Assembly of custom TALE-type DNA binding domains by modular cloning

Robert Morbitzer[1], Janett Elsaesser[1], Jens Hausner[2] and Thomas Lahaye[1,*]

[1]Institute of Biology, Genetics, University of Munich (LMU), 82152 Martinsried and [2]Institute of Biology, Martin-Luther-University, 06099 Halle, Germany

ABSTRACT

Transcription activator-like effector (TALE) DNA binding proteins show tremendous potential as molecular tools for targeted binding to any desired DNA sequence. Their DNA binding domain consists of tandem arranged repeats, and due to this repetitive structure it is challenging to generate designer TALEs (dTALEs) with user-defined specificity. We present a cloning approach that facilitates the assembly of multiple repeat-encoding DNA fragments that translate into dTALEs with pre-defined DNA binding specificity. This method makes use of type IIS restriction enzymes in two sequential cut-ligase reactions to build dTALE repeat arrays. We employed this modular approach for generation of a dTALE that differentiates between two highly similar DNA sequences that are both targeted by the *Xanthomonas* TALE, AvrBs3. These data show that this modular assembly system allows rapid generation of highly specific TALE-type DNA binding domains that target binding sites of predefined length and sequence. This approach enables the rapid and flexible production of dTALEs for gene regulation and genome editing in routine and high-throughput applications.

INTRODUCTION

DNA binding domains that can be tailored to interact with user-defined DNA sequences are crucial tools for molecular biology (1). Bacterial transcription activator-like effector proteins (TALEs) from the bacterial pathogen *Xanthomonas* target DNA via a novel type of DNA binding domain that is composed of tandem-arranged 33–35 amino acid repeat-modules, with each repeat binding to one base (2,3). Base preferences of individual repeats are specified by residues 12 and 13, known as the repeat variable diresidues (RVDs), that determine preferential pairing with A (NI), C (HD), G (NK) and T (NG) nucleotides, respectively (2–5). In principle, the use of this TALE code facilitates the assembly of repeat arrays that bind to any desired DNA sequence that is preceeded by a T nucleotide (2,6,7). Recent studies have also shown that TALE repeats function as sequence specific targeting modules not only in the context of a transcription factor but also when fused to a *Fok*I nuclease domain (4,8–10). This suggests that the TALE DNA binding domain enables applications comparable to zinc finger (ZF) technology (11). ZFs that are assembled into an array are known to influence the DNA specificity of adjacent ZFs (12) and, due to this context dependency, ZF arrays of desired DNA specificity require experimental validation. In contrast, there is no evidence so far that base-preferences of TALE repeats are context dependent and recent studies have demonstrated that *in vitro* assembled repeat arrays target the pre-defined nucleotide sequences (4,5,8,10,13).

A major hurdle to the routine application of the TALE DNA binding domain is that the assembly of genes encoding tandemly arranged repeats is difficult to achieve via standard cloning approaches. To address this need, we developed a rapid, efficient, low-cost approach for engineering TALE-type DNA binding domains with custom specificities, which involves fusing individual TALE repeats into a desired array. Functional repeat arrays can be cloned directly into an expression vector or alternatively into a Gateway-compatible entry vector that facilitates recombination based transfer into any desired Gateway-compatible expression vector. We also demonstrate that TALE repeat arrays, which have been assembled by this cloning approach, target *in vivo* the pre-defined DNA sequences with high sequence specificity.

*To whom correspondence should be addressed. Tel: +49(0)89218074740; Fax: +49(0)89218074702; Email: lahaye@bio.lmu.de

MATERIAL AND METHODS

Vector construction

The *Bsa*I recognition site in pUC57 was mutagenized with primers RM1 and RM2 to obtain pUC57Δ*Bsa*I (for primer sequences see Supplementary Table S1). Repeat-modules were amplified from pENTR-D-*avrBs3* with primers adding *Bsa*I recognition and restriction sites at the 5′- and 3′-end and were blunt end cloned into pUC57Δ*Bsa*I (sequences of repeat modules are provided in Supplementary Figure S1; primer sequences are provided in Supplementary Table S1). RVD encoding nucleotides were modified via site-directed mutagenesis to obtain all six different RVDs (for nucleotide sequence of each RVD see Supplementary Figure S1). pENTR-D-*avrBs3* had been amplified with primer RM3 and RM4 to create pENTR-D-TALE-Δrep-*Bsa*I-AC. To create pUC57-AB-DEST and BC-DEST, first the *aadA* gene was amplified from pGWB441 (14) and cloned by blunt end ligation into the *Sca*I restriction site of pUC57Δ*Bsa*I using RM5 and RM6. Second, the *Bsa*I and *Bpi*I restriction and recognition sites were added using primer RM7 with RM8 and RM9 with RM10, respectively, and the PCR fragment obtained was ligated to a PCR fragment that had been amplified from pGWB441 using primers RM11 and RM12. pENTR-D-*avrBs3* had been amplified with primers RM13 and RM14 to create pENTR-D-TALE-Δrep-*Bpi*I-AC. The RVDs in pENTR-D-TALE-Δrep-*Bpi*I-AC were changed using site-directed mutagenesis. Mutations were introduced with the Phusion site-directed mutagenesis kit (New England Biolabs). All *Bsa*I and *Bpi*I recognition sites in pGWB5 (14) were mutagenized as described previously (15) to create pGWB5* (for sequence see Supplementary Figure S4). The set of plasmids that's required to generate genes encoding dTALE repeat arrays has been deposited in the non-profit plasmid repository Addgene (http://www.addgene.org).

Cut-ligation cloning protocol

For cut-ligation reaction 40 fmol of each plasmid, ligation buffer (Fermentas), 15 U of either *Bsa*I or *Bpi*I and 15 U high-concentrated T4 DNA ligase were used in a 20 μl volume. The reaction was incubated in a thermo cycler with the following program: 5 min 37°C, 5 min 20°C, 50 cycles, 10 min 50°C and 10 min 80°C. One microliter reaction mix was added to 50 μl chemical competent TOP10 cells, incubated for 15 min on ice and transformed by heat shock. Clones were analysed by colony PCR, restriction and sequencing.

In planta analysis

All entry clones were transferred by LR recombination (Invitrogen) into the expression vector pGWB5 (14) and transformed into *Agrobacterium tumefaciens* GV3101 (16) for *in planta* analysis. Sequences and generation of all promoter constructs have been described previously (5). GUS measurements were carried out as described previously (17).

RESULTS

Assembly of *TALE* genes by a one-step *Bsa*I cut-ligation

Our TALE repeat assembly kit is based on type IIS restriction enzymes that cleave outside of their recognition site and produce a 4–bp 5′ overhang (15). Since recognition and cleavage sites are spatially separated in type IIS restriction enzymes, proper construct design facilitates cleavage-mediated generation of theoretically any desired 4–bp overhang for a given DNA substrate. Thus jigsaw puzzle-like directional assembly of multiple DNA fragments is feasible (18,19). Another important aspect of type IIS mediated cloning [TIIS-cloning; synonym: Golden Gate cloning (15,18,19)] is that the desired ligation products lack the recognition sites of the given type IIS endonuclease. Thus cleavage and ligation can be carried out simultaneously (cut-ligation) rather than sequentially.

We aimed to establish a toolkit that enables TIIS-mediated assembly of genes encoding 20 or more TALE repeats. The length of the corresponding TALE target sites should facilitate specific targeting of a unique DNA sequence even in the context of a highly complex eukaryotic genome. To do so, we PCR-amplified and cloned DNA fragments encoding individual repeats (herein referred to as repeat-modules) from the *TALE* gene *avrBs3* into a modified pUC57 cloning vector (Supplementary Figure S1). To avoid vector-derived *Bsa*I cleavage products that potentially interfere with the envisaged multi-fragment ligation, we removed the *Bsa*I recognition site from pUC57 by site-directed mutagenesis, yielding pUC57Δ*Bsa*I. The cloned repeat-modules served as initial building blocks for the envisaged repeat assembly. The primers introduced *Bsa*I sites at both termini of the repeat-module in such a way that both recognition sites are cleaved off from the repeat-module (Figure 1a). Initially we cloned 10 distinct repeat-modules each producing a different combination of terminal overhangs upon *Bsa*I cleavage. The overhangs were designed so that each repeat-module would ligate only to specific repeat-modules in such a way as to generate a pre-defined linear repeat array. The first repeat-module of an array will ligate specifically to the 5′-end of the second repeat-module. The 3′-end of the second repeat-module will ligate specifically to the 5′-end of the third repeat-module and so on. Each repeat-module array consists of: (i) a 5′-adaptor repeat-module; (ii) a variable number of core repeat-modules; and (iii) a 3′-adaptor repeat-module. The position of each repeat-module within an array is defined exclusively by the given overlap. To get a sufficient number of different overlaps for the distinct repeat-modules, we made use of the degeneracy of the genetic code and incorporated different codons for identical amino acids at the repeat-module fusion points. In addition, the fusion point between the repeat-modules was varied (Supplementary Figure S1). Thus not every repeat-module does encode a complete 34-amino acid repeat. However, each repeat-module encodes one pair of RVDs and thus determines the base preference of the given repeat. The overlaps were designed in such a way that correctly assembled repeat-modules
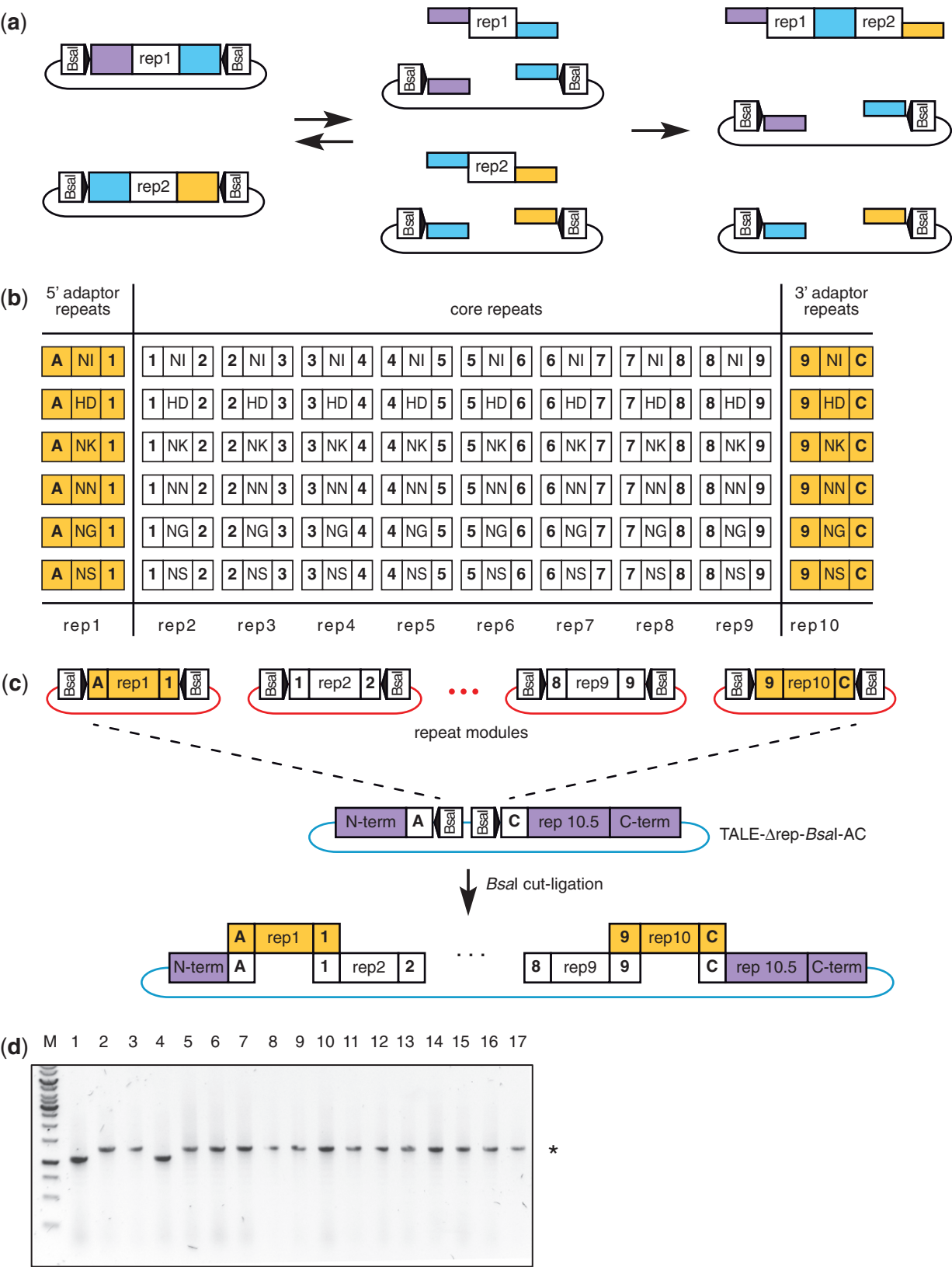
Figure 1. Assembly of dTALE arrays with 10 repeat-modules via *Bsa*I cut-ligation. (**a**) The cut-ligation concept is shown for two representative repeat-modules that are displayed as white boxes (rep1 and rep2). *Bsa*I recognition sites are shown as pentagons with black arrowheads pointing to the cleavage site. Coloured boxes represent *Bsa*I cleavage sites and colour identity indicates ends with compatible overlaps. The line connecting both *Bsa*I sites represents the vector backbone. *Bsa*I cleavage releases repeat-modules (left side) and creates distinct overlaps at the 5′- and 3′-ends on each repeat-module. Ligation of two repeat-modules results in ligation products that lack *Bsa*I recognition sites (far right side). By contrast, re-ligation of repeat-modules into their donor vectors results in plasmids that still contain *Bsa*I recognition sites. (**b**) Repeat-modules that are used for *dTALE*

(continued)

translate into an array of tandemly arranged 34-amino acid repeats. For each type of repeat-module, we generated variants encoding the RVDs NI (A), HD (C), NK (G) NN (G/A), NG (T) and NS (A/C/G/T) by site-directed mutagenesis (Figure 1b).

Next we generated a *TALE* gene deletion construct that lacks the repeat array encoding region and that allows *in vitro* integration of multiple repeat-modules by *Bsa*I cut-ligation thereby generating a functional, full-length designer *TALE* (*dTALE*) gene. The deletion was generated using an *avrBs3* gene that is flanked by *att*L sites (pENTR-D-*avrBs3*). Thus an assembled *dTALE* gene can be transferred by recombination into any desired Gateway-compatible expression vector. To generate this vector, 17 of the 17.5 repeat-modules of *avrBs3* were removed and in their place a cassette with two *Bsa*I recognition sites facing in opposite directions was inserted, yielding TALE-*Δ*rep-*Bsa*I-AC (Figure 1c, A and C denotes two distinct *Bsa*I-generated overlaps). The cassette is designed in such a way that *Bsa*I cleavage will release a fragment containing both *Bsa*I recognition sites and generate overhangs that will ligate to the 5′- and 3′-adaptor repeat-modules (Figure 1c). The repeat-deficient *TALE* gene (TALE-*Δ*rep-*Bsa*I-AC) was cloned into a vector containing a kanamycin selection marker, whereas all repeat-modules were cloned into vectors encoding an ampicillin resistance. Thus recovery of cloned repeat-modules can be easily avoided by using kanamycin-containing medium.

Next we carried out *Bsa*I cut-ligations with equimolar amounts of 10 repeat-modules and the repeat-deficient *TALE* gene to generate a functional *dTALE* gene. Restriction and sequence analysis of the observed plasmids showed that 90% of the *in vitro* generated *dTALE* genes contained the pre-defined repeat arrays (Figure 1d). Since dTALEs with 10 repeats are unlikely to be long enough to bind to a unique sequence within the context of a complex genome, we carried out cut-ligations with 20 repeat-modules using a *Bsa*I cut-ligation. However, many of the observed arrays had <20 repeats, and those that contained 20 repeats did not show the pre-defined order of repeat-modules. We anticipated that optimization of the procedure might enable us to produce repeat arrays containing 20 repeat-modules. However, given the problems that we experienced with the assembly of 20 repeat-modules it seemed unlikely that it would be simple to produce routinely arrays consisting of 30 or more repeat-modules by a single step cut-ligation with an acceptable efficiency.

Assembly of *dTALE* genes by two subsequent cut-ligations

Given that 10 repeats could be assembled efficiently in a *Bsa*I cut-ligation (Figure 1d), we decided to generate repeat arrays by ligation of two *Bsa*I-generated sub-arrays into a repeat-deficient *TALE* gene by a subsequent cut-ligation. The second cut-ligation is carried out with the type IIS enzyme *Bpi*I that, like *Bsa*I, produces 4-bp overhangs. For simplicity, we refer to the individual repeat-modules that are flanked by *Bsa*I recognition sites as level 1 modules and to corresponding sub-arrays as level 2 modules. We generated two distinct level 2 modules containing 10 and 7 level 1 modules, respectively. Both level 2 modules are assembled with mostly identical core repeat-modules but differ in their 5′- and 3′-terminal adaptor modules. Therefore in addition to A and C, we defined a new B overlap and generated 5′ and 3′ adaptor modules that connect both level 2 modules (Figure 2a and b). For the assembly of these level 2 modules, we generated two pUC57 derivatives (pUC57-AB-DEST; pUC57-BC-DEST), herein referred to as level 2 destination vectors in which level 1 modules can be assembled by *Bsa*I cut-ligation and from which level 2 modules can be subsequently released by *Bpi*I cleavage. In level 2 destination vectors, *Bsa*I and *Bpi*I are positioned in inverse orientation relative to each other but create identical overlaps (Figure 2b). The two types of level 2 modules (AB and BC level 2 modules, see Figure 2c) produce distinct overhangs when being released by *Bpi*I cleavage, which facilitates specific ligation of the two level 2 modules to each other. For the assembly of two level 2 modules into a functional *dTALE* gene by *Bpi*I-mediated cut-ligation, we generated a construct containing a repeat–deprived *TALE* gene, herein referred to as level 3 destination vector. The level 3 destination vector is basically identical to the above-described repeat-deficient *TALE* gene construct (TALE-*Δ*rep-*Bsa*I-AC) but contains two central *Bpi*I instead of *Bsa*I sites (TALE-*Δ*rep-*Bpi*I-AC). Using a *Bpi*I cut-ligation, two distinct level 2 modules are fused into a level 3 destination vector to encode a functional *dTALE* gene (Figure 2c). The level 3 destination vector also encodes the last half repeat that defines the C–terminal end of each TALE repeat array. In order to have the possibility also to select desired RVDs for this terminal half-repeat, we generated six different level 3 destination vectors that encode the RVDs NI (A), HD (C), NK (G), NN (G/A), NG (T) and NS (A/C/G/T).

In summary, with these materials in hand, we can generate any *dTALE* gene encoding 17.5 or 20.5 repeats with pre-defined RVD composition in just two subsequent

Figure 1. Continued

assembly are represented as boxes. 5′ and 3′ adaptor modules are shown in yellow. Core repeats are shown in white. The position (rep1–rep10), RVD (NI, HD, NK, NN, NG and NS) and overlaps (boldface font) of every module is written in the box. (**c**) Assembly of a *dTALE* gene with 10.5 repeats via *Bsa*I cut-ligation. Adaptor and core repeats are shown as yellow and white boxes, respectively. Boldface font indicates the unique overlaps of modules created by *Bsa*I cleavage. Red dots indicate cloned core repeats 3–8 that are not displayed. Purple boxes represent the regions encoding the N- and C-terminal parts of the TALE (N- and C-term) including the last half repeat (rep 10.5). Lines connecting the boxes represent the vector backbones that mediate either ampicillin (red line) or kanamycin (blue line) resistance. The *Bsa*I cut-ligation assembles repeat-modules 1–10 into the TALE-*Δ*rep-*Bsa*I-AC vector. The assembled *dTALE* gene does not contain any *Bsa*I recognition sites. Black dots indicates repeat modules of the assembled array (3–8) that are not displayed. (**d**) PCR was used to analyse colonies obtained in a representative *Bsa*I cut-ligation. PCR fragments from 17 colonies (1–17) were separated on a 1% agarose gel and stained with ethidium bromide. The expected size of the PCR product was 1.3 kb and is marked with an asterisk. M: GeneRuler 1 kb DNA Ladder (Fermentas).
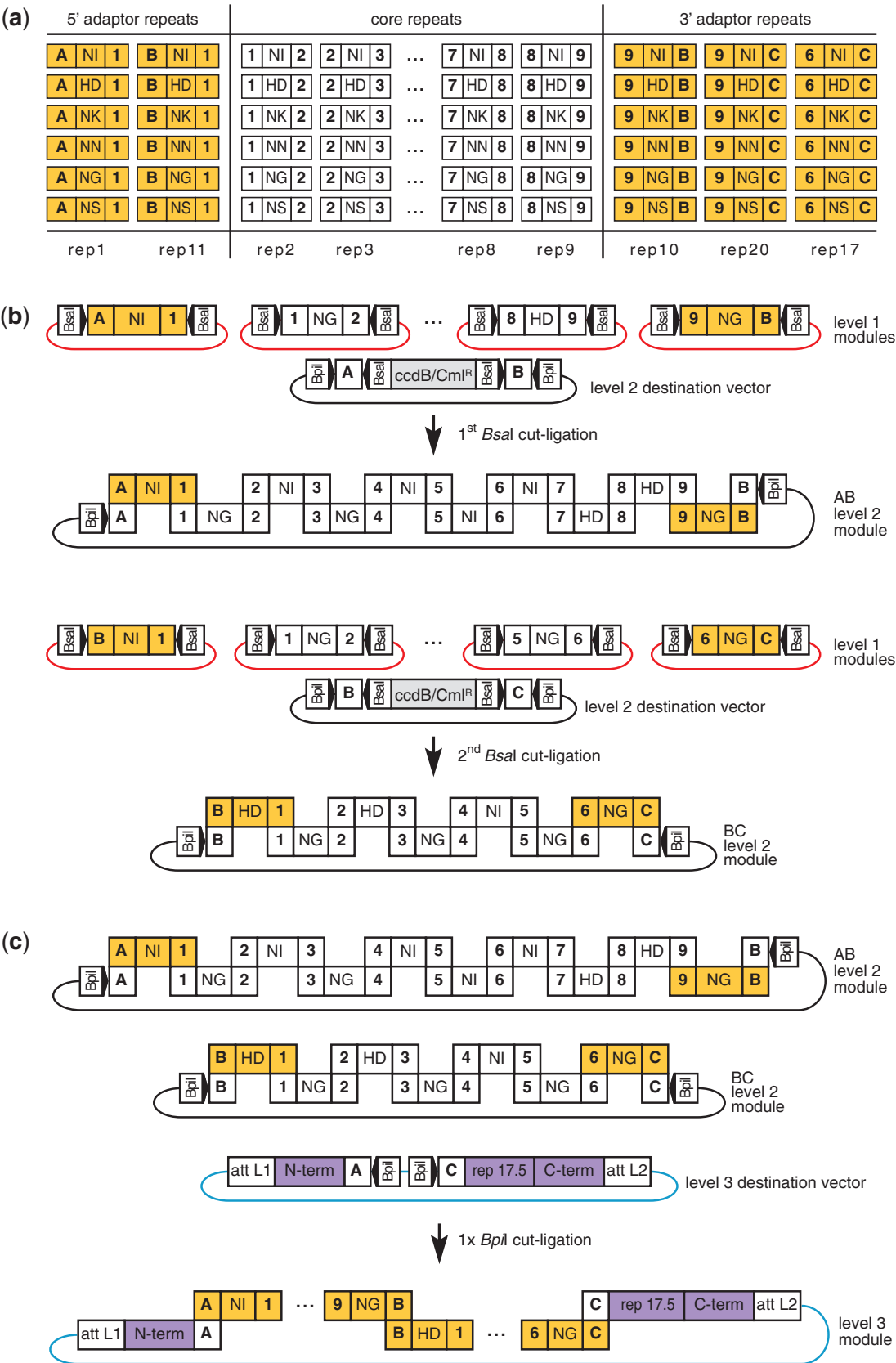
Figure 2. Assembly of a *dTALE* gene with a 17.5 repeat array by two subsequent cut-ligations. If not specified, the shapes and lines are as described in the legend of Figure 1. (**a**) Adaptor and core repeat-modules that are used for *dTALE* assembly. Distinct overlaps of repeat modules are defined in bold font (A, B, C, 1–9). (**b**) Two *Bsa*I cut-ligation facilitate assembly of two distinct level 2 modules that contain 7 and 10 repeat-modules, respectively. *Bsa*I and *Bpi*I recognition sites are shown as pentagons with black arrowheads pointing to their cleavage sites. Level 2 destination

(continued)

cut-ligations. Notably, the majority of naturally occurring *Xanthomonas* TALEs, including the well-studied AvrBs3 protein contains 17.5 repeats. Thus the chosen architecture enables us to directly compare the specificity of *Xanthomonas TALEs* and *in vitro* generated *dTALEs*.

We carried out a number of *Bsa*I cut-ligations and observed that desired AB level 2 modules (10 repeat-modules) and BC level 2 modules (7 repeat-modules) were observed in 50 and 95% (Figure 3a). The subsequent *Bpi*I cut-ligations produced on average 95% clones of correct size (Figure 3b). Sequence analysis of ∼100 assembled *dTALE* genes that showed the correct size in gel electrophoresis did not uncover a single mutation. This extremely high level of sequence fidelity is most likely due to the fact that our approach relies on sequence-validated plasmid DNA and does not involve PCR. Thus the generation of *dTALE* genes by two subsequent cut-ligations worked with high efficiency and fidelity.

Functional analysis of dTALEs with target-optimized RVD composition

To test the functionality and specificity of *in vitro* generated *dTALE* genes, we took advantage of two sequence-related 19-bp AvrBs3 target boxes [up-regulated by TALE AvrBs3 (UPT_{AvrBs3}) boxes] that are present in the pepper *Bs3* promoter ($Bs3_P UPT_{AvrBs3}$ box) (6,17) and the pepper *UPA20* promoter ($UPA20_P UPT_{AvrBs3}$ box) (20). The 19-bp AvrBs3 target sites in *Bs3* and *UPA20* are mostly identical but differ in four basepairs (Supplementary Figure S2). Given that AvrBs3 targets both *UPT* boxes, we wondered if we could generate a *dTALE* gene with an identical number of repeat-modules as *avrBs3* that, however, due to target-adapted design of the encoded RVDs, would specifically activate promoters containing the $UPA20_P UPT_{AvrBs3}$ box but not promoters containing the related $Bs3_P UPT_{AvrBs3}$ box. We generated the *dTALE[UPA20]* via two subsequent cut-ligations and transferred it via recombination into the plant-expression vector pGWB5 (GenBank: AB289768.2). In this *in planta*-expression vector, a given gene is driven by the constitutive cauliflower mosaic virus 35S promoter (35S).

To study target specificity of dTALE[*UPA20*] *in vivo* we made use of two previously established GUS-reporter constructs in which the two distinct UPT_{AvrB3} boxes from the *Bs3* and *UPA20* promoter are embedded in an identical promoter context (5). *Agrobacterium tumefaciens*-mediated delivery of the *TALE* and *dTALE* genes in *Nicotiana benthamiana* leaves showed that dTALE[*UPA20*] produced GUS staining only in combination with the promoter containing the matching $UPA20_P UPT_{AvrBs3}$ box but not with the promoter

containing the highly similar $Bs3_P UPT_{AvrBs3}$ box. In contrast, AvrBs3 produced GUS activity with both promoter constructs (Figure 4). Thus the *dTALE* gene produced by modular cloning was functional *in vivo*. Furthermore the target-adapted RVD composition enabled us to generate a dTALE that, in contrast to AvrBs3, discriminated between two highly similar target sequences.

A modified expression vector simplifies generation of *dTALE* expression constructs

In the above described approach, functional analysis of a given *dTALE* gene requires recombination-based transfer from the entry into a desired expression vector. Implementation of Gateway technology provides a high level of flexibility since assembled *dTALE* genes can be transferred into any Gateway-compatible expression vector. However, in principle, level 2 modules that encode repeat sub-arrays can also be assembled directly by *Bpi*I cut-ligation to a functional *dTALE* gene within the framework of a desired expression vector. This approach allows assembly in two rather than three steps and avoids the rather costly Gateway cloning step. Direct TIIs mediated cloning requires suitable expression vectors that must by devoid of recognition sites for the IIs enzyme that is used in the assembly of level 2 modules. Inspection of the *in planta* expression vector pGWB5 (17 961 bp) revealed nine *Bpi*I, and two *Bsa*I sites. We decided to remove both, *Bpi*I and *Bsa*I recognition sites since this would enable us to use this expression vector in cut-ligations with *Bpi*I or *Bsa*I.

To remove *Bpi*I and *Bsa*I recognition sites, we amplified 11 subfragments of pGWB5, each with primers that overlap with internal *Bpi*I or *Bsa*I sites but contain single nucleotide mismatches to eliminate the given recognition sequence (Supplementary Figure S3). Each of the generated 11 pGWB5-derived PCR-fragments contained at its far end a *Bsa*I recognition site that is cleaved off upon *Bsa*I treatment. We designed the overlaps that are generated upon *Bsa*I cleavage in such a way that the 11 pGWB5-derived fragments would assemble in the desired order in a *Bsa*I cut-ligation, yielding pGWB5*. The newly assembled pGWB5* is identical in its functional elements to pGWB5 but does not contain *Bpi*I or *Bsa*I recognition sites. We used Gateway recombination to transfer the repeat deprived *TALE* gene from the level 3 destination vector TALE-Δrep-*Bpi*I-AC into pGWB5*, yielding pGWB5*-TALE-Δrep-*Bpi*I-AC. This pGWB5-derivative is now a level 3 destination and *in planta* expression vector that can be used for assembly of level 2 modules in *Bpi*I cut-ligations (Figure 5a). Regardless of whether

Figure 2. Continued
vectors contain two pairs of *Bpi*I and *Bsa*I sites producing identical overlaps (A–C). The grey box represents a Gateway cassette with *ccdB* gene and chloramphenicol (Cml^R) resistance marker. The line connecting the *Bsa*I or *Bpi*I sites represents the vector backbone that mediates either ampicillin (red line) or spectinomycin resistance (black line). (**c**) A *Bpi*I cut-ligation facilitates assembly of two level 2 modules into a functional *dTALE* gene. White rectangles denote the two attachment sites (*att*L1 and *att*L2) used for LR recombination. Lines connecting the boxes represent the vector backbones that mediate either spectinomycin (black line) or kanamycin (blue line) resistance. *Bpi*I cleavage creates overlaps in the level 3 destination vectors (A and C) that are complementary to those at the 5′- and 3′-end of the AB and BC level 2 modules, respectively. The generated level 3 module, which encodes a functional dTALE with 17.5 repeats does not contain *Bpi*I recognition sites.
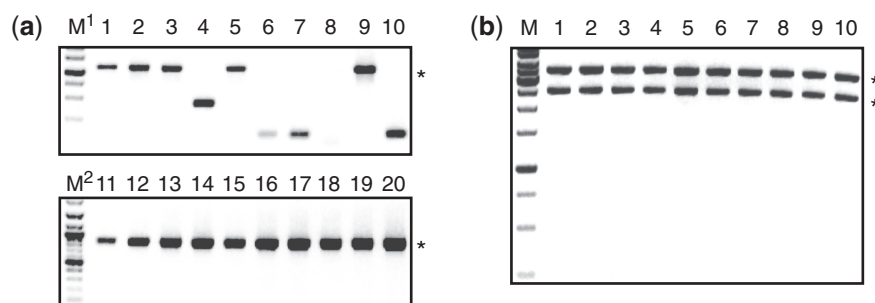
Figure 3. Cut-ligation efficiency for the generation of level 2 and modules. (**a**) Colony PCR was used to analyse colonies obtained in the cloning of AB and BC level 2 modules. PCR fragments from 10 colonies each (1–10 [AB] and 11–20 [BC]) were separated on a 1% agarose gel and stained with ethidium bromide. The expected sizes for the two distinct level 2 modules are 1.2 kb [AB] and 0.9 kb [BC], respectively (asterisk). M¹ and M²: GeneRuler 1 kb and 100 bp DNA ladder from fermentas. (**b**) Full-length *dTALE* genes that were generated by *Bpi*I cut-ligation were analysed by *Stu*I-*Age*I double digest. Fragments obtained from 10 distinct colonies (1–10) were separated on a 1% agarose gel and stained with ethidium bromide. The expected sizes were 2.5 and 3.5 kb (asterisks). M: GeneRuler 1 kb DNA ladder from fermentas.

the assembly of *dTALE* genes is in a three-step procedure (involving Gateway recombination into pGWB5) or a two-step procedure (involving *Bpi*I cut-ligation with pGWB5*-TALE-Δrep-*Bpi*I-AC), the resulting T-DNAs within the given expression vectors are identical.

We assembled ∼50 repeat arrays by *Bpi*I cut-ligations in pGWB5*-TALE-Δrep-*Bpi*I-AC and tested generally two clones. Each of these TALE arrays showed the desired composition of repeat-modules (Figure 5b). Thus the assembly of full-length *dTALE* genes into an expression vector allows rapid, and cost-effective cloning into expression vectors. Using the two-step cloning procedure, we assembled a *dTALE* gene that was identical in its RVDs to AvrBs3. *In planta* analysis showed that the assembled dTALE (dAvrBs3) was functionally indistinguishable from the *Xanthomonas* AvrBs3 protein (Figure 5c). Thus the described two-step generation of *dTALE* expression constructs provides a rapid and cost-efficient approach for generation *dTALE* genes.

DISCUSSION

TIIs mediated assembly of repeat-modules—a highly flexible approach

We developed a rapid, simple and highly cost-efficient approach that facilitates generation of *dTALE* genes that translate into proteins with custom specificity. The approach involves two subsequent cut-ligations fusing individual sequence-validated cloned repeat-modules into a desired array. In the present study, we fused the repeat-modules into the context of a transcription factor. However, the developed cloning approach is highly flexible in several aspects. For example, the designed repeat-modules are flexible with respect to the context into which they can be cloned. One could easily change the level 3 destination vectors to generate TALE nucleases of desired specificity instead of dTALE transcription factors with desired specificity. Our approach is also highly flexible with respect to the size of a repeat array. We have fused two repeat sub-arrays (level 2 modules) containing 7 and 10 repeat-modules into a functional *dTALE* gene. However, given that 10 repeat-modules
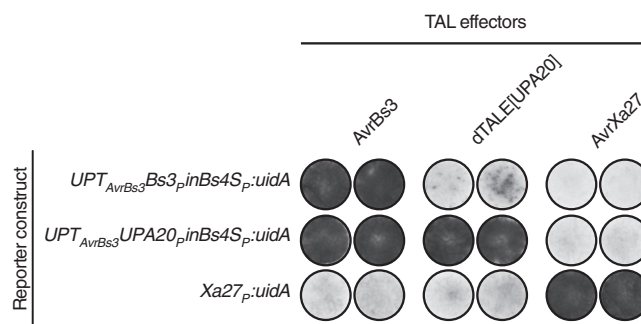

Figure 4. A *dTALE* with optimized RVD composition discriminates between the closely releated sequences present in pepper *UPA20* and *Bs3* promoter. The *uidA* reporter constructs under transcriptional control of the promoters shown at left were delivered via *A. tumefaciens* into *N. benthamiana* leaves in combination with the 35S promoter-driven *TALE* genes indicated above leaf discs. GUS assays were carried out at 40 hpi. Leaf discs were stained with 5-bromo-4-chloro-3-indolyl-β-ᴅ-glucuronic acid, cyclohexylammonium salt (X-Gluc) to visualize activity of the GUS reporter.

could be assembled in a cut-ligation with high efficiency, it should be possible, by creating the appropriate adaptor modules, to generate repeat arrays of 30 or more repeats by fusing multiple level 2 modules. By using identical core repeat-modules for each level 2 module, the amount of effort needed to generate large repeat arrays is limited.

The hierarchical modular cloning system that we present relies on the use of the type IIs enzymes *Bsa*I and *Bpi*I that are used in subsequent cut-ligations and that produce a full-length *dTALE* gene. In fact the described approach also facilitates the generation of higher order constructs that can combine multiple *dTALEs* or other functional units in one construct. For example, terminal *Bsa*I sites flanking the d*TALE* gene in level 3 destination vectors could be used to combine multiple level 3 modules into a desired level 4 vector. Given that genome editing with TALE nucleases generally requires two distinct proteins with different repeat arrays (11), this might represent a useful extension of the current approach.

Our assembly kit allows incorporation of repeats with six distinct RVDs including NK that was recently found to
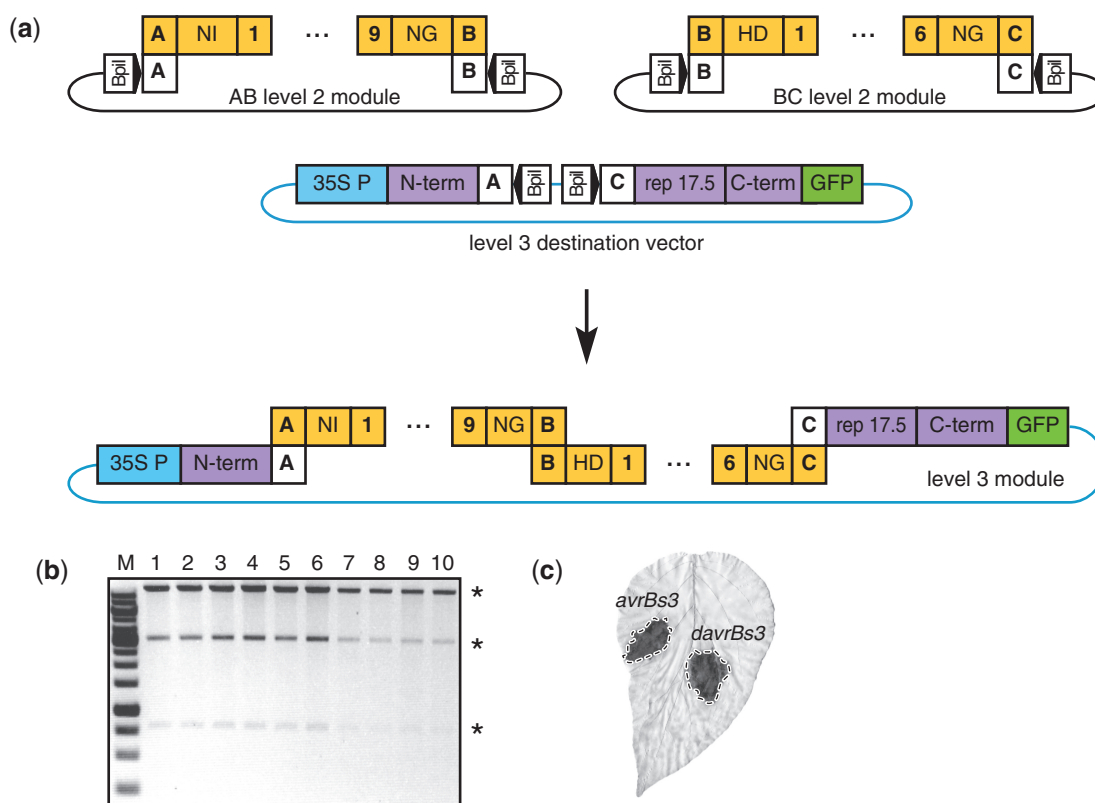
Figure 5. Direct assembly of two *TALE* repeat sub-arrays into a modified plant expression vector that lacks *Bpi*I recognition sites. If not specified, the shapes and lines are as described in the legend of Figure 1. (**a**) A *Bpi*I cut ligation facilitates direct assembly of two level 2 modules into an expression vector. Blue and green rectangles represent the 35S promoter (35S P) and the C-terminal epitope tag (GFP). The blue and black lines that connect the boxes represent the vector backbones that contain spectinomycin and kanamycin resistance markers, respectively. (**b**) Cut-ligation efficiency observed in *Bpi*I-mediated generation of *dTALE* gene expression constructs. Plasmids obtained in the cloning of two level 2 modules into the level 3 destination vector were analysed by *Hin*dIII and *Sac*I digestion. Fragments obtained from 10 colonies (1–10) were separated on a 1% agarose gel and stained with ethidium bromide. The expected sizes were 0.8, 4.3 and 14.6 kb are marked (asterisks). M: GeneRuler 1 kb DNA Ladder (Fermentas). (**c**) Functional analysis of a *dTALE* that was assembled via cut ligation into the modified *in planta* expression vector pGWB5*. The *de novo* assembled *dTALE* gene (*davrBs3*) is identical in its RVD composition to the *Xanthomonas avrBs3* gene. *Agrobacterium tumefaciens* was transformed with either pGWB5* containing *davrBs3* or pGWB5 containing *avrBs3*. Dashed lines mark the inoculated areas. Using *A. tumefaciens* transient transformation the two *TALE* genes were delivered into the leaf of a *Capsicum annuum* genotype that contains the *Bs3* resistance gene (ECW-30R). Two days after infiltration, the leaves were cleared in ethanol to visualize the AvrBs3-triggered and *Bs3*-mediated hypersensitive response (dark areas).

interact with G bases preferentially (4,5) and NS, which has been shown to target A, C, G and T bases with almost identical affinity (2). Thus our assembly kit allows generation of dTALEs with high sequence specificity as well as degeneracy within defined positions of the given target site.

Previous studies have often made use of conserved pairs of restriction enzyme sites flanking the repeat region (e.g. *Bam*HI, *Sph*I) to move repeat arrays into vectors containing a *TALE* backbone (21). These conserved pairs of restriction sites are also present in the backbone sequence used in our modular assembly kit, derived from the *Xanthomonas avrBs3* gene. Thus constructs made via our approach are compatible for cloning into such existing vectors.

Design of TALE repeat arrays—how to maximize target specificity

Target specificity is a major issue in the generation of dTALE repeat arrays and is influenced by the repeat number and type of RVDs. The well-studied TALE

AvrBs3, that binds to a 19-bp target sequence, contains three NS-type RVDs, which have been shown to target A, C, G and T nucleotides with almost identical affinity (2). We assumed that target specificity of AvrBs3 could be improved if RVDs with ambiguous target specificity are replaced with RVDs with tight sequence specificity. To test this hypothesis we made use of two similar 19-bp AvrBs3 target sequences in the pepper *Bs3* and pepper *UPA20* promoter that differ in four basepairs. We generated a dTALE that has the same number of repeats as AvrBs3 but that does not contain RVDs with ambiguous target specificity and that was designed to differentiate between the two similar AvrBs3 target sites in the *Bs3* and *UPA20* promoters. Indeed, this dTALE with optimized RVD composition discriminated between the two similar AvrBs3 target sequences and activated specifically the pepper *UPA20* but not the *Bs3* promoter. These data demonstrate that a target-adapted RVD composition facilitates generation of repeat arrays with high specificity. Previously, we engineered two AvrBs3 derivatives with four additional repeat units that target a 23-bp target

box and that activate specifically either the pepper *Bs3* or *UPA20* promoter (5). Thus we could show that RVD composition as well as the size of the repeat array affects target specificity.

Alternative approaches for assembly of TALE repeats

A most recent manuscript provides an alternative protocol for modular assembly of *dTALE* genes that encode 12.5 repeats (13). The approach relies on four cloned repeat monomers (NI, HD, NN, NG) that are linked to adaptors with type IIs recognition sites via PCR, which facilitates assembly of a desired repeat array into a *TALE* gene deletion construct by two sequential ligations. Thus this approach facilitates assembly of TALE arrays using four cloned repeat monomers and *TALE* gene deletion constructs, respectively. At first sight, the method is very attractive, because only a few gene constructs are needed and thus the upfront work is quite limited. On the other hand, the assembly procedure is rather complex and laborious and involves amplification of twelve individual repeat-modules, subsequent PCR product purification, *Bsa*I cleavage (generates distinct 4-bp overlaps for each module), purification of cleavage products, ligation of three sub-arrays (each with four repeats), gel purification and subsequent PCR amplification of the three repeat sub-arrays, subsequent PCR product purification, cleavage of repeat sub-arrays and vector (contains *TALE* gene deletion that lacks the repeat array), purification of cleavage products and finally ligation of three sub-arrays into the *TALE* gene backbone. In summary, the approach involves two rounds of PCRs, five purification steps and two ligations. Our assembly kit facilitates assembly of 17.5 TALE repeats in two subsequent single-tube cut-ligations. Importantly our approach does not require purification or PCR steps, instead relying on ligation of sequence-validated plasmid inserts. Thus assembly of *dTALE* genes with this approach is less laborious and results in arrays with 17.5 instead of 12.5 repeats. Since our approach relies on ligation of sequence-validated plasmids it seems likely that the accuracy of these *dTALE* genes is higher as compared to *dTALE* genes that were produced by ligation of amplification products that have undergone two subsequent PCRs. While assembly of *dTALE* genes by our assembly kit has in many ways advantages, it needs to be emphasized that our kit required 80 gene constructs (72 repeat-modules, two level 2 destination and six level 3 destination vectors) and therefore had a substantial upfront workload. In its current state, our kit enables the production of many *dTALE* genes quickly, at low cost and with high efficiency. Since the established assembly approach does not involve any complex manual procedures it could be carried out by a pipetting robot. This would facilitate implementation of TALE technology in high throughput applications.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Klug,A. (2010) The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annu. Rev. Biochem.*, **79**, 213–231.
2. 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.
3. Moscou,M.J. and Bogdanove,A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, **326**, 1501.
4. 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.* (2011) A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.*, **29**, 143–148.
5. 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.
6. Römer,P., Strauss,T., Hahn,S., Scholze,H., Morbitzer,R., Grau,J., Bonas,U. and Lahaye,T. (2009) Recognition of AvrBs3-like proteins is mediated by specific binding to promoters of matching pepper *Bs3* alleles. *Plant Physiol.*, **150**, 1697–1712.
7. Römer,P., Recht,S., Strauß,T., Elsaesser,J., Schornack,S., Boch,J., Wang,S. and Lahaye,T. (2010) Promoter elements of rice susceptibility genes are bound and activated by specific TAL effectors from the bacterial blight pathogen, *Xanthomonas oryzae* pv. *oryzae*. *New Phytol.*, **187**, 1048–1057.
8. Christian,M., Cermak,T., Doyle,E.L., Schmidt,C., Zhang,F., Hummel,A., Bogdanove,A.J. and Voytas,D.F. (2010) TAL effector nucleases create targeted DNA double-strand breaks. *Genetics*, **186**, 757–761.
9. Li,T., Huang,S., Jiang,W.Z., Wright,D., Spalding,M.H., Weeks,D.P. and Yang,B. (2011) TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.*, **39**, 359–372.
10. Mahfouz,M.M., Li,L., Shamimuzzaman,M., Wibowo,A., Fang,X. and Zhu,J.-K. (2011) De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc. Natl Acad. Sci. USA*, **108**, 2623–2628.
11. Urnov,F.D., Rebar,E.J., Holmes,M.C. and Gregory,H.S.Z.D. (2010) Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.*, **11**, 636–646.
12. Ramirez,C.L., Foley,J.E., Wright,D.A., Muller-Lerch,F., Rahman,S.H., Cornu,T.I., Winfrey,R.J., Sander,J.D., Fu,F., Townsend,J.A. *et al.* (2008) Unexpected failure rates for

modular assembly of engineered zinc fingers. *Nat. Methods*, **5**, 374–375.

13. Zhang,F., Cong,L., Lodato,S., Kosuri,S., Church,G.M. and Arlotta,P. (2011) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat. Biotechnol.*, **29**, 149–153.

14. 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.

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

16. Koncz,C. and Schell,J. (1986) The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimeric genes carried by a novel type of *Agrobacterium* binary vector. *Mol. Gen. Genet.*, **204**, 383–396.

17. 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.

18. Engler,C., Gruetzner,R., Kandzia,R. and Marillonnet,S. (2009) Golden Gate Shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS ONE*, **4**, e5553.

19. Weber,E., Engler,C., Gruetzner,R., Werner,S. and Marillonnet,S. (2011) A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE*, **6**, e16765.

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

21. Sugio,A., Yang,B., Zhu,T. and White,F.F. (2007) Two type III effector genes of *Xanthomonas oryzae* pv. *oryzae* control the induction of the host genes *OsTFIIAγ1* and *OsTFX1* during bacterial blight of rice. *Proc. Natl Acad. Sci. USA*, **104**, 10720–10725.