

Erstellung und Analyse von TALE-Transkriptionsfaktoren mit
benutzerdefinierter Erkennungsspezifität

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)

vorgelegt der
FAKULTÄT FÜR BIOLOGIE

der
LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

von Robert Morbitzer

geboren am 16.02.1983 in Berlin-Mitte

Gutachter:

1. Prof. Dr. Thomas Lahaye

2. Prof. Dr. Martin Parniske

Tag der mündlichen Prüfung: 19.07.2013

Martinsried, den 01.11.2012

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbstständig und ohne unerlaubte Hilfe angefertigt ist.

Martinsried, den 01.11.2012

.....

Erklärung

Hiermit erkläre ich, *

- dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist.
- dass ich mich anderweitig einer Doktorprüfung ohne Erfolg **nicht** unterzogen habe.
- dass ich mich mit Erfolg der Doktorprüfung im Hauptfach und in den Nebenfächern bei der Fakultät für der unterzogen haben.
- dass ich ohne Erfolg versucht habe, eine Dissertation einzureichen oder mich der Doktorprüfung zu unterziehen.

Martinsried, den 01.11.2012

.....

*) Nichtzutreffendes streichen

Zusammenfassung

Proteine, die spezifisch an DNA binden, sind Schlüsselkomponenten der Molekularbiologie. *Transcription activator like effector* (TALE)-Proteine phytopathogener Bakterien der Gattung *Xanthomonas* binden spezifisch an Promotoren pflanzlicher Wirtsgene und aktivieren diese transkriptionell. Dabei wird die Spezifität der DNA-Bindung durch die zentrale *Repeat*-Domäne der TALEs bestimmt. Diese besteht aus sich wiederholenden, 33-35 Aminosäuren langen *Repeat*-Einheiten, die sich hauptsächlich an den Aminosäure-Positionen 12 und 13 unterscheiden. Die präferentielle Interaktion dieser als *repeat variable diresidues* (RVDs) bezeichnen Aminosäuren mit einem Nukleotid der Erkennungssequenz führte zur Entschlüsselung des TALE-Codes. Um zu prüfen, ob neben den RVDs auch Nicht-RVDs der *Repeats* einen Einfluss auf die Erkennungsspezifität haben, wurden im Rahmen der vorliegenden Arbeit zwei TALE-Chimären generiert, bei denen die RVDs in den Nicht-RVD Hintergrund des anderen eingesetzt wurden. Die Analyse der Chimären zeigte, dass die Nicht-RVDs keinen Einfluss auf die DNA-Erkennungsspezifität haben. Folglich ermöglichen Änderung der RVDs es, TALEs (*designer* TALEs, dTALEs) mit gewünschter Erkennungsspezifität zur transkriptionellen Aktivierung beliebiger Gene zu erzeugen. Durch die Modifizierung der RVDs des *Xanthomonas* TALE-Proteins AvrBs3 konnten spezifisch die Promotoren des Tomaten *Bs4S*-, des *Arabidopsis* *EGL3*- und des *Arabidopsis* *KNAT1*-Promoter transkriptionell induziert werden. Mutationen in den TALE-Bindestellen verhinderten indes eine Aktivierung der jeweiligen Promotoren. *Xanthomonas* AvrBs3 induziert in Paprika unter anderem die Expression des *Bs3*- und *UPA20*-Gens. Durch die Erzeugung von zwei AvrBs3-Varianten mit vier zusätzlichen *Repeat*-Monomeren konnten die Promotoren von *Bs3* und *UPA20* jeweils spezifisch aktiviert werden. Dies zeigt, dass durch zusätzliche RVDs die Erkennungsspezifität von TALEs erhöht werden kann.

Um *Repeat*-Module und damit dTALEs mit gewünschter Erkennungsspezifität herzustellen, wurde im Rahmen dieser Arbeit zudem ein Klonierungsansatz unter Verwendung von Typ IIS-Restriktionsenzymen entwickelt. Dieser erlaubt eine schnelle, effiziente und kostengünstige Erzeugung von dTALEs.

Summary

Proteins, binding specifically to desired DNA sequences, are key tools for molecular biology. Transcription activator-like effectors (TALEs) from the bacterial genus *Xanthomonas* bind to host-gene promoters in a sequence specific manner, thereby activating them transcriptionally. DNA-binding specificity is defined by the central Repeat-domain, more precisely by Repeat-variable diresidues (RVDs) of tandem-arranged 33-35-amino acid Repeat units. The preferential interaction of these RVDs (amino acid position 12 and 13 of each Repeat) with one type of nucleotide of the recognition sequence uncovered the TALE-Code. To examine, whether beside RVDs also non-RVDs influence the recognition specificity, two chimeras were created, by inserting RVDs of one in the non-RVDs background of the other and vice versa. These studies revealed that non-RVDs had no effect on the DNA-binding specificity. Suggesting, that modifying the RVDs should be sufficient to generate designer TALEs (dTALEs), capable of activating transcription of user-defined target genes. By changing only the RVDs of the TALE AvrBs3 the tomato *Bs4S*-, the *Arabidopsis EGL3*- or the *Arabidopsis KNAT1*-promoter were induced transcriptionally. Mutations within the targeted DNA sequences abolished promoter activation. *Xanthomonas* AvrBs3 activates the pepper *Bs3*- and *UPA20* gene. By adding four additional Repeat units to AvrBs3, either the pepper *Bs3* or *UPA20* promoter were activated specifically. These data show that addition of Repeat units facilitates increased TALE-specificity.

To generate Repeat-modules and thereby dTALEs with pre-defined DNA binding specificities a cloning approach was developed to assemble multiple Repeat-encoding DNA fragments (Repeat-arrays). This method makes use of type IIS restriction enzymes in two sequential cut-ligase reactions to build dTALE Repeat arrays and enables the rapid, flexible and low-cost production of dTALEs for gene regulation and genome editing in routine applications.

Inhaltsverzeichnis

Eidesstattliche Erklärung	II
Erklärung	II
Zusammenfassung	III
Summary	IV
Inhaltsverzeichnis	V
Abbildungsverzeichnis	VII
Tabellenverzeichnis	VII
Abkürzungsverzeichnis	VIII
1 Einleitung	1
1.1 Grundlagen der DNA-Bindung durch DNA-bindende Proteine	1
1.2 DNA-Bindedomänen mit modulierbarer Spezifität	1
1.2.1 Zinkfingerproteine	1
1.2.1.1 Vorkommen und Funktion	1
1.2.1.2 Struktur der DNA-Bindedomäne	2
1.2.1.3 Modulierbarkeit der DNA-Bindedomäne.....	3
1.2.1.4 Biotechnologische Anwendungen von ZF-Fusionsproteinen	4
1.2.2 Meganukleasen	5
1.2.2.1 Vorkommen und Funktion	5
1.2.2.2 Struktur von LAGLIDADG-Meganukleasen	5
1.2.2.3 Modulierbarkeit der LMN-DNA-Bindedomäne	6
1.2.2.4 Biotechnologische Anwendungen von LMN.....	7
1.2.3 <i>Xanthomonas-Transcription activator like effectors</i> (TALEs).....	7
1.2.3.1 Grundlagen der Xanthomonas-TALE-Pflanze-Interaktion	7
1.2.3.2 TALE-Code.....	9
1.2.3.3 TALE-Struktur	10
1.2.3.4 Biotechnologische Anwendungen von TALEs	11
1.3 Zielstellung der vorliegenden Arbeit	12
2 Ergebnisse	13
2.1 Übersicht über Veröffentlichungen welche Grundlage der vorliegenden Arbeit sind	13

2.2	Übersicht über Veröffentlichungen welche nicht Grundlage der vorliegenden Arbeit sind	13
2.3	Manuskript 1	14
2.4	Manuskript 2	44
3	Diskussion	61
3.1	Die Spezifität der TALE-DNA-Bindedomäne ist modulierbar	61
3.1.1	Nicht-RVDs haben keinen Einfluss auf die Erkennungsspezifität	61
3.1.2	C-terminal angefügte <i>Repeats</i> erhöhen die Erkennungsspezifität	62
3.1.3	Der RVD NK besitzt im Vergleich zu NN eine höhere Spezifität für G	65
3.1.4	Die Optimierung der RVD-Abfolge erhöht die Erkennungsspezifität	66
3.2	TALE-basierte Repeat-Module mit benutzerdefinierter Erkennungsspezifität können über verschiedene Methoden generiert werden	67
3.2.1	Die PCR-basierte Modifikation von TALE-DNA-Bindedomänen mit <i>Xanthomonas</i> -Nukleotidsequenz ist schwer durchführbar	67
3.2.2	Die Veränderung der <i>codon usage</i> ermöglicht die PCR-basierte Modifikation von TALE-DNA-Bindedomänen	68
3.2.3	Multi-Fragment-Assemblierungen ermöglichen die schnelle und effiziente Erstellung von <i>Repeat</i> -Modulen	69
3.2.4	Die für die Mehrfragment-Ligation benutzen 4 bp-Überhänge können durch die Verwendung der Degenerierung des genetischen Codes und die Verschiebung des Fusionspunktes optimiert werden	72
3.2.5	Vergleichende Analyse bisher verwendeter Strategien zur Erzeugung von TALE-DNA-Bindedomänen	74
3.2.5.1	Plasmid-basierte Repeat-Bibliotheken sind weniger fehleranfällig als PCR-basierte Repeat-Bibliotheken	74
3.2.5.2	Repeat-Monomer-basierte Mehrfragment- und Zweifragment-Ligationen unterscheiden sich hinsichtlich der Dauer der Vorarbeiten, der Repeat-Modul-Erstellung und der Flexibilität der Repeat-Länge	74
3.2.5.3	Durch die Verwendung von Repeat-Polymeren nähern sich Mehrfragment- und Zweifragment-Ligation hinsichtlich der Dauer der Repeat-Modul-Erstellung und der Vorarbeiten an	78
4	Literaturverzeichnis	82
5	Anhang	93
	Danksagung	104
	Lebenslauf	105

Abbildungsverzeichnis

Abb. 1 Struktur eines C ₂ H ₂ -Zinkfingermotivs.....	2
Abb. 2 Schematische Abbildung der Zinkfinger-DNA-Interaktion.....	3
Abb. 3 Darstellung der I-Crel-Meganuklease mit der 24 Basenpaar- Erkennungsstelle.....	6
Abb. 4 Grundlagen der <i>Xanthomonas</i> -Pflanzen-Interaktion.....	8
Abb. 5 TALE-Code	10
Abb. 6 Struktur des TALE-Repeat-DNA-Komplexes.	11
Abb. 7 Darstellung der Fehlpaarungsquote in Abhängigkeit der Repeat-Anzahl.	64
Abb. 8: Prinzip der MultiSite Gateway®-Klonierung.....	70
Abb. 9: Prinzip der In-Fusion-HD-Klonierung.	70
Abb. 10: Prinzip der <i>Golden Gate</i> -Klonierung.	72
Abb. 11: Prinzip der Zweifragment-Ligation.	75
Abb. 12: Aminosäuresequenz der Repeats von AvrBs3 sortiert nach RVD-Typ.....	93
Abb. 13: Aminosäuresequenz der Repeats von AvrBs4 sortiert nach RVD-Typ.....	94
Abb. 14: Vergleich der Nukleotid-Identität der <i>avrBs3</i> -Repeats 1 bis 17 mit <i>Xanthomonas codon usage</i> (A) und <i>N. benthamiana codon usage</i> (B) in Prozent.	99
Abb. 15: Der Fehlpaarungstoleranz-Analyse zugrundeliegende TALE-Promotor- Interaktionen.	102

Tabellenverzeichnis

Tabelle 1: Vergleich der Mehrfragment-Ligation nach Morbitzer et al. (2011) und Zweifragment-Ligation nach Huang et al. (2011) hinsichtlich Dauer und Zahl der Klonierungsschritte für die Erstellung eines TALEs mit 17.5 Repeats.	76
Tabelle 2: Vergleich der Monomer-, Dimer- bzw. Polymer-basierten Mehrfragment- und Zweifragment-Ligationen zur Erstellung eines TALEs mit 17.5 Repeats.	79
Tabelle 3: Anzahl der bei der Mehrfragment- bzw. Zweifragment-Ligation benötigten Plasmide.....	80
Tabelle 4: Zusammenfassung der Fehlpaarungstoleranz-Analyse.	103

Abkürzungsverzeichnis

<i>avr</i> , Avr	Avirulenz
AAD	<i>acidic activation domain</i> , saure Aktivierungsdomäne
As	Aminosäuren
bp	Basenpaare
<i>Bs</i>	<i>bacterial spot</i>
C-terminal	carboxyterminal
DNA	Desoxyribonukleinsäure
dTALE	<i>designer</i> TALE (TALE mit benutzerdefinierter Erkennungsspezifität)
<i>EGL3</i>	<i>enhancer of glabra 3</i>
<i>GUS</i>	beta-glucuronidase
<i>hax</i>	Homolog von AvrBs3 in <i>Xanthomonas</i>
<i>KNAT1</i>	<i>knotted-like from arabidopsis thaliana 1</i>
LMN	LAGLIDADG-Meganukleasen
M1	Manuskript 1
M2	Manuskript 2
NHEJ	<i>Non-homologous end joining</i>
<i>N. benthamiana</i>	<i>Nicotiana benthamiana</i>
N-terminal	aminoterminal
NLS	<i>nuclear localisation signal</i> , Kern-Lokalisierungssignal
PCR	polymerase chain reaction, Polymerase-Ketten-Reaktion
RVD	<i>repeat-variable diresidues</i>
TALE	<i>transcription activator like effector</i>
TALEN	<i>transcription activator like effector-Nuklease</i>
<i>UPA</i>	upregulated by AvrBs3
<i>UPT</i>	up regulated by TALE
ZF	Zinkfinger
ZFN	Zinkfingernukleasen

1 Einleitung

1.1 Grundlagen der DNA-Bindung durch DNA-bindende Proteine

Interaktionen zwischen Proteinen und DNA sind wichtig für viele biologische Prozesse, wie z.B. während der Transkription, Translation, Replikation und Rekombination. Dabei wird grundsätzlich in sequenzspezifische Protein-DNA-Interaktionen (Proteine zur Chromosomenorganisation z.B. Histone oder *high-mobility group* Proteine) und sequenzspezifische Protein-DNA-Interaktionen unterschieden (z.B. Transkriptionsfaktoren). Bei sequenzspezifischen Interaktionen wird die Abfolge der Basen anhand der chemischen Eigenschaften der großen Furche der DNA erkannt. Dies erfolgt über Wasserstoffbrückenbindungen, die zwischen dem DNA-bindenden Protein und den in die große Furche ragenden funktionellen Gruppen der DNA ausgebildet werden. Die Interaktion in der großen Furche erfolgt in der Regel über eine Protein α -Helix (Rohs et al., 2010).

Für molekularbiologische und biotechnologische Anwendungen (z.B. artifizielle Transkriptionsfaktoren, Rekombinasen, Nukleasen und epigenetische Modulatoren) sind besonders DNA-bindende Proteine mit Basenerkennung geeignet, deren Sequenzspezifität veränderbar ist. Ein solches Protein sollte aus Modulen bestehen, welche die Erkennung von einem oder wenigen Nukleotiden spezifisch vermittelt. Durch die Kombination mehrerer Module sollte dann eine definierte Zielsequenz angesteuert werden können. Die bisher am Besten charakterisierten modularen DNA-bindenden Proteine sind Zinkfingerproteine (ZF).

1.2 DNA-Bindedomänen mit modulierbarer Spezifität

1.2.1 Zinkfingerproteine

1.2.1.1 Vorkommen und Funktion

Die Mehrzahl der ZF gehören der Klasse der C_2H_2 -Zinkfinger an. Diese Klasse wurde bei Untersuchungen der TFIIIA-Interaktion mit der 5S RNA in *Xenopus* entdeckt und ist ein auch beim Menschen weit verbreitetes Proteinmotiv. So kodieren 3% der menschlichen Gene für C_2H_2 -ZF (Klug, 2010; Miller et al., 1985).

Viele der C_2H_2 -ZF stellen das DNA-Bindemodul von Transkriptionsaktivatoren und Transkriptionsrepressoren dar. So ist das C_2H_2 -ZF-Motiv z.B. in der Hälfte der humanen Transkriptionsfaktoren vorhanden (Messina et al., 2004). Neben der DNA-

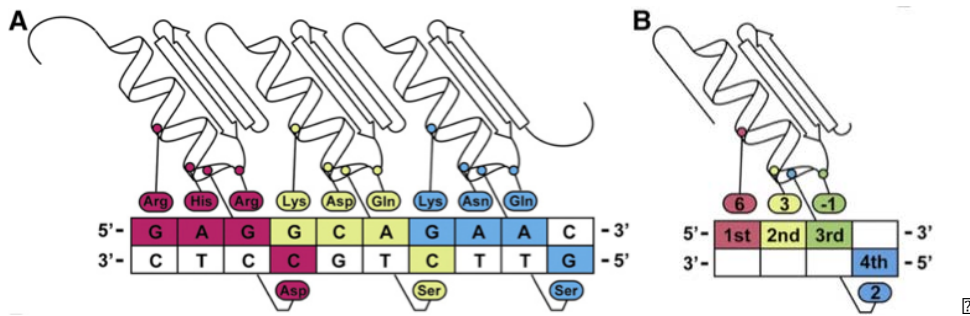


Abb. 2: Schematische Abbildung der Zinkfinger-DNA-Interaktion. (A) Darstellung der Interaktionen eines Zinkfingers-Modul (ZF 1 in lila, ZF 2 in gelb und ZF 3 in blau) mit der DNA-Erkennungssequenz. Farbige Quadrate stellen Nukleotide dar, die direkt mit einer Aminosäure (farbige Ovale mit Namen der Aminosäure im Dreibuchstabencode) der ZF interagieren. Die Position der Aminosäure im ZF ist durch einen farbigen Punkt gekennzeichnet. (B) Darstellung der Interaktionen eines ZF mit der DNA. Die Zahlen innerhalb des ZF bezeichnen die Position der Aminosäure-Reste der mit der DNA interagierenden α -Helix bzw. *loop*-Region. Zwischen dem ZF und der DNA ausgebildete Wasserstoffbrücken sind durch gleiche Farben der ZF-Aminosäure und DNA-Nukleotide gekennzeichnet (aus Sera, 2009).

StDStN

Durch flexible Aminosäuresequenzen zwischen den einzelnen ZF (*linker*), die nicht Teil der ZF-kodierenden Sequenz sind, ist es möglich, mehrere Zinkfinger aneinander zu koppeln und die Zahl der gebundenen Nukleotide zu erhöhen. Benachbarte ZF sind allerdings bezüglich ihrer Basenspezifität nicht völlig unabhängig, da die Aminosäure an Position 2 der α -Helix mit dem letzten Nukleotid des vorherigen *triplets* interagiert (Abb. 2) (Isalan et al., 1997). Somit lässt sich die DNA-Spezifität mehrerer zusammengesetzter ZF nicht zuverlässig vorhersagen. Folglich sind aufwändige Sichtungs- und Optimierungsprozesse notwendig, um zusammengesetzte ZFs mit gewünschter Basenspezifität zu identifizieren.

Die Spezifität eines ZF ist durch Mutagenese der Positionen -1, 2, 3 und 6 der α -Helix veränderbar (Berg, 1988). Alternativ können ZF mit neuen Bindeeigenschaften auch über Phagen-Display aus Peptid- bzw. Protein-Bibliotheken über ihre Zielsequenz affinitätsgereinigt werden (Choo und Klug, 1994a, b).

Basierend auf diesen ZF mit neuen Bindeeigenschaften gibt es verschiedene Methoden, um ZF-Module aus drei bzw. vier ZF herzustellen. So werden z.B. aus einer Bibliothek an ZF diejenigen zusammengefügt, deren einzelne Bindeeigenschaften die Nukleotid-*triplets* der Zielsequenz repräsentieren (Modulare Assemblierung) (Beerli und Barbas, 2002; Liu et al., 2002). Allerdings wird bei

diesem Ansatz die Interferenz eines ZF mit dem vorherigen *triplet* nicht berücksichtigt, so dass die resultierenden, zusammengesetzten ZF-Module oft nicht aktiv sind (Ramirez et al., 2008).

Alternative Methoden berücksichtigen die Kontextabhängigkeit und die finale Position des ZF im ZF-Modul (Greisman und Pabo, 1997; Hurt et al., 2003; Isalan et al., 2001; Maeder et al., 2008). Da diese Methoden auf speziellem Wissen zur Herstellung großer ZF-Bibliotheken beruhen und die Nutzung laborintensive Selektionsmethoden beinhaltet, sind diese für viele Wissenschaftler nur schwer zugänglich (Cathomen und Joung, 2008).

Des Weiteren gibt es eine Assemblierungsmethode, bei der ausgehend von einer Bibliothek mit ZF-Dimeren jeweils zwei dieser Dimere zusammengesetzt, getestet und weiter optimiert werden (Moore et al., 2001; Urnov et al., 2005). Diese Ressource (ca. 4000 ZF-Dimere) ist allerdings nur für den akademischen Bereich in Kooperation mit Sangamo BioSciences (USA) nutzbar (Cathomen und Joung, 2008). Sie ist auch die Grundlage kommerzieller ZF, die über Sigma-Aldrich (USA) vertrieben werden (Baker, 2012).

1.2.1.4 Biotechnologische Anwendungen von ZF-Fusionsproteinen

Durch die Verknüpfung mehrerer ZF können DNA-Sequenzen *in vitro* und *in vivo* spezifisch gebunden werden. Somit besteht die Möglichkeit, ZF für diagnostische, biochemische und gentherapeutische Anwendungen zu nutzen. Die Fusion von Aktivierungs- bzw. Repressor-Domänen ermöglicht, die Genaktivität von Zielgenen spezifisch zu verändern (Choo et al., 1994; Papworth et al., 2003; Reynolds et al., 2003). Die größte biotechnologische Bedeutung wird ZF-Nukleasen (ZFN) zugeschrieben. Dies sind Fusionen aus ZF-Modulen und der Restriktionsdomäne des Typ IIS Restriktionsenzym FokI (Kim et al., 1996). Für die Nukleaseaktivität ist die Dimerisierung der FokI-Monomere notwendig. Daher müssen zwei ZFN in einem definierten Abstand (5 bis 7 Basenpaare) an gegenüberliegenden DNA-Strängen binden, um ein aktives FokI-Dimer zu konstituieren (Smith et al., 2000). Der erste ZFN-vermittelte *in vivo knock-out* wurde in *Drosophila* durchgeführt (Bibikova et al., 2002). Seitdem wurden ZFN erfolgreich in Mäusen, Menschen, Motten, Nematoden, Ratten, Seeigeln, Zebrafischen und einigen Pflanzenspezies (*Arabidopsis*, Mais und Tabak) angewandt (Carroll, 2011). ZFN-induzierte Doppelstrangbrüche bedingen die Aktivierung zellulärer

Reparatursysteme. So kann beispielsweise durch *Non-homologous end joining* (NHEJ), der Doppelstrangbruch geschlossen werden. Dieser Prozess ist jedoch häufig fehlerhaft und resultiert oft in Nukleotid-Insertionen bzw. -Deletionen (Moore und Haber, 1996). Durch die simultane Verwendung zweier ZFN-Paare können auch große Bereiche (10^2 bis 15^6 Basenpaare) deletiert werden (Lee et al., 2010). Neben NHEJ können Doppelstrangbrüche auch über homologe Rekombination geschlossen werden. Voraussetzung ist die Verfügbarkeit einer Donor-DNA, die Homologie zum 5'- und 3'-Bereich des Bruchpunktes besitzt. Über den Ansatz der homologen Rekombination können Genaustausche bzw. Geninsertionen realisiert werden (Capecchi, 1989).

1.2.2 Meganukleasen

1.2.2.1 Vorkommen und Funktion

Homing Endonukleasen/Meganukleasen sind DNA schneidende Enzyme, die in Mikroben aller biologischen Gattungen sowie mit Phagen- bzw. Viren-assoziierten Organismen vorkommen. Durch die Generierung von DNA-Doppelstrangbrüchen an definierten Stellen im Genom sind sie in der Lage, die sie kodierende Sequenz zu mobilisieren. Diese Aktivität führt zu ortsspezifischen Rekombinationsereignissen, in deren Folge es zu Insertionen oder Deletionen in der DNA kommen kann. Meganukleasen werden, basierend auf ihrer Sequenz und Struktur, in fünf Klassen eingeteilt (LAGLIDADG, GIY-YIG, HNH, His-Cys box und PD-(D/E)XK). Die biotechnologisch relevanteste Klasse ist die Klasse der LAGLIDADG-Meganukleasen (LMN), auf die im Folgenden ausschließlich eingegangen werden soll (Stoddard, 2011).

1.2.2.2 Struktur von LAGLIDADG-Meganukleasen

Neben der Funktion der DNA-Bindung besitzen Meganukleasen auch die für die Erzeugung von Doppelstrangbrüchen notwendige Nukleaseaktivität. Allerdings sind beide Funktionen nicht in getrennten Domänen organisiert, sondern ungleichmäßig über das gesamte Protein verteilt (Abb. 3). Neben den als Homodimere vorkommenden und folglich eine palindromische Erkennungssequenz bindenden LMN gibt es auch LMN, die aus zwei monomeren Proteinen zusammengesetzt sind, die über einen Peptidlinker verbundene sind. Dieser Typ ist in der Lage,

asymmetrische Sequenzen zu binden. Beide LMN-Typen bilden einen Proteinkern, der durch eine charakteristische $\alpha\beta\alpha\beta\alpha$ -Struktur gekennzeichnet ist.

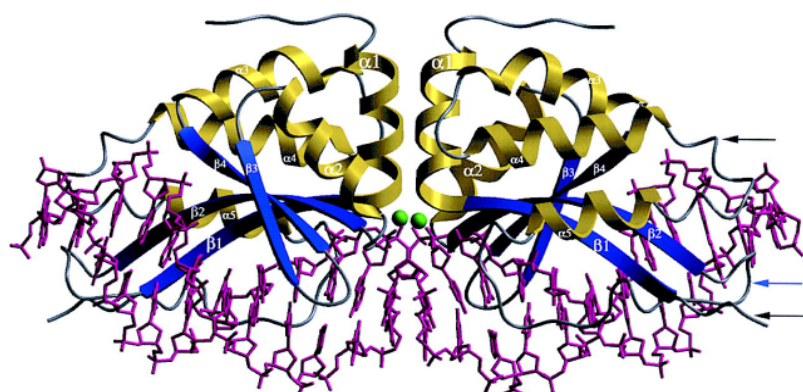


Abb. 3: Darstellung der I-CreI-Meganuklease mit der 24 Basenpaar-Erkennungsstelle. Die strukturellen Domänen der beiden I-CreI-Monomere sind in blau (β -Faltblätter) und gelb (α -Helices) dargestellt. Die durch Interaktionen mit den β -Faltblättern gebundene DNA ist in lila dargestellt (aus Jurica et al., 1998).

LMN sind in der Lage, DNA-Sequenzen zwischen 14 und 40 Nukleotiden zu binden (Silva et al., 2011). Durch die Ausbildung von Dimeren vergrößern Meganukleasen ihre Interaktionsoberfläche mit der DNA, wodurch z.B. die I-CreI-Meganuklease trotz ihrer relativ geringen Größe (163 Aminosäuren) eine 22bp lange DNA-Sequenzen erkennen kann. Für die Bindung kommt es zu einer Interaktion der antiparallelen β -Faltblätter mit der großen Furche der DNA im Bereich der Erkennungsstelle (Abb. 3) (Jurica et al., 1998). Im Rahmen der Erkennung werden eine Reihe von spezifischen und unspezifischen Interaktionen ausgebildet, die ungleichmäßig über die DNA-Erkennungsstelle verteilt sind (Chevalier et al., 2003).

1.2.2.3 Modulierbarkeit der LMN-DNA-Bindedomäne

Für die Veränderung der Spezifität von LMN-DNA-Bindedomänen sind bisher aufwendige Mutagenese- und Selektionsprozesse notwendig, die auf der Bindung bzw. dem Schneiden von gewünschten Erkennungssequenzen beruhen (Pingoud und Wende, 2011). Allerdings können dadurch nur geringe Veränderungen zur ursprünglichen Erkennungssequenz der Wildtyp-LMN erreicht werden. Bisher sind durch diesen Prozess einige Meganukleasen mit neuer Erkennungsspezifität *de novo* generiert worden (Arnould et al., 2006; Smith et al., 2006).

Die Kombination von Domänen bzw. Untereinheiten verschiedener LMN ist eine weitere Methode, neue LMN mit neuen Erkennungssequenzen herzustellen

(Chevalier et al., 2002; Epinat et al., 2003). Dabei ist es möglich die LMN-Domänen für die linke und rechte Hälfte der Erkennungsstelle getrennt voneinander zu optimieren und anschließend zusammenzuführen.

Eine weitere Möglichkeit der Modulierung der Erkennungsspezifität beruht auf der Nutzung vorhandener Kristallstrukturen. Anhand dieser Strukturdaten können *in silico* Veränderungen an den DNA-Kontaktstellen der LMN vorhergesagt werden, die eine veränderte DNA-Spezifität bedingen (Ashworth et al., 2006). Neben einer „Umprogrammierung“ können über diesen Weg auch Kontaktstellen eliminiert werden, um beispielsweise LMN herzustellen, die mit kürzeren Bindestellen interagieren.

Da sich Änderungen der DNA-Spezifität meistens auch auf die Nukleaseaktivität auswirken, ist die Modulierung der Erkennungsspezifität in der Regel nur schwer durch ein nichtspezialisiertes Labor zu realisieren. Demzufolge werden die meisten LMN kommerziell hergestellt und sind durch die damit verbundenen Kosten für viele Grundlagenforscher nicht nutzbar (Webb, 2011).

1.2.2.4 Biotechnologische Anwendungen von LMN

Unter den Meganukleasen bilden die LMN die Grundlage der meisten Enzyme für biotechnologische Anwendungen (Stoddard, 2011). Bisher wurden LMN gegen drei humane Gene sowie ein Gen in Mais erstellt (Arnould et al., 2007; Chapdelaine et al., 2010; Gao et al., 2010; Grizot et al., 2009). Besonders die hohe Spezifität und damit verbundene geringe Zytotoxizität machen LMN für medizinische Anwendungen interessant. Da die DNA-Bindefunktion und die Nukleasefunktion jedoch nicht von getrennten Proteindomänen ausgeübt werden, ist es nicht möglich, die DNA-Bindesepezifität mit einer anderen enzymatischen Eigenschaft zu kombinieren.

1.2.3 *Xanthomonas-Transcription activator like effectors* (TALEs)

1.2.3.1 Grundlagen der *Xanthomonas*-TALE-Pflanze-Interaktion

Während der Infektion durch das pflanzenpathogene Bakterium *Xanthomonas* werden zwischen 30 und 40 Effektorproteine mittels Typ III-Sekretionssystem in die Pflanzenzelle übertragen (Furutani et al., 2009; Roden et al., 2004; Thieme et al., 2005). Eine Klasse dieser Effektorproteine sind *transcription activator like effector* (TALE) Proteine, die nach dem Import in den Zellkern an definierte

Zielsequenzen binden (Abb. 4). Die damit verbundene Induktion von Wirtsgenen fördert die Kolonisierung und/oder Verbreitung von *Xanthomonas* in der Pflanze (Büttner und Bonas, 2010).

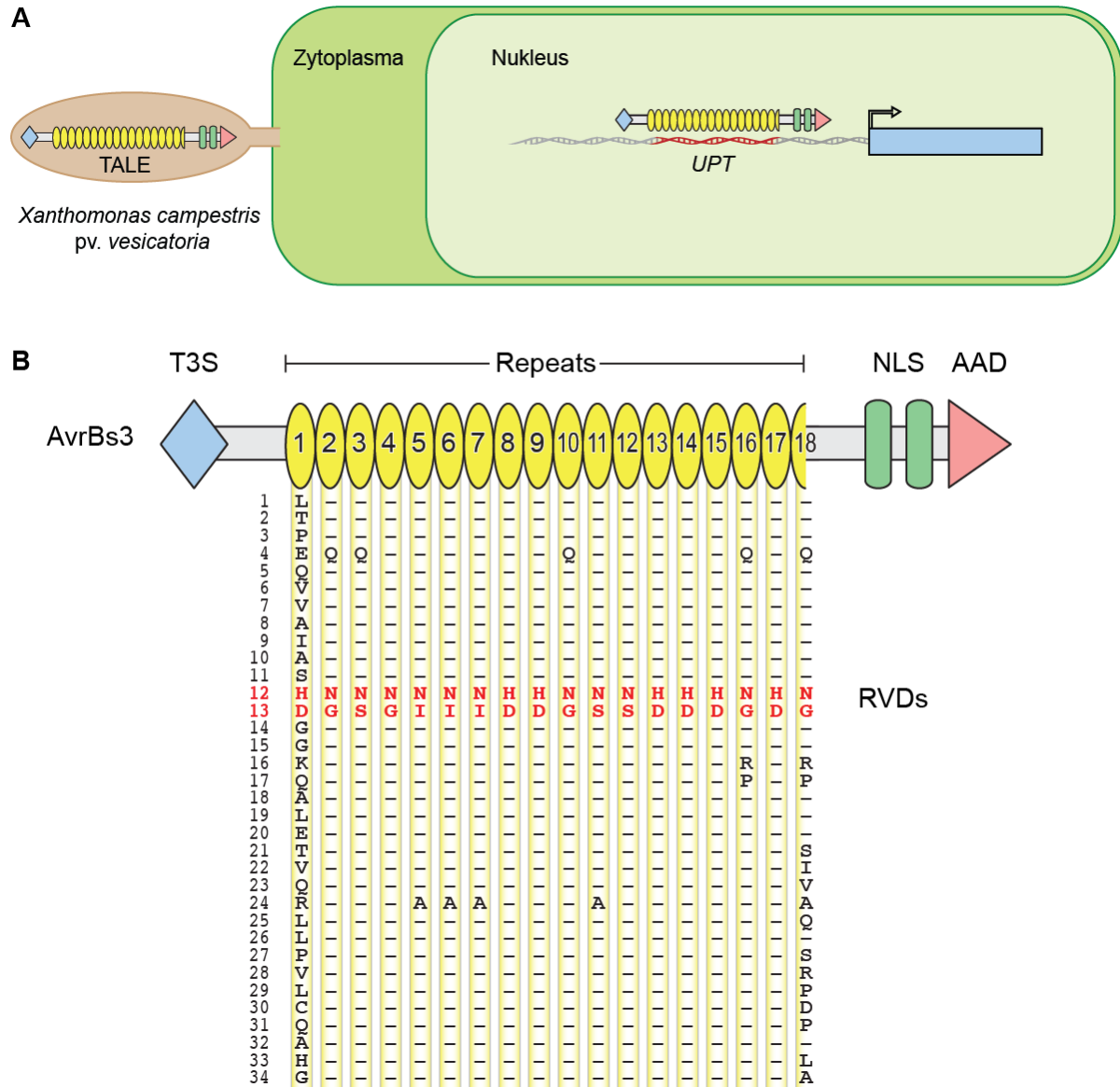


Abb. 4: Grundlagen der *Xanthomonas*-Pflanzen-Interaktion. (A) Rolle von TALEs in *Xanthomonas*-Infektionen. Die *Xanthomonas* TALEs werden durch das vorhandene Typ III-Sekretionssignal (blaue Raute, T3S) mit Hilfe des Typ III-Sekretionssystems in das Zytoplasma der Wirtszelle transloziert. Kernlokalisierungssignale (grüne Ovale, NLS; *nuclear localisation signals*) vermitteln den Kernimport. Hintereinandergeschaltete *Repeats* (gelbe Ovale) sind für die Bindung an eine DNA-Zielsequenz notwendig (*UPT*; *up regulated by TALE*) (roter DNA-Abschnitt). Dies führt zur Induktion von Wirtsgenen (blaues Rechteck) durch die Aktivierungsdomäne (AAD; *acidic activation domain*) (rotes Dreieck). (B) Schematische Darstellung des TALE-Proteins AvrBs3. Die Aminosäureabfolge der *Repeats* ist unter den von 1 bis 18 nummerierten *Repeats* mithilfe des Einbuchstabencodes dargestellt. Die innerhalb der *Repeats* am schwächsten konservierten Aminosäuren, die sogenannten *Repeat-variable diresidues (RVDs)*, sind rot und fett gedruckt dargestellt. Zum ersten *Repeat* identische Aminosäuren sind in nachfolgenden *Repeats* mit einem waagerechten Strich gekennzeichnet. B verändert nach (Bogdanove et al., 2010).

TALEs besitzen im aminoterminalen (N-terminalen) Bereich ein für die Translokation notwendiges Typ III-Sekretionssignal sowie im carboxyterminalen (C-terminalen) Bereich Kernlokalisierungssignale (*nuclear localisation signal*, NLS) und eine für Transkriptionsfaktoren typische saure Aktivierungsdomäne (*acidic activation domain*, AAD). Der zentrale Bereich besteht aus einer variablen Anzahl von Wiederholungen eines 33 bis 35 Aminosäure langen Motivs (*Repeats*). Dabei unterscheiden sich die *Repeats* hauptsächlich an den Positionen 12 und 13 (Abb. 4), welche als *Repeat-variable diresidues* (RVDs) bezeichnet werden (Boch et al., 2009; Moscou und Bogdanove, 2009). Weiter Positionen, in denen sich z.B. die *Repeats* von AvrBs3 und AvrBs4 unterscheiden, sind Position 4, 15, 16 und 24 (Abb. 12 und Abb. 13). Bisher sind etwa 20 verschiedene RVDs von TALEs bekannt, wobei die häufigsten RVDs HD, NG, NI und NN (HD: Histidin und Aspartat; NG: Asparagin und Glycin; NI: Asparagin und Isoleucin; NN: Asparagin und Asparagin) etwa 75% der insgesamt in TALEs vorkommenden RVDs ausmachen (Abb. 5) (Moscou und Bogdanove, 2009). Für die *Repeats* konnte bereits die Fähigkeit zur spezifischen DNA-Bindung gezeigt werden (Kay et al., 2007; Römer et al., 2007). Basierend darauf sowie durch das Vorhandensein der AAD kommt es zur transkriptionellen Induktion von Wirtsgenen. Der Transkriptionsstart wird dabei durch die *UPT*-Box (*UPT*; *up regulated by TALE*) des TALE definiert und befindet sich etwa 40 bis 60 Nukleotide 3' zur Erkennungssequenz. Durch das Verschieben der Erkennungssequenz kann auch der Transkriptionsstart verändert werden (Römer et al., 2009a; Römer et al., 2009b).

1.2.3.2 TALE-Code

Nachdem für einige TALEs die Aktivierung von Ziel promotoren gezeigt werden konnte (Gu et al., 2005; Kay et al., 2007; Römer et al., 2007; Sugio et al., 2007; Yang et al., 2006), führte der Vergleich der Nukleotide der Erkennungssequenz mit den RVDs des korrespondierenden TALEs zur Entschlüsselung des sogenannten TALE-Codes (Abb. 5) (Boch et al., 2009). Zum gleichen Ergebnis führte auch die computergestützte Analyse einer Reihe von TALEs und von den Promotoren ihrer Zielgene (Moscou und Bogdanove, 2009). Dem TALE-Code zufolge interagiert ein RVD eines TALEs mit einem Nukleotid der Erkennungssequenz. Anhand der vorhandenen Daten wurden Präferenzen von RVDs zu bestimmten Nukleotiden bestimmt (HD mit C, NG mit T, NI mit A und NN mit G bzw. A siehe Abb. 5). Die in

Moscou et al. (2009) beschriebenen RVD-Nukleotid-Interaktionen beinhalten drei RVDs (NK, HN, NA), die ausschließlich mit G interagieren. Allerdings beruhen diese RVD-Nukleotid-Interaktionen nur auf einer oder zwei Interaktionen (Abb. 5).



Abb. 5: TALE-Code. Darstellung der in Moscou et al. (2009) gezeigten RVD-Nukleotid-Assoziationen. Die Korrelation eines Nukleotides (C, A, T und G) mit dem RVD des zugehörigen *Repeats* ist in prozentualer Häufigkeit angegeben. Unter jedem RVD ist dessen absolute Häufigkeit in den untersuchten TALEs angegeben (aus Bochtler, 2012).

Unter Verwendung des TALE-Codes ist es möglich, Zielgene für bekannte TALEs vorherzusagen und TALEs mit benutzerdefinierter Erkennungsspezifität (*designer* TALEs; dTALEs) zur Induktion von Zielgenen zu generieren. So wurden bereits artifizielle *UPT*-Boxen für die *Xanthomonas campestris* pv. *amoraciae*-TALEs Hax2, Hax3 und Hax4 erstellt, für die bis zu diesem Zeitpunkt *in planta* keine Promotorinteraktionen bekannt waren (Boch et al., 2009). Des Weiteren konnte in Boch et al. (2009) die Funktionalität von dTALEs gezeigt werden, deren *Repeats* in zufälliger Reihenfolge zusammengesetzt wurden.

Sowohl in Moscou et al. (2009) als auch in Boch et al. (2009) wurde außerdem die Notwendigkeit eines T Nukleotids (Position -1; T₋₁) vor der Erkennungsstelle identifiziert. Auch in vorangegangenen Arbeiten wurde bereits gezeigt, dass die Stärke der DNA-Bindung eines Fragmentes, das nur die *Repeats* enthält, zehn Mal schwächer ist als ein Fragment, das weitere N- und C-terminale Aminosäuren besitzt (Kay et al., 2007). Dies führte zur Postulierung eines sogenannten nullten *Repeat*, der für die Interaktion mit dem T₋₁ verantwortlich ist.

1.2.3.3 TALE-Struktur

Jeder *Repeat* besteht aus zwei Helices, deren Verbindungssequenz die RVDs beinhaltet und mit der DNA interagiert. Die Aminosäure an Position 12 stabilisiert dabei die Struktur der Schleife. Durch die Aminosäure an Position 13 wird ein nukleotidspezifischer Kontakt zur großen Furche der DNA hergestellt. Weiterhin wurde gezeigt, dass die *Repeats* eine rechtsgängige Superhelix um die DNA bilden

(Abb. 6). Vor dem ersten *Repeat* werden zwei weitere *Repeats* ausgebildet (*Repeat* -1 und *Repeat* 0), die allerdings auf Aminosäureebene nur sehr geringe Homologie zu den nachfolgenden *Repeats* haben. Ein Tryptophan aus der die beiden *Repeats* verbindenden Schleife vermittelt die Interaktion zum T₋₁ (Deng et al., 2012; Mak et al., 2012).

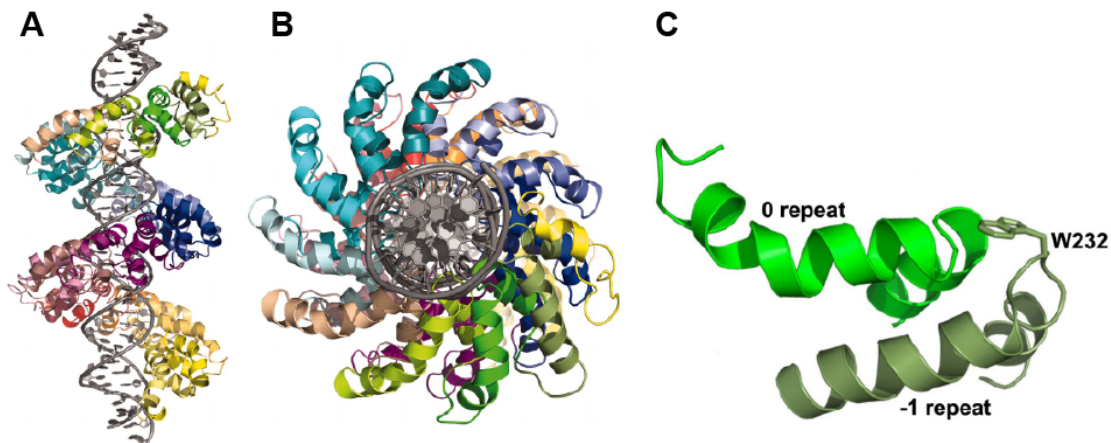


Abb. 6: Struktur des TALE-Repeat-DNA-Komplexes. Darstellung der TALE-*Repeats* von PthXo1 (farbige Proteinstrukturen) die sich rechtsgängig um die DNA (grau) winden von der Seite (A) bzw. von oben (B). (C) Darstellung des *Repeat* 0 und *Repeat* -1 sowie des Tryptophan an Position 232, das mit dem T₋₁ Nukleotid vor der *UPT*-Box interagiert. (aus Mak et al., 2012)

1.2.3.4 Biotechnologische Anwendungen von TALEs

Neben der Nutzung zur spezifischen Genaktivierung wird die TALE-DNA-Bindedomäne auch verwendet um Endonukleasen an definierte Sequenzabschnitte zu dirigieren. Dafür werden TALE-*Repeats* mit gewünschter Spezifität an die sequenzunspezifische FokI-Nuklease translational fusioniert. Die Funktionalität der resultierenden TALE-Nukleasen (TALENs) wurden erfolgreich in Hefe (Li et al., 2011b), Pflanzen (Cermak et al., 2011; Mahfouz et al., 2011), *C.elegans* (Wood et al., 2011), Zebrafisch (Huang et al., 2011; Sander et al., 2011), Ratte (Tesson et al., 2011) und humanen somatischen (Cermak et al., 2011; Miller et al., 2011; Mussolino et al., 2011) sowie pluripotenten Stammzellen (Hockemeyer et al., 2011) nachgewiesen.

Eine weitere Anwendung von TALEs besteht in der Nutzung als Transkriptionsrepressoren durch den Austausch der Aktivierungsdomäne gegen Repressordomäne (Cong et al., 2012; Mahfouz et al., 2012).

1.3 Zielstellung der vorliegenden Arbeit

Das Hauptziel der vorliegenden Arbeit bestand in der Erstellung von TALE-basierten DNA-Bindedomänen mit benutzerdefinierter Erkennungsspezifität. Dafür musste zuerst überprüft werden, ob die DNA-Bindesepezifität von TALE-Transkriptionsfaktoren ausschließlich von RVDs determiniert wird, oder ob auch andere Bereiche der *Repeats* einen Einfluss auf die DNA-Bindesepezifität haben. Auf der Basis dieser Erkenntnisse sollten dann TALE-Transkriptionsfaktoren mit benutzerdefinierter DNA-Bindesepezifität *de-novo* generiert und anschließend analysiert werden. Ein weiterer Aspekt war die Erarbeitung von Wegen zur Optimierung der DNA-Bindedomäne von TALE-Transkriptionsfaktoren, um hochspezifisch Sequenzen in komplexen Organismen anzusteuern.

Parallel zu diesen Arbeiten sollte ein modulares Klonierungssystem zur Erstellung von Genen entwickelt werden, die *Repeat*-Module von TALE-Transkriptionsfaktoren mit gewünschter Spezifität kodieren.

2 Ergebnisse

2.1 Übersicht über Veröffentlichungen welche Grundlage der vorliegenden Arbeit sind

Morbitzer, R., Römer, P., Boch, J., und Lahaye, T. (2010). Regulation of selected genome loci using *de novo*-engineered transcription activator-like effector (TALE)-type transcription factors. *P Natl Acad Sci USA* 107, 21617-21622.

Eigenanteil: Planung, Konstrukt-Erstellung (mit Ausnahme der Plasmide pENTR-D-*Bs3P_{UPT}_{AvrBs4}* und pGWB3-*Xa27_{P:uidA}*), Durchführung und Auswertung aller Experimente, Erstellung aller Abbildungen, Entwurf der Abbildungsunterschriften und des Material und Methodenteils

Morbitzer, R., Elsaesser, J., Hausner, J., und Lahaye, T. (2011). Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Res.* 39, 5790-5799.

Eigenanteil: Planung, etwa 30% der Konstrukt-Erstellung, Durchführung und Auswertung aller Experimente, Erstellung aller Abbildungen, Entwurf der Abbildungsunterschriften und des Material und Methodenteils

2.2 Übersicht über Veröffentlichungen welche nicht Grundlage der vorliegenden Arbeit sind

Mussolino, C., Morbitzer, R., Lutge, F., Dannemann, N., Lahaye, T., und Cathomen, T. (2011). A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res.* 39, 9283-9293.

Bultmann, S., Morbitzer, R., Schmidt, C. S., Thanisch, K., Spada, F., Elsaesser, J., Lahaye, T. und Leonhardt, H. (2012). Targeted transcriptional activation of silent *oct4* pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Res.* 40, 5368-5377.

2.3 Manuskript 1

Morbitzer, R., Römer, P., Boch, J., und Lahaye, T. (2010). Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *P Natl Acad Sci USA* 107, 21617-21622.

Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors

Robert Morbiter^a, Patrick Römer^a, Jens Boch^b, and Thomas Lahaye^{a,1}

^aGenetics, Department of Biology I, Ludwig-Maximilians-University Munich, 82152 Martinsried, Germany; and ^bInstitute of Biology, Martin-Luther-University, 06099 Halle, Germany

Edited* by Brian J. Staskawicz, University of California, Berkeley, CA, and approved November 1, 2010 (received for review September 6, 2010)

Proteins that can be tailored to bind desired DNA sequences are key tools for molecular biology. Previous studies suggested that DNA-binding specificity of transcription activator-like effectors (TALEs) from the bacterial genus *Xanthomonas* is defined by repeat-variable diresidues (RVDs) of tandem-arranged 34/35-amino acid repeat units. We have studied chimeras of two TALEs differing in RVDs and non-RVDs and found that, in contrast to the critical contributions by RVDs, non-RVDs had no major effect on the DNA-binding specificity of the chimeras. This finding suggests that one needs only to modify the RVDs to generate designer TALEs (dTALEs) to activate transcription of user-defined target genes. We used the scaffold of the TALE AvrBs3 and changed its RVDs to match either the tomato *Bs4*, the *Arabidopsis* *EGL3*, or the *Arabidopsis* *KNAT1* promoter. All three dTALEs transcriptionally activated the desired promoters in a sequence-specific manner as mutations within the targeted DNA sequences abolished promoter activation. This study is unique in showing that chromosomal loci can be targeted specifically by dTALEs. We also engineered two AvrBs3 derivatives with four additional repeat units activating specifically either the pepper *Bs3* or *UPA20* promoter. Because AvrBs3 activates both promoters, our data show that addition of repeat units facilitates TALE-specificity fine-tuning. Finally, we demonstrate that the RVD NK mediates specific interaction with G nucleotides that thus far could not be targeted specifically by any known RVD type. In summary, our data demonstrate that the TALE scaffold can be tailored to target user-defined DNA sequences in whole genomes.

zinc-finger proteins | AvrBs4 | AvrXa27

The ability to achieve site-specific manipulation of genomes has widespread implications for basic and applied research, and much effort has been devoted to identify protein folds that can be tailored to bind user-defined DNA sequences (1). Recent studies uncovered that transcription activator-like effector proteins (TALEs) from the plant pathogenic bacterial genus *Xanthomonas* contain a unique type of DNA-binding domain (2, 3). TALEs are modular proteins that are composed of (i) an N-terminal translocation signal, (ii) a central DNA binding domain, and (iii) a C-terminal region containing nuclear localization signals and a transcriptional activation domain (4, 5). The TALE DNA binding domain is composed of a variable number of tandemly arranged, imperfect 34/35 amino acid repeat units. Repeat unit polymorphisms are found predominantly at positions 12 and 13, referred to as the repeat-variable diresidues (RVDs), but occur also within non-RVD repeat unit positions (4, 5). Recent studies showed that repeats with different RVDs recognize different DNA base pairs and that consecutive RVDs in a TALE correspond directly to the sequence of nucleotides in the binding site, one-to-one in a linear and contiguous fashion. This TALE code explains the specificity of TALEs and enables the prediction of corresponding DNA target sequences (6–8).

Plant pathogenic xanthomonads inject TALEs into host cells via a syringe-like type III secretion system to promote bacterial growth or spread. Within the host cell, TALEs translocate to the nucleus, bind via their DNA binding domain to corresponding *UPT* (Up-

regulated by TALE) boxes and activate downstream host genes. The primary function of TALEs is to transcriptionally activate host-susceptibility (*S*) genes that, by mostly unknown means, promote growth or spread of the bacterial pathogen (4, 5). However, in the course of plant-pathogen coevolution, some plants evolved disease resistance (*R*) genes that mediate the perception of individual TALEs. AvrBs3 from the pepper (*Capsicum* spec.) and tomato (*Solanum lycopersicum*) pathogenic strain *Xanthomonas campestris* pv. *vesicatoria* (*Xcv*) is one TALE for which corresponding plant *S* and *R* genes have been studied. In the context of a susceptible plant genotype, AvrBs3 has been shown to interact with a corresponding *UPT*_{AvrBs3} box present in the promoter of the pepper transcription factor *UPA20* (2). Expression of the host *S* gene *UPA20* induces hypertrophy of mesophyll cells, which is believed to promote the release of *Xcv* from infected leaves (2, 9). If the given pepper plant contains the *R* gene *Bs3*, AvrBs3 triggers a programmed cell death, referred to as the hypersensitive response (HR) (3, 10). Certain pepper lines have an allele of *Bs3* known as *Bs3-E*, which confers resistance to strains carrying the AvrBs3 derivative AvrBs3Δrep16 that has a deletion of repeat units 11 to 14, compared with AvrBs3 (3, 10). AvrBs3 and AvrBs3Δrep16 interact specifically with distinct *UPT* boxes in the *Bs3* and *Bs3-E* promoters, respectively, and activate transcription of the corresponding coding sequences, leading to HR (10).

The *Xcv* TALE protein AvrBs4 is 97% identical to AvrBs3 (11), yet it does not trigger either the pepper *Bs3* or the *Bs3-E* gene because its distinct RVD composition does not support interaction with these promoters (3). AvrBs4 does, however, trigger an HR in the *Capsicum pubescens* accession PI 235047 that contains the matching *R* gene *Bs4* (12). AvrBs4 is also recognized in tomato plants that contain the corresponding nucleotide binding site leucine-rich repeat type R protein *Bs4* (13). Importantly, the tomato *R* gene *Bs4* is distinct from all other known plant *R* genes that mediate recognition of TALEs because it is not transcriptionally activated by matching TALEs (14). Furthermore, the tomato *Bs4* gene is currently the only TALE-specific *R* gene that mediates recognition of several TALEs with distinct RVD composition (14). For clarity, we refer in this article to the above-described *R* genes as *Bs4S* (*S. lycopersicum*) and *Bs4C* (*C. pubescens*).

Previously, it was shown that the TALE code could be used to predict DNA target sequences of *Xanthomonas* wild-type TALEs, as well as in vitro-assembled TALEs with arbitrary RVD composition (6–8). In the present study we intended to use the TALE code in a reverse manner to generate designer TALEs

Author contributions: R.M. and T.L. designed research; R.M. performed research; P.R. and J.B. contributed new reagents/analytic tools; and T.L. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

¹To whom correspondence should be addressed. E-mail: lahaye@bio.lmu.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1013133107/-DCSupplemental.

(dTALs) that would activate user-defined chromosomal target genes. We demonstrate that the scaffold of the TALE protein AvrBs3 can be designed with different consecutive RVDs to target desired DNA sequences. In addition, we show that variation of the repeat unit number facilitates fine-tuning of TALE target specificity. Finally, we show that the RVD NK facilitates specific targeting of G nucleotides, which thus far could not be addressed by a specific RVD.

Results

Codon-Optimized TALE Genes of AvrBs3 and AvrBs4 Specifically Trigger the Corresponding Capsicum R Genes Bs3 and Bs4C. The molecular dissection of TALE DNA-target specificity requires TALE genes that can be modified by PCR mutagenesis in a defined manner. However, because of the high overall GC content of *Xanthomonas* TALE genes and their nearly identical, tandemly arranged 102-bp repeat sequences, PCR mutagenesis of *Xanthomonas* TALE genes is challenging. For example, the TALE gene *avrBs3* has an overall GC content of 67% and is composed of 17.5 tandemly arranged 102-bp repeats that are >90% identical to each other at the nucleotide level (15). To overcome these limitations, we generated codon-optimized versions of the *Xanthomonas* TALE genes *avrBs3* and *avrBs4*, herein defined as *avrBs3** and *avrBs4**, respectively (SI Text). Both *avrBs3** and *avrBs4** are codon-optimized for expression in tobacco, have a GC content of only 50%, and show on average 72% homology between the 102-bp repeats. To functionally test codon-optimized TALE genes, we made use of the *Capsicum* R genes *Bs3* and *Bs4C* that mediate specific recognition of AvrBs3 and AvrBs4, respectively (3, 12). To do so, we cloned *avrBs3** and *avrBs4** into plant-expression vectors under transcriptional control of the constitutive cauliflower mosaic virus 35S (*35S*) promoter. Subsequently we used *Agrobacterium tumefaciens* T-DNA transfer to deliver the codon-optimized TALE genes in *Capsicum* plants containing the *Bs3* or *Bs4C* genes.

As shown in Fig. S1, *in planta* expression of *avrBs3** triggered an HR only in genotypes that contain the *Bs3* gene. Reciprocally, *in planta* expression of *avrBs4** triggered an HR only in genotypes that contain the *Bs4C* gene. Thus, codon-optimized *avrBs3** and *avrBs4** genes showed *in planta* the expected specificity and triggered HR only in plants containing the corresponding plant R gene.

Repeat-Polymorphisms in Non-RVD Residues Have No Major Impact on RVD Specificity. Previous studies of *Xanthomonas* TALEs and matching UPT boxes suggested that RVD composition determines recognition specificity (6–8, 16). However, repeat units differ not only in RVD but also in non-RVD residues and circumstantial evidence suggests that non-RVDs that are polymorphic between repeat units might have coevolved with defined RVD motifs. For example, in the *Xcv* TALE AvrBs3, all seven repeat units with the RVD motif HD contain an arginine at position 24, but the three repeat units with the RVD motif NI contain an alanine at position 24 (SI Text). To clarify if distinct RVD motifs function only in the context of matching non-RVD residues, we generated chimeric proteins consisting of the *Xcv* TALEs AvrBs3 and AvrBs4 that both contain 17.5 repeat units (11, 15). AvrBs3 and AvrBs4 differ in 38 repeat-unit residues and 18 of those differences are located within the RVDs (SI Text). To test if only the RVDs but not other repeat unit residues define DNA binding-specificity, we replaced all RVDs in AvrBs3 by the RVDs from AvrBs4, thereby generating AvrBs3RVD4*. Reciprocally we replaced all RVDs in AvrBs4 by the RVDs from AvrBs3, thereby generating AvrBs4RVD3* (Fig. S2A). Both, *avrBs3RVD4** and *avrBs4RVD3** were cloned in T-DNA vectors under transcriptional control of the *35S* promoter and were transiently expressed in pepper plants containing the *Bs3* or the *Bs4C* gene. AvrBs3* and AvrBs4RVD3* triggered an HR in *Bs3* but not in *Bs4C* genotypes (Fig. S3). Reciprocally, AvrBs4* and AvrBs3RVD4* triggered an HR in *Bs4C*- but not in *Bs3*-

containing plants. In summary, these data indicate that only the RVDs of AvrBs3 and AvrBs4 dictate recognition specificity in their interaction with the pepper R genes *Bs3* and *Bs4C*, respectively, but non-RVDs have no or only a minor impact on TALE specificity.

To compare the biological activity of wild-type and chimeric TALEs in a more quantitative fashion, we generated *Bs3*-promoter derivatives that contain either the UPT_{AvrBs3} box ($Bs3pUPT_{AvrBs3}$) or the UPT_{AvrBs4} box ($Bs3pUPT_{AvrBs4} \Delta UPT_{AvrBs3}$), and cloned these promoters in front of an *uidA* reporter gene (Fig. S2B and SI Text). The UPT_{AvrBs4} box was deduced using the TALE code and was cloned into a *Bs3* promoter in which the UPT_{AvrBs3} box had been deleted. The *promoter:uidA* reporter constructs were delivered into *Nicotiana benthamiana* leaves in combination with the *35S* promoter-driven TALE genes *avrBs3**, *avrBs3RVD4**, *avrBs4**, and *avrBs4RVD3**, respectively. GUS (β -glucuronidase) assays showed that the UPT_{AvrBs3} box-containing *Bs3* promoter is transcriptionally activated only by AvrBs3* and AvrBs4RVD3*, whereas the UPT_{AvrBs4} box containing *Bs3*-promoter derivative is transcriptionally activated only by AvrBs4* and AvrBs3RVD4*, respectively (Fig. S2C). The chimeras showed in both cases a specificity that was equal to the *Xanthomonas* wild-type proteins with identical RVD composition. This finding indicates that, as long as the general structure of the 34-aa TALE repeat is used, DNA specificity is predominantly controlled by RVD rather than non-RVD residues. Quantitative GUS assays showed that the chimeric AvrBs4RVD3* and AvrBs3RVD4* proteins produced slightly lower GUS activity than the corresponding AvrBs3* and AvrBs4* proteins, respectively (Fig. S2D). These data suggest that certain non-RVDs may have coevolved with defined RVD residues and may contribute a minor role in DNA binding. However, overall we conclude that non-RVD residues have insufficient impact on the DNA-binding specificity of TALEs to warrant factoring them into the design of de novo-engineered TALEs.

De Novo-Engineered dTALEs Specifically Activate Promoters with Matching UPT Boxes. We were able to demonstrate that codon-optimized TALE genes have the expected biological function (Fig. S1) and that repeat-polymorphic non-RVDs have no significant impact on TALE DNA-binding specificity (Figs. S2 and S3). These findings suggested that dTALEs with desired specificity can be generated simply by modulation of TALE RVDs. As our first dTALE target promoter, we selected the well-studied tomato *Bs4S* promoter ($Bs4Sp$), which is particularly well-suited for promoter-GUS studies because of its low basal activity (7). Thus far, no *Xanthomonas* TALE is known that transcriptionally activates *Bs4Sp*. A dTALE, designated dTALE[*Bs4S*] (dTALE inducing the *Bs4S* promoter), was designed against a 19-bp sequence (TATA-TAACTTTGTCCAAAA; $UPT_{dTALe[Bs4S]}$ box) located 46-bp upstream of the ATG start codon of the tomato *Bs4S* coding sequence (13) (SI Text). The TALE code was used to translate the nucleotide sequence of the $UPT_{dTALe[Bs4S]}$ box into corresponding RVDs, and a dTALE comprised of the structural scaffold of AvrBs3 and the desired RVDs was engineered (SI Text). The function of dTALE[*Bs4S*] was tested by *A. tumefaciens*-mediated codelivery of $35S:dTALE[Bs4S]$ and a *Bs4S* promoter *uidA* reporter construct ($Bs4Sp:uidA$) into *N. benthamiana* leaves. As hypothesized, the dTALE[*Bs4S*] but not the related AvrBs3* protein produced blue staining in combination with the $Bs4Sp:uidA$, demonstrating the selective activation of the *Bs4S* promoter by dTALE[*Bs4S*] (Fig. 1A). To further test the specificity of dTALE[*Bs4S*] for the $UPT_{TALE[Bs4S]}$ box of the *Bs4S* promoter, we generated two promoter-deletion derivatives that lack either nucleotides 6 to 8 ($Bs4Sp\Delta 6-8$) or nucleotides 10 to 12 ($Bs4Sp\Delta 10-12$) of the $UPT_{TALE[Bs4S]}$ box and cloned these in front of an *uidA* reporter gene (Fig. S4). *A. tumefaciens*-mediated delivery of $35S:dTALE[Bs4S]$ with either $Bs4Sp\Delta 6-8:uidA$ or $Bs4Sp\Delta 10-12:uidA$ failed to induce the expression of the GUS reporter, indicating that dTALE[*Bs4S*] does indeed target the selected 19-bp motif present in the

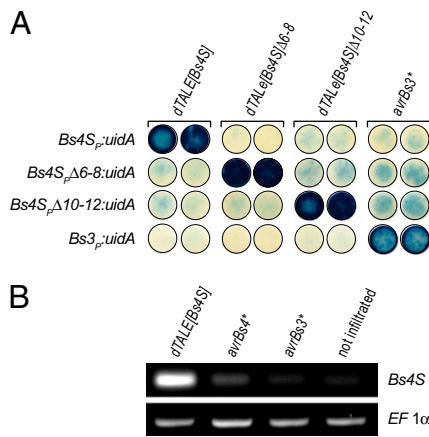


Fig. 1. De novo-engineered dTALEs specifically activate promoters with matching UPT boxes. (A) *In planta* functional analysis of different TALE-promoter combinations. The *uidA* reporter constructs under transcriptional control of the promoters indicated at left were codelivered via *A. tumefaciens* into *N. benthamiana* leaves with the 35S promoter-driven TALE genes indicated above leaf discs. GUS assays were carried out 40 hpi. Leaf discs were stained with 5-bromo-4-chloro-3-indolyl- β -D-glucuronic acid, cyclohexylammonium salt (X-Gluc) to visualize activity of the GUS reporter. Asterisks (*) indicate TALE genes with optimized codon usage and corresponding gene products. (B) The dTALE[*Bs4S*] transcriptionally activates the endogenous *Bs4S* gene in tomato. Semiquantitative RT-PCR was carried out on RNA from tomato leaf tissue 36 h after *A. tumefaciens*-mediated delivery of the depicted 35S promoter-driven TALE genes. Isolated RNA was used for cDNA synthesis and semiquantitative RT-PCR was performed with primers specific for the tomato *Bs4S* gene. The constitutively expressed gene elongation factor 1 α (*EF1 α*) served as a normalization control.

Bs4S promoter (Fig. 1A). In a reciprocal experiment, we also generated deletion derivatives of the dTALE[*Bs4S*] that lack repeat units 6 to 8 (dTALE[*Bs4S*] Δ 6–8) or repeat units 10 to 12 (dTALE[*Bs4S*] Δ 10–12). These units should be compatible with the *Bs4S* promoter-deletion derivatives *Bs4S_P Δ 6–8* and *Bs4S_P Δ 10–12*, respectively (Fig. S4). Indeed, *A. tumefaciens*-mediated expression of dTALE[*Bs4S*] Δ 6–8 and dTALE[*Bs4S*] Δ 10–12 induced the GUS reporter only with the matching *Bs4S_P Δ 6–8* and *Bs4S_P Δ 10–12* promoter GUS constructs, respectively (Fig. 1A). In summary, these data demonstrate the high specificity of dTALE[*Bs4S*] and its derivatives for the selected target DNA sequence in the *Bs4S* promoter and the relative ease of designing highly sequence-specific transcriptional activators based on the TALE DNA-binding scaffold.

To demonstrate that the dTALE[*Bs4S*] is also capable of transcriptionally activating the *Bs4S* endogene in tomato, 35S promoter-driven TALE genes dTALE[*Bs4S*], *avrBs4**, and *avrBs3** were delivered into tomato leaves via *A. tumefaciens*-mediated T-DNA transfer. Subsequent semiquantitative RT-PCR analysis demonstrated the transcriptional activation of the *Bs4S* endogene by the dTALE[*Bs4S*] in tomato leaves (Fig. 1B). In contrast, *avrBs4** and *avrBs3** did not transcriptionally activate the *Bs4S* endogene, consistent with previous studies done with *AvrBs3* and *AvrBs4* (14). These data are unique in demonstrating that a dTALE can be designed and delivered to effectively target a user-defined chromosomal gene.

Designer TALEs Specifically Activate Promoters of Targeted *Arabidopsis thaliana* Endogenes. To further support our hypothesis that the TALE code can be used to create dTALEs with desired target specificity, we designed two additional dTALEs targeted to the *A. thaliana* genes *EGL3* (17) and *KNAT1* (18). We choose these as targets because previous studies had shown that constitutive overexpression of *EGL3* or *KNAT1* results in an increase of trichomes (19) or dramatically altered leaf morphology

(18), respectively. Designer TALE[*EGL3*] and dTALE[*KNAT1*] were raised against distinct 19-bp sequences present in each of the *EGL3* and *KNAT1* promoters, respectively (SI Text). The nucleotide sequences of the *UPT*_{dTALE[*EGL3*]} (TATATACAGTGTACACA-CA) and *UPT*_{dTALE[*KNAT1*]} (TATATACCTAGTTCGTTTT) box were translated into corresponding RVDs and incorporated into the structural scaffold of *AvrBs3* (SI Text).

To test functionality of these two dTALEs, 35S promoter-driven T-DNA constructs with C-terminal *GFP* tags were generated and tested via transient promoter-GUS reporter assays in *N. benthamiana* leaves. To avoid problems in the GUS assay because of potentially leaky *A. thaliana* *EGL3* and *KNAT1* promoters, we introduced the two *UPT* boxes into the *Bs4S* promoter yielding the *Bs4S*-promoter derivatives *UPT*_{dTALE[*EGL3*]}*inBs4S_P* and *UPT*_{dTALE[*KNAT1*]}*inBs4S_P* and *UPT*_{dTALE[*EGL3*]}*inBs4S_P::uidA* and *UPT*_{dTALE[*KNAT1*]}*inBs4S_P::uidA*, respectively (Fig. S5). These *Bs4S* promoter derivatives were then codelivered with the 35S promoter-driven dTALE[*EGL3*] gene by *A. tumefaciens* into *N. benthamiana* leaves. GUS staining was apparent with the matching *UPT*_{dTALE[*EGL3*]}*inBs4S_P* but not with the *UPT*_{dTALE[*KNAT1*]}*inBs4S_P* promoter construct (Fig. S5C). Reciprocally, dTALE[*KNAT1*] produced GUS activity only in combination with the promoter containing the matching *UPT*_{dTALE[*KNAT1*]} box and not with the promoter that contains the *UPT*_{dTALE[*EGL3*]} box. These data show that each of the two dTALEs specifically activated the promoter containing the corresponding UPT box.

Next, dTALE[*EGL3*] and dTALE[*KNAT1*] were tested to see if they can function as transcriptional activators of *A. thaliana* endogenes. We generated stable transgenic lines of the *A. thaliana* ecotype Col-0 by introducing the 35S promoter driven dTALE[*EGL3*] and dTALE[*KNAT1*] genes by floral dipping (20). Putative transformants were selected on kanamycin-supplemented medium and microscopically screened for expression of GFP. Analysis of leaf tissue of the dTALE-GFP-expressing plants by semiquantitative RT-PCR showed that transgenic lines expressing dTALE[*EGL3*] or dTALE[*KNAT1*] have in comparison with a nontransgenic *A. thaliana* plant increased transcript levels of *EGL3* and *KNAT1*, respectively (Fig. 2). However, the expected morphological changes were not detected, which might be because of counterselection against the potentially detrimental phenotypes in the recovered transgenic lines or a posttranscriptional regulatory process.

Using the Patmatch algorithm (<http://is.gd/g31n9>), we scanned the *Arabidopsis* genome for sequences with homology to the boxes in the *EGL3* and *KNAT1* promoters that were targeted by our dTALEs. We identified no perfect match in the *Arabidopsis* genome but three sequences, AT4G27740, AT3G15640, and AT3G59220, and one sequence, AT1G77980, containing two mismatches from the given *EGL3* and *KNAT1* target sequences, respectively (Fig. S5D). RT-PCR analysis demonstrated that AT3G59220 was the only off-target site with increased transcript levels in the transgenic lines (Fig. S5D). These data imply that

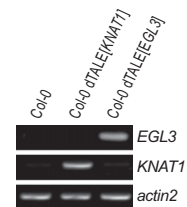


Fig. 2. The dTALE[*EGL3*] and dTALE[*KNAT1*] transcriptionally activate the *EGL3* and *KNAT1* endogenes in *A. thaliana*, respectively. Semiquantitative RT-PCR was carried out on RNA from the *Arabidopsis* ecotype Columbia (Col-0) and corresponding lines containing 35S promoter-driven dTALE[*EGL3*] or dTALE[*KNAT1*] transgenes. RNA was isolated from leaves and used for cDNA synthesis. Semiquantitative RT-PCR was performed with *EGL3* and *KNAT1* specific primers. The constitutively expressed gene *actin2* served as an internal normalization control.

some mismatches strongly impair DNA recognition, whereas others seem to have no obvious effect and are in accordance with previous findings (10).

In summary, these data demonstrate that custom dTALEs reliably activate user-defined target genes. However, these data also demonstrate that potential off-sites need to be routinely analyzed, as with other DNA binding scaffolds.

De Novo-Engineered AvrBs3 Derivatives with Four Additional Repeat Units Specifically Activate Either the *Bs3* or the *UPA20* Promoter. A central question in the design of dTALEs is the correlation between repeat number and DNA-binding specificity. We have used the 17.5 repeat-unit TALE AvrBs3 as a structural scaffold for our dTALEs. This repeat-unit number is most predominant among naturally occurring *Xanthomonas* strains, although TALE proteins do exist with a range of repeat-unit numbers, including some with more than 30 repeat units (4). We wished to examine how the number of repeat units in a given TALE correlates with its degree of target specificity.

To address this question, we made use of the pepper *Bs3* and *UPA20* promoters, two well-studied AvrBs3 target promoters (2, 3, 10, 21) that are almost identical within their UPT_{AvrBs3} boxes but that differ significantly up- and downstream of this sequence motif (Fig. S6). Based on the assumption that the addition of repeat units might alter the DNA-binding specificity of AvrBs3*, we extended AvrBs3* by four C-terminal repeat units with RVD motifs matching to either the *Bs3* promoter (target box: $UPT_{AvrBs3}Bs3_P$) or the *UPA20* promoter (target box: $UPT_{AvrBs3}UPA20_P$) resulting in the AvrBs3* derivatives AvrBs3+4R_{Bs3}* and AvrBs3+4R_{UPA20}*, respectively (Fig. S6). To study targeting of the $UPT_{AvrBs3}Bs3_P$ and the $UPT_{AvrBs3}UPA20_P$ boxes in the same promoter context, we introduced both UPT boxes into the *Bs3* promoter and fused these promoter constructs in front of the *uidA* reporter gene. *A. tumefaciens*-mediated delivery of the 35S promoter-driven *avrBs3+4R_{Bs3}** TALE gene in *N. benthamiana* leaves produced GUS activity only in combination with the promoter containing the matching $UPT_{AvrBs3}Bs3_P$ box but not with the promoter containing the $UPT_{AvrBs3}UPA20_P$ box (Fig. S6D). Reciprocally, the *avrBs3+4R_{UPA20}** TALE gene induced GUS expression only in combination with the promoter containing the matching $UPT_{AvrBs3}UPA20_P$ box and not with the promoter containing the $UPT_{AvrBs3}Bs3_P$ box (Fig. S6D). AvrBs3* induced both promoters containing either the $UPT_{AvrBs3}Bs3_P$ or the $UPT_{AvrBs3}UPA20_P$ box. To further test specificity of AvrBs3* and its derivatives, we also analyzed the rice *Xa27* promoter, which was previously shown to be transcriptionally activated by the matching *Xanthomonas oryzae* pv. *oryzae* TALE AvrXa27 (22) but not by the *Xcv* TALE AvrBs3 (23). *A. tumefaciens*-mediated delivery of the 35S promoter-driven dTALE genes *avrBs3+4R_{Bs3}** or *avrBs3+4R_{UPA20}** showed that neither of those produced GUS activity in combination with the rice *Xa27* promoter *uidA* reporter construct (Fig. S6D).

To analyze the specificity of the elongated TALEs on the *UPA20* and *Bs3* endogenes, we expressed the corresponding TALE genes via *A. tumefaciens*-mediated delivery in pepper. Semiquantitative RT-PCR demonstrated that expression of AvrBs3+4R_{Bs3}* resulted in increased *Bs3* but not *UPA20* transcript levels (Fig. 3). Reciprocally, expression of AvrBs3+4R_{UPA20}* resulted in increased *UPA20* but not *Bs3* transcript levels. In contrast, expression of AvrBs3 resulted in increased levels of both *Bs3* and *UPA20* (Fig. 3).

In summary, our data show that the addition of repeat units to the TALE protein AvrBs3 alters its target specificity.

RVD NK Mediates Specific Targeting of G Nucleotides. Experimental analysis of TALEs and matching UPT boxes uncovered RVDs that target A, C, or T nucleotides, specifically (7). The RVD type NN binds to G and A nucleotides, but no RVD type was shown to target exclusively G nucleotides. Bioinformatic analysis of *Xanthomonas* TALEs and potential UPT boxes in the rice genome

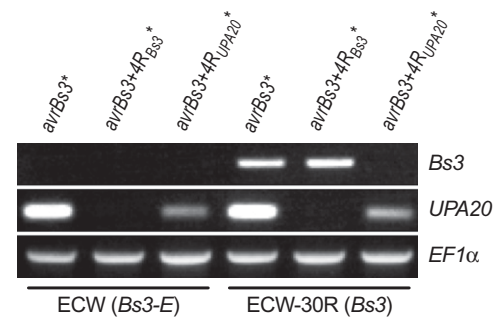


Fig. 3. AvrBs3 derivatives with four additional C-terminal repeat units discriminate between the pepper *Bs3* and *UPA20* promoter. The depicted 35S promoter-driven TALE genes were delivered via *A. tumefaciens* into the pepper genotypes ECW or ECW-30R. Semiquantitative reverse-transcription RT-PCR was carried out on RNA isolated from leaf material 36 h upon infection with primers specific for the *Bs3* or *UPA20* gene. The constitutively expressed gene *EF1α* served as an internal normalization control.

uncovered two interaction sites in which an NK type RVD aligned to a G nucleotide (6). To test if NK would bind G nucleotides in a specific manner, we made use of three tandem-arranged NI-type RVDs in the TALE AvrBs3 (repeat units 5–7) (Fig. S7 and *SI Text*) and the corresponding tandem-arranged A nucleotides in the UPT_{AvrBs3} box of the pepper *Bs3* promoter. We generated *Bs3*-promoter derivatives in which the three A nucleotides were replaced by three C, G, or T nucleotides and fused these *Bs3*-promoter derivatives in front of the *uidA* reporter gene, generating $UPT_{AvrBs3}Bs3_{P:5-7C:uidA}$, $UPT_{AvrBs3}Bs3_{P:5-7G:uidA}$, and $UPT_{AvrBs3}Bs3_{P:5-7T:uidA}$, respectively (Fig. S7). Furthermore, we constructed an AvrBs3* derivative in which the three subsequent NI type RVDs were replaced by NK type RVDs, generating AvrBs3 5–7 NK*. Target specificity of AvrBs3 5–7 NK* was studied by *A. tumefaciens*-mediated delivery of the 35S promoter-driven *avrBs3 5–7 NK** gene in combination with the different *Bs3*-promoter derivatives. We found that AvrBs3 5–7 NK* induced GUS expression only in combination with the *Bs3*-promoter derivative $UPT_{AvrBs3}Bs3_{P:5-7G:uidA}$ that contains the three tandem-arranged G nucleotides (Fig. S8). In contrast, AvrBs3*, which contains three tandem-arranged NI-type RVDs, produced blue staining only in combination with the *Bs3* wild-type promoter ($Bs3_P:uidA$) that contains three corresponding tandem-arranged A nucleotides. In summary, our findings show that the NK-type RVD facilitates specific targeting of G nucleotides.

Discussion

RVD Composition of the Repeat Unit Array and dTALE Target Specificity.

In this article, we demonstrate that the TALE scaffold can be used for transcriptional activation of user-defined chromosomal loci. Target specificity is a central goal in the design and use of dTALEs, raising the question as to whether the AvrBs3 scaffold can facilitate targeting of single-copy sequences in complex genomes. Extensive screening has uncovered in total 18 AvrBs3-induced genes in the pepper genome (2, 3, 9, 21), which has an estimated genome size of 3×10^9 bp (24) and which is similar in size to the human genome (25). These findings may suggest that the AvrBs3 scaffold does not facilitate targeting of single-copy sequences in the human genome or genomes of comparable size. However, it needs to be noted that AvrBs3 contains three NS-type RVDs, which have been shown to target A, C, G, and T nucleotides with almost identical affinity (7). Hence, sequence-defined AvrBs3 scaffold-based dTALEs should have an ~64-fold (4^3) higher target specificity than the *Xanthomonas* AvrBs3 protein if RVDs with ambiguous target specificity are replaced with RVDs with tight sequence specificity. Our finding, that the RVD NK specifically mediates targeting of G nucleotides (Fig. S8), is thus an important contribution for construction of dTALEs

with increased sequence specificity, especially in the context of G-rich sequences.

Repeat-Unit Number and TALE Target Specificity. In addition to RVD composition, we speculated that another important feature of dTALE target specificity is the number of repeat units in a given TALE or dTALE. This postulate is supported by our analysis of two distinct AvrBs3 derivatives with four additional repeat units that mediate specific targeting of the *Bs3* and *UPA20* promoters, although the wild-type AvrBs3 protein activates both promoters (Fig. 3). These data suggest that dTALEs with a high number of repeat units generally have superior target specificity, compared with dTALEs with fewer repeat units.

The modified target specificity of AvrBs3 derivatives with additional repeat units also raises the question of the molecular basis of this modified target specificity. Previous studies had shown that AvrBs3 physically interacts with and transcriptionally activates the pepper *Bs3* and *UPA20* promoters (2, 3). However, the two elongated TALEs that contain the complete repeat unit array of AvrBs3 specifically activate the *Bs3* or the *UPA20* promoters (Fig. 3), suggesting that these AvrBs3 derivatives bind specifically to either the *Bs3* or the *UPA20* promoter. We envision two molecular scenarios of why those AvrBs3-derivatives have a modified-promoter specificity compared with their progenitor AvrBs3. One possibility is that additional, nonmatching repeat units at the C terminus cause these terminal repeat units to “loop out,” which may hinder the C-terminal TALE activation domain to get in physical proximity and activate the given promoter. Alternatively, we envision that the TALE scaffold is rigid and, hence, the C-terminal, nonmatching RVDs would destabilize the structure and prevent the entire TALE from binding to a given target sequence. Further studies will have to clarify which model adequately explains our experimental findings.

TALE Repeat Units: A Promising Sequence-Specific Targeting Device for Nucleases and Other DNA-Modifying Enzymes. Previous studies of *Xanthomonas* TALEs have shown that TALE-mediated promoter activation strictly correlates with physical interaction of the given promoter target sequences (2, 3, 8, 10, 21). Hence, dTALE repeat-unit arrays are likely to function as a sequence-specific targeting modules not only in the context of a transcription factor, but also when fused to other functional domains as, for example, nucleases, methylases, recombinases, or transcription-repressor domains. Most recently, two reports have indeed demonstrated that translational fusions of TALE proteins to the nonsequence-specific FokI nuclease result in sequence-specific TALE nucleases (TALENs) (26, 27), suggesting that TALEs might pose an alternative to the widely used zinc-finger nucleases (1).

The report by Christian et al. (26) described also custom-designed TALE repeat-unit arrays fused to FokI, targeting 12- and 13-bp sequences from the *A. thaliana* *ADHI* or the zebrafish *gridlock* genes, respectively. The authors could show that homodimeric TALEs, which bind to two identical sequences from either the *ADHI* or the *gridlock* gene, placed on opposite sides of a central spacer function in a sequence-specific manner. Although none of the two recent reports on TALEs demonstrated cleavage of chromosomal loci by TALEs, it seems possible that this goal can be achieved in the near future.

TALE DNA Binding Domain: A Promising Alternative for Zinc-Finger Arrays? Our work on dTALEs and the most recent report on custom-designed TALENs (26) suggest that the TALE DNA-binding domain might pose an alternative to the currently used zinc-finger arrays. However, a major difference between tandem-arranged zinc fingers and TALE repeat units is that the latter modules seem to act in a context-independent manner (6). In consequence, tandem-arranged TALE repeat units have a highly predictable sequence specificity but zinc-finger arrays of desired specificity have to be identified from complex expression libraries. The laborious selection procedure for custom zinc-finger arrays and the more restrictive requirements for candidate-binding sites are limitations on the wide-spread use of this technology, which we do not expect to plague TALE scaffold proteins.

Materials and Methods

Generation of Codon-Optimized TALE Genes. The *avrBs3** gene, optimized for minimal sequence homology between the 102-bp repeats (for sequence, see *SI Text*), was synthesized by GenScript with a tobacco codon usage, and cloned into the MCS of pUC57 yielding pUC57 *avrBs3**. To obtain the GATEWAY entry clone pENTR-D-*avrBs3**, amplification was carried out with Phusion high-fidelity DNA polymerase (New England Biolabs), primers RM1 and RM2 (for primer sequences, see *Table S1*) and pUC57 *avrBs3** as template DNA. *SspI* and *FseI* restriction sites flanking the repeat unit array were introduced into pENTR-D-*avrBs3** via PCR using primers RM3 and RM4, yielding pENTR-D-*avrBs3*-L*. The repeat-arrays of *avrBs4**, *avrBs3RVD4**, and *avrBs4RVD3** were synthesized in the same way as *avrBs3** and cloned using *PvuII* and *FseI* into the *SspI* and *FseI* digested pENTR-D-*avrBs3*-L* yielding pENTR-D-*avrBs4**, pENTR-D-*avrBs3RVD4**, and pENTR-D-*avrBs4RVD3**, respectively. To create pENTR-D-TAL[*Bs4S*], pENTR-D-TAL[*EGL3*], and pENTR-D-TAL[*KNAT1*], each repeat-array was synthesized partially from repeat number 6 to 17.5 flanked by *XhoI* and *FseI* restriction sites. These fragments were cloned into pENTR-D-*avrBs3*-R6-XhoI*, which was generated from pENTR-D-*avrBs3** using primers RM4 and RM5. Repeat deletion derivatives of pENTR-D-TAL[*Bs4S*] were obtained using primers RM6 and RM7 or RM8 and RM9, resulting in pENTR-D-TAL[*Bs4S*] Δ 6–8 or pENTR-D-TAL[*Bs4S*] Δ 10–12, respectively. To generate pENTR-D-*avrBs3*+4R_{Bs3}* and pENTR-D-*avrBs3*+4R_{Upa20}*, an *XhoI* site was introduced into pENTR-D-*avrBs3** using primers RM10 and RM11, yielding pENTR-D-*avrBs3*-R17-XhoI*. Terminal *XhoI* sites were introduced into synthesized four repeat-unit fragments with primers RM12 and RM13 or RM14 and RM15 and cloned via *XhoI* into pENTR-D-*avrBs3*-R17-XhoI*, yielding pENTR-D-*avrBs3*+4R_{Bs3}* or pENTR-D-*avrBs3*+4R_{Upa20}*, respectively. Mutations were introduced with the Phusion site directed mutagenesis kit (New England Biolabs). All entry clones were transferred by LR recombination (Invitrogen) into the expression vector pGWB5 (28) and transformed into *A. tumefaciens* GV3101 (29) for *in planta* analysis. Transgenic *A. thaliana* for dTALE[*EGL3*] and dTALE[*KNAT1*] were obtained as described previously (20).

Generation and Analysis of Promoter Mutants. Sequences of all promoter constructs are provided in the *SI Text*. Insertion of different *UPT* boxes was done by Phusion site directed mutagenesis (New England Biolabs). Promoter constructs were transferred by LR recombination into the binary vector pGWB3 (28) and transformed into *A. tumefaciens* GV3101 (29) for *in planta* analysis. Analysis of promoters via semiquantitative RT-PCR and GUS measurements were carried out as described earlier (3, 10).

ACKNOWLEDGMENTS. We thank D. Horvath, M. Bhakta, O. de Lange, and A. Strauß for helpful comments on the manuscript; T. Strauß, J. Elsaesser, B. Rosinsky, and M. Schulze for technical support; and T. Nakagawa (Shimane University, Matsue, Japan) for providing the pGWB vector series. Work in our laboratory has been supported by grants from the 2Blades Foundation, the Exzellenznetzwerk Biowissenschaften (Ministry of Culture of Saxonia-Anhalt), and the Deutsche Forschungsgemeinschaft (SPP1212, SFB 648, and LA1338/2-2).

1. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD (2010) Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* 11:636–646.
2. Kay S, Hahn S, Marois E, Hause G, Bonas U (2007) A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science* 318:648–651.
3. Römer P, et al. (2007) Plant pathogen recognition mediated by promoter activation of the pepper *Bs3* resistance gene. *Science* 318:645–648.
4. Boch J, Bonas U (2010) *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol* 48:419–436.

5. Bogdanove AJ, Schornack S, Lahaye T (2010) TAL effectors: Finding plant genes for disease and defense. *Curr Opin Plant Biol* 13:394–401.
6. Moscou MJ, Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. *Science* 326:1501.
7. Boch J, et al. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326:1509–1512.
8. Römer P, et al. (2010) Promoter elements of rice susceptibility genes are bound and activated by specific TAL effectors from the bacterial blight pathogen, *Xanthomonas oryzae* pv. *oryzae*. *New Phytol* 187:1048–1057.

9. Marois E, Van den Ackerveken G, Bonas U (2002) The *Xanthomonas* type III effector protein AvrBs3 modulates plant gene expression and induces cell hypertrophy in the susceptible host. *Mol Plant Microbe Interact* 15:637–646.
10. Römer P, et al. (2009) Recognition of AvrBs3-like proteins is mediated by specific binding to promoters of matching pepper *Bs3* alleles. *Plant Physiol* 150:1697–1712.
11. Bonas U, Conrads-Strauch J, Balbo I (1993) Resistance in tomato to *Xanthomonas campestris* pv. *vesicatoria* is determined by alleles of the pepper-specific avirulence gene *avrBs3*. *Mol Gen Genet* 238:261–269.
12. Stall RE, Jones JB, Minsavage GV (2009) Durability of resistance in tomato and pepper to xanthomonads causing bacterial spot. *Annu Rev Phytopathol* 47:265–284.
13. Schornack S, et al. (2004) The tomato resistance protein Bs4 is a predicted non-nuclear TIR-NB-LRR protein that mediates defense responses to severely truncated derivatives of AvrBs4 and overexpressed AvrBs3. *Plant J* 37:46–60.
14. Schornack S, Peter K, Bonas U, Lahaye T (2005) Expression levels of *avrBs3*-like genes affect recognition specificity in tomato *Bs4*- but not in pepper *Bs3*-mediated perception. *Mol Plant Microbe Interact* 18:1215–1225.
15. Bonas U, Stall RE, Staskawicz B (1989) Genetic and structural characterization of the avirulence gene *avrBs3* from *Xanthomonas campestris* pv. *vesicatoria*. *Mol Gen Genet* 218:127–136.
16. Schornack S, Minsavage GV, Stall RE, Jones JB, Lahaye T (2008) Characterization of AvrHah1, a novel AvrBs3-like effector from *Xanthomonas gardneri* with virulence and avirulence activity. *New Phytol* 179:546–556.
17. Koornneef M, Dellaert LW, van der Veen JH (1982) EMS- and radiation-induced mutation frequencies at individual loci in *Arabidopsis thaliana* (L.) Heynh. *Mutat Res* 93:109–123.
18. Lincoln C, Long J, Yamaguchi J, Serikawa K, Hake S (1994) A *knotted1*-like homeobox gene in *Arabidopsis* is expressed in the vegetative meristem and dramatically alters leaf morphology when overexpressed in transgenic plants. *Plant Cell* 6:1859–1876.
19. Zhang F, Gonzalez A, Zhao M, Payne CT, Lloyd A (2003) A network of redundant bHLH proteins functions in all TGT1-dependent pathways of *Arabidopsis*. *Development* 130:4859–4869.
20. Clough SJ, Bent AF (1998) Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16:735–743.
21. Kay S, Hahn S, Marois E, Wieduwild R, Bonas U (2009) Detailed analysis of the DNA recognition motifs of the *Xanthomonas* type III effectors AvrBs3 and AvrBs3 Δ rep16. *Plant J* 59:859–871.
22. Gu K, et al. (2005) *R* gene expression induced by a type-III effector triggers disease resistance in rice. *Nature* 435:1122–1125.
23. Römer P, Recht S, Lahaye T (2009) A single plant resistance gene promoter engineered to recognize multiple TAL effectors from disparate pathogens. *Proc Natl Acad Sci USA* 106:20526–20531.
24. Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* 9:208–219.
25. Consortium IHGS; International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431:931–945.
26. Christian M, et al. (2010) TAL effector nucleases create targeted DNA double-strand breaks. *Genetics* 186:757–761.
27. Li T, et al. (2010) TAL nucleases (TALNs): Hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res*, 10.1093/nar/gkq704.
28. Nakagawa T, et al. (2007) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J Biosci Bioeng* 104:34–41.
29. Koncz C, Schell J (1986) The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimeric genes carried by a novel type of *Agrobacterium* binary vector. *Mol Gen Genet* 204:383–396.

Supporting Information

Morbitzer et al. 10.1073/pnas.1013133107

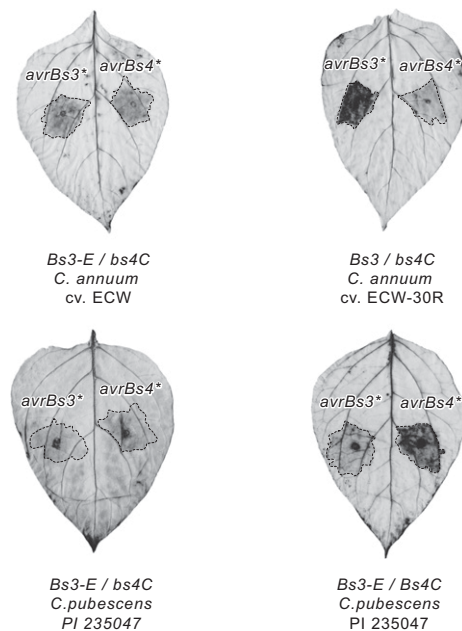


Fig. S1. Codon-optimized transcription activator-like effectors (TALE) genes of *avrBs3* and *avrBs4* trigger a hypersensitive response (HR) in *Bs3*- and *Bs4C*-resistant *Capsicum* species, respectively. *In planta* functional analysis of codon-optimized (*) TALE genes. The 35S promoter-driven TALE genes indicated on the left and right sides of leaves were delivered via *Agrobacterium tumefaciens* into the four *Capsicum* genotypes shown. The allelic configuration at the *Bs3* locus [alleles: *Bs3* (recognition of AvrBs3), *Bs3-E* (recognition of AvrBs3 Δ rep16)] and the *Bs4C* locus [alleles: *Bs4C* (recognition of AvrBs4), *bs4C* (no recognition of AvrBs3, AvrBs3 Δ rep16, and AvrBs4)] is indicated for each plant genotype. Dashed lines mark the inoculated areas. Two days after infiltration, the leaves were cleared in ethanol to visualize the hypersensitive response (dark areas).

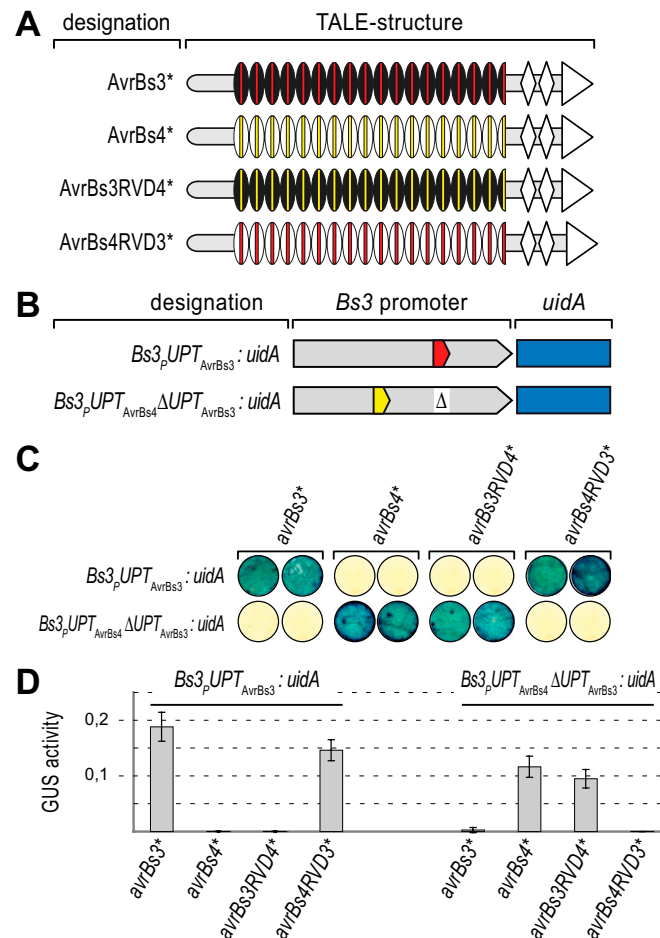


Fig. S2. TALE nonrepeat-variable diresidues (non-RVD) residues have no measurable influence on TALE specificity. (A) Graphic representation of chimeric TALEs and corresponding *uidA* reporter constructs. TALEs are displayed as gray bars. Black and white ovals represent the repeat units of AvrBs3 and AvrBs4, respectively. Red and yellow vertical lines within the ovals represent the RVDs of AvrBs3 and AvrBs4, respectively. The chimeric TALE AvrBs3RVD4 contains the RVDs of AvrBs4 but is otherwise identical to AvrBs3. Reciprocally, the chimeric TALE AvrBs4RVD3 contains the RVDs of AvrBs3 but is otherwise identical to AvrBs4. TALE nuclear localization signals (NLSs) and activation domain (AD) are displayed as white diamonds and white arrowheads, respectively. Asterisks (*) indicate that proteins are translated from genes with optimized codon usage. (B) Graphic representation of *Bs3* promoter derivatives. Gray arrows represent the promoter of the pepper *Bs3* gene ($Bs3_p$). Red and yellow arrowheads depict the UPT_{AvrBs3} and the UPT_{AvrBs4} boxes, respectively. The white area within the $Bs3_pUPT_{AvrBs4} \Delta UPT_{AvrBs3}$ promoter indicates the deleted UPT_{AvrBs3} box (Δ). Blue boxes represent the *uidA* reporter gene, encoding the β -glucuronidase (GUS) reporter protein. The nucleotide sequences of the promoters are provided as *SI Text*. (C) RVDs of AvrBs3 and AvrBs4 retain promoter-activation specificity regardless of the repeat-array backbone. Depicted *uidA* reporter constructs were delivered into *Nicotiana benthamiana* leaves in combination with the depicted *35S*-driven TALE genes. Asterisks (*) indicate TALE genes with optimized codon usage. Leaf discs were stained with 5-bromo-4-chloro-3-indolyl- β -D-glucuronic acid, cyclohexylammonium salt (X-Gluc) to visualize activity of the GUS reporter. Samples were taken at 40 hpi. (D) Quantitative analysis of promoter activation via wild-type and chimeric TALEs. GUS activity (pmol 4-MU-min⁻¹g protein⁻¹) was determined 27 h after *A. tumefaciens*-mediated codelivery of the depicted *uidA* reporter constructs and the *35S*-driven TALE gene shown below columns. Error bars denote SD. Asterisks (*) indicate TALE genes with optimized codon usage.

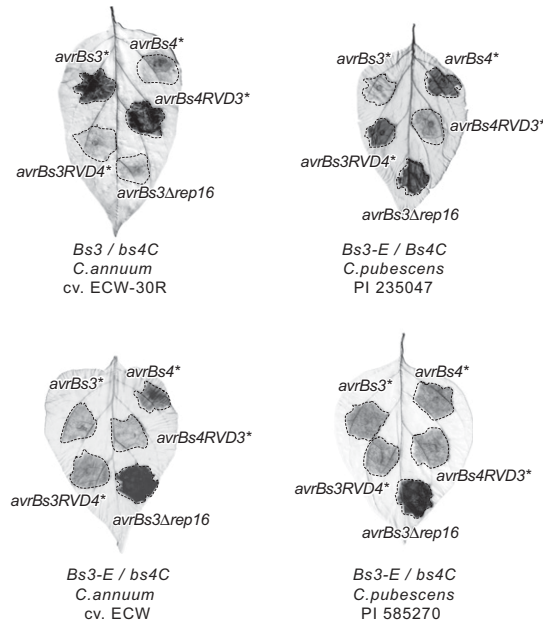


Fig. S3. *In planta* functional analysis of chimeric TALE genes. The 35S promoter-driven TALE genes indicated on the left and right side of the leaves were delivered via *A. tumefaciens* into leaves of the *Capiscum annuum* cultivars ECW and ECW-30R or the *Capiscum pubescens* accessions PI 585270 and PI 235047. The allelic configuration of the plant genotypes at the *Bs3* locus (alleles: *Bs3* or *Bs3-E*) and the *Bs4* locus (alleles: *Bs4C* or *bs4C*) locus is shown below the leaves. Dashed lines mark the inoculated areas. Two days after inoculation, the leaves were harvested and cleared with ethanol to visualize the HR (dark areas). Asterisks (*) indicate that AvrBs3 was translated from a codon-optimized gene.

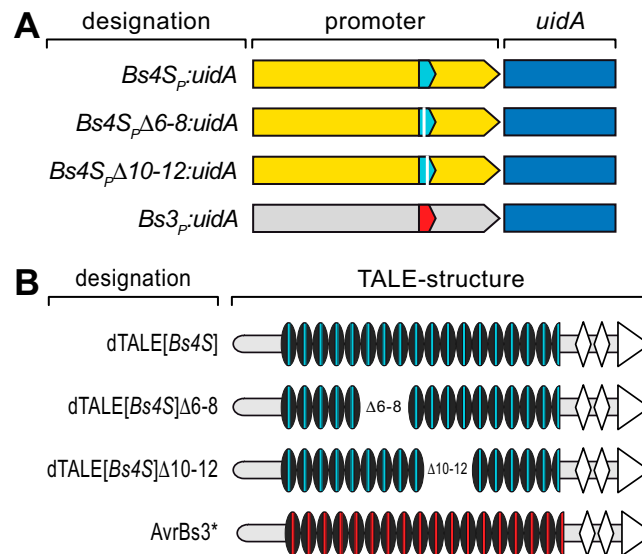


Fig. S4. Graphic display of *Bs4S* promoter:*uidA* reporter constructs and matching dTALEs. Corresponding experimental data (Fig. 1). (A) Arrows represent the promoters of the tomato *Bs4S* (yellow) and the pepper *Bs3* gene (gray). Cyan and red arrowheads depict the $UPT_{dTALE[Bs4S]}$ and UPT_{AvrBs3} boxes, respectively. White vertical bars within the cyan arrowheads represent deletions in the $UPT_{dTALE[Bs4S]}$ box. Blue boxes represent the *uidA* reporter gene, encoding the GUS protein. The nucleotide sequences of the promoters are provided as *SI Text*. (B) Graphic display of the TALEs used in this experiment. The TALEs are displayed as gray bars. Repeat units, NLSs, and AD are depicted as black ovals, white diamonds, and white arrowheads, respectively. Cyan and red colored vertical lines within the ovals represent the RVDs of designer TALEs (dTALE) [Bs4S] and AvrBs3*, respectively. Repeat unit deletions in dTALE[Bs4S] Δ 6-8 and dTALE[Bs4S] Δ 10-12 are indicated. Asterisks (*) indicate that the AvrBs3 is translated from a codon-optimized gene.

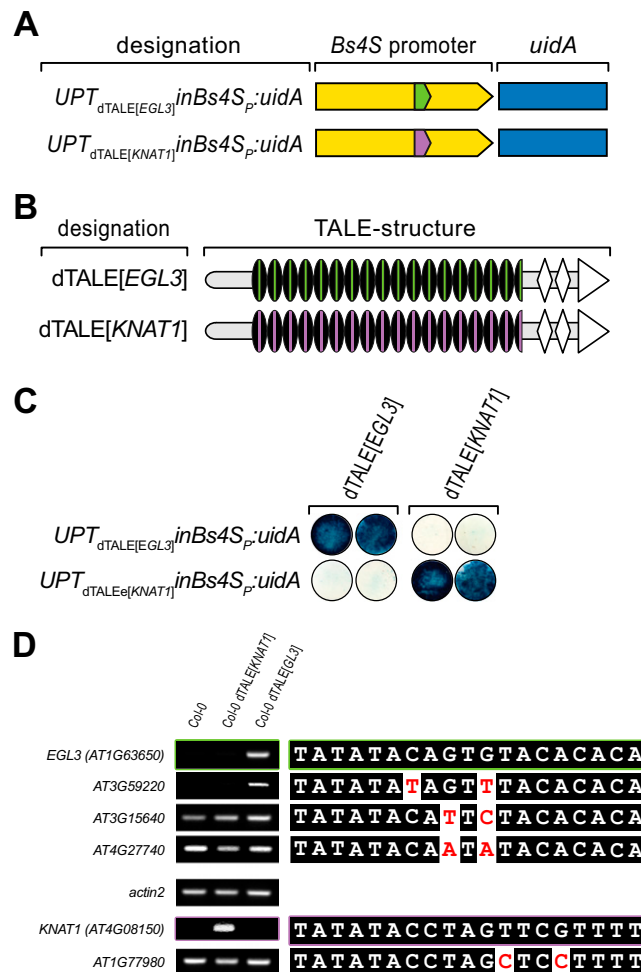


Fig. S5. Designer TALEs activate promoters of user-defined target genes in *Arabidopsis thaliana*. (A) Graphic display of promoter:*uidA* reporter constructs used in this experiment. Yellow arrows represent the promoter of the tomato *Bs4S* gene. Green and purple arrowheads depict the inserted $UPT_{dTAL[EGL3]}$ and $UPT_{dTAL[KNAT1]}$ boxes, respectively. Blue boxes represent the *uidA* reporter gene encoding the GUS protein. The nucleotide sequences of the promoters are provided as *SI Text*. (B) Graphic display of engineered dTALEs. The dTALEs are displayed as gray bars. Repeat units, NLSs, and AD are displayed as black ovals, white diamonds, and white arrowheads, respectively. Green and purple colored vertical lines within the ovals represent the RVDs of $dTAL[EGL3]$ and $dTAL[KNAT1]$, which direct binding to promoter elements of the *Arabidopsis EGL3* and *KNAT1* genes, respectively. Although not displayed graphically, both dTALEs contain a C-terminal GFP-tag that facilitates analysis of expression *in vivo*. (C) *In planta* functional analysis of different TALE-promoter combinations. The *uidA* reporter constructs under transcriptional control of the promoters indicated at left were delivered via *A. tumefaciens* into *N. benthamiana* leaves in combination with the 35S promoter-driven TALE genes shown above leaf discs. GUS assays were carried out 40 hpi as described in Fig. 1. (D) Analysis of potential $dTAL[EGL3]$ and $dTAL[KNAT1]$ off-targets in the *Arabidopsis* genome. Semiquantitative RT-PCR was carried out on RNA from the *Arabidopsis* ecotype Columbia (Col-0) and corresponding lines containing the 35S promoter-driven $dTAL[EGL3]$, or the $dTAL[KNAT1]$ transgene. RNA was isolated from leaves and used for cDNA synthesis. Semiquantitative RT-PCR was performed with locus-specific primers (*SI Text*). The constitutively expressed gene *actin2* served as an internal normalization control. The boxes targeted by $dTAL[EGL3]$ and $dTAL[KNAT1]$ are shown on the right side and are highlighted with green and purple frames, respectively. Sequences of potential off-targets are displayed below the two distinct, user-defined boxes. Differences between the targeted box and the potential off-targets are displayed in boldface red font. Details on the genomic context of the targeted boxes and potential off targets are provided in the *SI Text*.

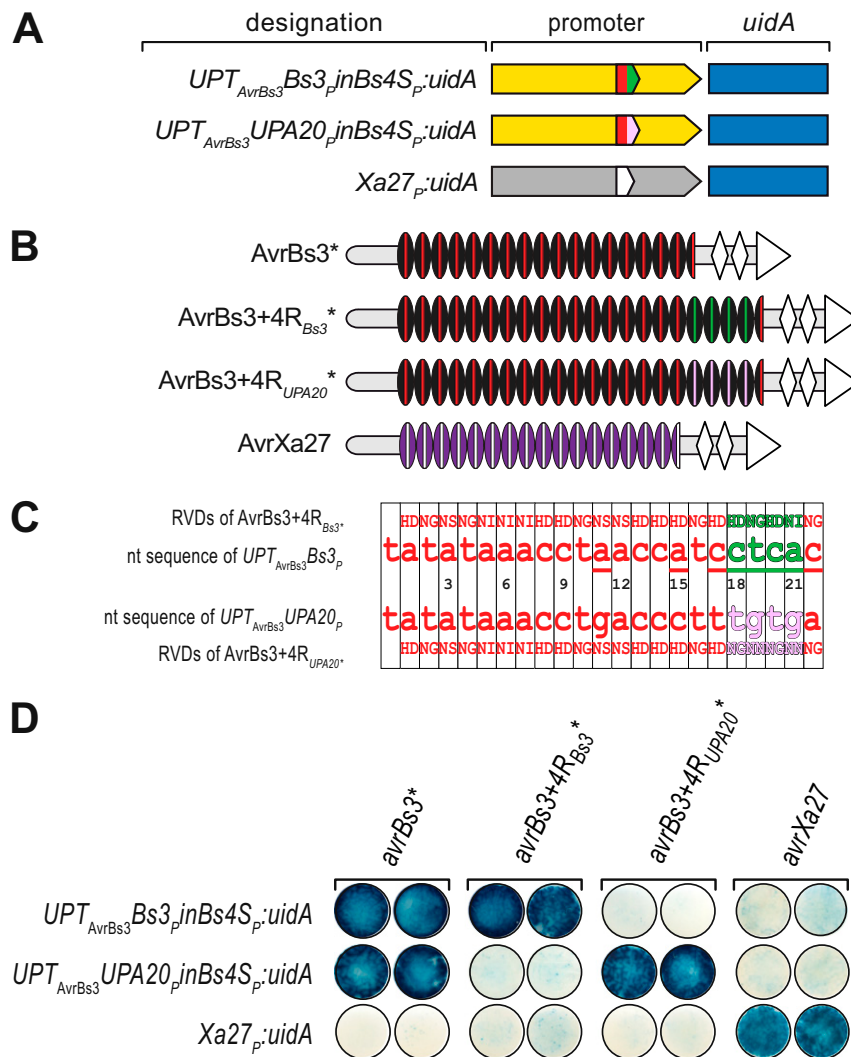


Fig. S6. Graphic display of AvrBs3 derivatives with additional repeat units and corresponding *uidA* reporter constructs. For corresponding experimental data, see also Fig. 3. (A) Arrows represent the promoters of the tomato *Bs4S* (yellow) and the rice *Xa27* gene (gray). Arrowheads depict *UPT* boxes. AvrBs3-targeted *UPT* boxes from the pepper *Bs3* and *UPA20* are displayed in red. *Bs3*- and *UPA20*-derived promoter regions 3' of their UPT_{AvrBs3} boxes are displayed in green and pink, respectively. The $UPT_{AvrXa27}$ box from the rice *Xa27* gene is displayed in white. A blue box represents the *uidA* reporter gene, encoding the GUS protein. The nucleotide sequences of the promoters are provided in the *SI Text*. (B) Graphic display of TALEs used in this experiment. The TALEs are displayed as gray bars. Repeat units, NLSs, and AD are displayed as ovals, white diamonds, and white arrowheads, respectively. Black- and purple-colored ovals represent the repeat units of AvrBs3 and AvrXa27, respectively. Red and white vertical lines represent RVDs of the TALEs AvrBs3 and AvrXa27, respectively. Green and pink vertical lines represent C-terminal RVDs that are specific to AvrBs3+4R_{Bs3}* and AvrBs3+4R_{UPA20}*, respectively. (C) Alignment of RVDs and *UPT*-boxes of the depicted TALEs and promoter-reporter constructs. RVDs are shown as uppercase letters using the single-letter code. Nucleotides are shown as lowercase letters. Horizontal boxes mark repeat units, corresponding RVDs, and aligned nucleotides. Numbers mark TALE repeat units and aligned nucleotides. Lowercase red letters represent nucleotides of the naturally occurring *UPT* boxes present in the pepper *Bs3* ($UPT_{AvrBs3}Bs3_p$) and the *UPA20* ($UPT_{AvrBs3}UPA20_p$) promoters. Lowercase green and pink letters represent nucleotides that are located 3' of the UPT_{AvrBs3} boxes in the *Bs3* and *UPA20* promoters, respectively. Nucleotides that differ between the *Bs3* and *UPA20* promoter are underlined. Uppercase red letters represent RVDs of the TALE AvrBs3. Green and pink big letters represent C-terminal RVDs that are specific to the dTALEs AvrBs3+4R_{Bs3}* and AvrBs3+4R_{UPA20}*, respectively and matching nucleotides of the *Bs3* and *UPA20* promoters. (D) AvrBs3 derivatives with additional repeat units discriminate between *UPA20* and *Bs3* promoter-derived *UPT* boxes with extensive 3' homology. The *uidA* reporter constructs under transcriptional control of the promoters shown at left were delivered via *A. tumefaciens* into *N. benthamiana* leaves in combination with the 35S promoter-driven TALE genes indicated above leaf discs. GUS assays were carried out 40 hpi, as described in Fig. 1. Asterisks (*) indicate that TALEs are translated from codon-optimized genes.

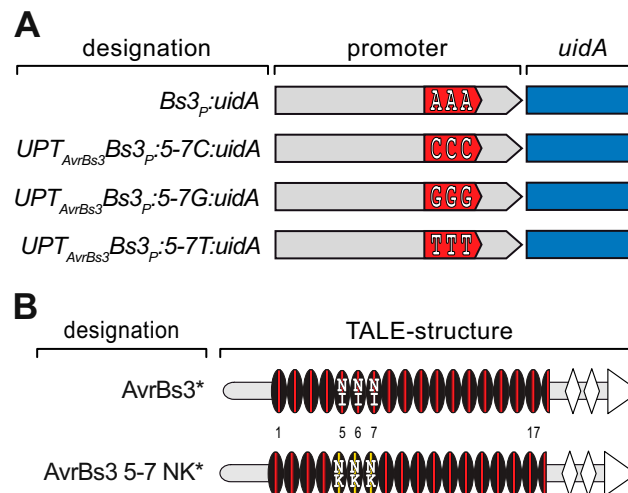


Fig. S7. Graphic display of AvrBs3 derivatives with three NK repeat units and corresponding *uidA* reporter constructs. For corresponding experimental data, see Fig. S8. (A) Gray arrows represent the pepper *Bs3* promoter. The red arrowhead with three tandem-arranged A nucleotides (AAA) depicts the wild-type *UPT*_{AvrBs3} box of the *Bs3* promoter. *Bs3*-promoter derivatives in which three tandem-arranged A nucleotides of the *UPT*_{AvrBs3} box are replaced by three C, G, or T nucleotides are depicted (CCC, GGG, and TTT). The nucleotide sequences of the promoters are provided in the *SI Text*. Blue boxes represent the *uidA* reporter gene encoding the GUS protein. (B) Graphic display of the TALEs used in this experiment. The TALEs are displayed as gray bars. Repeat units, NLSs, and AD are depicted as black ovals, white diamonds, and white arrowheads, respectively. Red vertical lines within black ovals represent the RVDs of AvrBs3. Three tandem-arranged NI type RVDs that are present in repeat-unit residues 5 to 7 of the wild-type AvrBs3 protein are depicted. Yellow vertical lines within black ovals represent RVD residues within the AvrBs3 scaffold that were changed to NK-type residues. Asterisks (*) indicate that TALEs are translated from codon-optimized genes.

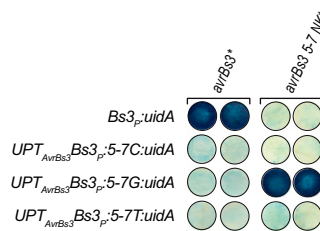


Fig. S8. The TALE RVD NK mediates specific targeting of G nucleotides. *In planta* functional analysis of different TALE-promoter combinations. The *uidA* reporter constructs under transcriptional control of the promoters indicated at left were codelivered via *A. tumefaciens* into *N. benthamiana* leaves with the 35S promoter-driven TALE genes indicated above leaf discs. GUS assays were carried out 40 hpi, as described in Fig. 1. Asterisks (*) indicate TALE genes with optimized codon usage and corresponding gene products.

Table S1. Nucleotide sequences of primers used in this study

Number	Designation	Nucleotide sequence
RM 1	AvrBs3*-CACC-F/RobM	CACCATGGATCCCATTAGGTCTAGG
RM 2	AvrBs3*-ns-R/RobM	CTGTGGAAGCAATTCCATAAGCC
RM 3	AvrBs3*-N-term-R/RobM	CCCGCATGCAATATTCAGCGGTGCCCCCGTCAAAA
RM 4	AvrBs3*-C-term-F/RobM	GGGAGTCATGGCCGGCCAGCACTTGAGTCGATTGTT
RM 5	AvrBs3*-R6-XhoI-R/RobM	CGGTCTCGAGAGCCTGTTTACTCCG
RM 6	TAL[Bs4S]-R5-R/RobM	GCCATGAGCCTGGCATAGTACAGGTAGAAGCGC
RM 7	TAL[Bs4S]-R9-F/RobM	CTTACACCAGAGCAAGTAGTTGCTATTGCATC
RM 8	TAL[Bs4S]-R9-R/RobM	TCCATGTGCCTGGCATAGA AACTGGCAGGAGC
RM 9	TAL[Bs4S]-R13-F/RobM	CTAACCCCGAACAGGTGGTGGCTATTGCATCCC
RM 10	AvrBs3*-R17-XhoI-F/RobM	ACTCGAGACAGTCCAGAGATTGCTGCCG
RM 11	AvrBs3*-R17-XhoI-R/RobM	CCCCTCGAGTGCCTGCTTCCACCATCATGCGAAGCGATAGCTACTACC
RM 12	4R _{Bs3} XhoI-F/RobM	CAAGCACTCGAGACCGTGCAGAGACTTCTG
RM 13	4R _{Bs3} XhoI-R/RobM	CGTCTCGAGAGCAGGTCTTCTCCACCG
RM 14	4R _{upa20} XhoI-F/RobM	GGCAAGCAGGCACCTCGAGACAGTTCAGAGACTG
RM 15	4R _{upa20} XhoI-R/RobM	CCCCTCGAGTGCCTGTTTGCCTCCGTTGTTGCTTGCATAGCC
RM 16	UPT _{TAL[Bs4S]} D6-8-R/RobM	GACAATATATACAAGAAAGAAGAAATTAACAAGTACC
RM 17	UPT _{TAL[Bs4S]} D10-12-R/RobM	GAAGTTATATACAAGAAAGAAGAAATTAACAAGTACC
RM 18	UPT _{TAL[Bs4S]} F/RobM	CAAAATATCATCAATTGATCTCATCCATAC
RM 19	UPT _{AvrBs4in3} -F/PR	TATAATTAATAATCCACTTCTGGTTAAACAATGAACACGTTTGC
RM 20	UPT _{AvrBs4in3} -R/PR	GGTGTGCAAAATTTGGTTTAAACCC
RM 21	DUPT _{AvrBs3inBs3p} -F/RobM	TCACAACCTCAAGTTATCATCCCC
RM 22	DUPT _{AvrBs3inBs3p} -R/RobM	ATAAAATTGGTCAGGCCAAACGTGTTC
RM 23	UPT _{AvrBs3/Bs3inBs4Sp} -F/PR	CAATTTTATTATATAAACCTAACCTCCTCACACGTTTCAAGTGGTACTTGT
RM 24	UPT _{AvrBs3/upa20inBs4Sp} -F/RobM	CTTCATCTTTATATAAACCTGACCCTTTGTGACATGTTTCAAGTGGTACTTGT
RM 25	UPT _{dTAL[EGL3]} inBs4Sp-F/RobM	TTTTGTGATTATATACAGTGTACACACACCTAAAAGTTTCAAGTGGTACTTGT
RM 26	UPT _{dTAL[KNAT1]} inBs4Sp-F/RobM	GACGAATTCTATATACCTAGTTCTGTTTTTCTCCGTTTCAAGTGGTACTTGT
RM 27	Bs4p-R/PR	GTGAAAGCTTGATTAACATTCGCTTTG
RM 28	Bs4 RT1 F/RobM	CAAGATGATAAAAAGACTAGAGCATGGAG
RM 29	Bs4 RT1 R/RobM	TGGAAGAGATGCAATCTACGATCTGCT
RM 30	RS-EFrt-F1	AGTCAACTACCACTGGTCAC
RM 31	RS-EFrt-R1	GTGCAGTAGTACTTAGTGGTC
RM 32	KNAT1 RT1 F	CAAGCTTACTTGGACTGCCAAAAAG
RM 33	KNAT1 RT1 R	TTAACCAACATGTCACAGTATGCTTC
RM 34	EGL3 RT1 F/RobM	CGCCCGAGAAAATTGAATGAACG
RM 35	EGL3 RT1 R/RobM	CCATGCAACCCCTTTGAAGTGCC
RM 36	actin2 RT F/RobM	GACTACGAGCAGATGGAAACC
RM 37	actin2 RT R/RobM	CTGGACCTGCCTCATCATACTCGG
RM 38	UPA20 RT F/RobM	GCACTACTACATCACAAAGCAG
RM 39	UPA20 RT R/RobM	CTAGAGGCTAGCTATAGTCAG
RM 40	Cand-7-01-fwd	ATGAATCAGAATTGCTTTAAATCTTGTTC
RM 41	Cand-7-01-rev	TGATTCTTGTGCTACATTTGTTCTTTCC
RM 42	AT5G20270.1 K F/RobM	ACAACCAACGTGGTCACGCCAGTCC
RM 43	AT5G20270.1 K R/RobM	GCAAAGAGTCCAGGGAAATGTAACAAGATGG
RM 44	AT3G15640 G F/RobM	AGAATCGTCTCCTCTCAGCTCAAAACC
RM 45	AT3G15640 G R/RobM	ATCCGGAGGACCACAGGACCAACC
RM 46	AT3G59220 G F/RobM	GAAGGTGAAGGAGCTGTCGTTAGAAATGG
RM 47	AT3G59220 G R/RobM	CAAACCCGTTCTTAGCATTCTGATAGTCATC
RM 48	AT4G27740 G F/RobM	ATGGCAGCAAACAAAACCTTCTCCTACG
RM 49	AT4G27740 G R/RobM	TCTTTTGGTCAACTTGAGCTTCTCGATGAC

Other Supporting Information Files

[SI Text \(DOC\)](#)

Supporting Information - Text

Nucleotide sequences of codon-optimized TALE *avrBs3* and *avrBs4* genes. (*)

indicates that the given *TALE* genes are optimized for expression in tobacco and thus differ from the corresponding *Xanthomonas* wildtype genes. Please note that corresponding *TALE* genes with tobacco and *Xanthomonas* codon usage translate into identical amino acid sequences.

*avrBs3**

```

ATGGATCCCATAGGTCTAGGACTCCTTCACCTGCTAGGGAGCTTTGCGCTGGTCTCAGCCAGATGGAGTTCAACCAACTGCAGA
TAGGGGAGTTTCTCCTCCTGCTGGAGGACCCTTGATGGTCTTCTCCTGCTAGGAGAACAATGTCTAGAACAAGATTGCCTTCTCCTC
CTGCTCCATCTCCTGCTTTCTCTGCTGGTTCATTTTCTGATTTGTTGAGACAGTTGATCCTTCTTTGTTAATACTTCACATTTTT
GATCTTTTCCACCATTGGTGTCTCATCTACTGAGGCTGCTACAGGTGAATGGGATGAAGTTCAATCTGGACTTAGAGCTGCTGA
TGCTCCTCCTCCTACTATGAGAGTTGCTGTTACAGCTGCTAGGCCACCAAGAGCTAAGCCTGCTCCTAGGAGAAGAGCTGCTCAAC
CATCTGATGCTTCTCCTGCTCAAGTTGATCTTAGAACATTGGGATATTACACAACAACAGCAGGAGAAGATTAAGCCTAAGGTT
AGATCTACAGTTGCTCAACATCATGAAGCTCTTGTGGTTCATGGTTTCACTCATGCTCATATTGTTGCTCTTTCTCAACATCCTGC
TGCTCTTGGTACTGTTGCTGTTAAGTATCAGGATATGATTGCTGCTCTCCAGAAGCTACTCATGAAGCTATTGTTGGTGTGGTA
AGCAGTGGTCTGGTCTAGGGCTCTTGGCCCTTGTACTGTTGCTGGGAATTGAGAGGGCCTCCACTGCAATTAGATACAGGT
CAGTTGCTCAAGATAGCTAAACGGGGAGGAGTACTGCCGTGGAAGCTGTGCATGCTGGAGGAATGCTTTGACGGGGGACCGCT
GAATCTTACACCTGAGCAGGTTGTCGCAATCGCTAGTCACGATGGCGGCAACAGGCATGGAGACCGTGCAACGCTTCTCCAG
TCTTGTGTCGCAAGCTCATGGACTGACACCGCAACAAGTGGTAGCAATCGCAAGCAATGGCGGAGGGAAGCAGGCTCTTGAAGCTGTA
CAGAGTTTCTACCTGTCTATGCCAGGGGACCGGGCTGACACCAACAAGTTGTAGCTATCGCTAGCAATAGTGGCGGAAAGCA
GGCATTGGAGACTGTTCAAAGGCTACTCCAGTTCTATGTCAGGCTCATGGACTGACACCTGAACAAGTGGTGTGCTATCGCTAGCA
ACGGTGGTGAAAGCAAGCACTTGAGACGGTTCAACGACTCCTGCCCCTACTGTGCCAAGCTCATGGACTTACGCCTGAGCAGGTA
GTTGCTATTGCTTCTAATATAGGAGGAAAGCAGGCTCTCGAAACTGTGCAAGCGCTTCTACCTGTACTATGCCAGGCTCATGGCCT
TACCCAGAACAGGTAGTCGCAATGCTTCAATATCGGAGGTAACAGGCTCTTGAGACCGTTCAAGCCTTGTGGCAGTGTAT
GTCGAAGCATGGGCTAACTCCTGAGCAAGTTGTCGAATAGCTTCAAAATAGGGGGAAAACAAGCCCTTGAGACAGTCCAAAGCC
TTGTTGCCGGTCTATGCCAGGCACATGGATTGACACCGGAACAGGTTGTTGCGATCGCTAGCCACGATGGTGGCAAGCAGGCACT
TGAAACAGTTCAGAGACTGTTACCAGTACTATGTCAAGCTCACGGACTCACCCCTGAACAGGTGGTGGCGATCGCTTACACAGATG
GAGGCAAGCAGGCTCTAGAGACTGTGCAGCGTCTTTGCCAGTCTCTGTCAAGCCATGGATTGACACCAGCAAGTTGTAGCT
ATTGCGTCTAATGGTGGAGGTAACAAGCCCTGGAAACCGTCCAGAGACTTCTTCCAGTGTCTTTGTCAAGCACATGGACTCACACC
TGAACAAGTTGTTGCTATCGCAAGCAATCTGGAGGTAACAGGCGCTTGAGACAGTTCAGCTCTTCTCCAGTGTCTGTCAAG
CTCATGGACTCACTCCTGAACAGGTGGTGGCTATCGCAAGCAACTCTGGAGGCAACAAGCATTGGAGACCGTGCAGAGACTCTG
CCTGTCAAGCCATGGCTTGACACCACAGCAGGTGGTGGCAATGTCATCAAACGGAGGAGGACGGCCAGCACTTGAGTCGATTGT
TGCGCAATTGAGCAGACCCGATCCTGCTTTGGCTGCGCTAACGAATGATCACCTCGTTGCTTTAGCATGCCTAGGAGGTCGGCCGG
CTCTGGATGCTGTCAAGAAAGGGCTTCCACATGCCCTGTCTTTAAGCGTACTAACCGACGAATCCCAGAGCCACCAGCCAT
AGAGTAGCTGACCACGGCAAGTTGTTAGAGTCTTAGGCTTTTTCCAATGCCATTTCTATCCCCTCAGGCTTTCGACGATGCAAT
GACTCAATTGCAATGCTTAGACATGGTCTTCTTCAATTGTTTCGAGAGTTGGTGTCAACGAACTCGAAGCAGTGTGGAATC
TCCCTCCTGCAAGTCAGAGATGGGACCGCATTTTGAAGCAAGTGGTATGAAAAGGGCTAAACCGTCTCCAAACAGTACCCAAACG
CCTGATCAGGCTTCTTTCACGCATTTGCTGATTCGCTGGAGCGAGATCTTGACGCTCCGTACCAATGCATGAAGGAGACCAAC
ACGAGCGTCATCTGAAAGCGATCTAGGAGCGATCGTGAGTGACCGGTCATCTGCTCAACAATCATTTGAGGTTAGGGTTCCTG

```

AACAAAGAGATGCTTTGCATTTGCCTTTGTCTATGGAGGGTTAAGAGACCTAGAACTTCTATTGGAGGAGGTTGCCGGATCCTGGA
 ACTCCTACTGCTGATCTTGTCTGCTTCTCAACTGTTATGAGAGAGCAAGATGAAGATCCTTTTGTCTGGAGCTGCTGATGATTT
 CCTGCTTTTAAATGAGGAAGAGCTTGCTTTGGCTTATGGAATTGCTTCCACAGTAA

*avrBs4**

ATGGATCCCATTAGGTCTAGGACTCCTTACCTGCTAGGGAGCTTTTGCCTGGTCTCAGCCAGATGGAGTTCAACCAACTGCAGA
 TAGGGGAGTTTCTCCTCCTGCTGGAGGACCCTTGATGGTCTTCTGCTAGGAGAAACAATGTCTAGAACAAGATTGCCCTTCTCCTC
 CTGCTCCATCTCCTGCTTTCTCTGCTGGTTCATTTTCTGATTTGTTGAGACAGTTCGATCCTTCTTTGTTTAAATACTTCACTTTTT
 GATCTTTTCCACCATTGGTGTCTATCATACTAGGCTGCTACAGGTGAATGGATGAAGTTCAATCTGGACTTAGAGCTGCTGA
 TGCTCCTCCTACTATGAGAGTTGCTGTTACAGCTGCTAGGCCACCAAGAGCTAAGCCTGCTCCTAGGAGAAGAGCTGCTCAAC
 CATCTGATGCTTCTCCTGCTGCTCAAGTTGATCTTAGAACAATTGGGATATTCACAACAACAGCAGGAGAAGATTAAGCCTAAGGTT
 AGATCTACAGTTGCTCAACATCATGAAGCTCTTGTGGTGTGCTGCTTACTCATGCTCATATTGTTGCTCTTTCTCAACATCCTGC
 TGCTCTTGGTACTGTTGCTGTTAAGTATCAGGATATGATTGCTGCTTCCAGAAGCTACTCATGAAGCTATTGTTGGTGTGGTA
 AGCAGTGGTCTGGTCTAGGGCTCTTGGGCTTGTACTGTTGCTGGGAATTGAGAGGGCCTCCACTGCAATTAGATACAGGT
 CAGTTGCTCAAGATAGCTAAACGGGGAGGAGTACTGCCGTGGAAGCTGTGCATGCCTGGAGGAATGCTTTGACGGGGGCACCGCT
 GAATCTTACACCTGAGCAGGTTGTCGCAATCGCTAGTAATATTGGCGGCAACAGGCATTGGAGACCGTGAAGCTCTTCTCCAG
 TCTTGTGTCAAGCTCATGGACTGACACCGGATCAAGTGGTAGCAATCGCAAGCAATGGCGGAGGGAAGCAGGCTCTTGAACCTGTA
 CAGAGGTTGCTACCTGTCTATGCCAGGCGCACGGGCTGACACCAGAACAAGTTGTAGCTATCGCTAGCAATATTGGCGGAAAGCA
 GGCATTGGAGACTGTTCAAAGGCTACTCCAGTCTATGTCAAGGCTCATGGACTGACACCTGAACAAGTGGTGTGCTATCGCTAGCA
 ACATTGGTGGAAAGCAAGCACTTGGAGACGGTTCACAGCTCCTGCCCTACTGTGCCAAGCTCATGGACTTACGCCTGAGCAGGTA
 GTTGTCTATTGCTTCTAATGGTGGAGGAAAGCAGGCTCTCGAAACTGTGCAACGCTCTTCTACCTGTACTATGCCAGGCTCATGGCCT
 TACCCAGAAGCAGGTAGTCGCAATTGCTTCAAGTGGAGGAGGTAACAGGCTCTTGGAGACCGTTCAAAGATTGTTGCCAGTGTAT
 GTCGAAGCACATGGGCTAATCCCTCAACAAGTTGTTGCAATAGCTTCAAATATAGGGGGAAAACAAGCCCTTGAGACAGTCCAAAG
 TTGTTGCCGGTCTATGCCAGGCACATGGATTGACACCGCAACAGGTTGTTGCGATCGCTAGCAATTCTGGTGGCAAGCAGGCACT
 TGAACAGTTCAGAGACTGTTACCAGTACTATGTCAAGCTCACGGACTCACCCCTGAACAGGTGGTGGCGATCGCTTCAAATGGT
 GAGCAAGCAGGCTCTAGAGACTGTGACGCTCTTTGCGAGCTCTGTCAAGCCCATGGATTGACACCACAGCAAGTTGTAGCT
 ATTGCGTCTAATATTGGAGGTAACAAGCCCTGGAAACCGTCCAGAGACTTCTTCCAGTGTCTTGTCAAGACCATGGACTCACACC
 TCAGCAAGTGTGCTATCGCAAGCAATCTGGAGGTAACAGGCGCTTGAGACAGTTCAAAGACTTCTTCCAGTGTCTGTCAAG
 CTCATGGACTCACTCCTCAGCAGGTGGTGGCTATCGCAAGCAACGGTGGAGGCAACAAGCATTTGGAGACCGTGCAGAGACTTCTG
 CCGGTGCTTTGCCAGGCCCATGGTCTAACACCCCAACAGGTCGTGCTATCGCTTACATGATGGTGGAAAGCAAGCTCTTGAAC
 TGTACAAAGGCTGTTACCAGTCTCTGCCAGGCTCATGGATTGACTCCACAACAAGTAGTCGCTATAGCCAGTCTGATGGTGGAA
 AACAGGCTCTTGGAGACTTACAGAGATTATACCTGTCTGTGTCAAGCCACGGACTGACACCACAACAAGTGTGTGCTATTGCC
 TCTAATTCTGGTGGTAAACAGGCCCTAGAGACCGTTCAAAGGTTACTTCCAGTGTGTGTCAAGCCATGGACTAACACCCGAACA
 AGTGTGTGCTATTGCTTCTCATGATGGAGGAAAACAAGCTTTGGAGACGGTACAGAGACTTACTTACTGTACTGTGTCAGGCTCATG
 GTCTTACTCCTCAACAGGTAGTAGCTATCGCTTCAAGTGGAGGTGGAAAGCAGGCACTGGAGACAGTCCAGAGATTGCTGCCGGTT
 CTCTGTCAAGCCCATGGCTTGGACACCACAGCAGGTGGTGGCAATGCATCAAACGGAGGAGGACGGCCAGCACTTGGTGTGATTGT
 TGCGCAATTGAGCAGACCCGATCCTGCTTTGGCTGCGCTAACGAATGATCACCTCGTTGCTTTAGCATGCCTAGGAGTTCGGCCGG
 CTCTGGATGCTGTCAAGAAAGGGCTTCCACATGCCCTGCTCTTATTAAGCGTACTAACCGACGAATCCCAGAGCGCACACGCCAT
 AGAGTAGCTGACCACGGCAAGTGTAGAGTCTTAGGCTTTTTCCAATGCCATCTCATCCCGCTCAGGCTTTCGACGATGCAAT
 GACTCAATTGGAATGCTTAGACATGGTCTTCTTCAATTGTTTCGACAGATTGGTGTACCGAAGTCAAGCAGTGTAGTGGAACTC
 TCCCTCCTGCAAGTCAAGATGGGACCGCATTTTGCAGCAAGTGGTATGAAAAGGGCTAAACCGTCTCCAACAGTACCCAAACG
 CCTGATCAGGCTTCTTTCACGCATTTGCTGATTCGCTGGAGCGAGATCTTGCAGCTCCGTACCAATGCATGAAGGAGACCAAC
 ACGAGCTCATCTCGAAAGCGATCTAGGAGCGATCGTGCAGTGACCGGTCATCTGCTCAACAATCATTTGAGGTTAGGGTTCCTG
 AACAAAGAGATGCTTTGCATTTGCCTTTTGTCTATGGAGGGTTAAGAGACCTAGAACTTCTATTGGAGGAGGTTGCCGGATCCTGGA
 ACTCCTACTGCTGATCTTGTCTGCTTCTCAACTGTTATGAGAGAGCAAGATGAAGATCCTTTTGTCTGGAGCTGCTGATGATTT
 CCTGCTTTTAAATGAGGAAGAGCTTGCTTTGGCTTATGGAATTGCTTCCACAGTAA

Sequence polymorphisms in RVD and non-RVD repeat unit residues of AvrBs3.

Sequence alignment of the AvrBs3 repeat units 1-17. Amino acids are displayed in single-letter code. Vertical numbers at the top of sequence (read from top to bottom) indicate the position of the given residue within the repeat. The repeat variable diresidues (RVDs at position 12 and 13 are displayed in boldface. An alanine (A) residue at repeat unit position 24 that is found in all NI-type RVDs is highlighted in red font. Alignments were constructed with ClustalW. Identical amino acids (white text on black background) and 50% similar amino acids (white on grey background) were shaded using Boxshade.

	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	3	3	3	3			
	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4
rep01	L	T	P	E	Q	V	V	A	I	A	S	H	D	G	G	K	Q	A	L	E	T	V	Q	R	L	L	P	V	L	C	Q	A	H	G
rep02	L	T	P	Q	Q	V	V	A	I	A	S	N	G	G	G	K	Q	A	L	E	T	V	Q	R	L	L	P	V	L	C	Q	A	H	G
rep03	L	T	P	Q	Q	V	V	A	I	A	S	N	S	G	G	K	Q	A	L	E	T	V	Q	R	L	L	P	V	L	C	Q	A	H	G
rep04	L	T	P	E	Q	V	V	A	I	A	S	N	G	G	G	K	Q	A	L	E	T	V	Q	R	L	L	P	V	L	C	Q	A	H	G
rep05	L	T	P	E	Q	V	V	A	I	A	S	N	I	G	G	K	Q	A	L	E	T	V	Q	A	L	L	P	V	L	C	Q	A	H	G
rep06	L	T	P	E	Q	V	V	A	I	A	S	N	I	G	G	K	Q	A	L	E	T	V	Q	A	L	L	P	V	L	C	Q	A	H	G
rep07	L	T	P	E	Q	V	V	A	I	A	S	N	I	G	G	K	Q	A	L	E	T	V	Q	A	L	L	P	V	L	C	Q	A	H	G
rep08	L	T	P	E	Q	V	V	A	I	A	S	H	D	G	G	K	Q	A	L	E	T	V	Q	R	L	L	P	V	L	C	Q	A	H	G
rep09	L	T	P	E	Q	V	V	A	I	A	S	H	D	G	G	K	Q	A	L	E	T	V	Q	R	L	L	P	V	L	C	Q	A	H	G
rep10	L	T	P	Q	Q	V	V	A	I	A	S	N	G	G	G	K	Q	A	L	E	T	V	Q	R	L	L	P	V	L	C	Q	A	H	G
rep11	L	T	P	E	Q	V	V	A	I	A	S	N	S	G	G	K	Q	A	L	E	T	V	Q	A	L	L	P	V	L	C	Q	A	H	G
rep12	L	T	P	E	Q	V	V	A	I	A	S	N	S	G	G	K	Q	A	L	E	T	V	Q	R	L	L	P	V	L	C	Q	A	H	G
rep13	L	T	P	E	Q	V	V	A	I	A	S	H	D	G	G	K	Q	A	L	E	T	V	Q	R	L	L	P	V	L	C	Q	A	H	G
rep14	L	T	P	E	Q	V	V	A	I	A	S	H	D	G	G	K	Q	A	L	E	T	V	Q	R	L	L	P	V	L	C	Q	A	H	G
rep15	L	T	P	E	Q	V	V	A	I	A	S	H	D	G	G	K	Q	A	L	E	T	V	Q	R	L	L	P	V	L	C	Q	A	H	G
rep16	L	T	P	Q	Q	V	V	A	I	A	S	N	G	G	G	R	P	A	L	E	T	V	Q	R	L	L	P	V	L	C	Q	A	H	G
rep17	L	T	P	E	Q	V	V	A	I	A	S	H	D	G	G	K	Q	A	L	E	T	V	Q	R	L	L	P	V	L	C	Q	A	H	G

The AvrBs3 and AvrBs4 repeat domains differ in 18 RVD and 20 non-RVD residues. Pairwise alignment of the AvrBs3 and AvrBs4 repeat units. Amino acids (aa) are shown in the single-letter code. Vertical numbers at the top (read from top to bottom) indicate the location within the 34-aa repeat unit. Repeat-variable diresidues are at positions 12 and 13 of each repeat unit and are displayed in boldface. Alignments were constructed with ClustalW. Identical amino acids (white text on black background) and 50% similar amino acids (white on grey background) were shaded using Boxshade.

```

0000000001111111112222222222233333
1234567890123456789012345678901234
AvrBs3-rep01 LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG
AvrBs4-rep01 LTPEQVVAIASNIGGKQALETVQRLLPVLCQAHG
AvrBs3-rep02 LTPQQVVAIASNNGGGKQALETVQRLLPVLCQAHG
AvrBs4-rep02 LTPDQVVAIASNNGGGKQALETVQRLLPVLCQAHG
AvrBs3-rep03 LTPQQVVAIASNSGGKQALETVQRLLPVLCQAHG
AvrBs4-rep03 LTPEQVVAIASNIGGKQALETVQRLLPVLCQAHG
AvrBs3-rep04 LTPEQVVAIASNNGGGKQALETVQRLLPVLCQAHG
AvrBs4-rep04 LTPEQVVAIASNIGGKQALETVQRLLPVLCQAHG
AvrBs3-rep05 LTPEQVVAIASNIGGKQALETVQRLLPVLCQAHG
AvrBs4-rep05 LTPEQVVAIASNNGGGKQALETVQRLLPVLCQAHG
AvrBs3-rep06 LTPEQVVAIASNIGGKQALETVQRLLPVLCQAHG
AvrBs4-rep06 LTPEQVVAIASNNGGGKQALETVQRLLPVLCQAHG
AvrBs3-rep07 LTPEQVVAIASNIGGKQALETVQRLLPVLCQAHG
AvrBs4-rep07 LIPEQVVAIASNIGGKQALETVQRLLPVLCQDHHG
AvrBs3-rep08 LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG
AvrBs4-rep08 LTPQQVVAIASNSGGKQALETVQRLLPVLCQAHG
AvrBs3-rep09 LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG
AvrBs4-rep09 LTPEQVVAIASNNGGGKQALETVQRLLPVLCQAHG
AvrBs3-rep10 LTPQQVVAIASNNGGGKQALETVQRLLPVLCQAHG
AvrBs4-rep10 LTPQQVVAIASNIGGKQALETVQRLLPVLCQDHHG
AvrBs3-rep11 LTPEQVVAIASNSGGKQALETVQRLLPVLCQAHG
AvrBs4-rep11 LTPQQVVAIASNSGGKQALETVQRLLPVLCQAHG
AvrBs3-rep12 LTPEQVVAIASNSGGKQALETVQRLLPVLCQAHG
AvrBs4-rep12 LTPQQVVAIASNNGGGKQALETVQRLLPVLCQAHG
AvrBs3-rep13 LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG
AvrBs4-rep13 LTPQQVVAIASHDGGKQALETVQRLLPVLCQAHG
AvrBs3-rep14 LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG
AvrBs4-rep14 LTPQQVVAIASHDGGKQALETVQRLLPVLCQAHG
AvrBs3-rep15 LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG
AvrBs4-rep15 LTPQQVVAIASNSGGKQALETVQRLLPVLCQAHG
AvrBs3-rep16 LTPQQVVAIASNNGGGKQALETVQRLLPVLCQAHG
AvrBs4-rep16 LTPQQVVAIASHDGGKQALETVQRLLPVLCQAHG
AvrBs3-rep17 LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG
AvrBs4-rep17 LTPQQVVAIASNNGGGKQALETVQRLLPVLCQAHG
AvrBs3-rep18 LTPQQVVAIASNNGGRRPALESIVA
AvrBs4-rep18 LTPQQVVAIASNNGGRRPALESIVA

```


*UPT*_{dTALE[KNATI]}*inBs4Sp*. Lowercase letter marks the sequence from the *A. thaliana* *KNATI* promoter that was inserted into the *Bs4S* promoter. The dTALE[*KNATI*] target sequence (*UPT*_{dTALE[KNATI]}) is underlined.

ATTTTCAAGTTTTTATTATTTTAATCTTTTGATGGCTTTTGTTAGTGGAAAAAGGGCAACGGTAAACAA
ATAAATTAATTTGTCAACTCTTCTTTGTTGTCCTTTGACAAAGTCAGCTTTCATTAGTGTGACAGAGAA
TTTTATTTAGATTATTTTTTATTGTTGAAGACTGATCAAAGCGAATGTTAATACAAGCTTTCACgacga
attctatatacctagttcgttttttcttccGTTTCAAGTGGTACTTGTTTAATCTTCTTTCTTGTATA
TAACCTTGTCCAAAATATCATCAATTGATCTCATCCATACAATTTATTTTAAATCGAATCT

*UPT*_{AvrBs3Bs3p}*inBs4Sp*. Lowercase letter marks the sequence from the *C. annuum* *Bs3* promoter that was inserted into the *Bs4S* promoter. The AvrBs3 target sequence that is present in the *Bs3* promoter (*UPT*_{AvrBs3Bs3p}) is underlined.

ATTTTCAAGTTTTTATTATTTTAATCTTTTGATGGCTTTTGTTAGTGGAAAAAGGGCAACGGTAAACAA
ATAAATTAATTTGTCAACTCTTCTTTGTTGTCCTTTGACAAAGTCAGCTTTCATTAGTGTGACAGAGAA
TTTTATTTAGATTATTTTTTATTGTTGAAGACTGATCAAAGCGAATGTTAATACAAGCTTTCACccta
tttattatataaacctaaccatcctcacaacGTTTCAAGTGGTACTTGTTTAATCTTCTTTCTTGTAT
ATAACTTTGTCCAAAATATCATCAATTGATCTCATCCATACAATTTATTTTAAATCGAATCT

*UPT*_{AvrBs3UPA20p}*inBs4Sp*. Lowercase letter marks the sequence from the *C. annuum* *UPA* promoter that was inserted into the *Bs4S* promoter. The AvrBs3 target sequence that is present in the *UPA20* promoter (*UPT*_{dAvrBs3UPA20p}) is underlined.

ATTTTCAAGTTTTTATTATTTTAATCTTTTGATGGCTTTTGTTAGTGGAAAAAGGGCAACGGTAAACAA
ATAAATTAATTTGTCAACTCTTCTTTGTTGTCCTTTGACAAAGTCAGCTTTCATTAGTGTGACAGAGAA
TTTTATTTAGATTATTTTTTATTGTTGAAGACTGATCAAAGCGAATGTTAATACAAGCTTTCACcctca
tctttatataaacctgacccttggatgacatGTTTCAAGTGGTACTTGTTTAATCTTCTTTCTTGTATA
TAACCTTGTCCAAAATATCATCAATTGATCTCATCCATACAATTTATTTTAAATCGAATCT

Xa27-promoter from the rice cultivar IRBB27 (GenBank: AY986492.1). The AvrXa27 target sequence is underlined.

CTTTTACTCTCTATCACATCGGATGTTTGGACACTAATTATAAATATTTAAACGTAGACTATTAATAAAA
CCCATCTATAATCTTGTATTAATTCGCGAGACGAATCTATTGAGCCTAATTAATCCATGATTAGCCTAT
GTGATGCTATAATAAACATTCTCTAATTATAAATTAATTGGGCTTAAAAAATTTGTCTCGCGTATTAGC
TTTCATTTATATAATTAGTTTTATAAATAGTCTATATTTAATACTCTAAATTAGTGTCTAAATACAGGG
ACTAAAGTTAAGTCACTGGATCCAAACACCACCTAAGGTTTTCTTGTGTACTTGTGAATTGTGGTTGAC
TACGACTACTAGTCTATAAATAGAAAGAGACCCATAGAGAGCATCAGAGCAAAGTACTCCTAAAAG
ACAGCCACACACTGAGACACCCAAGAAGCTGCCTCCA

Repeat variable diresidues of dTALEs and corresponding DNA recognition sequences.

RVDs of the dTALE[*Bs4S*] that targets the $UPT_{\text{dTALE}[Bs4SP]}$ box of the tomato *Bs4* promoter

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
HD	NG	NS	NG	NI	NI	HD	NG	NG	NG	NN	NG	HD	HD	NI	NI	NI	NI
A	T	A	T	A	A	C	T	T	T	G	T	C	C	A	A	A	A

RVDs of the dTALE[*EGL3*] that targets the $UPT_{\text{dTALE}[EGL3]}$ box of the *A. thaliana* *EGL3* promoter

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
HD	NG	NS	NG	NI	HD	NI	NN	NG	NN	NG	NS	HD	NI	HD	NI	HD	NI
A	T	A	T	A	C	A	G	T	G	T	A	C	A	C	A	C	A

RVDs of the dTALE[*KNATI*] that targets the $UPT_{\text{dTALE}[KNATI]}$ of the *A. thaliana* *KNATI* promoter

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
HD	NG	NS	NG	NI	HD	HD	NG	NI	NN	NG	NG	HD	NN	NG	NG	NG	NG
A	T	A	T	A	C	C	T	A	G	T	T	C	G	T	T	T	T

Analysis of dTALE[*EGL3*] and dTALE[*KNATI*] target and potential off-target sites in the Arabidopsis genome.

User defined target sequences of dTALE[*EGL3*] and dTALE[*KNATI*] are highlighted with yellow background. The ATG start codon is displayed in boldface black. Non-translated regions (UTR and Introns) are shown as lowercase letters. Exons are displayed in uppercase font. 5' UTR regions are shown in red font. Primer regions used for RT-PCR studies are underlined.

EGL3 (AT1G63650)

```

ttatcatgaaacgtagagaataggacatatcattgatagaaaaagatttttagagatggaaaaataact
aatttctcttaacaaaaataaatttccaaaaggagagtggctgctggaacaccgtattcAAAagggtta
tacctagacactacttttttttttttttaccatcttgttttataataatttatttAAAattcctcgt
aagttcagtggtttgtgat tatatacagtggtacacaca cctaaaaaagtatccataaagagatttgg
cattcggacaccgtccttttctcactcaaaagctcaattcaaccacaacagatagtaacttgaag
cacaaatatttaataattttccattatagaaggttaaaaaaaataaataatattctgggttttgggtct
tcttctttctcctcaggaaaccaaaccaatgtcatgt aaaaatcactaactcatgactagtcataacc
tttaaatcacttctgcttttttctgcttctcttttgatttattcaagacaacaaagttttgcttttg
acaaaaaatctgatcttttggcttctgctttgtttgttacatggggatgaagaaacaATGGCAACCGGAG
AAAACAGAACGGTGCCGGACAATCTAAAGAAACAGCTCGCAGTTTCAGTTCGAAACATTCCAATGGAGTT
ATGGAATCTTCTGGTCTGTCTCTGCTTCTCAACCAGGGgtacaatctcgtaatctttatccttcaatta
acttttcttgaaaataaagttttgcttgagacattcttgacttttagtccaaaactcagataaaacttta
tccttacttcaagatttcgatttttggttatgttttctcgattaatccaatttttcttggtatctttt
tgttttctctaaacgaagAGTGTGGAGTGGGGAGATGGATATTACAATGGAGACATAAAGACAAGGAA
GACGATTCAAGCAGCAGAAGTCAAAATTGACCAGTTAGGTCTTGAGAGAAGTGAGCAGCTTAGAGAGCT
TTATGAATCTCTCTCCCTCGCTGAATCCTCAGCTTCCGGTAGCTCTCAGGTCACTAGACGAGCTTCCGC
CGCCGCTCTCTCACCGGAGGACCTCACCGACACCGAGTGGTACTACTTAGTATGCATGTCTTTCGTCTT
CAACATCGGTGAAGGGtaattccgctctcttctcataatcctttcatttcaccttttAAAAcaaaa
acctatctttcagaaaaatttcttttatgaagaatatcgtttttaaattatagtatttttattcgt
gacattataagaaaatgtaaacgataccaaaattctaaaaaaaagtataagggtatgttttattttgt
ctcataaaaaaagacaagaacaccctttattagcttgctccttcatttcttctccatgtgagcttat
tcagaaatccgacatgtacgagtttaattattaacaatgcaattatcctttatatttacttaattttt
ttatgtttgtaggtttttttaaagtataaccttttttaactcttagaagacatacatagatagaatgcg
aaagcacaacggaaataagtttttacttcatgttcataattgcattatacatttgctattcaagaaag
tagtagaaaacttttgttcccattaagataaagattttgagatacaaaacataaagggtatttccgctct
ttctaagttagAATCCCCGGAGGAGCGTTATCCAATGGAGAACCAATATGGCTTTGTAACGCTGAAACC
GCCGATAGCAAAGTCTTCACTCGTCTCTTCTAGCTAAAGtaagttttttgcaattggaaacgctaacg
ctagaattcaagaacgctataactgaaaacgctcttttttttttttttttttctgcttggcgcttttgcct
ttaaagAGTGCTTCGCTTCAGACAGTGGTTTGCTTCCCGTTTCTTGAGGAGTCCTTGAGATTGGCAGC
ACCGAACATgtaagcctataatttctatttttttccaaactatcaaaacatttcttaaagtagccaa
aacttatcttttgttcttttagATTAAGAGGACATGAACGTGATACAAAGTGTTAAGACGTTGTTCCCT
TGAAGCTCCTCCATATACTACAATATCGACAAGATCAGACTATCAAGAAATTTTTGATCCCTTAAGTGA
CGATAAATACACTCCGGTGTATAACCGAAGCTTTTCCAACAACCTTACTAGCGGGTTTGAGCAAGA
ACCTGAGGATCATGATTCGTTTCATCAACGATGGTGGTGCCTCAGGTACAAAGCTGGCAGTTTGTGGG
TGAAGAAATCAGTAACTGCATTCACCAATCGTTAAATCAAGCGATTGCGTTTCCCAAACGTTTGTGG
AACAACCGGGGAGACTTGCTTGCGATCCAAGGAAGAGTAGGATTCACCGTTAGGTGAGATTCAAGAACA
GAGTAACCATGTAAATATGGACGACGATGTTTACCAAGGCGTGATATCGACGATTTTCAAAAACAAC
GCATCAGCTAATACTCGGACCGCAGTTTCAGAACTTCGATAAGCGGTCTAGCTTCAAGGTGGAAGCG
ATCATCATCTGTGAAAACATTTGGGAGAGAAAATCGCAGAAGATGATAAAGAAGATACTTCTCGAGGTTCC

```

TTTGATGAACAAGAAAGAAGAGTTGTTACCGGACACACCAGAGGAAACCGGGAACCATGCCTTGTCCGA
 GAAGAAACGCCCGCAGAAAATTGAATGAACGGTTTATGACATTGAGATCAATCATTCCTCAATTAGTAA
 Ggtaaccgattctagcaaatctgtttcagtaacgttttatgtttatttcttgataataggaagacatgt
 ttaaataataatgctctgaaggaactaagcttgagtctcttgctgattctacagATTGATAAAGTGTCTGA
 TTCTTGATGATACAATTGAGTATCTTCAAGATTTACAGAAACGGGTTCAAGAGTTGGAATCTTGTAGAG
 AATCTGCTGATACAGAGACACGGATAACGATGGAAGGAAACCGGATGATGAGGAGGAAAGAG
 CATCAGCGAATTGTATGAACAGCAAAAGGAAGGAGTGAATGTGAATGTAGGAGAAGATGAACCAGCTG
 ATATCGGTTATGCTGGTCTAACGGATAACTTAAGGATCAGTTCATTAGGTAACGAGGTGGTTATTGAGC
 TTAGATGCGCTTGAGAGAAAGGGATATTGCTTGAGATAATGGATGTGATTAGTGTATCTCAACTTGGATT
 CTCACTCGGTTCACTCGTCAACCGGAGACGGTTTACTCTGCTTAACTGTCAATTGCAAGGtacagtcat
 taaaccaagcttgatttatgttactcaaccggaatatgagcaaaccgaaatgtgatatgtaatgtttaa
 tgttgattgcagCATAAAGGGACAAAAATAGCAACAACAGGAATGATCCAAGAGGCACTTCAAAGGGTT
 GCATGGATATGTTAA

AT3G59220 (potential dTALE[*EGL3*] off-site)

tatatatagtttacacacacacgcccactaccbaatacaaaagcactgacttcttccattaacaagcttttc
 ttcgtttatcatcaccatcaccatcatcatcatcatcatcatcgatcatcatcatcatcatcatc**ATG**AC
 GTATGAGAACAATAGTGTACCAAGAATTGTAATCAAGAAAGTTCTAGCAAAGCTTGAGAAGGAAGGTGA
 AGGAGCTGTCTGTTAGAAATGGCATCACAAAgtagtctcttcttcttcttcttctcaaaagcaaaatcca
 atgccttcaccttcaagtttcatatatttttttttcttttttgggttgattcagGATTGACCAGAAGTTA
 TTGGACCCGTTCTGTGTTGCTAGTTGAATTTTCTGtaagtttcttcttccccgaatctgtcaattaa
 aagtgtggttcttttattatgtagcatgtcataatcatttgctctgtttggtgatcagTTTCACTCTC
 AGCTGGATTCCCAGATCATCTCACAGAGgtcaatttccacataaccttcttacacacgaagtgtctct
 gttctgaaatgtctataatgtaattgatcaactgtgtgcagGTTTTGAAAGTGTACATACATGCTACA
 GGGAGGTATCATTACAAAGATCCTAAAGGTCAATAAGGTACAATTCAAGCCGGAGATGTTTCAAGgtgaa
 aacaaaaacagaaacacaaacaaagaacactcgaataactctctgagttttcttgtttcaactcaatttga
 ctttgattgacagTGGATGACAGCAGGAAGAGGAATCATTTCATTCGAAATTTCCGGAAGAAGAAGTAAA
 CAATTGGTTTACAGCTTTGGATCAATCTCCCTTCCACTGAAAAAATgtataagatcaactctttgatataca
 tctcttctcagaaagagttccagcttgaatctctctgttctcttgattttgacagACTGAACCAAAAT
 ATAAGGAACATCGAGTTTAGACATTCTCCGAGCAGAAAGAAACCGGAGTTGAGGTCAAAGTCATAGCCG
 GAGATTCAATGGGAATCAAATCTCCAGTCTACACAAGAACACCAACAATGTTCTTACTTTTACCCTCA
 AGCCAGGATCTCAAACCCACCAAAACCGTTCCAGAAATCATGGACCGCTTTCGCTTACATTATAGAAGGCG
 ATGAAGGTGTTTTCGGTTCCCTGAACTCTTCCGCAATATCGGCCACCATGTTGTTGTGTTTTGGGCCAG
 GGGATTTAGTTAGCGTGTGGAACAAGTCAACTTCGAGGTCAATTGAGGTTTTTGTGATTGCAGGGGAAC
 CTATCGGCGAGCCTGTGGTTTCAAGTGTGGTCCCTTTGTGATGAATTCACAGGCGGAGATCGATATGGCTT
 TTGATGACTATCAGAATGCTAAGAACGGGTTTGAATGGCCAAGTGTAG

AT3G15640 (potential dTALE[*EGL3*] off-site)

tatatatacattctacacacacaaactactataccacgattttttttatgctaacattacctctagctaca
 gacagaagtagaaagaaacaaaaccatcaggtattacatttgaagttttgaattgtggaattctcaa
 cgaggttcgaggaacatgaatttgggttgctaactaaagcaacttgcttgctgtgtgcataactgga
 ctgaggaaaaaataatccttacatgtccatttcatcataatttcattttgtagtatatgcatcattgta
 tcattatttccacttgtatcactaattccacttgaaatgtcacacaatatagggtccccgacacgtgt
 gtactatgtagtttccacacgtaccggttaatatcgagattccaagtaaaaaagtgacc**aattc**ta**aatt**
agaatctctctaattctcttcccaagaagaatcatcactgaggaagaataaaacatttggggagtttc
ttcgtaatccttcaaggtgtgacctgagaatcacaacacaaaacaatctcaactcgaattcctttgac
tcatcatcatcaccATG****TGGAGAAGAATCGTCTCTCTCAGCTCAAACCCTAGCCGCCGATGTCGTCG
 CTGCTTCTCTCGTCGATCGATAGCCGCCACCACAGACCTGTGCGTTTTCTATCTCGCCGCCAATCGAT
 CAGCATTTCGGCATCTCTTTGCTTATCCCTCGTCTGTTTTCAGCTCCGATTCAGgtgattcatcgttgc
 tatagtttacagcattttcagatctcggctcaggactctacctgcgctggaattagattcagattcacta
 tgttttcttcacgcataattagaggatagactaaatctgtgtgttggtctcttagattggtgttgc
 attttgtttcatctctcaattggaaggtgaatgattgtttattacgattgtattttgcagTAGAGA
 CTCTGCGACGAAGAAGGTGGAGGATGTAATGCCCATTCGACTGGACACGAGAAGGAAGAGCTTGAAG
 CTGAACCTTGAGgtgattgcttatgtttactttatgtattgcaattttaagatctgagtatgattacaat

tgtgtttgtgatcatgacttcacttctcgataatTTTTctggtatagcaatgccaatggagtgatattt
 ctgcttgataatacacttcagttgaggaaaatttgattttgactttttcattgttttcttgtgatagGG
 AAGGAGGCTGGATGATATTGACTTCCCTGAAGGTCCTTTTGGAACTAAGGtttagataagtttcttttac
 cagagatatgtcgtgactcctccttgaatatcctaaccttgtttttactgcaaagtcatcagtgtttc
 taggcatgatatagttccttatgttcatcactaacacacgggcatacgtgaaacaaaccaaggacactta
 aactagctcgagtattatgaatggctttatgataaatgtagctgittcaatgatctcaattaactagaa
 taagtatgtatccttgaacttattgcttcttcaaaactaatatgagtcagctgggttactgacttata
 tcgatcaacttgcaattccttggctatgggttcataaattgctttagcaggattactatatcatgcagcaa
 tcactttgttatgtcgttttggagctaaatcatcttgatattacactggtcataatttctggatttttat
 tttttttgtgaataatgtaattagtatatttccctcaatcaagcactagcttgggggtgaatcatata
 ttggttttgggtgaattaaataactctaagctgtgtttgttaatttacgggtggctcactgcctatgat
 gcttttgtgcacgataatgacgaaaagattgctggatgtcacatacaacatgtgcaagtgtctgtagatc
 atggaaattcagttcctaacatccataaaaaacttgctccaaattactgaagatgttaggagtatatg
 actgttttgcatgctatgacaagacatttgattttttcattttcttgtgtcattttatcttttgagtga
 gcagtatcattttcttttggctacacaattggatacttattttttcattttcttgatacacaatttg
 gatgtcattttgtattgccttagGAAGCTCCTGCTATTGTGAAGTCTACTATGACAAGCGAATTGTGG
 GCTGCCCTGGTGGTGAAGGCGgtaagttcttcccttcttcaaaacttagtcgcatattatggacaatct
 ttaatcggaatcgtttgttgcattgaaatatacaagacacattattaccaatctaactttctcctga
 aaactgcagAGGACGAGCAGATGTTGTCTGGTCTGGCTGGGAGAAAGGAAAGTCTTTGAATGCCCGG
 TTTGCACTCAGTACTTTGAGgtaagtgaaagtcgtatcagatacttactcagtaaaataaatcgaccta
 gcctctcatttcatagtcataaggccagacgaatctgtcaaaggattagagcaagaactcattaagcttt
 cgaattcagaatattgaattggttgaataatgtatgcttgcagCTGGAAGTGGTTGGTCTGGTGGTCC
 TCCGGATGGTCACGGTGACGAAGACGATGAGCACCACCCTGA

AT4G27740 (potential dTALE[*EGL3*] off-site)

tatatacaatatacacacacacacacaaactgagaattagaatgtacatgtgaatgggctgggtgcagtaaatg
 cattattacgtgtaaatgtacatgtgaatgggctggtagcagtaaacgcataatctcaaaaccaccggg
 caccgatcgggtgcaagtctcagtcagaacgtgattggaccaagtctcgggtgataaagacacacagtc
 ccataccgcgtaaaataatgaccacacggatgggcccataatatacaaaatttgggtccttttaactct
 tgtttcacaagtgaattattcaaatgggcctttattatacacccataaaatggcccattagtccaacaac
 atacaataatgcaagtggctcggggctcaacgcactagaaaattccacaagaactaaacgataagaacaat
 aatcttcatagtcctcatccaatcttacaactttctgatttcttctacattcatcgaccaATGGC
 AGCAAACAAAACCTCTTCTACGTATTTTTGCAGAACTGCGAGAATCCACTAGCTCTCGGTGAAGATCT
 CATCTCCAAAAAATTTGTGgtaacaaaaaaaaaaaaactaaagaatggtatagctttacaaaaaaaaatc
 actttttcatttcttatcggatgattttctagagatgatcaagtctcttcataatctccattgggtata
 aaatgcaaaatgtttttgaatatggattatagGGAGCTTCAGGGCCAGCGTTTATGTTTTACACGCGA
 TGAACGTGGTGGTTGGACCGAAGATTGGGAGGAACTGATAACCGGATCGTATGTAGTGGCAGATGTGA
 TGTGCAGTAAGTGTGGTGGAGACGTTGGGTGGAAATATGTTGAGACATTCGATCTTAAACAGAGGTATA
 AAGAAGGCATGTTTGTTCATCGAGAAGCTCAAGTTGACCAAAAGATATTGA

KNATI (AT4G08150)

ccttgacgaattctatataccttagttcgtttttcttctcctcaaatatatctttttcaatttatttggtt
 tttctttgggtgcaacttcacctcaaaaaatttctctcttttttatattaatttgagttaggccttt
 ttgatttcatagatgagtcgtctagtcgtctggatttgatgtggttatagtccttacagagacctttgat
 tgaataagaacaaaagcaagaatacatatcctcttcatcttacacctccttttttatttttcta
 gggttttatttttttaatttattttttcttttgatttttatattctctctctctcaaatcttta
 ctcatctgggtATGGAAGAATACCAGCATGACAACAGCACCCTCCTCAAAGAGTAAGTTTCTTGTACT
 CTCCAATCTCTTCTTCCAACAAAAACGATAACACAAGTGATACCAACAACAACAACAATAATAATA
 GTAGCAATTATGGTCTCTGGTTACAATAATACTAACAACAACAATCATCACCACCAACACATGTTGTTTC
 CACATATGAGTCTCTTCTCCCTCAAACAACCGAAGATTGCTTCCGATCTGATCATGATCAACCCAACA
 ACAACAACAACCCACTCTGTTAAATCTGAAGCTAGCTCCTCAAGAAATCAATCATTACTCCATGTTAATGA
 GAGCCATCCACAATACTCAAGAAGCTAACAACAACAACAATGACAACGTAAGCGATGTTGAAGCCATGA
 AGGCTAAAATCATTGCTCATCCTCACTACTCTACCCTCTACAAGCTTACTTGGACTGCCAAAAGgttt
 ggttttattcttattggttatgttataatataactcttttttaaagatataatataatagttatgtgtg

acatatatatgtatggtcatgggtcatgggttatatatagATTGGAGCTCCACCTGATGTGGTTGATAGAA
TTACGGCGGCACGGCAAGACTTTGAGGCTCGACAACAGCGGTCAACACCGTCTGTCTCTGCCCTCTTA
GAGACCCGGAGTTAGATCAATTCATGgttcattactaaccaaagtgaattttaaattaaatgggtcgg
tctttgtgtgtatgtggttagaagtcttgatcttcttttttattttttggtatgaatagGAAGCATA
CTGTGACATGTTGGTTAAATATCGTGAGGAGCTAACAAAGGCCATTCAGGAAGCAATGGAGTTTATACG
TCGTATTGAATCTCAGCTTAGCATGTTGTGTGAGAGTCCCATTCACATCCTCAACAATCCTGGtcatg
tcatcaatctcatatacatatatgcacatgtaagaaatggctggagaaaaaaagtgtagtgga
aggtagggtcaatgatggattgcataaattggcatgtttctttggttcatggtaaattggttaacatt
gagatattaaaaaataatcatagtctacatacatctttcaaatgtaattaagattataagataa
ttaagataattgagatgcattattgttggaagtatgtaagatacaaatagggtgtgcatattgttctt
gttttccccattgtggaatccatccaaagatgggtggggcttattttgtttttttcttttactcgg
cccaaaagtactatcacaagaacttttggtgctgctcgtgaagatttatttgggacattatcttctcg
atcttcttgatgctaaacatcaaatctctactcggattaattagtctaatcataaaggtagatgacta
gaatgttggtaaatgactaaaccatttaacttcatgaaatataattatggtgaagcagATGGGAAGAGTG
ACAATATGGGATCATCAGACGAAGAACAAGAGAATAACAGCGGAGGGGAAACAGAATTACCGAAATAG
ACCCGAGGGCCGAAGATCGGGAACCTCAAGAACCATTGCTGAAGAAGTATAGTGGATACCTAAGCAGTT
TGAAGCAAGAACTATCCAAGAAGAAAAAGAAAGGTAAACTTCCTAAAGAAGCACGGCAGAAGCTTCTCA
CGTGGTGGGAGTTGCATTACAAGTGGCCATATCCTTCTgtacgtataaaattttctctcatctctttt
ccctcatattttttaaataatataactacaaattaagtgttttaaaaacctatctttaagctagcaattc
cctctgaagccgctatggttaggccaagtaggtacttgcatgaccgattttgtgtttaaactcttctgacg
gtttccaacaacatggttttaaatccgaatgctcacaacactagaattagtcactaggatacatactt
agtcatggtgttctaacccttaacataactatcttaacacgtatatcaatcacttttttttaaaaaat
gtaagttgtgatgaaatttttatccatataatttgaaacaatttaattttttttagtatttaattag
ttgaaatgaaataactaatctaaccgactgtaatgatagtgaaaaaacatgtttaaattgtgtaaat
ataaatgacatggttggtaaaaagataaataagacaacataatagttagtcttacttgatgtcaagaaaa
atgaagttagtcatttttacttagagaatcgttcaaaaacaacaaactttaggtcattttccatttttatg
tctagaatcatttaaaaaataatcttgcattatctctagcattgctttttatacattgattactacataa
atgaagatggttggtttgagttgtgtggcagGAGTCAGAGAAAGGTAGCGTTGGCGGAATCAACGGGGTT
AGATCAGAAACAAATCAACAATTGGTTCATAAACCAAAGAAAGCGTCACTGGAAACCATCTGAAGACAT
GCAGTTCATGGTGATGGATGGTCTGCAGCACC CGCACCACGCAGCTCTGTACATGGATGGTCATTACAT
GGGTGATGGACCTTATCGTCTCGGTCCATAAagacatccaaaagcttttagccacataataacaacctc
gttgccttctgtttacaactcatgttttgaattcccctacatcagtttgctactttagcttctttgt
ttcacgccttttgaatgtcttatgctggttcggggagtttgagacttcttagtcagaaatatcccttc
attttattttctctttttttcggtgtattttgtttttttgttttctcactaatgtttttatttaataat
gtgtgaaaggaaaaatgctattctgaaaagggtgttgctccatttttttattttatagggcatgaaatgta
tggttttgaaagacaaaacccaaatgatctaattggattcttcatgacctttccacaagctaaaacaaa
atccacaaaatgtaaaacatcgatagttgataagtctcaaaacctctttgcttcgaacacggtttcag
gcataatattatgacttatgcaaagcattacatcataatacataggtcatggaatcactaccagaaaa
aaaaaactattgtcgaagattaagttaaagattctccctttgcaccactcgtcactctgatgcgga
tccaaattaaaccaaaccgggaccataccaaatcagccggtacatttatcaaacggattgatgaagagc
atthttggag

AT1G77980 (potential dTALE[*KNATI*] off-site)

ATGGGTGAGTGAATTTGGAGATAAAAAGAAATAGAAAACACAACGAATCGACAAGTTACATTTCTCCAAA
AGACGAAATGGTCTTATAAAGAAAGCTTATGAATTGTCCATTCTTTGTGACATCGACATTGCTCTCCTC
ATGTTCTCTCCTCTGATCGCTTAGCTCTTTTCGGGCAAAAACAAGgtttttgtgtgtgtgtctta
taacattttctttcgggttaattgtacaaaacatggaaacattcatttgatctatatttattcgttaaac
tcttgtcagGATTGAAGACGTTTTCTCTAGATACATCAATCTTTCTGATCAAGAACGAGAGAAGtaagc
tgacctttcaaggctttaatagattctcctaaagatattccactcatttggttcttaattttctctttca
ttttgttttagTGCTCTAGTATTCCCCGATCAAAGTCGACGCCGgtatgtatcttactcttattccgat
aaataattaacatataataaaccgtatatttataatcatcaaaaatctaattgagtttttcatattttga
caactctatcttgcagAGATTTCCAGAGCAAAGAGgtatgtaattagattaagcatattgtgcagtggt
catgtgtgta~~tatat~~cttagctccttttcaatgaaggacaattatacatttcagattattgatctcga
aattttcttgcacgtgacagTATCTACTCCGACTTTGCAGCAACTCAAGGCTGAGAATGATATTGCT
CTTCAACTTACCAAgctctcaaatctttctcactttcattattatataactacgttattttgcaataat

tcggaagttgattttctttatctttctcttattggatgattttaaaaagCCCTACAGCTATCAACTCC
GACGTCGAGgtgaatcttctatctttttctcttagctaataactttactttatgttatttactctctaa
tcatcaatcttagGAACTTGAGCATGAAGTTTATAAGTTACAACAACAACCTTCTCATGGCAGAGGAAGA
ACTAAGgtcttctttttataacaaaaacacagtgaatgtgatacaatgtatgaaaatttatgtaacta
atacttatatccaaattatataatataattagGAAATACGAACCAGATCCAATCAGGTTTACAACAATGG
AAGAGTACGAAACTTGTGAGAAGCAGTCATGGACACCTTAACACGTGTCAATCAACGACGGgtcaatā
atthaaactccccctttatctttttcttaagcataatgtattcaacaatatctattagtaactaaataatt
tcttgagacgcagGAACATATATTGAGTCAAGACCAATTATCTTCATATGAAGCATCTGCTTTACAACA
ACAGCAAAGCATGGGTGGGCCTTTTGGAAAACGACGTCGTTGGAGGATGGCTGACTGAAAATGGGCCTAA
CGAAGCCCATTATTTCGACGCATCGGCACATTCAGCAATGTATGAACTTTATTGCAAGGAAGTAGCTC
AAGCTCGAACCAAAAACAACATTATGGGTGAATCCAATGTGTCAAAATCATAACGGTGACATGTTTCAAGA
ATGGGCCCAAGCCTATAATTCTACTACGGCCATAACCCTTCGACTCTATTTCCCTCCTATGCAGCATCA
Ggttctatattttctctttttccgaaataaattaaaattgtttcactaattttttttttgcttttaaa
tatgttagCATGGACTGGTGGTTGATCCCAATATAGAAGAAATTGAAATACCGGTTATGAAGAAGGATG
CACAAGCAGACCACGAGGTCTCCGACTATGATATAAGAATGCCTCAGCTCAGTAGCCAATAA

Supporting figures - legends

Fig. S1 Codon optimized *TALE* genes of *avrBs3* and *avrBs4* trigger an HR in *Bs3* and *Bs4C* resistant *Capsicum* species, respectively. *In planta* functional analysis of codon-optimized (*) *TALE* genes. The 35S promoter-driven *TALE* genes indicated on the left and right sides of leaves were delivered via *A. tumefaciens* into the four *Capsicum* genotypes shown. The allelic configuration at the *Bs3* locus (alleles: *Bs3* [recognition of AvrBs3], *Bs3-E* [recognition of AvrBs3 Δ rep16]) and the *Bs4C* locus (alleles: *Bs4C* [recognition of AvrBs4], *bs4C* [no recognition of AvrBs3, AvrBs3 Δ rep16 and AvrBs4]) is indicated for each plant genotype. Dashed lines mark the inoculated areas. Two days after infiltration, the leaves were cleared in ethanol to visualize the hypersensitive response (dark areas).

Fig. S2 TALE non RVD residues have no measurable influence on TALE specificity. (A) Graphical representation of chimeric TALEs and corresponding *uidA* reporter constructs. TALEs are displayed as gray bars. Black and white ovals represent the repeat units of AvrBs3 and AvrBs4, respectively. Red and yellow vertical lines within the ovals represent the Repeat-variable diresidues (RVDs) of AvrBs3 and AvrBs4, respectively. The chimeric TALE AvrBs3RVD4 contains the RVDs of AvrBs4 but is otherwise identical to AvrBs3. Reciprocally, the chimeric TALE AvrBs4RVD3 contains the RVDs of AvrBs3 but is otherwise identical to AvrBs4. TALE NLSs and AD are displayed as white diamonds and white arrowheads, respectively. Asterisks (*) indicate that proteins are translated from genes with optimized codon usage. (B) Graphical representation of *Bs3* promoter derivatives. Gray arrows represent the promoter of the pepper *Bs3* gene (*Bs3_P*). Red and yellow arrowheads depict the *UPT*_{AvrBs3} and the *UPT*_{AvrBs4} boxes, respectively. The white area within the *Bs3_PUPT*_{AvrBs4} Δ *UPT*_{AvrBs3} promoter indicates the deleted *UPT*_{AvrBs3} box (Δ). Blue boxes represent the *uidA* reporter gene, encoding the β -glucuronidase (GUS) reporter protein. The nucleotide sequences of the promoters are provided as supporting information (SI Text). (C) Repeat-variable diresidues of AvrBs3 and AvrBs4 retain promoter activation specificity regardless of the repeat array backbone. Depicted *uidA* reporter constructs were delivered into *N. benthamiana* leaves in combination with the depicted 35S-driven *TALE* genes. Asterisks (*) indicate *TALE* genes with optimized codon usage. Leaf discs were stained with 5-bromo-4-chloro-3-indolyl- β -D-glucuronic acid, cyclohexylammonium salt (X-Gluc) to visualize activity of the GUS reporter. Samples were taken at 40 hpi.

(D) Quantitative analysis of promoter activation via wildtype and chimeric TALEs. GUS activity (pmol 4-MU/min/g protein) was determined 27 hours after *A. tumefaciens*-mediated co-delivery of the depicted *uidA* reporter constructs and the 35S-driven *TALE* gene shown below columns. Error bars denote standard deviations. Asterisks (*) indicate *TALE* genes with optimized codon usage.

Fig. S3 *In planta* functional analysis of chimeric *TALE* genes. The 35S promoter-

driven *TALE* genes indicated on the left and right side of the leaves were delivered via *A. tumefaciens* into leaves of the *C. annuum* cultivars ECW and ECW-30R or the *C. pubescens* accession numbers PI 585270 and PI 235047. The allelic configuration of the plant genotypes at the *Bs3* locus (alleles: *Bs3* or *Bs3-E*) and the *Bs4* locus (alleles: *Bs4C* or *bs4C*) locus is shown below the leaves. Dashed lines mark the inoculated areas. Two days after inoculation, the leaves were harvested and cleared with ethanol to visualize the HR (dark areas). Asterisks (*) indicate codon-optimized *TALE* genes.

Fig. S4 Graphical display of *Bs4S* promoter::*uidA* reporter constructs and matching dTALEs. Corresponding experimental data; see Fig. 1. (A) Arrows represent the promoters of the tomato *Bs4S* (yellow) and the pepper *Bs3* gene (gray). Cyan and red arrowheads depict the $UPT_{dTALe[Bs4S]}$ and UPT_{AvrBs3} boxes, respectively. White vertical bars within the cyan arrowheads represent deletions in the $UPT_{dTALe[Bs4S]}$ box. Blue boxes represent the *uidA* reporter gene, encoding the β -glucuronidase (GUS) protein. The nucleotide sequences of the promoters are provided as supporting information (SI Text). (B) Graphical display of the TALEs used in this experiment. The TALEs are displayed as gray bars. Repeat units, NLSs and AD are depicted as black ovals, white diamonds and white arrowheads, respectively. Cyan and red colored vertical lines within the ovals represent the repeat-variable diresidues of dTALE[*Bs4S*] and AvrBs3, respectively. Repeat unit deletions in dTALE[*Bs4S*] Δ 6-8 and dTALE[*Bs4S*] Δ 10-12 are indicated. Asterisks (*) indicate that the AvrBs3 is translated from a codon optimized gene.

Fig. S5 dTALEs activate promoters of user defined target genes in *A. thaliana*. (A) Graphical display of promoter::*uidA* reporter constructs used in this experiment. Yellow arrows represent the promoter of the tomato *Bs4S* gene. Green and purple arrowheads depict the inserted $UPT_{dTALe[EGL3]}$ and $UPT_{dTALe[KNAT1]}$ boxes, respectively. Blue boxes represent the *uidA* reporter gene, encoding the β -glucuronidase (GUS) protein. The nucleotide sequences of the promoters are provided as supporting information (SI Text). (B) Graphical display of engineered dTALEs. The dTALEs are displayed as gray bars. Repeat units, NLSs and AD are displayed as black ovals, white diamonds and white arrowheads, respectively. Green and purple colored vertical lines within the ovals represent the repeat-variable diresidues of dTALE[*EGL3*] and dTALE[*KNAT1*] which direct binding to promoter elements of the *Arabidopsis EGL3* and *KNAT1* genes, respectively. Although not displayed graphically, both dTALEs contain a C-terminal GFP-tag that facilitates analysis of expression *in vivo*. (C) *In planta* functional analysis of different TALE-promoter combinations. *uidA* reporter constructs under transcriptional control of the promoters indicated at left were delivered via *A. tumefaciens* into *N. benthamiana* leaves in combination with the 35S-promoter-driven *TALE* genes shown above leaf discs. GUS assays were carried out 40 hpi as described in Fig 1. (D) Analysis of potential dTALE[*EGL3*] and dTALE[*KNAT1*] off-targets in the *Arabidopsis* genome. Semi-quantitative reverse-transcription PCR (RT-PCR) was carried out on RNA from the *Arabidopsis* ecotype Columbia (Col-0) and corresponding lines containing the 35S-promoter-driven dTALE[*EGL3*], or the dTALE[*KNAT1*] transgene. RNA was

isolated from leaves and used for cDNA synthesis. Semi-quantitative RT-PCR was performed with locus-specific primers (see SI Text). The constitutively expressed gene *actin2* served as an internal normalization control. The boxes targeted by dTALE[*EGL3*] and dTALE[*KNAT1*] are shown on the right side and are highlighted with green and purple frames, respectively. Sequences of potential off-targets are displayed below the two distinct, user-defined boxes. Differences between the targeted box and the potential off-targets are displayed in boldface red font. Details on the genomic context of the targeted boxes and potential off targets are provided in SI Text.

Fig. S6 Graphical display of AvrBs3 derivatives with additional repeat units and corresponding *uidA* reporter constructs. Corresponding experimental data; see Fig. S8 and Fig. 4. (A) Arrows represent the promoters of the tomato *Bs4S* (yellow) and the rice *Xa27* gene (gray). Arrowheads depict *UPT* boxes. AvrBs3-targeted *UPT* boxes from the pepper *Bs3* and *UPA20* are displayed in red color. *Bs3*- and *UPA20*-derived promoter regions 5' of their UPT_{AvrBs3} boxes are displayed in green and pink color, respectively. The $UPT_{AvrXa27}$ box from the rice *Xa27* gene is displayed in white color. A blue box represents the *uidA* reporter gene, encoding the β -glucuronidase (GUS) protein. The nucleotide sequences of the promoters are provided as supporting information (SI Text). (B) Graphical display of TALEs used in this experiment. The TALEs are displayed as gray bars. Repeat units, NLSs and AD are displayed as ovals, white diamonds and white arrowheads, respectively. Black and purple colored ovals represent the repeat units of AvrBs3 and AvrXa27, respectively. Red and white vertical lines represent RVDs of the TALEs AvrBs3 and AvrXa27, respectively. Green and pink vertical lines represent C-terminal RVDs that are specific to AvrBs3+4R_{*Bs3*}* and AvrBs3+4R_{*UPA20*}*. (C) Alignment of RVDs and *UPT*-boxes of the depicted TALEs and promoter-reporter constructs. RVDs are shown as upper case letters using the single-letter code. Nucleotides are shown as lower case letters. Horizontal boxes mark repeat units, corresponding RVDs and aligned nucleotides. Numbers mark TALE repeat units and aligned nucleotides. Lowercase red letters represent nucleotides of the naturally occurring *UPT* boxes present in the pepper *Bs3* ($UPT_{AvrBs3Bs3P}$) and the *UPA20* ($UPT_{AvrBs3UPA20P}$) promoters. Lowercase green and pink letters represent nucleotides that are located 3' of the UPT_{AvrBs3} boxes in the *Bs3* and *UPA20* promoters. Nucleotides that differ between the *Bs3* and *UPA20* promoter are underlined. Uppercase red letters represent RVDs of the TALE AvrBs3. Green and pink big letters represent C-terminal RVDs that are specific to the dTALEs AvrBs3+4R_{*Bs3*}* and AvrBs3+4R_{*UPA20*}* and matching nucleotides of the *Bs3* and *UPA20* promoters. (D) AvrBs3 derivatives with additional repeat units discriminate between *UPA20*- and *Bs3*-promoter derived *UPT* boxes with extensive 5' homology. *uidA* reporter constructs under transcriptional control of the promoters shown at left were delivered via *A. tumefaciens* into *N. benthamiana* leaves in combination with the 35S-promoter-driven TALE genes indicated above leaf discs. GUS assays were carried out 40 hpi as described in Fig 1.

Fig. S7 Graphical display of AvrBs3 derivatives with three NK repeat units and corresponding *uidA* reporter constructs. Corresponding experimental data; see

Fig. 5. (A) Gray arrows represent the pepper *Bs3* promoter. The red arrowhead with three tandem arranged A nucleotides ('AAA') depicts the wildtype UPT_{AvrBs3} box of the *Bs3* promoter. *Bs3*-promoter derivatives in which three tandem arranged A nucleotides of the UPT_{AvrBs3} box are replaced by three C, G or T nucleotides are depicted ('CCC', 'GGG' and 'TTT'). The nucleotide sequences of the promoters are provided as supporting information (SI Text). Blue boxes represent the *uidA* reporter gene, encoding the β -glucuronidase (GUS) protein. (B) Graphical display of the TALEs used in this experiment. The TALEs are displayed as gray bars. Repeat units, NLSs and AD are depicted as black ovals, white diamonds and white arrowheads, respectively. Red vertical lines within black ovals represent the repeat-variable diresidues of AvrBs3. Three tandem-arranged NI type RVDs that are present in repeat unit residues 5-7 of the wildtype AvrBs3 protein are depicted. Yellow vertical lines within black ovals represent RVD residues within the AvrBs3 scaffold that were changed to NK-type residues. Asterisks (*) indicate that TALEs are translated from codon optimized genes.

Fig. S8 The TALE RVD NK mediates specific targeting of G nucleotides. *In planta* functional analysis of different TALE-promoter combinations. *uidA* reporter constructs under transcriptional control of the promoters indicated at left were co-delivered via *A. tumefaciens* into *N. benthamiana* leaves with the 35S-promoter-driven *TALE* genes indicated above leaf discs. GUS assays were carried out 40 hpi as described in Fig 1. Asterisks (*) indicate *TALE* genes with optimized codon usage and corresponding gene products.

2.4 Manuskript 2

Morbitzer, R., Elsaesser, J., Hausner, J., und Lahaye, T. (2011). Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Res.* 39, 5790-5799.

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

Robert Morbitzer¹, Janett Elsaesser¹, Jens Hausner² and Thomas Lahaye^{1,*}

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

Received February 3, 2011; Revised February 28, 2011; Accepted March 1, 2011

ABSTRACT

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

INTRODUCTION

DNA binding domains that can be tailored to interact with user-defined DNA sequences are crucial tools for molecular biology (1). Bacterial transcription activator-like effector proteins (TALEs) from the bacterial pathogen *Xanthomonas* target DNA via a novel type of DNA binding domain that is composed of tandem-arranged 33–35 amino acid repeat-modules, with each

repeat binding to one base (2,3). Base preferences of individual repeats are specified by residues 12 and 13, known as the repeat variable diresidues (RVDs), that determine preferential pairing with A (NI), C (HD), G (NK) and T (NG) nucleotides, respectively (2–5). In principle, the use of this TALE code facilitates the assembly of repeat arrays that bind to any desired DNA sequence that is preceded by a T nucleotide (2,6,7). Recent studies have also shown that TALE repeats function as sequence specific targeting modules not only in the context of a transcription factor but also when fused to a *FokI* nuclease domain (4,8–10). This suggests that the TALE DNA binding domain enables applications comparable to zinc finger (ZF) technology (11). ZFs that are assembled into an array are known to influence the DNA specificity of adjacent ZFs (12) and, due to this context dependency, ZF arrays of desired DNA specificity require experimental validation. In contrast, there is no evidence so far that base-preferences of TALE repeats are context dependent and recent studies have demonstrated that *in vitro* assembled repeat arrays target the pre-defined nucleotide sequences (4,5,8,10,13).

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

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

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2011. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.5>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

MATERIAL AND METHODS

Vector construction

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

Cut-ligation cloning protocol

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

In planta analysis

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

RESULTS

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

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

We aimed to establish a toolkit that enables TIIS-mediated assembly of genes encoding 20 or more TALE repeats. The length of the corresponding TALE target sites should facilitate specific targeting of a unique DNA sequence even in the context of a highly complex eukaryotic genome. To do so, we PCR-amplified and cloned DNA fragments encoding individual repeats (herein referred to as repeat-modules) from the *TALE* gene *avrBs3* into a modified pUC57 cloning vector ([Supplementary Figure S1](#)). To avoid vector-derived *BsaI* cleavage products that potentially interfere with the envisaged multi-fragment ligation, we removed the *BsaI* recognition site from pUC57 by site-directed mutagenesis, yielding pUC57 Δ *BsaI*. The cloned repeat-modules served as initial building blocks for the envisaged repeat assembly. The primers introduced *BsaI* sites at both termini of the repeat-module in such a way that both recognition sites are cleaved off from the repeat-module (Figure 1a). Initially we cloned 10 distinct repeat-modules each producing a different combination of terminal overhangs upon *BsaI* cleavage. The overhangs were designed so that each repeat-module would ligate only to specific repeat-modules in such a way as to generate a pre-defined linear repeat array. The first repeat-module of an array will ligate specifically to the 5'-end of the second repeat-module. The 3'-end of the second repeat-module will ligate specifically to the 5'-end of the third repeat-module and so on. Each repeat-module array consists of: (i) a 5'-adaptor repeat-module; (ii) a variable number of core repeat-modules; and (iii) a 3'-adaptor repeat-module. The position of each repeat-module within an array is defined exclusively by the given overlap. To get a sufficient number of different overlaps for the distinct repeat-modules, we made use of the degeneracy of the genetic code and incorporated different codons for identical amino acids at the repeat-module fusion points. In addition, the fusion point between the repeat-modules was varied ([Supplementary Figure S1](#)). Thus not every repeat-module does encode a complete 34-amino acid repeat. However, each repeat-module encodes one pair of RVDs and thus determines the base preference of the given repeat. The overlaps were designed in such a way that correctly assembled repeat-modules

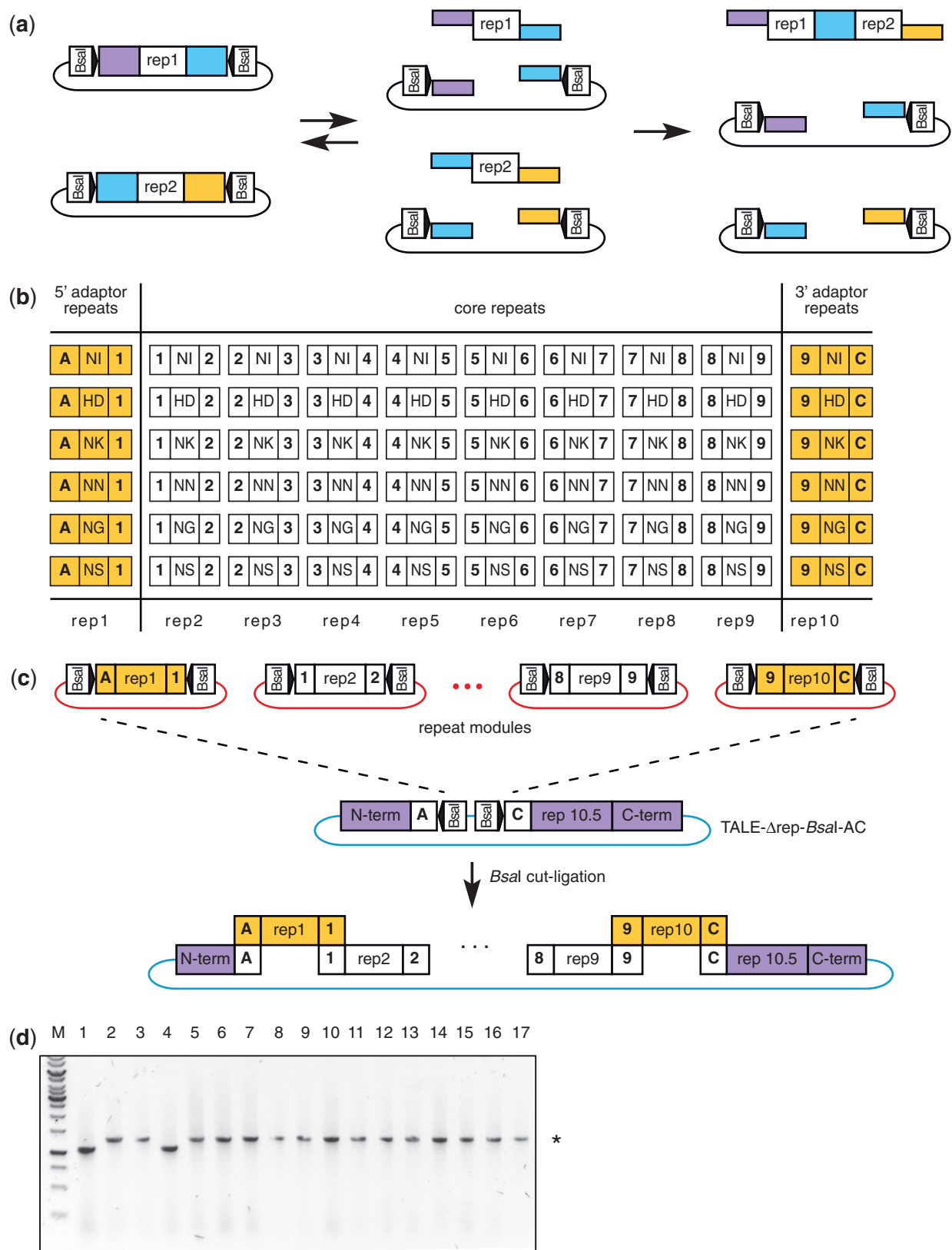


Figure 1. Assembly of dTALE arrays with 10 repeat-modules via *BsaI* cut-ligation. (a) The cut-ligation concept is shown for two representative repeat-modules that are displayed as white boxes (rep1 and rep2). *BsaI* recognition sites are shown as pentagons with black arrowheads pointing to the cleavage site. Coloured boxes represent *BsaI* cleavage sites and colour identity indicates ends with compatible overlaps. The line connecting both *BsaI* sites represents the vector backbone. *BsaI* cleavage releases repeat-modules (left side) and creates distinct overlaps at the 5'- and 3'-ends on each repeat-module. Ligation of two repeat-modules results in ligation products that lack *BsaI* recognition sites (far right side). By contrast, re-ligation of repeat-modules into their donor vectors results in plasmids that still contain *BsaI* recognition sites. (b) Repeat-modules that are used for dTALE

4 *Nucleic Acids Research*, 2011

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

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

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

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

Given that 10 repeats could be assembled efficiently in a *BsaI* cut-ligation (Figure 1d), we decided to generate repeat arrays by ligation of two *BsaI*-generated sub-arrays into a repeat-deficient *TALE* gene by a subsequent cut-ligation. The second cut-ligation is carried out with the type IIS enzyme *BpiI* that, like *BsaI*, produces 4-bp overhangs. For simplicity, we refer to the individual repeat-modules that are flanked by *BsaI* recognition sites as level 1 modules and to corresponding sub-arrays as level 2 modules. We generated two distinct level 2 modules containing 10 and 7 level 1 modules, respectively. Both level 2 modules are assembled with mostly identical core repeat-modules but differ in their 5'- and 3'-terminal adaptor modules. Therefore in addition to A and C, we defined a new B overlap and generated 5' and 3' adaptor modules that connect both level 2 modules (Figure 2a and b). For the assembly of these level 2 modules, we generated two pUC57 derivatives (pUC57-AB-DEST; pUC57-BC-DEST), herein referred to as level 2 destination vectors in which level 1 modules can be assembled by *BsaI* cut-ligation and from which level 2 modules can be subsequently released by *BpiI* cleavage. In level 2 destination vectors, *BsaI* and *BpiI* are positioned in inverse orientation relative to each other but create identical overlaps (Figure 2b). The two types of level 2 modules (AB and BC level 2 modules, see Figure 2c) produce distinct overhangs when being released by *BpiI* cleavage, which facilitates specific ligation of the two level 2 modules to each other. For the assembly of two level 2 modules into a functional *dTALE* gene by *BpiI*-mediated cut-ligation, we generated a construct containing a repeat-deprived *TALE* gene, herein referred to as level 3 destination vector. The level 3 destination vector is basically identical to the above-described repeat-deficient *TALE* gene construct (TALE- Δ rep-*BsaI*-AC) but contains two central *BpiI* instead of *BsaI* sites (TALE- Δ rep-*BpiI*-AC). Using a *BpiI* cut-ligation, two distinct level 2 modules are fused into a level 3 destination vector to encode a functional *dTALE* gene (Figure 2c). The level 3 destination vector also encodes the last half repeat that defines the C-terminal end of each *TALE* repeat array. In order to have the possibility also to select desired RVDs for this terminal half-repeat, we generated six different level 3 destination vectors that encode the RVDs NI (A), HD (C), NK (G), NN (G/A), NG (T) and NS (A/C/G/T).

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

Figure 1. Continued

assembly are represented as boxes. 5' and 3' adaptor modules are shown in yellow. Core repeats are shown in white. The position (rep1–rep10), RVD (NI, HD, NK, NN, NG and NS) and overlaps (boldface font) of every module is written in the box. (c) Assembly of a *dTALE* gene with 10.5 repeats via *BsaI* cut-ligation. Adaptor and core repeats are shown as yellow and white boxes, respectively. Boldface font indicates the unique overlaps of modules created by *BsaI* cleavage. Red dots indicate cloned core repeats 3–8 that are not displayed. Purple boxes represent the regions encoding the N- and C-terminal parts of the *TALE* (N- and C-term) including the last half repeat (rep 10.5). Lines connecting the boxes represent the vector backbones that mediate either ampicillin (red line) or kanamycin (blue line) resistance. The *BsaI* cut-ligation assembles repeat-modules 1–10 into the TALE- Δ rep-*BsaI*-AC vector. The assembled *dTALE* gene does not contain any *BsaI* recognition sites. Black dots indicate repeat modules of the assembled array (3–8) that are not displayed. (d) PCR was used to analyse colonies obtained in a representative *BsaI* cut-ligation. PCR fragments from 17 colonies (1–17) were separated on a 1% agarose gel and stained with ethidium bromide. The expected size of the PCR product was 1.3 kb and is marked with an asterisk. M: GeneRuler 1 kb DNA Ladder (Fermentas).

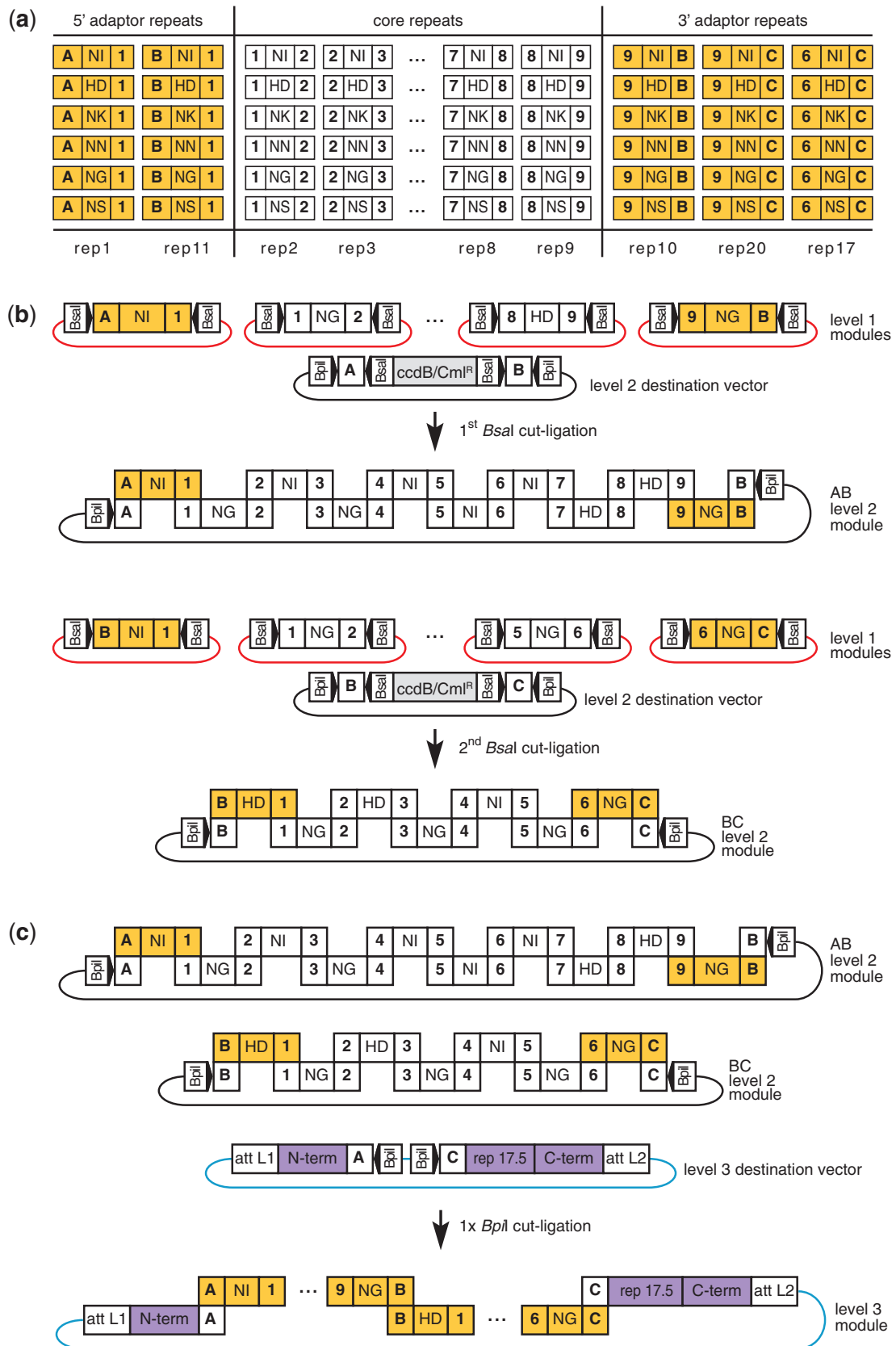


Figure 2. Assembly of a *dTALE* gene with a 17.5 repeat array by two subsequent cut-ligations. If not specified, the shapes and lines are as described in the legend of Figure 1. (a) Adaptor and core repeat-modules that are used for *dTALE* assembly. Distinct overlaps of repeat modules are defined in bold font (A, B, C, 1–9). (b) Two *BsaI* cut-ligation facilitate assembly of two distinct level 2 modules that contain 7 and 10 repeat-modules, respectively. *BsaI* and *BpiI* recognition sites are shown as pentagons with black arrowheads pointing to their cleavage sites. Level 2 destination

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

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

Functional analysis of dTALEs with target-optimized RVD composition

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

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

containing the highly similar *Bs3*_P*UPT*_{AvrBs3} box. In contrast, AvrBs3 produced GUS activity with both promoter constructs (Figure 4). Thus the dTALE gene produced by modular cloning was functional *in vivo*. Furthermore the target-adapted RVD composition enabled us to generate a dTALE that, in contrast to AvrBs3, discriminated between two highly similar target sequences.

A modified expression vector simplifies generation of dTALE expression constructs

In the above described approach, functional analysis of a given dTALE gene requires recombination-based transfer from the entry into a desired expression vector. Implementation of Gateway technology provides a high level of flexibility since assembled dTALE genes can be transferred into any Gateway-compatible expression vector. However, in principle, level 2 modules that encode repeat sub-arrays can also be assembled directly by *Bpi*I cut-ligation to a functional dTALE gene within the framework of a desired expression vector. This approach allows assembly in two rather than three steps and avoids the rather costly Gateway cloning step. Direct TIIs mediated cloning requires suitable expression vectors that must be devoid of recognition sites for the IIs enzyme that is used in the assembly of level 2 modules. Inspection of the *in planta* expression vector pGWB5 (17961 bp) revealed nine *Bpi*I, and two *Bsa*I sites. We decided to remove both, *Bpi*I and *Bsa*I recognition sites since this would enable us to use this expression vector in cut-ligations with *Bpi*I or *Bsa*I.

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

Figure 2. Continued

vectors contain two pairs of *Bpi*I and *Bsa*I sites producing identical overlaps (A–C). The grey box represents a Gateway cassette with *ccdB* gene and chloramphenicol (Cm^R) resistance marker. The line connecting the *Bsa*I or *Bpi*I sites represents the vector backbone that mediates either ampicillin (red line) or spectinomycin resistance (black line). (e) A *Bpi*I cut-ligation facilitates assembly of two level 2 modules into a functional dTALE gene. White rectangles denote the two attachment sites (*att*L1 and *att*L2) used for LR recombination. Lines connecting the boxes represent the vector backbones that mediate either spectinomycin (black line) or kanamycin (blue line) resistance. *Bpi*I cleavage creates overlaps in the level 3 destination vectors (A and C) that are complementary to those at the 5'- and 3'-end of the AB and BC level 2 modules, respectively. The generated level 3 module, which encodes a functional dTALE with 17.5 repeats does not contain *Bpi*I recognition sites.

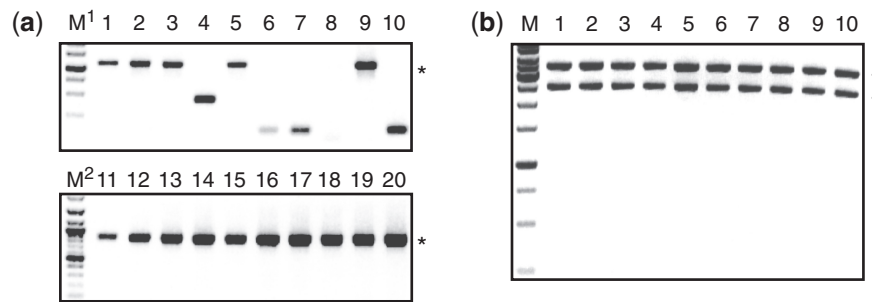


Figure 3. Cut-ligation efficiency for the generation of level 2 and modules. (a) Colony PCR was used to analyse colonies obtained in the cloning of AB and BC level 2 modules. PCR fragments from 10 colonies each (1–10 [AB] and 11–20 [BC]) were separated on a 1% agarose gel and stained with ethidium bromide. The expected sizes for the two distinct level 2 modules are 1.2 kb [AB] and 0.9 kb [BC], respectively (asterisk). M¹ and M²: GeneRuler 1 kb and 100 bp DNA ladder from fermentas. (b) Full-length *dTALE* genes that were generated by *BpiI* cut-ligation were analysed by *StuI-AgeI* double digest. Fragments obtained from 10 distinct colonies (1–10) were separated on a 1% agarose gel and stained with ethidium bromide. The expected sizes were 2.5 and 3.5 kb (asterisks). M: GeneRuler 1 kb DNA ladder from fermentas.

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

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

DISCUSSION

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

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

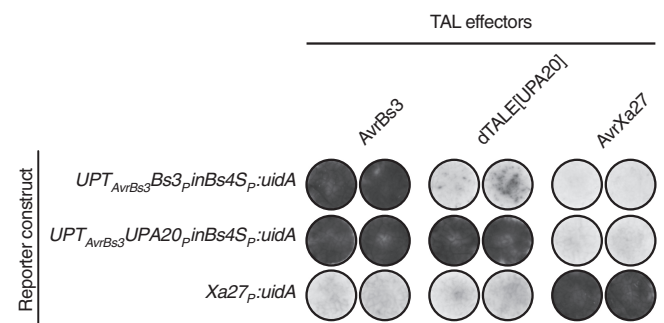


Figure 4. A *dTALE* with optimized RVD composition discriminates between the closely related sequences present in pepper *UPA20* and *Bs3* promoter. The *uidA* reporter constructs under transcriptional control of the promoters shown at left were delivered via *A. tumefaciens* into *N. benthamiana* leaves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

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

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

Our assembly kit allows incorporation of repeats with six distinct RVDs including NK that was recently found to

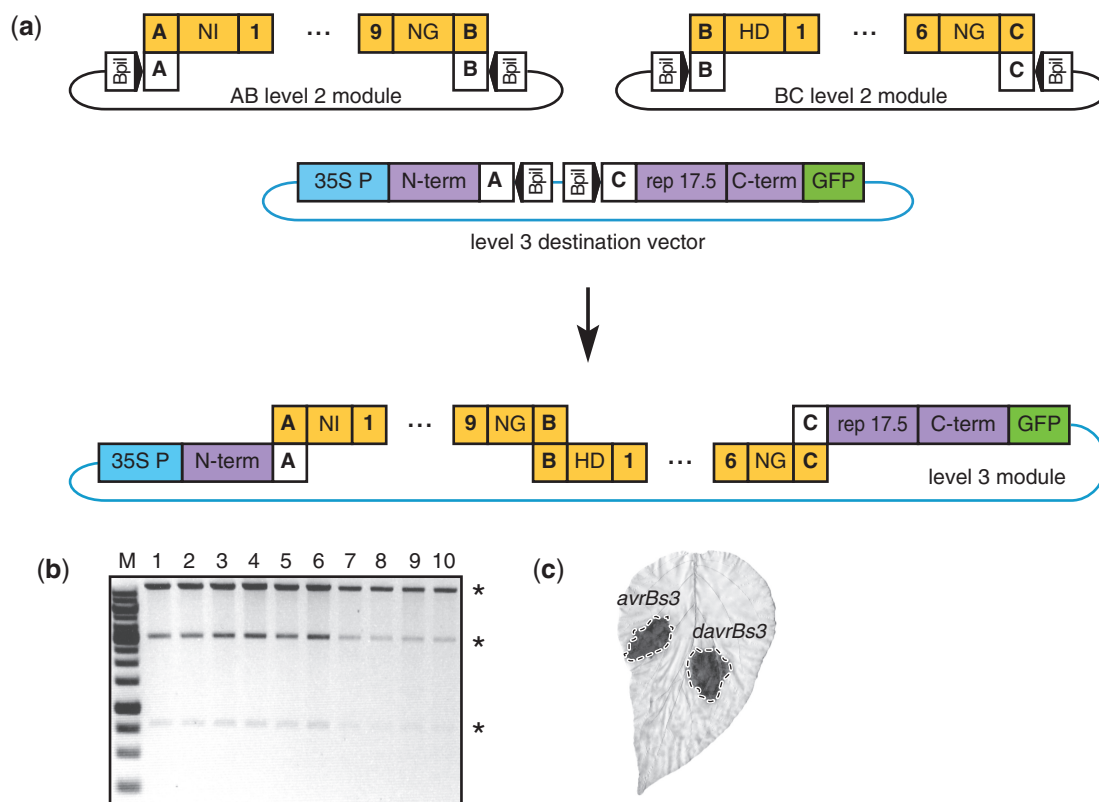


Figure 5. Direct assembly of two *TALE* repeat sub-arrays into a modified plant expression vector that lacks *BpI* recognition sites. If not specified, the shapes and lines are as described in the legend of Figure 1. (a) A *BpI* cut ligation facilitates direct assembly of two level 2 modules into an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

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

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

Design of *TALE* repeat arrays—how to maximize target specificity

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

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

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

Alternative approaches for assembly of TALE repeats

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

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are indebted to D. Horvath and O. de Lange for helpful comments on the manuscript. We thank S. Marillonnet for helpful comments on the assembly approach. We thank T. Strauß, M. S. Füßl and K. H. Braun for technical support; and T. Nakagawa (Shimane University, Matsue, Japan) for providing the pGWB vector series. R.M. designed method and research and performed research; J.E. designed and performed research; J.H. performed research; T.L. designed method and research and wrote the article.

FUNDING

The 2Blades Foundation; Deutsche Forschungsgemeinschaft (SPP1212, SFB 648 and LA1338/2-2 to T.L.). Funding for open access charge: The 2Blades Foundation.

Conflict of interest statement. None declared.

REFERENCES

- Klug,A. (2010) The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annu. Rev. Biochem.*, **79**, 213–231.
- Boch,J., Scholze,H., Schornack,S., Landgraf,A., Hahn,S., Kay,S., Lahaye,T., Nickstadt,A. and Bonas,U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509–1512.
- Moscou,M.J. and Bogdanove,A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, **326**, 1501.
- Miller,J.C., Tan,S., Qiao,G., Barlow,K.A., Wang,J., Xia,D.F., Meng,X., Paschon,D.E., Leung,E., Hinkley,S.J. *et al.* (2011) A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.*, **29**, 143–148.
- Morbitzer,R., Römer,P., Boch,J. and Lahaye,T. (2010) Regulation of selected genome loci using de novo engineered transcription activator-like effector (TALE)-type transcription factors. *Proc. Natl Acad. Sci. USA*, **107**, 21617–21622.
- Römer,P., Strauss,T., Hahn,S., Scholze,H., Morbitzer,R., Grau,J., Bonas,U. and Lahaye,T. (2009) Recognition of AvrBs3-like proteins is mediated by specific binding to promoters of matching pepper *Bs3* alleles. *Plant Physiol.*, **150**, 1697–1712.
- Römer,P., Recht,S., Strauß,T., Elsaesser,J., Schornack,S., Boch,J., Wang,S. and Lahaye,T. (2010) Promoter elements of rice susceptibility genes are bound and activated by specific TAL effectors from the bacterial blight pathogen, *Xanthomonas oryzae* pv. *oryzae*. *New Phytol.*, **187**, 1048–1057.
- Christian,M., Cermak,T., Doyle,E.L., Schmidt,C., Zhang,F., Hummel,A., Bogdanove,A.J. and Voytas,D.F. (2010) TAL effector nucleases create targeted DNA double-strand breaks. *Genetics*, **186**, 757–761.
- Li,T., Huang,S., Jiang,W.Z., Wright,D., Spalding,M.H., Weeks,D.P. and Yang,B. (2011) TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.*, **39**, 359–372.
- Mahfouz,M.M., Li,L., Shamimuzzaman,M., Wibowo,A., Fang,X. and Zhu,J.-K. (2011) De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc. Natl Acad. Sci. USA*, **108**, 2623–2628.
- Urnov,F.D., Rebar,E.J., Holmes,M.C. and Gregory,H.S.Z.D. (2010) Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.*, **11**, 636–646.
- Ramirez,C.L., Foley,J.E., Wright,D.A., Muller-Lerch,F., Rahman,S.H., Cornu,T.I., Winfrey,R.J., Sander,J.D., Fu,F., Townsend,J.A. *et al.* (2008) Unexpected failure rates for

10 *Nucleic Acids Research, 2011*

- modular assembly of engineered zinc fingers. *Nat. Methods*, **5**, 374–375.
13. Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G.M. and Arlotta, P. (2011) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat. Biotechnol.*, **29**, 149–153.
 14. Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., Toyooka, K., Matsuoka, K., Jinbo, T. and Kimura, T. (2007) Development of series of Gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J. Biosci. Bioeng.*, **104**, 34–41.
 15. Engler, C., Kandzia, R. and Marillonnet, S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE*, **3**, e3647.
 16. Koncz, C. and Schell, J. (1986) The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimeric genes carried by a novel type of *Agrobacterium* binary vector. *Mol. Gen. Genet.*, **204**, 383–396.
 17. Römer, P., Hahn, S., Jordan, T., Strauß, T., Bonas, U. and Lahaye, T. (2007) Plant-pathogen recognition mediated by promoter activation of the pepper *Bs3* resistance gene. *Science*, **318**, 645–648.
 18. Engler, C., Gruetzner, R., Kandzia, R. and Marillonnet, S. (2009) Golden Gate Shuffling: a one-pot DNA shuffling method based on type II restriction enzymes. *PLoS ONE*, **4**, e5553.
 19. Weber, E., Engler, C., Gruetzner, R., Werner, S. and Marillonnet, S. (2011) A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE*, **6**, e16765.
 20. Kay, S., Hahn, S., Marois, E., Wieduwild, R. and Bonas, U. (2009) Detailed analysis of the DNA recognition motifs of the *Xanthomonas* type III effectors AvrBs3 and AvrBs3Δrep16. *Plant J.*, **59**, 859–871.
 21. Sugio, A., Yang, B., Zhu, T. and White, F.F. (2007) Two type III effector genes of *Xanthomonas oryzae* pv. *oryzae* control the induction of the host genes *OstFIIA1* and *OstFX1* during bacterial blight of rice. *Proc. Natl Acad. Sci. USA*, **104**, 10720–10725.

Figure S1

A-1_NS
g f c t c t c t t h a c c g c a g a g a g t u g t u g t c g a t c g c a g a a h a t a g a g t t g g c a a a g c a g a g c g t c g h a g a c g t g f c a g c g c t c t c c g f a g a c c
L **T** **P** **Q** **V** **A** **I** **A** **S** **N** **S** **G** **K** **Q** **A** **L** **E** **T** **V** **Q** **R** **L** **L** **P**

B-1_NS
g f c t c t c t a g c g a t c g c c h a g c h a t h a g d g r h g d a a g c a g c c t g g a g c c g t g g a c c g t g c a g c g g t c t c c g f a g a c c
A **I** **A** **S** **N** **S** **G** **K** **Q** **A** **L** **E** **T** **V** **Q** **R** **L** **L** **P**

2_NS
g g t c t d f t t c c c g t c t g t g d e a g c d e a h g d c t g a c c c e g a g e a g t r g t g f c c h a c f g c c h a t a g c g r c g r c h a t a g c g r c g r c h a g a g c
L **P** **V** **I** **C** **Q** **A** **H** **G** **L** **T** **P** **E** **Q** **V** **V** **A** **I** **A** **S** **N** **S** **G** **K** **Q** **A** **L** **E** **T** **V** **Q** **R** **L**

3_NS
g f c t c g a t t g t t g c c g t g t g t g c c a a g c c c h a t h g g c t g a c c c e g a g a g t g t g t g f c c h a c f g c c h a t a g c g r c g r c h a t a g c g r c g r c h a g a g c
R **L** **L** **P** **V** **I** **C** **Q** **A** **H** **G** **L** **T** **P** **E** **Q** **V** **V** **A** **I** **A** **S** **N** **S** **G** **K** **Q** **A** **L** **E** **T** **V**

4_NS
g f c t c t g t c c a g c g t g t t g c c g t u g t t g t g c d e a g c c c a t u g c t g a c c c e g a g a g t g t g t g f c c h a t a g c g r c g r c h a t a g c g r c g r c h a g a g c
T **V** **Q** **R** **L** **L** **P** **V** **I** **C** **Q** **A** **H** **G** **L** **T** **P** **E** **Q** **V** **V** **A** **I** **A** **S** **N** **S** **G** **K** **Q** **A** **L** **E**

5_NS
g f c t c t t h a g a c g t t g a g t g g t g t t g f c c g t g c t g t g c c a g c c c a t u g c t g a c c c e g a g a g t g t g t g f c c h a t a g c g r c g r c h a t a g c g r c t g g g a g a c
L **E** **T** **V** **Q** **R** **L** **L** **P** **V** **I** **C** **Q** **A** **H** **G** **L** **T** **P** **E** **Q** **V** **V** **A** **I** **A** **S** **N** **S** **G** **K** **Q** **A** **L**

6_NS
g f c t c t c t g a g h c g t g f a g t g f t g f c t g f c t g f c c g a g d e a h g d c t g a c c c e g a g a g t g t g t g f c c h a t a g c g r c g r c h a t a g c g r c g r c h a g a g c
A **L** **E** **T** **V** **Q** **R** **L** **L** **P** **V** **I** **C** **Q** **A** **H** **G** **L** **T** **P** **E** **Q** **V** **V** **A** **I** **A** **S** **N** **S** **G** **K** **Q** **A** **L**

7_NS
g f c t c t g r g c a g c a g t g t c t g a g a c g t c t g a g a c g t t g a g a c g t t g c c g a g t e t g c c g a g c c a t u g c t g a c c c e g a g a g t g t g t g f c c h a t a g c g r c g r c h a t a g c g r c g r c h a g a g c
G **K** **Q** **A** **L** **E** **T** **V** **Q** **R** **L** **L** **P** **V** **I** **C** **Q** **A** **H** **G** **L** **T** **P** **E** **Q** **V** **V** **A** **I** **A** **S** **N** **S** **G** **K** **Q** **A** **L** **E** **T** **V** **Q** **R** **L** **L** **P**

8_NS
g f c t c t c a h g c h a t h a g c g r t g c a g a g a g t g c t t g a a g c c g t g c a g d e t t g c c g t g t c g t c a g c c a t u g c c t g a c c c e g a g a g t u g t u g r c a t c g g a g a c c
A **S** **N** **S** **G** **K** **Q** **A** **L** **E** **T** **V** **Q** **A** **L** **L** **P** **V** **I** **C** **Q** **A** **H** **G** **L** **T** **P** **E** **Q** **V** **V** **A** **I**

9_NS
g f c t c t c a h t c g c c h a t h a t h a g d g r t g r c h a a g c a g t c g t g r a g a c c g t g c a g d e t g t g f c c g t g t c a g c c a t u g c c t g h a c c c g f a h a c a g g a g a c c
A **I** **A** **S** **N** **S** **G** **K** **Q** **A** **L** **E** **T** **V** **Q** **R** **L** **L** **P** **V** **I** **C** **Q** **A** **H** **G** **L** **T** **P** **E** **Q** **V** **V** **A** **I**

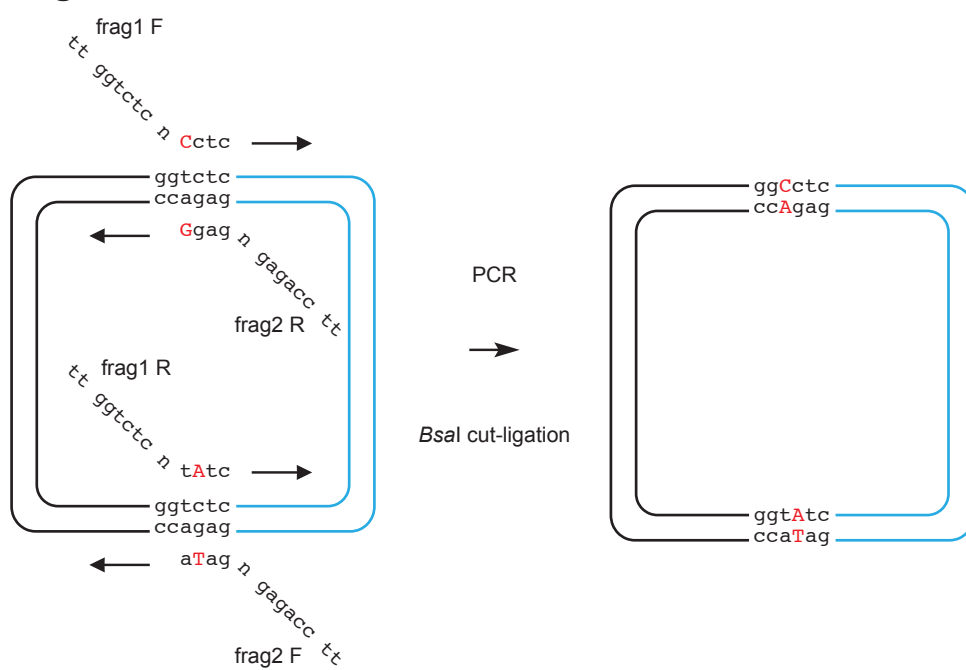
10-B_NS
g f c t c t a c a t c g t u g t c g c c a t c g c c h a g c a t h a g d g r t g g a a a g c a g c o t t g a a c o t t g c a g t c g t g f c c g t g f c t g t c a g c c a t u g c c t g h a c c c e g a g a g t u g t u g r c a t c g g a g a c c
Q **Q** **V** **V** **A** **I** **A** **S** **N** **S** **G** **K** **Q** **A** **L** **E** **T** **V** **Q** **R** **L** **L** **P** **V** **I** **C** **Q** **A** **H** **G** **L** **T** **P** **E** **Q** **V** **V** **A**

10-C_NS
g f c t c t a a a g t u g t u g c c a t c g c c h a g c a t h a g d g r t g g a a a g c a g c o t t g a a c o t t g c a g t c g t g f c c g t g f c t g t c a g c c a t u g c c t g h a c c c e g a g a g t u g t u g r c a t c g g a g a c c
Q **Q** **V** **V** **A** **I** **A** **S** **N** **S** **G** **K** **Q** **A** **L** **E** **T** **V** **Q** **R** **L** **L** **P** **V** **I** **C** **Q** **A** **H** **G** **L** **T** **P**

7-C_NS
g f c t c t c g r g c a a g c a g t g c t g g a g a c g t t g c a g c g g c t g t c c c g t g t c t g t g c a g a g c c a t u g c c t g a c c c e g a g a g t u g t u g r c a t c g g a g a c c g r c g r c h a t a g c g r c g r c h a g a g c
G **K** **Q** **A** **L** **E** **T** **V** **Q** **R** **L** **L** **P** **V** **I** **C** **Q** **A** **H** **G** **L** **T** **P** **E** **Q** **V** **V** **A** **I** **A** **S** **N** **S** **G** **K** **Q** **A** **L** **E** **T** **V** **Q** **R** **L** **L** **P**

Figure S2



Figure S3

SUPPLEMENTARY – FIGURE LEGENDS

Figure S1 | Sequences of repeat modules. Nucleotide sequences of the different 5' and 3' adaptor repeats and core repeats are shown as lower case letters. Corresponding amino acid sequences are shown below in single-letter code as upper case letters. The amino acid signature “LTP” that marks the N-terminus of each repeat is displayed in bold-italic font. Note that the only NS type repeat modules are displayed. The 6 nucleotides encoding the RVDs NS are marked in blue letters. The other RVDs are encoded as follows: NI-AATATT, NN-AATAAT, NG-AATGGC, HD-CACGAC and NK-AATAAG. The *BsaI* recognition and cleavage sites are shown in green and red letters, respectively.

Figure S2 | Alignment of RVDs and *UPT*-boxes of the depicted TALEs and promoter-reporter constructs. **(a)** The *de novo* generated dTALE[*UPA20*] has a target-optimized RVD composition for the *UPT_{AvrBs3}UPA20_P* box. RVDs are shown as uppercase letters using the single-letter code. Lowercase letters denote nucleotides present in the *UPT* boxes of the pepper *Bs3* (*UPT_{AvrBs3}Bs3_P*) and *UPA20* (*UPT_{AvrBs3}UPA20_P*) promoters that align to consecutive RVDs in *AvrBs3* and dTALE[*UPA20*] with matching (black) or non-matching (red) base preferences, respectively. Numbers mark TALE repeat units and aligned nucleotides. **(b)** Nucleotide polymorphisms between the *UPT_{AvrBs3}Bs3_P* and *UPT_{AvrBs3}UPA20_P* boxes. Nucleotides that differ between the *Bs3* and *UPA20* promoter are underlined.

Figure S3 | The combined use of IIIs-mediated cut-ligations and PCR-based mutagenesis of vector subfragments provides a highly efficient approach for simultaneous site-directed mutagenesis of multiple sites. The concept of the approach is shown here for the mutagenesis of two sites only. For each internal *BsaI* or *BpiI* recognition site (or any other site that should be mutated) primers are deduced with single nucleotide exchanges within the recognition site, indicated by red uppercase letters. Each primer contains at its 5' end a *BsaI* recognition site. The internal recognition sites are mutated by two independent PCR reactions producing fragment 1 (primers: frag1F and frag1R) and fragment 2 (primers: frag2F and frag2R). Subsequently the terminal *BsaI* sites of both PCR-amplified and mutated vector subfragments are fused to a circular vector by a *BsaI* cut ligation. Arrows represent primer parts identical to the vector sequence.

Figure S4 | Deduced Nucleotide sequence of pGWB5*. The plasmids that were used for the assembly of pGWB5* were sequence validated. However pGWB5* itself has not been sequenced to validate its nucleotide sequence.

SUPPLEMENTARY

TABLE S1 Primers used in this study		
Number	Designation	Sequence
RM 1	BsaI MutF/RobM Pho	TGAGCGTGGATCTCGCGGTATC
RM 2	BsaI Mut R/RobM	CCGGCTCCAGATTTATCAGCAATAAACC
RM 3	C-term C <i>BsaI</i> F/RobM	GGGGTCTCCACCCCAGCAGGTGGTGGCC
RM 4	N-term A <i>BsaI</i> R/RobM	GGGGTCTCGCGTAAGGTTTCAGGGGGGCACCCG
RM 5	SpecR Scal F/RobM	GGGAGTACTGGCACGAACCCAGTGGACATAAGC
RM 6	SpecR Scal R/RobM	CCCAGTACTGTCATGCATGATATATCTCCAATTTGTG
RM 7	pUC57 BB A R/RobM	TTGGTCTCACGTATAGTCTTCCGATATCTAGATGCATTCCGCG
RM 8	pUC57 BB B F/RobM Pho	GGGGTCTCTAGCGTAGTCTTCCGGCCCCGTCGACTGCAGAGGCC
RM 9	pUC57 BB B R/RobM	TTGGTCTCACGCTTAGTCTTCCGATATCTAGATGCATTCCGCG
RM 10	pUC57 BB C F/RobM Pho	GGGGTCTCTACCCTAGTCTTCCGGCCCCGTCGACTGCAGAGGCC
RM 11	GW EcoRV F/RobM	ATCACAAGTTTGTACAAAAAGCTGAACGAG
RM 12	GW EcoRV R/RobM	ATCAACCACTTTGTACAAGAAAGCTGAACG
RM 13	C-term C <i>Bpil</i> F/RobM	GGGAAGACACACCCCAGCAGGTGGTGGCC
RM 14	N-term A <i>Bpil</i> R/RobM	GGGAAGACGGCGTAAGGTTTCAGGGGGGCACCCG
RM 15	A-1_NS-fwd	GGTCTTACGCCGACGAGGTGGTG
RM 16	A-1_NS-rev	GGTCTCCGGAAGCAGCCGCTGCACCG
RM 17	B-1_NS-fwd	GGTCTTAGCGATCGCCAGCAATAGCGG
RM 18	B-1_NS-rev	GGTCTCCGGAAGCAGCCGCTGCACC
RM 19	2_NS-fwd	GGTCTTTCCGGTGCTGTGCCAGGCC
RM 20	2_NS-rev	GGTCTACAATCGCTGCACCGTCTCC
RM 21	3_NS-fwd	GGTCTCGATTGTTGCCGGTGTGTGCC
RM 22	3_NS-rev	GGTCTCGGACAGTCTCCAGCGCCTGC
RM 23	4_NS-fwd	GGTCTCCTGTCCAGCGGCTGTTGCCGG
RM 24	4_NS-rev	GGTCTCTCTCAAGCGCCTGCTTGCCACC
RM 25	5_NS-fwd	GGTCTCTTGAGACGGTGCAGCGGCTG
RM 26	5_NS-rev	CGGTGGCAAGCAGGCTCTGGGAGACC
RM 27	6_NS-fwd	GGTCTCCTCTGGAGACGGTGCAGCGGC
RM 28	6_NS-rev	GGTCTCTGCCCCGCTATTGCTGGCG
RM 15	7_NS-fwd	GGTCTCTGGGCAAGCAGGCGCTGGAG
RM 15	7_NS-rev	GGTCTCAGCTTGCATGGCCACCACC
RM 15	8_NS-fwd	GGTCTCCAAGCAATAGCGGTGGCAAGC
RM 15	8_NS-rev	GGTCTCCGATTGCCACCACCTGCTCC
RM 15	9_NS-fwd	GGTCTCCAATCGCCAGCAATAGCGGTGG
RM 15	9_NS-rev	GGTCTCCCTGTTGCGGGTTCAGGCC
RM 15	10-B_NS-fwd	GGTCTCAACAGGTGGTGGCCATCGCC
RM 15	10-B_NS-rev	GGTCTCACGCTACCACCTGCTGGGGTG
RM 15	10-C_NS-fwd	GGTCTCAACAGGTGGTGGCCATCGCC
RM 15	10-C_NS-rev	GGTCTCGGGGTGTCAGGCCATGGCCTGGC
RM 15	7-C_NS-fwd	GGTCTCTGGGCAAGCAGGCGCTGGAG

3 Diskussion

3.1 Die Spezifität der TALE-DNA-Bindedomäne ist modulierbar

3.1.1 Nicht-RVDs haben keinen Einfluss auf die Erkennungsspezifität

Auf der Grundlage des TALE-Codes (Abb. 5), demzufolge die DNA-Erkennungsspezifität eines TALEs durch seine RVDs determiniert wird, wurden TALE-Bindestellen vorhergesagt und erste, *in vitro* erstellte TALEs (dTALEs) mit zufälliger DNA-Erkennungsspezifität generiert (Boch et al., 2009). Die in natürlichen TALEs vorkommenden *Repeat*-Monomere eines RVD-Typs (z.B. alle NI-*Repeats* von AvrBs4) sind jedoch nicht identisch, da es weitere Positionen gibt, die sich innerhalb eines *Repeat*-Moduls unterscheiden (Abb. 12). Als *Repeat*-Modul wird nachfolgend die DNA-Sequenz bezeichnet, welche die vollständige DNA-Bindedomäne kodiert. Im Rahmen der vorliegenden Arbeit sollte geklärt werden, ob diese Nicht-RVD-Unterschiede einen Einfluss auf die Erkennungsspezifität und/oder die transkriptionelle Aktivität von TALEs haben. Ausgehend von zwei TALEs mit unterschiedlicher RVD- und Nicht-RVD-Zusammensetzung wurden dafür Chimären erstellt in denen die RVDs der TALEs untereinander ausgetauscht wurden. Dies bedingt, dass RVDs in den Nicht-RVD-Hintergrund des anderen eingefügt werden (M1 Abb. S2). Die Koexpressionen dieser TALEs mit korrespondierenden GUS-Reportergenkonstrukten zeigten, dass nur geringe Unterschiede in den Reporterexpressionen für die zwei getesteten TALEs mit gleicher RVD-Abfolge auftraten (M1 Abb. S2). Diese Befunde weisen darauf hin, dass Nicht-RVDs höchstens einen geringen Einfluss auf die transkriptionelle Aktivierung haben. Die geringen Unterschiede der Expressionslevel des Reportergens in Abhängigkeit der beiden RVD-Kompositionen sind wahrscheinlich auf die unterschiedliche Position der Bindeelemente auf dem Reporterkonstrukt relativ zum ATG zurückzuführen. Da der Transkriptionsstart durch die Position des TALEs bestimmt wird (Römer et al., 2009a), kommt es wahrscheinlich zu unterschiedlichen Transkripten. Diese können sich in ihrer Stabilität unterscheiden und damit auch Ursache für die verschiedenen Expressionslevel des Reportergens sein. Eine andere Erklärung bieten die unterschiedlichen RVD-Kompositionen der TALEs. So ist es sehr wahrscheinlich, dass verschiedene RVDs verschiedene DNA-Affinitäten haben, was direkt die Aktivierungsstärke beeinflusst. Um die erste Hypothese zu untersuchen,

sollten Reportergerkonstrukte generiert werden, bei denen die Bindeelemente eine identische Position haben. Um den Einfluss unterschiedlicher RVDs auf die Stärke der transkriptionellen Aktivierung zu testen, sollte TALEs verwendet werden, deren RVD-Abfolge sich nur an vorher definierten *Repeat*-Positionen unterscheidet. So könnte der zu untersuchende RVD z.B. als *Trimer* eingesetzt (M1 Abb. S8) werden oder etwa alternierend mit konstanten RVDs getestet werden.

In weiteren Untersuchungen könnten bestimmte Nicht-RVD-Konfigurationen in Kombination mit verschiedenen RVDs getestet werden. Dies könnte zur Identifizierung eines optimalen *Repeat*-Monomers in Abhängigkeit des RVDs führen und eine Möglichkeit zur Modulierung der transkriptionellen Aktivität darstellen. Auch in diesen Experimenten sollten TALEs mit mehreren *Repeats* der zu untersuchenden RVD/Nicht-RVD-Zusammensetzung getestet werden, um einen messbaren Einfluss der zu untersuchenden *Repeats* relativ zur Gesamtaktivität des TALEs zu erreichen. Im Rahmen der vorliegenden Arbeit konnte nur ein sehr geringer Einfluss der Nicht-RVDs auf die transkriptionelle Aktivität bestimmt werden, der in weiteren Analysen nochmals hinterfragt werden sollte. Die Spezifität der Chimären wird allerdings ausschließlich von den RVDs bestimmt, da diese durch andere Nicht-RVDs nicht verändert wurde.

3.1.2 C-terminal angefügte *Repeats* erhöhen die Erkennungsspezifität

Nachdem in der vorliegenden Arbeit gezeigt wurde, dass die Erkennungsspezifität nur durch den RVD eines *Repeat*-Monomers bestimmt wird, sollte untersucht werden, ob durch zusätzlich angefügte *Repeat*-Monomere die Erkennungsspezifität eines *Repeat*-Moduls erhöht werden kann.

Dafür wurde die Bindedomäne von AvrBs3, welche u.a. die transkriptionelle Aktivierung des Paprika *Bs3*- und *UPA20*-Promotors vermittelt (Kay et al., 2007; Römer et al., 2007), um 4 *Repeats* verlängert. Die Paprika *Bs3* und *UPA20* Promotoren sind innerhalb der AvrBs3-Bindestelle nahezu identisch. Sequenzunterschiede außerhalb der AvrBs3-Bindestelle erlauben es jedoch, zusätzliche für den *Bs3*- und *UPA20*-Promotor sequenzspezifische *Repeats* an AvrBs3 anzufügen. Diese angefügten *Repeats* ermöglichten in der Tat eine differentielle Induktion der Promotoren durch TALEs mit zusätzlichen C-terminalen *Repeats* (M1 Abb. S6). Auf molekularer Ebene lässt sich die veränderte Erkennungsspezifität auf zwei Weisen erklären. Es ist denkbar, dass die *Repeats*

eine starre Struktur bilden. Die zusätzlichen, nicht-passenden C-terminalen *Repeats* bedingen daher möglicherweise eine vollständige Unterdrückung der TALE-DNA-Interaktion. Folglich kommt es auch nicht zur transkriptionellen Aktivierung. Alternativ ist es denkbar, dass die *Repeats* eine gewisse Flexibilität aufweisen, wodurch nur der C-terminale Bereich des TALEs nicht mit der DNA interagiert. Diese fehlende DNA-Interaktion könnte sterische Veränderungen im C-terminalen Bereich des TALEs hervorrufen, welche die Positionierung der Aktivierungsdomäne zur DNA beeinflussen. Dann würde es trotz Bindung, zumindest eines Teils der Bindedomäne, nicht zu einer transkriptionellen Aktivierung kommen. Welche der beiden Hypothesen zutrifft, ließe sich durch *in vitro* DNA-Bindungsstudien (z.B. *electrophoretic mobility shift assays* oder *microscale thermophoresis*) klären. Auch wäre zu klären, ob N-terminal angefügte *Repeat*-Polymere ebenso einen Einfluss auf die Spezifität ausüben wie die in der vorliegenden Arbeit getesteten C-terminalen *Repeats*. Diese N-terminalen Verlängerungen lassen sich allerdings schwerer realisieren als C-terminale Verlängerungen. Weil das T an Position -1 von TALE-Bindestellen invariant ist (Moscou und Bogdanove, 2009), kann die Zahl der angefügten *Repeats* in einem vorgegebenen Promotorkontext nicht frei gewählt werden. Die Erhöhung der Erkennungsspezifität kann durch N-terminale *Repeats* sehr wahrscheinlich nur durch eine vollständige Unterdrückung der Bindung erfolgen. Bei einer erfolgreichen Bindung des TALEs an die Erkennungssequenz müssten die N-terminal angefügten, nicht passenden *Repeats* abstehen. Dies hätte aber sicher keinen Einfluss auf die sterische Ausrichtung der Aktivierungsdomäne, da diese durch die im C-terminalen Bereich passenden *Repeats* fixiert sein sollte. Auch für TALEs mit N-terminal verlängerten Bindedomänen sollten in weiteren Untersuchungen *in vitro*-Bindungsstudien durchgeführt werden. Anhand dieser könnte erneut zwischen verhinderter Bindung und verhinderter Aktivierung unterschieden werden.

Bei der Verlängerung der DNA-Bindedomäne wird in der Regel eine Erhöhung der DNA-Spezifität angestrebt. Möglicherweise binden diese verlängerten DNA-Bindedomänen an weiteren im Genom vorhandenen Zielsequenzen, an die das kürzere Ausgangsprodukt nicht bindet. Dies könnte dazu führen, dass der verlängerte TALE zwar bezogen auf die untersuchte Erkennungsstelle spezifischer wird, allerdings auf genomweiter Ebene insgesamt an Spezifität verliert. Um zu prüfen, ob zusätzliche *Repeats* neben der gewünschten erhöhten lokalen Spezifität

auch auf genomweiter Ebene eine erhöhte Spezifität aufweisen, könnten vergleichende Gesamt-Transkriptom-Shotgun-Sequenzierungsanalysen für unterschiedlich lange TALEs durchgeführt werden, welche die gleiche Bindestelle ansteuern. Diese Daten würden auch eine Aussage darüber ermöglichen, ob längere DNA-Bindedomänen auch mehr Fehlpaarungen zulassen als kürzere und somit auch zeigen, wie das Verhältnis zwischen Spezifität und Fehlpaarungstoleranz pro angefügten *Repeat* ist.

Einen Anhaltspunkt zur Beantwortung dieser Frage bietet u.U. die Analyse der bekannten, natürlich vorkommenden TALEs und ihrer Erkennungsstellen (Abb. 7).

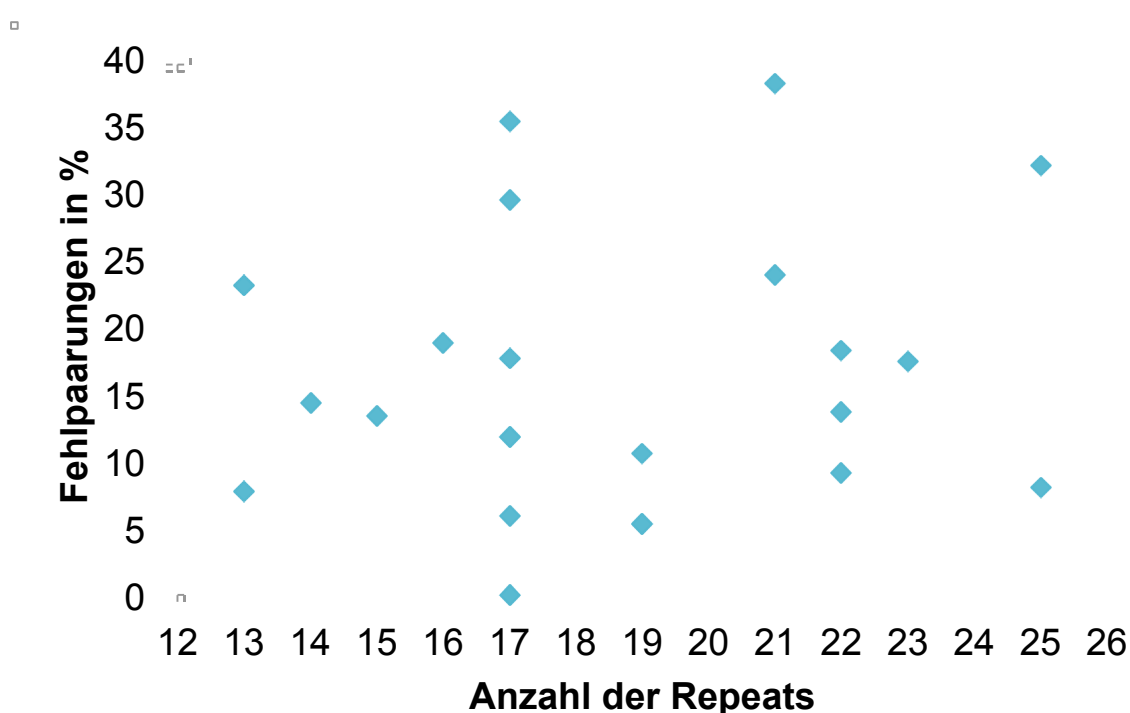


Abb. 7: Darstellung der Fehlpaarungsquote in Abhängigkeit der *Repeat*-Anzahl. Dafür wurden die vollständigen RVD der jeweiligen TALEs mit den interagierenden Nukleotiden der Zielpromotoren verglichen und der Anteil an Fehlpaarungen zu den Gesamtpaarungen (Anzahl der *Repeats*) ermittelt. Anschließend wurde die Fehlpaarungsquote (y-Achse) gegen die *Repeat*-Anzahl (x-Achse) aufgetragen. Als Fehlpaarung wurden alle RVD-Nukleotid-Paarungen angesehen, die nicht den folgenden entsprechen: HD/ND/N*-C, NI-A, NG/HG-T, NN-G/A. Die der Abbildung zugrundeliegenden TALE-Promotor-Analysen sind in

Abb. 16 aufgelistet.

Anhand der Analyse ergibt sich eine durchschnittliche Fehlpaarungsquote von 15%. Das bedeutet, dass etwa jeder siebte RVD zu einem nicht optimalen Nukleotid paart. Es lässt sich allerdings auf Grundlage dieser Daten kein Zusammenhang zwischen der *Repeat*-Anzahl und der Fehlpaarungsquote feststellen. Des Weiteren liegen

diesen positiven Interaktionen unterschiedliche Reporter (z.B. Induktion von Zelltodreportern oder GUS-Reportern) zugrunde, die jeweils Unterschiede bezüglich ihrer Sensitivität und Quantifizierbarkeit aufweisen. Demzufolge sind direkte Vergleiche über die Aktivität der TALEs auf den jeweiligen Promotoren auf Basis der verfügbaren Daten nur bedingt möglich. Somit kann nicht beurteilt werden, ob es mit steigender Fehlpaarungsquote zu einer graduellen Reduktion der Induktionsstärke kommt.

3.1.3 Der RVD NK besitzt im Vergleich zu NN eine höhere Spezifität für G

Um eine DNA-Sequenz spezifisch zu binden, müssen für jede der vier Basen jeweils korrespondierende, hochspezifische RVDs verwendet werden. Für die Basen C (HD), A (NI) und T (NG) waren RVDs bekannt, die präferentiell mit diesen Basen paaren. Der mit der Base G interagierende RVD NN paart jedoch in gleicher Frequenz mit A (Boch et al., 2009). Somit konnte G in Zielsequenzen nicht spezifisch mit einem korrespondierenden RVD angesteuert werden. Für zahlreiche TALE-RVDs war die Erkennungsspezifität nicht eindeutig geklärt, da sie aus RVD-Nukleotid-Gegenüberstellungen mit sehr geringer Häufigkeit abgeleitet wurden (Moscou und Bogdanove, 2009). So zeigen nur 2 RVD-DNA-Interaktionen eine Interaktion des RVD NK mit dem Nukleotid G an (Moscou und Bogdanove, 2009). Um die Signifikanz dieser Korrelation zu testen, wurde ein *in vitro* erzeugter *dTALE* verwendet, der mehrere NK RVDs enthielt. Im erzeugten *dTALE* wurden die RVD von drei aufeinanderfolgenden *Repeats* zu NK geändert (M1 Abb. S7).

Dieser *dTALE* war nur dann in der Lage, einen transkriptionellen Reporter zu aktivieren, wenn an den komplementären Positionen der TALE-Erkennungsstelle drei hintereinander angeordnete G-Nukleotide vorhanden waren (M1 Abb. S8). Diese *in vivo*-Daten stimmen mit *in vitro*-SELEX-Experimenten (Systematische Evolution von Liganden durch exponentielle Anreicherung) überein, die ebenfalls eine höhere Spezifität von NK zu G als von NN zu G zeigten (Miller et al., 2011).

Durch die Fusion der FokI-Nuklease an TALE-basierten DNA-Bindedomänen können TALE-Nukleasen (TALEN) erzeugt werden. Diese sind in der Lage sequenzspezifische DNA-Doppelstrangbrüche zu induzieren (Christian et al., 2010; Li et al., 2011a). In TALEN-Assays wurden für TALENs, deren Bindedomänen sich nur durch die Verwendung von NK und NN unterschieden, differenzielle Nukleaseaktivitäten bestimmt. Dabei besaßen die TALENs mit NN anstatt NK in

in vitro- und *in vivo*-Experimenten eine fünf-fach bzw. zwei-fach höhere Nukleaseaktivität (Huang et al., 2011). Die aktuellen Daten lassen vermuten, dass NN zwar nur eine geringe Spezifität für G hat, aber eine große Affinität zu G, die in einer höheren Aktivität resultiert. Im Gegensatz dazu besitzt NK zwar eine hohe Spezifität für G, allerdings ist diese nicht mit einer hohen Affinität gepaart, da die Aktivität von TALEs mit NK vergleichsweise geringer ist.

Eine Ursache für die erhöhte Spezifität von NK könnten unterschiedliche Bindungseigenschaften der RVDs sein. Deren Grundlage müsste in der unterschiedlichen Struktur der Aminosäuren Asparagin (N) und Lysin (K) liegen, da gezeigt werden konnte, dass die Nukleotiderkennung von TALEs über die große Furche der DNA nur durch die Aminosäure an Position 13 der *Repeats* erfolgt (Deng et al., 2012; Mak et al., 2012). Asparagin bildet demnach eine Wasserstoffbrückenbindung mit dem N7-Stickstoff des G-Nukleotides aus. Keine der in den beiden vorhandenen TALE-DNA-Kristallstrukturen untersuchten Bindedomänen besitzt ein RVD NK. Daher bieten diese auch keine Erklärung für die unterschiedlichen Eigenschaften der beiden RVDs. Ein Vergleich der beiden RVDs auf struktureller Ebene ist anhand empirischer Ergebnisse bisher nicht möglich.

Es sind weitere vergleichende Untersuchungen von NN und NK nötig, insbesondere die Ermittlung von Bindungskonstanten bzw. Kristallstrukturen, um die unterschiedlichen Eigenschaften von NK bzw. NN für die G-Nukleotid-Erkennung erklären zu können.

3.1.4 Die Optimierung der RVD-Abfolge erhöht die Erkennungsspezifität

Die RVD-Abfolge von natürlich vorkommende TALEs ist in Bezug auf ihre Bindestelle(n) nur sehr selten optimal. So sind von den für Abb. 7 analysierten 31 TALE-DNA-Interaktionen nur zwei perfekt. Das heißt, dass es keine suboptimalen Nukleotid-RVD-Paarungen gibt. Des Weiteren besitzen 17 der 20 untersuchten TALEs den RVD NS, der keine Spezifität vermittelt. Daher sollte in der vorliegenden Arbeit als weitere Möglichkeit zur Veränderung der DNA-Bindesepezifität die Optimierung der RVD-Komposition der DNA-Bindedomäne untersucht werden. Dabei wurde für jedes Nukleotid der Erkennungssequenz ein spezifischer und passender RVD verwendet. In Morbitzer et al. (2011) konnte gezeigt werden, dass eine optimale RVD-Komposition bei gleichbleibender Länge des *Repeat*-Polymers zu einer erhöhten Erkennungsspezifität führt. Dafür reichten vier

Nukleotidunterschiede aus (Position 11, 15, 17, 18 von insgesamt 18), die sich überwiegend im 3'-Bereich der Zielsequenz befanden, um zwei ansonsten identische TALE-Bindestellen spezifisch in einem Aktivierungs-Assay zu unterscheiden (M2 Abb. 4 und Abb. S2).

In weitergehenden Experimenten sollte untersucht werden, ob im 5'-, zentralen oder 3'-Bereich der Erkennungssequenz Fehlpaarung besser oder schlechter toleriert werden. Dieses Wissen wäre für die Selektion einer spezifischen Erkennungsstelle hilfreich, um zusätzliche und daher unerwünschte Erkennungsstellen (*off-targets*) zu vermeiden, da Fehlpaarungen gezielt in diesem Bereich akkumuliert werden könnten.

Zusammenfassend zeigten unsere Ergebnisse, dass eine Erhöhung der Erkennungsspezifität eines TALEs möglich ist durch eine Verlängerung der Bindedomäne und die Optimierung der RVD-Abfolge mittels Verwendung von spezifischen und zur Nukleotidsequenz passenden RVDs.

3.2 TALE-basierte *Repeat*-Module mit benutzerdefinierter Erkennungsspezifität können über verschiedene Methoden generiert werden

3.2.1 Die PCR-basierte Modifikation von TALE-DNA-Bindedomänen mit *Xanthomonas*-Nukleotidsequenz ist schwer durchführbar

Die *in vitro*-Erstellung von TALE-basierten *Repeat*-Modulen ist mit Hilfe klassischer Klonierungsmethoden schwer durchführbar, da die *Repeat*-Monomere aus *Xanthomonas* eine hohe Sequenzidentität aufweisen. So sind z.B. die *Repeats* von *AvrBs3* auf Nukleotidebene zu 92% sequenzidentisch. Darüberhinaus sind die *Repeat*-Monomere für *Repeat* 1, 8, 9, 13, 14, 15 und 17, *Repeat* 5, 6 und 7 sowie *Repeat* 2 und 10 identisch (Abb. 15). Dies macht eine PCR-basierte (polymerase chain reaction, Polymerase Ketten-Reaktion; PCR) Modifikation des *avrBs3-Repeat*-Moduls schwer durchführbar, da keine für jedes *Repeat*-Monomer spezifischen Oligonukleotide abgeleitet werden können. Daher lässt sich die RVD-Komposition und damit die Erkennungsspezifität von *AvrBs3* nicht gezielt modifizieren. Des Weiteren beträgt der GC-Gehalt des *Repeat*-Moduls von *avrBs3* 70%, wodurch PCR-Amplifikationen der kodierenden Sequenz zusätzlich erschwert werden. Daher sollten andere Wege gefunden werden, um TALE-basierte DNA-Bindedomänen mit benutzerdefinierter Erkennungsspezifität zu generieren.

3.2.2 Die Veränderung der *codon usage* ermöglicht die PCR-basierte Modifikation von TALE-DNA-Bindedomänen

Eine erste Lösung dieses Problems bestand in der Synthese eines *avrBs3*-Gens das einen niedrigeren GC-Gehalt aufwies und auf maximale Varianz der verschiedenen *Repeat*-Monomere optimiert war (M1 SI-Text). Dies wurde durch die Nutzung der Degeneration des genetischen Codes erreicht. Diese Modifikationen erlaubten erstmals eine gerichtete Mutagenese und Amplifikation eines definierten *Repeat*-Polymers (M1 Abb. S4, S6 und S7). Auf der Grundlage des codon-optimierten *avrBs3*-Genes wurde auch der erste *in vitro* generierte *dTALE* erstellt (M1 Abb.1).

Mithilfe einer für den Zielorganismus optimalen *codon usage* ist es möglich, den TALE-Expressionslevel zu erhöhen. Dies wurde anhand vergleichender Expressionsstudien von TALEs mit unterschiedlicher *codon usage* herausgefunden. Dafür wurde *AvrBs3* und *AvrBs4* (ein weiterer TALE von *Xanthomonas* mit abweichender RVD-Komposition), entweder mit *Xanthomonas*- oder *N. benthamiana*-*codon usage* (Abb. 14), Agrobakterium-vermittelt transient in *N. benthamiana* exprimiert und anschließend immunologisch detektiert. Dabei konnten größere Proteinmengen für die ausgehend von *N. benthamiana*-*codon usage* exprimierten TALEs nachgewiesen werden. Die durch *N. benthamiana*-*codon usage* erhöhten Proteinmengen von *AvrBs3* führten auch zur Ausbildung einer Zelltodreaktion, die bei der Expression mit *Xanthomonas codon usage* nicht zu beobachten ist (Morbitzer und Lahaye, unveröffentlicht). Da die Expression von *AvrBs4* mit codon-optimierter Nukleotidsequenz nicht zur Ausbildung einer Zelltodreaktion führte, handelt es sich um eine spezifische Abwehrreaktion. Mittels Gesamt-Transkriptom-Shotgun-Sequenzierungsanalysen könnten die der *AvrBs3* vermittelten Abwehrreaktion zugrundeliegenden Gene identifiziert werden.

Die Verringerung der Redundanz der Nukleotidsequenz des *Repeat*-Moduls durch die Veränderung der *codon usage* bietet zwar die Möglichkeit der PCR-basierten Modifikation, allerdings kann es auch zur Veränderung der Expressionslevel kommen. Für eine gezielte Anwendung von TALEs ist die in Abhängigkeit der *codon usage* erzeugte Proteinmenge ein entscheidender Faktor. Um die Induktion von *off-targets* zu vermeiden, sollte die TALE-Proteinmenge, u.U. auch zu lasten der Induktion des Zielgens nicht zu hoch sein.

3.2.3 Multi-Fragment-Assemblierungen ermöglichen die schnelle und effiziente Erstellung von *Repeat*-Modulen

Ziel der vorliegenden Arbeit war die Etablierung eines Klonierungssystems, mit dessen Hilfe DNA-Bindedomänen von TALEs schnell und effizient erstellt werden können. Um dies zu erreichen, mussten eine Vielzahl von *Repeat*-Monomeren in gewünschter Reihenfolge übergangslos zu einem *Repeat*-Modul verbunden werden. Mehrere *Repeat*-Monomere werden im folgenden als *Repeat*-Polymer bezeichnet, falls dies noch nicht das vollständige *Repeat*-Modul kodiert.

Derzeit gibt es verschiedene kommerzielle Plattformen wie MultiSite Gateway® von Life Technologies (USA, Abb. 8) und In-Fusion-HD von Clontech (USA, Abb. 9) um eine Multi-Fragment-Assemblierung durchzuführen.

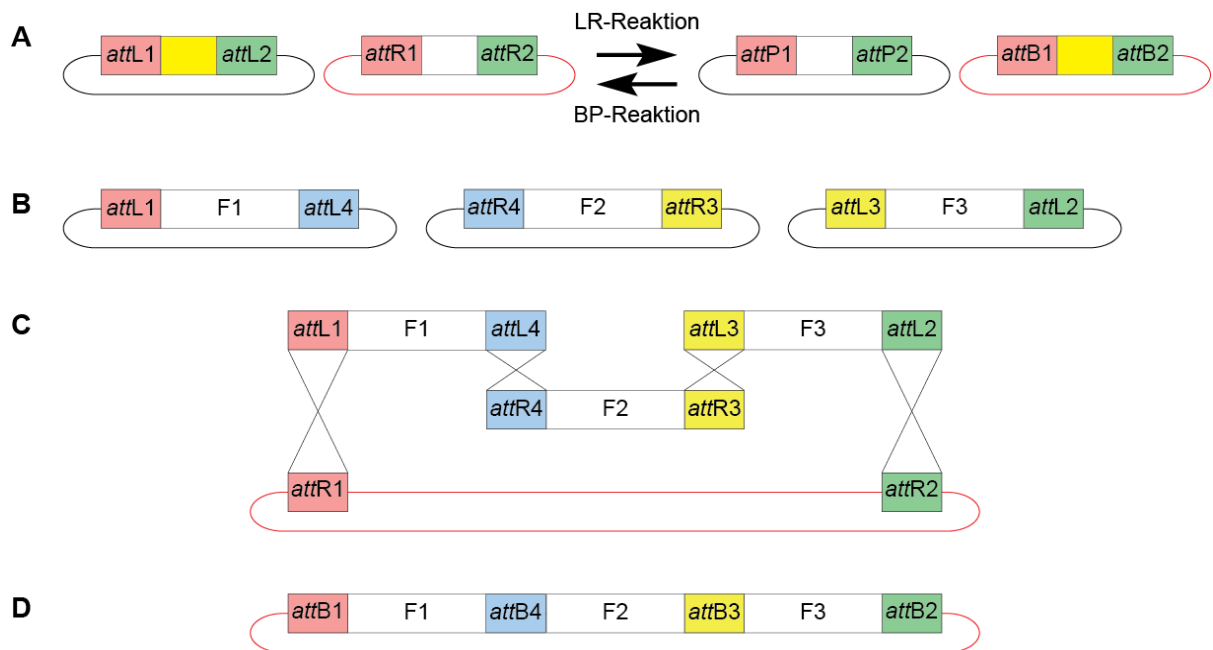


Abb. 8: Prinzip der MultiSite Gateway®-Klonierung. Die Gateway®-Technologie beruht auf der sequenzspezifischen Rekombination des Bakteriophagen λ in das *E. coli*-Genom. Diese ist abhängig von vier *att*-Stellen (*attB*, *attP*, *attL*, *attR*) und drei Proteinen, die an der Integration und Entfernung des Bakteriophagen λ beteiligt sind. Katalysiert wird die Reaktion von der viralen Integrase und dem bakteriellen *integration host factor* (BP-Reaktion) und zusätzlich der viralen Excisionase (LR-Reaktion) (Hartley et al., 2000). (A) Während der Gateway-Klonierung kommt es zum rekombinationsbasierten Austausch der innerhalb der *att*-Stellen befindlichen Fragmente (gelbes und weißes Rechteck) und des außerhalb der *att*-Stellen befindlichen Rückgrats (schwarze und rote Linie zwischen den *att*-Stellen). Als Rekombinationsprodukte entstehenden dabei aus *attL* und *attR* *attB*- und *attP*-Stellen (LR-Reaktion), sogenannte „Hybridsequenzen“. Die Umkehrreaktion wird als BP-Reaktion bezeichnet. Für das MultiSite Gateway®-System wurden die *att*-Stellen hinsichtlich ihrer Länge und Nukleotidsequenz optimiert, um gleichzeitig mehrere rekombinationsbasierte Klonierungen ausführen zu können. Dadurch können mittels fünf verschiedenen *att*-Stellen bis zu vier Fragmente gleichzeitig kloniert werden. (B) Für die Klonierung von drei Fragmenten (Rechtecke mit F1, F2 und F3) müssen diese Fragmente mit unterschiedlichen *attL*- bzw. *attR*-Stellen flankiert sein (farbige Rechtecke). (C) Während der LR-Reaktion kommt es zur Rekombination zwischen *attL*- und *attR*-Stellen gleicher Nummerierung (*att*-Stellen mit identischer Farbe). (D) Dabei entsteht ein Fusionsprodukt aus den drei in die LR-Reaktion eingesetzten Fragmenten, die über *attB*-Stellen verknüpft sind und in einem neuen Vektor vorliegen.

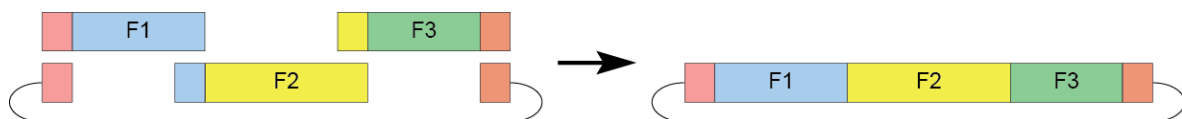


Abb. 9: Prinzip der In-Fusion-HD-Klonierung. Durch In-Fusion-HD können mehrere linearisierte Fragmente (Rechtecke mit F1, F2 und F3), die eine 15 bp lange Sequenzidentität mit dem 5'- bzw. 3'-Fragment besitzen (überlappende Bereiche der Fragmente sind in gleicher Farbe dargestellt) über homologe Rekombination mittels In-Fusion-Enzym in einen Zielvektor ligiert werden. Die Linearisierung der Fragmente erfolgt in der Regel über PCR-Amplifikation, so dass über Oligonukleotide, die jeweiligen 15 bp, überlappenden Sequenzen angefügt werden können. Der Zielvektor kann über PCR-Amplifikation oder mittels Restriktionsenzymen linearisiert werden.

Ein Hauptproblem für die Assemblierung von DNA-Bindedomänen ist die Tatsache, dass durch die Klonierung keine zusätzlichen Nukleotidsequenzen eingefügt werden dürfen, da diese Teil der kodierenden Sequenz werden. Da bei der Verwendung von MultiSite Gateway® nach der Rekombination 21 Nukleotide (*attB*-Stellen) hinzugefügt werden, ist diese Methode für die Erstellung von TALE-DNA-Bindedomänen ungeeignet. Im Gegensatz dazu werden bei der In-Fusion-HD-Klonierung keine zusätzlichen Nukleotide eingefügt. Die für jeden *Repeat* spezifischen Nukleotide, der 15 bp langen überlappenden Sequenz sollten ausreichen, um *Repeat*-spezifische Überhänge zu generieren. Dadurch könnten einige *Repeats* in definierter Reihenfolge zusammen gefügt werden. Da nach Herstellerangaben bis zu 8 Fragmente gleichzeitig ligiert werden können, müsste die Erstellung einer 17,5 *Repeats* langen DNA-Bindedomäne in einem mehrstufigen Prozess erfolgen. Sowohl für die Erstellung der einzelnen *Repeat*-Monomere als auch für die Fusion der *Repeat*-Polymere sind allerdings PCR-Schritte notwendig. Dadurch ist diese Methode anfällig für Fehler, die durch die verwendete Polymerase bzw. die verwendeten Oligonukleotide eingeführt werden. Des Weiteren betragen bei der In-Fusion-HD-Klonierung die Kosten pro Reaktion mehr als 10€. Eine vor allem preiswertere Alternative zu den kommerziell vertriebenen MultiSite Gateway® und In-Fusion-HD-Systemen bietet die *Golden Gate*-Klonierung (Abb. 10).

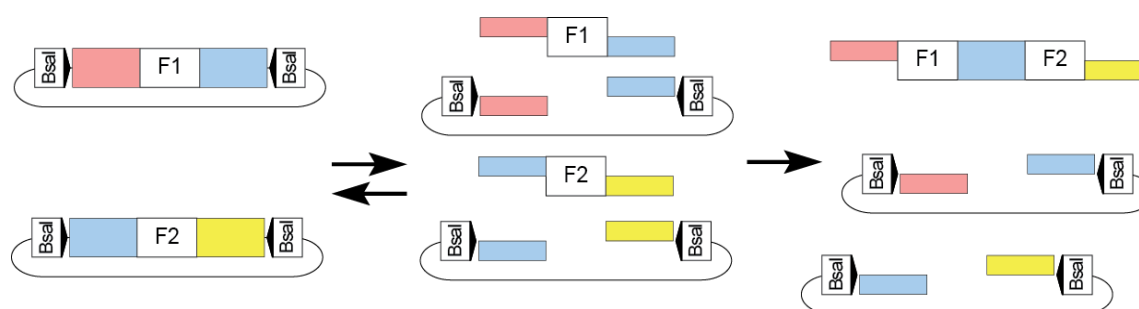


Abb. 10: Prinzip der *Golden Gate*-Klonierung. In einer *Golden Gate*-basierten Mehrfragment-Ligation werden in einem Schritt gleichzeitig mehrere *Repeat*-Monomere in definierter Reihenfolge ligiert (Engler et al., 2008). Die *Golden Gate*-Klonierungsmethode beruht auf der Nutzung von Restriktionsenzymen deren Erkennungsstelle (Kästchen mit „Bsal“) und Schnittstelle (farbige Kästchen) räumlich von einander getrennt sind. Diese werden als Typ IIS-Restriktionsenzyme bezeichnet und bieten die Möglichkeit Überhänge mit gewünschten Sequenzen zu generieren. Über sequenzspezifische Ligation können mehrere DNA-Fragmente in definierter Reihenfolge und übergangslos, d.h. ohne zusätzliche Nukleotide der Restriktionsschnittstellen, in der kodierenden Sequenz verknüpft werden. Dafür werden die Typ IIS-Erkennungsstellen 5' und 3' des DNA-Fragmentes so angeordnet, dass nach der Spaltung die Erkennungsstellen nicht mehr Teil des zu ligierenden DNA-Fragmentes sind. Werden die dabei entstehenden 4 bp 5'-Überhänge so gewählt, dass zwei aufeinanderfolgende DNA-Fragmente kompatible Überhänge (gleiche Farbe) besitzen, entsteht ein Ligationsprodukt, das keine Typ IIS-Erkennungsstelle besitzt. Da das gewünschte Reaktionsprodukt, aber nicht die Ausgangsvektoren gegenüber dem Restriktionsenzym, stabil sind, können Restriktion und Ligation einer Vielzahl von Fragmenten effizient in einer Reaktion ablaufen (Engler et al., 2009).

Die *Golden Gate*-Klonierung ist eine preiswerte und sehr effiziente Methode, um übergangslos und in gerichteter Orientierung mehrere Fragmente zu klonieren. Deshalb wurde auf Grundlage dieser Klonierungsmethode versucht, TALE-basierte *Repeat*-Module zu klonieren.

3.2.4 Die für die Mehrfragment-Ligation benutzen 4 bp-Überhänge können durch die Verwendung der Degenerierung des genetischen Codes und die Verschiebung des Fusionspunktes optimiert werden

Der im Rahmen der vorliegenden Arbeit entwickelte Ansatz zu Erstellung von *Repeat*-Domänen beruht auf der Ligation von *Repeat*-Monomer-Fragmenten mit Fragment spezifischen Überhängen durch Typ IIS-Restriktionsenzymen. Die gerichtete Ligation mehrerer Fragmente in gewünschter Reihenfolge wird durch die Überhänge der *Repeat*-Monomere definiert. In der vorliegenden Arbeit wurden Typ IIS-Enzyme verwendet, die einen 4 bp-Überhang generieren. Daher können 4^4 (256) verschiedene 4 bp Überhänge generiert werden. Da keines der 16 mit 4 bp

möglichen Palindrome als Überhang verwendet werden sollte, verringert sich die Zahl der möglichen Überhänge auf 240 (Engler et al., 2008).

Die verwendeten Überhänge sollten möglichst unterschiedlich sein, wobei dies besonders für die beiden inneren Nukleotide des Überhangs gilt. Durch Exonukleaseaktivität kann es während einer Multifragment-Ligation teilweise zum Verlust des äußersten Nukleotids kommen (Engler et al., 2009). Deshalb vermitteln die inneren Nukleotide die größte Spezifität und jede Kombination sollte wenn möglich nur einmal vorkommen. In der vorliegenden Arbeit wurden daher elf verschiedene Überhänge benutzt, deren Kombination der beiden inneren Nukleotide immer unterschiedlich ist (M2 Abb. S1).

Bei der Verknüpfung von Modulen, die eine translationale Fusion bilden, darf es während der Optimierung der Überhänge nicht zu Veränderungen der resultierenden Aminosäureabfolge kommen. Die Homologie zwischen den 4 bp-Überhängen kann unter Beibehaltung der kodierten Aminosäure durch die Verwendung der Degeneration des genetischen Codes verringert werden. Allerdings reicht die Variabilität der Nukleotidabfolge, die durch Nutzung der Degenerierung des genetischen Codes erzielt werden kann, nicht aus, um die inneren Nukleotide der Überhänge immer unterschiedlich zu gestalten. Eine weitere Erhöhung der Varianz der Überhänge erfolgte durch die Verschiebung der Fusionspunkte zwischen zwei Fragmenten (M2 Abb. S1). Dadurch wurden Nukleotide benutzt, welche für verschiedene Aminosäuren kodieren und somit eine größere Diversität der Überhänge erzeugt werden. Auch in anderen Publikationen wurde gezeigt, dass unterschiedliche Überhänge durch die Degeneration des genetischen Codes (Li et al., 2011b; Zhang et al., 2011) bzw. zusätzlich durch die Verschiebung der Fusionspunkte zwischen zwei Fragmenten erzeugt werden können (Cermak et al., 2011; Geissler et al., 2011; Weber et al., 2011).

Durch die Kombination aus unterschiedlichem Fusionspunkt und Degeneration des genetischen Codes können sehr variable 4 bp-Überhänge geschaffen werden. Dadurch kann die Effizienz und Spezifität der Mehrfragment-Ligation und folglich die Anzahl der miteinander verbundenen Fragmente stark erhöht werden.

3.2.5 Vergleichende Analyse bisher verwendeter Strategien zur Erzeugung von TALE-DNA-Bindedomänen

3.2.5.1 Plasmid-basierte Repeat-Bibliotheken sind weniger fehleranfällig als PCR-basierte Repeat-Bibliotheken

Die in der vorliegenden Arbeit erstellten *Repeat*-Monomere liegen in einer Plasmid-Bibliothek vor und werden in einer Restriktions-Ligations-Reaktion zu *Repeat*-Polymeren bzw. einem *Repeat*-Modul zusammengesetzt. Dies bietet den Vorteil, dass auf sequenzierte Plasmide zurückgegriffen werden kann und während der Klonierung keine Amplifikationsschritte mehr erfolgen. Im Gegensatz dazu wird in Zhang et al. (2011) eine PCR-basierte Bibliothek der *Repeat*-Monomere verwendet, deren Herstellung PCR-, Restriktions- und zwei Reinigungsschritte beinhaltet. Daher kann nicht ausgeschlossen werden, dass es während der Erstellung immer wieder zu Fehlern durch die Polymerase kommen kann. Ein weiteres Problem stellen Fehler innerhalb der Oligonukleotidsequenzen dar, sie die Restriktionsenzym-Erkennungsstelle bzw. den Überhang verändern. Folglich werden *Repeat*-Monomere entweder nicht gespalten oder sie besitzen einen fehlerhaften Überhang, wodurch sie falsch oder gar nicht ligiert werden.

Nur eine Plasmid-basierte *Repeat*-Monomer Bibliothek stellt nach einmaliger Erstellung und Validierung eine dauerhafte Basis für die Erstellung von *Repeat*-Modulen dar.

3.2.5.2 Repeat-Monomer-basierte Mehrfragment- und Zweifragment-Ligationen unterscheiden sich hinsichtlich der Dauer der Vorarbeiten, der Repeat-Modul-Erstellung und der Flexibilität der Repeat-Länge

Repeat-Monomer-basierte Mehrfragment-Ligationen benötigen mehr Vorarbeiten als *Repeat*-Monomer-basierte Zweifragment-Ligationen

Ein großer Unterschied zwischen der in der vorliegenden Arbeit entwickelten Mehrfragment-Ligation und der Zweifragment-Ligation (Abb. 11) z.B. aus Huang et al. (2011) besteht in der Anzahl der notwendigen Plasmide. Um jedes Nukleotid der DNA spezifisch adressieren zu können, benötigt man für vier verschiedene RVDs (HD, NI, NG, NN) jeweils ein *Repeat*-Monomer und einen Zielvektor. Nach der Erstellung dieser acht Plasmide können TALE-basierte DNA-Bindedomänen generiert werden.

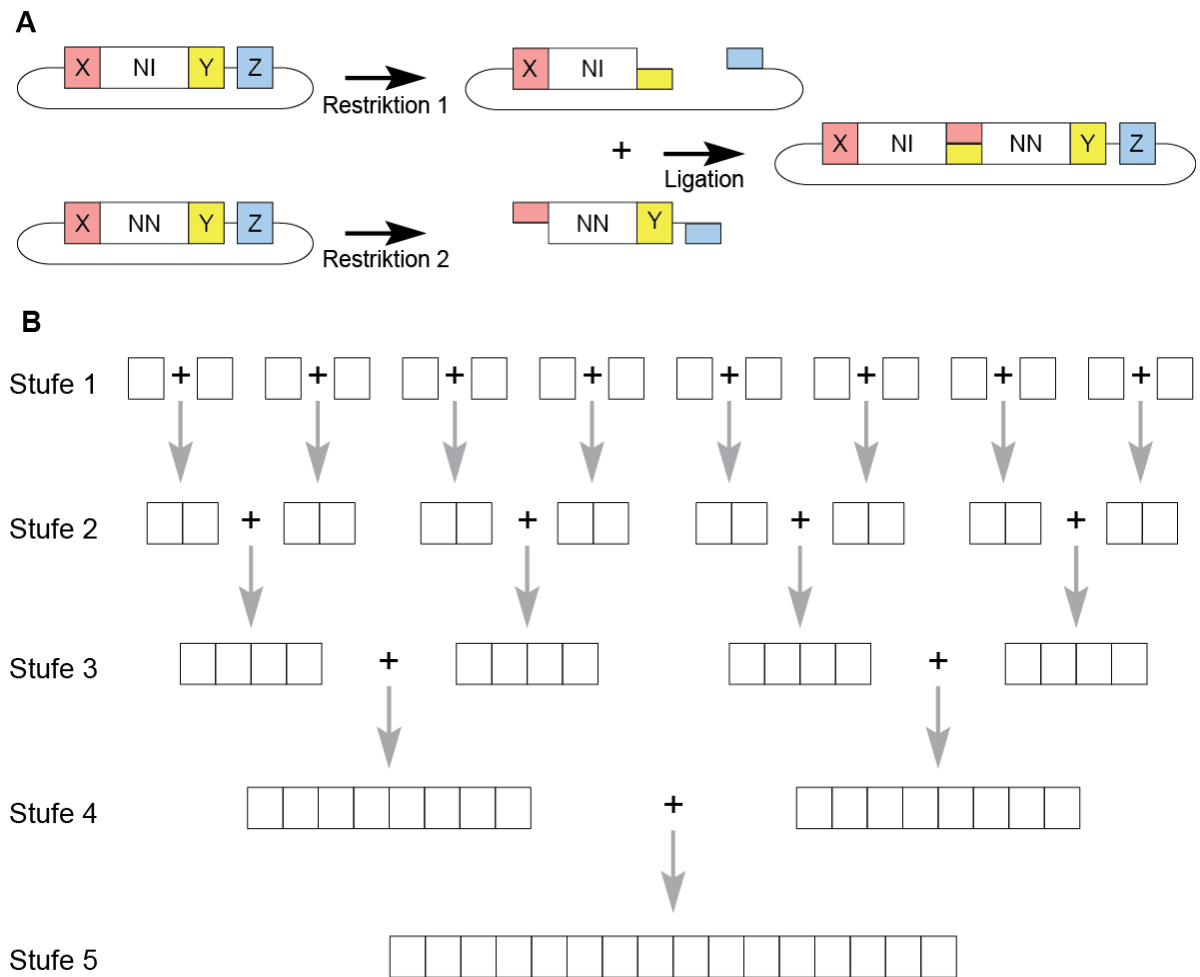


Abb. 11: Prinzip der Zweifragment-Ligation. Schematische Darstellung der Zweifragment-Ligation. (A) Die Assemblierung eines *Repeat*-Moduls kann durch die schrittweise Ligation von zwei Fragmenten geschehen (Zweifragment-Ligation). Die Zweifragment-Ligation beruht auf der Verwendung von Fragmenten (Rechtecke mit den kodierten RVDs z.B. NI und NN), die durch stille Nukleotidveränderungen als Teil der kodierenden Sequenz 5' (X) und 3' (Y) von Restriktionsenzymen flankiert sind. Beide Enzyme sind Isocaudomere und generieren sequenzidentische 4 bp-Überhänge. Nach Ligation der X/Y-Spaltprodukte ist die Erkennungssequenz für beide Enzyme zerstört. Des Weiteren muss 3' zur Erkennungsstelle für das Restriktionsenzym Y eine weitere Restriktionsenzym-Erkennungsstelle (Z) positioniert sein. Ausgehend von den *Repeat*-Monomeren können in einem ersten Restriktions- und anschließendem Ligations-Schritt *Repeat*-Dimere erzeugt werden. Dafür wird das Plasmid des N-terminalen Monomers mit den Restriktionsenzymen Y und Z und das des C-terminalen Monomers mit den Restriktionsenzymen X und Z gespalten. Nach der Ligation der Spaltprodukte über den durch die Restriktionsenzyme X/Y entstandenen Überhang und den durch das Restriktionsenzym Z entstandenen Überhang ist ein *Repeat*-Dimer entstanden, welches in der nächsten Stufe über die gleiche Strategie mit einem anderen *Repeat*-Monomer oder vorher generierten *Repeat*-Dimer bzw. -Polymer ligiert werden kann. (B) Schematische Darstellung der in mehreren Stufen verlaufenden Ligation eines *Repeat*-Moduls.

Im Gegensatz dazu benötigt die Mehrfragment-Ligation bedeutend mehr Vorarbeiten. Durch die parallele Klonierung mehrerer *Repeat*-Monomere zu einem *Repeat*-Polymer wird für jede Position im *Repeat*-Polymer ein spezifischer *Repeat*-Monomer benötigt. Daher sind für das im Rahmen der vorliegenden Arbeit generierte System zwölf *Repeat*-Monomere und ein Zielvektor pro RVD notwendig (M2 Abb. 1).

Dies führt bei vier verschiedenen RVDs zu 52 benötigten Plasmiden, deren Erstellung mehr Zeit erfordert als die für die acht Plasmide der Zweifragment-Ligation. Auch die Erweiterung der Klonierungssysteme hinsichtlich der verwendeten RVDs ist bei der Huang et al. (2011)-basierten Zweifragment-Ligation mit je zwei Plasmiden pro zusätzlichen RVD schneller. Im Vergleich dazu benötigt das in der vorliegenden Arbeit beschriebene System 13 zusätzliche Plasmide.

Durch die geringeren Vorbereitungszeiten zur Erstellung der *Repeat*-Monomere bietet die in Huang et al. (2011) beschriebene Zweifragment-Ligation die Möglichkeit, mit sehr wenigen Plasmiden und daher ohne lange Vorarbeiten relativ schnell einige *dTALEs* zu erstellen.

Repeat-Monomer-basierte *Repeat*-Module lassen sich über Mehrfragment-Ligationen schneller als über Zweifragment-Ligationen erstellen

Der Hauptunterschied zwischen Zwei- und Mehrfragment-Ligationen, die auf *Repeat*-Monomeren beruhen, besteht in der Anzahl der nötigen Klonierungsschritte, um ein *Repeat*-Modul definierter Länge zu generieren (Tabelle 1).

Bei der im Rahmen der vorliegenden Arbeit etablierten Methode zur Assemblierung eines 17.5 *Repeat*-Moduls sind theoretisch drei Restriktions-Ligations-Reaktionen, drei Transformationen und drei Plasmidpräparationen notwendig. Inklusive der anschließenden Klonierung der *Repeat*-Polymere in einen Zielvektor sind dafür nur fünf Tage notwendig. Im Gegensatz dazu sind für die in Huang et al. (2011) und Sander et al. (2011) vorgestellten Zweifragment-Ligation 13 Tage notwendig, um in sechs Stufen 17.5 *Repeat*-DNA-Bindedomänen in einen Zielvektor zu klonieren (Abb. 11). Dies ist fast dreimal so lang und bedarf zudem ein vielfaches an Klonierungsschritten verglichen mit der hier vorgestellten Mehrfragment-Ligation (Tabelle 1).

Tabelle 1: Vergleich der Mehrfragment-Ligation nach Morbitzer et al. (2011) und Zweifragment-Ligation nach Huang et al. (2011) hinsichtlich Dauer und Zahl der Klonierungsschritte für die Erstellung eines TALEs mit 17.5 *Repeats*.

Zahl der -Ligation	benötigten Tage	Restriktionen	Ligationen	Transfor- mationen	Plasmidprä- parationen
Mehrfragment-	5		3	3	3
Zweifragment-	13	33	33	16	16

Zwar halbiert sich die Anzahl der zu bearbeitenden *Repeat*-Polymere in jeder Stufe der Zweifragment-Ligation (Abb. 11), allerdings ist die hohe Zahl an Reaktionen zu Beginn der *Repeat*-Modul-Erstellung ein Hemmnis für die parallele Erstellung vieler dTALEs.

Der notwendige hohe Arbeitsaufwand macht sie für die parallele und automatisierte Herstellung einer großen Anzahl von TALE-basierten DNA-Bindedomänen ungeeignet. Dieser Nachteil besteht bei der Mehrfragment-Ligation nicht. Nach den notwendigen Vorarbeiten kann mit einem überschaubaren Arbeitsaufwand parallel eine Vielzahl von TALE-basierten DNA-Bindedomänen erstellt werden.

Die *Repeat*-Länge ist bei Mehrfragment-Ligationen im Vergleich zu Zweifragment-Ligationen stärker vorbestimmt

Die Variabilität der *Repeat*-Anzahl ist besonders für die Anwendung von TALE-Bindedomänen im Kontext von TALE-Nukleasen wichtig, da hier zwei Bindedomänen in einem definierten Abstand zueinander binden müssen (Christian et al., 2010; Li et al., 2011a). Dieser setzt sich aus der Länge der beiden Bindedomänen und der sogenannten *linker* Region zwischen diesen zusammen. Für die Vermeidung von *off-target*-Restriktionen ist es wichtig, Nukleasen zu verwenden, die einen definierten und engen Aktivitätsbereich hinsichtlich ihres Abstandes besitzen. Einzig die Notwendigkeit des T₋₁ vor beiden Erkennungsstellen führt zu einer Einschränkung der möglichen Bindestellen. Bezogen auf einen definierten Schnittpunkt der TALE-Nuklease ist die Wahrscheinlichkeit für ein T an der Position T₋₁ auf beiden Seiten 6,25% ($1/4 \cdot 1/4$). Ist allerdings die Länge der *Repeats* um ein weiteres *Repeat* flexibel, so vervierfacht sich die Wahrscheinlichkeit (25% $2/4 \cdot 2/4$) für eine TALE-Nuklease-Bindestelle bezogen auf einen definierten Schnittpunkt. Ab vier verschiedenen *Repeat*-Längen beträgt die Wahrscheinlichkeit 100%, auf jedes Nukleotid ein TALE-Nuklease-Paar generieren zu können.

Bei der Mehrfragment-Ligation ist die Anzahl der *Repeats* des *Repeat*-Moduls durch das Design festgelegt. Die *Repeat*-Anzahl ergibt sich dabei aus der Summe der in den *Repeat*-Polymeren verwendeten *Repeat*-Monomere. Das hier verwendete Design erlaubt, *Repeat*-Module mit 18 bzw. 21 *Repeats* aus zwei *Repeat*-Polymeren herzustellen. In Cermak et al. (2011) und Geissler et al. (2011) können *Repeat*-Polymere mit einer flexiblen Anzahl an Monomeren erzeugt werden, so dass die Länge der *Repeat*-Module variabel für 12 - 31 bzw. 1 - 24 Nukleotide ist. Auch das in

der vorliegenden Arbeit vorgestellte System wurde nach seiner Veröffentlichung zur Generierung von *Repeat*-Modulen für die Erkennung von theoretisch 2 - 32 Nukleotiden erweitert (Elsaesser, Morbitzer, Ramirez und Lahaye, unveröffentlichte Daten).

Im Gegensatz zur Mehrfragment-Ligation bietet das System der Zweifragment-Ligationen keine Beschränkungen hinsichtlich der Anzahl der *Repeat*-Monomere. Da jedes *Repeat*-Polymer immer wieder als Ausgangspunkt für einen weiteren Restriktions-Ligations-Schritt dienen kann, ist die Länge des *Repeat*-Moduls zwischen zwei und theoretisch unendlich vielen *Repeats* frei wählbar.

Um wenige TALE-basierte DNA-Bindedomänen zu erstellen ist die Zweifragment-Ligation optimal, da diese mit sehr geringen Vorarbeiten und gleichzeitig großer Flexibilität hinsichtlich der Länge erstellt werden können. Sollen allerdings regelmäßig eine größere Anzahl an Bindedomänen parallel erstellt werden, ist diese Methode ungeeignet. Dann empfiehlt sich trotz größerer Vorarbeiten die Mehrfragment-Ligation, da sie eine geringere Anzahl von nacheinander ablaufenden Klonierungen beinhaltet und damit insgesamt viel schneller durchführbar ist.

3.2.5.3 Durch die Verwendung von Repeat-Polymeren nähern sich Mehrfragment- und Zweifragment-Ligation hinsichtlich der Dauer der Repeat-Modul-Erstellung und der Vorarbeiten an

Die Verwendung von *Repeat*-Polymeren führt zur Beschleunigung der Mehrfragment- und Zweifragment-Ligations-basierten *Repeat*-Modul-Erstellung

Der entscheidende Vorteil der Mehrfragment-Ligation gegenüber der Zweifragment-Ligation ist die geringere Dauer und die geringere Anzahl an notwendigen Klonierungsschritten (Tabelle 2). Eine Möglichkeit den Zeitaufwand für die Mehrfragment-Ligation zu verringern, besteht in der Verwendung von *Repeat*-Dimeren anstelle von *Repeat*-Monomeren (Li et al., 2012). Dafür wurde eine Bibliothek aus allen RVD-Kombinationen von zwei aufeinanderfolgenden *Repeat*-Monomeren erstellt (*Repeats* 1 und 2, *Repeats* 3 und 4, usw.). Dadurch halbiert sich die Anzahl der zu ligierenden Fragmente und es sollte möglich sein, ein *Repeat*-Modul aus 17.5 *Repeats* in einem Schritt zu klonieren wodurch sich die Dauer von fünf Tagen (M2) auf drei Tage verkürzt. Da für 17.5 *Repeats* schon zehn Fragmente zu ligieren sind und bei mehr als zehn Fragmenten die Ligationseffizienz stark sinkt,

ist es schwer möglich, längere *Repeat*-Module in einem einstufigen Verfahren herzustellen. In Li et al., (2012) wurde mit diesem System allerdings nur DNA-Bindedomänen mit 12.5 *Repeats* erstellt.

Tabelle 2: Vergleich der Monomer-, Dimer- bzw. Polymer-basierten Mehrfragment- und Zweifragment-Ligationen zur Erstellung eines TALEs mit 17.5 *Repeats*.

Mehrfragment-Ligation			Zweifragment-Ligation	
Monomer (Morbitzer et al. 2011)	Dimer (Li et al. 2012)		Monomer (Huang et al. 2011)	Polymer (Reyon et al. 2012)
3	1	Anzahl der: Klonierungen	16	4
2	1	Klonierungs- schritte	5	3
5	3	Tage für die TALE-Klonierung	13	7

Repeat-Polymere wurden auch im Rahmen von Zweifragment-Ligationen verwendet (Reyon et al., 2012). Durch die Verwendung von *Repeat*-Tetrameren anstatt *Repeat*-Monomeren lässt sich ein 17.5 *Repeat*-Modul auch innerhalb von 7 statt 13 Tagen in einem Zielvektor assemblieren (Tabelle 2), da bei der Erstellung die Stufen 1 und 2 gespart werden können (Abb. 11). Allerdings werden damit immer noch zwei Tage mehr benötigt, als bei der in der vorliegenden Arbeit beschriebenen Mehrfragment-Ligation.

Durch die Verwendung von *Repeat*-Polymer-Plasmid-Bibliotheken verkürzt sich der zeitliche Aufwand bei beiden Ligationen deutlich. Allerdings ist diese Methode nicht geeignet, *Repeat*-Module mit viel mehr als 17.5 *Repeats* in einer Mehrfragment-Ligation zu erstellen.

Repeat-Polymer basierte Zweifragment-Ligationen benötigen im Vergleich zu Mehrfragment-Ligationen ein Vielfaches an Plasmiden

Um den zeitlichen Nachteil der Zweifragment-Ligation bei der Erstellung von *Repeat*-Modulen gegenüber der in der vorliegenden Arbeit beschriebenen Mehrfragment-Ligation zu kompensieren, müssen *Repeat*-Polymere als Ausgangspunkt benutzt werden.

Basierend auf den in Sander et al. (2011) generierten *Repeat*-Monomeren wurde in Reyon et al. (2012) eine Plasmid-Bibliothek erstellt, die alle theoretisch möglichen RVD-Kombinationen für *Repeat*-Dimere, *Repeat*-Trimere und *Repeat*-Tetramere enthält und damit 376 Plasmide umfasst. Die Implementierung von weiteren RVDs

führt zu einer enormen Erhöhung der notwendigen Plasmide (Tabelle 3). Für jeden weiteren RVD wird die Anzahl der zu generierenden Plasmide fast verdoppelt. Bei sechs verschiedenen RVDs, wie sie in der vorliegenden Arbeit verwendet wurden, wären dann 17 *96-well* Platten notwendig, um die 1632 Plasmide zu lagern. Die Erstellung der Plasmid-Bibliothek ist jedoch mit erheblichem Mehraufwand verbunden und die dafür notwendigen *96-well* Platten sind manuell schwer zu handhaben. Daher basiert die in Reyon et al. (2012) beschriebene Methode auf der Verwendung von Pipettierrobotern, die allerdings den wenigsten Anwendern zur Verfügung stehen und deren Anschaffung nur für die kommerzielle Herstellung von TALE-basierten DNA-Bindedomänen sinnvoll ist.

Auch für die *Repeat*-Polymer-basierte Mehrfragment-Ligation erhöht sich die Anzahl der benötigten Plasmide etwa um das drei- bis vierfache (Tabelle 3) und erschwert die manuelle Handhabbarkeit.

Tabelle 3: Anzahl der bei der Mehrfragment- bzw. Zweifragment-Ligation benötigten Plasmide.

Mehrfragment-Ligation		Anzahl der Plasmide bei	Zweifragment-Ligation	
Monomer (Morbitzer et al. 2011)	Dimer (Li et al. 2012)		Monomer (Huang et al. 2011)	Polymer (Reyon et al. 2012)
52	144	4 RVDs	8	376
65	225	5 RVDs	10	835
78	324	6 RVDs	12	1632

Die Verwendung von *Repeat*-Polymeren beschleunigt sowohl die Mehrfragment- als auch die Zweifragment-Ligation. Allerdings ist für die Klonierung der Polymer-Bibliotheken ein hoher zeitlicher Aufwand für Vorarbeiten notwendig. Besonders bei der Zweifragment-Ligation steigt die Zahl der notwendigen Plasmide bei der Hinzunahme weiterer RVDs stark an, so dass zusätzlich zu den Vorarbeiten auch die manuelle Handhabung nur begrenzt möglich ist und diese Methode dadurch relativ anfällig für Fehler wird.

Die Festphasenkopplung hebt die Nachteile der Zweifragment-Ligation hinsichtlich der Dauer der *Repeat*-Modul Erstellung auf

Der zeitliche Nachteil der Zweifragment-Ligation gegenüber der Mehrfragment-Ligation kommt durch die schrittweise Assemblierung der *Repeat*-Polymere und die

notwendigen Zwischenklonierungen zustande. Dieser Nachteil wird durch die Biotin-vermittelte Kopplung des ersten *Repeat*-Moduls an eine Streptavidinmatrix aufgehoben. Dadurch kann die Dauer der Zweifragment-Ligation auf drei Tage verkürzt werden, da alle Ligationen nacheinander, aber ohne Zwischenklonierungsschritte erfolgen können (Reyon et al., 2012). In 90% der Fälle reichten sechs Plasmidpräparationen aus, um einen positiven Klon zu erhalten, so dass die Effizienz mit der in der vorliegenden Arbeit beschriebenen Mehrfragment-Ligation vergleichbar ist. Die in Reyon et al. (2012) beschriebene Methode ist durch die notwendigen Restriktions-, Reinigungs- und Ligationsschritte jedoch sehr arbeitsintensiv und lässt sich nur durch den Einsatz von Pipettierrobotern für die parallele Erstellung vieler TALE-basierter Bindedomänen verwenden.

Durch die Biotin-vermittelte Festphasenkopplung und das Entfallen der Zwischenklonierungsschritte, bietet die in Reyon et al. (2012) beschriebene Methode eine in Bezug auf Zeit und Effizienz mit der *Repeat*-Dimer basierten Mehrfragment-Ligation nach Li et al. (2012) vergleichbare Methode.

Für einen Anwender der TALEs mit 17.5 *Repeats* generieren will, eignet sich besonders die *Repeat*-Dimer basierte Mehrfragment-Ligation. Diese erlaubt es mit einem Minimum an menschlichem Einsatz, Laborausstattung und Zeit bei einer noch handhabbaren Anzahl an Plasmiden TALE-basierte DNA-Bindedomänen zu generieren. Die einzige Limitation dieser Methode liegt in der Anzahl der zu ligierenden *Repeats*, allerdings sind meines Erachtens 17.5 *Repeat* bzw. u.U. bis zu 21.5 *Repeats* für die meisten Anwendungen ausreichend. Der Vorteil der Festphasen gekoppelte Zweifragment-Ligation besteht in der Flexibilität hinsichtlich der Länge des *Repeat*-Moduls. Allerdings ist dafür ein hohes Maß an Vorarbeiten für die manuell schwer zu handhabende Plasmid-Bibliothek nötig. Zudem ist diese Methode ohne die Verwendung von Pipettierrobotern sehr arbeitsintensiv, wodurch die parallele Erstellung vieler *Repeat*-Module stark eingeschränkt ist.

4 Literaturverzeichnis

Arnould, S., Chames, P., Perez, C., Lacroix, E., Duclert, A., Epinat, J.C., Stricher, F., Petit, A.S., Patin, A., Guillier, S., *et al.* (2006). Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J Mol Biol* 355, 443-458.

Arnould, S., Perez, C., Cabaniols, J.P., Smith, J., Gouble, A., Grizot, S., Epinat, J.C., Duclert, A., Duchateau, P., und Paques, F. (2007). Engineered I-Crel derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. *J Mol Biol* 371, 49-65.

Ashworth, J., Havranek, J.J., Duarte, C.M., Sussman, D., Monnat, R.J., Jr., Stoddard, B.L., und Baker, D. (2006). Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature* 441, 656-659.

Baker, M. (2012). Gene-editing nucleases. *Nat Methods* 9, 23-26.

Beerli, R.R., und Barbas, C.F., 3rd (2002). Engineering polydactyl zinc-finger transcription factors. *Nat Biotechnol* 20, 135-141.

Berg, J.M. (1988). Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. *P Natl Acad Sci USA* 85, 99-102.

Bibikova, M., Golic, M., Golic, K.G., und Carroll, D. (2002). Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics* 161, 1169-1175.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., und Bonas, U. (2009). Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science* 326, 1509-1512.

Bochtler, M. (2012). Structural basis of the TAL effector–DNA interaction. *Biological Chemistry* 393, 1055-1066.

Bogdanove, A.J., Schornack, S., und Lahaye, T. (2010). TAL effectors: finding plant genes for disease and defense. *Curr Opin Plant Biol* 13, 394-401.

Brayer, K.J., Kulshreshtha, S., und Segal, D.J. (2008). The protein-binding potential of C₂H₂ zinc finger domains. *Cell Biochem Biophys* 51, 9-19.

Brayer, K.J., und Segal, D.J. (2008). Keep your fingers off my DNA: protein-protein interactions mediated by C₂H₂ zinc finger domains. *Cell Biochem Biophys* 50, 111-131.

Büttner, D., und Bonas, U. (2010). Regulation and secretion of *Xanthomonas* virulence factors. *Fems Microbiol Rev* 34, 107-133.

Capecchi, M.R. (1989). Altering the Genome by Homologous Recombination. *Science* 244, 1288-1292.

Carroll, D. (2011). Zinc-finger nucleases: a panoramic view. *Curr Gene Ther* 11, 2-10.

Cass, D., Hotchko, R., Barber, P., Jones, K., Gates, D.P., und Berglund, J.A. (2011). The four Zn fingers of MBNL1 provide a flexible platform for recognition of its RNA binding elements. *BMC Mol Biol* 12, 20.

Cathomen, T., und Joung, J.K. (2008). Zinc-finger nucleases: The next generation emerges. *Mol Ther* 16, 1200-1207.

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J., und Voytas, D.F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* 39, e82

Chapdelaine, P., Pichavant, C., Rousseau, J., Paques, F., und Tremblay, J.P. (2010). Meganucleases can restore the reading frame of a mutated dystrophin. *Gene Ther* 17, 846-858.

Chevalier, B., Turmel, M., Lemieux, C., Monnat, R.J., Jr., und Stoddard, B.L. (2003). Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-Crel and I-Msol. *J Mol Biol* 329, 253-269.

Chevalier, B.S., Kortemme, T., Chadsey, M.S., Baker, D., Monnat, R.J., und Stoddard, B.L. (2002). Design, activity, and structure of a highly specific artificial endonuclease. *Mol Cell* 10, 895-905.

Choo, Y., und Klug, A. (1994a). Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. *P Natl Acad Sci USA* 91, 11168-11172.

Choo, Y., und Klug, A. (1994b). Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage. *P Natl Acad Sci USA* 91, 11163-11167.

Choo, Y., Sanchez-Garcia, I., und Klug, A. (1994). In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence. *Nature* 372, 642-645.

Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J., und Voytas, D.F. (2010). Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics* 186, 757-U476.

Cong, L., Zhou, R.H., Kuo, Y.C., Cunniff, M., und Zhang, F. (2012). Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun* 3, 968.

Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.K., Shi, Y., und Yan, N. (2012). Structural Basis for Sequence-Specific Recognition of DNA by TAL Effectors. *Science* 335, 720-723.

Engler, C., Gruetzner, R., Kandzia, R., und Marillonnet, S. (2009). Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *Plos One* 4, e5553.

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

Epinat, J.C., Arnould, S., Chames, P., Rochaix, P., Desfontaines, D., Puzin, C., Patin, A., Zanghellini, A., Paques, F., und Lacroix, E. (2003). A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res.* 31, 2952-2962.

Fairall, L., Schwabe, J.W., Chapman, L., Finch, J.T., und Rhodes, D. (1993). The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature* 366, 483-487.

Furutani, A., Takaoka, M., Sanada, H., Noguchi, Y., Oku, T., Tsuno, K., Ochiai, H., und Tsuge, S. (2009). Identification of Novel Type III Secretion Effectors in *Xanthomonas oryzae* pv. *oryzae*. *Mol Plant Microbe In* 22, 96-106.

Ganss, B., und Jheon, A. (2004). Zinc finger transcription factors in skeletal development. *Crit Rev Oral Biol Med* 15, 282-297.

Gao, H., Smith, J., Yang, M., Jones, S., Djukanovic, V., Nicholson, M.G., West, A., Bidney, D., Falco, S.C., Jantz, D., *et al.* (2010). Heritable targeted mutagenesis in maize using a designed endonuclease. *The Plant journal : for cell and molecular biology* 61, 176-187.

Geissler, R., Scholze, H., Hahn, S., Streubel, J., Bonas, U., Behrens, S.E., und Boch, J. (2011). Transcriptional Activators of Human Genes with Programmable DNA-Specificity. *Plos One* 6, e19509.

Greisman, H.A., und Pabo, C.O. (1997). A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* 275, 657-661.

Grizot, S., Smith, J., Daboussi, F., Prieto, J., Redondo, P., Merino, N., Villate, M., Thomas, S., Lemaire, L., Montoya, G., *et al.* (2009). Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. *Nucleic acids Res.* 37, 5405-5419.

Gu, K., Yang, B., Tian, D., Wu, L., Wang, D., Sreekala, C., Yang, F., Chu, Z., Wang, G.L., White, F.F., *et al.* (2005). R gene expression induced by a type-III effector triggers disease resistance in rice. *Nature* 435, 1122-1125.

Hartley, J.L., Temple, G.F., und Brasch, M.A. (2000). DNA cloning using *in vitro* site-specific recombination. *Genome Res* 10, 1788-1795.

Hockemeyer, D., Wang, H.Y., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller, J.C., *et al.* (2011). Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol* 29, 731-734.

Huang, P., Xiao, A., Zhou, M.G., Zhu, Z.Y., Lin, S., und Zhang, B. (2011). Heritable gene targeting in zebrafish using customized TALENs. *Nat Biotechnol* 29, 699-700.

Hurt, J.A., Thibodeau, S.A., Hirsh, A.S., Pabo, C.O., und Joung, J.K. (2003). Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *P Natl Acad Sci USA* 100, 12271-12276.

Isalan, M., Choo, Y., und Klug, A. (1997). Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *P Natl Acad Sci USA* 94, 5617-5621.

Isalan, M., Klug, A., und Choo, Y. (2001). A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. *Nat Biotechnol* 19, 656-660.

Jurica, M.S., Monnat, R.J., Jr., und Stoddard, B.L. (1998). DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-Crel. *Mol Cell* 2, 469-476.

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

Kim, Y.G., Cha, J., und Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *P Natl Acad Sci USA* 93, 1156-1160.

Klug, A. (2010). The Discovery of Zinc Fingers and Their Applications in Gene Regulation and Genome Manipulation. *Annu Rev Biochem* 79, 213-231.

Lee, H.J., Kim, E., und Kim, J.S. (2010). Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res* 20, 81-89.

Li, L.X., Piatek, M.J., Atef, A., Piatek, A., Wibowo, A., Fang, X.Y., Sabir, J.S.M., Zhu, J.K., und Mahfouz, M.M. (2012). Rapid and highly efficient construction of TALE-based transcriptional regulators and nucleases for genome modification. *Plant Mol Biol* 78, 407-416.

Li, T., Huang, S., Jiang, W.Z., Wright, D., Spalding, M.H., Weeks, D.P., und Yang, B. (2011a). TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.* 39, 359-372.

Li, T., Huang, S., Zhao, X.F., Wright, D.A., Carpenter, S., Spalding, M.H., Weeks, D.P., und Yang, B. (2011b). Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res.* 39, 6315-6325.

Liu, Q., Xia, Z., Zhong, X., und Case, C.C. (2002). Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *The Journal of biological chemistry* 277, 3850-3856.

Maeder, M.L., Thibodeau-Beganny, S., Osiaik, A., Wright, D.A., Anthony, R.M., Eichinger, M., Jiang, T., Foley, J.E., Winfrey, R.J., Townsend, J.A., *et al.* (2008). Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell* 31, 294-301.

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

Mahfouz, M.M., Li, L.X., Shamimuzzaman, M., Wibowo, A., Fang, X.Y., und Zhu, J.K. (2011). De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *P Natl Acad Sci USA* 108, 2623-2628.

Mak, A.N., Bradley, P., Cernadas, R.A., Bogdanove, A.J., und Stoddard, B.L. (2012). The crystal Structure of TAL Effector *PthXo1* bound to its DNA Target. *Science* 335, 716-719.

Messina, D.N., Glasscock, J., Gish, W., und Lovett, M. (2004). An ORFeome-based analysis of human transcription factor genes and the construction of a microarray to interrogate their expression. *Genome Res* 14, 2041-2047.

Miller, J., McLachlan, A.D., und Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *The EMBO journal* 4, 1609-1614.

Miller, J.C., Tan, S.Y., Qiao, G.J., Barlow, K.A., Wang, J.B., Xia, D.F., Meng, X.D., Paschon, D.E., Leung, E., Hinkley, S.J., *et al.* (2011). A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* 29, 143-U149.

Moore, J.K., und Haber, J.E. (1996). Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*. *Mol Cell Biol* 16, 2164-2173.

Moore, M., Klug, A., und Choo, Y. (2001). Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units. *P Natl Acad Sci USA* **98**, 1437-1441.

Morbitzer, R., Elsaesser, J., Hausner, J., und Lahaye, T. (2011). Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Res.* **39**, 5790-5799.

Morbitzer, R., Römer, P., Boch, J., und Lahaye, T. (2010). Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *P Natl Acad Sci USA* **107**, 21617-21622.

Moscou, M.J., und Bogdanove, A.J. (2009). A Simple Cipher Governs DNA Recognition by TAL Effectors. *Science* **326**, 1501-1501.

Mussolino, C., Morbitzer, R., Lutge, F., Dannemann, N., Lahaye, T., und Cathomen, T. (2011). A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res.* **39**, 9283-9293.

Papworth, M., Moore, M., Isalan, M., Minczuk, M., Choo, Y., und Klug, A. (2003). Inhibition of herpes simplex virus 1 gene expression by designer zinc-finger transcription factors. *P Natl Acad Sci USA* **100**, 1621-1626.

Pavletich, N.P., und Pabo, C.O. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **252**, 809-817.

Pingoud, A., und Wende, W. (2011). Generation of novel nucleases with extended specificity by rational and combinatorial strategies. *Chembiochem* **12**, 1495-1500.

Ramirez, C.L., Foley, J.E., Wright, D.A., Muller-Lerch, F., Rahman, S.H., Cornu, T.I., Winfrey, R.J., Sander, J.D., Fu, F., Townsend, J.A., *et al.* (2008). Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat Methods* **5**, 374-375.

Reynolds, L., Ullman, C., Moore, M., Isalan, M., West, M.J., Clapham, P., Klug, A., und Choo, Y. (2003). Repression of the HIV-1 5' LTR promoter and inhibition of HIV-1 replication by using engineered zinc-finger transcription factors. *P Natl Acad Sci USA* 100, 1615-1620.

Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D., und Joung, J.K. (2012). FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol.* 30, 460-465.

Roden, J.A., Belt, B., Ross, J.B., Tachibana, T., Vargas, J., und Mudgett, M.B. (2004). A genetic screen to isolate type III effectors translocated into pepper cells during *Xanthomonas* infection. *P Natl Acad Sci USA* 101, 16624-16629.

Rohs, R., Jin, X.S., West, S.M., Joshi, R., Honig, B., und Mann, R.S. (2010). Origins of Specificity in Protein-DNA Recognition. *Annu Rev Biochem* 79, 233-269.

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

Römer, P., Recht, S., und Lahaye, T. (2009a). A single plant resistance gene promoter engineered to recognize multiple TAL effectors from disparate pathogens. *P Natl Acad Sci USA* 106, 20526-20531.

Römer, P., Strauss, T., Hahn, S., Scholze, H., Morbitzer, R., Grau, J., Bonas, U., und Lahaye, T. (2009b). Recognition of AvrBs3-like proteins is mediated by specific binding to promoters of matching pepper *Bs3* alleles. *Plant Physiol* 150, 1697-1712.

Sander, J.D., Cade, L., Khayter, C., Reyon, D., Peterson, R.T., Joung, J.K., und Yeh, J.R.J. (2011). Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat Biotechnol* 29, 697-698.

Sera, T. (2009). Zinc-finger-based artificial transcription factors and their applications. *Adv Drug Deliver Rev* 61, 513-526.

Silva, G., Poirot, L., Galetto, R., Smith, J., Montoya, G., Duchateau, P., und Paques, F. (2011). Meganucleases and Other Tools for Targeted Genome Engineering: Perspectives and Challenges for Gene Therapy. *Curr Gene Ther* 11, 11-27.

Smith, J., Bibikova, M., Whitby, F.G., Reddy, A.R., Chandrasegaran, S., und Carroll, D. (2000). Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res.* 28, 3361-3369.

Smith, J., Grizot, S., Arnould, S., Duclert, A., Epinat, J.C., Chames, P., Prieto, J., Redondo, P., Blanco, F.J., Bravo, J., *et al.* (2006). A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res.* 34, e149.

Stoddard, B.L. (2011). Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. *Structure* 19, 7-15.

Sugio, A., Yang, B., Zhu, T., und White, F.F. (2007). Two type III effector genes of *Xanthomonas oryzae* pv. *oryzae* control the induction of the host genes OsTFIIAgamma1 and OsTFX1 during bacterial blight of rice. *P Natl Acad Sci USA* 104, 10720-10725.

Tesson, L., Usal, C., Menoret, S., Leung, E., Niles, B.J., Remy, S., Santiago, Y., Vincent, A.I., Meng, X.D., Zhang, L., *et al.* (2011). Knockout rats generated by embryo microinjection of TALENs. *Nat Biotechnol* 29, 695-696.

Thieme, F., Koebnik, R., Bekel, T., Berger, C., Boch, J., Buttner, D., Caldana, C., Gaigalat, L., Goesmann, A., Kay, S., *et al.* (2005). Insights into genome plasticity and pathogenicity of the plant pathogenic bacterium *Xanthomonas campestris* pv. *vesicatoria* revealed by the complete genome sequence. *J Bacteriol* 187, 7254-7266.

Urnov, F.D., Miller, J.C., Lee, Y.L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D., und Holmes, M.C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* 435, 646-651.

Webb, S. (2011). Genome editing goes global. *BioTechniques* 51, 371-373.

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

Wood, A.J., Lo, T.W., Zeitler, B., Pickle, C.S., Ralston, E.J., Lee, A.H., Amora, R., Miller, J.C., Leung, E., Meng, X.D., *et al.* (2011). Targeted Genome Editing Across Species Using ZFNs and TALENs. *Science* 333, 307-307.

Yang, B., Sugio, A., und White, F.F. (2006). Os8N3 is a host disease-susceptibility gene for bacterial blight of rice. *P Natl Acad Sci USA* 103, 10503-10508.

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

5 Anhang

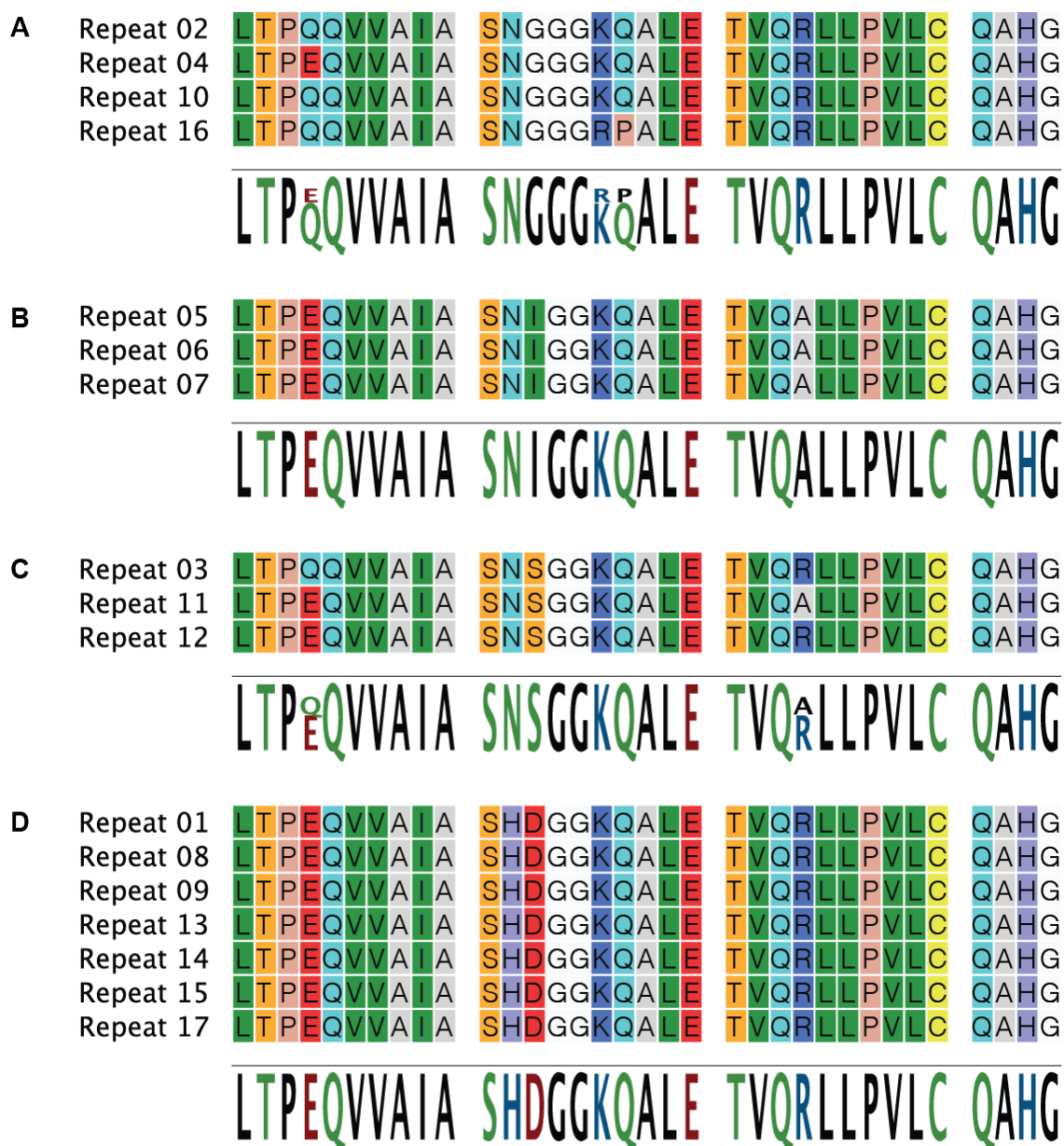


Abb. 12: Aminosäuresequenz der *Repeats* von AvrBs3 sortiert nach RVD-Typ. Um die Nicht-RVDs der *Repeats* eines RVD-Typs (A-NG, B-NI, C-NS, und D-HD) zu vergleichen, wurden die *Repeats* von AvrBs3 nach ihren RVDs sortiert. Unter den *Repeats* eines RVD-Typs ist die Konsensus Sequenz abgebildet.

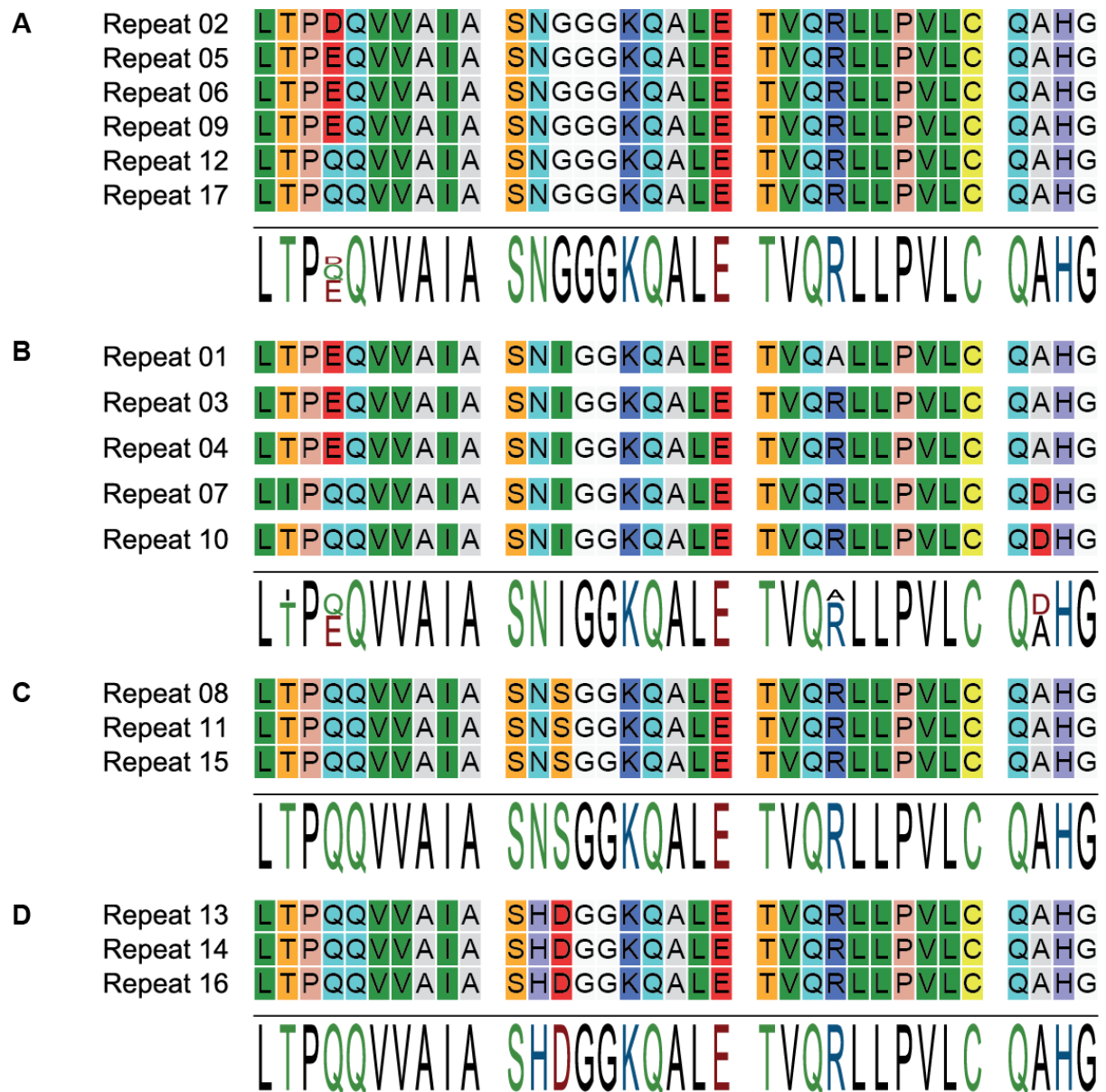


Abb. 13: Aminosäuresequenz der *Repeats* von AvrBs4 sortiert nach RVD-Typ. Um die Nicht-RVDs der *Repeats* eines RVD-Typs (A-NG, B-NI, C-NS, und D-HD) zu vergleichen, wurden die *Repeats* von AvrBs4 nach ihren RVDs sortiert. Unter den *Repeats* eines RVD-Typs ist die Konsensus Sequenz abgebildet.

5 Anhang

avrBs3* 1 ATGGATCCATTAGGTCAGGACTCCTTCACTGCTAGGGAGCTTTGCTGGTCTCAG
avrBs3 1 ATGGATCCATTCTGTCGGCACCAAGTCTGCTCCGAGCTTCTGCCGGACCCAA

avrBs3* 61 CCAGATGGAGTTCAACCAACTGCAGATAGGGAGTTCTCCTCCTGCTGGAGGACCACT
avrBs3 61 CCCGATGGGTTTCAAGCCGACTGCAGATCCTGGGTTGCTCCGCTGCTGGCCGGCCCTG

avrBs3* 121 GATGGCTTCTCCTGCTAGGAGAACATGCTAGAACAGATGCTTCTCCTCCTGCTCCA
avrBs3 121 GATGGCTTCCCCTCGGCGGACGATGCTCCGGACCGGCTGCCATCTCCCTGCTCCC

avrBs3* 181 TCTCCTGCTTTCTCTGCTGGTTCAATTTCGATTGTTGAGACAGTTCGATCCTTCTTTG
avrBs3 181 TCACCTGGTTCTCGCGGGCAGCTTTCAGTGACCTGTTACGTTCAGTTCGATCCGTCACTT

avrBs3* 241 TTTAATACTTCACTTTTGGATTCTTGCCACCATTGGTGCTCATCATACTGAGGCTGCT
avrBs3 241 TTTAATACATCGCTTTTGGATTCAATGCTCCCTTCGGCGCTCACCATACAGAGGCTGCC

avrBs3* 301 ACAGTGAATGGGATGAGTTCAATCTGGACTTAGAGCTGCTGATGCTCCTCCTCCTACT
avrBs3 301 ACAGCGAGTGGGATGAGGTCAATCGGTCTGCGGCGAGCCGACGCCCCACCAC

avrBs3* 361 ATGAGAGTTGCTGTTACAGCTGCTAGGCCACCAAGAGCTAAGCCTGCTCCTAGGAGAA
avrBs3 361 ATGCGCGTGGCTGTACTGCCCGCGGCGCGCGCAAGCCGGCGCGGACGACT

avrBs3* 421 GCTGCTCAACCATCTGATGCTTCTCCTGCTGCTCAAGTTGATCTTAGAACATTGGGATA
avrBs3 421 GCTGCGCAACCTCCGACGCTTCGCGCGCGCAGGTGGATCTACGACGCTCGGCTAC

avrBs3* 481 TCACAACAACAGCAGGAGAAGATTAGCCTAAGGTTAGATCTACAGTGGTCAACATCAT
avrBs3 481 AGCCAGCAGCAACAGGAGAAGATCAACCCGAAGTTCTTCGACAGTGGCGAGCACAC

avrBs3* 541 GAAGCTCTTGTGGTCAATGTTTCACTCATGCTCATATGTTGCTCTTCTCAACATCCT
avrBs3 541 GAGGCACTGCTCGGCATGGGTTTACACACGCGCACATCGTTGCGCTCAGC CAACACCG

avrBs3* 601 GCTGCTCTTGGTACTGTGCTGTTAAGTATCAGGATATGATTGCTGCTCTCCAGAGCT
avrBs3 601 GCAGCGTTAGGGACCGTCTGCTCAAGTATCAGGACATGATCGCAGCGTTGCCAGAGCG

avrBs3* 661 ACTCATGAAGCTATGTTGGTGTGGTAAAGCAGTGGTCTGGTGTAGCGCTCTTGAGGCC
avrBs3 661 ACACACGAAGCGATCGTTGGCGTCCGCAACAGTGGTCCGGCGACCGCTCTTGAGGCC

avrBs3* 721 TTGCTTACTGTTGCTGGGGAATTGAGAGGCCCTCCACTGCAATTAGATACAGGTCACTG
avrBs3 721 TTGCTCACGGTGGCGGAGAGTTGAGAGTCCACCCTTACAGTTGGACACAGGCCAACTT

avrBs3* 781 CTCAAGATAGCTAAACGGGGAGGAGTGACTGCCGTGGAAGCTGTGCATGCCGAGGAAT
avrBs3 781 CTCAAGATTGCAAAACGTTGGCGGTGACCGCAGTGGAGGCAGTGCATGCATGGCGCAAT

avrBs3* 841 GCTTTCAGCGGGGCAACCGCTGAATCTTACACCTGAGCAGGTTGTCCCAATCGCTAGTCAC
avrBs3 841 GCACTCAGCGGGTGCACCCTGAACCTGACCCGAGCAGGTTGGTGGCATCGCCAGCCAC

5 Anhang

avrBs3* 901 GATGGCGGCAAA CAGGCAT TGGAGAC CGTGCAACGC CTCTC CCAGTCTGTGTCAAGCT
avrBs3 901 GATGGCGGCAAG CAGGCCT TGGAGAC CGTGAGCGG CTGTG CCGGTGCTGTGCCAGGC

avrBs3* 961 CATGGA CTGACA CCGCAA CAAGTGGT AGCAATCGCA AGCAATGGCGA AGGGAAGCAGGC T
avrBs3 961 CATGGC CTGACC CCGCAG CAGGTGGT GGCATCGCC AGCAATGGCGT GGC AAGCAGGC G

avrBs3* 1021 CTTGAAACT GTACAGAGGT TGCTACCT GTCTATGCCAGGCG CACGGG CTGACACCACAA
avrBs3 1021 CTG GAGACGGTGCAGCGG CTGTTGCCGGT GCTGTGCCAGGC CATGGC CTGACC CCGCAG

avrBs3* 1081 CAAGTGTAGCT ATCGCTAGCAATAGTGGC GGA AAGCAGGCAT TGGAGACT GTTCAAAGG
avrBs3 1081 CAGGTGGTGGCCATCGCCAGCAATAGCGGTGGCAAGCAGGCCTGGAGAGGGTGCAGCGG

avrBs3* 1141 CTACTCCAGTTCTATGT CAGGCTCATGGA CTGACA CCTGAACAAGTGGT TGTATCGCT
avrBs3 1141 CTGTTGCCGGTGTGTGCCAGGCC CATGGC CTGACC CCGGAGCAGGTGGTGGCCATCGCC

avrBs3* 1201 AGCAACGGTGGTGGAAAGCAAGCACTTGAGACGGTCAACGACTCCTGCCCGTACTGTGC
avrBs3 1201 AGCAATGGCGGTGGCAAGCAGGCCTTGAGACGGTGACGCGGCTGTGCCGGTGTGTGC

avrBs3* 1261 CAAGTCATGGACTTAGCCCTGAGCAGGTAGTTGCTATGCTTCTAATATAGGAGGAAG
avrBs3 1261 CAGGCCATGGCCTGACCCCGAGCAGGTGGTGGCCATCGCCAGCAATATGGTGGCAAG

avrBs3* 1321 CAGGCTCTCGAAACTGTGCAAGCGCTCTACCTGTA CTATGCCAGGCTCATGGCCTTACC
avrBs3 1321 CAGGCCTGAGACGGTGCAGGCGTGTGCCGGTGTGTGCCAGGCCCATGGCCTGACC

avrBs3* 1381 CCAGAACAGGTAGTCCGAATTGCTTCGAATATCGGAGGTAAACAGGCTCTTGAGACCGTT
avrBs3 1381 CCGGAGCAGGTGGTGGCCATCGCCAGCAATATGGTGGCAAGCAGGCCTGGAGACGGT

avrBs3* 1441 CAAGCCTTGTTGCCAGTGTATGTCAAGCACATGGGCTAACTCCTGAGCAAGTTGTTGCA
avrBs3 1441 CAGGCCTTGTTGCCGTGCTGTGCCAGGCCCATGGCCTGACCCCGGAGCAGGTGGTGGCC

avrBs3* 1501 ATAGCTTCA AATATAGGGGAAAACAGCCCTTGAGACAGTCCAAGCCTTGTTGCCGGTCT
avrBs3 1501 ATGCCAGCAATATGGTGGCAAGCAGGCGCTGAGACGGTGCAGGCCTTGTTGCCGGT

avrBs3* 1561 CTATGCCAGGCACATGGAT TGACA CCGGAA CAGGTGTGCGATCGCTAGCCACGATGGT
avrBs3 1561 CTGTGCCAGGCCATGGCC TGACC CCGGAG CAGGTGTGCGCATCGCCAGCCACGATGGC

avrBs3* 1621 GGCAAGCAGGC ACTTGAAACAGTT CAGA GACTGTTACCAGTACTATGTC AAGCTCACGGA
avrBs3 1621 GGCAAGCAGGCCTGAGACGGTGCAGCGGCTGTTGCCGGTGTGTGCCAGGCCATGGC

avrBs3* 1681 CTCACCCCTGAA CAGGTGGTGGCGATCGCTTCA CACGATGGAGGCAAGCAGGCTCTAGAG
avrBs3 1681 CTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCGCTGGAG

avrBs3* 1741 ACTGTGCAGCGTCTTTTGCCAGTCTCTGT CAAGCCATGGAT TGACA CCAAGCAAGTT
avrBs3 1741 ACGGTGCAGCGCTGTTGCCGGTGTGTGCCAGGCCATGGCCTGACCCCGCAGCAGGT

avrBs3* 1801 GTAGCTATTGCGTCTAATGGTGGAGGTAAA CAAGCCCTGGAAACCGTCCAGAGACTTCTT
avrBs3 1801 GTGGCCATCGCCAGCAATGGCGGTGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTG

avrBs3* 1861 CCAAGTCTTTGTCAAGCACATGGACTCACACCTGAACAAGTTGTTGCTATCGCAAGCAAT
avrBs3 1861 CCGGTGCTGTGCCAGGCCATGGCCTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCAAT

avrBs3* 1921 TCTGGAGGTAAACAGGCGCTTGAGACAGTTCAAGCTCTCTTCCAGTGTCTGTCAAGCT
avrBs3 1921 AGCGGTGGCAAGCAGGCGCTGAGACGGTGCAGGCGCTGTGCCGGTGTGTGCCAGGC

avrBs3* 1981 CATGGA CTACTCCTGAA CAGGTGGTGGCTATCGCAAGCAACTCTGGAGGCAACAAGCA
avrBs3 1981 CATGGCCTGACCCCGAGCAGGTGGTGGCCATCGCCAGCAATAGCGGTGGCAAGCAGGC

avrBs3* 2041 TTGGAGACCGTGACAGACTTCTGCCGGTGTGTGCCAGGCCCATGGCTAACACCCGAG
avrBs3 2041 CTGGAGACGGTGCAGCGGCTGTTGCCGGTGTGTGCCAGGCCCATGGCTGACCCCGGAG

avrBs3* 2101 CAGGTGCTGTATCGCTTACATGATGGTGGAAAGCAAGCTCTTGA AACTGTACAAAGG
avrBs3 2101 CAGGTGGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGG

5 Anhang

avrBs3* 2161 CTGTTACCAAGTTCTCTGCCAGGCTCATGGATGACTCCAGAACAAGTAGTCGCTATAGCC
avrBs3 2161 CTGTTGCCGGTGTCTGTGCCAGGCCCATGGCTGACCCTGGAGCAGGTGGTGCCATCGCC

avrBs3* 2221 AGTCATGATGGTGGAAAACAGGCTCTTGAGACAGTTCAGAGATTATTACCTGTCTCTGTCT
avrBs3 2221 AGCCACGATGGCAGCAACAGGGCTCTGAGACGCTGCAGCGGCTGTTGCCGTCTGTGC

avrBs3* 2281 CAAGCCACGGACTGACACCAGAGCAAGTGTGCTATGCTTCTCAGCAGCGGTGGTAAA
avrBs3 2281 CAGGCCCATGGCTGACCCTGGAGCAGGTGGTGGCCATCGCCAGCCAGATGGCGGCAG

avrBs3* 2341 CAGGCCTAGAGACCCTCAAGGTTACTTCCAGTGTGTGTCAAGCCCATGGACTAACA
avrBs3 2341 CAGGCCTGAGAGCGTGCAGCGGCTGTGCGGTGCTGTGCAGGCCCATGGCTGACC

avrBs3* 2401 CCGCAACAAGTTGTTGCTATTGCTTCTAACGGTGGAGGAAGACCTGCTTGGAGACGGTA
avrBs3 2401 CCGCAGCAGGTGTGTGCCATCGCCAGCAATGGCGGCGGCAGGCCGGCTGGAGACGGTG

avrBs3* 2461 CAGACTCTTACCTGTACTGTCTCAGGCTCATGGCTTACTCCTGAGCAGGTAGTAGCT
avrBs3 2461 CAGCGCTGTTGCCGTCTGTGCAGGCCCATGGCTGACCCTGGAGCAGGTGGTGGC

avrBs3* 2521 ATCGCTTCGCATGATGGTGGAAAGCAGGCACTGGAGACAGTCCAGAGATGCTGCCGGTT
avrBs3 2521 ATCGCAGCCACGATGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTGCCGGTG

avrBs3* 2581 CTCTGTCAAGCCCATGGCTTGACACCAAGCAGGTTGGTGGCAATTGCATCAAACGGAGGA
avrBs3 2581 CTGTGCCAGGCCCATGGCTGACCCTGGAGCAGGTGGTGGCCATCGCCAGCAATGGCGGC

avrBs3* 2641 GGACGGCCAGCACTTGAGTCGATTGTTGGCAATTGAGCAGACCCGATCCTGCTTTGGCT
avrBs3 2641 GGACGGCCGCGCTGAGAGCATTGTTGCCAGTTATCTCGCCCTGATCCGCGTTGGCC

avrBs3* 2701 GCGCTAACGAATGATCACCTCGTTGCTTTAGCATGCCTAGGAGGTCCGCCGCTCTGGAT
avrBs3 2701 GCGTTGACCAACGACACCTCGTCCCTTGGCTGCCTGGCGGACCTCCTGCGCTGGAT

avrBs3* 2761 GCTGTCAAGAAAGGGGTTCCACATGCCCTGCTCTTATTAAGCGTACTAACCGACGAATC
avrBs3 2761 GCAGTGAAAAGGGATGCCGCACGGCGCGGCTTGATCAAAAAGAACCAATCGCCGATAT

avrBs3* 2821 CCAGACGGCACCAGCCATAGAGTAGCTGACCACGCGCAAGTGTTAGAGTCTTAGGCTTT
avrBs3 2821 CCCGAAACGACATCCATCGCTTGCAGACCACGCGCAAGTGGTTCGCTGCTGGGTTT

avrBs3* 2881 TTCCAAATGCCATCTCATCCCGCTCAGGCTTTCGACGATGCAATGACTCAATTCGGAATG
avrBs3 2881 TTCCAGTGCCATCCACCCAGGCGAAGCATTGATGACGCCATGACGAGTTCGGGATG

avrBs3* 2941 TCTAGACATGGTCTTCTTCAATGTTTTCGCAGAGTGGTGTCCACCGAACTCGAAGCACCT
avrBs3 2941 AGCAGGCACGGGTGTACAGCTCTTTCGCAGAGTGGCGTCCACCGAACTCGAAGCCCGC

avrBs3* 3001 AGTGGAACTCTCCCTCCTGCAAGTCAGAGATGGGACCGCATTTTGCAAGCAAGTGGTATG
avrBs3 3001 AGTGGAACTCTCCCTCCAGCCTCGCAGCGTTGGGACCGTATCCTCAGGCATCAGGGATG

5 Anhang

```

avrBs3* 3061 AAAAGGGCTAAACCGTCTCCAACAGTACC CAAACGCCTGATCAGGCTTCTCTTCACGCA
avrBs3 3061 AAAAGGGCCAAACCGTCCCTACTTCAACT CAAACGCCGATCAGGCGTCTTTGCATGCA

avrBs3* 3121 TTTGCTGATTCGCTGGAGCGAGATCTTGACGCTCCGTCA CCAATGCATGAAGGAGACCAA
avrBs3 3121 TTCGCCGATTCGCTGGAGCGTGACCTTGATGCCCTAGCCCAATGCACGACGGAGATCAG

avrBs3* 3181 ACACGAGCGTTCATCTCGAAAGCGATCTAGGAGCGATCGTGCAAGTGACCGGTCCATCTGCT
avrBs3 3181 ACGCGGGCAAGCAGCCGTAAACGGTCCCGATCGGATCGTGCTGTCACCGGTCCCTCCGCA

avrBs3* 3241 CAACAATCATTTGAGGTAGGGTTCCGGAACAAAGAGATGCTTTGCATTTGCCCTTTGTCA
avrBs3 3241 CAGCAATCGTTCGAGGTGCGCGTTCCCGAACAGCGCGATGCGCTGCATTTGCCCTTCAGT

avrBs3* 3301 TGGAGGGTTAAGAGACC TAGAACTTCTATGGAGGAGGTTGCCGGATCCTGGAACCTCCT
avrBs3 3301 TGGAGGGTAAAACGCCGCGCTACAGTATCGGGGGCGGCTCCCGGATCCTGGTACGCC

avrBs3* 3361 ACTGCTGCTGATCTTGCTGCTTCTTCAACTGTTATGACAGAGCAAGATGAAGATCCTTTT
avrBs3 3361 ACGGCTGCCGACCTGGCAGCGTCCAGCACCGTGATGCGGGAACAAGATGAGGACCCCTTC

avrBs3* 3421 GCTGGAGCTGCTGATGATTTCCCTGCTTTTAATGAGGAAGAGCTTGCTTGGCTTATGGAA
avrBs3 3421 GCAGGGGCAGCGGATGATTTCCCGGCATTCAACGAAGAGGAGCTCGCATGGTTGATGGAG

avrBs3* 3481 TTGCTTCCACAGTAA
avrBs3 3481 CTATTGCCTCAGTGA

```

Abb. 14 Vergleich der Nukleotidsequenz von *avrBs3* mit *N. benthamiana codon usage (avrBs3*)* und *Xanthomonas codon usage (avrBs3)*. Die zwischen beiden Nukleotidsequenzen identischen Nukleotide sind in weißer Schrift auf schwarzem Grund und unterschiedliche in schwarzer Schrift auf weißem Grund dargestellt. Die Homologie der Gesamtnukleotidsequenz beträgt 72% und im Bereich der *Repeats* 74%.

5 Anhang

A

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1		94.12	93.14	95.10	93.14	93.14	93.14	100.00	100.00	94.12	92.16	94.12	100.00	100.00	100.00	93.14	100.00
2	94.12		99.02	99.02	94.12	94.12	94.12	94.12	94.12	100.00	96.08	98.04	94.12	94.12	94.12	97.06	94.12
3	93.14	99.02		98.04	95.10	95.10	95.10	93.14	93.14	99.02	97.06	99.02	93.14	93.14	93.14	96.08	93.14
4	95.10	99.02	98.04		95.10	95.10	95.10	95.10	95.10	99.02	97.06	99.02	95.10	95.10	95.10	96.08	95.10
5	93.14	94.12	95.10	95.10		100.00	100.00	93.14	93.14	94.12	98.04	96.08	93.14	93.14	93.14	91.18	93.14
6	93.14	94.12	95.10	95.10	100.00		100.00	93.14	93.14	94.12	98.04	96.08	93.14	93.14	93.14	91.18	93.14
7	93.14	94.12	95.10	95.10	100.00	100.00		93.14	93.14	94.12	98.04	96.08	93.14	93.14	93.14	91.18	93.14
8	100.00	94.12	93.14	95.10	93.14	93.14	93.14		100.00	94.12	92.16	94.12	100.00	100.00	100.00	93.14	100.00
9	100.00	94.12	93.14	95.10	93.14	93.14	93.14	100.00		94.12	92.16	94.12	100.00	100.00	100.00	93.14	100.00
10	94.12	100.00	99.02	99.02	94.12	94.12	94.12	94.12	94.12		96.08	98.04	94.12	94.12	94.12	97.06	94.12
11	92.16	96.08	97.06	97.06	98.04	98.04	98.04	92.16	92.16	96.08		98.04	92.16	92.16	92.16	93.14	92.16
12	94.12	98.04	99.02	99.02	96.08	96.08	96.08	94.12	94.12	98.04	98.04		94.12	94.12	94.12	95.10	94.12
13	100.00	94.12	93.14	95.10	93.14	93.14	93.14	100.00	100.00	94.12	92.16	94.12		100.00	100.00	93.14	100.00
14	100.00	94.12	93.14	95.10	93.14	93.14	93.14	100.00	100.00	94.12	92.16	94.12	100.00		100.00	93.14	100.00
15	100.00	94.12	93.14	95.10	93.14	93.14	93.14	100.00	100.00	94.12	92.16	94.12	100.00	100.00		93.14	100.00
16	93.14	97.06	96.08	96.08	91.18	91.18	91.18	93.14	93.14	97.06	93.14	95.10	93.14	93.14	93.14		93.14
17	100.00	94.12	93.14	95.10	93.14	93.14	93.14	100.00	100.00	94.12	92.16	94.12	100.00	100.00	100.00	93.14	

B

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1		64.71	77.45	76.47	69.61	71.57	67.65	76.47	78.43	71.57	75.49	75.49	71.57	71.57	77.45	69.61	74.51
2	64.71		77.45	73.53	71.57	68.63	67.65	75.49	68.63	70.59	70.59	70.59	73.53	73.53	66.67	70.59	68.63
3	77.45	77.45		80.39	72.55	68.63	68.63	76.47	69.61	75.49	78.43	73.53	75.49	72.55	73.53	74.51	73.53
4	76.47	73.53	80.39		71.57	66.67	71.57	80.39	74.51	71.57	79.41	77.45	76.47	74.51	73.53	75.49	74.51
5	69.61	71.57	72.55	71.57		75.49	76.47	67.65	71.57	69.61	72.55	68.63	78.43	65.69	68.63	71.57	73.53
6	71.57	68.63	68.63	66.67	75.49		78.43	67.65	72.55	70.59	74.51	65.69	68.63	70.59	75.49	65.69	72.55
7	67.65	67.65	68.63	71.57	76.47	78.43		65.69	68.63	69.61	72.55	67.65	75.49	69.61	69.61	67.65	73.53
8	76.47	75.49	76.47	80.39	67.65	67.65	65.69		77.45	71.57	75.49	70.59	78.43	80.39	72.55	71.57	75.49
9	78.43	68.63	69.61	74.51	71.57	72.55	68.63	77.45		69.61	74.51	76.47	77.45	72.55	72.55	69.61	78.43
10	71.57	70.59	75.49	71.57	69.61	70.59	69.61	71.57	69.61		75.49	72.55	71.57	72.55	80.39	75.49	73.53
11	75.49	70.59	78.43	79.41	72.55	74.51	72.55	75.49	74.51	75.49		77.45	71.57	72.55	75.49	68.63	70.59
12	75.49	70.59	73.53	77.45	68.63	65.69	67.65	70.59	76.47	72.55	77.45		67.65	68.63	66.67	71.57	74.51
13	71.57	73.53	75.49	76.47	78.43	68.63	75.49	78.43	77.45	71.57	71.57	67.65		71.57	73.53	73.53	78.43
14	71.57	73.53	72.55	74.51	65.69	70.59	69.61	80.39	72.55	72.55	72.55	68.63	71.57		78.43	71.57	77.45
15	77.45	66.67	73.53	73.53	68.63	75.49	69.61	72.55	72.55	80.39	75.49	66.67	73.53	78.43		70.59	74.51
16	69.61	70.59	74.51	75.49	71.57	65.69	67.65	71.57	69.61	75.49	68.63	71.57	73.53	71.57	70.59		68.63
17	74.51	68.63	73.53	74.51	73.53	72.55	73.53	75.49	78.43	73.53	70.59	74.51	78.43	77.45	74.51	68.63	

Abb. 15 Vergleich der Nukleotid-Identität der *avrBs3-Repeats* 1 bis 17 mit *Xanthomonas codon usage* (A) und *N. benthamiana codon usage* (B) in Prozent.

5 Anhang

>AvrBs3drep16-UPA12

HD NG NS NG NI NI NI HD HD NG HD NG HD NG
A A G T A A A A C T C T C C

>AvrBs3drep16-UPA14

HD NG NS NG NI NI NI HD HD NG HD NG HD NG
A A A T A A A C C T C C C T

>AvrBs3drep16-UPA19

HD NG NS NG NI NI NI HD HD NG HD NG HD NG
A T A T A C A C C C C T C C

>AvrBs3drep16-Bs3-E

HD NG NS NG NI NI NI HD HD NG HD NG HD NG
A T A T A A A C C T C T C T

>AvrHah1-Bs3

NN IG NI NI NI HD HD NG NN NI HD HD HD NG
A T A A A C C T A A C C C T

>AvrBs3drep109-Bs3

HD NG NS NG NI NI NI HD HD NG NS NS NG HD NG
A T A T A A A C C T A A C C A

>Tal1C-OsHen1

HD HD HD HD HD NG HD NN HD NG HG NN HD NG NG NG
C C C C C T C G C T T C C C T T

>AvrXa27-Xa27

NI NN NG NG NS NN NN NI NN NI NG HD HD NI NG NG
A G A A G A A G A G A C C C A T A

>AvrBs3-Bs3

HD NG NS NG NI NI NI HD HD NG NS NS HD HD HD NG HD NG
A T A T A A A C C T A A C C C T C C

>AvrBs3-UPA20

HD NG NS NG NI NI NI HD HD NG NS NS HD HD HD NG HD NG
A T A T A A A C C T G A C C C T C T

>AvrBs3-UPA12

HD NG NS NG NI NI NI HD HD NG NS NS HD HD HD NG HD NG
A A G T A A A A C T C T C C C C T C T

>AvrBs3-UPA14

HD NG NS NG NI NI NI HD HD NG NS NS HD HD HD NG HD NG
A A A T A A A C C T C C C C C C T C C

>AvrBs3-UPA19

HD NG NS NG NI NI NI HD HD NG NS NS HD HD HD NG HD NG
A T A T A C A C C C C T C C C C C C C

>AvrBs3-UPA21

HD NG NS NG NI NI NI HD HD NG NS NS HD HD HD NG HD NG
A T A T A A A A C C A A A A C C C T

5 Anhang

>AvrBs3-UPA25

HD NG NS NG NI NI NI HD HD NG NS NS HD HD HD NG HD NG
A T A A A A C C T G C C C C T C T

>Tal6C-Os07g47790

NI HD NN NI NN HA NN NS NS NI HD HA HA HA HD HD HD HA HD NG
C C A A G C A A A A C G C C C C A A C T

>Tal9A(PthXo8)-OsHen1

HD HD HD NG NG NN HD HD NG NI NI NN HD HI ND HD NI HD NG NG
C C C T T C C T A A A C C C C A C T T

>pthXo7-OsTFIIAg1

NI NG NI NI NG NN HD HD NG NI NI NI NG HD HG NN NS NN HD HD NG NG
A T A A T C C A A A T C C C C C T C

>pthXo6-OsTFX1 (Os09g29820)

NI H* NI NN NN NN NN NN HD NI HD HG HD NI N* NS NI NI HG HD NS NS NG
A T A A A A G G C C T C A C C A A C C A T

>Tal4C-tal-C4c

NS HG NI NG NS NN ND NG NG NG HN NN HD NS NI NN HD HD NG NG HG HD NG
A T A A A C T G A C A A G C C T T C T

>pthXo1-Os8N3 (Os08g42350)

NN HD NI HG HD NG N* HD HD NI NG NG NI HD NG NN NG NI NI NI NI N* NS N*
G C A T C T C C C T A C T G T A A C A C

>AvrXa7-Os11N3

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*
A T A A A C C C C C C C C A A C C A G C C C C

>XOCTAL4c-Os06g37080

NS HG NI NG NS NN ND N* NG N* HN NN HD NS NI NN HD HD NG NG HG HD NG
A T A A A C C T G A C A A G C C T T C T

>XOCTAL4a-Os03g37840

NI NG NI NG NN NG NN NG HD NN NN HG HD NN NS NN HD HD NG NA NN HD HD HD HD NG
A T A T A T T C G G T C A G G C C A G G C C C C T

>XOCTAL6-Os05g07700

NN HD NI NI NN HA NN NS NS NI HD HA HA HA HD HD HD HA HD NG
G C A A A C C C A C A G C C C C C C T

>XOCTAL6-Os07g47790

NN HD NI NI NN HA NN NS NS NI HD HA HA HA HD HD HD HA HD NG
G C A A G C A A A A C G C C C C A A C T

>XOCTAL6-Os12g42970

NN HD NI NI NN HA NN NS NS NI HD HA HA HA HD HD HD HA HD NG
A A A G C G A A A C C C T C C C T C C

>Tal7a-Os01g68740

NI HG NI NI NI NN HD NS NN NS NN HD NN NI HD NN NS NG
A T A A A A C A A C G C A A C G A T

5 Anhang

>Tal8a-Os01g68740

NI HG NI NI NI NN HD NS NN NS NN HD NN NI HD NN NS NG
A T A A A A C A A C G C A A C G A T

>PthXo7-Os01g13610

NI NG NI NI NG NN HD HD NG NI NI NI NG HD HG NN NS NN HD HD NG NG
A **C** A A **C** **C** C C T A A A **C** **A** **C** A A **C** C C **C** T

>XOCTAL5b-Os05g25770

HH NN HD HD NG N* HD NN HD ND N* NG NS NS NN HD HD NG
T G C C T C C A C C C **C** A A G C C T

>PthXo3-Os11N3

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*
A T A T A A A C **C** C C C T C C A **A** C **C** A **C** G **C** **C** **C** T **A** **A**

Abb. 16: Der Fehlpaarungstoleranz-Analyse zugrundeliegende TALE-Promotor-Interaktionen. Die Nukleotide suboptimaler RVD-Nukleotid-Interaktionen sind rot hinterlegt. Die RVD-Nukleotid-Interaktionen des letzten halben *Repeats* wurden nicht in die Analyse eingeschlossen, sind aber abgebildet.

Tabelle 4: Zusammenfassung der Fehlpaarungstoleranz-Analyse.

TALE	Promotor	Repeats	Fehlpaarungen	Fehlpaarungsquote in %
AvrHah1	Bs3	13	1	8
AvrBs3 Δ rep16	Bs3-E	13	1	8
AvrBs3 Δ rep16	UPA12	13	3	23
AvrBs3 Δ rep16	UPA14	13	3	23
AvrBs3 Δ rep16	UPA19	13	3	23
AvrBs3 Δ rep109	Bs3	14	2	14
Tal1C	OsHen1	15	2	13
AvrXa27	Xa27	16	3	19
Tal7a	Os01g68740	17	0	0
Tal8a	Os01g68740	17	0	0
XOCTAL5b	Os05g25770	17	1	6
AvrBs3	Bs3	17	2	12
AvrBs3	UPA20	17	2	12
AvrBs3	UPA25	17	2	12
AvrBs3	UPA12	17	3	18
AvrBs3	UPA14	17	3	18
AvrBs3	UPA19	17	5	29
AvrBs3	UPA21	17	6	35
PthXo8	OsHen1	19	1	5
XOCTAL6	Os05g07700	19	1	5
XOCTAL6	Os07g47790	19	1	5
XOCTAL6	Os12g42970	19	1	5
Tal6C	Os07g47790	19	2	11
PthXo7	OsTFIIAy1	21	5	24
PthXo7	Os01g13610	21	8	38
PthXo6	OstFX1	22	2	9
Tal4C	Os06g37080	22	4	18
XOCTAL4c	Os06g37080	22	3	14
PthXo1	Os8N3 (Xa13)	23	4	17
XOCTAL4a	Os03g37840	25	2	8
AvrXa7	Os11N3	25	8	32

Danksagung

Ich möchte mich bei Prof. Dr. Thomas Lahaye für die hervorragende Bereuung, die vielen Diskussionen und die Möglichkeit zur wissenschaftlichen Selbstständigkeit während der Erstellung der vorliegenden Arbeit und die Erstellung des Gutachtens bedanken. Auch für die anderen kleinen und großen Hilfen sind meine Familie und ich dir zutiefst dankbar.

Bei Prof. Dr. Martin Parniske bedanke ich mich für die Erstellung des Gutachtens und das sich unsere Arbeitsgruppe bei ihm ins gemachte Nest setzen durfte, wodurch die Kontinuität meiner Arbeit trotz des Umzugs nach München gewährleistet war.

Bei meiner Arbeitsgruppe bedanke ich mich für die angenehme und fruchtbare Arbeitsatmosphäre sowie die vielen kleinen Hilfestellungen, die zum Gelingen der vorliegenden Arbeit beigetragen haben.

Auch allen weiteren Hallenser und Münchner Kollegen danke ich dafür, diese Zeit nur sehr selten als reine Arbeit empfunden zu haben.

Ein besonderer Dank gilt Annett für das Korrekturlesen dieser Arbeit. Auch bei Tina bedanke ich mich für das Korrekturlesen und die seelische und moralische Unterstützung in den letzten Monaten.

Meiner Familie danke ich für ihre immerwährende Liebe, Unterstützung und Vertrauen.

Claudia danke ich für die Zeit in der sie genug Liebe, Kraft, Mut und Vertrauen hatte diesen Weg gemeinsam zu gehen.

Meinem Sohn danke ich für die Freude und Kraft die er in mein Leben bringt, sowie die Erkenntnis das Leben viel mehr als arbeiten ist. Danke, dass du es mir so einfach machst.