Synthetic Biology, 2017, 2(1): ysx004

doi: 10.1093/synbio/ysx004
Research article

Exploiting the sequence diversity of TALE-like repeats to vary the strength of dTALE-promoter interactions

Orlando de Lange[1,2,†], Niklas Schandry[1,†], Markus Wunderlich[1], Kenneth Wayne Berendzen[3], and Thomas Lahaye[1,*]

[1]Department of General Genetics, Center for Plant Molecular Biology (ZMBP), Eberhard Karls Universitat Tübingen, Tübingen, Germany, [2]Department of Electrical Engineering, University of Washington, Seattle, WA, USA and [3]ZMBP Central Facilities, Center for Plant Molecular Biology (ZMBP), Eberhard Karls Universität Tübingen, Tübingen, Germany

*Corresponding author: E-mail: thomas.lahaye@zmbp.uni-tuebingen.de

†The first two authors are joint first authors.

Abstract

Designer transcription activator-like effectors (dTALEs) are programmable transcription factors used to regulate user-defined promoters. The TALE DNA-binding domain is a tandem series of amino acid repeats that each bind one DNA base. Each repeat is 33–35 amino acids long. A residue in the center of each repeat is responsible for defining DNA base specificity and is referred to as the base specificying residue (BSR). Other repeat residues are termed non-BSRs and can contribute to TALE DNA affinity in a non-base-specific manner. Previous dTALE engineering efforts have focused on BSRs. Non-BSRs have received less attention, perhaps because there is almost no non-BSR sequence diversity in natural TALEs. However, more sequence diverse, TALE-like proteins are found in diverse bacterial clades. Here, we show that natural non-BSR sequence diversity of TALEs and TALE-likes can be used to modify DNA-binding strength in a new form of dTALE repeat array that we term variable sequence TALEs (VarSeTALEs). We generated VarSeTALE repeat modules through random assembly of repeat sequences from different origins, while holding BSR composition, and thus base preference, constant. We used two different VarSeTALE design approaches combing either whole repeats from different TALE-like sources (inter-repeat VarSeTALEs) or repeat subunits corresponding to secondary structural elements (intra-repeat VarSeTALEs). VarSeTALE proteins were assayed in bacteria, plant protoplasts and leaf tissues. In each case, VarSeTALEs activated or repressed promoters with a range of activities. Our results indicate that natural non-BSR diversity can be used to diversify the binding strengths of dTALE repeat arrays while keeping target sequences constant.

Key words: TAL effector; plant synthetic biology; programmable transcription factor.

1. Introduction

Transcriptional activator-like effector (TALE) repeat arrays are a popular form of programmable DNA-binding domain. Reprogrammed TALE transcription factors are referred to as dTALEs (designer TALEs) and are used within the field of synthetic biology (1) as well as fundamental research (2).

Naturally TALEs are pathogenicity factors secreted by *Xanthomonas* spp. into host plant cells to bind specific promoters and activate host transcription (3). Sequence-specific DNA binding is conferred by the TALE repeat array. Natural TALE repeat arrays are formed of 10–30 repeats (4), each of which pairs with one DNA base. Collectively, the repeats of a TALE array form a right-handed superhelix, enfolding the DNA, allowing repeats

Submitted: 1 February 2017; **Received (in revised form):** 13 June 2017. **Accepted:** 16 June 2017

to contact their target bases (5,6). Although each repeat is 33–35 amino acids long, polymorphisms are mostly restricted to the residues occupying Positions 12 and 13, termed the repeat variable diresidue (RVD). Within the RVDs, the residue at Position 13 is oriented in close proximity with target bases (5,6) and defines the base preference of each repeat (7–9). The sequence of all consecutive base specificying residues (BSRs) from the N-terminal to the C-terminal end of the repeat array, known as the BSR composition, determines the 5′–3′ DNA sequence specificity. Non-BSRs are generally seen as fixed scaffolds to house BSRs. Natural *TALE* genes seem to evolve rapidly with respect to the number and BSR composition of the repeats they encode, forming TALEs with new target preferences (10). Non-BSRs are contrastingly conserved, a factor that may speed the evolution of *TALEs* through repeat recombinations (11). dTALEs, like natural TALEs, mostly differ in repeat number and BSR composition, with non-BSRs held constant. In most currently available dTALE assembly kits, repeat arrays are assembled from the four most common naturally occurring BSRs: Asn for A, Asp for C, Gly for T and Asn for G (4). This keeps dTALE design simple but does not exploit the full potential of the TALE repeat array for DNA binding. Extensive studies on uncommon or unnatural BSRs have been undertaken and revealed new base specificities (12,13). In fact, the base preferences of all possible BSRs have been assayed in dTALE repeats (14).

In this study, we explore the potential for non-BSR polymorphisms as additional parameters in dTALE design. TALE structural data show that non-BSRs do not contact DNA bases (5,6). They do, however, determine the superhelical shape of the repeat array through interactions between non-BSRs of each repeat and those of neighboring repeats (5,6). In addition, non-BSRs can make contacts with the DNA backbone, contributing to the overall affinity of the TALE-DNA interaction (5,6).

The power of non-BSRs to modify DNA-binding parameters other than base specificity has already been demonstrated in a few cases. A recent study showed that repeat polymorphisms at two non-BSR positions can increase the binding affinity between a dTALE repeat array and its target DNA (15). Another study showed that extended TALE repeats, with additional non-BSR amino acids, can be used as a tool to provide optional target base skipping (16). Both of those studies drew their inspiration from natural repeat sequence variants found among TALEs of *Xanthomonas* spp. The advantage of sticking to natural variation is that it limits the non-BSR sequence space for testing to those that have been pre-screened by natural selection. Unfortunately, as mentioned, there is little non-BSR variation among the repeats of *Xanthomonas* TALEs. However, TALE-like proteins are produced by other bacteria, including members of the *Ralstonia solanacearum* species complex, whose TALE-likes show far greater non-BSR polymorphism than those of TALEs (11). TALE-likes are also found in *Burkholderia rhizoxinica* (17,18) and two unknown marine bacteria (19), and experimental evidence shows that their repeats can be embedded as functional elements into *Xanthomonas* TALEs. Non-BSR polymorphisms abound in TALE-like repeats at almost every position (Supplementary Figure S1), yet in all studied cases BSRs confer the same specificities in these TALE-like repeats as they do in *Xanthomonas* TALE repeats. Insights from studies on TALE repeat array structures and these observations from comparative studies of TALE-likes inspired us to use natural non-BSR polymorphisms to vary repeat array binding strength while keeping target sequences constant.

We used the natural pool of sequence diversity in TALE and TALE-like repeats to assemble sequence diverse repeat arrays, termed variable sequence TALEs (VarSeTALEs). The VarSeTALEs we created have conserved BSR compositions but differ substantially at non-BSR positions. For this work, we used sequences from previously characterized TALE-like proteins from *R. solanacearum* (11,20) and *B. rhizoxinica* (18) strains. We also drew on the full diversity of *Xanthomonas* TALEs, which is generally not used (most dTALEs previously published are derived from two TALEs: AvrBs3 (21), Tal1c (22) and Hax3 (13)).

Our design goal for this study was to generate dTALEs that target the same DNA sequence but do so with a range of binding affinities. We indeed observed that VarSeTALEs mediate a range of promoter activation or repression levels in reporter assays. This is, to the best of our knowledge, the first report on the use of natural TALE-like sequence diversity to tune activities of dTALE repeat arrays while keeping BSR composition constant.

2. Materials and methods

2.1 VarSeTALE design

Intra-repeat VarSeTALEs were designed by randomly selecting sets of sequences from a set of unique TALE, RipTAL and Bat sub-repeat modules (Supplementary Table S1), corresponding to secondary structural elements based on alignments to solved TALE and TALE-like repeat array structures (5,6,23) . Each intra-repeat VarSeTALE 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 7–10 (see Supplementary Figure S2 for sequences and further details).

Inter-repeat VarSeTALEs were designed by randomly selecting from a set of unique TALE, RipTAL and Bat whole repeat sequences (Supplementary Figure S2). Each inter-repeat VarSeTALE 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 VarSeTALEs 5 and 6 are the combinations of 1 and 3 and 2 and 4, respectively (see Supplementary Figure S2 and Supplementary Sequence Files for further details).

All *VarSeTALE* repeat blocks were synthesized (Genscript) with *Xanthomonas euvesicatoria* codon usage and flanked by BpiI restriction sites to facilitate assembly into *dTALEs* as described previously (20).

The BSR 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 BSRs are known to recognize adenine bases (14), and for this reason, the BSR composition differs slightly between VarSeTALEs. Specifically, the BSRs of Repeats 1 and 3 differ between inter- and intra-repeat VarSeTALEs, and thus, separate reference dTALEs are provided for each.

2.2 Molecular cloning

For the repressor assays displayed in Figure 2, *VarSeTALE* repeat arrays were cloned into a derivative of *Escherichia coli* expression vector *pBT102* bearing truncated AvrBs3 N- and C-terminal domains, via Golden Gate cloning as described previously (19). The promoter sequence of the cognate reporter (Supplementary Figure S4 and Supplementary Sequence Files) was introduced into *pSMB6* via polymerase chain reaction (PCR) as previously described (19).

For protoplast activation assays, *VarSeTALE* repeat arrays were cloned into a *pENTR/D-TOPO* derivative containing an *avrBs3* CDS lacking repeats with BpiI restriction sites in their place, as described previously (20). *CDSs* of VarSeTALEs were then moved

into T-DNA vector *pGWB605* (24) via Gateway LR reaction (ThermoFischer Scientific). The resulting gene is a cauliflower mosaic virus 35-S (*CaMV35-S*) promoter-driven 3′ *GFP* fusion. The reporter was the 360 bp fragment of the *Capsicum annuum Bs3* promoter cloned into *pENTR-Bs3p-mCherry* (Supplementary Figure S4 and Supplementary Sequence Files) (25).

2.3 *E. coli* repressor assay

The assay was carried out as described previously (19). Briefly, *TALE* genes and *mCherry* reporter genes, carried on separate plasmids and driven by different constitutive promoters, are co-transformed into *E. coli (TOP10)* cells. Colonies were allowed to grow to saturation on plate for 24 h and then single colonies were used to inoculate 150 μl scale liquid cultures in 96-well clear-bottom plates. Optical density (OD) at 600 nm and mCherry fluorescence were measured after 3.5 h growth using a Tecan Safire2 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.

2.4 Protoplast transfections and flow cytometry

Arabidopsis root cell culture protoplasts were prepared and transfected as described (11). *35-S::TALE-GFP* (3 μg) plasmid was co-transfected with 5 μg of mCherry reporter plasmid. The reporter gene was downstream of the *Bs3* promoter, which exhibits low basal expression in plant cells (3), contains the binding site of TALE AvrBs3 and used as the basis for all dTALEs in this study. The DNA-binding domain of the negative control dTALE has no cognate binding site in the *Bs3* promoter. Green fluorescent protein (GFP) and mCherry fluorescence were measured in a MoFlo XDP (Beckman Coulter) with a separate blue (488 nm, elliptical focus) and yellow (561 nm, spherical focus) laser for each fluorophore. GFP peak emission was captured by a 534/30 bandpass and mCherry peak emission by a 625/26 bandpass. Viable cells were identified by gating out dead cells by comparing narrow-scatter log-area versus 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 with the FL2 (585/29) over that of un-transfected cells. Similar, mCherry expressing cells were identified by comparing FL7 (625/26) with FL6 (580/23). Finally, a gate [GFP or mCherry] was made to capture all transfected cells and exported.

For data analysis, the raw GFP (FL1) and mCherry (FL7) measurements of the gated population were log10 transformed. Using log(GFP) as a measurement of VarSeTALE expression levels and log(mCherry) as a read-out for promoter activity, we estimated log(mCherry)/log(GFP), hereafter referred to as 'activity' using a linear model, assuming a fixed intercept for all effectors. Analysis of variance and *post hoc* testing were used to assess statistically significant differences between the individual VarSeTALEs. To arrive at the activities displayed in Figure 3, we subtracted from each measured event the activity of a negative control dTALE, dBat1, which cannot bind to the reporter construct. Thus, all activities are relative to background promoter activity.

The linear model was constructed using the lm() function, and pairwise testing and group assignment were performed using the lsmeans() and cld() functions from the *lsmeans* and *multcomp* R packages (26,27).

2.5 Plant material and *Agrobacterium* leaf infiltrations

Pepper (*C. 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 pMP90 by electroporation and selection on YEB medium (5 g/l beef extract, 1 g/l yeast extract, 5 g/l peptone, 5 g/l sucrose, 0.5 g/l MgCl2, Bactoagar 15 g/l for solid medium) containing the appropriate antibiotics. *Agrobacterium* strains were grown as liquid culture for 24 h 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 6-week-old pepper plants. After 48 h, infiltrated patches were cut out and stored at −80 °C for RNA extraction.

2.6 Isolation of RNA and quantitative real-time reverse transcriptase PCR 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 1 μg of the total RNA using the iSCRIPT cDNA Synthesis Kit (Bio-Rad, Hercules, CA, USA). Quantitative PCR reactions were performed using SYBR® Green technology (MESA GREEN qPCR Mastermix, Eurogentec, Germany) on a Bio-Rad CFX384 system (Bio-Rad). *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 complementary DNA 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.

3. Results

3.1 VarSeTALE repeat arrays contain large numbers of non-BSR polymorphisms

In this study, we generated VarSeTALEs, which are dTALE repeat arrays bearing several repeats with sequences drawn from different TALE and TALE-like origins. Specifically, we generated VarSeTALEs repeat modules using sequences from *Xanthomonas* TALEs, *Ralstonia* TALE-likes (RipTALs (20)) and *Burkholderia* TALE-likes (BurrH/Bat1 and Bat2 (17,18)). Sequences of TALE-likes repeats used as the raw material for design for the repeat arrays generated in this study are displayed in Supplementary Figure S1.

We explored two alternative approaches for VarSeTALE design. We combined either whole repeats (inter-repeat VarSeTALEs) or repeat subunits (intra-repeat VarSeTALEs). For inter-repeat VarSeTALEs, the highly conserved leucine residue at Position 29 (Supplementary Figure S1) was used as the break point between repeats of different origins. Repeat subunits used in our intra-repeat VarSeTALEs are shown in Figure 1 and correspond to secondary structural elements (9): short-helix (AA 4–10), BSR loop (AA 11–15), long-helix (AA 16–28) and inter-repeat loop (AA 29–1). Figure 1 illustrates the two design approaches using example sequences.

The VarSeTALEs generated in this study are AvrBs3 derivatives, bearing between 3 and 10 sequence diverse repeats in place of AvrBs3 repeats. In the case of intra-repeat VarSeTALEs, only 3–4 repeats per array were replaced, whereas 5–10 were replaced to create inter-repeat VarSeTALEs. All intra-repeat VarSeTALEs have exactly the same BSR composition as AvrBs3, while, as a contingency of our cloning strategy, intra-repeat VarSeTALEs use NI BSRs (A-specifying) in two positions where

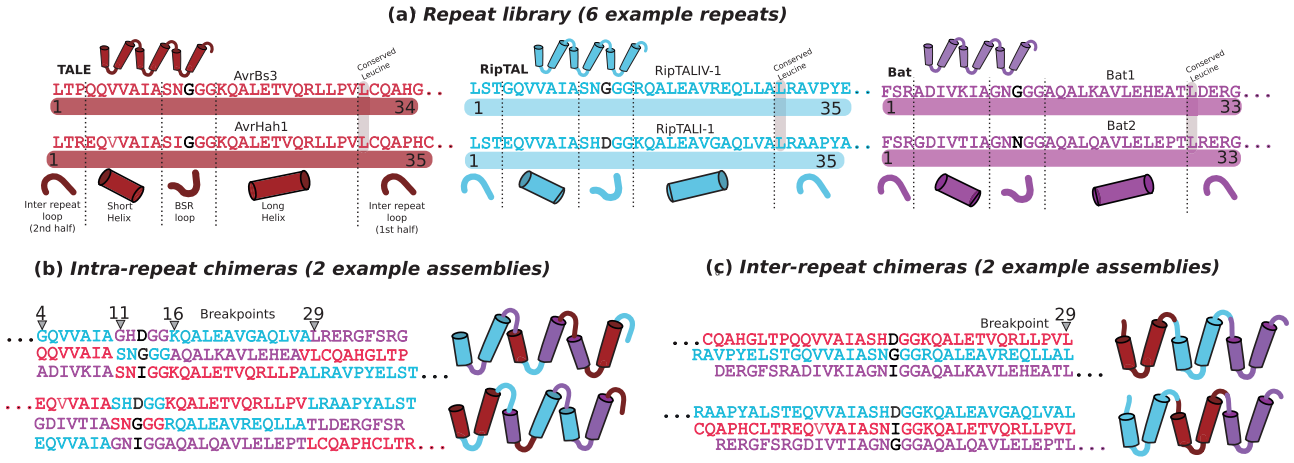

Figure 1. (a) Starting material was an *in silico* repeat library of non-identical repeat sequences from TALEs and TALE-likes. Sequences were sorted based on bacterial origin (TALEs of *Xanthomonas*, RipTALs of *Ralstonia solanacearum* and Bats of *Burkholderia*). Color coding reflects these groupings throughout this figure. Numbers indicate residue positions within each repeat, as classically defined. Throughout this figure, BSR residues are left uncolored, as BSRs are kept constant. To facilitate intra-repeat VarSeTALE design, the known TALE repeat structure 5, 6 was divided up into predicted secondary structural elements. To facilitate inter-repeat VarSeTALE design, we searched for a repeat position, close to a helix-loop transition, which is conserved across all TALE-like repeats within the library; this was leucine 29. Repeat subunits (b) or whole repeats (c) were randomly shuffled to design sequences encoding blocks of sequence diverse repeats. These repeats were synthesized and cloned into otherwise standard dTALE repeat arrays, generating VarSeTALEs for functional testing.

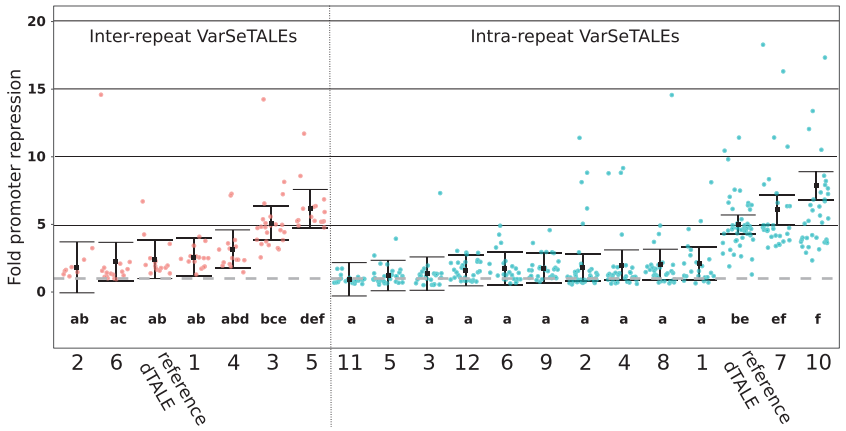

Figure 2. Six inter-repeat (red) and 12 intra-repeat (blue) VarSeTALEs and control dTALEs (Ref.) were tested for their ability to repress transcription from a bacterial promoter containing a cognate-binding element, in a promoter driving expression of an *mCherry* reporter. Repression indicated on the *y*-axis is displayed as fold *mCherry* promoter repression (Base 10), relative to an unrelated, negative control, dTALE (Supplementary Figure 2C). A dashed line at 1 indicates basal reporter activity without repression. Each dot corresponds to a single colony picked into a 96-well assay plate. Black squares indicate the mean for each VarSeTALE. Errorbars show the 95% confidence intervals. Reference dTALEs were assembled entirely from AvrBs3-derived repeats (21). VarSeTALEs are ordered within their groups based on increasing repression strength below each plot the identifier of each VarSeTALE is given. Letters indicate significance groups from generalized linear hypothesis testing conducted for all VarSeTALEs; samples sharing a letter are not significantly different. Outlier values above 20 are not shown but were included in the calculations (4 values of intra-repeat VarSeTALE 10 are not plotted, these are at 22.0. 22.6, 23.2 and 26.0).

other A-specifying BSRs are found in AvrBs3. Please refer to materials and methods for further details and figure S2 for full amino-acid sequences of all VarSeTALEs generated.

We hypothesized that the non-BSR polymorphisms of VarSeTALEs will result in differing binding strengths on their target DNA boxes, mediating either weaker or stronger promoter regulation as a consequence. We used three experimental approaches that infer relative binding strengths from differential promoter regulation: an *E. coli* promoter repression assay (Figure 2), an *Arabidopsis* protoplast transactivation assay (Figure 3) and *Agrobacterium* delivery into *C. annuum* (bell pepper) to activate a genomic promoter (Figure 4).

3.2 Differential promoter repression by VarSeTALEs in *E. coli*

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 (29,19). In this assay, a TALE binds to a modified *Trc* promoter driving constitutive *mCherry* expression in *E. coli*. dTALE promoter binding is assumed to impair promoter activity by occluding the RNA polymerase complex. We were previously able to demonstrate that in this assay repression correlates to DNA-binding affinity as measured *in vitro* (19). VarSeTALE and reporter plasmids were co-transformed into *E. coli*, and the resulting colonies were used to inoculate separate cultures in wells of a 96-well plate. After 3.5 h of further growth

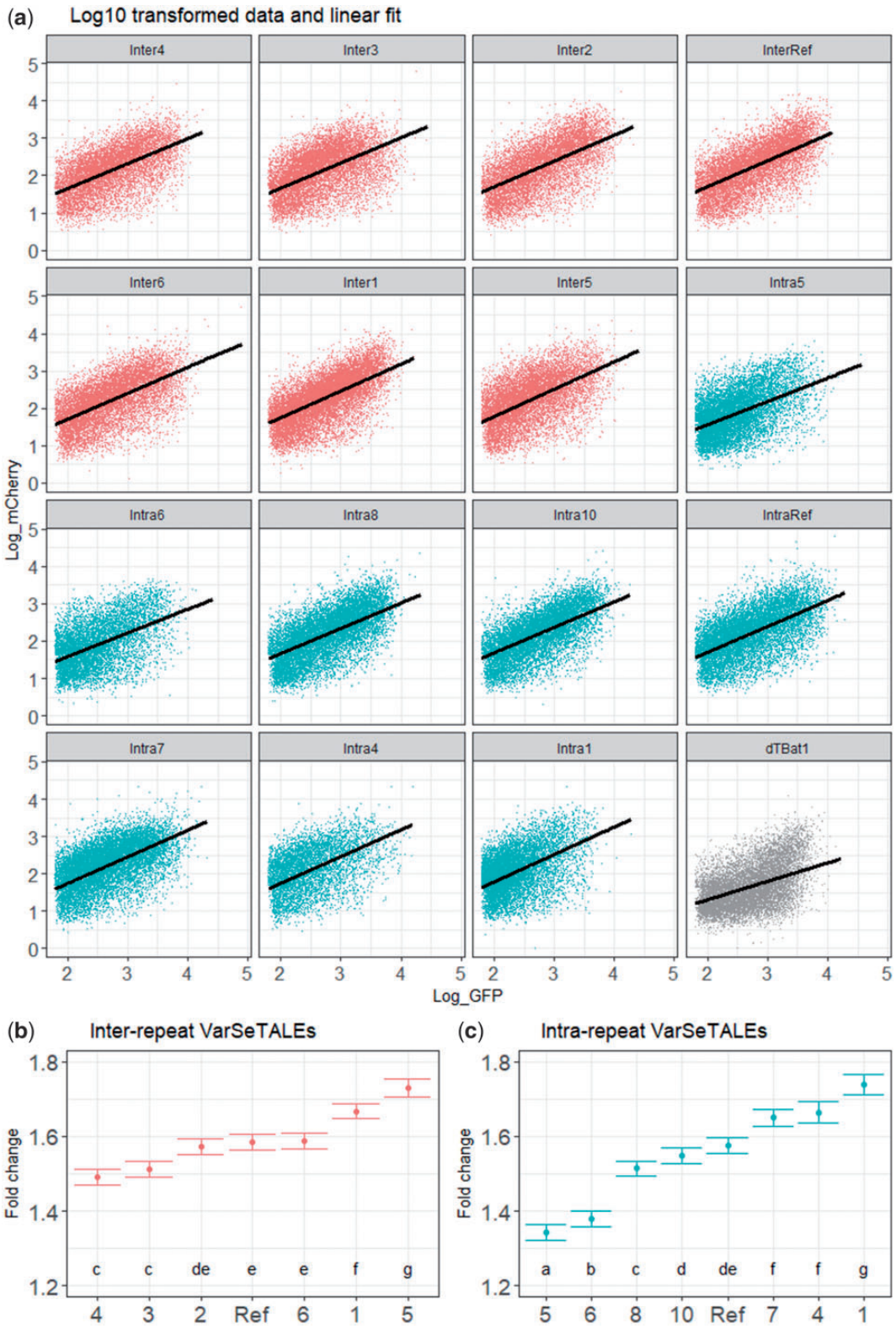

Figure 3. Results of flow cytometry reporter assays. (**a**) Overview of the activation of the reporter constructs by different inter- and intra-repeat VarSeTALEs and negative control dTALE (colored red, blue and gray, respectively). Raw GFP and mCherry counts were log10 transformed to increase linearity before plotting (dots) and linear modeling (black line). Subsequently, the estimated log10(mCherry) per log10(GFP) was adjusted for the activity of the negative control (dTBat1 (18)), back transformed and plotted ordered by increasing fold activation, with 95% confidence interval (indicated by error bars). (**b**) Intra-repeat VarSeTALEs and (**c**) inter-repeat VarSeTALEs. Letters are used to indicate statistically significant difference across panels (b and c). Constructs that do not share a letter have different mean fold changes and given a significance threshold of 0.05.

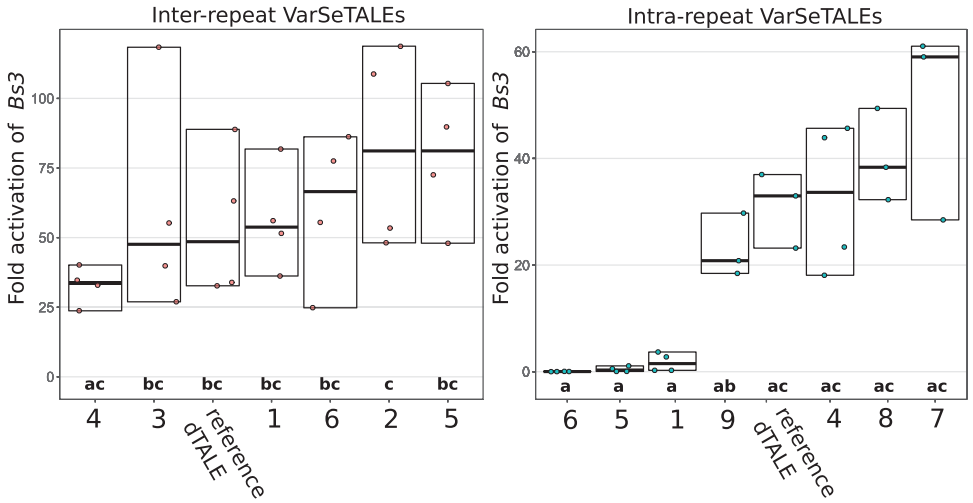

Figure 4. Leaves of *Capsicum annuum* plants containing the basally transcriptionally silent *Bs3* gene, a natural target of TALE AvrBs3 and therefore of the VarSeTALEs in this study, were infiltrated with *Agrobacterium tumefaciens* strains to deliver T-DNA constructs encoding VarSeTALEs and reference dTALEs. Expression of the *Bs3* gene was quantified with qPCR and compared with the mean expression of *Bs3* in the negative control, which was a dTALE (dTALE Bat1 (18)) unable to bind to the promoter of the *Bs3* gene. Circles represent individual replicates (red for inter- and blue for intra-repeat VarSeTALEs), which are connected into a data block to visualize the range of expression obtained for each VarSeTALE with a thick horizontal line at the sample median. Below each data block, the identifier of each VarSeTALE is given. Letters indicate significance groups based on pairwise *t*-tests; the means of samples sharing a letter are not significantly different.

mCherry expression and cell density (OD 600 nm) were measured in a plate reader. Results are shown in Figure 2.

Our expectation was that VarSeTALEs would mediate a range of reporter activities. No prediction was made as to the activities of individual VarSeTALEs. Instead, we expected that due to the spread of sequence polymorphisms the whole set of VarSeTALEs would capture a range of reporter repression levels. That is indeed what we observed (Figure 2). For both the intra- and inter-repeat VarSeTALEs, the range of repression strengths ranged from barely detectable to above the activity of the reference dTALE, as inferred from comparison of sample medians.

For both designs, a range of repression strengths were achieved, but the range was smaller for inter-repeat VarSeTALEs. Ten of the 12 intra-repeat VarSeTALEs (Figure 2, blue) mediated significantly weaker reporter repression than the reference dTALE, though intra-repeat VarSeTALE 10 mediated significantly stronger repression than the reference. The inter-repeat VarSeTALEs displayed a similar relationship to their reference dTALE but with a slightly smaller total range of median fold repression strengths: 4.1 compared with 4.8 for intra-repeat VarSeTALEs.

Since 10 of the intra-repeat VarSeTALEs displayed repression strengths that were not significantly different from one another, some were set aside in the next experiments. Intra-repeat VarSeTALEs 1, 4, 5, 6, 7, 8, 9 and 10 were chosen to capture the full range of activities measured for the repressor reporter (Figure 2).

3.3 Differential promoter transactivation by VarSeTALEs in *Arabidopsis* protoplasts

The *E. coli* repression assay used in Figure 2 gives a straightforward read out of stoichiometric promoter repression, which should correlate directly to DNA-binding affinity. However, when dTALEs are used in eukaryotes for regulation of synthetic genetic circuits, they are fused to activation or repression domains (9). In this context, the relationship between promoter regulation and DNA binding is less direct. So, we next tested the ability of VarSeTALEs to activate a promoter driving a fluorescent reporter in eukaryotic cells (Figure 3). We chose to work in *Arabidopsis* root cell culture protoplasts to exploit the natural C-terminal domain of AvrBs3, which encodes a strong *in planta* transactivation domain (30). Each VarSeTALE was GFP-tagged to allow us to monitor VarSeTALE expression. This enables us to derive a relative estimate of the transactivation strength for each VarSeTALE (see Materials and Methods and Supplementary Material S4 for further details).

As previously shown (Figure 2), we found that VarSeTALEs mediated a range of transactivation strengths (Figure 3). This time the difference between intra- and inter-repeat VarSeTALEs was more pronounced. The seven intra-repeat VarSeTALEs we assayed spanned a range of transactivation parameters with the same maximum and a lower minimum than the inter-repeat VarSeTALEs.

Interestingly, the relative performances of individual VarSeTALEs often differed in the transactivation assay and repressor assay. Both intra-repeat VarSeTALE 5 and inter-repeat VarSeTALE 5 do occupy the same relative positions, as the worst and best performers, respectively, in both assays. For all other constructs, there is no obvious connection between repression and transactivation performance. VarSeTALE expression level was not measured and controlled for in the repressor assay, thus expression differences between *E. coli* cells and *Arabidopsis* protoplasts may account for some of the discrepancies in observed reporter activation or repression. However, the differences in performance between the two assays are perhaps not surprising when one considers the conceptual difference between an assay of stoichiometric repression and one of promoter transactivation. A strong VarSeTALE–DNA interaction may lead to strong stoichiometric repression (19) (Figure 2). By contrast, promoter activation involves recruitment of the transcriptional machinery and unwinding of the double helix coupled to strand-disassociation downstream to allow transcription. In such a scenario, a high affinity, particularly a low K_{off}, may be disadvantageous. A study that derived DNA-binding affinities, as well as fold activations for a set of 20 dTALEs, differing in BSR composition found an overall positive correlation between DNA-binding affinity and promoter activation, but this correlation disappeared for the highest affinity TALE-DNA pairings (31).Thus, some of the observed

discrepancies are likely assay dependent and could be a consequence of the differences between stoichiometric repression and promoter transactivation.

3.4 Differential genomic promoter activation by VarSeTALEs in C. *annuum* leaf tissue

The key specification we were hoping to achieve from our designs is that VarSeTALEs with the same BSR composition can bind and regulate a promoter to a range of levels, and in this they met our expectations. We therefore next tested whether this property was preserved in the activation of a chromosomally embedded gene, a common application of dTALEs (2,32). The *Bs3* gene of bell pepper (*C. annuum* ECW30-R) contains a target site for TALE AvrBs3 in its promoter (3). We introduced constitutively expressed, CaMV35-S promoter-driven *VarSeTALE* genes into bell pepper leaves via *A. tumefaciens* transient transformation and quantified *Bs3* transcript levels via qPCR, which provides a proxy for promoter activation levels (Figure 4). The same set of VarSeTALEs was used as for the experiments in Figure 3, with the exception of intra-repeat VarSeTALE 9 in place of intra-repeat VarSeTALE 10.

We expected to see a range of activation levels of the usually transcriptionally silent *Bs3* gene. This is indeed what we observed (Figure 4), with VarSeTALEs of both design types. However, only in the case of the intra-repeat VarSeTALEs 6, 5 and 1, compared with all inter-repeat VarSeTALEs, with the exception of 4, we observed statistically significant differences. This is likely to be to a great extent a reflection of the high variability between replicates, arising from the variation in *Agrobacterium* infection and DNA delivery between leaf samples. However, as previously mentioned, the intra-repeat VarSeTALEs mediated a greater range of activation levels, from barely detectable (intra-repeat VarSeTALEs 1, 5 and 6) to almost 60-fold activation (intra-repeat VarSeTALE 7). By contrast, inter-repeat VarSeTALEs mediated 30- to 80-fold activation.

Again the relative performances of VarSeTALEs measured in this assay do not correspond well to the results of the transactivation reporter assay (Figure 3). In this case, both assays provide a measure of promoter transactivation in plant cells. Interestingly, the weakest activators in both transactivation assays (intra-repeat VarSeTALEs 5 and 6 and inter-repeat VarSeTALE 4) are also among the weaker repressors (Figure 2). This suggests that those VarSeTALEs are poor DNA binders leading to consistently weak promoter regulation. The data set in this study is not extensive enough to allow detailed analysis of the effects of specific sets of non-BSR polymorphisms even though our overall results do indicate that VarSeTALE repeat arrays, containing high numbers of non-BSR polymorphisms, bind DNA and regulate the same promoters to different levels.

4. Discussion

4.1 VarSeTALEs harness non-BSR polymorphism to tune promoter regulation

Our goal was to harness natural non-BSR polymorphisms as a means to vary TALE-DNA-binding affinity without changing base preference. We created sequence diverse dTALE repeat arrays, termed VarSeTALEs, by drawing on natural TALE and TALE-like repeat diversity. We combined either whole repeats (inter-repeat VarSeTALEs) or repeat subunits (intra-repeat VarSeTALEs) to create the sets of sequence diverse repeats used in the study. Using three different experimental approaches, we demonstrated that sets of 6–12 intra or inter-repeat VarSeTALEs can regulate the same target promoter to different levels (Figures 2–4). The observed differences in promoter activity are consistent with a range of VarSeTALE-DNA-binding affinities based on previous work with the *E. coli* TALE-repressor assay directly comparing repression with DNA-binding affinity (19). This study was limited to measures of relative promoter regulation strengths, and we did not test for possible alterations of base preference. Our data, while not exhaustive, do show that the VarSeTALE approach can be used to vary promoter regulation by dTALEs, while keeping BSR composition constant.

4.2 Comparing VarSeTALE design approaches: intra- versus inter-repeat

We used two different approaches to design VarSeTALEs (Figure 1), but in each case, the goal was the same: to vary binding strength while retaining target sequence recognition. Sets of both intra- and inter-repeat VarSeTALEs mediated a range of repression (Figure 2) or activation strengths (Figures 3 and 4). However, the intra-repeat constructs consistently outperformed the inter-repeat VarSeTALEs in plasmid reporter assays, because they covered both a greater absolute range and mediated effect strengths both above and below that of their reference dTALE in each different assay (Figures 2 and 3). Generally, inter-repeat VarSeTALEs tended to more closely match the performance of their cognate reference dTALE (Figures 2–4). The intra-repeat VarSeTALE design approach seems to have better achieved the goal of varying repeat array binding strength.

The larger observed effect range of sets of intra-repeat VarSeTALEs compared with inter-repeat VarSeTALEs may stem from the greater number of repeat sequence origins they represent. Each intra-repeat VarSeTALE repeat was assembled from four different subunits: short helix, long helix, BSR loop and inter-repeat loop (Figure 1). Each of those subunits is derived from a different TALE or TALE-like. The non-BSRs of a given TALE or TALE-like repeat array have evolved together, and it seems reasonable to assume that bringing together sequences from phylogenetically distant repeat arrays would disrupt natural intra-molecular interactions. We can assume that most novel combinations of non-BSR polymorphisms will disrupt interactions that normally hold together the TALE-like repeat structure leading to poorer DNA binding. This assumption is supported by previous work showing that rearrangements of the highly polymorphic repeats of Bat1 often impaired repeat array function (18). Indeed, out of our initial set of 12 intra-repeat VarSeTALEs, most were very poor repressors, not significantly different from the negative control dTALE (Figure 2). However, a recent study has also demonstrated that non-BSR polymorphisms that disrupt inter-repeat interactions can increase the structural flexibility of a dTALE repeat array superhelix, enhancing DNA binding (15). Intra-repeat VarSeTALE repeat arrays contain a greater number of novel non-BSR residue pairings than inter-repeat VarSeTALEs, which may explain the diversity of promoter regulation strengths we observed for these constructs.

While the intra-repeat VarSeTALEs in this study were able to mediate a range of promoter-regulation strengths, there was not an even distribution of activities. Of the 12 intra-repeat VarSeTALEs tested in the *E. coli* repressor assay (Figure 2), 10 mediated 1- to 2.5-fold repression compared with 5-fold repression mediated by the reference dTALE. Only 2 of the 12 tested intra-repeat VarSeTALEs repressed the reporter to a greater extent than the reference dTALE was able to.

4.3 Applications for VarSeTALEs: controlling synthetic gene circuits, reverse genetics and transgene stability

The creation of synthetic genetic circuits is a central practice of synthetic biology (33). Promoters are used as key regulation points within synthetic genetic circuits, tuning circuit flux through downstream gene expression. They also serve as integration points for inputs from other genes encoding transcription factors. By now, numerous studies have explored the potential for dTALE-promoter interactions to regulate synthetic genetic promoters, creating analog (29) or digital (1) control of gene expression as well as Boolean logic gates (34). Unsurprisingly, therefore, libraries of TALE-promoter pairs with different binding affinities have been characterized to serve as reusable modules in synthetic genetic circuit design (35,36). We believe that VarSeTALEs make a useful addition to those existing dTALE tools, filling a slightly different role. VarSeTALEs would be useful in cases where promoter sequence cannot be altered, but additional tuning is still desirable. VarSeTALEs could be added to existing synthetic genetic circuits without requiring any redesign of constituent promoters. VarSeTALEs could be used to fine-tune gene expression in synthetic genetic circuits, where control over expression level can include reducing as well as increasing expression to balance reaction rates.

Reverse genetics could be another application of VarSeTALEs. In this approach, the expression of a gene of unknown function is modified to observe effects on phenotype and therefore gain insights into gene function. Sets of VarSeTALEs could be built to target the same native promoter with different activation or repression strengths. If a permissive promoter position has been identified a set of VarSeTALEs could be transformed into the organism of interest, rapidly producing a set of transgenic lines differing in expression of the native gene of interest. This approach would be applicable for activator or repressor dTALEs, both of which have already been used in a range of host organisms (12,35,37).

An additional benefit of VarSeTALEs is that the DNA sequences encoding their repeats are more diverse. The runs of DNA repeats that encode conventional TALE repeat arrays are problematic for PCR-based manipulation (38) and are susceptible to recombinatorial sequence deletion in some systems (11,39). In the latter case, the problem of recombination can be alleviated by lowering repeat sequence similarity (40) through codon redundancy, but the added diversity that comes from amino acid level polymorphism provides an alternative solution. Where *dTALE* genes are intended to remain stably as transgenes over multiple generations VarSeTALEs may serve better than conventional dTALEs.

4.4 Future improvements to VarSeTALE design

We envision that the VarSeTALEs assembled in this study as an initial proof-of-concept and encourage the development of better tools to search within the total VarSeTALE design space. The VarSeTALEs in this study only capture a small subset of TALE and TALE-like repeat diversity. Especially since the recent characterization of TALE-like DNA binding proteins from marine bacteria (19) further expands the sequence pool of TALE-like repeats. The number of possible combinations of TALE and TALE-like repeats and repeat subunits is huge and random searches are a very slow method to arrive at those with desired DNA-binding properties. High-throughput assembly and screening could allow selection of promising candidates out of a VarSeTALE repeat library. Alternatively, further study of non-BSRs could be used to build rational design rules, helping users to select promising combinations. For example, a recent study used a mix of *in vitro* binding assays and molecular dynamics simulations to understand the functional impact of certain non-BSR polymorphisms at two positions within dTALE arrays (15). The VarSeTALE set in this study have poor resolution over some areas, e.g. most intra-repeat VarSeTALEs tested in this study mediated much weaker activity than the reference dTALE, with few only moderately weaker or stronger. Further studies on the structural and functional impacts of non-BSR polymorphisms could allow improved VarSeTALE design to achieve any desired repeat DNA-binding strength.

A core assumption of our design approach was that non-BSRs alter overall DNA binding affinity but do not change the target base preference of TALE repeats. This is based on previous work on TALE-like repeat arrays, all of which displayed broadly the same BSR-target base associations despite considerable non-BSR polymorphism (18–20). Yet high-throughput screens have shown that the base preference of dTALE repeats is often slightly altered by neighboring repeats (14). It therefore seems likely that the alterations to intra- and inter-repeat molecular interactions inherent to VarSeTALE design will have a range of subtle effects on base preference. A range of experimental approaches have been developed to screen base preference of dTALE repeat arrays using pools of random oligonucleotides as binding targets (31,41,42). These methods could be applied to VarSeTALEs to provide more information on base preference. This would be important to accurately predict off-targets in a genomic context.

We encourage further work to explore the VarSeTALE design concept while equally inviting interested parties to use the exact sequences in this study (provided in Supplementary Figure S2) as chassis for creating novel sets of VarSeTALEs by simply replacing BSRs used here with those matching a DNA target of interest. We would stress, however, that upon generating VarSeTALEs with a new BSR composition that their relative performances should be tested in the system of interest, because, as we have shown, relative activities of some VarSeTALEs differed considerably in the different assay systems we used in this study. However, what we anticipate is that using a set of VarSeTALEs, either those presented here or independently derived, will capture a range of promoter-regulation levels without the requirement for any rational engineering.

Supplementary data

Supplementary data are available at *SYNBIO* online.

Funding

N.S. and O.d.L. were funded by Deutsche Forschungsgemeinschaft (DFG) grant LA 1338/6 to T.L.. N.S. was funded in part by the DFG via the Institutional Strategy of the University of Tübingen [DFG ZUK 63 to NS].

Conflict of interest statement. None declared.

References

1. Lebar,T., Bezeljak,U., Golob,A., Jerala,M., Kadunc,L., Pirš,B., Stražar,M., Vučko,D., Zupančič,U., Benčina,M. et al. (2014) A bistable genetic switch based on designable DNA-binding domains. *Nat. Commun.*, 5, 5007.

2. 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.*, 82, 1–11.
3. 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, 645–648.
4. Boch,J. and Bonas,U. (2010) *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Annu. Rev. Phytopathol.*, 48, 419–436.
5. 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*, 335, 716–719.
6. 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*, 335, 720–723.
7. 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.
8. Moscou,M.J. and Bogdanove,A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, 326, 1501.
9. 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.
10. Pérez-Quintero,A.L., Lamy,L., Gordon,J.L., Escalon,A., Cunnac,S., Szurek,B. and Gagnevin,L. (2015) QueTAL: a suite of tools to classify and compare TAL effectors functionally and phylogenetically. *Front. Plant Sci.*, 6, 1–16.
11. Schandry,N., de Lange,O., Prior,P. and Lahaye,T. (2016) TALE-like effectors are an ancestral feature of the *Ralstonia solanacearum* species complex and converge in DNA targeting specificity. *Front. Plant Sci.*, 7, 1225.
12. 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.
13. Streubel,J., Blücher,C., Landgraf,A. and Boch,J. (2012) TAL effector RVD specificities and efficiencies. *Nat. Biotechnol.*, 30, 593–595.
14. 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. et al. (2015) Improved specificity of TALE-based genome editing using an expanded RVD repertoire. *Nat. Methods*, 12, 465–471.
15. Tochio,N., Umehara,K., Uewaki,J., Flechsig,H., Kondo,M., Dewa,T., Sakuma,T., Yamamoto,T., Saitoh,T. and Togashi,Y. (2016) Non-RVD mutations that enhance the dynamics of the TAL repeat array along the superhelical axis improve TALEN genome editing efficacy. *Sci. Rep.*, 6, 37887.
16. Richter,A., Streubel,J., Blücher,C., Szurek,B., Reschke,M., Grau,J. and Boch,J. (2014) A TAL effector repeat architecture for frameshift binding. *Nat. Commun.*, 5, 3447.
17. Juillerat,A., Bertonati,C., Dubois,G., Guyot,V., Thomas,S., Valton,J., Beurdeley,M., Silva,G.H., Daboussi,F. and Duchateau,P. (2014) BurrH: a new modular DNA binding protein for genome engineering. *Sci. Rep.*, 4, 3831.
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. ——, ——, 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.
20. ——, 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.
21. 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, 5790–5799.
22. Cermak,T., Doyle,E.L., Christian,M., Wang,L., Zhang,Y., Schmidt,C., Baller,J.A., Somia,N.V., Bogdanove,A.J. and Voytas,D.F. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.*, 39, e82.
23. Stella,S., Molina,R., López-Méndez,B., Juillerat,A., Bertonati,C., Daboussi,F., Campos-Olivas,R., Duchateau,P. and Montoya,G. (2014) BuD, a helix-loop-helix DNA-binding domain for genome modification. *Acta Crystallogr. D Biol. Crystallogr.*, 70, 2042–2052.
24. 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.
25. Strauss,T., van Poecke,R.M.P., Strauss,A., Römer,P., Minsavage,G.V., Singh,S., Wolf,C., Strauss,A., Kim,S., Lee,H.-A. et al. (2012) RNA-seq pinpoints a *Xanthomonas* TAL-effector activated resistance gene in a large-crop genome. *Proc. Natl. Acad. Sci. USA*, 109, 19480–19485.
26. Hothorn,T., Bretz,F. and Westfall,P. (2008) Simultaneous inference in general parametric models. *Biometrical J.*, 50, 346–363.
27. Lenth,R.V. (2016) Least-squares means: the R package lsmeans. *J. Stat. Softw.*, 69, 1–33.
28. Politz,M.C., Copeland,M.F. and Pfleger,B.F. (2013) Artificial repressors for controlling gene expression in bacteria. *Chem. Commun. (Camb)*, 49, 4325–4327.
29. 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, 523–534.
30. 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. et al. (2013) Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic Acids Res.*, 41, 4118–4128.
31. 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.
32. Endy,D. (2005) Foundations for engineering biology. *Nature*, 438, 449–453.
33. Lienert,F., Torella,J.P., Chen,J.-H., Norsworthy,M., Richardson,R.R. and Silver,P.A. (2013) Two- and three-input TALE-based AND logic computation in embryonic stem cells. *Nucleic Acids Res.*, 41, 9967–9975.
34. 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.
35. Schreiber,T. and Tissier,A. (2016) Libraries of synthetic TALE-activated promoters: methods and applications. In: *Synth. Biol. Metab. Eng. Plants Microbes Part B: Metabolism in Plants*, 361–378, Editor: O'Connor, S. 1st edn. Elsevier Inc.

36. 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. USA*, 107, 21617–21622.
37. 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.
38. 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. et al. (2014) Genetic rearrangements of variable di-residue (RVD)-containing repeat arrays in a baculoviral TALEN system. *Mol. Ther. Methods Clin. Dev.*, 1, 14050.
39. 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. et al. (2013) Optimization of scarless human stem cell genome editing. *Nucleic Acids Res.*, 41, 9049–9061.
40. Rogers,J.M., Barrera,L.A., Reyon,D., Sander,J.D., Kellis,M., Keith Joung,J. and Bulyk,M.L. (2015) Context influences on TALE–DNA binding revealed by quantitative profiling. *Nat. Commun.*, 6, 7440.
41. Miller,J.C., Tan,S., Qiao,G., Barlow,K. a., Wang,J., Xia,D.F., Meng,X., Paschon,D.E., Leung,E., Hinkley,S.J. et al. (2010) A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.*, 29, 143–148.