to tune TALE DNA binding properties without altering RVDs. We generated a set of synonymous TALE-effector-based transcription factors (VarSeTALEs) with a range of activities on a target promoter but synonymous in terms of target sequence. The sequences of our VarSeTALEs (SI) are a small subset of the possible Inter- and Intra-repeat chimeras that could be derived from naturally occurring TALE repeats (further sequences in SI?). We therefore encourage further exploration of the VarSeTALE sequence space whilst equally inviting interested parties to use the exact sequences in this study as chassis for the creation of novel sets of VarSeTALEs by simply replacing RVDs used here with those matching a promoter of interest. We would stress however, that upon generating a new set of VarSeTALEs (new RVD composition) that their actual performance should be tested in the system of interest, since, as we have shown, relative activities of some VarSeTALEs differed considerably in the three different assay systems. However, what we anticipate is that using a set of

VarSeTALEs, either those presented here or independently derived, will capture a range of reporter activity levels without the requirement for any rational engineering. We hope this approach will prove useful both within synthetic biology and molecular biology more generally.

An additional benefit of VarSeTALEs is that their repeats are more diverse at the DNA level. The runs of DNA repeats that encode conventional TALE repeat arrays are problematic for PCR based manipulation[28] and are susceptible to recombinatorial sequence deletion in some systems[29]. In the later case the problem of recombination an be alleviated by lowering repeat sequence similarity[30] through codon redundancy, but the added diversity that comes from amino acid level polymorphism provides an alternative solution.

In this study we generated VarSeTALEs from a manual assembly of different TALE and TALE-like repeat sequences without the aide of additional design rules. We also only drew on naturally occurring sequences. The development of VarSeTALE design rules could assist the generation of VarSeTALEs with predictive power. The use of designed peptide sequences could be used to further augment the properties of VarSeTALEs by introducing diversity at repeat positions for which there is none naturally. In addition, the recent characterization of TALE-like DNA binding proteins from marine bacteria expand the repeat sequence pool of TALE-likes[16].

We explored two different approaches for introducing diversity (Figure 1). In each case our design approach did not attempt to predict the binding strength of each VarSeTALE for its target. The prediction made was that a set of distinct VarSeTALEs will differ in binding strength from each other and from the reference AvrBs3 repeat array. To generate intra-repeat chimeras each secondary structural element within a repeat was treated as a module. In the inter-repeat chimera approach whole repeats were treated as modules. Sets of both intra- and inter- repeat chimeras mediated a range of repression (Figure 3) or activation strengths (Figures 3 and 4). Inter-repeat chimeras covered a smaller range of activation strengths (Figures 3 and 4), but a greater range of repression strengths (Figure 3). Additionally, most intra-repeat chimeras were very poor repressors, barely distinguishable from the negative control (Figure 3). An initial set of 13 (Figure 3) was reduced to a set of seven for later assays (Figures 3 and 4). In contrast all six inter-repeat chimeras generated mediated measurable activation or repression in all assay systems. The inter-repeat chimera approach may be more reliable. We have previously observed that rearranging repeats within a polymorphic TALE-like repeat array is linked to poor activity in transcriptional reporter assays[18]. We speculated then that the introduction of novel repeat interfaces may impair protein folding. If so this could explain the poor performance of many Intra-repeat chimeric VarSeTALes, which, through design, contain numerous novel interfaces both within and between repeats.

VarSeTALE design can be used to generate sets of synonymous TALEs differing in activity on a target promoter. Reverse genetics is one application for VarSeTALEs, allowing the phenotypic effect of a range of endogene expression levels to be assayed.

If a permissive promoter position has been identified a set of VarSeTALEs could be transformed into the organism of interest to arrive directly at a set of transgenic lines differing in expression of the endogene of interest. This approach would be applicable for designer activator or repressor TALEs, both of which have already been used in a range of host organisms[22,31,32]. Synthetic genetic circuits are another potential application, where libraries of promoters are typically used to tune expression levels [33]. VarSeTALEs offer a way to tune promoter strengths in trans based on targeting the same cis-element in each case, simplifying design. The engineerable TALE-DNA interface can therefore be seen as a tool to tune transcription that can be incorporated into the design of synthetic circuits.

Methods

VarSeTALE design:
Intra-repeat chimeras were designed by randomly selecting sets of sequences from a set of unique TALE, RipTAL and Bat sub-repeat modules, corresponding to presumed secondary structural elements (Table S1). Each Intra-repeat chimera contains a block of 3 or 4 such randomly-assembled repeats, replacing an equal number of AvrBs3 repeats at positions 1-4, 5-7 or 7-B. See Figure S2 for sequences and further details.

Inter-repeat chimeras were designed by randomly selecting from a set of unique TALE, RipTAL and Bat whole repeat sequences (Figure S2). Each Inter-repeat chimera contains a block of 5 or 10 such repeats, replacing an equal number of AvrBs3 repeats at positions 1-5, 6-10 or 1-10. Inter-repeat chimeras 5 and 6 are the combinations of 1 and 3, and 2 and 4 respectively. See Figure S2 for sequences and further details.

All chimeric repeat blocks were synthesized (Genscript) with *Xanthomonas euvesicatoria codon* usage and flanked by BpiI restriction sites to facilitate assembly into dTALEs as described previously[19].

Molecular cloning:
For the repressor assays displayed in Figure 3 VarSeTALE repeat arrays were cloned into a derivative of E. coli expression vector pBT102 bearing truncated AvrBs3 N- and C-terminal domains, via Golden Gate cloning as described previously[16]. The promoter sequence of the cognate reporter (Figure S4) was introduced into pSMB6 via PCR as previously described[16].

For protoplast activation assays VarSeTALE repeat arrays were cloned into a pENTR-D derivative containing an *avrBs3* CDS lacking repeats with BpiI restriction sites in their place, as described previously[19]. *CDSs* of VarSeTALEs were then moved into T-DNA vector pGWB605[34] via Gateway LR reaction (ThermoFischer Scientific). The resulting gene is a CaMV35-S promoter driven 3' *GFP* fusion. The reporter was the 360bp fragment of the *C. annuum* Bs3 promoter cloned into pENTR derivated pENTR-Bs3p-mCherry (Figure S4).

***E. coli* repressor assay:**
The assay was carried out as described previously[16]. Briefly, TALEs and mCherry reporter genes, carried on separate plasmids and driven by different constitutive promoters, are co-transformed into *E. coli* cells. Colonies are allowed to grow to saturation on plate for 24h and then single colonies were used to incoluate 150µl scale liquid cultures in 96-well clear-bottom plates . Optical density and mCherry fluorescence were measured after 3.5hours growth using a Tecan plate reader and used to calculate a repression value for each construct, comparing in each case to the combination of the reporter with a dTALE lacking any binding site in the reporter. Data analysis was carried out in R.

Protoplast transfections & flow cytometry:
Arabidopsis root cell culture protoplasts were prepared and transfected as described (RipTAL paper). 3µg of 35-S::TALE-GFP plasmid was co-transfected with 5µg of mCherry reporter plasmid. The reporter gene was downstream of the Bs3 promoter which exhibits low basal expression in plant cells (Ref: Schandry et al, Frontiers in plant sciences?), contains the binding site of TALE AvrBs3, used as the basis for all dTALEs in this study. The negative control dTALE lacked a binding site in the Bs3 promoter. GFP and mCherry fluorescence were measured in a MoFlo XDP (Beckman Coulter) with a separate blue (488nm, elliptical focus) and yellow (561nm, spherical focus) laser for each fluorophore. GFP peak emission was captured by a 534/30 bandpass, mCherry peak emission by a 625/26 bandpass. Viable cells were identified by gating out dead cells by comparing narrow-scatter log-area vs. large-angle scatter log-area. This was followed by elimination of large cell clumps by comparing large-angle scatter log-area to large-angle scatter pulse width. Thereafter each GFP population was identified as cells having more fluorescence emission in the FL1 (534/30) compared to the FL2 (585/29) over that of un-transfected cells. Similar, mCherry expressing cells were identified by comparing FL7 (625/26) to FL6 (580/23). Alternatively, a gate [GFP or mCherry] was made to capture all transfected cells in FlowJo and the intensity values exported and processed for correlation analysis using JMP SAS, and then were compared to the results for the control dTALE.

Plant material and agroinfiltrations
Pepper (*Capsicum annuum*) plants of cultivar ECW–30R containing the resistance gene Bs3 were grown in the greenhouse at 19°C, with 16 h of light and 30% humidity. Vector constructs were introduced into *Agrobacterium tumefaciens* strain GV3101 by electroporation and selection on YEB medium containing the appropriate antibiotics. Agrobacteria were grown as liquid culture for 24 hours in YEB medium, harvested by centrifugation and resuspended in sterile water at an OD of 0.4 for infiltration. The suspension was injected into the lower side of leaves from six-week-old pepper plants. After 48 hours infiltrated patches were cut out and stored at -80°C for RNA extraction.

Isolation of RNA and quantitative real-time RT-PCR (qPCR) analysis

RNA was isolated from 50 mg frozen leaf powder with the GeneMATRIX Universal RNA Purification Kit (EURX, Gdansk, Poland). Reverse transcription was performed with one µg of the total RNA using the iSCRIPT cDNA Synthesis Kit (Biorad, Hercules, CA). Quantitative PCR reactions were performed using SYBR® Green technology (MESA GREEN qPCR Mastermix, Eurogentec, Germany) on an Bio-Rad CFX384 system (Biorad, Hercules, CA). Bs3 cDNA was amplified with primers Bs3 RT F7 and Bs3 RT R7, EF1-α cDNA with primers EF1a F2 and EF1a R2, ß-TUB cDNA with the primers ß-TUB F2 and ß-TUB R2. Data were analyzed employing the Bio-Rad CFX Manager 3.1 software with EF1-α or β-TUBULIN as a reference gene.

SI

Table S1: Unique sub-repeat modules used for Intra-repeat chimera assembly.xlsx

Table S2: Primer sequences used for qPCR experiments

Figure S1: Alignment of natural TALE repeat variation used as the basis for VarSeTALE design

Figure S2: Sequences of VarSeTALEs and reference dTALEs.

Figure S3: Correlation analysis (PDF)

Figure S4: Sequences of reporter plasmids (pSMB6 EBE AvrBs3 and pENTR-Bs3p-mCherry)

References

(1) Thakore, P. I., Black, J. B., Hilton, I. B., and Gersbach, C. A. (2016) Editing the epigenome: technologies for programmable transcription and epigenetic modulation. *Nat. Methods 13*, 127–137.
(2) Marraffini, L. A. (2015) CRISPR-Cas immunity in prokaryotes. *Nature 526*, 55–61.
(3) Hutin, M., Pérez-Quintero, A. L., Lopez, C., and Szurek, B. (2015) MorTAL Kombat: the story of defense against TAL effectors through loss-of-susceptibility. *Front. Plant Sci. 6*.
(4) Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., van der Oost, J., Regev, A., Koonin, E. V., and Zhang, F. (2015) Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell 163*, 759–771.
(5) Gao, F., Shen, X. Z., Jiang, F., Wu, Y., and Han, C. (2016) DNA-guided genome editing using the Natronobacterium gregoryi Argonaute. *Nat. Biotechnol.* 1–7.
(6) Moscou, M. J., and Bogdanove, A. J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science 326*, 1501.

(7) Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.-K. J.-K., Shi, Y., and Yan, N. (2012) Structural Basis for Sequence-Specific Recognition of DNA by TAL Effectors. *Science (80-.). 335*, 720–723.
(8) Mak, A. N.-S., Bradley, P., Cernadas, R. A., Bogdanove, A. J., and Stoddard, B. L. (2012) The crystal structure of TAL Effector PthXo1 bound to its DNA target. *Science (80-.).*
(9) Morbitzer, R., Elsaesser, J., Hausner, J., and Lahaye, T. (2011) Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Res. 39*, 5790–9.
(10) Schmid-Burgk, J. L., Schmidt, T., Kaiser, V., Höning, K., and Hornung, V. (2013) A ligation-independent cloning technique for high-throughput assembly of transcription activator–like effector genes. *Nat. Biotechnol. 31*, 76–81.
(11) Lebar, T., Bezeljak, U., Golob, A., Jerala, M., Kadunc, L., Pirš, B., Stražar, M., Vučko, D., Zupančič, U., Benčina, M., Forstnerič, V., Gaber, R., Lonzarić, J., Majerle, A., Oblak, A., Smole, A., and Jerala, R. (2014) A bistable genetic switch based on designable DNA-binding domains. *Nat. Commun. 5*, 5007.
(12) Gao, H., Wang, Y., Fei, X., Wright, D. a., and Spalding, M. H. (2015) Expression activation and functional analysis of HLA3, a putative inorganic carbon transporter in Chlamydomonas reinhardtii. *Plant J.* n/a–n/a.
(13) Lebar, T., and Jerala, R. (2015) Benchmarking of TALE- and CRISPR/dCas9-Based Transcriptional Regulators in Mammalian Cells for the Construction of Synthetic Genetic Circuits.
(14) Boch, J., and Bonas, U. (2010) Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu. Rev. Phytopathol. 48*, 419–36.
(15) Miller, J. C., Zhang, L., Xia, D. F., Campo, J. J., Ankoudinova, I. V, Guschin, D. Y., Babiarz, J. E., Meng, X., Hinkley, S. J., Lam, S. C., Paschon, D. E., Vincent, A. I., Dulay, G. P., Barlow, K. a, Shivak, D. a, Leung, E., Kim, J. D., Amora, R., Urnov, F. D., Gregory, P. D., and Rebar, E. J. (2015) Improved specificity of TALE-based genome editing using an expanded RVD repertoire. *Nat. Methods 12*, 465–471.
(16) De Lange, O., Wolf, C., Thiel, P., Krüger, J., Kleusch, C., Kohlbacher, O., and Lahaye, T. (2015) DNA-binding proteins from marine bacteria expand the known sequence diversity of TALE-like repeats. *Nucleic Acids Res. 43*, 10065–10080.
(17) de Lange, O., Schreiber, T., Schandry, N., Radeck, J., Braun, K. H., Koszinowski, J., Heuer, H., Strauß, A., and Lahaye, T. (2013) Breaking the DNA-binding code of Ralstonia solanacearum TAL effectors provides new possibilities to generate plant resistance genes against bacterial wilt disease. *New Phytol. 199*, 773–786.
(18) de Lange, O., Wolf, C., Dietze, J., Elsaesser, J., Morbitzer, R., and Lahaye, T. (2014) Programmable DNA-binding proteins from *Burkholderia* provide a fresh perspective on the TALE-like repeat domain. *Nucleic Acids Res. 42*, 7436–7449.
(19) de Lange, O., Schreiber, T., Schandry, N., Radeck, J., Braun, K. H., Koszinowski, J., Heuer, H., Strauß, A., and Lahaye, T. (2013) Breaking the DNA-binding code of *Ralstonia solanacearum* TAL effectors provides new possibilities to generate plant resistance genes against bacterial wilt disease. *New Phytol. 199*, 773–86.
(20) Streubel, J., Blücher, C., Landgraf, A., and Boch, J. (2012) TAL effector RVD specificities and efficiencies. *Nat. Biotechnol. 30*, 593–5.
(21) Meckler, J. F., Bhakta, M. S., Kim, M.-S., Ovadia, R., Habrian, C. H., Zykovich, A., Yu, A., Lockwood, S. H., Morbitzer, R., Elsaesser, J., Lahaye, T., Segal, D. J., and

Baldwin, E. P. (2013) Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic Acids Res. 41*, 4118–4128.
(22) Morbitzer, R., Römer, P., Boch, J., and Lahaye, T. (2010) Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *Proc. Natl. Acad. Sci. U. S. A. 107*, 21617–22.
(23) Politz, M. C., Copeland, M. F., and Pfleger, B. F. (2013) Artificial repressors for controlling gene expression in bacteria. *Chem. Commun. (Camb). 49*, 4325–7.
(24) de Lange, O., Binder, A., and Lahaye, T. (2014) From dead leaf, to new life: TAL effectors as tools for synthetic biology. *Plant J. 78*, 753–771.
(25) Szurek, B., Marois, E., Bonas, U., and Van den Ackerveken, G. (2001) Eukaryotic features of the *Xanthomonas* type III effector AvrBs3: protein domains involved in transcriptional activation and the interaction with nuclear import receptors from pepper. *Plant J. 26*, 523–34.
(26) Bultmann, S., Morbitzer, R., Schmidt, C. S., Thanisch, K., Spada, F., Elsaesser, J., Lahaye, T., and Leonhardt, H. (2012) Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Res. 40*, 5368–5377.
(27) Römer, 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*, 645–8.
(28) Hommelsheim, C. M., Frantzeskakis, L., Huang, M., and Ülker, B. (2014) PCR amplification of repetitive DNA: a limitation to genome editing technologies and many other applications. *Sci. Rep. 4*, 5052.
(29) Lau, C.-H., Zhu, H., Tay, J. C.-K., Li, Z., Tay, F. C., Chen, C., Tan, W.-K., Du, S., Sia, V.-K., Phang, R.-Z., Tang, S.-Y., Yang, C., Chi, Z., Liang, C.-C., Ning, E., and Wang, S. (2014) Genetic rearrangements of variable di-residue (RVD)-containing repeat arrays in a baculoviral TALEN system. *Mol. Ther. — Methods Clin. Dev. 1*, 14050.
(30) Yang, L., Guell, M., Byrne, S., Yang, J. L., De Los Angeles, A., Mali, P., Aach, J., Kim-Kiselak, C., Briggs, A. W., Rios, X., Huang, P. Y., Daley, G., and Church, G. (2013) Optimization of scarless human stem cell genome editing. *Nucleic Acids Res. 41*, 9049–9061.
(31) Cong, L., Zhou, R., Kuo, Y.-C., Cunniff, M., and Zhang, F. (2012) Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat. Commun. 3*, 968.
(32) 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.
(33) Schaumberg, K. A., Antunes, M. S., Kassaw, T. K., Xu, W., Zalewski, C. S., Medford, J. I., and Prasad, A. (2015) Quantitative characterization of genetic parts and circuits for plant synthetic biology. *Nat. Methods 13*, 94–100.
(34) Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., Toyooka, K., Matsuoka, K., Jinbo, T., and Kimura, T. (2007) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J. Biosci. Bioeng. 104*, 34–41.