nature COMMUNICATIONS

ARTICLE

Received 7 Nov 2013 | Accepted 13 Feb 2014 | Published 11 Mar 2014

DOI: 10.1038/ncomms4447

A TAL effector repeat architecture for frameshift binding

Annekatrin Richter[1,*], Jana Streubel[1,*], Christina Blücher[1], Boris Szurek[2], Maik Reschke[1], Jan Grau[3] & Jens Boch[1]

Transcription activator-like effectors (TALEs) are important *Xanthomonas* virulence factors that bind DNA via a unique tandem 34-amino-acid repeat domain to induce expression of plant genes. So far, TALE repeats are described to bind as a consecutive array to a consecutive DNA sequence, in which each repeat independently recognizes a single DNA base. This modular protein architecture enables the design of any desired DNA-binding specificity for biotechnology applications. Here we report that natural TALE repeats of unusual amino-acid sequence length break the strict one repeat-to-one base pair binding mode and introduce a local flexibility to TALE–DNA binding. This flexibility allows TALEs and TALE nucleases to recognize target sequence variants with single nucleotide deletions. The flexibility also allows TALEs to activate transcription at allelic promoters that otherwise confer resistance to the host plant.

[1] Department of Genetics, Martin Luther University Halle-Wittenberg, Weinbergweg 10, D-06120 Halle (Saale), Germany. [2] UMR 186 IRD-UM2-Cirad 'Résistance des Plantes aux Bioagresseurs', BP 64501, 34394 Montpellier cedex 5, France. [3] Institute of Computer Science, Martin Luther University Halle-Wittenberg, von-Seckendorff-Platz 1, D-06120 Halle (Saale), Germany. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to J.B. (email: jens.boch@genetik.uni-halle.de).

Transcription activator-like effectors (TALEs) employ a programmable DNA-binding domain and have revolutionized biotechnology approaches that use sequence-specific DNA-binding proteins[1–3]. Natural TALEs are injected by plant pathogenic *Xanthomonas* spp. bacteria via a type-III secretion system into plant cells[4]. Inside the plant cell they function as transcription factors that activate target gene expression and support bacterial colonization[2,5]. TALEs contain a unique DNA-binding domain of tandem near-identical 34-amino-acid repeats. The repeats are highly conserved and mainly differ in two adjacent amino acids (positions 12 and 13) termed repeat-variable diresidue (RVD)[5,6]. In the array each repeat recognizes one nucleotide in the target DNA sequence, and the RVD specifies which base is bound[6,7]. In addition to the repeat region, the N-terminal part of TALEs contains four degenerated repeats (termed repeat 0, −1, −2 and −3) that contribute to DNA-binding[8]. Repeat −1 specifies thymine that typically precedes TALE target sequences[7,9,10]. The C-terminal part of TALEs contains two functional nuclear localization signals as well as an acidic activation domain that is important for gene activation[4,11].

Xanthomonas oryzae pv. *oryzae* (*Xoo*) and *Xanthomonas oryzae* pv. *oryzicola* (*Xoc*) cause bacterial leaf blight and bacterial leaf streak, respectively, which are two of the most devastating diseases of the staple crop rice[12,13]. Both pathogens contain a particularly large number (7–28) of *TALE* genes per strain[14] and the TALE-dependent modulation of host gene expression profiles is an important feature of *Xoo* and *Xoc* diseases. In some cases, loss of a single *TALE* gene severely compromised bacterial virulence[5]. The best-studied group of TALE virulence targets is the *SWEET* gene family in rice. Three members of this gene family are targets of several TALEs from different *Xoo* strains. *OsSWEET11* and *OsSWEET13* are targeted by PthXo1 and probably PthXo2, respectively, whereas *OsSWEET14* is targeted by four different TALEs, TalC, AvrXa7, PthXo3 and TAL5, originating from different *Xoo* strains[15–21]. Furthermore, it was shown that *OsSWEET12*, *OsSWEET13* and *OsSWEET15* support bacterial colonization if induced by designer TALEs[16,19]. Considering that >100 *Xanthomonas* TALEs are known the overall number of identified plant target genes is low. One approach to overcome this is by using computational algorithms to predict possible TALE targets[22–24]. Using the TALE RVD–DNA code the promoter sequences of host plants are scanned for potential TALE target sites. Promising target candidates contain a TALE target site and are TALE-dependently activated[23,24].

Plants have evolved resistances that are based on TALE-dependent activation of genes that trigger a resistance response towards *Xanthomonas* strains delivering the matching TALE[25–27]. A different resistance mechanism is based on mutation of potential TALE target DNA sequences[17,20,28,29]. These mutations can render virulence targets non-responsive to TALEs, thereby efficiently preventing the contribution of a given TALE to bacterial virulence. If target gene upregulation is important for virulence of the pathogen, the plants become resistant. Accordingly, rice varieties carrying insertions, deletions or substitutions in the promoter sequences of *OsSWEET11* and *OsSWEET12* are not susceptible to *Xoo* strains carrying the TALEs PthXo1 and PthXo2, respectively[15,17,28–30].

The modular TALE architecture allows a free combination of repeats to generate any desired DNA-binding specificity[7,31,32]. This flexibility resulted in adoption of the TALE repeats as specific DNA-binding domain that can be fused to executor domains to generate different biotechnological tools[2]. The most widespread use is based on fusions of TALEs with a nuclease domain (termed TALEN) to edit eukaryotic genomes at specific sites[3,31,33,34]. Other executor domains were applied for gene activation, gene repression, chromatin modification, fluorescent tagging of chromosomal loci and chromatin affinity purification[35–40].

The three-dimensional (3D) structures of TALE–DNA complexes have been solved[8–10,41]. TALE repeats form a right-handed superhelix that wraps around the DNA double strand. The helix–loop–helix structure of a single repeat exposes the RVD amino acid at position 13 to interact with a leading strand base from the DNA major groove. In contrast, the RVD amino acid at position 12 stabilizes the RVD loop by interaction with the carbonyl of position 8 of the same repeat[9,10,41]. This highly regular repeat architecture is likely necessary to position consecutive repeats correctly to a continuous string of DNA bases. The 3D structures show that in a TALE repeat non-RVD amino acids mediate inter- and intra-repeat interactions as well as connections to the DNA phosphate backbone. This probably supports the correct positioning of the repeat array to the DNA[9,10,41]. The TALE 3D structures also suggest that significant changes in the repeat length may have a profound impact on DNA binding. Two fairly common exceptions to the typical 34-amino-acid length of TALE repeats exist. These are either 33-amino-acid repeats that have a deletion of the RVD amino acid 13 or 35-amino-acid repeats containing a proline following amino acid 32. In both cases the DNA-binding specificity of TALEs is not altered[7,36,42]. Thirty-five-amino-acid repeats are also typical in TALE homologues from *Ralstonia solanacearum* for which a RVD-guided DNA-binding activity analogous to *Xanthomonas* TALEs was shown[43,44]. Interestingly, some naturally occurring *Xanthomonas* TALEs possess a single repeat of aberrant length besides the canonical 33–35-amino-acid repeats[5,45]. These aberrant repeats consist of 30 amino acids (with a 4-amino-acid deletion in the second helix of the repeat), 39 or 40 amino acids (duplication in the second helix of the repeat), or 42 amino acids (duplication in the first helix of the repeat)[5]. It is not known how repeats that differ in length from the normal 33–35 amino acids influence the DNA-binding behaviour of TALEs.

Here we analyse how the insertion of a repeat of aberrant length in the canonical repeat array influences TALE function. We find that such TALEs exhibit two possible binding conformations, the normal one and a novel one that tolerates single nucleotide deletions in the target sequence. We show that this binding behaviour expands the recognition specificity of TALEs and TALENs, respectively. Furthermore, we find that the flexible TALE-binding behaviour provides a potential evolutionary solution for *Xanthomonas* to overcome plant resistance.

Results

Natural TALEs with aberrant repeats. AvrXa7 and PthXo3 are natural TALEs from *Xoo* with 25.5 and 28.5 repeats, respectively, that both contain one aberrant repeat of 39 amino acids approximately in the middle of the repeat array (Fig. 1; Supplementary Figs 1 and 2 and Supplementary Table 1). Both TALEs are important virulence factors that support growth of *Xoo* strains on rice[15], which demonstrates that they are functional. Indeed, both TALEs induce expression of the rice sugar exporter *OsSWEET14* by binding to overlapping target boxes (Fig. 1b and Supplementary Figs 1 and 2)[15,18,19]. In contrast, computational predictions of AvrXa7 and PthXo3 target sequences rank the PthXo3 site in the *OsSWEET14* promoter at position 119 (ref. 23). We noticed a high number of non-matching RVD base combinations in the 3′ part of the PthXo3 target box following the aberrant repeat (Fig. 1c,d and Supplementary Fig. 2). If the C-terminal half of the repeat array is shifted one nucleotide upstream by looping out the long repeat, the PthXo3 target box in *OsSWEET14* will rank at position 1 in target predictions (Fig. 1c and Supplementary Table 2; ref. 23).

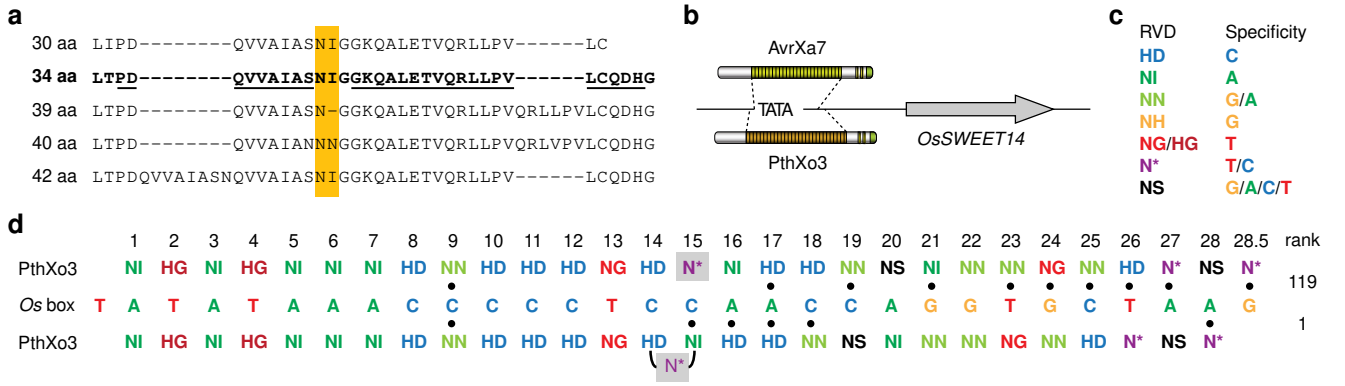


Figure 1 | Aberrant repeat length variants from *Xanthomonas* spp. TALEs. (**a**) Alignment of amino acid (aa) sequences of natural TALE repeats with different length. The repeat-variable diresidues (RVDs) are boxed in yellow. A typical 34-aa repeat is in bold face and the residues forming the two α-helices are underlined. (**b**) Cartoon of AvrXa7 and PthXo3 binding to their overlapping target sequences at the TATA-box in the *OsSWEET14* (*Os11g31190*) promoter. (**c**) TALE RVDs and their DNA base specificities. (**d**) PthXo3 RVDs with the aberrant repeat in normal or looped-out conformation aligned to the *OsSWEET14* promoter target sequence (*Os* box). A dot indicates a non-matching RVD–base combination. The rank of both alignments in a target site prediction is indicated.

To test the recognition specificities of AvrXa7 and PthXo3, we build reporter constructs with an optimal target box according to the TALE specificity code, and derivatives with nucleotide deletions (-1, -2) or insertions ($+1$, $+2$) at the nucleotide position behind the one corresponding to the aberrant repeat, respectively (Supplementary Figs 1a and 2a). The boxes are inserted upstream of a minimal promoter that has no basal activity[7]. In addition, reporter constructs with either the natural box of AvrXa7 in front of the minimal promoter or the *OsSWEET14* promoter fragment containing the natural AvrXa7/PthXo3 box are used (Supplementary Figs 1a and 2a, ref. 21). The reporter constructs are then co-transformed with expression constructs of AvrXa7 or PthXo3 using *Agrobacterium* into leaves of *Nicotiana benthamiana* plants. β-glucuronidase (GUS) assays reveal that both TALEs induce expression of the reporter containing either one of the target boxes although they exhibit several non-matching RVD–base combinations (Supplementary Figs 1b–g and 2b–h). Possible explanations for this surprising result are that AvrXa7 and PthXo3 contain a large number of repeats, several RVDs with broad specificity like NS, NN and N*, and one repeat of aberrant length. We aim to test the role of the aberrant repeat in a more controlled TALE design.

Aberrant repeats permit a novel TALE–DNA-recognition mode. We construct artificial TALEs[46] with a 17.5-repeat array that is designed to result in a maximum of non-matching RVD–base combinations upon a possible frameshift in the target sequence. TALEs are assembled with and without single aberrant repeats of 30-, 40- and 42-amino-acid length at position 8 (Figs 1a and 2a and Supplementary Table 1), and analysed *in planta* for the activation of GUS reporters with optimal and frameshift (-1, -2, $+1$ and $+2$) target boxes (Fig. 2a,b). The aberrant repeat sequences including their RVDs correspond to the natural *X. oryzae* TALE sequences (Supplementary Table 1).

TALEs with or without an aberrant repeat trigger GUS activity with the reporter construct containing the optimal box. This indicates that all TALEs tested recognize the optimal box and that aberrant repeats function similarly to normal ones in a repeat array. Significantly, only TALEs with an aberrant repeat result also in strong GUS activity with reporter constructs containing the -1 box (Fig. 2b). The same effect is apparent in artificial TALEs with a different repeat composition (Supplementary

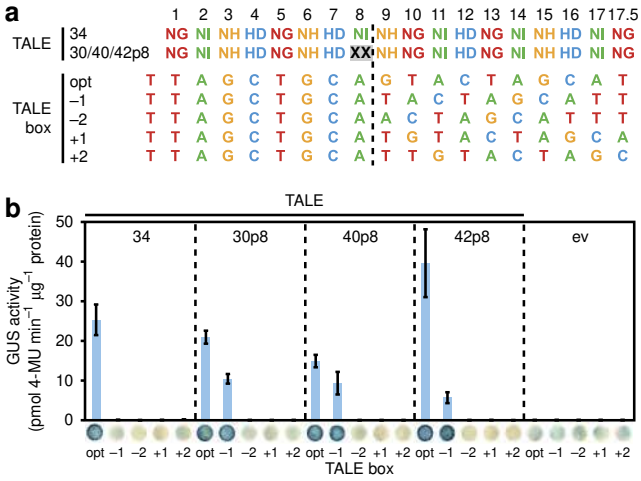


Figure 2 | Aberrant repeats allow a flexible recognition of target DNA sequences with a -1 nucleotide frameshift. (**a**) RVDs of artificial TALEs and target boxes. A TALE with all 34-aa repeats or TALEs with an aberrant 30-aa NI-repeat, 40-aa NN-repeat and 42-aa NI-repeat, respectively, at position 8 are constructed. An optimal target box (opt) and derivatives with deletions or insertions of one or two nucleotides ($-$ or $+1$ or 2) after position 8 (dashed line) are fused to a minimal promoter and a promoter-less GUS reporter gene. (**b**) GUS assays of TALEs and reporter constructs ($n=3$). 35S-driven GFP expression serves as empty vector (ev) control in quantitative and qualitative assays. Error bars indicate the s.d. in the quantitative assay. One representative leaf disk of the qualitative assay is shown.

Fig. 3). Thus, a TALE containing exclusively canonical 34-amino-acid repeats is not able to recognize a -1 box, because of many (in this case nine) non-matching RVD–base combinations in either the front or rear part of the box, depending on how the TALE RVDs are aligned to the box. In contrast, a TALE with an aberrant repeat efficiently activates the reporter. These data suggest that an aberrant repeat confers a flexible binding to TALEs. Interestingly, this flexible binding mode is possible with different aberrant repeats that are shorter (30 amino acids) or longer (39, 42 amino acids). The -2, $+1$ and $+2$ boxes are not recognized by any of the TALEs, indicating that other binding modes are not supported.

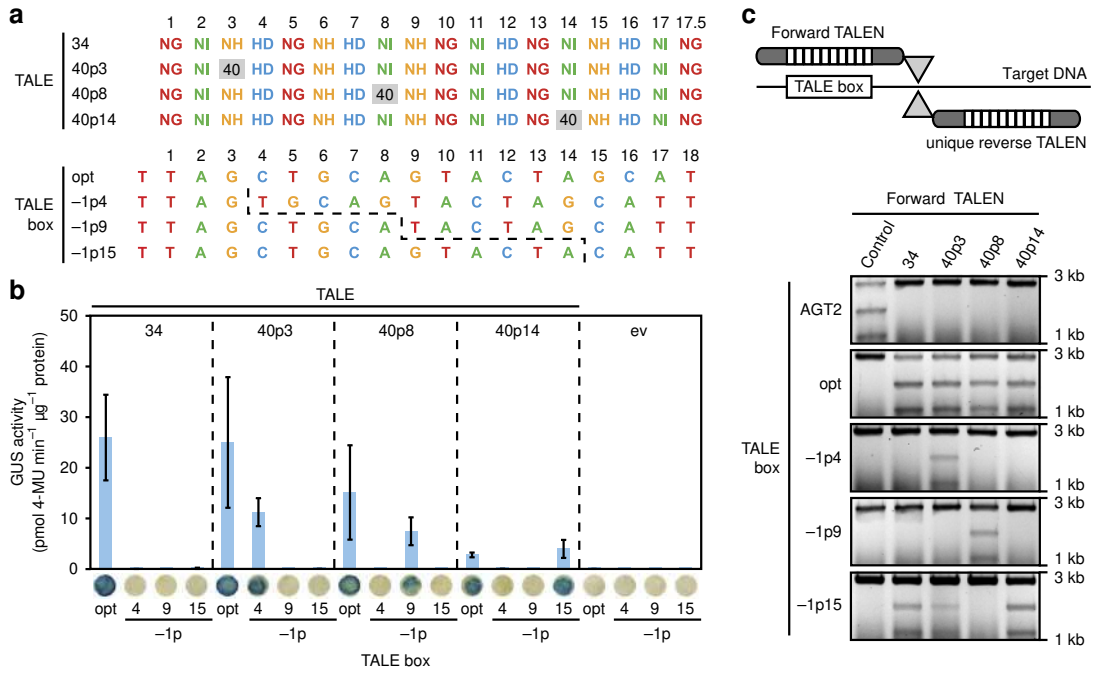


Figure 3 | Aberrant repeats function at different positions in the repeat array of TALEs and TALENs. (**a**) TALE RVDs and target boxes. Artificial TALEs or TALENs are constructed with all 34-aa repeats or an aberrant 40-aa NN-repeat inserted at position 3 (40p3), position 8 (40p8) or position 14 (40p14). The NN-repeat recognizes both G and A DNA bases. The boxes are either perfectly matching the specificity of normal repeats (opt) or have one base pair deleted to the right of the dashed line at position 4 (− 1p4), position 9 (− 1p9) or position 15 (− 1p15). (**b**) GUS assays of TALEs and reporter constructs ($n = 3$). 35S-driven GFP expression serves as empty vector (ev) control in quantitative and qualitative assays. Error bars indicate the s.d. in the quantitative assay. One representative leaf disk of the qualitative assay is shown. (**c**) Cartoon of TALENs bound to DNA and *in vitro* TALEN restriction assay. The TALEN pairs are placed such that the FokI domain (grey triangle) can dimerize and cut the DNA. The unique reverse TALEN is constant in all assays. The control TALEN recognizes a different target box (AGT2). Target DNA is incubated with *in vitro*-transcribed and translated TALEN pairs. Restriction fragments are documented on agarose gels.

Aberrant repeats function at different repeat array positions. As in nature single aberrant repeats are positioned more or less in the central region of the repeat array (Supplementary Table 1), we investigate whether TALEs tolerate only this arrangement. We insert the aberrant 40-amino-acid repeat at different positions (positions 3, 8 or 14) in the previously used 17.5 TALE repeat array that is highly susceptible to frameshift in its target sequence (Fig. 3a). As target boxes we use either the optimal one or sequences with frameshift at positions 4, 9 and 15, respectively (Fig. 3a). GUS reporter assays reveal that TALEs that carry the aberrant repeat at any of the three positions are functional. All TALEs with aberrant repeats recognize the optimal box and the one with a frameshift following the position corresponding to the aberrant repeat in the TALE (Fig. 3b). However, the TALE with an aberrant repeat at position 14 displays only weak activity on both the optimal and the corresponding − 1 frameshift box (Fig. 3b). This suggests that it is not favourable for the overall activity of TALEs if only a small number of repeats (in this case four) follow an aberrant repeat in the array.

Flexible TALENs with aberrant repeats. TALENs have become state-of-the-art tools for genome editing[3]. TALENs are TALE–DNA-binding domain fusions to the *Fok*I endonucleolytic domain, which act in pairs to enable *Fok*I dimerization and DNA cleavage[3]. We test whether aberrant repeats can change the DNA recognition behaviour of TALENs. TALENs with and without an aberrant repeat are assembled. The aberrant 40-amino acid repeat is placed at positions 3, 8 and 14, respectively, in a TALEN of 17.5 repeats (Fig. 3a). These 'forward' placed TALENs are combined with a common 'reverse' placed TALEN with 34-amino-acid repeats (Fig. 3c and Supplementary Fig. 4b) and linear DNA fragments containing the target boxes in *in vitro* restriction assays. The optimal box is cleaved using TALENs with or without an aberrant repeat emphasizing that the normal binding mode is also supported in TALEN assays. In contrast, the boxes with − 1 frameshift at position 4 and 9 are only cleaved in the presence of the TALENs with an aberrant repeat at position 3 and 8, respectively (Fig. 3c and Supplementary Fig. 5). This is in accordance with our observation for TALEs before (Fig. 3b) that aberrant repeats infer a local flexibility to the repeat array.

In contrast, the box with the − 1 frameshift at position 15 is recognized not only by the TALEN with an aberrant repeat at position 14 but also by the normal all 34-amino-acid repeat TALEN, and to a lesser degree by the TALEN with an aberrant repeat at position 3. Apparently, these TALENs tolerate three non-matching RVD–base combinations at the end of the repeat array (Fig. 3c). This mismatch tolerance of rear repeats in TALENs was not observed in the previous experiments using TALEs in reporter gene activation assays (Fig. 3b). It suggests that binding of the rear TALE repeats to DNA is less important for dimerization of the nuclease domains than for function of the natural activation domain.

Furthermore, we test whether the TALEN with the aberrant 40-amino-acid repeat at position 8 can perform cleavage of − 2, + 1 or + 2 frameshift boxes (Supplementary Fig. 4). We observe that the TALENs with or without the long repeat do not exhibit activity on these boxes (Supplementary Fig. 4c), suggesting that other binding modes are not supported by TALENs similarly to our observation for TALEs (Fig. 2). In summary, our experiments

demonstrate that aberrant repeats initiate a flexible DNA-binding behaviour *in vivo* and *in vitro*, and that they can be used to expand the recognition specificities of artificial TALEs as well as TALENs.

The aberrant repeat is excluded from the interaction. So far, it is not clear which repeat of the repeat array is excluded from the interaction, for example by looping out. Our results allow several possible explanations: either the aberrant repeat itself is excluded or the repeat up- or downstream of the aberrant repeat. Therefore, we test TALEs with aberrant repeats at position 8 combined with target boxes (-1) deleted in either the nucleotide at position 7, 8 or 9 in GUS reporter assays (Fig. 4a). We expect that the TALE will yield the highest reporter activity in combination with the -1 box that has the deletion exactly in opposite to the repeat that is excluded, because the other boxes will produce at least one non-matching RVD–base combination. Indeed, the target sequence with the nucleotide deleted at position 8 shows the highest activity of the three -1 boxes for the TALEs with 40- and 42-amino-acid aberrant repeats (Fig. 4b). The TALE with a 30-amino-acid aberrant repeat has a similarly high activity with the -1 boxes deleted at position 8 and 9, respectively (Fig. 4b). This suggests that it is the aberrant repeat itself that is excluded in frameshift binding.

To corroborate this finding, we use a TALE with a 40-amino-acid aberrant repeat placed at position 3 in either a 17.5 repeat or an 11.5 repeat array (Fig. 4c). It has been described that the initial repeats have a stronger impact on the overall TALE binding than later repeats[37,47]. Therefore, we reason that these TALEs will have a clear preference for the -1 box in which the nucleotide is missing that corresponds to the repeat that is excluded. The TALEs were combined with an optimal target box or boxes deleted in the nucleotide at position 3 or 4 (Fig. 4c). Indeed, the TALEs induce a higher reporter activity with the box carrying a -1 deletion at position 3 that corresponds to the position of the aberrant repeat than with the box with the deletion at position 4 (Fig. 4d). The shorter TALEs have a significantly weaker activity suggesting that the overall DNA recognition was partly compromised (Fig. 4d). In summary, we postulate that the aberrant repeat itself loops out of the repeat array when the TALE is bound to a target sequence with a single nucleotide deletion, but is inserted into the array when bound to an optimal box (Fig. 7).

Tandem aberrant repeats are not flexible. We analyse whether tandem aberrant repeats are accepted in a TALE repeat array. For this, we generate a TALE with aberrant 40-amino-acid repeats at position 8 and 9 in the frameshift-sensitive 17.5 repeat array used before (Fig. 5a). The TALE with the two aberrant repeats recognizes exclusively the box with a -1 frameshift at position 9 and neither the optimal one nor a -2 or $+1$ frameshift box (Fig. 5b). Apparently, the tandem aberrant repeats can only be compensated by one repeat being excluded from the interaction. This suggests that two neighbouring 40-amino-acid repeats can neither loop out simultaneously nor be arranged in the consecutive array of normal repeats. It cannot be excluded that aberrant repeats destabilize the TALE protein structure in certain repeat arrangements although we do not detect strong differences in protein amounts of the artificial TALEs used here and in the previous experiments with or without aberrant repeats, respectively (Supplementary Fig. 6).

Aberrant repeats can participate in TALE–DNA recognition. We want to clarify whether the RVD of the aberrant repeat participates in DNA base recognition when the TALE binds in the regular fashion to an optimal box. For this, we compare how

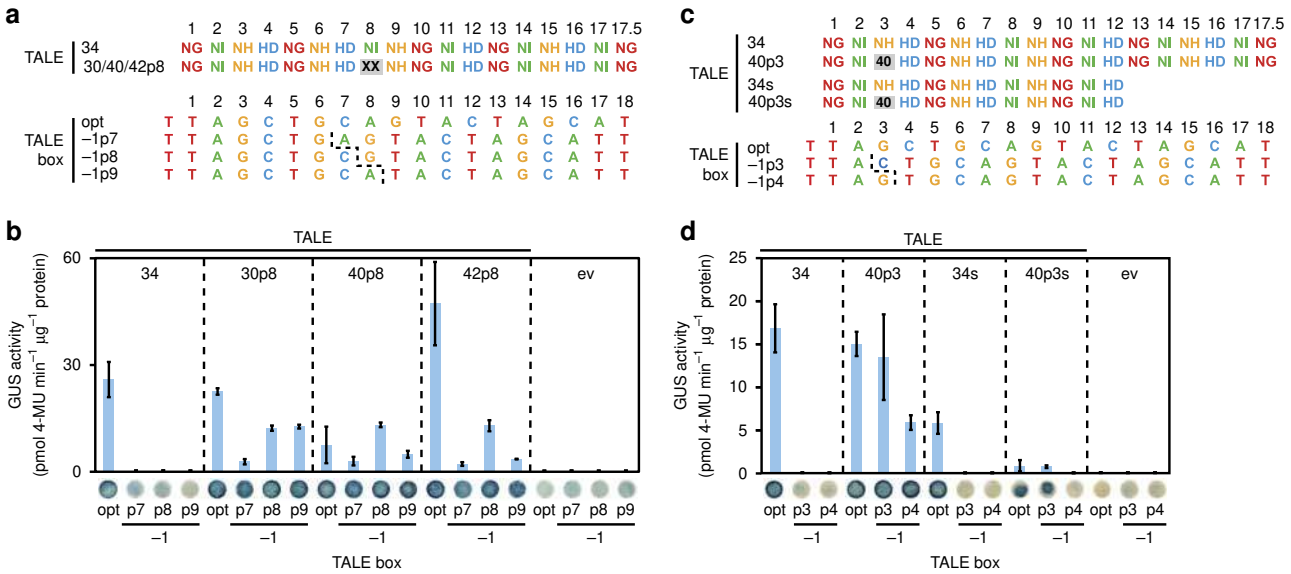


Figure 4 | Aberrant repeats allow recognition of target box frameshift close to their position. (**a**) RVDs of artificial TALEs and target boxes. A TALE with all 34-aa repeats or TALEs with an aberrant 30-aa NI-repeat, 40-aa NN-repeat and 42-aa NI-repeat, respectively, at position 8 are constructed. TALE boxes are either perfectly matching the specificity of normal repeats (opt) or have one base pair deleted to the right of the dashed line at position 7 (-1p7), position 8 (-1p8) or position 9 (-1p9). (**b**) GUS assays ($n=3$) of TALEs and reporter constructs described in (**a**). (**b**,**d**) 35S-driven GFP expression serves as empty vector (ev) control in quantitative and qualitative assays. Error bars indicate the s.d. in the quantitative assay. One representative leaf disk of the qualitative assay is shown. (**c**) TALE RVDs and target boxes. Artificial TALEs with either 17.5 or 11.5 repeats are constructed with a normal 34-aa NH-repeat or an aberrant 40-aa NN-repeat (grey box) inserted at position 3. The target boxes are either perfectly matching the specificity of normal repeats (opt) or have one base pair deleted to the right of the dashed line at position 3 (-1p3) or position 4 (-1p4). (**d**) GUS assays ($n=3$) of TALEs and reporter constructs described in (**c**).

TALEs with 34-amino-acid repeats and TALEs with an aberrant repeat deal with non-matching bases around the position of the aberrant repeat. We expect that a TALE with an aberrant repeat will display a reduced activity at the mismatch boxes if the RVD of the aberrant repeat participates in base recognition. TALEs with only 34-amino-acid repeats or an aberrant repeat of 30-, 40- or 42-amino-acid length, respectively, at position 8 are used. The aberrant repeat RVDs are NI or NN, which are compatible with the base adenine, but not thymine. We combine the TALEs with target boxes that contain 1–3 non-matching bases starting at position 8 or 9 (Fig. 6a). The TALE with normal 34-amino-acid repeats trigger significantly decreasing reporter activity with increasing number of mismatches (Fig. 6b). The TALEs with aberrant repeats show a very similar pattern. In particular, at the box with one mismatch at the position of the aberrant repeat (box 8.1; Fig. 6b) reporter activity is significantly less than at the optimal one. This indicates that all tested aberrant repeats contribute to DNA–base recognition when the TALE binds an optimal box (Fig. 7).

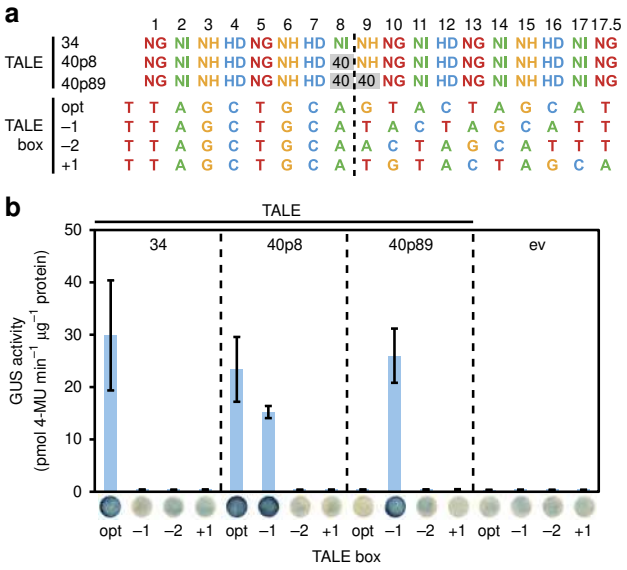


Figure 5 | Two aberrant repeats can be combined in tandem. (**a**) RVDs of artificial TALEs and target boxes. A TALE with all 34-aa repeats or TALEs with one or two 40-aa NN-repeats at position 8 (40p8) or 8 and 9 (40p89) are constructed. An optimal target box (opt) and derivatives with deletions of one or two nucleotides (−1 or −2) or an insertion of one nucleotide (+1) after position 8 (dashed line) are fused to a minimal promoter and a promoter-less GUS reporter gene. (**b**) GUS assays of TALEs and reporter constructs ($n=3$). 35S-driven GFP expression serves as empty vector (ev) control in quantitative and qualitative assays. Error bars indicate the s.d. in the quantitative assay. One representative leaf disk of the qualitative assay is shown.

Aberrant repeats influence AvrXa7 and PthXo3 target range. The natural aberrant repeats provide an extended flexibility to the repeat array that is not present in TALEs with standard repeats, and it is tempting to speculate that this has evolved to benefit *Xanthomonas* virulence. One possibility is that aberrant repeats enable *Xanthomonas* to recognize promoter variants in different plant cultivars or species. We investigate whether there are natural insertion/deletion (indel) mutations in the *OsSWEET14* promoter region targeted by PthXo3. We compare this region in full genomic alignments of *Oryza sativa japonica* cv. Nipponbare, *O. sativa indica*, *O. brachyantha* and *O. glaberrima*. While we do not observe a difference between *O. sativa japonica* and *O. sativa indica*, we indeed find a single base-pair insertion at position 18/19 of the PthXo3 box in *O. brachyantha*, which may be compensated by the aberrant repeat of PthXo3 (Supplementary Fig. 7). Interestingly, while binding of PthXo3 to its target box in *O. sativa* is preferred in the looped-out conformation (6 instead of 10 mismatches), binding in the normal conformation should be preferred in *O. brachyantha* (4 instead of 9 mismatches; Supplementary Fig. 7).

To study the contribution of the aberrant repeat in the natural TALE AvrXa7 on target recognition, we assemble an artificial TALE with the same RVD composition as AvrXa7 (termed ArtXa70) and a derivative (ArtXa71) with a normal 34-amino-acid

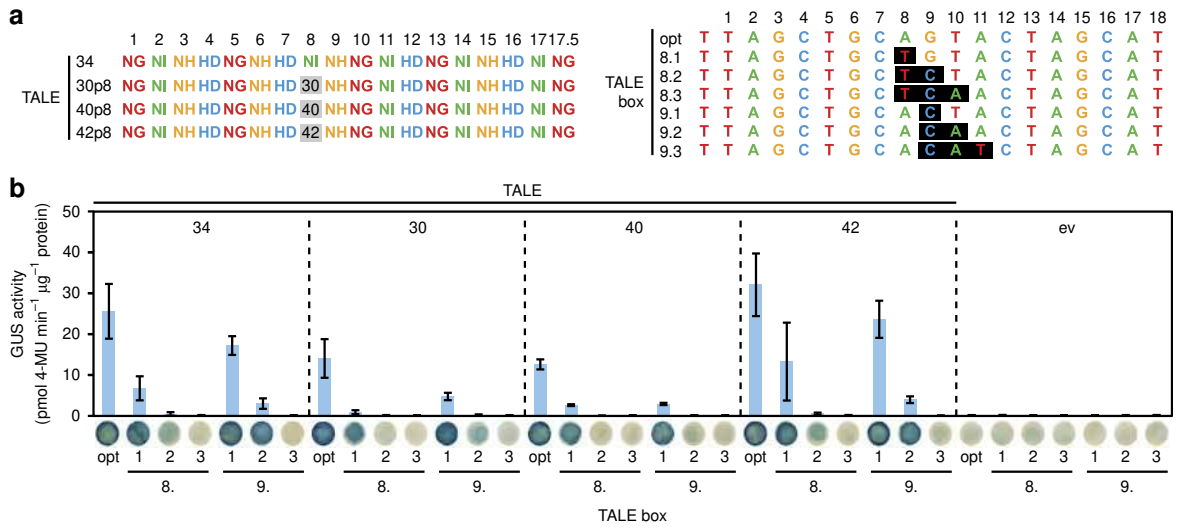


Figure 6 | Aberrant repeats participate in TALE-DNA base pair recognition. (**a**) TALE RVDs and target boxes. Artificial TALEs are constructed with all 34-aa repeats or aberrant repeats. An aberrant 30-aa NI-repeat, 40-aa NN-repeat and 42-aa NI-repeat is inserted at position 8, respectively. Target DNA boxes are either perfectly matching the specificity of normal repeats (opt) or have one to three non-matching bases (transversions) between position 8 and 11 (black boxes). (**b**) GUS assays of TALEs and reporter constructs ($n=3$). 35S-driven GFP expression serves as empty vector (ev) control in quantitative and qualitative assays. Error bars indicate the s.d. in the quantitative assay. One representative leaf disk of the qualitative assay is shown.

repeat instead of the aberrant 39-amino-acid-repeat at position 13 (Supplementary Fig. 8a,b). Both artificial TALEs recognize the optimal box in GUS assays similarly well as AvrXa7, but the TALE without the aberrant repeat (ArtXa71) is significantly compromised on the −1 target box (Supplementary Fig. 8c). This indicates that the aberrant repeat also contributes to the flexible target box recognition of the natural TALE AvrXa7 (Supplementary Fig. 1).

TALEs with aberrant repeats can break plant resistance. To test whether TALEs with aberrant repeats contribute to bacterial virulence, we design TALEs based on a naturally occurring promoter allele that confers plant resistance. Rice *Xa25* (*OsSWEET13*) encodes a SWEET protein that supports *Xanthomonas* virulence probably facilitated by the TALE PthXo2 (refs 16,19). A natural single base-pair insertion mutation in the recessive *xa25* allele results in plants resistant to *Xoo* likely by prohibiting binding of PthXo2, which does not contain any aberrant repeats[17] (Supplementary Fig. 9a). To test whether TALEs with aberrant repeats can overcome this resistance, we construct TALEs targeting the PthXo2-binding site in the recessive *xa25* allele either with 17.5 exclusively regular 34-amino-acid repeats (ArtXa251, Fig. 8 and Supplementary Fig. 9b) or one aberrant repeat at position 4 (ArtXa252, Fig. 8 and Supplementary Fig. 9c). The *xa25* and *Xa25* promoter regions are cloned from *O. sativa* cv. Nipponbare and *O. sativa* cv. Zhenshan 97, respectively[17]. Agrobacterium-mediated production of ArtXa251 triggers reporter gene activation of the *xa25* (Nipponbare; that is, optimal box), but not the *Xa25* (Zhenshan 97; that is, −1 frame-shift box) promoter (Supplementary Fig. 9d), supporting that normal TALEs are highly compromised by indel mutations. In contrast, the TALE with an aberrant repeat, ArtXa252, induces expression of both reporter constructs (Supplementary Fig. 9d).

We further analyse the contribution of TALEs with aberrant repeats to *Xanthomonas* spp. infections. The *Xoo* strain BAI3 causes disease on rice cultivars Azucena and Zhenshan 97 in part because it contains the TALE TalC that directs expression of the gene encoding the sugar exporter OsSWEET14 (ref. 21) (Fig. 8). Deletion of *talC* renders the bacteria unable to cause disease[21] (Fig. 8), but this can be compensated by artificial TALEs that trigger expression of *OsSWEET13* (ref. 19). We introduce ArtXa251 and ArtXa252 into *Xoo* BAI3Δ*talC* and infect rice cultivars Azucena (*xa25*) and Zhenshan 97 (*Xa25*), which differ in their *OsSWEET13* promoter region by the 1-bp indel mutation (Fig. 8). The *Xoo* strain containing ArtXa251 with exclusively normal repeats causes disease symptoms on *O. sativa* cv. Azucena, but not *O. sativa* cv. Zhenshan 97 (Fig. 8), because of

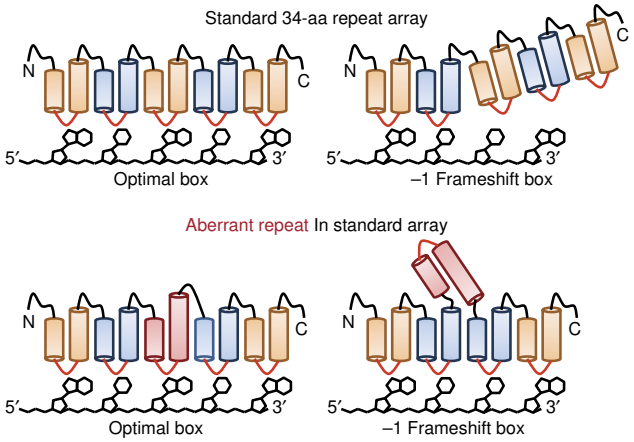


Figure 7 | Cartoons of repeat arrays in the normal and the looped-out conformation. Model of TALE repeats consisting of standard 34-aa repeats and repeat arrays containing one repeat of aberrant length, respectively, aligned to an optimal box or a −1 frameshift box. The aberrant repeat is shown in dark red.

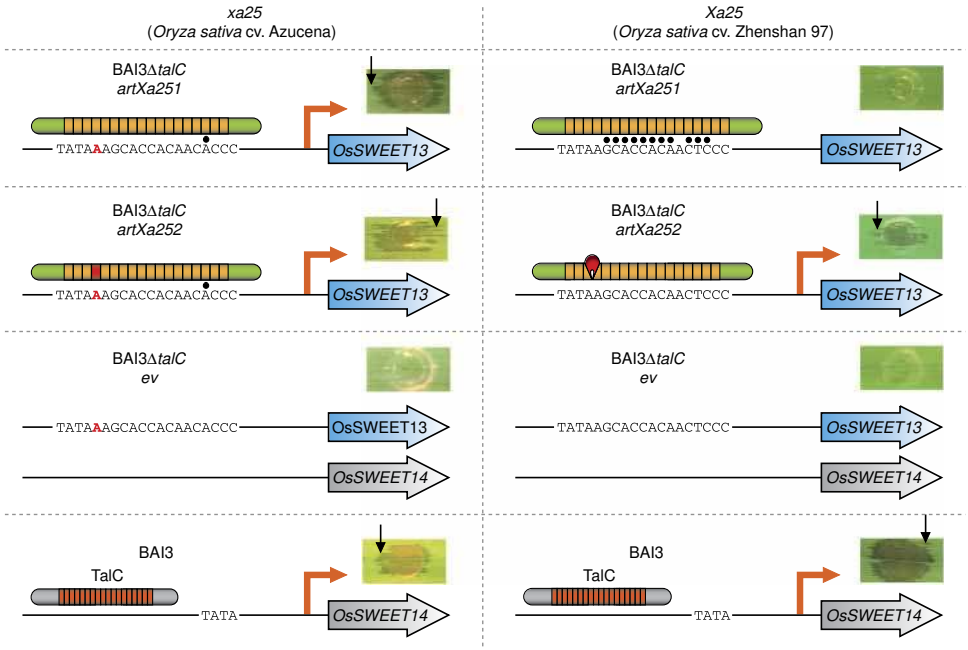


Figure 8 | TALEs with aberrant repeat can overcome plant resistance. Disease phenotypes of *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) strains on *Oryza sativa* leaves. Leaves of 3–4 week-old Azucena or Zhenshan 97 plants are inoculated with strain BAI3 or BAI3Δ*talC* carrying empty vector (*ev*) or TALE *artXa251* or *artXa252*. Pictures are taken 4 days post inoculation. Azucena and Zhenshan 97 contain alleles of *OsSWEET13*, termed *xa25* and *Xa25*, respectively, which differ in their promoter sequences by an additional nucleotide (red). The strain harbouring ArtXa252 with a 40-aa repeat (red) causes water soaking disease symptoms (black arrows) on both alleles. Black dots indicate RVD–base -mismatches. The experiments are performed twice with similar results.

the indel mutation in the target box. In contrast, *Xoo* containing ArtXa252 with the aberrant repeat causes disease symptoms on both rice cultivars (Fig. 8). This demonstrates that aberrant repeats can enable *Xanthomonas* to compensate for indels in target promoters and break natural plant resistances. Apparently, aberrant repeats are an evolutionary solution for *Xanthomonas* spp. to overcome small indel mutations that otherwise efficiently block TALE binding.

Discussion

We have described an exceptional and surprising recognition pattern of TALEs that has implications for the general concept of TALE–DNA interaction as well as the evolutionary adaptation of these important virulence factors. The near-identical amino-acid sequence of *Xanthomonas* TALE repeats implies that this regularity is required for binding to the highly symmetric structure of the DNA double helix[9,10,14]. Nevertheless, in nature the glycine and the leucine at positions 14 and 29, respectively, are the only amino acids conserved in all TALE repeat sequences[5]. Even more striking are natural repeat variants that deviate from the typical 34- or 35-amino-acid length of TALE repeats, because they are expected to impose a structural problem to the overall repeat array. Here, we show that either shorter (30 amino acids) or longer repeat variants (39, 40 or 42 amino acids) change the DNA-binding behaviour of TALEs. They can either insert into the repeat array like normal repeats or be excluded from the interaction to facilitate a shift of the following repeats forward by one position depending on the best fit of all RVDs to a given target sequence (Figs 2 and 7). This behaviour is not possible for TALEs with normal 34-amino-acid repeats. In the absence of structural data, we favour a model that the aberrant repeat loops out of the repeat array, because this seems to be sterically the easiest solution; alternative scenarios are possible though. Surprisingly, all aberrant repeats tested functioned in a comparable way although we have not compared them in all assays in our study. In part this might be explained by the fact that the aberrant repeat sequences are all related to a normal 34-amino-acid TALE repeat. They contain either a short deletion in the second α-helix (30-amino-acid repeat) or a duplicated first (42-amino-acid repeat) or second α-helix (39/40-amino-acid repeats) flanking the RVD (Fig. 1). Possibly, these duplications of structural elements still allow the hydrophobic repeat-to-repeat interactions that stabilize the normal repeat array[9,10].

At the same time, the aberrant repeats probably do not fit perfectly into the array, thereby weakening inter-repeat interactions and causing a local flexibility and structural tension that allows the aberrant repeat to loop out. In general, TALEs exhibit a highly flexible structure. Molecular dynamics simulations show that the TALE repeat region has a high conformational plasticity[48]. In addition, structural data and computational simulations indicated that the TALE repeat superstructure condenses upon interaction with cognate DNA resulting in more densely packed repeats than in the DNA-free form[9,41,48]. Likely, the flexible nature of the repeat region allows the unique binding mode conferred by aberrant repeats.

How does an aberrant repeat influence the dynamics of TALE protein–DNA interaction? Insertion of aberrant repeats in the rear part of the repeat array compromised TALE-mediated gene induction in our experiments (Fig. 3). We postulate that the aberrant repeat weakens the protein–DNA interaction and that a minimum number of repeats following the aberrant one is required to allow subsequent condensation of the rear repeats onto the DNA double helix. The N-terminal domain of TALEs mediates unspecific interaction to DNA as well as recognition of the initial thymine[7,8,10]. In addition, the initial repeats after the N-terminal domain are more important for binding than downstream repeats[37,47]. Together, these observations support a model where the N-terminal TALE region scans the DNA and the repeats subsequently condense at target sequences onto the DNA in a consecutive fashion starting from the N-terminal region. Possibly, an aberrant repeat that is excluded from binding in a frameshift scenario only allows condensation of further repeats, if their number is sufficiently high to condense onto the DNA in a separate event. Accordingly, the aberrant repeat might separate the repeat domain into two binding domains. In general, the tolerance of aberrant repeats in a repeat array likely depends on the overall binding energies of preceding and following repeats. Surprisingly, two tandem aberrant repeats only recognize a −1 frameshift box and not a normal one nor a −2 frameshift box (Fig. 5). This implies that the loop out conformation has specific requirements and does not allow looping out of more than one repeat. Furthermore, the additional structural perturbation of two near aberrant repeats does not allow a normal repeat arrangement. It is presently difficult to estimate how many aberrant repeats may be accepted in a TALE repeat domain.

The novel binding mode comprises interesting potential for biotechnology applications that are otherwise difficult to achieve. TALEs and TALENs with aberrant repeats can be used to simultaneously recognize allelic variants that differ by single nucleotide frameshifts. For synthetic TALE biology[49], master regulators with aberrant repeats might be used to control two subsets of target sequences that differ by indel mutations. This is difficult to achieve with CRISPR/Cas9 recognition of DNA sequences, because its RNA-guided recognition mechanism cannot compensate for indel mutations[50]. In addition, it is tempting to speculate that the aberrant repeats can function as an insertion point for more complex peptide tags or proteins within the TALE repeat domain to generate unique fusion proteins.

A significant aspect of our study is that these aberrant repeats are naturally evolved variants. Their presence in important virulence factors implies that they confer a selective advantage. The battle between pathogen and host is characterized by a constant struggle for innovation and counteracting activities to prevent losing the evolutionary race. Inherently, TALEs are especially sensitive to frameshift mutations in their target sequence and this has been exploited by natural plant resistances[16,20,25]. An aberrant repeat contributes a new degree of flexibility to the DNA-binding activity of a TALE without strong penalty on overall activity. We showed that TALEs with aberrant repeats can enable *Xoo* to overcome a natural resistance that is caused by a 1-bp deletion. As another example, the aberrant repeat in PthXo3 and AvrXa7 might have evolved to recognize unknown target box variants in response to a specific mutation that occurred in the *OsSWEET14* promoter in some rice cultivars. Alternatively, our computational search of available *OsSWEET14* promoter sequences raises the possibility that this potent virulence factor might favour colonization of *Xoo* on the different rice species *O. sativa* and *O. brachyantha*. So far, the possibility that TALEs of one bacterial strain support virulence in different host species has not been explored, but it is typical for type III effectors that they function in different plants and even non-plants[51,52]. Further analysis of natural TALEs with aberrant repeats and their target promoters will clarify the role of these unique virulence factors for the pathogen. Our data reveal that designer resistances based on mutations in TALE target boxes[16] have to be considered carefully to be effective.

Methods

Bacterial strains and growth conditions. *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) BAI3 and BAI3Δ*talC* are used in this study[21,53]. Plasmids are introduced into *Xoo* by conjugation using pRK2013 as a helper plasmid in triparental matings[54].

Rifampicin (100 μg ml^{-1}) and gentamicin (20 μg ml^{-1})-resistant clones are selected upon plating on PSA medium and one isolate is chosen for further experiments. *Escherichia coli* is cultivated at 37 °C in lysogeny broth and *Agrobacterium tumefaciens* GV3101 at 30 °C in yeast extract broth supplemented with appropriate antibiotics.

Plant growth and inoculations. *Nicotiana benthamiana* plants are cultivated in the greenhouse with 16 h light, 40–60% humidity and day/night temperatures of 23 °C/19 °C. Plants inoculated with *Agrobacterium* strains are transferred to a Percival growth chamber (Percival Scientific) with 22 °C/18 °C day/night temperatures and 16 h light. Rice experiments are performed under greenhouse conditions with cycles of 12 h of light at 26 °C, 70% relative humidity and 12 h of dark at 25 °C, 70% relative humidity. *Oryza sativa* subsp. *japonica* cv. Azucena and *O. sativa* subsp. *indica* cv. Zhenshan 97 are used for virulence assays. Leaves of 3–4-week-old plants are infiltrated with a bacterial suspension at an optical density of 0.5 at 600 nm (OD_{600}) using a needle-less syringe[55]. Pictures from disease symptoms (water-soaked lesions) are taken 4 days post inoculation.

Construction of repeat modules with aberrant repeats. Primer pairs encoding natural repeats of aberrant length are designed such that forward and reverse primer overlap in their 3′ part (Supplementary Tables 1 and 3). Phusion polymerase is used to extend the primers and the resulting DNA fragments are subcloned into pUC57 via *SmaI* cut-ligation. The resulting plasmids are used as template to amplify the aberrant repeats with primers that add *BpiI* restriction sites matching the existing Golden TAL Technology kit to insert the aberrant repeats at position 2 or 3 into a hexa-repeat array[46].

Construction of artificial TALEs. TALEs are constructed using the Golden TAL Technology[46]. Up to six individual repeats with selected RVDs are subcloned in an assembly vector. To construct AvrXa7-derivatives with 25.5 repeats, we extend the original Golden TAL kit[46] with two novel assembly vectors (Supplementary Note 1). The repeat backbone apart from the RVDs is identical (Supplementary Table 4). The repeats are pre-assembled in 2–5 assembly vectors and inserted together with the Hax3 N- and C-terminal regions[56]. For expression *in planta*, N-terminal GFP–TALE fusion are assembled in a Golden Gate-compatible binary vector allowing expression of the constructs under control of the constitutive 35S promoter. For examples of DNA and amino-acid sequences see Supplementary Note 1. For expression in *Xanthomonas* spp., TALE–FLAG fusions are assembled in a Golden Gate-compatible broad host range vector[19].

GUS reporter constructs and reporter assay. *OsSWEET13* (*Xa25*/*xa25*, *Os12N3* and *Os12g29220*) 1 kb promoter fragments (Supplementary Note 1), an *OsSWEET14* (*Os11N3* and *Os11g31190*) 341-bp promoter fragment, and artificial and natural TALE boxes (Supplementary Table 3) together with the minimal *Bs4* promoter, respectively, are inserted into pENTR/D-TOPO (Invitrogen). The promoter derivatives are recombined into pGWB3 (ref. 57) via LR recombination. Transient GUS reporter assays are performed[7]. Briefly, *Agrobacterium* strains delivering TALE constructs and GUS reporter constructs are mixed 1:1 and inoculated into leaves of 5–7-week-old *N. benthamiana* plants with a total OD_{600} of 0.8. The T-DNAs integrate into the plant chromosomes. Leaf disks (0.9 cm diameter) are sampled 2 days later and GUS activity is determined. For qualitative GUS assays, leaf disks are stained in X-Gluc (5-bromo-4-chloro-3-indolyl-β-D-glucuronide) solution, destained in ethanol and dried between acetate foil. For quantitative GUS assays two leaf disks are pooled, the plant tissue homogenized, diluted and incubated with 4-methyl-umbelliferyl-β-D-glucuronide (MUG). Proteins are quantified using Bradford assay (Roth). Values from three plants are combined into one data point. All experiments are done at least twice with similar results.

Construction of TALENs. TALENs are constructed from modules matching the Golden TAL Technology[46]. The repeats are assembled in hexa-repeat modules and inserted together with modified short N- and C-terminal modules into a compatible ENTRY vector (pEGG). The N-terminal TALEN module contains amino acids 153–288 of Hax3, a SV40 nuclear localization sequence, and a tag (c-myc-tag for forward TALEN and FLAG-tag for reverse TALEN) sequence (see Supplementary Note 1). The C-terminal TALEN module contains amino acids 1–63 of the C-terminal region of Hax3 and a heterodimeric (DS for forward TALEN and RR for reverse TALEN) 'sharkey' *Fok*I endonuclease domain[58]. The TALEN are transferred via GATEWAY (Invitrogen) LR recombination into pDEST17 under control of a T7 promoter.

TALEN *in vitro* cleavage assay. TALENs are expressed using the TnT T7 Quick Coupled Transcription/Translation System (Promega) following the manufacturer's instructions. Five hundred nanograms of DNA from each TALEN construct are used. The target DNA fragment is generated by linearization of pENTR containing the respective target box upstream of the minimal *Bs4* promoter using the restriction enzyme *Alw*44I (Thermo Scientific) following the manufacturer's instructions and subsequent purification with the GeneJET PCR Purification Kit (Thermo Scientific). For the *in vitro* cleavage assay 4 μl of TnT reaction containing the TALEN pair proteins is mixed with 200 ng of target DNA in 1 × NEBuffer 3 (New England Biolabs) supplied with 2.5 μg μl^{-1} BSA to a total volume of 20 μl. After incubation for 60 min at 37 °C the reaction is inactivated at 65 °C for 20 min and centrifuged at 16,000*g* for 3 min. The supernatant (16 μl) is analysed on a 1% agarose gel. All experiments are at least done twice with similar results.

Immunoblotting. For plant immunoblotting TALEs are transiently expressed in *N. benthamiana* for 2 days. *Agrobacterium* strains are inoculated with an OD_{600} of 0.4. Two leaf disks are pooled, the plant tissue homogenized, resuspended in 90 μl Lämmli buffer and incubated at 95 °C for 10 min. Debris are pelleted and SDS-PAGE is performed with 15 μl of each sample. The proteins are transferred to a PROTRAN nitrocellulose membrane (Whatman). To detect GFP-tagged proteins membranes are incubated with anti-GFP rabbit serum (Life Technologies; dilution 1:2,000). For detection by enhanced chemiluminescence (ECL) ECL anti-rabbit IgG (GE Healthcare; dilution 1:10,000) is used.

References

1. DeFrancesco, L. Move over ZFNs. *Nat. Biotechnol.* **29,** 681–684 (2011).
2. Doyle, E. L., Stoddard, B. L., Voytas, D. F. & Bogdanove, A. J. TAL effectors: highly adaptable phytobacterial virulence factors and readily engineered DNA-targeting proteins. *Trends Cell Biol.* **23,** 390–398 (2013).
3. Gaj, T., Gersbach, C. A. & Barbas, 3rd C. F. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31,** 397–405 (2013).
4. Van den Ackerveken, G., Marois, E. & Bonas, U. Recognition of the bacterial avirulence protein AvrBs3 occurs inside the host plant cell. *Cell* **87,** 1307–1316 (1996).
5. Boch, J. & Bonas, U. *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Annu. Rev. Phytopathol.* **48,** 419–436 (2010).
6. Moscou, M. J. & Bogdanove, A. J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326,** 1501 (2009).
7. Boch, J. *et al.* Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326,** 1509–1512 (2009).
8. Gao, H., Wu, X., Chai, J. & Han, Z. Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. *Cell Res.* **22,** 1716–1720 (2012).
9. Deng, D. *et al.* Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* **335,** 720–723 (2012).
10. Mak, A. N., Bradley, P., Cernadas, R. A., Bogdanove, A. J. & Stoddard, B. L. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* **335,** 716–719 (2012).
11. Zhu, W., Yang, B., Chittoor, J. M., Johnson, L. B. & White, F. F. AvrXa10 contains an acidic transcriptional activation domain in the functionally conserved C terminus. *Mol. Plant Microbe Interact.* **11,** 824–832 (1998).
12. Nino-Liu, D. O., Ronald, P. C. & Bogdanove, A. J. *Xanthomonas oryzae* pathovars: model pathogens of a model crop. *Mol. Plant Pathol.* **7,** 303–324 (2006).
13. Verdier, V., Vera Cruz, C. & Leach, J. E. Controlling rice bacterial blight in Africa: needs and prospects. *J. Biotechnol.* **159,** 320–328 (2012).
14. Scholze, H. & Boch, J. TAL effectors are remote controls for gene activation. *Curr. Opin. Microbiol.* **14,** 47–53 (2011).
15. Antony, G. *et al.* Rice *xa13* recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene Os-*11N3*. *Plant Cell* **22,** 3864–3876 (2010).
16. Li, T., Huang, S., Zhou, J. & Yang, B. Designer TAL effectors induce disease susceptibility and resistance to *Xanthomonas oryzae* pv. *oryzae* in rice. *Mol. Plant* **6,** 781–789 (2013).
17. Liu, Q. *et al.* A paralog of the MtN3/saliva family recessively confers race-specific resistance to *Xanthomonas oryzae* in rice. *Plant Cell Environ.* **34,** 1958–1969 (2011).
18. Römer, P. *et al.* 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 (2010).
19. Streubel, J. *et al.* Five phylogenetically close rice *SWEET* genes confer TAL effector-mediated susceptibility to *Xanthomonas oryzae* pv. *oryzae*. *New Phytol.* **200,** 808–819 (2013).
20. Yang, B., Sugio, A. & White, F. F. *Os8N3* is a host disease-susceptibility gene for bacterial blight of rice. *Proc. Natl. Acad. Sci. USA* **103,** 10503–10508 (2006).
21. Yu, Y. *et al.* Colonization of rice leaf blades by an african strain of *Xanthomonas oryzae* pv. *oryzae* depends on a new TAL effector that induces the rice Nodulin-3 *Os11N3* Gene. *Mol. Plant Microbe Interact.* **24,** 1102–1113 (2011).
22. Doyle, E. L. *et al.* TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res.* **40,** W117–W122 (2012).
23. Grau, J. *et al.* Computational predictions provide insights into the biology of TAL effector target sites. *PLoS Comput. Biol.* **9,** e1002962 (2013).

24. Pérez-Quintero, A. L. *et al.* An improved method for TAL effectors DNA-binding sites prediction reveals functional convergence in TAL repertoires of *Xanthomonas oryzae* strains. *PLoS One* **8,** e68464 (2013).
25. Gu, K. *et al. R* gene expression induced by a type-III effector triggers disease resistance in rice. *Nature* **435,** 1122 (2005).
26. Römer, P. *et al.* Plant-pathogen recognition mediated by promoter activation of the pepper *Bs3* resistance gene. *Science* **318,** 645–648 (2007).
27. Strauß, T. *et al.* RNA-seq pinpoints a Xanthomonas TAL-effector activated resistance gene in a large-crop genome. *Proc. Natl Acad. Sci. USA* **109,** 19480–19485 (2012).
28. Chu, Z. *et al.* Targeting *xa13*, a recessive gene for bacterial blight resistance in rice. *Theor. Appl. Genet.* **112,** 455–461 (2006).
29. Chu, Z. *et al.* Promoter mutations of an essential gene for pollen development result in disease resistance in rice. *Genes Dev.* **20,** 1250–1255 (2006).
30. Yuan, T., Li, X., Xiao, J. & Wang, S. Characterization of *Xanthomonas oryzae*-responsive cis-acting element in the promoter of rice race-specific susceptibility gene *Xa13*. *Mol. Plant* **4,** 300–309 (2011).
31. Christian, M. *et al.* TAL effector nucleases create targeted DNA double-strand breaks. *Genetics* **186,** 757–761 (2010).
32. Morbitzer, R., Römer, P., Boch, J. & Lahaye, T. 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 (2010).
33. Li, T. *et al.* TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.* **39,** 359–372 (2010).
34. Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29,** 143–148 (2011).
35. Byrum, S. D., Taverna, S. D. & Tackett, A. J. Purification of a specific native genomic locus for proteomic analysis. *Nucleic Acids Res.* **41,** e195 (2013).
36. Cong, L., Zhou, R., Kuo, Y. C., Cunniff, M. & Zhang, F. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat. Commun.* **3,** 968 (2012).
37. Garg, A., Lohmueller, J. J., Silver, P. A. & Armel, T. Z. Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Res.* **40,** 7584–7595 (2012).
38. Mendenhall, E. M. *et al.* Locus-specific editing of histone modifications at endogenous enhancers. *Nat. Biotechnol.* **31,** 1133–1136 (2013).
39. Miyanari, Y., Ziegler-Birling, C. & Torres-Padilla, M. E. Live visualization of chromatin dynamics with fluorescent TALEs. *Nat. Struct. Mol. Biol.* **20,** 1321–1324 (2013).
40. Politz, M. C., Copeland, M. F. & Pfleger, B. F. Artificial repressors for controlling gene expression in bacteria. *Chem. Commun. (Camb)* **49,** 4325–4327 (2013).
41. Stella, S. *et al.* Structure of the AvrBs3-DNA complex provides new insights into the initial thymine-recognition mechanism. *Acta Crystallogr. D Biol. Crystallogr.* **69,** 1707–1716 (2013).
42. Streubel, J., Blücher, C., Landgraf, A. & Boch, J. TAL effector RVD specificities and efficiencies. *Nat. Biotechnol.* **30,** 593–595 (2012).
43. de Lange, O. *et al.* 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 (2013).
44. Li, L. *et al.* Characterization and DNA-binding specificities of Ralstonia TAL-like effectors. *Mol. Plant* **6,** 1318–1330 (2013).
45. Yang, B., Zhu, W., Johnson, L. B. & White, F. F. The virulence factor AvrXa7 of *Xanthomonas oryzae* pv. *oryzae* is a type III secretion pathway-dependent nuclear-localized double-stranded DNA-binding protein. *Proc. Natl Acad. Sci. USA* **97,** 9807–9812 (2000).
46. Geißler, R. *et al.* Transcriptional activators of human genes with programmable DNA-specificity. *PLoS One* **6,** e19509 (2011).
47. Meckler, J. F. *et al.* Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic Acids Res.* **41,** 4118–4128 (2013).
48. Wan, H., Hu, J. P., Li, K. S., Tian, X. H. & Chang, S. Molecular dynamics simulations of DNA-free and DNA-bound TAL effectors. *PLoS One* **8,** e76045 (2013).
49. Li, Y., Moore, R., Guinn, M. & Bleris, L. Transcription activator-like effector hybrids for conditional control and rewiring of chromosomal transgene expression. *Sci. Rep.* **2,** 897 (2012).
50. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31,** 827–832 (2013).
51. Salomon, D., Dar, D., Sreeramulu, S. & Sessa, G. Expression of *Xanthomonas campestris* pv. *vesicatoria* type III effectors in yeast affects cell growth and viability. *Mol. Plant Microbe Interact.* **24,** 305–314 (2011).
52. Deslandes, L. & Rivas, S. Catch me if you can: bacterial effectors and plant targets. *Trends Plant Sci.* **17,** 644–655 (2012).
53. Gonzalez, C. *et al.* Molecular and pathotypic characterization of new *Xanthomonas oryzae* strains from West Africa. *Mol. Plant Microbe Interact.* **20,** 534–546 (2007).
54. Figurski, D. & Helinski, D. R. Replication of an origin-containing derivative of plasmid RK2 is dependent on a plasmid function provided in trans. *Proc. Natl Acad. Sci. USA* **76,** 1648–1652 (1979).
55. Reimers, P. J. & Leach, J. E. Race-specific resistance to *Xanthomonas oryzae* pv. *oryzae* conferred by bacterial blight resistance gene *Xa-10* in rice (*Oryza sativa*) involves accumulation of a lignin-like substance in host tissues. *Physiol. Mol. Plant Pathol.* **38,** 39–55 (1991).
56. Kay, S., Boch, J. & Bonas, U. Characterization of AvrBs3-like effectors from a *Brassicaceae* pathogen reveals virulence and avirulence activities and a protein with a novel repeat architecture. *Mol. Plant Microbe Interact.* **18,** 838–848 (2005).
57. Nakagawa, T. *et al.* Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J. Biosci. Bioeng.* **104,** 34–41 (2007).
58. Guo, J., Gaj, T. & Barbas, 3rd C. F. Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. *J. Mol. Biol.* **400,** 96–107 (2010).

Acknowledgements

We thank T. Schreiber, S. Thieme and U. Bonas for providing the GoldenGate-compatible vectors; K. Schlien for extending the Golden TAL kit, ArtXa70 and ArtXa71 constructs; T. Schreiber and U. Bonas for suggestions on the manuscript; R. Szczesny for providing *Fok*I constructs; J. Leach for providing *avrXa7*; C. Kretschmer for technical assistance and U. Bonas for support. This work was supported by grants from the Deutsche Forschungsgemeinschaft (SPP 1212, BO 1496/6-1, BO 1496/7-1), and from the European Regional Development Fund of the European Commission. B.S. is supported by a grant from GRiSP for New Frontiers Research projects (MENERGEP).

Author contributions

A.R., J.S. and J.B. conceived and designed the experiments. A.R., J.S. and C.B. performed the experiments. A.R., J.S., C.B. and J.B. analysed the data. M.R. established the TALEN assays. J.G. performed the bioinformatic analyses. B.S. provided material and performed additional experiments. A.R., J.S. and J.B. wrote the manuscript with support from all authors.

Additional information

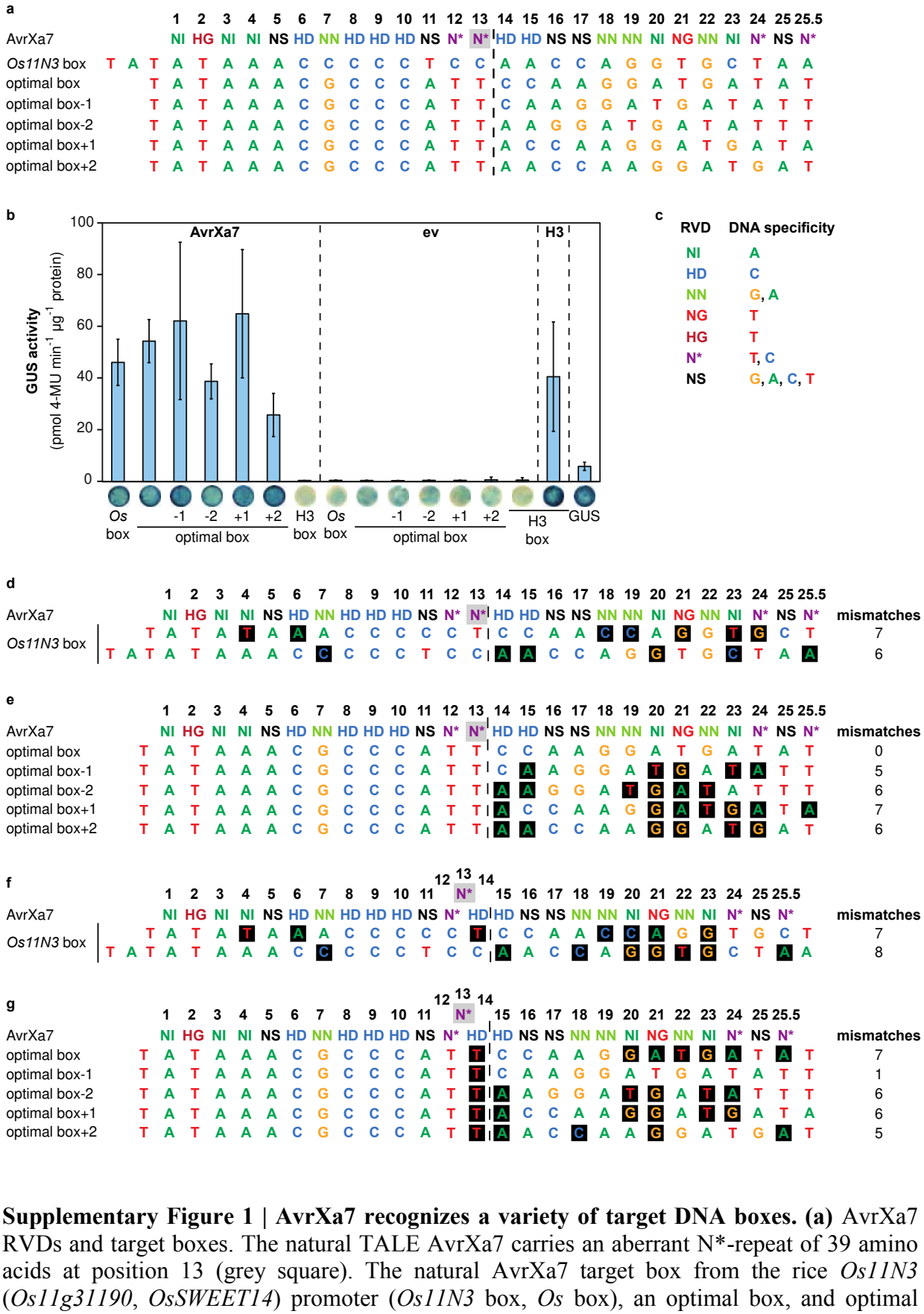
Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: J.B. is a part owner of the patent US 8420782 B2 'Modular DNA-Binding Domains And Methods Of Use' concerning the use of TALEs. The remaining authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Richter, A. *et al.* A TAL effector repeat architecture for frameshift binding. *Nat. Commun.* 5:3447 doi: 10.1038/ncomms4447 (2014).

Supplementary Figures



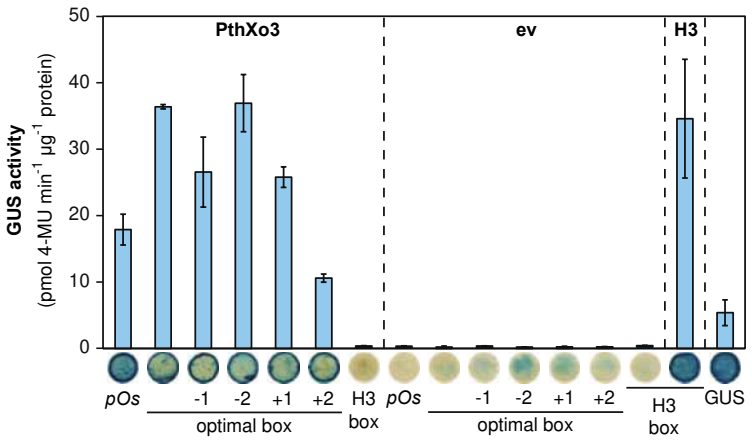
Supplementary Figure 1 | AvrXa7 recognizes a variety of target DNA boxes. (a) AvrXa7 RVDs and target boxes. The natural TALE AvrXa7 carries an aberrant N*-repeat of 39 amino acids at position 13 (grey square). The natural AvrXa7 target box from the rice *Os11N3* (*Os11g31190*, *OsSWEET14*) promoter (*Os11N3* box, *Os* box), an optimal box, and optimal

boxes containing one or two base pair deletions (optimal box-1, optimal box-2) or insertions (optimal box+1, optimal box+2) at position 14 (to the right of the dashed line) are fused to the minimal *Bs4* promoter and a promoterless GUS reporter gene. (**b**) GUS assay of AvrXa7 with target boxes. GUS reporter constructs are codelivered by *A. tumefaciens* into *N. benthamiana* leaf cells together with constitutive *35S*-driven *avrXa7*, and *GFP* (ev) constructs, respectively (n=3, error bars indicate s.d.). *35S::uidA* (GUS) and the natural TALE Hax3 (H3) with its perfect Hax3 box (H3 box) serve as controls. For qualitative assays, leaf disks are stained with X-Gluc, and a blue color indicates GUS activity. (**c**) RVD specificities. (**d-g**) Mismatch analysis of AvrXa7 on different target boxes. Non-matching RVD-DNA base combinations are boxed in black. (**d, e**) Aberrant repeat of AvrXa7 in regular base pair alignment. (**f, g**) Aberrant repeat of AvrXa7 in looped-out alignment. (**d, f**) AvrXa7 with natural box from the rice *Os11N3* promoter aligned either to the leftmost possible sequence described by Römer *et al*.[1] (bottom sequence) or to the second possible position described by Antony *et al*.[2] (top sequence). (**e, g**) AvrXa7 aligned to optimal boxes.

a

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	28.5
PthXo3	NI	HG	NI	HG	NI	NI	NI	HD	NN	HD	HD	HD	NG	HD	N*	NI	HD	HD	NN	NS	NI	NN	NN	NG	NN	HD	N*	NS	N*
pOs11N3	T	A	T	A	T	A	A	A	C	C	C	C	C	T	C	C	A	A	C	C	A	G	G	T	G	C	T	A	A	G
optimal box	T	A	T	A	T	A	A	A	C	G	C	C	C	T	C	T	A	C	C	G	A	A	G	G	T	G	C	T	A	T
optimal box-1	T	A	T	A	T	A	A	A	C	G	C	C	C	T	C	T	C	C	G	A	A	G	G	T	G	C	T	A	T	T
optimal box-2	T	A	T	A	T	A	A	A	C	G	C	C	C	T	C	T	C	G	A	A	G	G	T	G	C	T	A	T	T	T
optimal box+1	T	A	T	A	T	A	A	A	C	G	C	C	C	T	C	T	A	A	C	C	G	A	A	G	G	T	G	C	T	A
optimal box+2	T	A	T	A	T	A	A	A	C	G	C	C	C	T	C	T	A	A	A	C	C	G	A	A	G	G	T	G	C	T

b



c

RVD	DNA specificity
NI	A
HD	C
NN	G, A
NG	T
HG	T
N*	T, C
NS	G, A, C, T

d

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	28.5	mismatches
PthXo3	NI	HG	NI	HG	NI	NI	NI	HD	NN	HD	HD	HD	NG	HD	N*	NI	HD	HD	NN	NS	NI	NN	NN	NG	NN	HD	N*	NS	N*	
pOs11N3	T	A	T	A	T	A	A	A	C	C	C	C	C	T	C	C	A	A	C	C	A	G	G	T	G	C	T	A	A	G	10
optimal box	T	A	T	A	T	A	A	A	C	G	C	C	C	T	C	T	A	C	C	G	A	A	G	G	T	G	C	T	A	T	0
optimal box-1	T	A	T	A	T	A	A	A	C	G	C	C	C	T	C	T	C	C	G	A	A	G	G	T	G	C	T	A	T	T	8
optimal box-2	T	A	T	A	T	A	A	A	C	G	C	C	C	T	C	T	C	G	A	A	G	G	T	G	C	T	A	T	T	T	8
optimal box+1	T	A	T	A	T	A	A	A	C	G	C	C	C	T	C	T	A	A	C	C	G	A	A	G	G	T	G	C	T	A	6
optimal box+2	T	A	T	A	T	A	A	A	C	G	C	C	C	T	C	T	A	A	A	C	C	G	A	A	G	G	T	G	C	T	7

e

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	28.5	mismatches
PthXo3	NI	HG	NI	HG	NI	NI	NI	HD	NN	HD	HD	HD	NG	HD	N*	NI	HD	HD	NN	NS	NI	NN	NN	NG	NN	HD	N*	NS	N*	
pOs11N3	T	A	T	A	T	A	A	A	C	C	C	C	C	T	C	C	A	A	C	C	A	G	G	T	G	C	T	A	A	G	6
optimal box	T	A	T	A	T	A	A	A	C	G	C	C	C	T	C	T	A	C	C	G	A	A	G	G	T	G	C	T	A	T	7
optimal box-1	T	A	T	A	T	A	A	A	C	G	C	C	C	T	C	T	C	C	G	A	A	G	G	T	G	C	T	A	T	T	1
optimal box-2	T	A	T	A	T	A	A	A	C	G	C	C	C	T	C	T	C	G	A	A	G	G	T	G	C	T	A	T	T	T	8
optimal box+1	T	A	T	A	T	A	A	A	C	G	C	C	C	T	C	T	A	A	C	C	G	A	A	G	G	T	G	C	T	A	8
optimal box+2	T	A	T	A	T	A	A	A	C	G	C	C	C	T	C	T	A	A	A	C	C	G	A	A	G	G	T	G	C	T	6

f

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	28.5	mismatches
PthXo3	NI	HG	NI	HG	NI	NI	NI	HD	NN	HD	HD	HD	NG	HD	N*	NI	HD	HD	NN	NS	NI	NN	NN	NG	NN	HD	N*	NS	N*	
optimal box-2	C	C	T	A	T	A	T	A	A	C	G	C	C	C	T	C	T	C	G	A	A	G	G	T	G	C	T	A	T	8

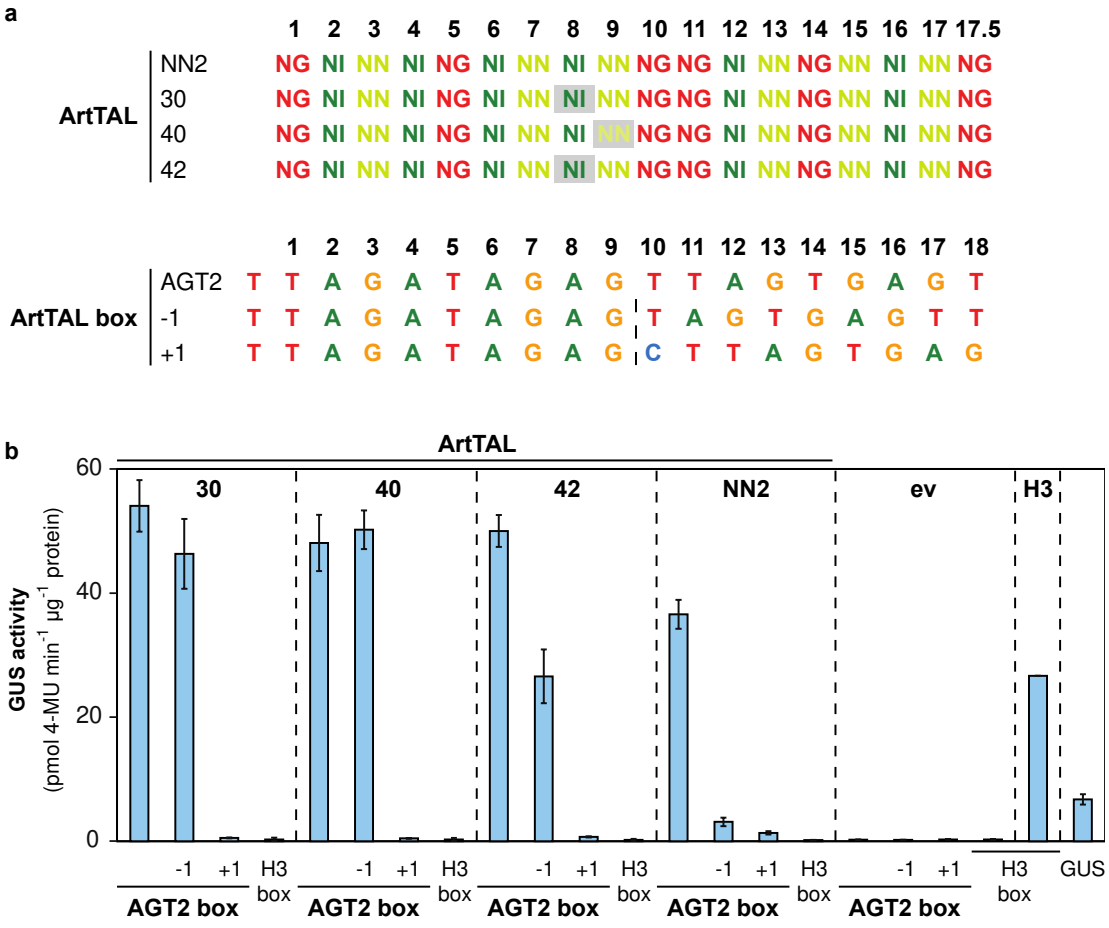
g

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	28.5	mismatches
PthXo3	NI	HG	NI	HG	NI	NI	NI	HD	NN	HD	HD	HD	NG	HD	N*	NI	HD	HD	NN	NS	NI	NN	NN	NG	NN	HD	N*	NS	N*	
optimal box+1	T	A	T	A	A	A	C	G	C	C	C	T	C	T	A	A	C	C	G	A	A	G	G	T	G	C	T	A	T	T	6

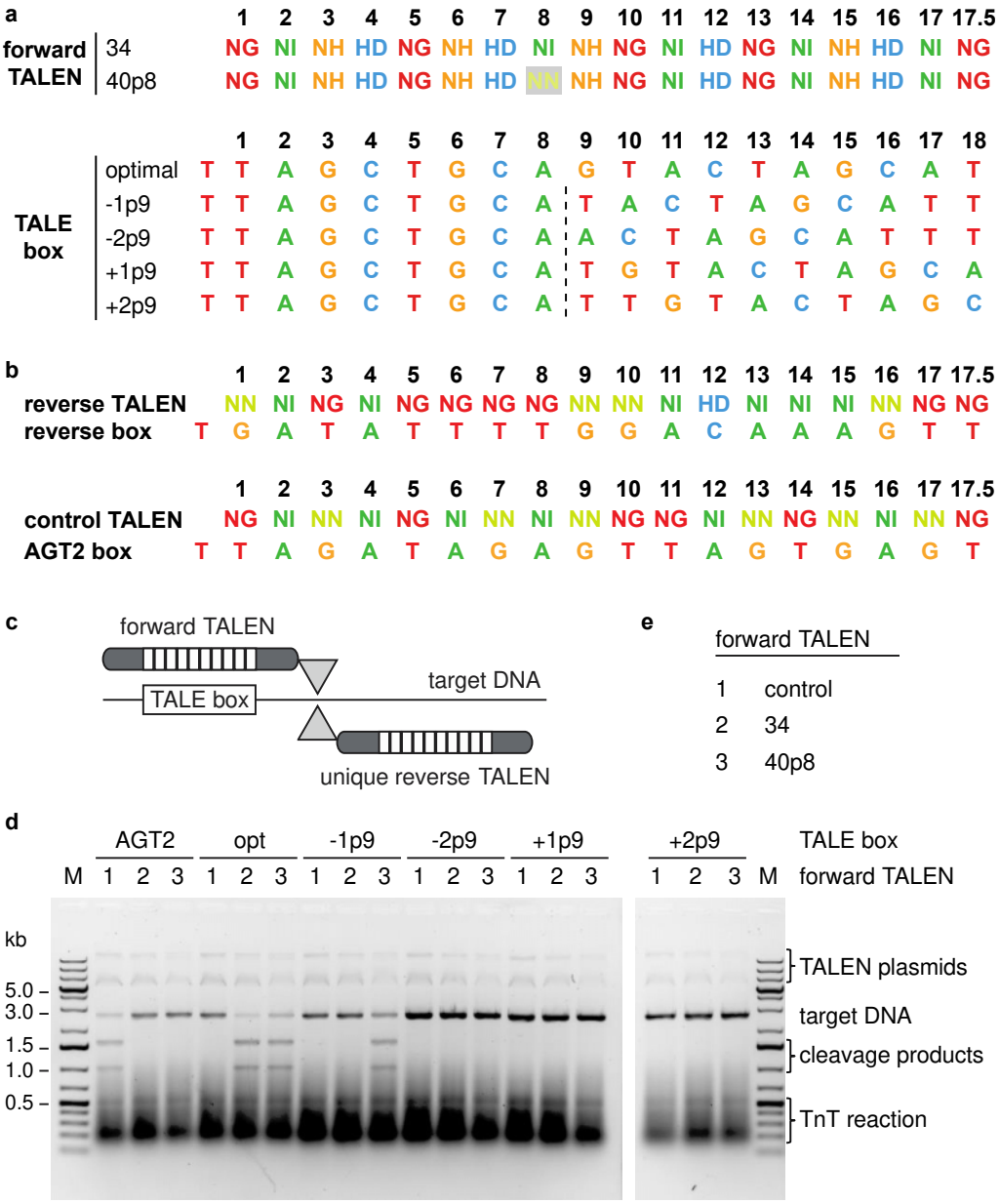
h

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	28.5	mismatches
PthXo3	NI	HG	NI	HG	NI	NI	NI	HD	NN	HD	HD	HD	NG	HD	N*	NI	HD	HD	NN	NS	NI	NN	NN	NG	NN	HD	N*	NS	N*	
optimal box+2	T	A	T	A	A	A	C	G	C	C	C	T	C	T	A	A	A	C	C	G	A	A	G	G	T	G	C	T	A	T	7

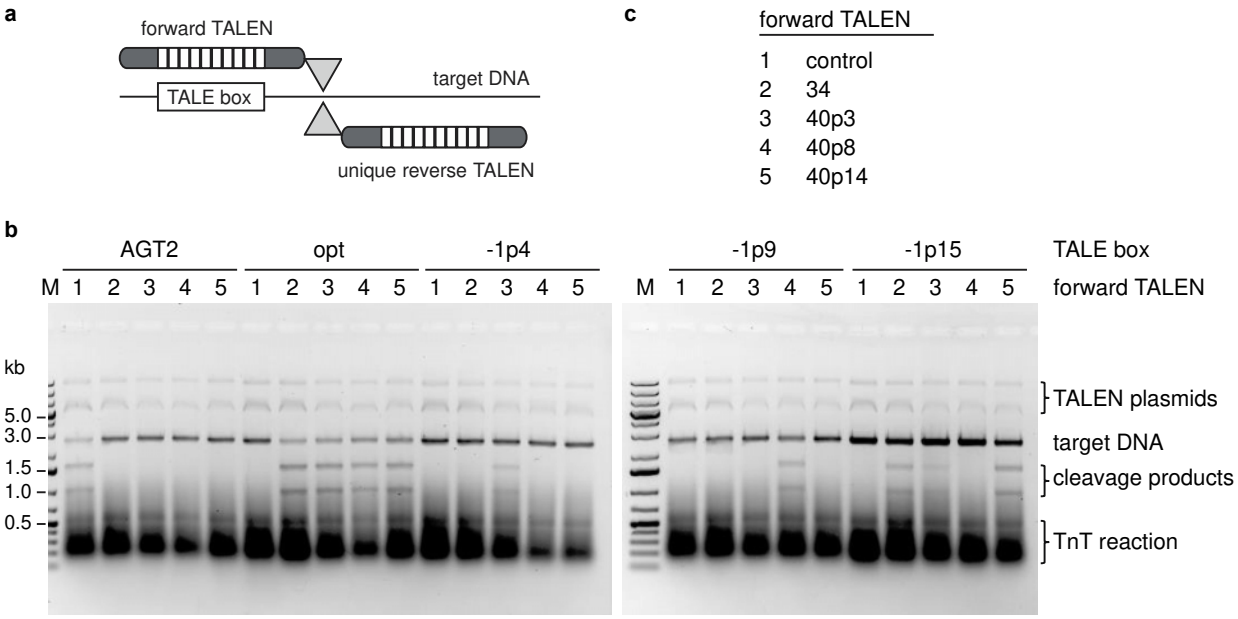
Supplementary Figure 2 | PthXo3 recognizes a variety of target DNA boxes. (**a**) PthXo3 RVDs and target boxes. The natural TALE PthXo3 carries an aberrant N*-repeat of 39 amino acids at position 15 (grey square). The natural PthXo3 target box in the rice *Os11N3* (*Os11g31190*, *OsSWEET14*) promoter[2] (*pOs11N3*, *pOs*) or an optimal box, and optimal boxes containing one or two base pair deletions (optimal box-1, optimal box-2) or insertions (optimal box+1, optimal box+2) at position 16 (to the right of the dashed line), upstream of the minimal *Bs4* promoter are fused to a promoterless GUS reporter gene. (**b**) GUS assay of PthXo3 with target boxes. GUS reporter constructs are codelivered by *A. tumefaciens* into *N. benthamiana* leaf cells together with constitutive *35S*-driven *pthXo3*, and *GFP* (ev) constructs, respectively (*n*=3, error bars indicate s.d.). *35S::uidA* (GUS) and the natural TALE Hax3 (H3) with its perfect Hax3 box (H3 box) serve as controls. For qualitative assays, leaf disks are stained with X-Gluc, and a blue color indicates GUS activity. (**c**) RVD specificities. (**d-h**) Mismatch analyses of PthXo3 on different target boxes. Non-matching RVD-DNA base combinations are boxed in black. (**d**) Aberrant repeat of PthXo3 in regular base pair alignment. (**e**) Aberrant repeat of PthXo3 in looped-out alignment. (**f-h**) Alternative alignments of PthXo3 to (**f**) the optimal box-2, (**g**) the optimal box+1 and (**h**) the optimal box+2.



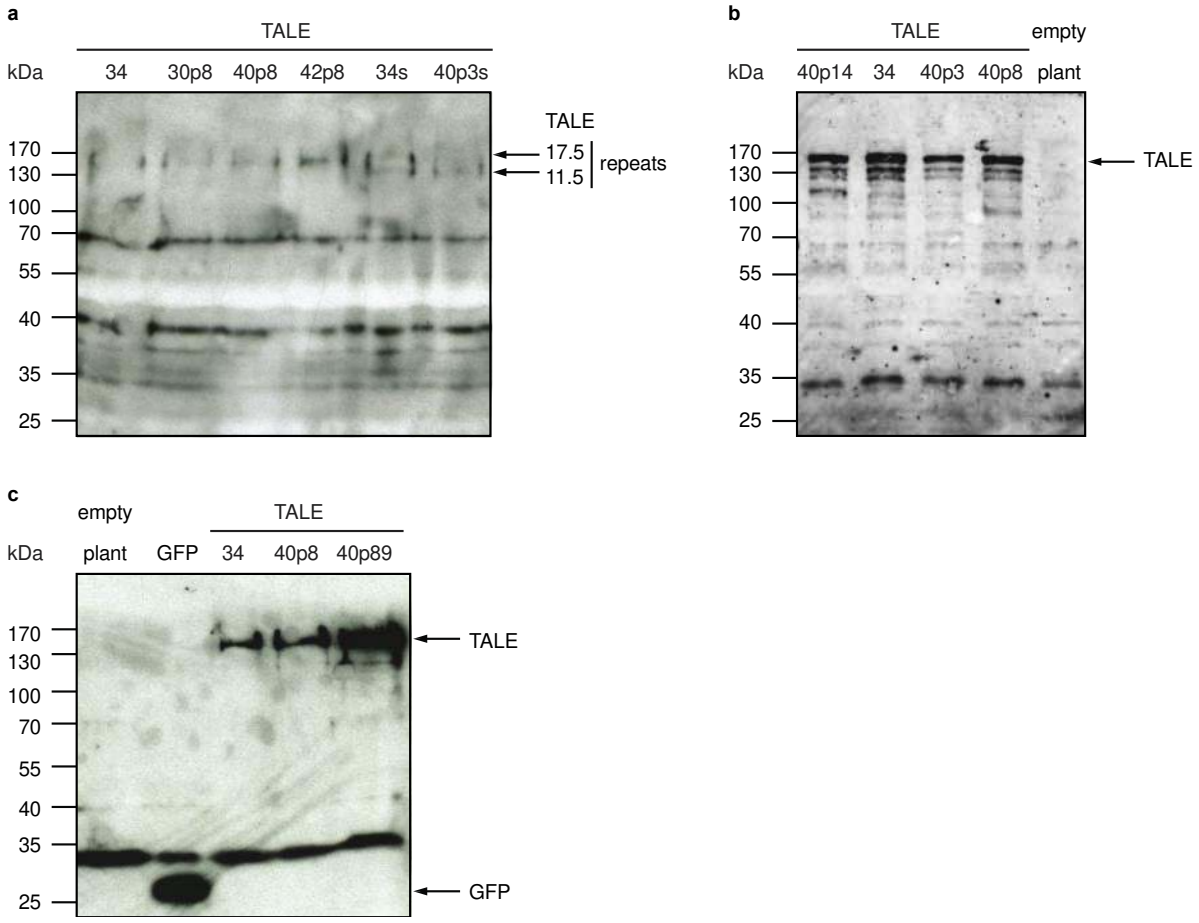
Supplementary Figure 3 | Aberrant repeats enable frame shift binding with an alternative TALE repeat array. (**a**) ArtTAL RVDs and target boxes. ArtTALs with all 34-amino acid (aa) repeats (NN2) according to Streubel *et al.*[3] or ArtTALs with an aberrant 30-aa NI-repeat, 40-aa NN-repeat, and 42-aa NI-repeat, respectively, at position 8 or 9 are constructed. The boxes are either perfectly matching the specificity of normal repeats (AGT2) or have one base pair deleted (AGT2-1) or inserted (AGT2+1) at position 10 (to the right of the dashed line). Target ArtTAL-boxes are fused to a minimal promoter and a promoterless *uidA* (GUS) reporter gene. (**b**) GUS assays of ArtTALs and reporter constructs. GUS reporter constructs are codelivered by *A. tumefaciens* into *N. benthamiana* leaf cells together with constitutive *35S*-driven *artTAL* genes, and *GFP* (ev), respectively (n=3, error bars indicate s.d.). *35S::uidA* (GUS) and the natural TALE Hax3 (H3) with its perfect Hax3-box (H3-box) serve as controls.



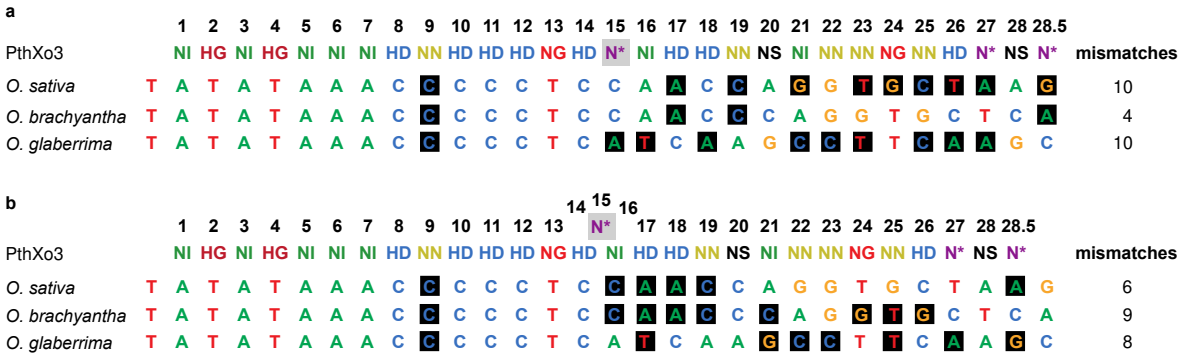
Supplementary Figure 4 | Aberrant repeats allow a -1 nucleotide frame shift tolerance in TALENs. (**a**) RVDs of forward TALENs and target boxes. TALENs are constructed with all 34-amino acid repeats or an aberrant 40-amino acid NN-repeat inserted at position 8 (40p8). The NN-repeat recognizes both, G and A DNA bases. The boxes are either perfectly matching the specificity of normal repeats (optimal) or have one or two base pairs either deleted (-1p9, -2p9) or inserted (+1p9, +2p9) at position 9 (to the right of the dashed line). The TALEN repeat array is designed such that for normal repeats mainly non-matching repeat-DNA base combinations occur following the position of the indels in the DNA sequence. (**b**) RVDs of reverse and control TALEN. The control TALEN recognizes a different target box (AGT2). (**c**) Cartoon of TALENs bound to DNA. The TALEN pairs are placed such that the FokI domain (grey triangle) dimerizes and cuts the DNA. The reverse TALEN is constant in all assays. (**d**) *In vitro* TALEN restriction assay. Target DNA is incubated with *in vitro* transcribed and translated TALEN pairs. The cleavage products are documented on agarose gels. Additional signals of low molecular weight and TALEN plasmids originate from the *in vitro* transcription and translation (TnT) reaction. GeneRuler 1 kb DNA Ladder (Thermo Scientific) is used as molecular weight marker (M). (**e**) Legend of forward TALENs.



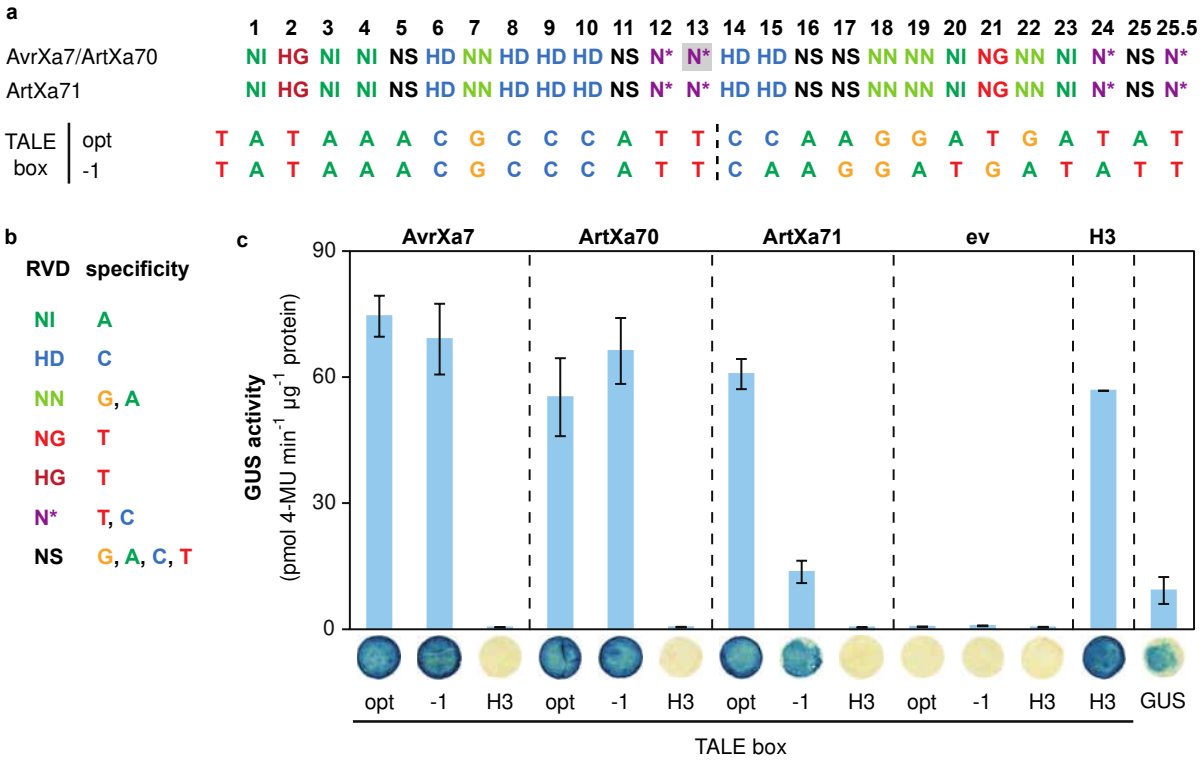
Supplementary Figure 5 | Full gel images of the *in vitro* TALEN restriction assay in Figure 3c. (**a**) Cartoon of TALENs bound to DNA. The TALEN pairs are placed such that the FokI domain (grey triangle) dimerizes and cuts the DNA. The reverse TALEN is constant in all assays. (**b**) *In vitro* TALEN restriction assay. Target DNA is incubated with *in vitro* transcribed and translated TALEN pairs. The cleavage products are documented on agarose gels. Additional low molecular weight signals and TALEN plasmids originate from the *in vitro* transcription and translation (TnT) reaction. GeneRuler 1 kb DNA Ladder (Thermo Scientific) is used as molecular weight marker (M). (**c**) Legend of forward TALENs. The control TALEN recognizes a different target box (AGT2).



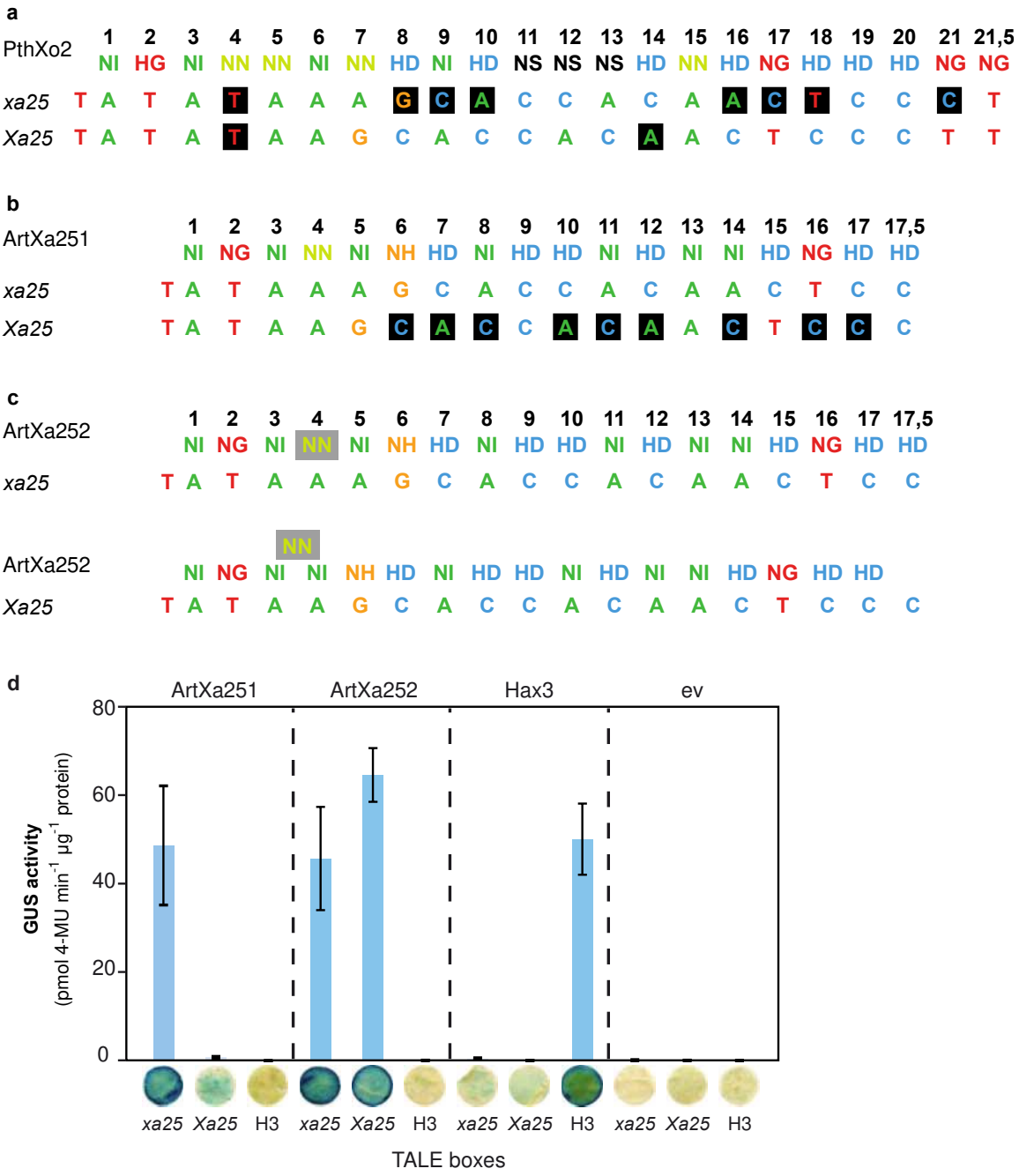
Supplementary Figure 6 | Immunoblotting of GFP-TALEs produced *in planta*. Detection of the TALE proteins used in GUS-assays of (a) Figure 2, 4 and 6, (b) Figure 3 and 4c-d, (c) Figure 5, respectively, using an anti-GFP antiserum.



Supplementary Figure 7 | PthXo3 RVDs aligned to predicted target boxes in the *OsSWEET14* promoters of *O. sativa*, *O. brachyantha* and *O. glaberrima*. (**a**) Aberrant repeat of PthXo3 in regular base pair alignment. (**b**) Aberrant repeat of PthXo3 in looped-out alignment.



Supplementary Figure 8 | The aberrant repeat contributes to frame shift box recognition of artificial AvrXa7-derivatives. (**a**) AvrXa7, ArtXa70 and ArtXa71 RVDs and target boxes. AvrXa7 derivatives are assembled with repeats resembling either the natural AvrXa7 repeats including the aberrant 39-aa N*-repeat at position 13 (grey square, ArtXa70) or all 34 aa repeats (ArtXa71). The optimal box (opt) and one containing an one base pair deletion (-1) at position 14 (to the right of the dashed line) are fused to the minimal *Bs4* promoter and a promoterless GUS reporter gene. (**b**) RVD specificities. (**c**) GUS assay. GUS reporter constructs are codelivered by *A. tumefaciens* into *N. benthamiana* leaf cells together with constitutive *35S*-driven *avrXa7*, *artXa70*, *artXa71*, and *GFP* (ev), respectively (n=3, error bars indicate s.d.). *35S::uidA* (GUS) and the natural TALE Hax3 (H3) with its Hax3 box (H3 box) serve as controls. For qualitative assays, leaf disks are stained with X-Gluc. A blue color indicates GUS activity.



Supplementary Figure 9 | An artificial TALE with aberrant repeat that targets the PthXo2 box recognizes different allelic *OsSWEET13* (*Xa25*) promoters. (**a**) PthXo2 RVDs aligned to *xa25* and *Xa25* promoter boxes of *O. sativa* cv. Nipponbare and *O. sativa* cv. Zhenshan 97[4]. (**b**) RVDs of the artificial TALE ArtXa251 aligned to *xa25* and *Xa25*. (**c**) RVDs of the artificial TALE ArtXa252 containing an aberrant 40-aa NN-repeat at position 4 (grey square) aligned to *xa25* and *Xa25* in normal and looped out conformation. (d) GUS assays (n=3) of artificial TALEs with *OsSWEET13* promoter sequences derived from *Oryza sativa* cv. Nipponbare (*xa25*) and Zhenshan 97 (*Xa25*) or the Hax3 box (H3) fused to a minimal promoter and a promoterless GUS-reporter gene. *35S*-driven expression of the natural TALE Hax3 or GFP as empty vector (ev) serve as controls in quantitative and qualitative assays. Error bars indicate the s.d. in the quantitative assay. One representative leaf disk of the qualitative assay is shown.

Supplementary Tables

Supplementary Table 1 | Aberrant repeat sequences.

TALE	Strain[a]	Feature[b]	Sequence[c]	Accession
30 amino acid-TALE repeats				
XOCORF_4248	*Xoc* BLS256	RVDs	NI-NN-HN-NN-NI-NG-HD-NN-HD-HG-HD-HG-HG-HD-HD-NG (15.5 repeats)	cmr.jcvi.org
		AA	LIPDQVVAIASNIGGKQALETVQRLLPVLC	
		DNA	CTGATCCCGGACCAGGTGGTGGCCATCGCCAGCAATATTGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGC	
39 amino acid-TALE repeats				
PthXo3	*Xoo* PXO61	RVDs	NI-HG-NI-HG-NI-NI-NI-HD-NN-HD-HD-HD-NG-HD-N*-NI-HD-HD-NN-NS-NI-NN-NN-NG-NN-HD-N*-NS-N* (28.5 repeats)	AAS46027
		AA	LTPDQVVAIASN*GGKQALETVQRLLPVQRLLPVLCQDHG	
AvrXa7	*Xoo* PXO86	RVDs	NI-HG-NI-NI-NS-HD-NN-HD-HD-HD-NS-N*-N*-HD-HD-NS-NS-NN-NN-NI-NG-NN-NI-N*-NS-N* (25.5 repeats)	AAG02079
		AA	LTPDQVVAIASN*GGKQALETVQRLLPVQRLLPVLCQDHG	
		DNA	CTGACCCCGGACCAGGTCGTGGCCATCGCCAGCAATGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTACAGCGGCTG TTGCCGGTGCTGTGCCAGGACCATGGC	
40 amino acid-TALE repeats				
XOO2866_MAFF	*Xoo* MAFF 311018	RVDs	NI-HG-NI-NI-HG-HD-NN-HD-HD-HD-NI-NI-NN-NI-HD-HD-HD-HG-NN-NN-HD-NS-NN-HD-NG-NS-N* (26.5 repeats)	YP_451895
		AA	LTPDQVVAIANNNGGKQALETVQRLLPVQRLVPVLCQDHG	
XOO3014_KACC	*Xoo* KACC 10331	RVDs	NI-HG-NI-NI-NS-HD-NN-HD-HG-NI-NI-HG-HD-NN-HD-HD-HD-NI-NI-NN-NI-HD-HD-HD-HG-NN-NN-HD-NS-NN- HD-N*-NS-N* (33.5 repeats)	YP_201653
		AA	LTPDQVVAIANNNGGKQALETVQRLLPVQRLVPVLCQDHG	
		DNA	CTGACCCCGGACCAGGTGGTGGCCATCGCCAACAATAACGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTACAGCGG CTGGTGCCGGTGCTGTGCCAGGACCATGGC	
42 amino acid-TALE repeats				
Tal9b	*Xoo* PXO99[A]	RVDs	HD-HD-NN-NN-NG-NG-HD-NS-HG-HD-NG-N*-HD-HD-HD-N*-NN-NI-NN-HD-HI-ND-HD-HG-NN-HG-N* (26.5 repeats)	YP_00191509
		AA	LTPDQVVAIASNQVVAIASNIGGKQALETVQRLLPVLCQDHG	
XOO1136_MAFF	*Xoo* MAFF 311018	RVDs	HD-HD-NN-NN-NS-NG-HD-S*-HG-HD-NG-N*-HD-HD-HD-N*-NN-NI-NN-HD-HI-ND-HD-HG-NN-HG-N* (26.5 repeats)	YP_450165
		AA	LTPDQVVAIASNQVVAIASNIGGKQALETVQRLLPVLCQDHG	
		DNA	CTGACCCCGGACCAGGTGGTGGCCATCGCCAGTAATCAAGTCGTCGCGATTGCCAGTAATATTGGCGGCAAGCAGGCGCTGGAGACGGTG CAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGC	

[a] *Xoc*, *Xanthomonas oryzae* pv. *oryzicola*; *Xoo*, *Xanthomonas oryzae* pv. *oryzae*.

[b] RVDs, repeat variable diresidues; AAs, amino acids; DNA, DNA sequence.

[c] RVDs and AAs are in one letter amino acid code; RVDs of aberrant repeats and their coding bases are in red.

Supplementary Table 2 | Prediction of PthXo3 target sequences in rice promoters[a].

TALE[b]	Rank[c]	Annotation	Start	ATG[d]	Sequence	Matches[e]	Score	p-value[f]
PthXo3	1	Os01g63580.1: glycerol-3-phosphate acyltransferase, putative, expressed	268	90	TATATTAACTCCCTCTCCCGCTGCTCCACT	M|||x||x||||:x||:xx:x|x|xx:	-27,37985934	4,42548E-08
PthXo3	2	Os01g17170.1: magnesium-protoporphyrin IX monomethyl ester cyclase, chloroplast precursor, putative, expressed	323	44	TATATACCCGCCCACCACCTCCCCTCCCCC	M||||xx|:|||x|:|||xxxxx|x|:x:	-27,38871869	4,42548E-08
PthXo3	3	Os01g17170.2: magnesium-protoporphyrin IX monomethyl ester cyclase, chloroplast precursor, putative, expressed	323	44	TATATACCCGCCCACCACCTCCCCTCCCCC	M||||xx|:|||x|:|||xxxxx|x|:x:	-27,38871869	4,42548E-08
PthXo3	4	Os01g69160.1: regulatory protein, putative, expressed	197	73	CATAAAAACGCCCTATAAAAAAAAATCCCAG	m||x|||:|||x:|xxx:|xx|x|::x	-27,64408828	4,42548E-08
PthXo3	5	Os03g08880.1: purine permease, putative, expressed	328	52	CATCAACACACCCTAATCCAAAGGTGCCAA	m||xx|x||x|||xxx||x:|::|:|::x	-27,85167952	8,85096E-08
PthXo3	6	Os02g49350.1: plastocyanin-like domain containing protein, putative, expressed	269	82	TATATAAACTCCCCCGGCCAGAGACACCGC	M||||||x||x|xx||xx|:xxx|:x:	-27,95934861	8,85096E-08
PthXo3	7	Os02g57120.1: HEAT repeat family protein, putative, expressed	268	96	TATATAAACCTCCTCCCCAAAAGCCCCACT	M||||||xx|||:x|xx:|:xxx|xx:	-27,98858864	8,85096E-08
PthXo3	8	Os02g03280.2: transmembrane BAX inhibitor motif-containing protein, putative, expressed	269	150	TATATAAACCCCCCCCCCCGCCTCCCCTCT	M||||||x||x|:x||:xxxxxx|:x:	-28,06685038	1,32764E-07
PthXo3	9	Os01g19330.1: MYB family transcription factor, putative, expressed	265	98	TATATAAATCCCCGCCCCAAGAGCTACCCC	M||||||xx|||x|:x|xxx|:x|x|:x:	-28,23286173	1,32764E-07
PthXo3	10	Os05g45430.1: TOO MANY MOUTHS precursor, putative, expressed	264	65	CACATAAACTCCCCCCACCCCATCTCCTCT	m|x||||x|||x|:|||xx|xx|x|:x:	-28,24252784	1,32764E-07
PthXo3	118	Os03g57240.1: ZOS3-19 - C2H2 zinc finger protein, expressed	273	140	TATAGAAACACCAACCTCCTCATTTCCCCC	M|||x|||x||xx|:x||xx|xx|x|:x:	-31,90820162	3,58464E-06
PthXo3	**119**	**Os11g31190.1: nodulin MtN3 family protein, putative, expressed**	**267**	**203**	**TATATAAACCCCCTCCAACCAGGTGCTAAG**	**M||||||x||||:|x|x:x:xxxxx:x**	**-31,91182322**	**3,58464E-06**
PthXo3	120	Os01g42820.1: RNA recognition motif containing protein, putative, expressed	264	121	CCGATAAAACCCCTCGACCGATCCCACCCC	mxx||||xx||||x|||::xxxxx|:x:	-31,9561958	3,76166E-06
PthXo3_short	**1**	**Os11g31190.1: nodulin MtN3 family protein, putative, expressed**	**267**	**204**	**TATATAAACCCCCTCCAACCAGGTGCTAA**	**M||||||x||||xxxxx|::|:|::x**	**-21,3989497**	**0**
PthXo3_short	2	Os02g57120.1: HEAT repeat family protein, putative, expressed	268	97	TATATAAACCTCCTCCCCAAAAGCCCCAC	M||||||xx|||x||x:|x:xx|:::	-21,69820182	0
PthXo3_short	3	Os01g46570.1: CTP synthase,	269	189	TATTTAAAAGCCCTCACCTCGAGTGCTCC	M||x|||x:||||||xxxx:|:|:x:	-23,44192339	0

		putative, expressed						
PthXo3_short	4	Os01g46570.2: CTP synthase, putative, expressed	269	189	TATTTAAAAGCCCTCACCTCGAGTGCTCC	M||x|||x:|||||||xxxx:|:|:x:	-23,44192339	0
PthXo3_short	5	Os03g09150.1: pumilio-family RNA binding repeat domain containing protein, expressed	266	205	TATATAAACCCCCCCGCCGCTAGTGCGGC	M|||||||x|||x|x||:xxx:|:|xx:	-23,92888744	0
PthXo3_short	6	Os02g40010.1: phosphoribosyl transferase, putative, expressed	267	85	TATATATACCTCCTCACCCCACAGGCCAC	M||||x||xx||||||xx|xxx:|:::	-24,84380499	0
PthXo3_short	7	Os02g40010.2: phosphoribosyl transferase, putative, expressed	267	85	TATATATACCTCCTCACCCCACAGGCCAC	M||||x||xx||||||xx|xxx:|:::	-24,84380499	0
PthXo3_short	8	Os02g40010.3: phosphoribosyl transferase, putative, expressed	267	85	TATATATACCTCCTCACCCCACAGGCCAC	M||||x||xx||||||xx|xxx:|:::	-24,84380499	0
PthXo3_short	9	Os02g40010.4: phosphoribosyl transferase, putative, expressed	267	85	TATATATACCTCCTCACCCCACAGGCCAC	M||||x||xx||||||xx|xxx:|:::	-24,84380499	0
PthXo3_short	10	Os09g13460.1: expressed protein	51	220	TATATAATCCCCCTCCCCCGAGGTCATTC	M||||x|x||||x||xx|::|xx:x:	-24,93731372	0

[a] Predictions were done by applying TALgetter[5] to *Oryza sativa* Nipponbare promoters -300 to +200 of the transcriptional start site.

[b] PthXo3, bound in normal mode; PthXo3_short, bound with aberrant repeat looped out.

[c] Ranks 1 to 10 for both binding modes and ranks 118-120 for the normal mode are shown.

[d] Distance of TALE box to ATG.

[e] Categories of matches per position: M/m, first position match/mismatch; |, match; :, weak match; x, mismatch

[f] Empirical p-values were computed by generating a random data base of at least the total length of all scanned input sequences from a homogeneous Markov model of order 2 trained on the input sequences. For these random data the likelihoods of sub-sequences were computed in the same manner as for the actual input sequences. The p-value of a putative target site was calculated by determining the percentile corresponding to the observed likelihood value on the distribution of likelihoods obtained for the random data base[5].

Supplementary Table 3 | Selected primers used in this study.

Name	Nucleotide sequence[a]
Cloning of aberrant repeats	
30-F	CTGATCCCGGACCAGGTGGTGGCCATCGCCAGCAATATTGGCGGCAAGCAGGCG
30-R	GCACAGCACCGGCAACAGCCGCTGCACCGTCTCCAGCGCCTGCTTGCCGCC
39-F	CTGACCCCGGACCAGGTCGTGGCCATCGCCAGCAATGGCGGCAAGCAGGCG
39-R	GCCATGGTCCTGGCACAGCACCGGCAACAGCCGCTGTACCGGCAACAGCCGCTGCACCGTCTCCAGCGCCTGCTTGCCGC
40-F	CTGACCCCGGACCAGGTGGTGGCCATCGCCAACAATAACGGCGGCAAGCAGGCG
40-R	GCCATGGTCCTGGCACAGCACCGGCACCAGCCGCTGTACCGGCAACAGCCGCTGCACCGTCTCCAGCGCCTGCTTGCCGC
42-F	CTGACCCCGGACCAGGTGGTGGCCATCGCCAGTAATCAAGTCGTCGCGATTGCCAGTAATATTGGCGGCAAGCAGGCG
42-R	GCCATGGTCCTGGCACAGCACCGGCAACAGCCGCTGCACCGTCTCCAGCGCCTGCTTGCCGCC
Cloning of aberrant repeats as Golden TAL Technology modules	
Position 2	
2-F_30	TTT**GAAGAC**TTCTGATCCCGGACCAGGTGG
2-R_30	TTT**GAAGAC**TTCGGTATCAGGCACAGCACCGGCAACAGCCGC
2-F_39-42	TTT**GAAGAC**TTCTGACCCCGGACCAGGTGGTG
2-R_39-42	TTT**GAAGAC**TTCGGTGTCAGGCCATGGTCCTGGCACAGCAC
Position 3	
3-F_30-42	TTT**GAAGAC**TTACCGGACCAGGTGGTGGCCATCGCC
3-R_30	TTT**GAAGAC**TTGGTGAGGCACAGCACCGGCAACAGCCGC
3-R_39-42	TTT**GAAGAC**TTGGTGAGGCCATGGTCCTGGCACAGCAC
Cloning of TALE boxes in reporter vector	
Forward primer[b]	
AvrXa7 nat_Box	CACCTATATAAACCCCCTCCAACCAGGTGCTAATTCTTTCTTGTATATAACTTTGTCC
AvrXa7 opt_Box	CACCTATAAACGCCCATTCCAAGGATGATATTTCTTTCTTGTATATAACTTTGTCC
AvrXa7 oBox-1	CACCTATAAACGCCCATTCAAGGATGATATTTCTTTCTTGTATATAACTTTGTCC
AvrXa7 oBox-2	CACCTATAAACGCCCATTAAGGATGATATTTCTTTCTTGTATATAACTTTGTCC
AvrXa7 oBox+1	CACCTATAAACGCCCATTACCAAGGATGATATTTCTTTCTTGTATATAACTTTGTCC
AvrXa7 oBox+2	CACCTATAAACGCCCATTAACCAAGGATGATATTTCTTTCTTGTATATAACTTTGTCC
AGT2-box-1_40	CACCTTAGATAGAGTAGTGAGTTTCTTTCTTGTATATAACTTTGTCC
AGT2-box+1_40	CACCTTAGATAGAGCTTAGTGAGTTTCTTTCTTGTATATAACTTTGTCC
sTAL-box	CACCTTAGCTGCAGTACTAGCATTTCTTTCTTGTATATAACTTTGTCC
sTAL-box-1	CACCTTAGCTGCATACTAGCATTTCTTTCTTGTATATAACTTTGTCC
sTAL-box-2	CACCTTAGCTGCAACTAGCATTTCTTTCTTGTATATAACTTTGTCC
sTAL-box+1	CACCTTAGCTGCATGTACTAGCATTTCTTTCTTGTATATAACTTTGTCC
sTAL-box40pos3	CACCTTAGTGCAGTACTAGCATTTCTTTCTTGTATATAACTTTGTCC
sTAL-box40pos14	CACCTTAGCTGCAGTACTACATTTCTTTCTTGTATATAACTTTGTCC
sTAL-box-1p6F	CACCTTAGCTGAGTACTAGCATTTCTTTCTTGTATATAACTTTGTCC
sTAL-box-1p7F	CACCTTAGCTGCGTACTAGCATTTCTTTCTTGTATATAACTTTGTCC
-1p2/C/F	CACCTTA CTGCAGTACTAGCATTTCTTTCTTGTATATAACTTTGTCC
1mm p8 F	CACCTTAGCTGCTGTACTAGCATTTCTTTCTTGTATATAACTTTGTCC
2mm p8 F	CACCTTAGCTGCTCTACTAGCATTTCTTTCTTGTATATAACTTTGTCC
3mm p8 F	CACCTTAGCTGCTCAACTAGCATTTCTTTCTTGTATATAACTTTGTCC

1mm F	CACCTTAGCTGCACTACTAGCATTTCTTTCTTGTATATAACTTTGTCC
2mm F	CACCTTAGCTGCACAACTAGCATTTCTTTCTTGTATATAACTTTGTCC
sTAL-box3mm	CACCTTAGCTGCACATCTAGCATTTCTTTCTTGTATATAACTTTGTCC
PthXo3box	CACCTATATAAACGCCCTCTACCGAAGGTGCTATTTCTTTCTTGTATATAACTTTGTCC
PthXo3box-1	CACCTATATAAACGCCCTCTCCGAAGGTGCTATTTCTTTCTTGTATATAACTTTGTCC
PthXo3box-2	CACCTATATAAACGCCCTCTCGAAGGTGCTATTTCTTTCTTGTATATAACTTTGTCC
PthXo3box+1	CACCTATATAAACGCCCTCTAACCGAAGGTGCTATTTCTTTCTTGTATATAACTTTGTCC
PthXo3box+2	CACCTATATAAACGCCCTCTAAACCGAAGGTGCTATTTCTTTCTTGTATATAACTTTGTCC
Reverse primer	
Bs4PR/B01_R	AGATTCGATTAAAAATAAATTGTATGGATGAGATC
Cloning of *OsSWEET13* (*Os12N3*, *Os12g29220*) promoters in reporter vector	
Forward primer	
Xa25Zh/NipT65 1kb /F	CACCAGGGATGTCTACTGCAGGTGAAAAC
Reverse primer	
Xa25Zh/NipT65 1kb /R	TTTTGTGTGCTAAAAGGGGGGTAATTG

[a] *Bpi*I restriction sites are shown in bold.

[b] Sequences bound by TALEs are in red. Forward primer start with CACC to enable cloning into pENTR/D-TOPO (Invitrogen). 3' black letters indicate sequences complementary to the *Bs4* promoter

Supplementary Table 4 | TALE repeat amino acid sequences[a]

RVD	Repeat sequence
NI	LTPEQVVAIASNIGGKQALETVQRLLPVLCQAHG
HD	LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG
NN	LTPEQVVAIASNNGGKQALETVQRLLPVLCQAHG
NG	LTPEQVVAIASNGGGKQALETVQRLLPVLCQAHG
HG	LTPEQVVAIASHGGGKQALETVQRLLPVLCQAHG
NS	LTPEQVVAIASNSGGKQALETVQRLLPVLCQAHG
NH	LTPEQVVAIASNHGGKQALETVQRLLPVLCQAHG
N*	LTPEQVVAIASN*GGKQALETVQRLLPVLCQAHG

[a] amino acids in one letter code. A star (*) indicates a missing amino acid in the alignment. RVDs are in green.

Supplementary Note 1 | DNA and amino acid sequences

GFP-TALE34

DNA sequence[a]

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTA
AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAG
TTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTG
CAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGC
TACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTC
GAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTG
GGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGC
ATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAG
CAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCC
CTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATC
ACTCTCGGCATGGACGAGCTGTACAAGCATATGGATCCCATTCGTTCGCGCACACCAAGTCCTGCCCGC
GAGCTTCTGTCCGGACCCCAACCCGATGGGGTTCAGCCGACTGCAGATCGTGGGGTGTCTCCGCCTGCC
GGCGGCCCCCTGGATGGCTTGCCCGCTCGGCGGACGATGTCCCGGACCCGGCTGCCATCTCCCCCTGCC
CCCTCACCTGCGTTCTCGGCGGACAGCTTCAGTGACCTGTTACGTCAGTTCGATCCGTCACTTTTTAAT
ACATCGCTTTTTGATTCATTGCCTCCCTTCGGCGCTCACCATACAGAGGCTGCCACAGGCGAGTGGGAT
GAGGTGCAATCGGGTCTGCGGGCAGCCGACGCCCCCCCACCCACCATGCGCGTGGCTGTCACTGCCGCG
CGGCCGCCGCGCGCCAAGCCGGCGCCGCGACGACGTGCTGCGCAACCCTCCGACGCTTCGCCGGCCGCG
CAGGTGGATCTACGCACGCTCGGCTACAGCCAGCAGCAACAGGAGAAGATCAAACCGAAGGTTCGTTCG
ACAGTGGCGCAGCACCACGAGGCACTGGTCGGCCATGGGTTTACACACGCGCACATCGTTGCGCTCAGC
CAACACCCGGCAGCGTTAGGGACCGTCGCTGTCAAGTATCAGGACATGATCGCAGCGTTGCCAGAGGCG
ACACACGAAGCGATCGTTGGCGTCGGCAAACAGTGGTCCGGCGCACGCGCTCTGGAGGCCTTGCTCACG
GTGGCGGGAGAGTTGAGAGGTCCACCGTTACAGTTGGACACAGGCCAACTTCTCAAGATTGCAAAGCGT
GGCGGCGTGACCGCAGTGGAGGCAGTGCATGCATGGCGCAATGCACTGACGGGTGCCCCCCTGAACCTT
ACCCCGGAGCAGGTGGTGGCCATCGCCAGCAATGGCGGTGGCAAGCAGGCGCTGGAGACGGTGCAGCGG
CTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCAATATT
GGTGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGACA
CCGGAGCAGGTGGTGGCCATCGCCAGCAATCATGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTG
TTGCCGGTGCTGTGCCAGGCCCATGGCCTCACCCCGGAGCAGGTGGTGGCCATCGCCAGCCACGATGGC
GGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGACTCCG
GAGCAGGTGGTGGCCATCGCCAGCAATGGCGGTGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTG
CCGGTGCTGTGCCAGGCCCATGGCCTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCAATCATGGCGGC
AAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCGCATGGCCTTACCCCGGAG
CAGGTGGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCG
GTGCTGTGCCAGGCCCATGGCCTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCAATATTGGTGGCAAG
CAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGACACCGGAGCAG
GTGGTGGCCATCGCCAGCAATCATGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTG
CTGTGCCAGGCCCATGGCCTCACCCCGGAGCAGGTGGTGGCCATCGCCAGCAATGGCGGTGGCAAGCAG
GCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGACTCCGGAGCAGGTG
GTGGCCATCGCCAGCAATATTGGTGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTG
TGCCAGGCCCATGGCCTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCG
CTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCGCATGGCCTTACCCCGGAGCAGGTGGTG
GCCATCGCCAGCAATGGCGGTGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGC
CAGGCCCATGGCCTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCAATATTGGTGGCAAGCAGGCGCTG
GAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGACACCGGAGCAGGTGGTGGCC
ATCGCCAGCAATCATGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAG
GCCCATGGCCTCACCCCGGAGCAGGTGGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCGCTGGAG
ACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGACTCCGGAGCAGGTGGTGGCCATC
GCCAGCAATATTGGTGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCC
CATGGCCTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCAATGGCGGTGGCAAGCAGGCGCTGGAGAGC
ATTGTTGCCCAGTTATCTCGCCCTGATCCGGCGTTGGCCGCGTTGACCAACGACCACCTCGTCGCCTTG
GCCTGCCTCGGCGGACGTCCTGCGCTGGATGCAGTGAAAAAGGGATTGCCGCACGCGCCGGCCTTGATC

AAAAGAACCAATCGCCGTATTCCCGAACGCACATCCCATCGCGTTGCCGACCACGCGCAAGTGGTTCGC
GTGCTGGGTTTTTTCCAGTGCCACTCCCACCCAGCGCAAGCATTTGATGACGCCATGACGCAGTTCGGG
ATGAGCAGGCACGGGTTGTTACAGCTCTTTCGCAGAGTGGGCGTCACCGAACTCGAAGCCCGCAGTGGA
ACGCTCCCCCCAGCCTCGCAGCGTTGGGACCGTATCCTCCAGGCATCAGGGATGAAAAGGGCCAAACCG
TCCCCTACTTCAACTCAAACGCCGGATCAGGCGTCTTTGCATGCATTCGCCGATTCGCTGGAGCGTGAC
CTTGATGCGCCTAGCCCAATGCACGAGGGAGATCAGACGCGGGCAAGCAGCCGTAAACGGTCCCGATCG
GATCGTGCTGTCACCGGTCCCTCCGCACAGCAATCGTTCGAGGTGCGCGTTCCCGAACAGCGCGATGCG
TTGCATTTGCCCCTCCTCAGCTGGGGTGTAAAACGCCCGCGTACCAGGATCGGCGGCCTCCTGGATCCT
GGTACGCCCATGGATGCCGACCTGGTAGCGTCCAGCACCGTGGTTTGGGAACAAGATGCGGACCCCTTC
GCAGGGACAGCGGATGATTTCCCGGCATTCAACGAAGAGGAGCTCGCATGGTTGATGGAGCTATTGCCT
CAGTGA

[a]The sequence encoding the N-terminal GFP-tag is underlined.

GFP-TALE34

Amino acid sequence[b]

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGV
QCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNIL
GHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSA
LSKDPNEKRDHMVLLEFVTAAGITLGMDELYKHMDPIRSRTPSPARELLSGPQPDGVQPTADRGVSPPA
GGPLDGLPARRTMSRTRLPSPPAPSPAFSADSFSDLLRQFDPSLFNTSLFDSLPPFGAHHTEAATGEWD
EVQSGLRAADAPPPTMRVAVTAARPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSQQQQEKIKPKVRS
TVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLT
VAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPEQVVAIASNGGGKQALETVQR
LLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNHGGKQALETVQRL
LPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNGGGKQALETVQRLL
PVLCQAHGLTPEQVVAIASNHGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLP
VLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNHGGKQALETVQRLLPV
LCQAHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVL
CQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNGGGKQALETVQRLLPVLC
QAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNHGGKQALETVQRLLPVLCQ
AHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQA
HGLTPEQVVAIASNGGGKQALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALI
KRTNRRIPERTSHRVADHAQVVRVLGFFQCHSHPAQAFDDAMTQFGMSRHGLLQLFRRVGVTELEARSG
TLPPASQRWDRILQASGMKRAKPSPTSTQTPDQASLHAFADSLERDLDAPSPMHEGDQTRASSRKRSRS
DRAVTGPSAQQSFEVRVPEQRDALHLPLLSWGVKRPRTRIGGLLDPGTPMDADLVASSTVVWEQDADPF
AGTADDFPAFNEEELAWLMELLPQ*

[b]RVDs are shown in red. The N-terminal GFP-tag is underlined.

Forward TALEN34

DNA sequence[c]

ATGGAACAAAAATTAATCTCAGAAGAAGACTTGGCCAAGAAGAAGAGGAAGGTGCAGGTGGATCTACGC
ACGCTCGGCTACAGCCAGCAGCAACAGGAGAAGATCAAACCGAAGGTTCGTTCGACAGTGGCGCAGCAC
CACGAGGCACTGGTCGGCCATGGGTTTACACACGCGCACATCGTTGCGCTCAGCCAACACCCGGCAGCG
TTAGGGACCGTCGCTGTCAAGTATCAGGACATGATCGCAGCGTTGCCAGAGGCGACACACGAAGCGATC
GTTGGCGTCGGCAAACAGTGGTCCGGCGCACGCGCTCTGGAGGCCTTGCTCACGGTGGCGGGAGAGTTG
AGAGGTCCACCGTTACAGTTGGACACAGGCCAACTTCTCAAGATTGCAAAGCGTGGCGGCGTGACCGCA
GTGGAGGCAGTGCATGCATGGCGCAATGCACTGACGGGTGCCCCCCTGAACCTTACCCCGGAGCAGGTG
GTGGCCATCGCCAGCAATGGCGGTGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTG
TGCCAGGCCCATGGCCTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCAATATTGGTGGCAAGCAGGCG
CTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGACACCGGAGCAGGTGGTG
GCCATCGCCAGCAATCATGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGC
CAGGCCCATGGCCTCACCCCGGAGCAGGTGGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCGCTG
GAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGACTCCGGAGCAGGTGGTGGCC
ATCGCCAGCAATGGCGGTGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAG
GCCCATGGCCTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCAATCATGGCGGCAAGCAGGCGCTGGAG

ACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCGCATGGCCTTACCCCGGAGCAGGTGGTGGCCATC
GCCAGCCACGATGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCC
CATGGCCTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCAATATTGGTGGCAAGCAGGCGCTGGAGACG
GTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGACACCGGAGCAGGTGGTGGCCATCGCC
AGCAATCATGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCAT
GGCCTCACCCCGGAGCAGGTGGTGGCCATCGCCAGCAATGGCGGTGGCAAGCAGGCGCTGGAGACGGTG
CAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGACTCCGGAGCAGGTGGTGGCCATCGCCAGC
AATATTGGTGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGC
CTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCGCTGGAGACGGTGCAG
CGGCTGTTGCCGGTGCTGTGCCAGGCGCATGGCCTTACCCCGGAGCAGGTGGTGGCCATCGCCAGCAAT
GGCGGTGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTG
ACCCCGGAGCAGGTGGTGGCCATCGCCAGCAATATTGGTGGCAAGCAGGCGCTGGAGACGGTGCAGCGG
CTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGACACCGGAGCAGGTGGTGGCCATCGCCAGCAATCAT
GGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTCACC
CCGGAGCAGGTGGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTG
TTGCCGGTGCTGTGCCAGGCCCATGGCCTGACTCCGGAGCAGGTGGTGGCCATCGCCAGCAATATTGGT
GGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGACCCCG
GAGCAGGTGGTGGCCATCGCCAGCAATGGCGGTGGCAAGCAGGCGCTGGAGAGCATTGTTGCCCAGTTA
TCTCGCCCTGATCCGTCGTTGGCCGCGTTAACCAACGACCACCTCGTCGCCTTGGCCTGCCTCGGCGGA
CGTCCTGCGCTGGATGCAGTGAAAAAGGGATTGCCGCACGCGCCGGCCTTGATCAAAAGAACCAATCGC
CGTATTCCCGAACGCACATCCCATCGCGTTGCCGGATC**CCAGCTGGTGAAGAGCGAGCTGGAGGAGAAG**
AAGTCCGAGCTGCGGCACAAGCTGAAGTACGTGCCCCACGAGTACATCGAGCTGATCGAGATCGCCAGG
AACCCCACCCAGGACCGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACAGG
GGAGAGCACCTGGGCGGAAGCAGAAAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGATTAC
GGCGTGATCGTGGACACAAAGGCCTACAGCGGCGGCTACAATCTGCCTATCGGCCAGGCCGACGCGATG
CAGAGCTACGTGGAGGAGAACCAGACCCGGAATAAGCACATCAACCCCAACGAGTGGTGGAAGGTGTAC
CCTAGCAGCGTGACCGAGTTCAAGTTCCTGTTCGTGAGCGGCCACTTCAAGGGCAACTACAAGGCCCAG
CTGACCAGGCTGAACCACATCACCAACTGCAATGGCGCCGTGCTGAGCGTGGAGGAGCTGCTGATCGGC
GGCGAGATGATCAAAGCCGGCACCCTGACACTGGAGGAGGTGCGGCGCAAGTTCAACAACGGCGAGATC
AACTTCAGATCTTGA

[c]The sequences encoding the N-terminal cmyc-tag and nuclear localisation sequence (NLS) are underlined and in blue, respectively. The sequence encoding the C-terminal Sharkey-DS *Fok*I is in bold.

Forward TALEN34

Amino acid sequence[d]

MEQKLISEEDLAKKKRKVQVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAA
LGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTA
VEAVHAWRNALTGAPLNLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQA
LETVQRLLPVLCQAHGLTPEQVVAIASNHGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQAL
ETVQRLLPVLCQAHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNHGGKQALE
TVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALET
VQRLLPVLCQAHGLTPEQVVAIASNHGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNGGGKQALETV
QRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQ
RLLPVLCQAHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQR
LLPVLCQAHGLTPEQVVAIASNHGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRL
LPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNGGGKQALESIVAQL
SRPDPSLAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGS**QLVKSELEEK**
KSELRHKLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDY
GVIVDTKAYSGGYNLPIGQADAMQSYVEENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQ
LTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINFRS*

[d]RVDs are shown in red. The N-terminal cmyc-tag and nuclear localisation sequence (NLS) are underlined and in blue, respectively. The C-terminal Sharkey-DS *Fok*I domain is in bold.

Reverse TALEN$_{Bs4}$

DNA sequence[e]

ATGGATTACAAGGATGACGATGACAAGGCCAAGAAGAAGAGGAAGGTGCAGGTGGATCTACGCACGCTC
GGCTACAGCCAGCAGCAACAGGAGAAGATCAAACCGAAGGTTCGTTCGACAGTGGCGCAGCACCACGAG
GCACTGGTCGGCCATGGGTTTACACACGCGCACATCGTTGCGCTCAGCCAACACCCGGCAGCGTTAGGG
ACCGTCGCTGTCAAGTATCAGGACATGATCGCAGCGTTGCCAGAGGCGACACACGAAGCGATCGTTGGC
GTCGGCAAACAGTGGTCCGGCGCACGCGCTCTGGAGGCCTTGCTCACGGTGGCGGGAGAGTTGAGAGGT
CCACCGTTACAGTTGGACACAGGCCAACTTCTCAAGATTGCAAAGCGTGGCGGCGTGACCGCAGTGGAG
GCAGTGCATGCATGGCGCAATGCACTGACGGGTGCCCCCCTGAACCTTACCCCGGAGCAGGTGGTGGCC
ATCGCCAGCAATAACGGTGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAG
GCCCATGGCCTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCAATATTGGTGGCAAGCAGGCGCTGGAG
ACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGACACCGGAGCAGGTGGTGGCCATC
GCCAGCAATGGCGGTGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCC
CATGGCCTCACCCCGGAGCAGGTGGTGGCCATCGCCAGCAATATTGGTGGCAAGCAGGCGCTGGAGACG
GTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGACTCCGGAGCAGGTGGTGGCCATCGCC
AGCAATGGCGGTGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCAT
GGCCTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCAATGGCGGTGGCAAGCAGGCGCTGGAGACGGTG
CAGCGGCTGTTGCCGGTGCTGTGCCAGGCGCATGGCCTTACCCCGGAGCAGGTGGTGGCCATCGCCAGC
AATGGCGGTGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGC
CTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCAATGGCGGTGGCAAGCAGGCGCTGGAGACGGTGCAG
CGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGACACCGGAGCAGGTGGTGGCCATCGCCAGCAAT
AACGGTGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTC
ACCCCGGAGCAGGTGGTGGCCATCGCCAGCAATAACGGTGGCAAGCAGGCGCTGGAGACGGTGCAGCGG
CTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGACTCCGGAGCAGGTGGTGGCCATCGCCAGCAATATT
GGTGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGACC
CCGGAGCAGGTGGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTG
TTGCCGGTGCTGTGCCAGGCGCATGGCCTTACCCCGGAGCAGGTGGTGGCCATCGCCAGCAATATTGGT
GGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGACCCCG
GAGCAGGTGGTGGCCATCGCCAGCAATATTGGTGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTG
CCGGTGCTGTGCCAGGCCCATGGCCTGACACCGGAGCAGGTGGTGGCCATCGCCAGCAATATTGGTGGC
AAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTCACCCCGGAG
CAGGTGGTGGCCATCGCCAGCAATAACGGTGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCG
GTGCTGTGCCAGGCCCATGGCCTGACTCCGGAGCAGGTGGTGGCCATCGCCAGCAATGGCGGTGGCAAG
CAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGACCCCGGAGCAG
GTGGTGGCCATCGCCAGCAATGGCGGTGGCAAGCAGGCGCTGGAGAGCATTGTTGCCCAGTTATCTCGC
CCTGATCCGTCGTTGGCCGCGTTAACCAACGACCACCTCGTCGCCTTGGCCTGCCTCGGCGGACGTCCT
GCGCTGGATGCAGTGAAAAAGGGATTGCCGCACGCGCCGGCCTTGATCAAAAGAACCAATCGCCGTATT
CCCGAACGCACATCCCATCGCGTTGCCGGATCC**CAGCTGGTGAAGAGCGAGCTGGAGGAGAAGAAGTCC**
GAGCTGCGGCACAAGCTGAAGTACGTGCCCCACGAGTACATCGAGCTGATCGAGATCGCCAGGAACCCC
ACCCAGGACCGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACAGGGGAGAG
CACCTGGGCGGAAGCAGAAAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGATTACGGCGTG
ATCGTGGACACAAAGGCCTACAGCGGCGGCTACAATCTGCCTATCGGCCAGGCCAGAGAGATGCAGAGA
TACGTGGAGGAGAACCAGACCCGGAATAAGCACATCAACCCCAACGAGTGGTGGAAGGTGTACCCTAGC
AGCGTGACCGAGTTCAAGTTCCTGTTCGTGAGCGGCCACTTCAAGGGCAACTACAAGGCCCAGCTGACC
AGGCTGAACCACATCACCAACTGCAATGGCGCCGTGCTGAGCGTGGAGGAGCTGCTGATCGGCGGCGAG
ATGATCAAAGCCGGCACCCTGACACTGGAGGAGGTGCGGCGCAAGTTCAACAACGGCGAGATCAACTTC
AGATCTTGA

[e]The sequences encoding the N-terminal FLAG-tag and nuclear localisation sequence (NLS) are underlined and in blue, respectively. The sequence encoding the C-terminal Sharkey-RR *Fok*I is in bold.

Reverse TALEN$_{Bs4}$

Amino acid sequence[f]

MDYKDDDDKAKKKRKVQVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALG
TVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVE
AVHAWRNALTGAPLNLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALE
TVQRLLPVLCQAHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALET
VQRLLPVLCQAHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNGGGKQALETV
QRLLPVLCQAHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNGGGKQALETVQ
RLLPVLCQAHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGKQALETVQR
LLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRL
LPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLL
PVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGKQALETVQRLLP
VLCQAHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNGGGKQALESIVAQLSR
PDPSLAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGS**QLVKSELEEKKS**
ELRHKLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGV
IVDTKAYSGGYNLPIGQAREMQRYVEENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLT
RLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINFRS*

[f]RVDs are shown in red. The N-terminal FLAG-tag and nuclear localisation sequence (NLS) are underlined and in blue, respectively. The C-terminal Sharkey-RR *Fok*I domain is in bold.

***Os12g29220* allelic promoters**

CLUSTAL 2.1 multiple DNA sequence alignment

> *Oryza sativa* cv Nipponbare
> *Oryza sativa* cv Zhenshan 97

```
AGGGATGTCTACTGCAGGTGAAAACAATCCTTCGACAAAAAATAAGTTACTTTTGGTAAA  60
AGGGATGTCTACTGCAGGTGAAAACAATCCTTCGACAAAAAATAAGTTACTTTTGGTAAA  60

GACAGTTAAATAATAAGCAGCTATATCACGCGCATGGGAGAATTGCATATTCAATTACAA 120
GACAGTTAAATAATAAGCAGCTATATCACGCGCATGGGAGAATTGCATATTCAATTACAA 120

TCATTATTTTTTTTTCAGAACACTGTCGGCGACATTGAGAATTAATCTACCCGTGCAAAC 180
TCATTATTTTTTTTTCAGAACACTGTCGGCGACATTGAGAATTAATCTACCCGTGCAAAC 180

AAAGAACAGAGAAACTATAGTATACCTACATGTATCTATCACCCAATAATTGCAAGATCA 240
AAAGAACAGAGAAACTATAGTATACCTACATGTATCTATCACCCAATAATTGCAAGATCA 240

TGTTACAAAACGGTTCTAATTAATATATAGAAACAAGGCAGAGAATTCTACCTTTCTTTT 300
TGTTACAAAACGGTTCTAATTAATATATAGAAACAAGGCAGAGAATTCTACCTTTCTTTT 300

GTCTAAGTACAATTATCTTTTTCTCCGCGATTAATATTTTTCGAGTAGTAAAATTTAAGT 360
GTCTAAGTACAATTATCTTTTTCTCCGCGATTAATATTTTTCGAGTAGTAAAATTTAAGT 360

CAAAAGCCGTATCAGGATTCAGGAATAATCCTTCACTGGGAGAGATCTCATGTGATTTGC 420
CAAAAGCCGTATCAGGATTCAGGAATAATCCTTCACTGGGAGAGATCTCATGTGATTTGC 420
```

```
TGTTGCACTCGGCGGCTATCTTTTACCGTTCCCAGCAGGAAGCTGCAGACGTTGGAGAGA 480
TGTTGCACTCGGCGGCTATCTTTTACCGTTCCCAGCAGGAAGCTGCAGACGTTGGAGAGA 480

TCGATCTCTACTGACAATGCACAAAGCAATTACTCACTAAATTGGCTATGGCTAGTGAGA 540
TCGATCTCTACTGACAATGCACAAAGCAATTACTCACTAAATTGGCTATGGCTAGTGAGA 540

GGTGCGCTGCGCACAAAGCCAATGCAACTTTTTTTGAAAATTAGCCAGGATTATCTCCAA 600
GGTGCGCTGCGCACAAAGCCAATGCAACTTTTTTTGAAAATTAGCCAGGATTATCTCCAA 600

CAGTAGCTCATTTTTGTAAAAGCCTAATTATTGTGCGTGTCCAAAAGACTTTCCTCAAAA 660
CAGTAGCTCATTTTTGTAAAAGCCTAATTATTGTGCGTGTCCAAAAGACTTTCCTCAAAA 660

GCAAATAAAGAAAAAAAATCTTTGCATAATTATTCTATGATTACTTTGATGCGTACGTGA 720
GCAAATAAAGAAAAAAAATCTTTGCATAATTATTCTATGATTACTTTGATGCGTACGTGA 720

ATGGCCATGGGTAGGAGGCAACCAAGTGATTCCCACCTAGCTAGCTTTGCTCCTATATAA 780
ATGGCCATGGGTAGGAGGCAACCAAGTGATTCGCACCTAGCTAGCTTTTCTCCTATATAA 780

AGCACCACAACTCCCTTCATTCCTCTCCAAGAGTTTTCAGCCAACACATTGAACTCTTCT 840
-GCACCACAACTCCCTTCATTCCTCTCCAAGAGTTTTCAGCCAACACATTGAACTCTTCT 839

TCAGAGCTCTCCCTTCCCTCCACAAAGGGGGTCTAGGGTTAGAGTGTGTGTGTCTGTGAC 900
TCAGAGCTCTCCCTTCCCTCCACAAAGGGG-TCTAGGGTTAGAGTGTGTGTGTCTGTGAC 898

AAGTTCCAAGCTAGCAACAACAAGCTCAATTCCTTGCTTGTTTGCTTCCATATTACACTA 960
AAGTTCCAAGCTAGCAACAACAAGCTCAATTCCTTGCTTGTTTGCTTCCATATTACACTA 958

CATCTCTTCCCTTCAATTACCCCCCTTTTAGCACACAAAA 1000
CATCTCTTCCCTTCAATTACCCCCCTTTTAGCACACAAAA 998
```

ArtXa251/ArtXa252 target sequences are underlined. The PthXo2 target sequence is in bold. Single nucleotide exchanges are shown in grey.

New assembly vectors for the "Golden TAL Technology" toolbox[6]

One to six repeats can be assembled using Golden Gate cloning into *Bpi*I sites (underlined) of an assembly vector. The repeats in the assembly vectors are flanked by *Bsa*I sites (black, bold) to facilitate assembly into the final TALE. Two new assembly vectors were designed to expand the maximum number of repeats per TALE from 23.5[6] to 29.5. Part of the assembly vector sequences are shown. *Bsa*I overhangs are in colors (C: blue, D: magenta, R: dark grey; according to[6]). *Bpi*I overhangs matching to the first and last repeat of the repeat assembly are shaded in light and dark grey, respectively. Relevant repeat codons are translated.

Assembly vector C-D

```
GTCGGTCTCAGTGCCAGGCGCATGGCCTTACAAGTCTTCCTT•••lacZ•••
            C  Q  A  H  G  L  T

•••GGCGAAGACTTTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTTGAGACCCAC
             L  E  T  V  Q  R  L  L  P  V  L
```

Assembly vector D-R

```
GTCGGTCTCACTGTGCCAGGCGCATGGCCTTACAAGTCTTCCTT•••lacZ•••
          L  C  Q  A  H  G  L  T

•••GGCGAAGACTTTGGAGAGCATGAGACCCAC
             L  E  S
```

Supplementary References

1. Römer, P. *et al.* 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 (2010).
2. Antony, G. *et al.* Rice *xa13* recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene Os-*11N3*. *Plant Cell* **22**, 3864-3876 (2010).
3. Streubel, J., Blücher, C., Landgraf, A. & Boch, J. TAL effector RVD specificities and efficiencies. *Nat. Biotechnol.* **30**, 593-595 (2012).
4. Liu, Q. *et al.* A paralog of the MtN3/saliva family recessively confers race-specific resistance to *Xanthomonas oryzae* in rice. *Plant Cell. Environ.* **34**, 1958-1969 (2011).
5. Grau, J. *et al.* Computational predictions provide insights into the biology of TAL effector target sites. *PLoS Comput. Biol.* **9**, e1002962 (2013).
6. Geißler, R. *et al.* Transcriptional activators of human genes with programmable DNA-specificity. *PLoS ONE* **6**, e19509 (2011).