Analysis of TALE repeats of aberrant length and identification of novel candidate target genes for *Xanthomonas oryzae* pv. *oryzicola* TALEs

Von der Naturwissenschaftlichen Fakultät
der Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation
von
Sebastian Becker, M.Sc.

2021

Referent: Prof. Dr. rer. nat. Jens Boch

Korreferent: Prof. Dr. rer. hort. Edgar Maiß

Tag der Promotion: 17.11.2020

Zusammenfassung

Transcription activator-like effectors (TALEs) sind Proteine pflanzenpathogener *Xanthomonaden*. Sie induzieren die Expression ihrer pflanzlichen Ziel-Gene indem sie die Funktion eukaryotischer Transkriptionsfaktoren imitieren. Die DNA-Bindung erfolgt dabei über 33-35 Aminosäuren lange Sequenzwiederholungen, die *repeats*, von denen jeder genau ein Nukleotid der DNA erkennt. In der Natur finden sich vereinzelt *repeats,* die deutlich von dieser Standardlänge abweichen. Die Anwesenheit eines solchen aberranten *repeats* erlaubt es einem TALE, eine zusätzliche Zielsequenz mit einer 1 bp-Leserasterverschiebung zu erkennen. In der vorliegenden Arbeit wurden diese aberranten *repeats* weiterführend analysiert. Dabei konnte sowohl gezeigt werden, dass TALEs mehr als ein oder zwei aberrante *repeats* tolerieren können, als auch, dass zwei aberrante *repeats* einen minimalen Abstand benötigen, um voneinander unabhängig zu fungieren. Sorgfältig platziert können zwei aberrante *repeats* es einem TALE dann erlauben, eine Vielzahl verschiedener Zielsequenzen mit unterschiedlichen Leserasterverschiebungen zu erkennen. Weiterhin wurden neue, natürlich vorkommende aberrante *repeats* identifiziert und analysiert. Dabei wurde demonstriert, dass nicht alle aberranten *repeats* die Erkennung einer Leserasterverschiebung vermitteln. Zwei von ihnen hatten keinerlei Einfluss auf die DNA-Bindung von TALEs, während eine dritte Variante nie an der DNA-Bindung teilnahm. Eine Analyse künstlich erzeugter aberranter *repeats* zeigte, dass ihre Funktionsweise durch die Art und die Position der veränderten Aminosäuren beeinflusst wird. Dies erlaubte, die Toleranz eines gewöhnlichen TALE *repeats* für Insertionen zu visualisieren. Es konnte auch demonstriert werden, dass eine funktionelle Domäne zwischen zwei TALE *repeats* inseriert werden kann, ohne die Funktion der inserierten Domäne oder die des TALEs zu beeinträchtigen. Darüber hinaus wurde der erste natürlich vorkommende TALE mit zwei aberranten *repeats* identifiziert und analysiert. Dieser TALE war nicht nur in der Lage, multiple Zielsequenzen mit unterschiedlichen Leserasterverschiebungen von 1 bis 4 bp zu binden, er induziert außerdem das bekannte Suszeptibilitätsgen *OsSWEET13* aus dem Reis cv. Nipponbare. Es konnte darüber hinaus gezeigt werden, dass ein neues Allel desselben Gens aus dem cv. Sadu Cho eine Resistenz gegen die beiden natürlich vorkommenden *OsSWEET13*-bindenden TALEs vermittelt.

Technologische Weiterentwicklungen haben die Verfügbarkeit komplett sequenzierter *X. oryzae* Stämme drastisch erhöht. Die Ziel-Gene vieler ihrer TALEs sind allerdings weiterhin unbekannt. Um dieses Problem zu adressieren, wurden die meisten natürlich vorkommenden *Xoc* TALEs künstlich nachgebaut. In GUS-Reporterstudien wurden daraufhin sieben, anhand von RNA-Seq Daten vorhergesagte, Ziel-Gene experimentell bestätigt. Basierend auf *de novo* Ziel-Gen Vorhersagen wurden außerdem 22 neue Ziel-Gen-Kandidaten identifiziert, deren Promotor direkt durch einen *Xoc* TALE induziert wird.

Summary

Transcription activator-like effectors (TALEs) are bacterial proteins most prominently found in plant pathogenic *Xanthomonas* species. TALEs mimic the function of eukaryotic transcription factors in order to upregulate the expression of their plant target genes. Their unique structure, an array of consecutively placed 33-35 amino acids long repeats, allows them to bind to DNA in a simple one repeat to one nucleotide manner. However, a few naturally occurring repeat-variants contain small duplications or deletions whose presence enables a TALE to break with this strict one repeat to one nucleotide binding mechanism, allowing for the recognition of an additional target sequence with a -1 nucleotide frameshift. In this work, these aberrant repeats were analysed in detail, demonstrating that a single TALE can tolerate more than one or two aberrant repeats by simultaneously looping out multiple repeats. It was also shown that two aberrant repeats need a minimal distance to act functionally independent and that they allow a TALE, if properly spaced, to flexibly recognize a multitude of different frameshift sequences. Novel natural aberrant repeat variants were identified and analysed, revealing that not all of them confer frameshift binding. Two variants did not alter the DNA-binding of the tested TALE while one variant was unable to participate in DNA-binding. By investigating a set of artificial aberrant repeats, it was found that the nature as well as the position of the altered amino acids influence the functionality of aberrant repeats. This analysis also allowed to visualize the tolerance of a standard TALE repeat to small insertions and resulted in an experiment that could show that it is possible to insert complete functional domains from other proteins between two TALE repeats. Furthermore, the first natural TALE to contain two aberrant repeats was identified and its ability to bind to multiple frameshift sequences was demonstrated. It was also shown that this TALE targets an allele of the susceptibility gene *OsSWEET13* from the rice cv. Nipponbare. In addition, a newly identified *OsSWEET13* allele from the cv. Sadu Cho was shown to confer resistance against both known TALEs addressing *OsSWEET13*.

Recent technological advancements lead to an increased availability of *X. oryzae* genomes and TALE sequences alike. Despite numerous efforts, for many of these TALEs, no plant target gene is known. To address this issue, most of the naturally occurring *X. oryzae* pv. *oryzicola* TALEs were artificially re-constructed. In subsequent GUS reporter studies seven genes previously predicted as TALE targets based on RNA-Seq data were confirmed experimentally. *De novo* TALE target predictions led also to the identification of 22 novel TALE target gene candidates whose promoters were directly induced by a *Xoc* TALE.

keywords: *Xanthomonas oryzae*, TALE target genes, repeats of aberrant length

Table of contents

List of figures

List of tables

List of abbreviations

4mC/5mC	4-methyl cytosine/5-methyl cytosine
4-MU	4-methylumbelliferone
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
A	adenine
A. tumefaciens	*Agrobacterium tumefaciens*
aa	amino acid(s)
AAD	acidic activation domain
ABA	abscisic acid
AIM	*Agrobacterium* infiltration media
Amp	ampicillin
APS	ammonium persulfate
BB	bacterial blight
bHLH TF	basic helix-loop-helix transcription factor
BLS	bacterial leaf streak
bp	base pair(s)
BRs	brassinosteroids
BSA	bovine serum albumin
C	cytosine
C. elegans	*Caenorhabditis elegans*
CDS	coding sequence
cv.	cultivar
DNA	deoxyribonucleic acid
DMR6	downy mildew resistant 6
e.g.	exempli gratia (for example)
EBE	effector binding element
ER	endoplasmic reticulum
ev	empty vector
Fig.	figure
G	guanine
GA	gibberellic acid
GA20ox5	gibberellin 20 oxidase 5
Gent	gentamicin
GFP	green fluorescent protein
GGC	golden gate cloning

GH3	gretchen hagen 3
GUS	β-glucuronidase
h	hour
IAA	indole-3-acetic acid
INDEL	insertions and deletions
iTALE	interfering TALE
JA	jasmonic acid
Kan	kanamycin
kbp	kilo base pairs
KO	knockout
LB	lysogeny broth
LRR-RLK	leucine-rich repeat receptor-like kinase
Lsi1	low silicon rice
MCS	multiple cloning site
min	minutes
MLN	mixed-linkage glucans
mm	mismatch(es)
ms	milliseconds
MU/MUG	4-methylumbelliferone/4-methylumbelliferyl-β-D-glucuronid
n.a.	not available
n.t.	not tested
NCR	non-canonical repeats
NLS	nuclear localization signal
OD	optical density
op box	optimal box
PCR	polymerase chain reaction
PIC	preinitiation complex
PPR	pentatricopeptide repeat
PSA	peptone saccharose agar (media)
PUF	pumilio family protein
pv.	pathovar
rev.	reverse
RLK	receptor-like kinase
RNA	ribonucleic acid
ROX3	regulator of the XA21-mediated immunity 3
rpm	rounds per minute
RVD	repeat variable di-residue

S. cerevisiae	*Saccharomyces cerevisiae*
SA	salicylic acid
SAUR	small auxin upregulated RNAs
SDS	Sodium dodecyl sulfate
SG	serine and glycine
Si	silicon
SNP	single nucleotide polymorphism
Spec	spectinomycin
spp.	*species pluralis* (species)
subsp.	subspecies
SWEET	sugar will eventually be exported transporter
T	thymine
T3SS	type III secretion system
T3SS-S	signal for T3SS-dependent translocation
Tab.	table
TALE	transcription activator-like effector
TALEN	TALE nuclease
TEMED	Tetramethylethylenediamine
TF	transcription factor
TFIIA	transcription factor IIA
TFB	transcription factor binding site
TPR	tetratricopeptide repeat
truncTALE	truncated TALE
v/v	volume per volume
w/v	weight per volume
WAK	wall-associated kinase
X. oryzae	*Xanthomonas oryzae*
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
X-Gluc	5-bromo-4-chloro-3-indolyl β-D-glucuronide
Xac	*Xanthomonas axonopodis* pv. *citri*
XET	xyloglucan endotransglycosylase
Xoc	*Xanthomonas oryzae* pv *oryzicola*
Xoo	*Xanthomonas oryzae* pv *oryzae*
XTH	xyloglucan endotransglucosylase/hydrolase
YEB	yeast extract broth
ZF	zinc finger

1 Introduction

1.1 The *Oryza sativa-Xanthomonas oryzae* pathosystem

Xanthomonas oryzae pv. *oryzicola* (*Xoc*) and *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) are two pathovars belonging to the same rice-infecting *Xanthomonas* species. Both pathogens have a close genomic relationship (Swings *et al*., 1990; Rademaker *et al*., 2000), and only few characteristics exist that allow for a phenotypic distinction between the two (Vera Cruz *et al*., 1984). Yet, despite their similarities, they still differ in numerous ways, ranging from the symptoms they cause and their infection routes to the effectors they encode.

While both pathogens are able to infect rice plants through wounded tissue in leaves or roots (Fig. 1.1 A), they can also exploit the already available natural openings. For this, *Xoc* relies on open stomata (Fig. 1.1 B; Xie *et al*., 1991; Mew *et al*., 1993) whereas *Xoo* uses hydathodes (Fig. 1.1 C; Mew *et al*., 1984; Guo and Leach, 1989). At the beginning of an infection, both pathovars multiply within the intercellular space of the parenchymatic tissue around their initial entry point. However, due to its nature as a systemic pathogen, *Xoo* quickly moves on to infect and colonize the xylem vessels, allowing for a rapid spreading throughout the plant by means of the vascular tissue (Tabei, 1967). *Xoc*, in contrast, is mainly limited to the apoplastic space between the parenchymatic cells that lie directly

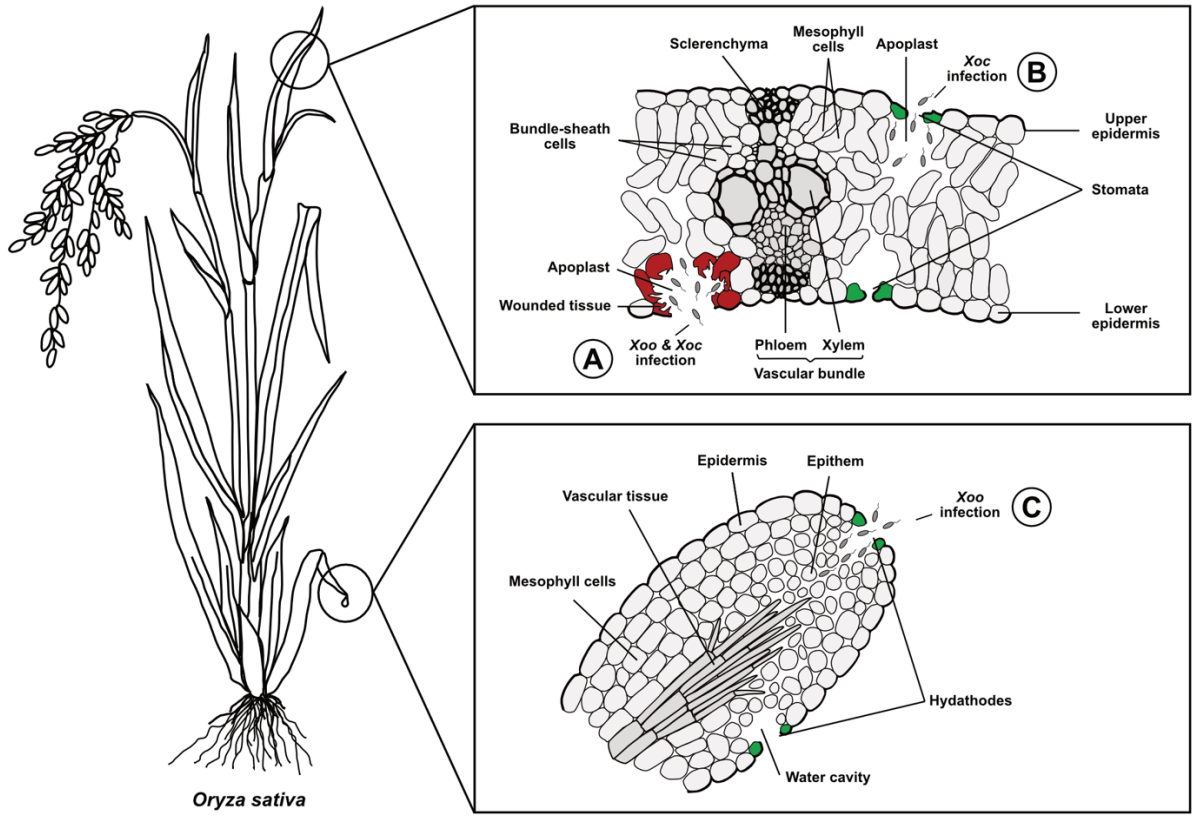


Figure 1.1: Methods of *X. oryzae* to enter and infect tissues within a rice plant. While *Xoo* and *Xoc* can enter a rice plant through **(A)** damaged tissues (red) e.g. of leaves, both pathogens can also use natural openings (green). For this *Xoc* relies on **(B)** open stomata while *Xoo* can use **(C)** Hydathodes to follow the guttation fluids into the plant.

underneath the epidermis. Only in later stages of an infection is *Xoc* also found in the extracellular space between the vascular bundles but, unlike *Xoo*, never actually within the xylem vessels (Mew *et al*., 1993). Both pathogens flourish best in the tropical and subtropical areas of South and East Asia but since there are also infection sites in Africa, Southwest Asia and Australia, they are found in all the major rice-producing regions of the world (Fig. 1.2 A; Jeger *et al*., 2018). *Xoo* infections were also reported for North, South and Central America (Jones *et al*., 1989; Leach and White, 1991), however, more recent studies indicate that at least the hypovirulent and TALE-deficient strains found in the U.S. belong to a genetically distinct lineage. Since they have not yet been assigned to a new pathovar, they are for now referred to as *X. oryzae*-US (*Xo*-US; Triplett *et al*., 2011). *Xoo* and *Xoc* can use infected plants, seeds or contaminated water as a source for a new infection (Vera Cruz *et al*., 2017) and both are able to overwinter within non-host plants or survive up to three months in soil or plant debris (Jeger *et al*., 2018). Both pathogens are transmitted mainly by wind and rain but spreading may also occur by humans moving through or small animals living within rice fields (Jeger *et al*., 2018).

While the two diseases, bacterial blight (BB) caused by *Xoo* and bacterial leaf streak (BLS) caused by *Xoc*, can be distinguished comparatively easy at early stages of an infection, their symptoms are more and more indiscernible as they progress (Fig. 1.2 B and C; Nino-Liu *et al*., 2006). If plants are infected at an early age, both pathogens are likely to cause a decrease in harvest yield but *Xoo* has a much more severe impact. The overall yield losses

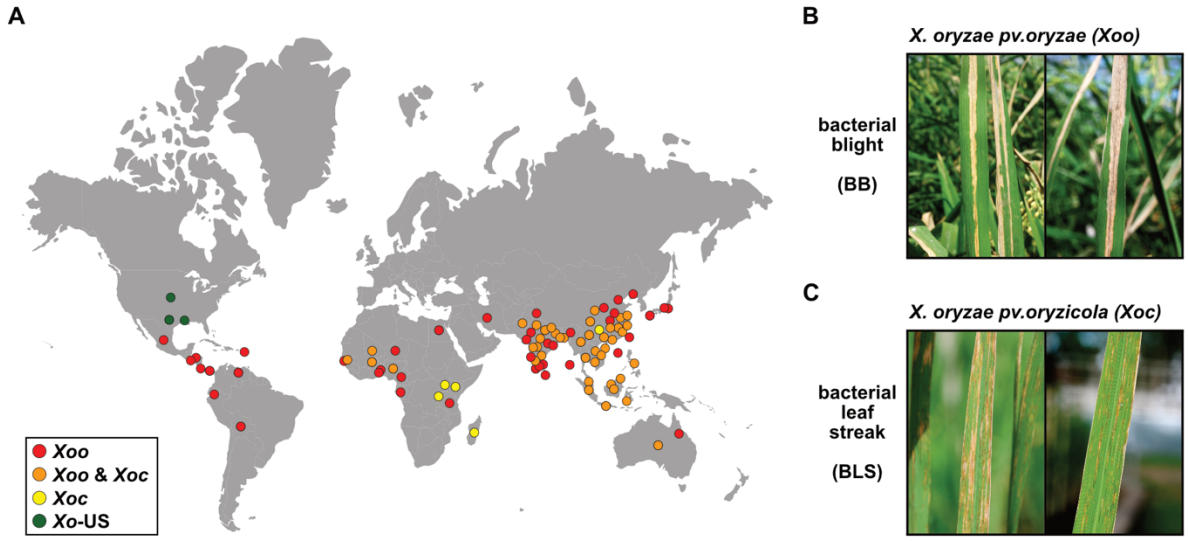


Figure 1.2: Distribution and symptoms of *Xanthomonas oryzae* spp. (A) Distribution map for the different *X. oryzae* spp. The distribution data was derived from https://www.cabi.org (full link in the Appendix). Each dot represents a country with reported infections. Multiple dots within a single country highlight the individual states or provinces that are affected in a more detailed representation (mainly countries with high rice production). The infections in the U.S. were originally reported to be caused by *Xoo*, however, recent studies demonstrated that these hypovirulent and TALE-deficient strains represent a genetically distinct lineage that is for now referred to as *X. oryzae*-US (*Xo*-US; Triplett *et al*., 2011). Symptoms of **(B)** bacterial blight (BB) caused by *X. oryzae* pv. *oryzae* (*Xoo*) and **(C)** bacterial leaf streak (BLS) caused by X. *oryzae* pv. *oryzicola* (*Xoc*). Pictures are taken from http://www.knowledgebank.irri.org (full link in the Appendix).

caused by BB can range from 2 to 100% and are depending on a wide range of different factors, among them the general susceptibility of the infected rice cultivar, the growth stage of the plant when it was infected, the nitrogen availability and the climatic conditions (Reddy *et al*., 1979; Mew *et al*., 1993). In contrast, reports for BLS only state yield losses of up to 30% - even under favorable conditions (Nino-Liu *et al*., 2006). However, BLS can have a big impact on a local level, e.g. in parts of Asia where some of the planted hybrid lines seem particularly susceptible towards this pathogen (Raymundo *et al*., 1999).

Taken together, both diseases have the potential to threaten the staple food production of more than 3 billion people worldwide, and thus a constant monitoring of their development and the ongoing effort to breed and select resistant rice cultivars is necessary. It is therefore critical to broaden our knowledge about the different *X. oryzae* pathovars and learn more about the mechanisms that are leading to their virulence.

1.2 The type III secretion system and TALEs

As it is the case with many plant pathogenic gram-negative bacteria, the virulence of *Xoo* and *Xoc* strongly relies on the presence of a type III secretion system (T3SS; Alfano and Collmer, 2004; Gurlebeck *et al*., 2006). With the help of this needle-like molecular structure, these pathogens are able to translocate effector proteins directly into the cells of their host plants to alter various processes from the inside (Nino-Liu *et al*., 2006). One large and very diverse group among those injected proteins is the family of transcription activator-like effectors (TALEs). TALEs occur in various *Xanthomonas* spp. infecting a wide range of different plants, many of which are of immense agronomical importance. Among these host plants are numerous cereals (barley, rye, wheat, and triticale), rice, citrus, the common bean, various *Brassicaceae*, cotton, pepper and mango. Interestingly, the number of TALEs in the different *Xanthomonas* species varies drastically. While *X. gardneri* strains contain, for example just one or two TALEs, the *Brassicaceae*-infecting *X. campestris* strains contain between 0 and 5, and the mango-infecting *X. campestris* strains up to 10 different ones (reviewed in Schornack *et al*., 2013). With up to 29 different TALEs per strain, however, the two rice-infecting pathovars *Xoo* and *Xoc* contain the by far largest TALE repertoires.

TALEs are bacterial effector proteins that mimic the function of eukaryotic transcription factors (TFs). They bind to promoter sequences in order to up-regulate the expression of their target genes, thus supporting bacterial growth and survival (Boch and Bonas, 2010; Doyle *et al*., 2013). Additionally, TALEs contain a signal in their N-terminus that facilitates their type III-dependent translocation (T3SS-S; Fig. 1.3 A and B). Once a TALE is successfully translocated into the cytosol of a plant cell, the two functional nuclear localization signals (NLS) in its C-terminal region (van den Ackerveken *et al*., 1996) interact

with the nuclear import protein importin α (and likely β) to mediate transportation through the nuclear envelope and into the nucleus (Fig. 1.3 A and B; Szurek *et al.*, 2001; Schornack *et al.*, 2013). The TALE then searches within the genome of the cell for potential target sites (also called TALE box or effector binding element (EBE)) and, upon encountering a matching sequence, binds to it. For this, each TALE contains a central repeat region that allows for the

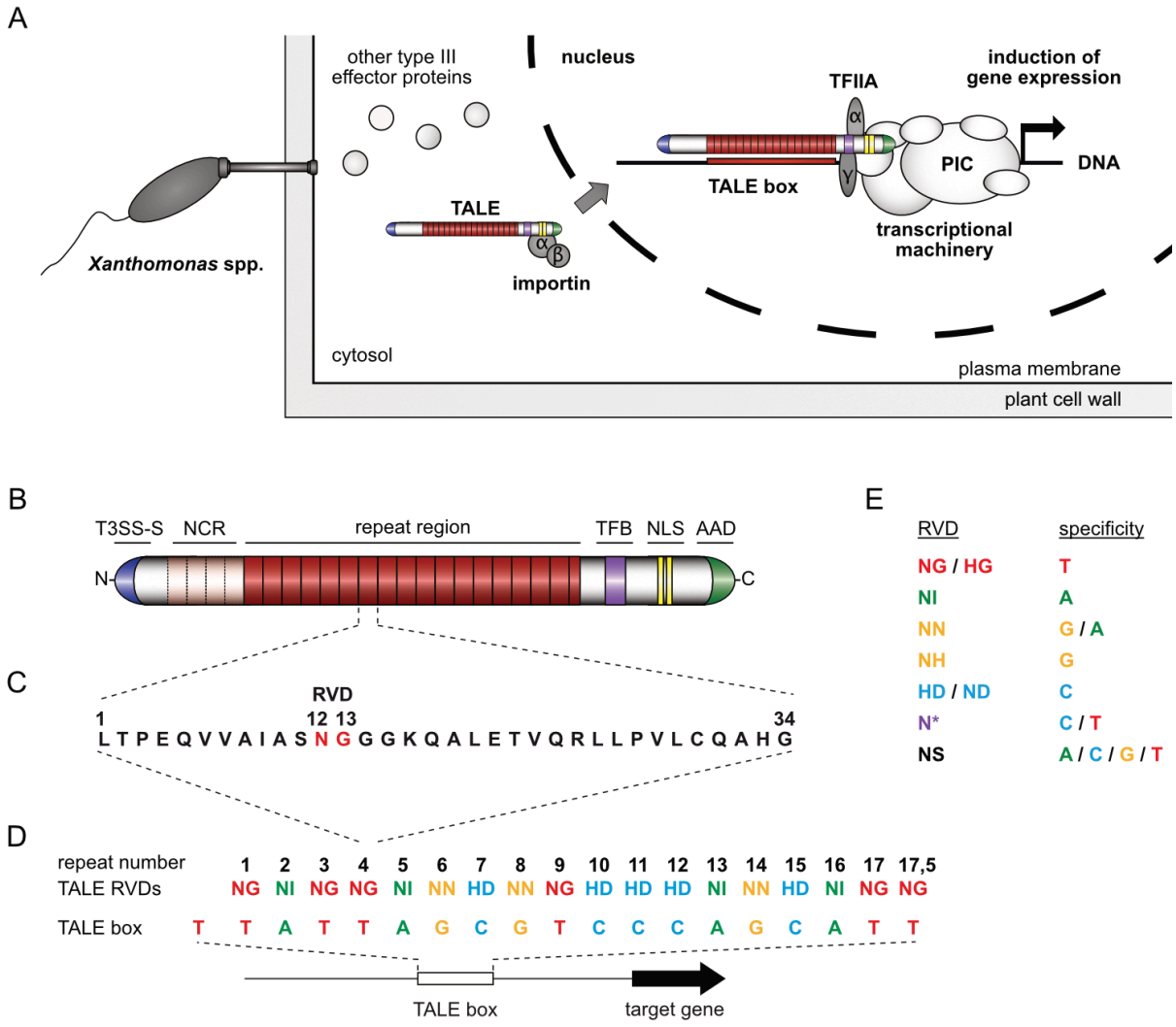


Figure 1.3: Transcription activator-like effector proteins and their mode of action. (A) *Xanthomonas* spp. use a type III secretion system to inject TALEs and other type III effector proteins into the cytosol of plant cells. After an importin-mediated transport into the nucleus, TALEs bind to matching sequences within the promoter regions of their target genes to stimulate gene expression. For this, they promote the assembly of the preinitiation complex (PIC) and recruit the transcriptional machinery of the cell. This is, at least partially, due to an interaction between TALEs and the α and γ subunits of transcription factor IIA (TFIIA). **(B)** General architecture of a TALE. The C-terminal part of TALEs contains two functional nuclear localization signals (NLS), a putative acidic activation domain (AAD) and a transcription factor binding site (TFB). The N-terminal region contains a type III secretion signal (T3SS-S) as well as four non-canonical repeats (NCR). Every TALE also contains a repeat region with a varying number of highly conserved repeats of 33-35aa placed in tandem. **(C)** Amino acid sequence of a typical TALE repeat with 34aa. Highlighted are the amino acids 12 and 13, which are responsible for the base specificity of a TALE and which are called repeat variable di-residues (RVD). **(D)** Each RVD confers specificity for one nucleotide within the target sequence of a TALE (TALE box). Nearly all known TALE targets are preceded by a thymine (T_0) that is specified by one of the non-canonical repeats (repeat -1). **(E)** Selection of representative RVDs and their specified nucleotides. Some RVD tolerate only a single nucleotide (e.g. NG, or NI) while others can tolerate two (e.g. NN), three or all four (e.g. NS).

specific recognition of target sequences (Kay *et al*., 2007; Römer *et al*., 2009b). This repeat region consists of a varying number of highly conserved repeats of 34aa length that are arranged in tandem (Fig. 1.3 B; Boch and Bonas, 2010). While the number of repeats observed in natural TALEs can range from 1.5 to 33.5, the majority of them contains 15.5 to 19.5 repeats (Boch and Bonas, 2010) and experiments with artificial TALEs could demonstrate that a minimum number of at least 10.5 repeats is required for full TALE activity (Boch *et al*., 2009). Experimental and computational studies indicated that the DNA-binding mechanism of TALEs is governed by a simple code in which one repeat confers binding to exactly one nucleotide of the DNA target sequence. This hypothesis was confirmed and the process found to be mediated by two highly variable residues located at position 12 and 13 within each repeat, referred to as repeat variable di-residues (RVDs; Fig. 1.3 C and D; Boch *et al*., 2009; Moscou and Bogdanove, 2009). Depending on the RVD, each repeat can specify either exactly one nucleotide (like HD→C, NI→A, NG→T, NK→G) or allow for a number of different ones - ranging from two (NN→G/A) to all four (NS→A/T/G/C) (Fig. 1.3 E; Cong *et al*., 2012; Streubel *et al*., 2012; Yang *et al*., 2014; Miller *et al*., 2015). To induce the expression of its target genes, each TALE contains a transcription factor-binding site (TFB) and a putative acidic activation domain (AAD) in its C-terminal region (Fig. 1.3 B; Zhu *et al*., 1998; Yuan *et al*., 2016). It was shown that the TFB of TALEs directly interacts with the α and γ subunits of the general transcription factor IIA (TFIIA) by mimicking its third and last subunit β (Yuan *et al*., 2016; Ma *et al*., 2018). By hijacking TFIIA, a crucial factor in the initiation of transcription in eukaryotes (Thomas and Chiang, 2006), TALEs are able to stimulate the recruitment of the preinitiation complex (PIC) in their direct vicinity, thereby leading to the assembly of the transcriptional machinery of the plant and ultimately to an increased expression of their target genes (Ma *et al*., 2018).

1.3 TALEs and their interaction with DNA

Structural studies demonstrated that TALEs wrap around the DNA in form of a right-handed superhelical structure and that each repeat contacts exactly one nucleotide in the target sequence (Fig. 1.4 A-C; Deng *et al*., 2012a, Mak *et al*., 2012). For this, each repeat forms two helices that are connected by a small loop region in which the RVDs and a single invariable glycine are placed (positions 12-14; Fig. 1.4 C-F; Deng *et al*., 2012a, Mak *et al*., 2012). With the help of the structural data, it was clarified that TALEs contact only the nucleotides of one DNA strand, the coding one, and that only the amino acid at position 13 directly interacts with the DNA bases, thereby being majorly responsible for the DNA-specificity of the TALE repeat (Fig. 1.4 C-F; Deng *et al*., 2012a, Mak *et al*., 2012). In contrast, the amino acid at position 12 fulfills only a RVD-loop stabilizing function by forming a direct

hydrogen bond with the 8th and 9th residue of the same repeat (Deng *et al*., 2012a, Mak *et al*., 2012). Additionally, several other residues were found to be crucial for the function and the structural integrity of TALEs. Most prominently are the four amino acids directly following the RVDs. All four help to stabilize the connection between the TALE and the DNA; the two glycines at position 14 and 15 by forming water-mediated hydrogen bonds with the DNA phosphates (Deng *et al*., 2014), the lysine and glutamine at positions 16 and 17 by forming direct- and water-mediated hydrogen bonds with the sugar-phosphate backbone of the DNA (Deng *et al*., 2012 and 2014; Wicky *et al*., 2013). Another important role is fulfilled by the hydrophobic residues of a TALE repeat since they mediate the inter- and intra-repeat interactions that are responsible for the stabilization of the TALE structure (Deng *et al*., 2014

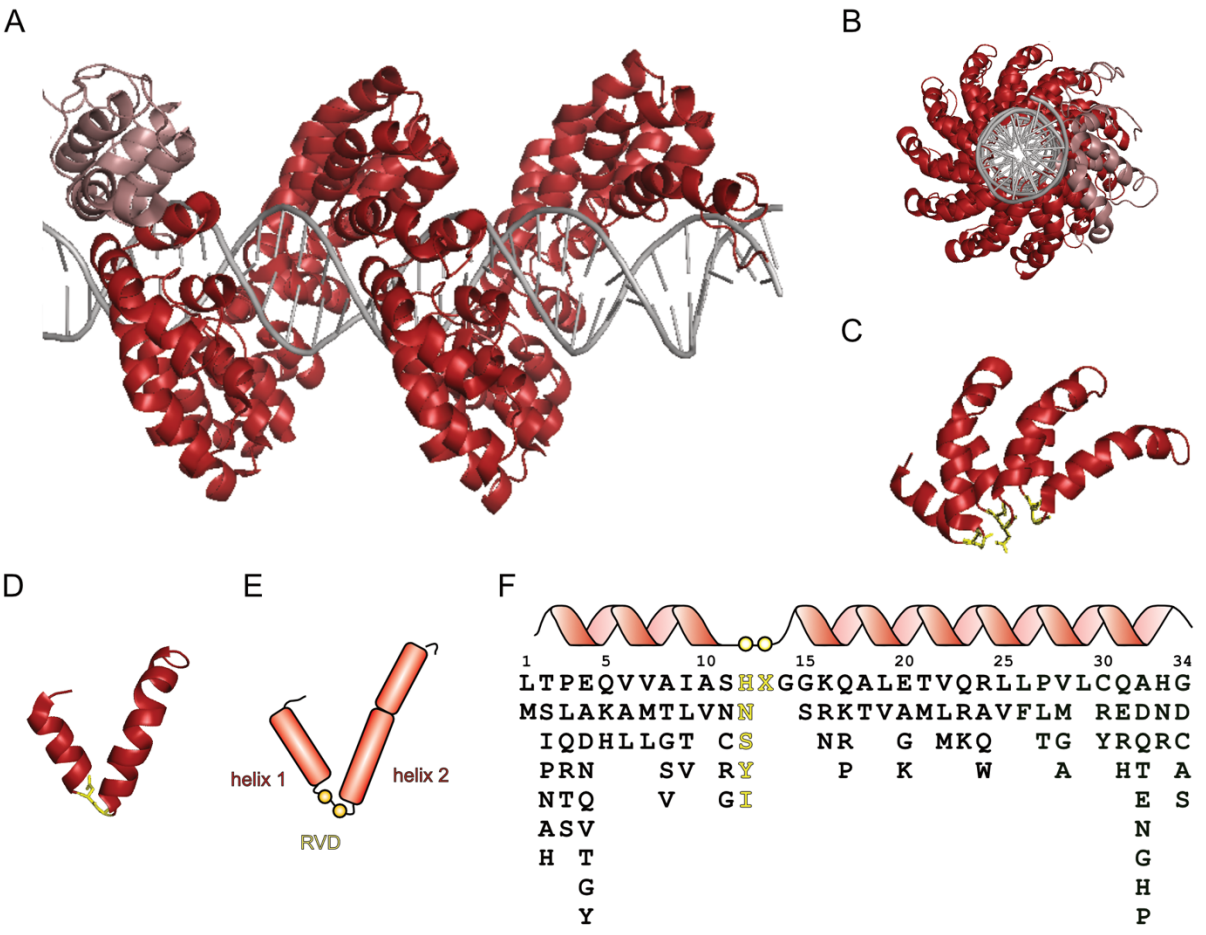


Figure 1.4: 3D structure of a TALE bound to its DNA target. 3D Structure of the TALE PthXo1 (23.5 repeats) bound to its DNA target sequence as viewed **(A)** from the front and **(B)** from the top of the protein (structural data originally published by Mak *et al*., 2012). The N-terminal part is incomplete and includes only the first two non-canonical repeats (light red), the C-terminal part is missing. **(C)** 3D structure of three TALE repeats. The RVD-loop (yellow) is oriented towards the DNA. **(D)** 3D structure, **(E)** schematic view and **(F)** amino acid sequence of a single TALE repeat. Each repeat consists of two helices that are connected by a loop structure in which the RVDs are located (yellow). The second helix of every TALE repeat is slightly kinked due to the presence of a proline at position 27. **(F)** Amino acid abundance found within >2000 analysed TALE repeats (modified after Boch and Bonas, 2010). The amino acids are sorted according to their frequency of occurrence and are shown aligned to the typical secondary structure of a TALE repeat.

and 2012). A last example is the proline in the center of the second helix of most TALE repeats. It creates an additional kink which is critical for the sequential packing and stacking of the repeats as well as for their association with the DNA double helix in general (Fig. 1.4 C-F; Mak *et al*., 2012).

Computational modeling indicated that the specificity of a TALE is mostly the result of a negative discrimination, meaning that RVDs accept nucleotides for as long as no major sterical hinderance takes place (Wicky *et al*., 2013). However, structural and experimental data suggested, at least partially, otherwise. It was shown that the RVD setup of a TALE does influence its overall activity even in absence of any obvious mismatches (Streubel *et al*., 2012; Meckler *et al*., 2013). As a result, RVDs were subcategorized into weak, intermediate and strong. The presence of the strong RVDs HD and NN, for example, increases the DNA-affinity of a TALE and at least a few of them are required to ensure activity in artificially generated TALEs (Streubel *et al*., 2012). The finding that every known natural TALE from *Xanthomonas* contains multiple strong RVDs (HD or NN) supports this experimental observation. The TALE crystal structures further elucidated the different contributions of RVDs to the overall DNA-affinity of TALEs. They showed that RVD/nucleotide combinations previously described as weak or intermediate are mostly connected by van der Waals interactions and therefore have indeed only a minor effect on the binding strength of a TALE (Deng *et al*., 2012a; Mak *et al*., 2012). RVDs previously described as strong on the other hand tend to form more stable hydrogen bonds to their respective nucleotides (Deng *et al*., 2012a; Wicky *et al*., 2013).

Computational, structural and experimental analysis also demonstrated that a general DNA-affinity is conferred by the N-terminal region of a TALE (Kay *et al*., 2007; Gao *et al*., 2012) and that this region is crucial for the initial binding of a TALE to DNA (Gao *et al*., 2012; Meckler *et al*., 2013) as well as for the process of searching and finding a target sequence (Cuculis *et al*., 2015). The N-terminal region of TALEs was further shown to contain four degenerated (or non-canonical) repeats (termed repeat 0, -1, -2 and -3) (Fig. 1.3 B; Gao *et al*., 2012). The amino acid sequences of these non-canonical repeats differ significantly from the consensus of a typical standard TALE repeat, yet, their 3D structures are remarkably similar (Gao *et al*., 2012). Repeat -1 also confers the specific recognition of the thymine that was shown to proceed nearly all natural TALE targets (T0; Fig. 1.3 B and D; Boch *et al*., 2009; Gao *et al*., 2012; Mak *et al*., 2012).

1.4 *Xanthomonas* TALEs and repeats of aberrant length

While the vast majority of *Xanthomonas* TALE repeats consist of 34aa, a few common exceptions exist. The last repeat of every TALE, for example, contains only the first 20aa of a standard repeat and is therefore referred to as last half repeat (Fig. 1.3 B; Fig. 1.5 A; Boch

and Bonas, 2010). However, despite its seemingly reduced length, it specifies for a nucleotide in the TALE target sequence like any normal repeat (Boch *et al*., 2009). This might be explained by the observation that the non-canonical amino acids in its second half still contribute to the typical helix-loop-helix structure associated with a standard TALE repeat, very much like the degenerated repeats found in the N-terminal region of a TALE (Deng *et al*., 2012a; Mak *et al*., 2012). Another exception in repeat length is found throughout most TALE-carrying *Xanthomonas* spp. It contains only 33aa and lacks the amino acid at position 13. However, the missing amino acid does not alter the function of the repeat but its RVD-specificity. Therefore, it is simply regarded as another RVD-type (most frequent example: N*; Boch and Bonas, 2010). TALE repeats with 35aa are the last common exception. They often occur in strains of *X. campestris* and *X. gardneri* and contain an additional proline at position 33 (Boch and Bonas, 2010). While some of these TALEs consist solely of 35aa repeats (e.g. Hax2 form *X. campestris* pv. *armoraciae;* Kay *et al*., 2005) others contain a mixture of 34aa and 35aa repeats (e.g. AvrHah1 from *X. gardneri*; Schornack *et al*., 2008). Interestingly, despite these differences in length and amino acid composition, an effect on the TALE-DNA interaction was not reported for any of these repeat variants.

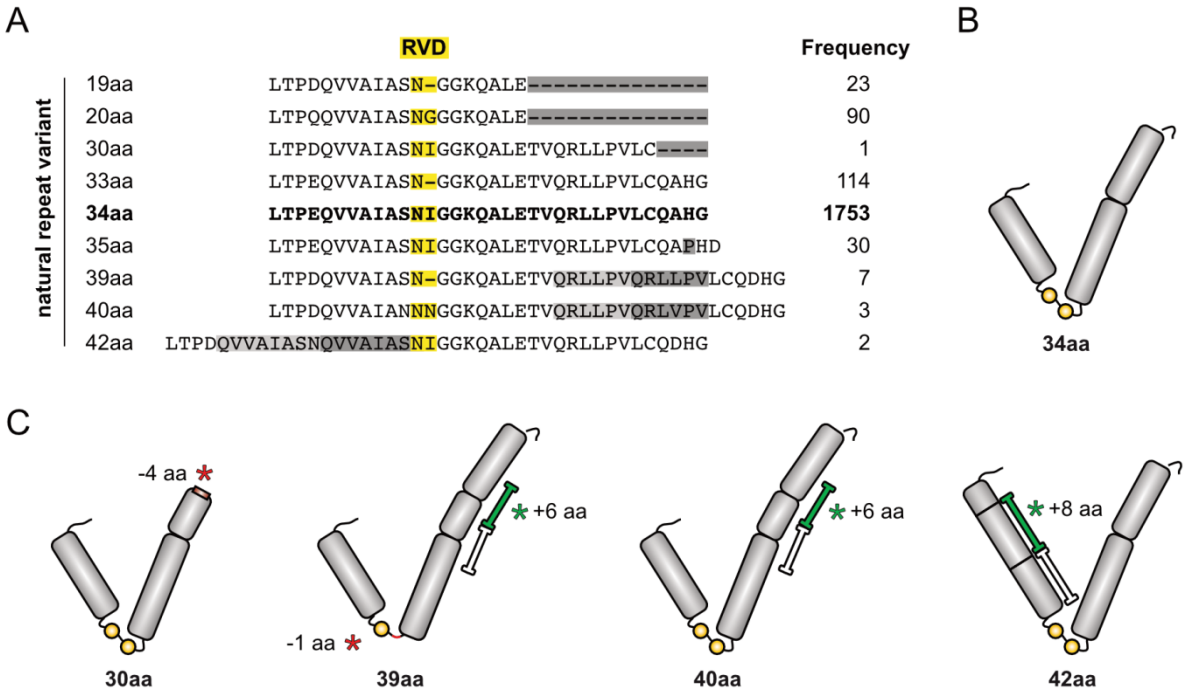


Figure 1.5: Repeat length variants found in *Xanthomonas* TALEs. (A) Repeat length variants and their frequency of occurrence as found in 113 analysed TALEs (figure is modified after Boch and Bonas, 2010). A standard repeat has a length of 34aa and is shown in bold. The 19aa and 20aa repeats correspond to the last half repeat found in the 113 analysed TALEs. Schematic view of **(B)** a standard repeat with a 34aa or **(C)** the four rare *Xanthomonas* TALE repeats of aberrant length (>20aa and <33aa or >35aa) that were shown to facilitate a frameshift recognition (Richter *et al*., 2014). Shown are the two alpha helices, the loop structure containing the RVD and the the proline that is found within the second helix of most TALE repeats (indicated by a kink). All helices are depicted elongated or shortened in accordance with the number of altered amino acids. Positions at which amino acids are missing are labeled red. Duplicated positions are shown green/white.

However, less frequently occurring repeat variants with more profound differences in length and a greater effect on the TALE-DNA interaction have been identified as well. In rare cases, a natural TALE may contain a single repeat of aberrant length. These aberrant repeats consist of 30aa, 39aa, 40aa or 42aa and are the result of small duplications or deletions within the repeat (Fig. 1.5; Boch and Bonas, 2010; Richter *et al*., 2014). Interestingly, they were shown to enable a TALE to flexibly recognize either a perfectly matching DNA target sequence or one with a single nucleotide deletion (-1 frameshift) placed corresponding to the position of the aberrant repeat (Fig. 1.6; Richter *et al*., 2014). These aberrant repeats contribute to DNA recognition and specificity when the TALE is combined with an optimal target sequence while they do not participate in DNA-binding when the TALE recognizes a -1 frameshift box (Richter *et al*., 2014). While the exact mechanism of how they confer this frameshift recognition is unknown, these observations strongly support a model in which the aberrant repeats are ejected (looped out) from the repeat array when the TALE binds to the -1 frameshift box (Fig. 1.6; Richter *et al*., 2014).

So far, all repeats of aberrant length have been identified in strains of *X. oryzae* and most of them are found in TALEs that are still without known function. Exceptions are the two TALEs AvrXa7 and PthXo3, which both contain the aberrant repeat variant with 39aa (Richter *et al*., 2014) and address a rice gene that is of crucial importance for the virulence of many *Xoo* strains. The presence of an aberrant repeat renders these two TALEs even more potent weapons for *Xanthomonas* since they allow for a much more flexible reaction to resistance mechanisms that are based on single nucleotide deletions (Richter *et al*., 2014).

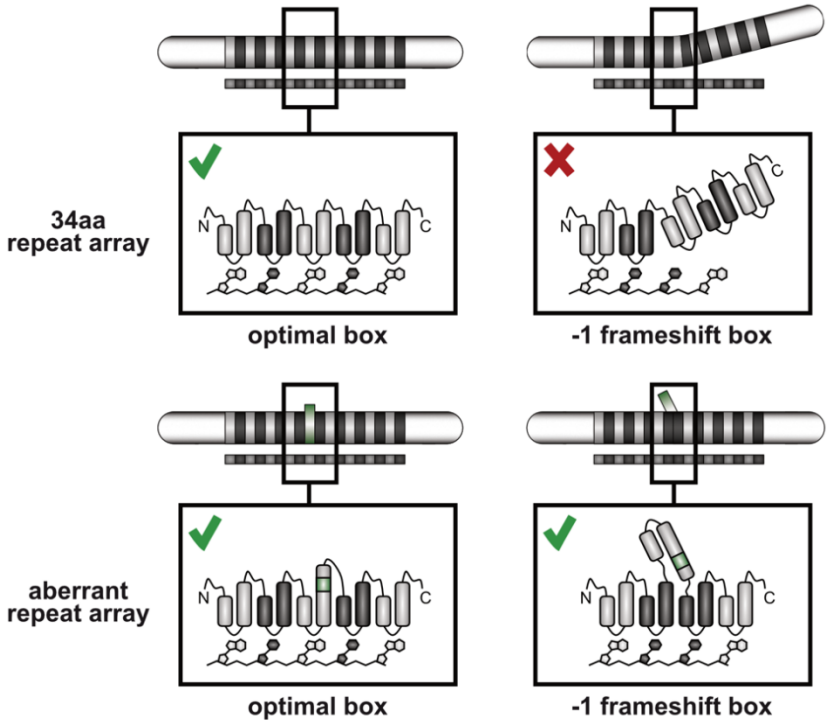


Figure 1.6: Model of a TALE repeat array containing either solely 34aa standard repeats (top) or a single repeat of aberrant length (bottom). The repeat arrays are shown in combination with either an optimal or a -1 nucleotide (-1 frameshift) box. In addition to the normal DNA-recognition, the aberrant repeat variants with 30aa, 40aa or 42aa (green) confer the recognition of a -1 frameshift box, presumably by looping out of the repeat array. Compatible TALE-box combinations are indicated by a green tick mark, incompatible ones by a red cross. The figure is partially based on Richter *et al*., 2014.

1.5 Target genes of TALEs from *Xanthomonas oryzae* and other *Xanthomonads*

When addressing the *X. oryzae* TALE repertoire, it is necessary to distinguish not only between the two pathovars *Xoo* and *Xoc* but also between the geographical origins of the specific strains. Strains from Asia have usually the highest TALE content, with numbers ranging from 24 to 29 in *Xoc* (Wilkins *et al*., 2015) and 11 to 19 in *Xoo* (Lee *et al*., 2005; Grau *et al*., 2016; Quibod *et al*., 2016). *Xoc* strains originating in Africa also contain between 20 and 23 TALEs (Wilkins *et al*., 2015), *Xoo* strains from this region, however, have a significantly smaller TALE repertoire with just 8 to 10 TALEs per strain but, interestingly, without any obvious reduction in their virulence (Gonzalez *et al*., 2007; Doucouré *et al*., 2018). In contrast, the recently isolated *Xoc* strain YNB0-17 from Asia is reported to contain also just nine TALE genes but this strain seems drastically impaired in its virulence (Cai *et al*., 2017). Of low virulence are also the *X. oryzae* strains isolated in North America (*Xo*-US), which appear to contain not even a single TALE gene (Triplett *et al*., 2011). The low pathogenicity of the *Xo*-US strains and the hypovirulence of the *Xoc* strain YNB0-17 could therefore be interpreted as a direct result of their reduced TALE repertoires. Consequently, the high number and the great diversity of TALEs in most other *X. oryzae* strains might be taken as an indication that some of them fulfill major functions in the *Xanthomonas*-rice interaction, particularly in those strains of Asian origin. However, in relation to the large number of TALEs found in these strains only few of their target genes were identified so far (Cernadas *et al*., 2014), and even fewer were shown to significantly contribute to the virulence of *X. oryzae*. While it is possible that not all TALEs possess a relevant natural function, many may also alter processes that are not easily detected under the artificial conditions of a greenhouse or by the limited phenotypic readouts used to estimate the pathogenicity of *Xanthomonas* in these experiments. Certain TALEs might also fulfill a function that is only important in a very specific genetic background of the host plant, or they could contribute only in a minor or in a redundant way to virulence, making the identification of their individual functions difficult.

Nonetheless, there are a few genes that are known to be specifically upregulated by TALEs from *X. oryzae* and whose upregulation was shown to positively affect either symptom formation or the growth and survival of the pathogen. Interestingly, all of them can be categorized into just two groups; they are either TFs or transporter proteins (Fig. 1.7). The first example is *OsTFX1*, a TF-encoding gene that is activated by TALEs from various *Xoo* strains (Fig. 1.7; Sugio *et al*., 2007; Cernadas *et al*., 2014). It is still unknown how its activation is contributing to pathogenicity but the deletion of the *OsTFX1*-inducing TALE PthXo6 in the *Xoo* strain PXO99 partially reduced the virulence of the strain (Sugio *et al*., 2007). The second example is *OsTFIIAγ1*, one of two copies of the small *γ* subunit of TFIIA in rice. It is activated by PthXo7 from the *Xoo* strain PXO99 and it was demonstrated that its

TALE-induced overexpression can negate the effects of the resistance gene *xa5* (Sugio *et al*., 2007). The first example for the second group is the sulfate transporter OsSULTR3;6 (Fig. 1.7). While its gene is specifically induced by every *Xoc* strain analysed so far, it is unknown why its overexpression leads to an increase in symptom formation (Cernadas *et al*., 2014; Wilkins *et al*., 2015). It was hypothesized that its induction could cause the removal of unwanted sulfate compounds from the apoplast, and thus prevent toxic effects for the bacteria (Fig. 1.7; Cernadas *et al*., 2014). The second example from this group is the OsSWEET family of sugar transporters (Fig. 1.7). From all known *X. oryzae* TALEs, the *OsSWEET*-gene activating ones have the most drastic impact on growth and symptom formation and are therefore referred to as major virulence factors (Yang and White, 2004; Streubel *et al*., 2013). Mutating the *OsSWEET*-inducing TALEs in a strain results in an extremely reduced symptom formation as well as hampered bacterial growth (Bai *et al*., 2000; Yang and White, 2004). Drastically decreased virulence of *Xoo* was also observed by modifying the TALE binding site in the promoter of the targeted *OsSWEET* genes (Li *et al*., 2012; Blanvillain-Baufumé *et al*., 2017). Introducing just one of the natural *OsSWEET*-inducing TALEs into a hypovirulent *Xoo*-US strain on the other hand resulted in this strain becoming highly virulent and causing disease symptoms typically associated with BB (Verdier *et al*., 2012). Experiments with artificial TALEs could even show that the overexpression of any one of the five *OsSWEET* genes within the third clade of this gene family is beneficial for *Xoo* and sufficient to promote virulence (Streubel *et al*., 2013), however, so far only three of them were actually found to be addressed by a natural TALE. These are *OsSWEET11* by PthXo1(Chu *et al*., 2006; Yang *et al*., 2006), *OsSWEET13* by

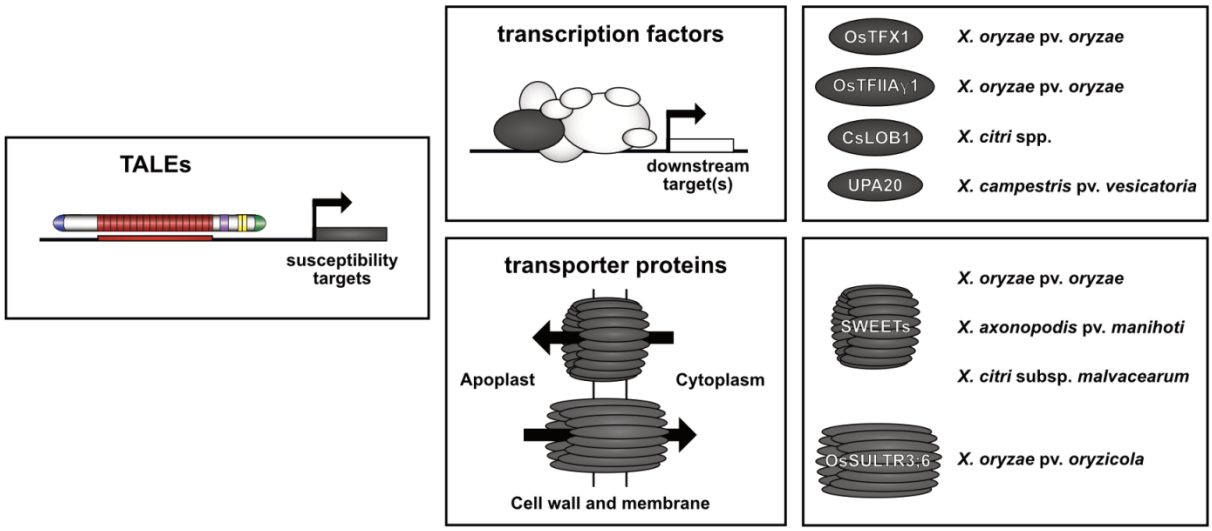


Figure 1.7: The two known types of susceptibility targets of *Xanthomonas* spp. TALEs. All confirmed susceptibility targets of *Xanthomonas* spp. TALEs can be categorized into two groups. The first group contains TFs that regulate the expression of one or multiple (mostly) unknown downstream genes or affect transcription in general. The second group contains transporter proteins that are assumed to either remove toxic substances from the apoplast by transporting them into the cell or export beneficial substances from within the cell into to apoplast where they are available for further use by the pathogen.

PthXo2 (Zhou *et al*., 2015) and *OsSWEET14* by PthXo3, AvrXa7, TalC and Tal5 (Yang and White, 2004; Anthony *et al*., 2010; Yu *et al*., 2011; Streubel *et al*., 2013). The fact that TALEs address the *OsSWEET14* promoter by at least two completely independent binding sites (Yang and White, 2004; Anthony *et al*., 2010; Yu *et al*., 2011; Streubel *et al*., 2013) suggests that *OsSWEET14* was targeted at least twice during evolution, thus further highlighting its significance. Since clade III OsSWEETs were shown to export preferentially sucrose and glucose out of cells, it is assumed that their TALE-induced overexpression leads to an increased availability of these sugars in the apoplast and/or the xylem, thus representing an easily accessible nutrition source for *Xoo* (Fig. 1.8; Chen *et al*., 2010, 2012 and 2014; Sonnewald, 2011; Baker, *et al*., 2012; Zhou *et al*., 2015).

In addition to the so far discussed susceptibility targets, several more genes have been shown to be specifically upregulated by *X. oryzae* TALEs. However, for none of them a positive effect on the virulence of *X. oryzae* was described and it remains unclear whether or not their TALE-dependent upregulation is of benefit for the pathogen. This list includes, among others, one additional transporter protein, four more TFs, a 16K transmembrane protein, several expressed proteins, a VQ-domain containing protein, two putative oxidases and one putative hydrolase. One noteworthy example from this list is *OsHEN1*, the so far only known rice gene targeted by TALEs from both pathovars, *Xoc* and *Xoo* (Moscou and Bogdanove, 2009; Cernadas *et al*., 2014; Wilkins *et al*., 2015). It is known that *HEN1* encodes for a methyl transferase that regulates the stability of small RNAs, however, while several studies addressed its function in rice and Arabidopsis (Chen *et al*., 2002;

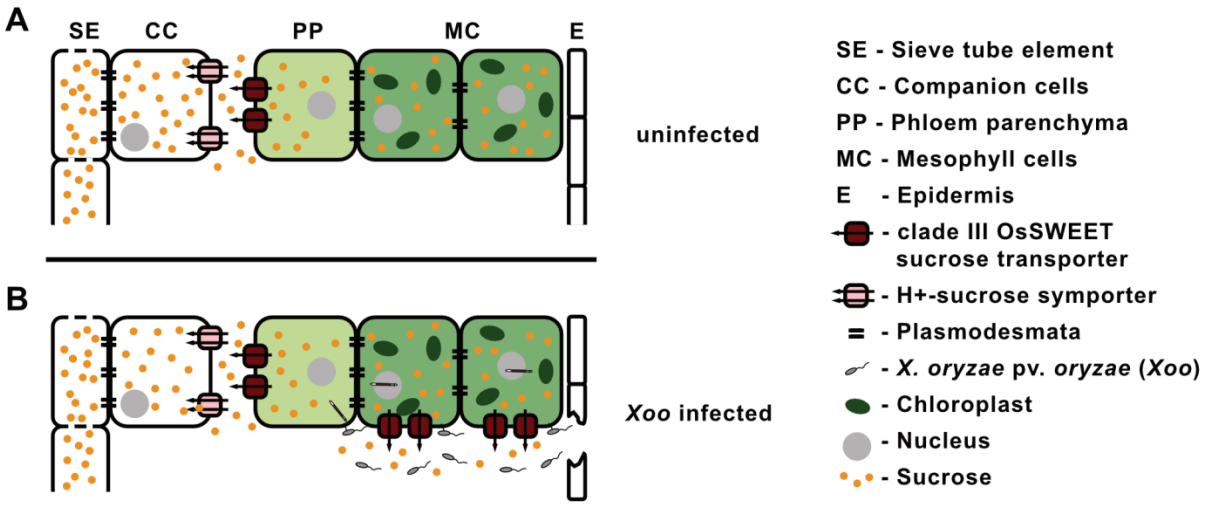


Figure 1.8: Role of clade III OsSWEET sugar transporters in the virulence of *X. oryzae* pv. *oryzae*. (A) Sucrose is produced in the photosynthetically active tissues and transported from cell to cell by plasmodesmata. From the phloem parenchyma the sucrose is exported into the apoplast by clade III OsSWEETs. Via H^+-sucrose symporter, the sucrose is transported actively and against the concentration gradient into the companion cells of the sieve tube elements (phloem). **(B)** Upon *Xoo* infection, TALEs are translocated into the plant cells to induce the expression of clade III *OsSWEET* genes. The presence of the OsSWEET sucrose exporters in the mesophyll cells leads to a sucrose efflux into the apoplast. Presumably this sucrose is used by *Xoo* as an easily accessible nutrition source. The figure shows only the infection of mesophyll cells and the subsequent colonisation of the apoplastic space but not the infection of the xylem. The figure is modified after Chen, 2014.

Tkaczuk *et al*., 2006; Abe *et al*., 2010), it still remains unknown how its induction is of benefit for *X. oryzae*. Despite these examples, most TALEs from *X. oryzae*, from *Xoo* and *Xoc* alike, remain without a known target gene. For efficient control of BB and BLS, however, it is vital to understand the infection strategies of the two *X. oryzae* pathovars as detailed as possible, especially the functionality of their effector repertoire. Therefore, it is necessary to intensify the efforts to identify additional TALE target genes and to investigate their individual functions in further detail.

Curiously, all so far discovered susceptibility targets that are addressed by TALEs from non-rice-infecting *Xanthomonas* spp. do also belong to the two above-mentioned groups, they are either transporter proteins or TFs. One prominent example is the TF-encoding gene *CsLOB1* from citrus (Fig. 1.7). Its induction by TALEs from various *X. citri* spp. (e.g. PthA4, PthAW, PthA*, PthB and PthC) was shown to be essential for the development of citrus canker disease and results in cell proliferation and pustule formation (Duan *et al*., 1999; Al-Saadi *et al*., 2007; Jalan *et al*., 2013; Hu *et al*., 2014; Pereira *et al*., 2014). Another example is *UPA20* from pepper, a TF-encoding gene that is induced by AvrBs3 from *Xanthomonas campestris* pv. *vesicatoria* (Fig. 1.7; Kay *et al*., 2007 and 2009). Its TALE-dependent overexpression causes cell enlargement (hypertrophy) and a swelling of the affected leaf-tissue, which is assumed to facilitate the development of wounds from which a further spreading of the pathogen might occur (Marois *et al*., 2002; Kay *et al*., 2007). Remarkably, two other non-rice-infecting *Xanthomonas* spp. also rely on the TALE-dependent exploitation of *SWEET* genes in their respective host plants (Fig. 1.7). The first, *Xanthomonas citri* subsp. *malvacearum*, requires the induction of *GhSWEET10* by Avrb6 for full virulence in cotton (Cox *et al*., 2017). The second, *X. axonopodis* pv. *manihoti*, induces the expression of *MeSWEET10a* via TAL20 to promote symptoms in cassava (Cohn *et al*., 2014). This renders SWEET proteins the only known virulence targets that are addressed by multiple *Xanthomonas* species in several different plant hosts. Since the expression and the presence of SWEETs was demonstrated to be of relevance for several other host-pathogen interactions as well, including Arabidpsis-*Botrytis cinerea* (Chong *et al*., 2014) and tomato-*Pseudomonas syringae* (Siemens *et al*., 2006; Chen *et al*., 2010), their exploitation might very well represent a universal mechanism that is utilized by bacterial and fungal pathogens alike.

1.6 *Xanthomonas oryzae* and TALE-based mechanisms of plant resistance

The major method to control *Xoo* in the field is the planting of rice lines that contain one of the more than 40 known resistance genes directed against BB (Suh *et al*., 2013). While their individual mechanisms vary, several of them were shown to be directly linked to the presence of specific TALEs. One mechanism for a plant to negate the function of a TALE is

to alter its binding site within the promoter of the targeted gene, thus preventing it from activating the respective susceptibility gene (Hutin *et al*., 2015a). This can happen by one or several single nucleotide polymorphisms (SNPs) that create mismatches (compare Fig. 1.9 [1-1] and [1-2A]), by single nucleotide INDELs that result in a frameshift (compare Fig. 1.9 [1-1] and [1-2B]) or by bigger insertions and deletions that change a huge part of the TALE binding site at once (compare Fig. 1.9 [1-1] and [1-2C]). Since most natural TALEs seem to tolerate at least one mismatch within their target sequence, the plant most likely needs to change several positions simultaneously in order to impair the binding of a TALE efficiently. However, minor alterations are more likely to occur and less likely to cause problems with the natural function of a gene but they are also more easily overcome by the TALE. Single amino acid substitutions frequently occur within TALE RVDs, suggesting a very high adaptation capability to minor alterations in their target site (Erkes *et al*., 2017). Bigger modifications in a TALE target site on the other hand are more difficult to counter but are also more likely to affect the regulation of the promoter and thereby the natural function of a plant gene (Hutin *et al*., 2015a). This is especially the case if a TALE targets an important regulatory region that cannot simply be altered e.g. the TATA-box. TALEs are, however, extremely sensitive towards frameshifts in their target sequence, and thus even the addition or the removal of just a single nucleotide is usually sufficient to render a TALE incapable of binding. Yet, due to their repetitive nature, TALEs are prone to rearrangements. This can lead to duplications or deletions of single or multiple repeats as well as recombination events between different TALEs (Erkes *et al*., 2017). As a result, TALEs are able to adapt even to frameshifts.

So far, only very few SNPs have been identified in the known susceptibility targets of TALEs and none were reported to actually facilitate resistance against one of the TALEs addressing this gene (Hutin *et al*., 2015a). However, there are three examples in which naturally occurring insertions or deletions within the target site of a TALE abolish its binding. The first is the resistance gene *xa25*, an allelic variant of the *OsSWEET13* promoter from the rice cultivar Nipponbare. This variant contains a single nucleotide insertion or deletion if compared to the rice cultivars Zhenshan or IR24, respectively, causing it to be inaccessible for the *OsSWEET13*-inducing TALE PthXo2 (Liu *et al*., 2011; Richter *et al*., 2014; Zhou *et al*., 2015). The second one is an allele of *OsSWEET14*, named *xa41(t)*, which contains an 18 bp deletion that confers resistance against at least two of the known TALEs targeting *OsSWEET14*, AvrXa7 and Tal5 (Hutin *et al*., 2015b). The last example is *xa13*, an allelic variant of *OsSWEET11* that was shown to contain a 243 bp insertion that renders it inaccessible for the TALE PthXo1 (Chu *et al*., 2006; Yang *et al*., 2006).

A more elaborate resistance mechanism based on TALEs is the modification of host components that are crucial for their general function, however, the only known example of this principle is *xa5* (Schornack *et al*., 2013). It encodes for a slightly modified version of the

general transcription factor TFIIAγ5 (Fig. 1.9 [2-1] and [2-2]; Iyer and McCouch, 2004). Interestingly, both versions differ only in a single amino acid, yet, *xa5* drastically reduces the gene-inducing capabilities of TALEs without affecting the regular transcription of the plant (Iyer and McCouch, 2004; Sugio *et al*., 2007; Gu *et al*., 2009). The fact that TFIIAγ is a direct interaction partner of the TFB located within the C-terminal region of TALEs (Yuan *et al*., 2016; Ma *et al*., 2018) strongly implies that this altered version of TFIIAγ5 cannot interact with TALEs anymore, and thereby prevents them from efficiently stimulating PIC formation.

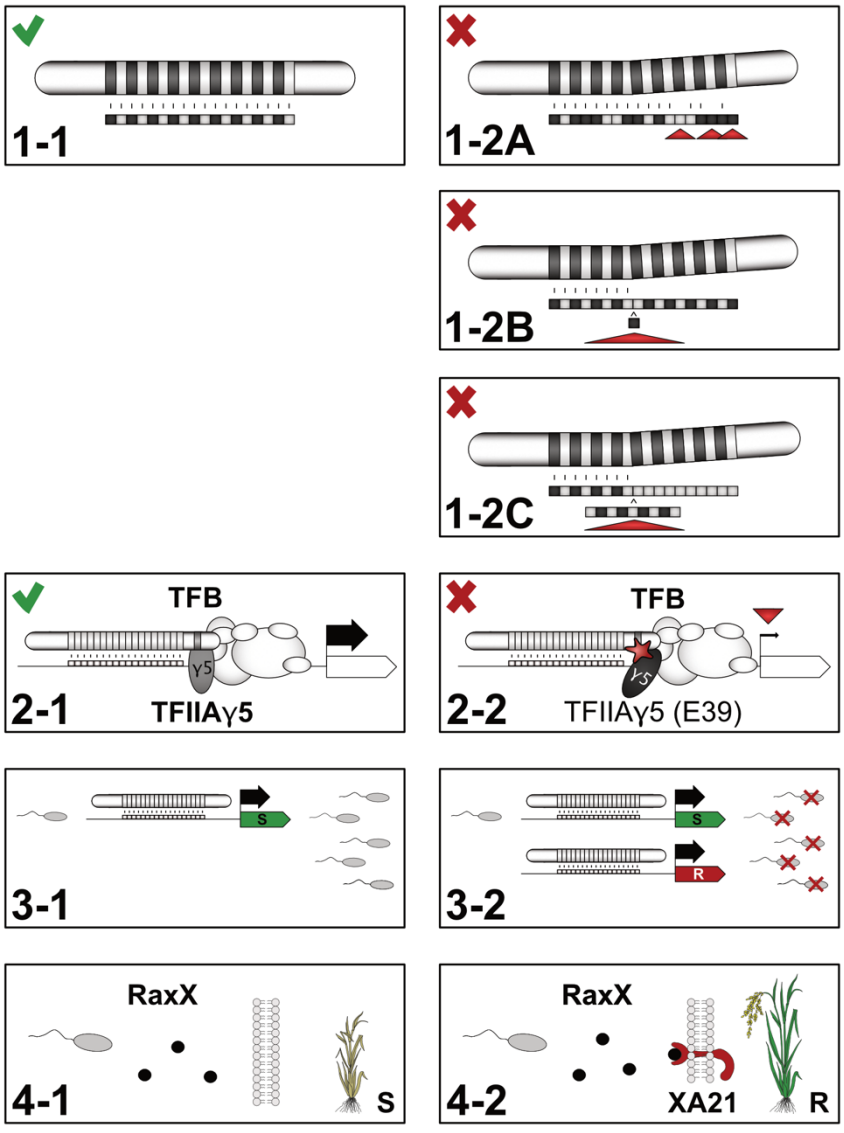


Figure 1.9: Rice resistance mechanisms aimed against *X. oryzae*. The shown resistance mechanisms are aimed at either *X. oryzae* TALEs and their function or the pathogen in general**. [1-1], [2-1]** and **[3-1]** The TALE-dependent induction of susceptibility targets leads to an increased growth of the pathogen, however, the plant evolved different mechanisms to avoid these effects. **[1-2A]** One or several single nucleotide polymorphisms (SNPs) within the target sequence of a TALE reduce its DNA-binding affinity and prevent the activation of the targeted susceptibility gene. **[1-2B]** A single nucleotide insertion or deletion (INDEL) prevents TALE binding by generating a frameshift that brings part of the TALE out-of-sync. **[1-2C]** A large insertion or deletion disrupts the TALE-binding site, preventing the TALE from binding and activating the targeted susceptibility gene. **[2-2]** By modifying crucial residues in a TALE-interaction partner, in this case the 39th residue of the TFIIA gamma subunit on chromosome 5, the plant impairs TALEs in their general gene-activating capabilities. **[3-2]** Like a susceptibility gene, executor *R*-genes contain a binding site for a TALE. However, instead of benefiting the pathogen, their activation causes a strong resistance reaction, most commonly by triggering a local cell-death, thus preventing the pathogen from further spreading. **[4-2]** Rice plants can also rely on an early detection of *X. oryzae* e.g. by recognizing the presence of specific peptides secreted by the pathogen. Upon detection by an adequate protein, usually a wall-associated or leucine-rich repeat containing receptor kinase (WAK/LRR-RK) a resistance reaction is triggered, thus preventing further growth and spreading of the pathogen. The illustrated example shows the recognition of the small bacterial peptide RaxX by the receptor like kinase XA21 (S – susceptible plant; R – resistant plant).

Remarkably, *Xanthomonas* found a way to counter this resistance mechanism by using yet another TALE. In addition to *TFIIAγ5* on chromosome 5, rice contains a close paralogue of *TFIIAγ* on chromosome 1, named *TFIIAγ1* (Iyer and McCouch, 2004). By inducing the expression of this paralogue, TALE PthXo7 from *Xoo* strain PXO99 is able to restore general TALE functionality despite the presence of the resistant version of *TFIIAγ5*, indicating that the increased availability of TFIIA*γ1* is sufficient to substitute for the loss of *TFIIAγ5* as an interaction partner (Sugio *et al*., 2007).

All so far mentioned resistance mechanisms have one disadvantage, for the plant and for breeders alike, they are not dominant in their effect. For as long as the target gene has at least one TALE-inducible allele, its induction is still sufficient for virulence. However, dominant resistance mechanisms against TALEs exist as well. One frequently observed way is a promoter trap in which a TALE induces the expression of an executor *R*-gene. Inducing the expression of an executor *R*-gene causes a resistance reaction e.g. in form of a local cell death, ultimately leading to the bacteria being contained, and thus prevented from further spreading (compare Fig. 1.9 [3-1] and [3-2]; Schornack *et al*., 2013). Examples for such *R*-genes in rice are *Xa10*, *Xa23* and *Xa27* which are activated by the TALEs AvrXa10, AvrXa23 and AvrXa27, respectively (Gu *et al*., 2005; Rouxel and Balesdent, 2010; Wang *et al*., 2014 and 2015). All three resistance proteins are relatively short and contain multiple hydrophobic membrane spanning domains (Zhang *et al*., 2015). *Xa10* encodes a protein of 126 amino acids that localizes to the membrane of the endoplasmic reticulum (ER) and that induces a local cell death reaction by triggering Ca^{2+} depletion (Tian *et al*., 2014). *Xa23* encodes for a protein of 113 amino acids that shows ~50% identity to *Xa10* and possesses a similar functional mechanism (Wang *et al*., 2017). *Xa27* has also 113 amino acids but it shows neither sequence homologies to *Xa10* or *Xa23* nor to any other proteins outside the genus *Oryza* (Gu *et al*., 2005; Bimolata *et al*., 2015). Besides the fact that the outcome of the *Xa27* resistance reaction shows a dosage dependency and that it is affected by the developmental stage of the rice plant, nothing is known about its mechanism (Gu *et al*., 2009; Liu *et al*., 2014; Zhang *et al*., 2015). In order to avoid detection by these promoter traps *Xanthomonas* would need to alter the specificity of the affected TALE, but doing so without losing too much activity on the initial susceptibility target. Alternatively, *Xanthomonas* could lose this particular TALE completely. However, if this TALE is crucial for virulence, its loss might affect the pathogen as severely as a triggered resistance gene. In addition to these examples, many other resistance genes were found whose functionality is completely independent from the presence of a TALE and several of them encode for either leucine-rich repeat (LRR) receptor kinases or wall-associated kinases (WAK), e.g. *Xa3*, *Xa4*, *Xa21* and *Xa26* (compare Fig. 1.9 [4-1] and [4-2]; Zhang *et al*., 2015). It is assumed that they recognize specific elicitor peptides or proteins from *Xoo* but so far only one example was characterized

in more detail: RaxX, a small, highly conserved and type I secreted bacterial protein that is specifically recognized by the rice protein XA21 (Pruitt *et al*., 2015).

Due to the high adaptability and the immense genetic diversity of *Xoo*, no resistance gene was shown effective against all *Xoo* strains and many have proven to be not permanent in their effect (Vera Cruz *et al*., 2000). As a result, researchers and breeders constantly search for novel resistance mechanisms. They investigated the utilization of resistance genes derived from other cereals like maize (Zhao *et al*. 2005) or stacked different resistance mechanisms, especially those known to recognize important bacterial virulence factors like *OsSWEET*-inducing TALEs (Huang *et al*., 1997; Römer *et al*., 2009a). Following a similar line of thought, the binding sites of multiple highly conserved *Xoo* TALEs were introduced into the promoter region of the resistance gene *Xa27*, thus preventing the pathogen from easily avoiding recognition by the loss of just a single TALE (Hummel *et al*., 2012; Liu *et al*., 2014). Interestingly, all of the more than 40 known BB resistance genes are ineffective against *Xoc* (Cai *et al*., 2017) and not even a single specific TALE/*R*-gene combination has been identified within the *Xoc*-rice pathosystem. Inoculation experiments even revealed that some *Xoc* strains are able to supress the function of selected BB resistance genes despite the presence of an adequate elicitor TALE from *Xoo* (Makino *et al*., 2006), thus giving one possible explanation for the apparent lack of functional *R*-genes against *Xoc*.

1.7 AnnoTALE – a unifying TALE nomenclature

The larger number and great diversity of TALE genes makes it extremely difficult to clearly name and categorize individual TALEs, especially in the rice-infecting *X. oryzae* pathovars *Xoo* and *Xoc*. Additionally, various different naming systems for TALEs exist. Historically, TALE names signified involvement in pathogenicity (“Pth” as in PthXo7) or resistance (avirulence, “Avr” as in AvrXa7). Nowadays on the other hand, with the increased availability of completely sequenced *Xanthomonas* genomes, TALEs/*TALE* genes are named simply by size (e.g. TalA, TalB, TalC in Yu *et al*., 2011) or, even more commonly, by numbering them according to their orientation and location within the respective genome (e.g. Tal1, Tal2a, Tal2b and Tal2c in Salzberg *et al*., 2008). However, due to their repetitive nature and their tendency to associate with transposable elements, *TALE* genes are often affected by genome rearrangements. As a result, duplications, deletions and inversions as well as changed genomic positions of *TALE* genes are regularly observed (Salzberg *et al*., 2008; Ferreira *et al*., 2015; Grau *et al*., 2016; Erkes *et al*., 2017). With the naming systems most commonly in use, this unavoidably leads to situations in which completely unrelated TALEs with different RVD sequences end up with identical names or numbers, while at the same time even identical TALEs can have different names. Aware of those difficulties, researchers started to sub-group individual TALEs based on similarity to prominent TALEs (e.g. PthXo2-

like, AvrXa7-like), however, since this covered only some of the known TALE sequences, the problem remained largely unsolved.

In search for a more lasting solution, scientists addressed the problem bioinformatically. The program AnnoTALE is one of the results (Grau *et al*., 2016). It allows to sort TALEs according to their RVD similarities and to group those that are likely to share the same target gene(s). In order to do so, the program uses a new and unifying naming system for TALEs. Every TALE is assigned a unique name that is based on two letters representing the TALEs class and a subsequent number for every individual TALE member within this class. Following this nomenclature, the first two TALEs of class TalAH are for example named TalAH1 and TalAH2 (Fig. 1.10). The advantage is obvious: while the class affiliation demonstrates a close similarity between the individual class members, the subsequent numbers ensures nonetheless that each unique TALE sequence is also assigned a unique name. By taking into account the genomic positions and the regions adjacent to TALE genes, the new AnnoTALE nomenclature was also used to unify and simplify the numbering of genomic TALE clusters within *Xoo* strains, thus simplifying the comparison of different *Xanthomonas* strains (Fig. 1.11; Grau *et al*., 2016). AnnoTALE can also highlight the differences between individual class members and display their phylogenetic distance while its "predict and intersect" application even allows to search within any given set of sequences for putative TALE binding sites (Grau *et al*., 2016). A second study pursued a similar approach and presented QueTAL (Pérez-Quintero *et al*., 2015). It can be used as well to compare and categorize TALE sequences into families and to illustrate their relationship in a phylogenetic tree. For this, the program generates families that are based either on similarities within the TALE repeat sequences (DisTAL) or on the likelihood that the TALEs share a target gene (FuncTAL) (Pérez-Quintero *et al*., 2015). However, due to the easier handling, the benefit of a definite nomenclature as well as several additional functions, the present work was done using AnnoTALE and its subsequent nomenclature. A list of all *Xoo* and *Xoc* TALE sequences and their names (old and new) is available online (http://www.jstacs.de/downloads/List_of_classes.txt).

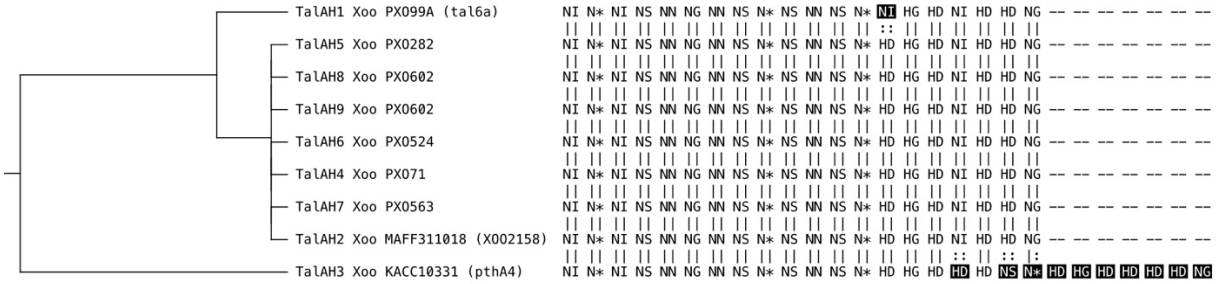


Figure 1.10: Example for the new AnnoTALE nomenclature. The members of a representative TALE class (TalAH) are shown. Their new name, pathovar and strain of origin are indicated as well as the former name or number (brackets). RVD changes between TALEs are highlighted in black.

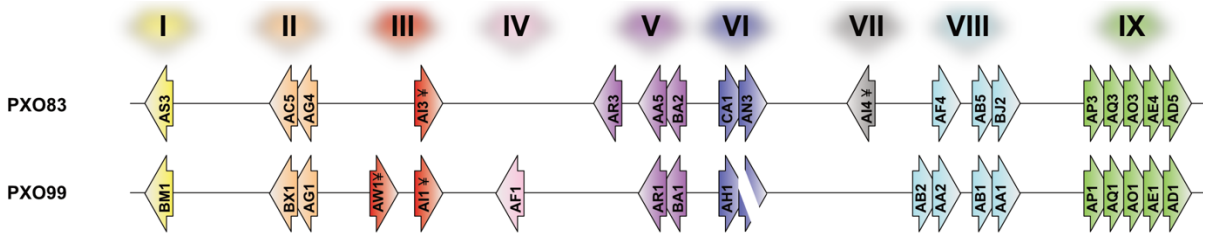


Figure 1.11: Distribution of TALE genes in the two *Xoo* strains PXO83 and PXO99. TALEs are indicated as arrows with regard to their orientation in the genome. TALE classes are named according to the AnnoTALE classification system. TALE classes were assigned into TALE clusters according to their genomic position and adjacent genes. Clusters are indicated by colours. TALEs with incomplete N- and C-terminal regions are labeled with ¥, otherwise fragmented TALE genes are indicated by a broken arrow. The figure is based on Grau *et al*., 2016.

1.8 TALEs and their application in biotechnology

After the code for the sequence specificity of TALEs was unravelled, the scientific community was quick to realise how much potential underlay the discovery. Their extremely regular binding mechanism and the simple and predictable way their DNA specificity is conferred, render TALEs something rare – DNA-binding proteins with an easily adjustable specificity. Additionally, due to numerous efforts, many of the original limitations for their application were successfully overcome. This includes the difficulties of cloning the highly repetitive TALE repeats (Cermak *et al*., 2011; Geißler *et al*., 2011; Li *et al*., 2011; Morbitzer *et al*., 2011; Weber *et al*., 2011; Zhang *et al*., 2011b), the necessity for a T_0 proceeding the target sequence (Doyle *et al*., 2013; Tsuji *et al*., 2013) or the initial lack of a RVD with an exclusive specificity for guanine (Streubel *et al*., 2012; Moore *et al*., 2014; Yang *et al*., 2014). As a result, researchers generated a whole set of TALE-based tools suited for numerous biotechnological applications. In addition to their use as activators (Morbitzer *et al*., 2010; Zhang *et al*., 2011a), the DNA-binding domain of TALEs was also fused to repressor domains to supress the expression of genes (Cong *et al*., 2012; Mahfouz *et al*., 2012; Politz *et al*., 2013), combined with a histone demethylase to modify the state of chromatin (Konermann *et al*., 2013; Mendenhall *et al*., 2013), used to visualize specific DNA sequences or chromatin alterations (Ma et al., 2013; Ren *et al*., 2017) and adjusted to specifically recognize methylated DNA sequences (Deng *et al*., 2012b; Valton *et al*., 2012; Tsuji *et al*., 2016). However, the most frequently used TALE-based tool remains a fusion to the nuclease domain of the restriction enzyme *Fok*I that is called TALE nuclease (TALEN; Christian *et al*., 2010; Li *et al*., 2011; Miller *et al*., 2011). Applied in pairs, with each protein containing one-half of a *Fok*I dimer, TALENs achieve a very high target specificity and their extensive utilization in genome editing ranges from plants and animals to human cell lines (reviewed in: Scharenberg *et al*., 2013; Schornack *et al*., 2013).

1.9 Aim of the study

This study aims to characterize the known aberrant repeats found in *Xanthomonas* TALEs and their effects on DNA-binding in further detail. This includes investigating the ability of TALEs to tolerate multiple aberrant repeats placed in tandem or distributed throughout the repeat region. In this regard, it will also be clarified if multiple aberrant repeats placed within a single TALE can confer the recognition of target sequences containing frameshifts of more than one nucleotide. To assess which regions of a TALE repeat allow for the insertion of additional amino acids, another part of this work will be to generate artificial aberrant repeats by introducing small duplications or insertions at different positions throughout a TALE repeat and to investigate how these alterations influence the DNA-binding of TALEs. Taken together these experiments should shed light on the impact individual repeats of aberrant length have on the ability of TALEs to flexibly recognize multiple target sequences. These findings might also help to understand adaptations of TALEs to „binding site"-based resistances in their target genes better and to improve their application in biotechnology, e.g. by enabling them to recognize different alleles of a single target gene.

Another part of this study will be the analysis of the available TALE sequences from all completely sequenced *X. oryzae* pv. *oryzicola* strains with regard to their natural target genes in rice. This will include confirming the known target genes by testing for a direct interaction between TALE and target gene promoter as well as identifying completely new ones. For this, a set of artificial *Xoc* TALEs will be constructed and analysed via GUS reporter assays. Expanding the knowledge about the TALEs from *Xoc* and their target genes in rice will help to improve our general understanding of this agronomically important pathogen-host interaction. More detailed information about the pathways and mechanisms that are specifically altered by TALEs should also allow for the identification of new ways to generate more resistant rice varieties.

2 Material and Methods

2.1 Materials

2.1.1 Media

Table 2.1: Used media and their composition.

media	contents	usage
AIM	10 mM MES (pH 5,4); 10 mM magnesium chloride; 150 µM acetosyringon	*A. tumefaciens* infiltration media
LB	1% (w/v) bacto tryptone; 0.5% (w/v) yeast extract; 1% (w/v); sodium chloride; pH 7.5	*E. coli* growth media
PSA	1% (w/v) petone; 1% (w/v) saccharose; 0.1% (w/v) glutamic acid; pH 7.0	*X. oryzae* growth media
YEB	0.5% (w/v) bacto meat extract; 0.1% (w/v) bacto yeast extract; 0.5% (w/v) bacto peptone; 0.5% (w/v) sucrose; 0.2% 1 M magnesium sulphate; pH 7.2	*A. tumefaciens* growth media

For solid media plates, 1.5% (w/v) agar was added to the liquid media. For bacterial selection, antibiotics were added. For the used antibiotics concentrations see table 2.1.2.

2.1.2 Antibiotics

Table 2.2: Used antibiotics and other media additives.

antibiotic/other additive	final concentration
ampicillin	100 µg/ml (solid media) or 50 µg/ml (liquid media)
gentamycin	20 µg/ml
hygromycin	50 µg/ml
kanamycin	100 µg/ml (for *A. tumefaciens*) or 25 µg/ml (for *E. coli*)
rifampicin	100 µg/ml
spectinomycin	30 µg/ml
X-gal	0.004% (w/v)

2.1.3 Bacterial strains

Table 2.3: Used bacterial strains.

bacterial strain	genotype/description	reference
Escherichia coli		
One Shot TOP10	F-, mcrA, Δ(mrr-hsdRMS-mcrBC), Φ80lacZΔM15, ΔlaxX74, recA1, araD139, Δ(ara-leu)7697, galU, galK, rpsL(StrR), endA1, nupG	Invitrogen
One Shot® ccdB Survival™ T1R	F- , mcrA, Δ(mrr-hsdRMS-mcrBC), Φ80lacZΔM15, ΔlaxX74, deoR, recA1, araD139, Δ(ara-leu)7697, galU, galK, rpsL(StrR), endA1, nupG, tonA::PtrcccdA	Invitrogen
Agrobacterium tumefaciens		
GV3101	Ti-Plasmid:pMP90 (pTiC58ΔT-DNA), rifampicin resistance, gentamycin resistance	Larebeke *et al*., 1974; Koncz and Schell, 1986
Xanthomonas oryzae pv *oryzae*		
BAI3	rifampicin resistance	Gonzalez *et al*., 2007
BAI3Δ*talC*	rifampicin resistance	Yu *et al*., 2011
Roth X1-8	rifampicin resistance	Triplett *et al*., 2011

E. coli was grown at 37°C using LB media. *A. tumefaciens* GV3101 was grown at 28°C using YEB media. *Xoo* strains were grown at 28°C using PSA media. Bacteria were cultivated using antibiotics matching the resistance cassettes on their plasmids.

2.1.4 Plant material

Nicotiana benthamiana (*N. benthamiana*) plants were grown under greenhouse conditions with 16 h of light, relative humidity of 40-60% and temperatures of 23°C and 19°C, during daytime and night, respectively. Plants were used for inoculation experiments after 4-6 weeks of growth.

Oryza sativa seeds seeds were placed on a wet tissue and incubated at room temperature without light for 3-4 days. After germination, the resulting rice plantlets were placed in soil and grown under greenhouse conditions with 16 h of light, relative humidity of ~70% and temperatures of 28°C and 25°C, during daytime and night, respectively. Plants were used for inoculation experiments after 3-5 weeks of growth. Experiments were performed using the four *O. sativa* cultivars (cv.) IR24, Nipponbare, Sadu Cho and Zhenshan.

2.1.5 Oligonucleotides

Table 2.4: List of used oligonucleotides.

oligonucleotide	sequence	description
oligonucleotides for sequencing		
M13F	GTAAAACGACGGCCAG	for pUC57
M13R	CAGGAAACAGCTATGAC	for pUC57
Assembly seq R	GCTCACATGTTCTTTCCTGC	for assembly vectors
gus 5´RACE	GAACTGATCGTTAAAACTGCCTG	for pGWB3GG
GGC_Module_R	CCTCTTCGCTATTACGCCAG	for pSKA2
GGC_Module_F	GTTAGCTCACTCATTAGGCAC	for pSKA2
oligonucleotides to modify the overhangs of single repeat modules (sorted by position)		
7-F 1-R	TTTGAAGACTTTTACCCCGGAGCAGGTGGTGGCC TTTGAAGACTTTCAGGCCATGGGCCTGGCACAGCAC	p1 repeat modules
2-F 2-R	TTTGAAGACTTCTGACCCCGGAGCAGGTGGTG TTTGAAGACTTCGGTGTCAGGCCATGGGCCTGGCACAGCAC	p2 repeat modules
3-F 3-R	TTTGAAGACTTACCGGAGCAGGTGGTGGCCATCGCC TTTGAAGACTTGGTGAGGCCA TGGGCCTGGCACAGCAC	p3 repeat modules
4-F 4-R	TTTGAAGACTTCACCCCGGAGCAGGTGGTGGCC TTTGAAGACTTGAGTCAGGCCA TGGGCCTGGCACAGCAC	p4 repeat modules
5-F 5-R	TTTGAAGACTTACTCCGGAGCAGGTGGTGGCCATC TTTGAAGACTTATGGGCCTGGCACAGCACCGG	p5 repeat modules
6-F 6III-	TTTGAAGACTTCCA TGGCCTGACCCCGGAGCAGGTGG TTTGAAGACTTTCCAGCGCCTGCTTGCC	p6 and stop repeat modules
oligonucleotides to modify the RVD of single repeat modules (sorted by RVD)		
rep alle links	ATGGCCACCACCTGCTCCGG	universal forward primer
rep_mut_AN_rechts	CGCCAGCGCCAACGGCGGCAAGCAGGCGCTGGAGACG	RVD AN
rep_mut_HA_rechts	CGCCAGCCACGCTGGCGGCAAGCAGGCGCTGGAGACG	RVD HA
rep_mut_HD	CGCCAGCCACGATGGCGGCAAGCAGGCGCTGGAGACG	RVD HD
rep_mut_HH_rechts	CGCCAGCCATCACGGCGGCAAGCAGGCGCTGGAGACG	RVD HH
rep_mut_HM_rechts	CGCCAGCCACATGGGCGGCAAGCAGGCGCTGGAGACG	RVD HM
rep_mut_HN_rechts	CGCCAGCCACAATGGCGGCAAGCAGGCGCTGGAGACG	RVD HN
rep_mut_HY_rechts	CGCCAGCCACTATGGCGGCAAGCAGGCGCTGGAGACG	RVD HY
rep mut N*	CGCCAGCAATGGCGGCAAGCAGGCGCTGGAGA	RVD N*
rep_mut_NA_rechts	CGCCAGCAATGCCGGCGGCAAGCAGGCGCTGGAGACG	RVD NA
rep_mut_NC_rechts	CGCCAGCAATTGCGGCGGCAAGCAGGCGCTGGAGACG	RVD NC
rep_mut_ND_rechts	CGCCAGCAACGATGGCGGCAAGCAGGCGCTGGAGACG	RVD ND
rep_mut_NG	CGCCAGCAATGGCGGCGGCAAGCAGGCGCTGGAGACG	RVD NG
rep_mut_NI	CGCCAGCAATATTGGCGGCAAGCAGGCGCTGGAGACG	RVD NI
rep_mut_NN	CGCCAGCAATAACGGCGGCAAGCAGGCGCTGGAGACG	RVD NN
rep_mut_QD_rechts	CGCCAGCCAAGATGGCGGCAAGCAGGCGCTGGAGACG	RVD QD
rep_mut_SN_rechts	CGCCAACAGTAACGGCGGCAAGCAGGCGCTGGAGACG	RVD SN
rep_mut_YG_rechts	CGCCAGCTATGGCGGCGGCAAGCAGGCGCTGGAGACG	RVD YG
oligonucleotides for the generation of single repeat modules for natural aberrant repeats (sorted by aberrant repeat variant)		
37er_v1_F_Rep1	CCCCATGGCCTGAAAGTCTTCAAA	37aa v1 p1 F
37er_v1_F_Rep2	CCCCATGGCCTGACACCGAAGTCTT	37aa v1 p2 F
37er_v1_F_Rep3	CCCCATGGCCTCACCAAGTCTTCAAA	37aa v1 p3 F
37er_v1_F_Rep4	CCCCATGGCCTGACTCAAGTCTTCAAA	37aa v1 p4 F
37er_v1_R	GCCATGGGCCTGGCACAGCA	37aa v1 R

37er_v2_F 37er_v2_R	GTGCAGCGGCTGTTGCCGGTGCTGTGCC CCGCTGCACCGTCTCCAGCGCCTGC	37aa v2
36aaVar1_F 36aaVar1_R	CCGGTGCTGTGCCAGGCCCATGGC CACCGGCAACAGCCGCTGCACCGTC	36aa v1
36erVar2_F 36erVar2_R	GCCCATGCCCATGGCCTGACACCGAAG CTGGCACAGCACCGGCAACAGCCGC	36aa v2
35erVar1_F 35erVar1_R	CCGCATGATCTGACACCGAAGTCTTC GGCCTGGCACAGCACCGGCAA	35aa v1
35erVar2_F 35erVar2_R	CCGCCGTACGGCCTGACACCGAAGTCTTC CTGGCACAGCACCGGCAACAGCCGCTG	35aa v2
F1_30er R1_30er	TTTGAAGACTTTTACCCCGGACCAGGTGGTGGCC TTTGAAGACTTTCAGGCACAGCACCGGCAACAGC	30aa p1
F2_30er R2_30er	TTTGAAGACTTCTGATCCCGGACCAGGTGGTGGC TTTGAAGACTTCGGTGTCAGGCACAGCACCGGCAACAGC	30aa p2
F3_30er R3_30er	TTTGAAGACTTACCGGACCAGGTGGTGGCCATCG TTTGAAGACTTGGTGAGGCACAGCACCGGCAACAGC	30aa p3
F4_30er R4_30er	TTTGAAGACTTCACCCCGGACCAGGTGGTGGCCA TTTGAAGACTTGAGTCAGGCACAGCACCGGCAACAGC	30aa p4
1_28er_F 1_28er_R	GTGCTGTGCCAGGCCCATGG CGTCTCCAGCGCCTGCTTGC	28aa
hexa mut 40er links	CGGCAACAGCCGCTGCACCGTCTCCAAAGTCTTCG	40aa AV universal L
hexa-mut 40er rechts LA	GTACAGCGGCTGGTGCCGGTGCTGTGCCAGG	40aa AV LA R
hexa mut 40er rechts AB	pho GTACAGCGGCTGGTGCCGGTGCTGTGCCAGGTGA	40aa AR AB R
hexa-mut 40er rechts BC	pho GTACAGCGGCTGGTGCCGGTGCTGTGCTGAGAC	40aa AV BC R
oligonucleotides for the generation of single repeat modules for artificial aberrant repeats (sorted by aberrant repeat variant)		
1_48er_F 1_48er_R	CACCGGCAACAGCCGCTGCACCGTCTCCAGCGCC CAGCGGCTGGTGCCGGTGCTGTGCCAGGCCCATGGC	48aa
Rep3_link08_7nA_F Rep3_link08_7nA_R	CAGGTGGTGGCCATCGCCAGCAATAACGGTGGCAA GCTGCCACCGCCACCCGAGCTCTCCGGTAAGTCTT	42aa art1
Rep3_8aa_12nA_F Rep3_8aa_12nA_R	AATAGCTCGGGTGGCGGTGGCAGCAATAACGGTGGCAA GCTGGCGATGGCCACCACCTGCTCCGG	42aa art2
link06_22nA_F link06_22nA_R	AGCGGTGGCGGTGGCAGCCAGCGGCTGTTG CACCGTCTCCAGCGCCTGCTTGCCACC	40aa art1
link06_28nA_F link06_28nA_R	AGCGGTGGCGGTGGCAGCCTGTGCCAGGCC CACCGGCAACAGCCGCTGCACCGTCTCCAGCG	40aa art2
36er Art v1 F 36er Art v1 R	CAGGTGGTGGCCATCGCCAG CTCCGGCTCCGGTAAGTCTTC	36aa art1
36er Art v2 F 36er Art v2 R	GCCATCGCCAGCAATAACGGTG GATGGCCACCACCTGCTCC	36aa art2
36er Art v3 F 36er Art v3 R	CAGGCGCTGGAGACGGTGCA CGCCTGCTTGCCACCGTTATTG	36aa art3
36er Art v4 F 36er Art v4 R	CAGCGGCTGTTGCCGGTGC CCGCTGCACCGTCTCCAGC	36aa art4
Rep3_link07_4nA_F Rep3_link07_04nA_R	CAGGTGGTGGCCATCGCCAGCAATAACGGTGGCAA GCTGCCACCGCCACCCGAGCTCTCCGGTAAGTCTT	3NN (7aa #4)
R3_L01_05nA_F R3_L01_05nA_R	GTGGTGGCCATCGCCAGCAA CGACTGCTCCGGTAAGTCTTC	3NN (1aa #5)
Rep3_link07_5nA_F Rep3_link07_5nA_R2	GTGGTGGCCATCGCCAGCAATAACGGTGGCAAGC GCTGCCACCGCCACCCGAGCTCTGCTCCGGTAAGT	3NN (7aa #5)
Rep3_link07_07nA_F Rep3_link07_07nA_R	GCCATCGCCAGCAATAACGGTGGCAAGCAGGC GCTGCCACCGCCACCCGAGCTCACCACCTGCT	3NN (7aa #7)
Rep3_link07_09nA_F	GCCAGCAATAACGGTGGCAAGCAGGCGCTGG	3NN (1/7aa #9) F
R3_L01_09nA_R	CGAGATGGCCACCACCTGCTCCGGTAAG	3NN (1aa #9) R
Rep3_link07_09nA_R	GCTGCCACCGCCACCCGAGCTGATGGCCACCA	3NN (7aa #9) R
Rep3_link07_10nA_F	AGCAATAACGGTGGCAAGCAGGCGCTGGAGACG	3NN (1/7aa #10) F
R3_L01_10nA_R	CGAGGCGATGGCCACCACCTGCTC	3NN (1aa #10) R
Rep3_link07_10nA_R	GCTGCCACCGCCACCCGAGCTGGCGATGGCCA	3NN (7aa #10) R
Rep3_link07_11nA_F Rep3_link07_11nA_R	AGCTCGGGTGGCGGTGGCAGCAATAACGGTGGCA GCTGGCGATGGCCACCACCTGCTCCGGT	3NN (7aa #11)
R3_L01_13nA_F	TCGGGTGGCAAGCAGGCGCTGGAGACG	3NN (1aa #13) F
Rep3_link07_13nA_F	AGCTCGGGTGGCGGTGGCAGCGGTGGCAAGCA	3NN (7aa #13) F
Rep3_link07_13nA_R	GTTATTGCTGGCGATGGCCACCACCTGCTCC	3NN (1/7aa #13) R
R3_L01_15nA_F	TCGAAGCAGGCGCTGGAGACGGTG	3NN (1aa #15) F
Rep3_link07_15nA_F	AGCTCGGGTGGCGGTGGCAGCAAGCAGGCGCT	3NN (7aa #15) F
Rep3_link07_15nA_R	GCCACCGTTATTGCTGGCGATGGCCACCACC	3NN (1/7aa #15) R
R3_L01_17nA_F	TCGGCGCTGGAGACGGTGCAGCGG	3NN (1aa #17) F
Rep3_link07_17nA_F	AGCTCGGGTGGCGGTGGCAGCGCGCTGGAGAC	3NN (7aa #17) F
Rep3_link07_17nA_R	CTGCTTGCCACCGTTATTGCTGGCGATGGCCAC	3NN (1/7aa #17) R
R3_L01_18nA_F	TCGCTGGAGACGGTGCAGCGGCTGTTG	3NN (1aa #18) F
Rep3_link07_18nA_F	AGCTCGGGTGGCGGTGGCAGCCTGGAGACGGT	3NN (7aa #18) F
Rep3_link07_18nA_R	CGCCTGCTTGCCACCGTTATTGCTGGCGATGGCC	3NN (1/7aa #18) R
R3_L01_19nA_F	TCGGAGACGGTGCAGCGGCTGTTGCCG	3NN (1aa #19) F

Rep3_link07_19nA_F	AGCTCGGGTGGCGGTGGCAGCGAGACGGTGCA	3NN (7aa #19) F
Rep3_link07_19nA_R	CAGCGCCTGCTTGCCACCGTTATTGCTGGCGATG	3NN (1/7aa #19) F
Rep3_link07_20nA_F Rep3_link07_20nA_R	AGCTCGGGTGGCGGTGGCAGCACGGTGCAGC CTCCAGCGCCTGCTTGCCACCGTTATTGCTG	3NN (7aa #20)
Rep3_link07_22nA_F Rep3_link07_22nA_R	AGCTCGGGTGGCGGTGGCAGCCAGCGGCTGTTGCCG CACCGTCTCCAGCGCCTGCTTGCCACCG	3NN (7aa #22)
Rep3_link07_24nA_F Rep3_link07_24nA_R	AGCTCGGGTGGCGGTGGCAGCCTGTTGCCGGT CCGCTGCACCGTCTCCAGCGCCTGCTTGCCA	3NN (7aa #24)
R3_L01_25nA_F	TCGTTGCCGGTGCTGTGCCAGGCCCATGG	3NN (1aa #25) F
Rep3_link07_25nA_F	AGCTCGGGTGGCGGTGGCAGCTTGCCGGTGCT	3NN (7aa #25) F
Rep3_link07_25nA_R	CAGCCGCTGCACCGTCTCCAGCGCCTGCTTG	3NN (1/7aa #25) R
Rep3_link07_26nA_F Rep3_link07_26nA_R	AGCTCGGGTGGCGGTGGCAGCCCGGTGCTGTG CAACAGCCGCTGCACCGTCTCCAGCGCCTGC	3NN (7aa #26)
Rep3_link07_28nA_F Rep3_link07_28nA_R	AGCTCGGGTGGCGGTGGCAGCCTGTGCCAGGCC CACCGGCAACAGCCGCTGCACCGTCTCCAGCG	3NN (7aa #28)
Rep1_link07_5vE_F Rep1_link07_5vE_R	AGCTCGGGTGGCGGTGGCAGCTGCCAGGCCCATGG CAGCACCGGCAACAGCCGCTGCACCGTCTCCAGCGC	3NN (7aa #29)
Rep3_link07_31nA_F Rep3_link07_31nA_R	AGCTCGGGTGGCGGTGGCAGCGCCCATGGCCT CTGGCACAGCACCGGCAACAGCCGCTGCACC	3NN (7aa #31)
Rep2_linker_fwd Rep2_linker09_rev	TTTGAAGACTTCTGAGCTCGGGTGGCGGT TTTGAAGACTTCGGTGTCAGCGAGCTACCGCTACCGCCA	2 (7aa) (due to mutation)
Rep2_Linkneu_F Rep2_Linkneu18_R	TTTGAAGACTTCTGAGCTCGGGTGGCGGTAGCGGCGGTTCGGGCAGCGGC TTTGAAGACTTCGGTGTCAGCGAGCTACCGCTGCCACCGCCGCTGCCCGA	2 (18aa)
Rep2_linker_L1_F Rep2_linker_L1_R	TTTGAAGACTTCTGAGCTCGGGTGGCGGT TTTGAAGACTTGCTGCCACCGCCACCCGA	2.1 (L1)
Rep2_linker_R1_F Rep2_linker_R1_R	GAAGACTTGGTGGCTCGAGCGGTGGCAGC TTTGAAGACTTCGGTGTCAGGCTGCCACCGCTC	2.3 (R1)
Rep2_GFP (L1/R1)_F Rep2_GFP (L1/R1)_R	TTTGAAGACTTCAGCATGGTGAGCAAGGGC TTTGAAGACTTCACCCTTGTACAGCTCGTCC	2.2 (GFP)
Rep2_HA (L1/R1)_F Rep2_HA (L1/R1)_R	TTTGAAGACTTCAGCTATCCTTATGACGTGCCTGACTATGCCAG TTTGAAGACTTCACCAGGTCCTCCCAGGCTGGCATAGTCA	2.2 (HA)
oligonucleotides for the generation of reporter constructs with TALE target boxes		
Bs4Pr-GG_R	TTTGGTCTCACCTTAGATTCGATTAAAAATAAATTGTATG	universal R primer
op_F	TTTGGTCTCACACCTTAGCTGCAGTACTAGCATATTCTTTCTTGTATATAACT TTGTCC	box “op”
op -1 p4_F	TTTGGTCTCACACCTTAGTGCAGTACTAGCATAATTCTTTCTTGTATATAACTT TGTCC	box “op -1 p4”
op -1 p6_F	TTTGGTCTCACACCTTAGCTCAGTACTAGCATAATTCTTTCTTGTATATAACT TTGTCC	box “op -1 p6”
op -1 p7_F	TTTGGTCTCACACCTTAGCTGAGTACTAGCATAATTCTTTCTTGTATATAACT TTGTCC	box “op -1 p7”
op -1 p8_F	TTTGGTCTCACACCTTAGCTGCGTACTAGCATATTCTTTCTTGTATATAACTT TGTCC	box “op -1 p8”
op -1 p9_F	TTTGGTCTCACACCTTAGCTGCATACTAGCATATTCTTTCTTGTATATAACTT TGTCC	box “op -1 p9”
op -1 p12_F	TTTGGTCTCACACCTTAGCTGCAGTATAGCATATTCTTTCTTGTATATAACTT TGTCC	box “op -1 p12”
op -2 p4/8_F	TTTGGTCTCACACCTTAGTGCGTACTAGCATTATTCTTTCTTGTATATAACTT TGTCC	box “op -2 p4/8”
op -2 p4/12_F	TTTGGTCTCACACCTTAGTGCAGTATAGCATTATTCTTTCTTGTATATAACTT TGTCC	box “op -2 p4/12”
op -2 p6/7_F	TTTGGTCTCACACCTTAGCTAGTACTAGCATTAATTCTTTCTTGTATATAACT TTGTCC	box “op -2 p6/7”
op -2 p7/8_F	TTTGGTCTCACACCTTAGCTGGTACTAGCATTATTCTTTCTTGTATATAACTT TGTCC	box “op -2 p7/8”
op -2 p8/9_F	TTTGGTCTCACACCTTAGCTGCTACTAGCATTATTCTTTCTTGTATATAACTT TGTCC	box “op -2 p8/9”
op -2 p8/12_F	TTTGGTCTCACACCTTAGCTGCGTATAGCATTATTCTTTCTTGTATATAACTT TGTCC	box “op -2 p8/12”
op -3 p4/8/12_F	TTTGGTCTCACACCTTAGTGCGTATAGCATGTATTCTTTCTTGTATATAACTT TGTCC	box “op -3 p4/8/12”
op -3 p7/8/9_F	TTTGGTCTCACACCTTAGCTGTACTAGCATGTATTCTTTCTTGTATATAACTT TGTCC	box “op -3 p7/8/9”
MidG6_F	TTTGGTCTCACACCTTAGCTGGGGGGGTAGCATTTCTTTCTTGTATATAACT TTGTCC	box “midG6 op”
MidG6(-1)F	TTTGGTCTCACACCTTAGCTGGGGGGTAGCATATTCTTTCTTGTATATAACT TTGTCC	box “midG6 -1”
MidG6(-2)_F	TTTGGTCTCACACCTTAGCTGGGGGTAGCATTATTCTTTCTTGTATATAACTT TGTCC	box “midG6 -2”
MidG6(-3)_F	TTTGGTCTCACACCTTAGCTGGGGTAGCATGTATTCTTTCTTGTATATAACTT TGTCC	box “midG6 -3”
MidG6(-4)_F	TTTGGTCTCACACCTTAGCTGGGTAGCATCGTATTCTTTCTTGTATATAACTT TGTCC	box “midG6 -4”
MidG6(-5)_F	TTTGGTCTCACACCTTAGCTGGTAGCATCCGTATTCTTTCTTGTATATAACTT TGTCC	box “midG6 -5”

MidG6 (-6)_F	TTTGGTCTCACACCTTAGCTGTAGCATCGCGTATTCTTTCTTGTATATAACTTTGTCC	box “MidG6 -6”
BK op_F	TTTGGTCTCACACCTATAGAGCACCAACAACTCCCTTTTCTTTCTTGTATATAACTTTGTCC	box “BK op”
BK -1p9_F	TTTGGTCTCACACCTATAGAGCACAACAACTCCCTTTTCTTTCTTGTATATAACTTTGTCC	box “BK -1 p9”
BK -1p12_F	TTTGGTCTCACACCTATAGAGCACCACAACTCCCTTTTCTTTCTTGTATATAACTTTGTCC	box “BK -1 p12”
BK -2p9/12_F	TTTGGTCTCACACCTATAGAGCACACAACTCCCTTTTCTTTCTTGTATATAACTTTGTCC	box “BK -2 p9/12”
BK -4p9-12_F	TTTGGTCTCACACCTATAGAGCACAACTCCCTTTTCTTTCTTGTATATAACTTTGTCC	box “BK -4 p9-12”
misM 1_F	TTTGGTCTCACACCTTACGAGCAGTACTAGCATTTCTTTCTTGTATATAACTTTGTCC	box “mm #1”
misM 2_F	TTTGGTCTCACACCTTAGCTGCAGATGTAGCATTTCTTTCTTGTATATAACTTTGTCC	box “mm #2”
misM 3_F	TTTGGTCTCACACCTTAGCTGCAGTACTTCGATTTCTTTCTTGTATATAACTTTGTCC	box “mm #3”
oligonucleotides for the amplification of 1 kbp promoter fragments (sorted by locus ID)		
SB_Pr44_F SB_Pr44_R	TTTGGTCTCACACCCAACTATGGGCCAGAATTC TTTGGTCTCACCTTCGCACTCCACCACACAAAA	*Os01g04190*
SB_Pr191_F SB_Pr191_R	TTTGGTCTCACACCGCGCGCCGTCGTGGAAAG TTTGGTCTCACCTTGGCCGGAATTCGCGACTCCG	*Os01g09470*
SB_Pr27_fw SB_Pr27_rv	TTTGGTCTCACACCAATCCCCTCTTTCTTTCATCTC TTTGGTCTCACCTTGCCGCCGCCGCCTT	*Os01g19870*
SB_Pr166_F SB_Pr166_R	TTTGGTCTCACACCAATCGCCAGAAGATAAGGA TTTGGTCTCACCTTACTCTGTTTAGTAAATACA	*Os01g24710*
SB_Pr113_F SB_Pr113_R	TTTGGTCTCACACCAAATGATATAGGAAATGTG TTTGGTCTCACCTTCAGCTTGGCGCCGCCTTT	*Os01g32439*
SB_Pr122_F SB_Pr122_R	TTTGGTCTCACACCTGTCTTTTACTGCACAAT TTTGGTCTCACCTTGGTGAGATGCGAGTACGTA	*Os01g40290*
(Os01g51040)_F_SB (Os01g51040)_R_SB	TTTGGTCTCACACCTGGTCGGGTTCAGCGG / TTTGGTCTCACCTTCCTCCACAACGACGACGG	*Os01g51040*
SB_Pr152_F SB_Pr152_R	TTTGGTCTCACACCTTTTGGACGAATGGAGTAG TTTGGTCTCACCTTGGTGTGTTCTGCCCCCGC	*Os01g53920*
SB_Pr45_F SB_Pr45_R	TTTGGTCTCACACCATCGTAGCTCAACCGATCA TTTGGTCTCACCTTGGCGCAGTAGAAATGAAATGC	*Os01g55600*
SB_Pr106_F SB_Pr106_R	TTTGGTCTCACACCAACGAACAAGTTCTCGATC TTTGGTCTCACCTTGTCGGGCTGGCGCCATTGC	*Os01g58330*
SB_Pr164_F SB_Pr164_R	TTTGGTCTCACACCTATTGTTGGATTTTGTGAG TTTGGTCTCACCTTTTCGCAGCAGCAGCCTCG	*Os01g65590*
SB_Pr38_F SB_Pr38_R	TTTGGTCTCACACCCTGTCAGTGTCAGGGTTGGC TTTGGTCTCACCTTGCCGCCGGTGACGATGATC	*Os02g02190*
SB_Pr18_fw SB_Pr18_rv	TTTGGTCTCACACCAACTCTCAAGCCCCTCATTAACC TTTGGTCTCACCTTGGCCGGAGCAGCCGAA	*Os02g02620*
SB_Pr105_F SB_Pr105_R	TTTGGTCTCACACCTGGTGGAAGCAACCCCCCC TTTGGTCTCACCTTCCCCCTCCCAGATCTACCCT	*Os02g14720*
SB_Pr162_F SB_Pr162_R	TTTGGTCTCACACCGCCCATGTTTTTGGAAGT TTTGGTCTCACCTTATTTCCTTCTGTAGAGGTG	*Os02g29230*
SB_Pr59_F SB_Pr59_R	TTTGGTCTCACACCATACAACAGGAGGACAAC TTTGGTCTCACCTTGCTACAACTACAAGTGCAAC	*Os02g34970*
SB_Pr126_F SB_Pr126_R	TTTGGTCTCACACCGATATATTTAGGGAGCTTC TTTGGTCTCACCTTGGCGAGCGAAGTGGAGGG	*Os02g36850*
SB_Pr168_F SB_Pr168_R	TTTGGTCTCACACCTCATGCTTTCTCTCTCCTC TTTGGTCTCACCTTCCTTGCTGTTGCTGCAACT	*Os02g38386*
SB_Pr53_F SB_Pr53_R	TTTGGTCTCACACCTGAGTACAAGTAGGTAAGAAG TTTGGTCTCACCTTCTTCTCGGCACGCGGC	*Os02g43760*
SB_Pr56_F SB_Pr56_R	TTTGGTCTCACACCGATGTTCACGTGCACTGC TTTGGTCTCACCTTATTACTAACCACTTAGCAAG	*Os02g47660*
SB_Pr51_F SB_Pr51_R	TTTGGTCTCACACCATATTTATAGCGTAAAAATTTTA TTTGGTCTCACCTTCATTGCCTTGCTCTATGAAT	*Os02g48570*
SB_Pr175_F SB_Pr175_R	TTTGGTCTCACACCGCCCAATGCAGTTGACAA TTTGGTCTCACCTTGCATGCAGCTTCTTCCAA	*Os02g49710*
SB_Pr102_F SB_Pr102_R	TTTGGTCTCACACCCATCAAGCAGCAGGAGTT TTTGGTCTCACCTTGATGGTATGGTGGTTAAT	*Os02g57860*
SB_Pr190_F SB_Pr190_R	TTTGGTCTCACACCTCCCTTAAATAAAAGGAC TTTGGTCTCACCTTTGCGAAACTGTGATGAGA	*Os03g03350*
SB_Pr57_F SB_Pr57_R	TTTGGTCTCACACCTTTCATAGCCTCTCTGTAG TTTGGTCTCACCTTGGTCGCCGCCTGCAATAT	*Os03g07540*
SB_Pr92_F SB_Pr92_R	TTTGGTCTCACACCTGATCGAGAAAGAAAACT TTTGGTCTCACCTTGGCTATCTGCTGCCCGT	*Os03g11050*
SB_Pr182_F SB_Pr182_R	TTTGGTCTCACACCGGACCTTTACAGAATTTGCATGC TTTGGTCTCACCTTCTCCGTGGAGGACGACGAG	*Os03g11980*
SB_Pr54_F SB_Pr54_R	TTTGGTCTCACACCTTACAAATTGAAAAAAAGAAGAG TTTGGTCTCACCTTCGATCGATCTCCAAAAC	*Os03g37840*

SB_Pr196_F SB_Pr196_R	TTTGGTCTCACACCCCATAGCTATCAAGTTGT TTTGGTCTCACCTTGCAGCCATTAAATTTCACT	*Os03g40100*
SB_Pr147_F SB_Pr147_R	TTTGGTCTCACACCTTTGATATTAAGTTAGTTTC TTTGGTCTCACCTTGGTCGATCAGTAGCTGGT	*Os03g42130*
SB_Pr144_F SB_Pr144_R	TTTGGTCTCACACCGTGTCCATAGCATTAAATA TTTGGTCTCACCTTGTTAACAATTTTAGGTCGA	*Os03g59790*
SB_Pr50_F SB_Pr50_R	TTTGGTCTCACACCCCAATCAACGCCCCAGCCGT TTTGGTCTCACCTTTGTGAGTGCGAGTGGCAACGGC	*Os03g60720*
SB_Pr150_F SB_Pr150_R	TTTGGTCTCACACCAAGCATCTAACTGTACTG TTTGGTCTCACCTTGGCCGGCGTCGACTCGAG	*Os03g62010*
SB_Pr195_F SB_Pr195_R	TTTGGTCTCACACCCAGCAATTAAGCTCTGAAAATTCC TTTGGTCTCACCTTGGCAGATCGGGCCGGAC	*Os04g18830*
SB_Pr160_F SB_Pr160_R	TTTGGTCTCACACCCAGTGACCTTGCAGGCTC TTTGGTCTCACCTTGGCCGGCCGATCGATTCTT	*Os04g27060*
SB_Pr165_F SB_Pr165_R	TTTGGTCTCACACCACAAATGCATTTTTTTAGC TTTGGTCTCACCTTTGTGCTAGCGCGCGCCCG	*Os04g45730*
(Os04g47890)_F_SB (Os04g47890)_R_SB	TTTGGTCTCACACCTGCAGGTAGACTGGGAAA TTTGGTCTCACCTTCACCAACAACGGTTTGCTTC	*Os04g47890*
SB_Pr155_F SB_Pr155_R	TTTGGTCTCACACCGTATTTAAACATATTCCTT TTTGGTCTCACCTTCTCCTAGGCCTTCAGCTC	*Os04g48160*
SB_Pr58_F SB_Pr58_R	TTTGGTCTCACACCGACTTTATGAATCTTAATATATG TTTGGTCTCACCTTGGGAGACGAAGCGCGAG	*Os04g49194*
SB_Pr97_F SB_Pr97_R	TTTGGTCTCACACCCATTATTTGTAGATGTCAC TTTGGTCTCACCTTTTCTCTGCTCAAGTCATC	*Os04g53210*
(Os04g58560)_F_SB (Os04g58560)_R_SB	TTTGGTCTCACACCTAAATTCTCTTGGTACGC TTTGGTCTCACCTTATTCCATCCGCAATTGG	*Os04g58560*
SB_Pr60_F SB_Pr60_R	TTTGGTCTCACACCCTCCACCAGTGATCCTAT TTTGGTCTCACCTTCCGATCGGTTAGTTCTTGG	*Os05g27590*
SB_Pr177_F SB_Pr177_R	TTTGGTCTCACACCGGTCGCCATGAGCTACGA TTTGGTCTCACCTTGTGGGGCCCACGCGCCA	*Os05g28800*
SB_Pr23_fw SB_Pr23_rv	TTTGGTCTCACACCAACTCAATAGTAGCGGTGGCCG TTTGGTCTCACCTTGCCGGCGCCCGCAGA	*Os05g33390*
SB_Pr197_F SB_Pr197_R	TTTGGTCTCACACCTAGCAACCAAATATATATCT TTTGGTCTCACCTTCTTGGTAACCACCATGGA	*Os05g38320*
SB_Pr48_F SB_Pr48_R	TTTGGTCTCACACCCTGGCCTTAGTGTTAATCAACC TTTGGTCTCACCTTGGCACGCGTGTCGTGTAGTA	*Os05g42150*
SB_Pr179_F SB_Pr179_R	TTTGGTCTCACACCGGACACGTTAGTATGTACTC TTTGGTCTCACCTTGGAGAATTATTGGAGATGGAC	*Os05g51690*
SB_Pr133_F SB_Pr133_R	TTTGGTCTCACACCCGCCATGGAGTGCTCACTG TTTGGTCTCACCTTCTCCCCCTCCAAAAACCTA	*Os06g02130*
SB_Pr40_F SB_Pr40_R	TTTGGTCTCACACCACGATTCAATGGCGGGAA TTTGGTCTCACCTTCTTACCAGTAAGTAGCTCTAG	*Os06g07260*
SB_Pr135_F SB_Pr135_R	TTTGGTCTCACACCCCACCTCCAGGTAGTTGAT TTTGGTCTCACCTTAGTGGATAATTCAGTCTAAG	*Os06g10670*
SB_Pr161_F SB_Pr161_R	TTTGGTCTCACACCTTGGGAGTCCATAGCAACA TTTGGTCTCACCTTCTTAATTTGATCTGTGTCTC	*Os06g12410*
SB_Pr107_F3 SB_Pr107_R2	TTTGGTCTCACACCGGGTCCACACCTTCTCTCCTC TTTGGTCTCACCTTGTGCTTCCGCCGCCGCCGCCGC	*Os06g12530*
SB_Pr114_F2 SB_Pr114_R	TTTGGTCTCACACCCCCAAGACAAGAAGTCTAGCAGC TTTGGTCTCACCTTGGCGGGCGCGGTTGGTGG	*Os06g17020*
SB_Pr125_F SB_Pr125_R	TTTGGTCTCACACCTTTTATATCGTTGCTGAC TTTGGTCTCACCTTTGCTGCAAGGCTTGGGTT	*Os06g17910*
SB_Pr154_F SB_Pr154_R	TTTGGTCTCACACCTAACTCACGTGCTGTGCA TTTGGTCTCACCTTTATTCTTTTCGTCGCGGAG	*Os06g36270*
SB_Pr55_F SB_Pr55_R	TTTGGTCTCACACCCAGCTCAGCCTAAAATGA TTTGGTCTCACCTTGGCTGAAAAGTGAAAAGGC	*Os06g37080*
SB_Pr189_F SB_Pr189_R	TTTGGTCTCACACCTGTGAGCTCTTGTTATAAAG TTTGGTCTCACCTTCGTCGCGGCGGAGGTCGT	*Os06g43650*
SB_Pr128_F SB_Pr128_R	TTTGGTCTCACACCTTTTTGAAATGTAGGATGC TTTGGTCTCACCTTAGTAGTCAAATCAGATCTC	*Os07g01490*
(Os07g05720)_F_SB (Os07g05720)_R_SB	TTTGGTCTCACACCAAAATATCTCAAAATTAAATTTAAAAAC TTTGGTCTCACCTTCCCGACGATGACG	*Os07g05720*
SB_Pr146_F SB_Pr146_R	TTTGGTCTCACACCGAATACTTTGTATGTGGC TTTGGTCTCACCTTGGCCGGGTGCTCTTCTTCT	*Os07g09470*
SB_Pr24_fw SB_Pr24_rv	TTTGGTCTCACACCCCCCACGAATTCGAAAGAAAG TTTGGTCTCACCTTCGTCGGAGTCCGGCGA	*Os07g17280*
SB_Pr142_F SB_Pr142_R	TTTGGTCTCACACCCATATTATTGTGAAGATTT TTTGGTCTCACCTTGGCCACCCTCTGCACCTC	*Os07g20290*
SB_Pr169_F SB_Pr169_R	TTTGGTCTCACACCTAGTCAACTACGGGCACCA TTTGGTCTCACCTTGGTAGCAAGAACAGAAGCT	*Os07g23470*
SB_Pr84_F SB_Pr84_R	TTTGGTCTCACACCAGTTAAATTTGAGTTGTTTTG TTTGGTCTCACCTTCTCCTCCGCTGCTGCCCC	*Os07g26520*
SB_Pr187_F SB_Pr187_R	TTTGGTCTCACACCGATGACCGTGTACGTCCT TTTGGTCTCACCTTGGTTGGACAACAAGCTGGG	*Os07g28990*
SB_Pr19_fw	TTTGGTCTCACACCAACAATGGGGCTCC	*Os07g29310*

SB_Pr19_rv	TTTGGTCTCACCTTGATCGATCGATATATATATATAG	
SB_Pr149_F SB_Pr149_R	TTTGGTCTCACACCCGGCACTTATTTCAATACAAG TTTGGTCTCACCTTTCCTCGAGCTCTCGGCGGT	*Os07g33740*
SB_Pr85_F SB_Pr85_R	TTTGGTCTCACACCTTCCTTTTGCGTCTACTTT TTTGGTCTCACCTTCTTCAGAGAGAGAGAAAG	*Os07g36430*
SB_Pr43_F SB_Pr43_R	TTTGGTCTCACACCCTATCCAAAAATCAATAGGTA TTTGGTCTCACCTTCTCGCCCCTCCTCCTCCTCCTGA	*Os07g39280*
SB_Pr173_F SB_Pr173_R	TTTGGTCTCACACCTCCTCCCCCCACTCCCAC TTTGGTCTCACCTTGGCAGGCAAGGGCTAGGAG	*Os07g39590*
SB_Pr121_F SB_Pr121_R	TTTGGTCTCACACCAATAACAATAAGAAGATATT TTTGGTCTCACCTTTTCAGTGTCCAGAGGACA	*Os07g44890*
SB_Pr20_fw SB_Pr20_rv	TTTGGTCTCACACCATCGTACGTCAAGCATGTCC TTTGGTCTCACCTTCTTTGCTGCTGCTGCGCG	*Os07g47790*
SB_Pr93_F SB_Pr93_R	TTTGGTCTCACACCATGAGCTTAATTAACTTAATTTAA TTTGGTCTCACCTTCGGAATCCCTCAGGTCG	*Os08g01260*
SB_Pr181_F SB_Pr181_R	TTTGGTCTCACACCTTTTAAGGGATGTATTTGTA TTTGGTCTCACCTTGGTTAGAGTAAGGAGGTT	*Os08g14940*
SB_Pr143_F SB_Pr143_R	TTTGGTCTCACACCCGAGGGCCCCTTGGCCTC TTTGGTCTCACCTTCACAGAAGCATATCACCTC	*Os08g33120*
SB_Pr183_F SB_Pr183_R	TTTGGTCTCACACCATGCTCTCTGATCGCTGT TTTGGTCTCACCTTGCCCAGCTAATTAACCAC	*Os08g37432*
SB_Pr153_F SB_Pr153_R	TTTGGTCTCACACCAATTGCCCAAAATAGATGTTTGG TTTGGTCTCACCTTCCCGATCGATCTCCAGCT	*Os08g41950*
SB_Pr100_F SB_Pr100_R	TTTGGTCTCACACCTAGAAGAATCCCTGAGTG TTTGGTCTCACCTTGGCGTCAACGTCAACGC	*Os09g25200*
SB_Pr101_F SB_Pr101_R	TTTGGTCTCACACCCGCGGAGCGGCAGCCGC TTTGGTCTCACCTTTCCTCCGGCGGTTTCGA	*Os09g27090*
SB_Pr36_F SB_Pr36_R	TTTGGTCTCACACCCCTAAACACAAAATCAAGGGAA TTTGGTCTCACCTTTTCCTTGCTGATAAATCTTGC	*Os09g29100*
SB_Pr99_F SB_Pr99_R	TTTGGTCTCACACCGACATTACTACGTACAGC TTTGGTCTCACCTTGGTCGCCGTCGCCGGCC	*Os09g29130*
SB_Pr138_F SB_Pr138_R	TTTGGTCTCACACCGTGGACGAGTTCACGAGTA TTTGGTCTCACCTTCGCCGGCCGATCGATCAGC	*Os09g30120*
SB_Pr89_F SB_Pr89_R	TTTGGTCTCACACCGAAACTGAGGAAGCATTT TTTGGTCTCACCTTGGCGGATCGGATCGGACA	*Os09g32100*
SB_Pr41_F SB_Pr41_R	TTTGGTCTCACACCCATGAGTGCATCTTGCATCT TTTGGTCTCACCTTGGCTGGTTCTGGTTGTTCTT	*Os09g37350*
SB_Pr42_F SB_Pr42_R	TTTGGTCTCACACCAGTAACTCTCCAAATTTTATTTG TTTGGTCTCACCTTGGCTGGTTCTTGGTCTTGG	*Os09g37470*
SB_Pr137_F SB_Pr137_R	TTTGGTCTCACACCCATCATACTCACACACTCA TTTGGTCTCACCTTTGCTGAATGAATCCTAGCA	*Os10g05910*
SB_Pr96_F SB_Pr96_R	TTTGGTCTCACACCCCGTATAAGAAATATTTTG TTTGGTCTCACCTTCGGCGAGCAAGAATAAT	*Os10g13820*
SB_Pr22_fw SB_Pr22_rv	TTTGGTCTCACACCGATCTACTGTTTCCTTTTT TTTGGTCTCACCTTGATTGCCAAGGTGTGGA	*Os10g25550*
SB_Pr172_F SB_Pr172_R	TTTGGTCTCACACCATCGGCATCTATAAGCCG TTTGGTCTCACCTTTTTGAGCTTATTCCGGTTG	*Os10g31380*
SB_Pr178_F SB_Pr178_R	TTTGGTCTCACACCCAAATCTGTGCTGCGTGG TTTGGTCTCACCTTTGTGGACTTACTGGCTAGC	*Os10g31640*
SB_Pr49_F SB_Pr49_R	TTTGGTCTCACACCCAAGAATGTGCATGTCTTAAC TTTGGTCTCACCTTTTTGTCAGTTATAGTGATGTG	*Os10g39840*
SB_Pr108_F SB_Pr108_R	TTTGGTCTCACACCATTTCGCCTGTAAATTGCG TTTGGTCTCACCTTGTATGGTCATGTTTTGTATA	*Os11g02440*
SB_Pr37_F SB_Pr37_R	TTTGGTCTCACACCATTCAATCCACACGATTAAC TTTGGTCTCACCTTCTCGATCGATCCAGCAAT	*Os11g04130*
SB_Pr132_F SB_Pr132_R	TTTGGTCTCACACCAGATGGGGCTTACCGCCCG TTTGGTCTCACCTTCACCCTCCCTTGCAAAACA	*Os11g13440*
SB_Pr157_F SB_Pr157_R	TTTGGTCTCACACCGTGCATATATACCAACTAA TTTGGTCTCACCTTCTCTCTCGACTAATTCACC	*Os11g32650*
SB_Pr127_F SB_Pr127_R	TTTGGTCTCACACCGAACTACCCCGTGACGCT TTTGGTCTCACCTTGACATATTCGATTCCTCCAT	*Os11g39480*
SB_Pr129_F SB_Pr129_R	TTTGGTCTCACACCGGATAGCAACAACCTAACC TTTGGTCTCACCTTTGTTACGATCTTGTCGTTA	*Os11g44170*
SB_Pr26_fw SB_Pr26_rv	TTTGGTCTCACACCTTTGTATATTAAGTTTTAGTATTC TTTGGTCTCACCTTCTTGTGCACTCTTTCT	*Os11g46070*
SB_Pr115_F SB_Pr115_R	TTTGGTCTCACACCCTATCCAAAAATCAATAG TTTGGTCTCACCTTCTCGCCCCTCCTCCTCCT	*Os11g47350*
SB_Pr180_F SB_Pr180_R	TTTGGTCTCACACCATCAAAGCCATGCATCTCGAG TTTGGTCTCACCTTCCTATGGTCGTTCTCGTGTAC	*Os12g10080*
SB_Pr192_F SB_Pr192_R	TTTGGTCTCACACCATGTGAATGTTTGTAGAACG TTTGGTCTCACCTTGTCTCCCATGAATGATCA	*Os12g13810*
SB_Pr130_F SB_Pr130_R	TTTGGTCTCACACCTGATTAGTAGTAGAGATGT TTTGGTCTCACCTTTGCTCCACAATATTTGGTT	*Os12g17340*
SB_Pr103_F SB_Pr103_R	TTTGGTCTCACACCTAAAGTACTATGAATGAT TTTGGTCTCACCTTGGCCGCCCGGCGATTGG	*Os12g24320*

SB_Pr136_F SB_Pr136_R	TTTGGTCTCACACCAAATGATTGTATTAGAAAAC TTTGGTCTCACCTTGGCTGGGTGCTTTGTGGTC	*Os12g42040*
SB_Pr46_F SB_Pr46_R	TTTGGTCTCACACCATCTCAGATTCTCTCTGCTTG TTTGGTCTCACCTTCCTCTACTCCTCCTTACTG	*Os12g42200*
SB_Pr25_fw SB_Pr25_rv	TTTGGTCTCACACCGATAATGTATAATTTTATTTGTCTAACAAG TTTGGTCTCACCTTGCCCAGCGACGAGGA	*Os12g42970*
OsSWEET13_F OsSWEET13_R	TTTGGTCTCACACCAGGGATGTCTACTGCAGGTG TTTGGTCTCACACCCATTTTTGTGTGCTAAAAGGGGGGTA	1 kbp *OsSWEET13* promoter fragment
OsSWEET14_F OsSWEET14_R	TTTGGTCTCACACCGTGTGTGCCACTCCAACTGATA TTTGGTCTCACCTTTGCAGCAAGATCTTGATTAACTAGCT	1 kbp *OsSWEET14* promoter fragment
OsSWEET14_Res_F OsSWEET14_Res_R	ATCAAGCCTTCAAGCAAAGCAAAC GAGGGGGTTTATATAGTGCTAGG	1 kbp *OsSWEET14 xa41(t)* mimic

2.1.6 Plasmids

Table 2.5: List of used plasmids.

plasmid name	description	resistance	reference
general plasmids			
pGWB3GG	acceptor vector; for GUS reporter constructs	Kan	Annekatrin Richter
pUC57	acceptor vector; for blunt end cloning	Amp	Geißler *et al*., 2011
pSKA2	acceptor vector; for TALE expression constructs (*A. tumefaciens*)	Spec	Sabine Thieme
pSKX1	acceptor vector; for TALE expression constructs (*X. oryzae*)	Gent	Sabine Thieme
pNTM(Xoo)AG4	GGC module; N-terminal region from TalAG4 (of *Xoo* strain PXO83) in pUC57	Amp	Mücke *et al*., 2019
pCTM(Xoo)AO3	GGC module; C-terminal region from TalAO3 (of *Xoo* strain PXO83) in pUC57	Amp	Mücke *et al*., 2019
pNTH3	GGC module; N-terminal region from Hax3 in pJet1.2	Amp	Kay *et al*., 2005
pCTH3	GGC module; C-terminal region from Hax3 in pUC57	Amp	Kay *et al*., 2005
novel single repeat modules for natural aberrant repeats (sorted by aberrant repeat type, RVD and position)			
pSR2HD(28)	2HD (28aa) (single repeat module for position 2; RVD HD; 28aa repeat)	Amp	this study
pSR3HD(28)	3HD (28aa)	Amp	this study
pSR1NG(28)	1NG (28aa)	Amp	this study
pSR2NG(28)	2NG (28aa)	Amp	this study
pSR3NG(28)	3NG (28aa)	Amp	this study
pSR2NI(28)	2NI (28aa)	Amp	this study
pSR3NI(28)	3NI (28aa)	Amp	this study
pSR1NN(28)	1NN (28aa)	Amp	this study
pSR2NN(28)	2NN (28aa)	Amp	this study
pSR3NN(28)	3NN (28aa)	Amp	this study
pSR1HD(30)	1HD (30aa)	Amp	this study
pSR2HD(30)	2HD (30aa)	Amp	this study
pSR3HD(30)	3HD (30aa)	Amp	this study
pSR4HD(30)	4HD (30aa)	Amp	this study
pSR1NG(30)	1NG (30aa)	Amp	this study
pSR2NG(30)	2NG (30aa)	Amp	this study
pSR3NG(30)	3NG (30aa)	Amp	this study
pSR4NG(30)	4NG (30aa)	Amp	this study
pSR1NI(30)	1NI (30aa)	Amp	this study
pSR2NI(30)	2NI (30aa)	Amp	Richter *et al*., 2014
pSR3NI(30)	3NI (30aa)	Amp	this study
pSR4NI(30)	4NI (30aa)	Amp	this study
pSR1NN(30)	1NN (30aa)	Amp	this study
pSR2NN(30)	2NN (30aa)	Amp	this study
pSR3NN(30)	3NN (30aa)	Amp	this study
pSR4NN(30)	4NN (30aa)	Amp	this study
pSR2HD(35v1)	2HD (35aa v1)	Amp	this study
pSR2NN(35v1)	2NN (35aa v1)	Amp	this study
pSR2HD(35v2)	2HD (35aa v2)	Amp	this study
pSR2NN(35v2)	2NN (35aa v2)	Amp	this study
pSR2HD(36v1)	2HD (36aa v1)	Amp	this study
pRHD3(36)	3HD (36aa v1)	Amp	Stefanie Mücke
pSR1NI(36v1)	1NI (36aa v1)	Amp	this study
pSR2NI(36v1)	2NI (36aa v1)	Amp	this study
pSR3NI(36v1)	3NI (36aa v1)	Amp	this study
pSR2NN(36v1)	2NN (36aa v1)	Amp	this study
pSR3NN(36v1)	3NN (36aa v1)	Amp	this study
pRNS1(36)	1NS (36aa v1)	Amp	Stefanie Mücke
pSR2NS(36v1)	2NS (36aa v1)	Amp	this study
pSR2HD(36v2)	2HD (36aa v2)	Amp	this study
pSR2NI(36v2)	2NI (36aa v2)	Amp	this study
pSR2NN(36v2)	2NN (36aa v2)	Amp	this study
pSR1HD(37v1)	1HD (37aa v1)	Amp	this study
pSR2HD(37v1)	2HD (37aa v1)	Amp	this study
pSR3HD(37v1)	3HD (37aa v1)	Amp	this study

pSR4HD(37v1)	4HD (37aa v1)	Amp	this study
pSR1NN(37v1)	1NN (37aa v1)	Amp	this study
pSR2NN(37v1)	2NN (37aa v1)	Amp	this study
pSR3NN(37v1)	3NN (37aa v1)	Amp	this study
pSR4NN(37v1)	4NN (37aa v1)	Amp	this study
pSR1HD(37v2)	1HD (37aa v2)	Amp	this study
pSR2HD(37v2)	2HD (37aa v2)	Amp	this study
pSR3HD(37v2)	3HD (37aa v2)	Amp	this study
pSR4HD(37v2)	4HD (37aa v2)	Amp	this study
pSR1NN(37v2)	1NN (37aa v2)	Amp	this study
pSR2NN(37v2)	2NN (37aa v2)	Amp	this study
pSR4NN(37v2)	4NN (37aa v2)	Amp	this study
pSR1HD(40)	1HD (40aa)	Amp	this study
pSR2HD(40)	2HD (40aa)	Amp	this study
pSR3HD(40)	3HD (40aa)	Amp	this study
pSR4HD(40)	4HD (40aa)	Amp	this study
pSR5HD(40)	5HD (40aa)	Amp	this study
pSR1NG(40)	1NG (40aa)	Amp	this study
pSR2NG(40)	2NG (40aa)	Amp	this study
pSR3NG(40)	3NG (40aa)	Amp	this study
pSR4NG(40)	4NG (40aa)	Amp	this study
pSR5NG(40)	5NG (40aa)	Amp	this study
pSR1NI(40)	1NI (40aa)	Amp	this study
pSR2NI(40)	2NI (40aa)	Amp	this study
pSR3NI(40)	3NI (40aa)	Amp	this study
pSR4NI(40)	4NI (40aa)	Amp	this study
pSR5NI(40)	5NI (40aa)	Amp	this study
pSR1NN(40)	1NN (40aa)	Amp	this study
pSR2NN(40)	2NN (40aa)	Amp	Richter *et al*., 2014
pSR3NN(40)	3NN (40aa)	Amp	Richter *et al*., 2014
pSR4NN(40)	4NN (40aa)	Amp	this study
pSR5NN(40)	5NN (40aa)	Amp	this study
pAV_L-A(40)	combined with normal repeat at last position = 40aa repeat at last position in L-A	Kan	this study
pAV_A-B(40)	combined with normal repeat at last position = 40aa repeat at last position in A-B	Kan	this study
pAV_B-C(40)	combined with normal repeat at last position = 40aa repeat at last position in B-C	Kan	this study
pSR1HD(42)	1HD (42aa)	Amp	this study
pSR2HD(42)	2HD (42aa)	Amp	this study
pSR3HD(42)	3HD (42aa)	Amp	this study
pSR4HD(42)	4HD (42aa)	Amp	this study
pSR5HD(42)	5HD (42aa)	Amp	this study
pSR6HD(42)	6HD (42aa)	Amp	this study
pSR1NG(42)	1NG (42aa)	Amp	this study
pSR2NG(42)	2NG (42aa)	Amp	this study
pSR3NG(42)	3NG (42aa)	Amp	this study
pSR4NG(42)	4NG (42aa)	Amp	this study
pSR6NG(42)	6NG (42aa)	Amp	this study
pSR1NI(42)	1NI (42aa)	Amp	this study
pSR2NI(42)	2NI (42aa)	Amp	Richter *et al*., 2014
pSR3NI(42)	3NI (42aa)	Amp	this study
pSR4NI(42)	4NI (42aa)	Amp	this study
pSR5NI(42)	5NI (42aa)	Amp	this study
pSR6NI(42)	6NI (42aa)	Amp	this study
pSR1sNI(42)	1NI (42aa) stop	Amp	this study
pSR2sNI(42)	2NI (42aa) stop	Amp	this study
pSR3sNI(42)	3NI (42aa) stop	Amp	this study
pSR4sNI(42)	4NI (42aa) stop	Amp	this study
pSR5sNI(42)	5NI (42aa) stop	Amp	this study
pSR1NN(42)	1NN (42aa)	Amp	this study
pSR2NN(42)	2NN (42aa)	Amp	this study
pSR3NN(42)	3NN (42aa)	Amp	this study
pSR4NN(42)	4NN (42aa)	Amp	this study
pSR5NN(42)	5NN (42aa)	Amp	this study
pSR6NN(42)	6NN (42aa)	Amp	this study
novel single repeat modules for artificial aberrant repeats (sorted by aberrant repeat type and RVD)			
pSR2link7aa	2link7aa (7aa SG linker as substitute for a single repeat at position 2)	Amp	this study
pSR2link18aa	2link18aa (18aa SG linker as substitute for a single repeat at position 2)	Amp	this study
pSR3NN(36art1)	3NN (36aa art1) (single repeat module for position 3; RVD NN; 36aa art1 repeat)	Amp	this study
pSR3NN(36art2)	3NN (36aa art2)	Amp	this study
pSR3NN(36art3)	3NN (36aa art3)	Amp	this study
pSR3NN(36art4)	3NN (36aa art4)	Amp	this study
pSR3NN(7ins4)	3NN (7aa insertion #4)	Amp	this study
pSR3NN(7ins5)	3NN (7aa insertion #5)	Amp	this study
pSR3NN(7ins9)	3NN (7aa insertion #9)	Amp	this study

pSR3NN(7ins10)	3NN (7aa insertion #10)	Amp	this study
pSR3NN(7ins11)	3NN (7aa insertion #11)	Amp	this study
pSR3NN(7ins13)	3NN (7aa insertion #13)	Amp	this study
pSR3NN(7ins15)	3NN (7aa insertion #15)	Amp	this study
pSR3NN(7ins17)	3NN (7aa insertion #17)	Amp	this study
pSR3NN(7ins18)	3NN (7aa insertion #18)	Amp	this study
pSR3NN(7ins18)	3NN (7aa insertion #19)	Amp	this study
pSR3NN(7ins20)	3NN (7aa insertion #20)	Amp	this study
pSR3NN(7ins22)	3NN (7aa insertion #22)	Amp	this study
pSR3NN(7ins24)	3NN (7aa insertion #24)	Amp	this study
pSR3NN(7ins25)	3NN (7aa insertion #25)	Amp	this study
pSR3NN(7ins26)	3NN (7aa insertion #26)	Amp	this study
pSR3NN(7ins28)	3NN (7aa insertion #28)	Amp	this study
pSR3NN(7ins29)	3NN (7aa insertion #29)	Amp	this study
pSR3NN(7ins31)	3NN (7aa insertion #31)	Amp	this study
pSR3NN(1ins5)	3NN (1aa insertion #5)	Amp	this study
pSR3NN(1ins9)	3NN (1aa insertion #9)	Amp	this study
pSR3NN(1ins10)	3NN (1aa insertion #10)	Amp	this study
pSR3NN(1ins13)	3NN (1aa insertion #13)	Amp	this study
pSR3NN(1ins15)	3NN (1aa insertion #15)	Amp	this study
pSR3NN(1ins17)	3NN (1aa insertion #17)	Amp	this study
pSR3NN(1ins18)	3NN (1aa insertion #18)	Amp	this study
pSR3NN(1ins19)	3NN (1aa insertion #19)	Amp	this study
novel single repeat modules needed for the construction of *Xoc* TALEs (sorted by RVD and position)			
pSR3AN	3AN (single repeat module for position 3; RVD AN)	Amp	this study
pSR1HA	1HA	Amp	this study
pSR2HA	2HA	Amp	this study
pSR3HA	3HA	Amp	this study
pSR4HA	4HA	Amp	this study
pSR5HA	5HA	Amp	this study
pSR6HA	6HA	Amp	this study
pSR1HH	1HH	Amp	this study
pSR2HH	2HH	Amp	this study
pSR3HH	3HH	Amp	this study
pSR4HH	4HH	Amp	this study
pSR5HH	5HH	Amp	this study
pSR6HH	6HH	Amp	this study
pSR2HM	2HM	Amp	this study
pSR1HN	1HN	Amp	this study
pSR2HN	2HN	Amp	this study
pSR5HN	5HN	Amp	this study
pSR5HY	5HY	Amp	this study
pSR1sN*	1N* stop (single repeat module for position 1; RVD N*; stop repeat)	Amp	this study
pSR2sN*	2N* stop	Amp	this study
pSR3sN*	3N* stop	Amp	this study
pSR4sN*	4N* stop	Amp	this study
pSR2NA	2NA	Amp	this study
pSR3NC	3NC	Amp	this study
pSR5NC	5NC	Amp	this study
pSR1ND	1ND	Amp	this study
pSR2ND	2ND	Amp	this study
pSR3ND	3ND	Amp	this study
pSR4ND	4ND	Amp	this study
pSR5ND	5ND	Amp	this study
pSR6ND	6ND	Amp	this study
pSR1NV	6NV	Amp	this study
pSR3SN	3SN	Amp	this study
pSR5SN	5SN	Amp	this study
pSR1YG	1YG	Amp	this study
multi repeat modules for the construction of aberrant repeat-carrying TALEs			
pH043	1NG 2NI 3NH 4HD 5NG 6NH in assembly vector L-A	Kan	this study
pH055	1HD 2NI 3NN 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH045	1NG 2NI 3NH 4HD 5NI 6NG in assembly vector B-R	Kan	this study
pH056	1HD(40aa) 2NI 3NN 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH097	1HD(42aa) 2NI 3NN 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH120	1NG 2NI 3NH 4HD 5NG(stop) in assembly vector L-A	Kan	this study
pH210	1NH 2HD(30aa) 3NI 4NN 5NG 6NI in assembly vector A-B	Kan	this study
pH123	1HD 2NG 3NI 4NH 5HD 6NI in assembly vector B-C	Kan	this study
pH124	1NG(stop) in assembly vector C-R	Kan	this study
pH137	1NH 2HD(28aa) 3NI 4NN 5NG 6NI in assembly vector A-B	Kan	this study
pH268	1NH 2HD(36aa v1) 3NI 4NN 5NG 6NI in assembly vector A-B	Kan	this study
pH269	1NH 2HD(36aa v2) 3NI 4NN 5NG 6NI in assembly vector A-B	Kan	this study
pH274	1NH 2HD(35aa v1) 3NI 4NN 5NG 6NI in assembly vector A-B	Kan	this study

pH275	1NH 2HD(35aa v2) 3NI 4NN 5NG 6NI in assembly vector A-B	Kan	this study
pH333	1NH 2HD(37aa v1) 3NI 4NN 5NG 6NI in assembly vector A-B	Kan	this study
pH334	1NH 2HD(37aa v2) 3NI 4NN 5NG 6NI in assembly vector A-B	Kan	this study
pH017	1NN 2NN 3NN 4NN 5NN 6NN in assembly vector A-B	Kan	this study
pH023	1NN 2NN 3NN(40aa) 4NN 5NN 6NN in assembly vector A-B	Kan	this study
pH029	1NN 2NN 3NN(40aa) 4NN(40aa) 5NN 6NN in assembly vector A-B	Kan	this study
pH030	1NN 2NN(40aa) 3NN(40aa) 4NN(40aa) 5NN 6NN in assembly vector A-B	Kan	this study
pH031	1NN 2NN(40aa) 3NN(40aa) 4NN(40aa) 5NN(40aa) 6NN in assembly vector A-B	Kan	this study
pH063	1NN(40aa) 2NN(40aa) 3NN(40aa) 4NN(40aa) 5NN(40aa) 6NN in assembly vector A-B	Kan	this study
pH073	1NN(40aa) 2NN(40aa) 3NN(40aa) 4NN(40aa) 5NN(40aa) 6NN in assembly vector A-B(40aa p6)	Kan	this study
pH059	1HD(40aa) 2NI(40aa) 3NN 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH100	1HD(42aa) 2NI(42aa) 3NN 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH062	1HD(40aa) 2NI(40aa) 3NN(40aa) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH103	1HD(42aa) 2NI(42aa) 3NN(42aa) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH116	1NI 2NH 3HD 4NI 5NG(stop) in assembly vector B-R	Kan	this study
pH115	1HD 2NN 3NG 4NI 5HD 6NG in assembly vector A-B	Kan	this study
pH153	1HD(40aa) 2NN(40aa) 3NG 4NI 5HD 6NG in assembly vector A-B	Kan	this study
pH154	1HD(42aa) 2NN(42aa) 3NG 4NI 5HD 6NG in assembly vector A-B	Kan	this study
pH209	1HD(30aa) 2NN(30aa) 3NG 4NI 5HD 6NG in assembly vector A-B	Kan	this study
pH332	1HD(37aa v2) 2NN(37aa v2) 3NG 4NI 5HD 6NG in assembly vector A-B	Kan	this study
pH290	1NH 2HD(36aa v1) 3NN(36aa v1) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH213	1NH 2HD(28aa) 3NN(28aa) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH119	1HD(48aa) 2NI 3NN 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH065	1NN(40aa) 2NN 3NN 4NN 5NN 6NN in assembly vector A-B	Kan	this study
pH066	1NN(40aa) 2NN(40aa) 3NN 4NN 5NN 6NN in assembly vector A-B	Kan	this study
pH067	1NN(40aa) 2NN 3NN(40aa) 4NN 5NN 6NN in assembly vector A-B	Kan	this study
pH064	1NN(40aa) 2NN 3NN 4NN(40aa) 5NN 6NN in assembly vector A-B	Kan	this study
pH068	1NN(40aa) 2NN 3NN 4NN 5NN(40aa) 6NN in assembly vector A-B	Kan	this study
pH069	1NN(40aa) 2NN 3NN 4NN 5NN 6NN in assembly vector A-B(40aa p6)	Kan	this study
pH034	1NG 2NI 3NN 4HD 5NG 6NH in assembly vector L-A	Kan	this study
pH036	1NG 2NI 3NN 4HD(40aa) 5NG 6NH in assembly vector L-A	Kan	this study
pH159	1HD 2NI(40aa) 3NN 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH176	1HD 2NI 3NN 4NG 5NI 6HD in assembly vector A-B(40aa p6)	Kan	this study
pH160	1HD 2NI (40aa) 3NN 4NG 5NI 6HD in assembly vector A-B(40aa p6)	Kan	this study
pH264	1HD 2NI 3NN 4NG 5NI(stop) in assembly vector A-B	Kan	this study
pH265	1NG 2NI 3NH 4HD 5NI 6NG in assembly vector B-C	Kan	this study
pH262	1NG 2NI 3NN 4NG 5NH(stop) in assembly vector L-A	Kan	this study
pH291	1NS 2NG 3NI(36aa v1) 4NN(stop) in assembly vector L-A	Kan	this study
pH295	1NI 2NN(36aa v1) 3NN 4HD 5NI 6HD in assembly vector A-B	Kan	this study
pH296	1HD 2NS 3HD 4NI 5NI 6HD in assembly vector B-C	Kan	this study
pH294	1NG 2HD 3HD 4HD 5NG 6NG in assembly vector C-R	Kan	this study
pH270	1NH 2HD 3NN(36aa art1) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH271	1NH 2HD 3NN(36aa art2) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH272	1NH 2HD 3NN(36aa art3) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH273	1NH 2HD 3NN(36aa art4) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH278	1HD 2NN(36aa v1) 3NG 4NI 5HD 6NG in assembly vector A-B	Kan	this study
pH279	1HD 2NN(36aa v2) 3NG 4NI 5HD 6NG in assembly vector A-B	Kan	this study
pH194	1NH 2HD 3NN(40aa) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH179	1NH 2HD 3NN(40aa art1) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH175	1NH 2HD 3NN(40aa art2) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH148	1NH 2HD 3NN(42aa art1) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH143	1NH 2HD 3NN(42aa art2) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH195	1NH 2HD 3NN(42aa) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH144	1NH 2HD 3NN(7ins4) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH142	1NH 2HD 3NN(7ins5) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH216	1NH 2HD 3NN(7ins9) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH221	1NH 2HD 3NN(7ins10) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH143	1NH 2HD 3NN(7ins11) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH177	1NH 2HD 3NN(7ins13) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH217	1NH 2HD 3NN(7ins15) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH218	1NH 2HD 3NN(7ins17) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH222	1NH 2HD 3NN(7ins18) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH224	1NH 2HD 3NN(7ins19) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH178	1NH 2HD 3NN(7ins20) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH231	1NH 2HD 3NN(7ins22) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH219	1NH 2HD 3NN(7ins24) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH225	1NH 2HD 3NN(7ins25) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH223	1NH 2HD 3NN(7ins26) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH226	1NH 2HD 3NN(7ins28) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH174	1NH 2HD 3NN(7ins29) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH220	1NH 2HD 3NN(7ins31) 4NG 5NI 6HD in assembly vector A-B	Kan	this study

pH246	1NH 2HD 3NN(1ins5) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH247	1NH 2HD 3NN(1ins9) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH234	1NH 2HD 3NN(1ins10) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH248	1NH 2HD 3NN(1ins13) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH238	1NH 2HD 3NN(1ins15) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH249	1NH 2HD 3NN(1ins17) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH250	1NH 2HD 3NN(1ins18) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH267	1NH 2HD 3NN(1ins19) 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH083	1HD 2link7aa 3NN 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH106	1HD 2link18aa 3NN 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH132	1HD 2L1/HA/R1 3NN 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH133	1HD 2L1/GFP/R1 3NN 4NG 5NI 6HD in assembly vector A-B	Kan	this study
pH164	1HD 2L1/GFP/R1 3NI 4NN 5NG 6NI in assembly vector A-B	Kan	this study
pH162	1NG 2L1/GFP/R1 3NI 4NH 5HD 6NG in assembly vector L-A	Kan	this study
pH163	1NH 2HD 3NI 4NN 5NG 6NI in assembly vector A-B	Kan	this study
pH577	1NG 2L1/GFP/R1 3NI 4NH 5HD 6NI in assembly vector B-C	Kan	this study
pH580	1NH(stop) in assembly vector A-B	Kan	this study
pH581	1HD 2L1/GFP/R1 3NI 4NN 5NG 6NI in assembly vector B-C	Kan	this study
pH582	1HD(stop) in assembly vector C-D	Kan	this study
pH579	1NG 2L1/GFP/R1 3NI 4NH 5HD 6NI in assembly vector D-E	Kan	this study
pH587	1NG(stop) in assembly vector E-R	Kan	this study
pH371	1HD(stop) in assembly vector B-C	Kan	this study
pH578	1NG 2L1/GFP/R1 3NI 4NH 5HD 6NI in assembly vector C-D	Kan	this study
multi repeat modules for the construction of *Xoo* TALEs (sorted by TALE class and assembly vector)			
pH313	Tal2(K-74) (CD); 1NN 2NI 3NN 4NN 5NG(stop) in assembly vector C-D	Kan	this study
pH315	Tal5(MAI1) (LA); 1NN 2HD 3NN 4HD 5NG 6HD in assembly vector L-A	Kan	this study
pH316	Tal5(MAI1) (AB); 1HD 2NG 3HD 4NI 5NI 6NN in assembly vector A-B	Kan	this study
pH317	Tal5(MAI1) (BR); 1HD 2HD 3N* 4NG 5HD 6NI in assembly vector B-R	Kan	this study
pH320	TalC (LA); 1NS 2NG 3NS 4HD 5NI 6NG in assembly vector L-A	Kan	this study
pH321	TalC (AB); 1NN 2NG 3HD 4NI 5NN 6N* in assembly vector A-B	Kan	this study
pH322	TalC (BC); 1NI 2NN 3HD 4NG 5NI 6NN in assembly vector B-C	Kan	this study
pH323	TalC (CR); 1N* 2HD 3NN 4NG(stop) in assembly vector C-R	Kan	this study
pH324	TalAC2 (BC); 1N* 2N*(40aa) 3HD(stop) in assembly vector B-C	Kan	this study
pH325	TalAC2 (CR); 1HD 2NS 3NS 4NN 5N*(stop) in assembly vector C-R	Kan	this study
pH286	TalAM (LA); 1NI 2NG 3NI 4NN 5NN 6NI in assembly vector L-A	Kan	this study
pH287	TalAM (AB); 1NN 2HD 3NI 4HD 5NS 6NS in assembly vector A-B	Kan	this study
pH288	TalAM (BC); 1NS 2HD 3NN 4HD 5NG 6HD in assembly vector B-C	Kan	this study
pH289	TalAM (CR); 1HD 2HD 3NG 4NG(stop) in assembly vector C-R	Kan	this study
pH308	TalBH (LA); 1NI 2NG 3NI 4NG 5NI 6NI in assembly vector L-A	Kan	this study
pH309	TalBH (AB); 1NI 2HD 3NN 4HD 5HD 6HD in assembly vector A-B	Kan	this study
pH310	TalBH (BC); 1NG 2HD 3N*(40aa) 4NI 5HD 6HD in assembly vector B-C	Kan	this study
pH311	TalBH (CD); 1NN 2NS 3NI 4NN 5NN 6NG in assembly vector C-D	Kan	this study
pH312	TalBH (DR); 1NN 2HD 3N* 4NS 5N*(stop) in assembly vector D-R	Kan	this study
pH280/pLABK2	TalBK (LA); 1NI 2HG 3NI 4NN 5NI 6NN in assembly vector L-A	Kan	Stefanie Mücke
pH281/pABBK2	TalBK (AB1); 1HD 2NI 3HD(36aa v1) 4HD 5NS(stop) in assembly vector A-B	Kan	Stefanie Mücke
pH282	TalBK (AB2); 1HD 2HD 3HD(36aa v1) 4HD 5NS(stop) in assembly vector A-B	Kan	this study
pH283	TalBK (BR); 1NS(36aa v1) 2HD 3NI 4NI 5HD(stop) in assembly vector B-R	Kan	this study
pH284/pBCBK2	TalBK (BC); 1NS(36aa v1) 2HD 3NI 4NI 5HD 6NG in assembly vector B-C	Kan	Stefanie Mücke
pH285/pCRBK2	TalBK (CR); 1HD 2HD 3HD 4NG 5NG(stop) in assembly vector C-R	Kan	Stefanie Mücke
pH299	TalDS (LA); 1NI 2NG 3NI 4NI 5NS(stop) in assembly vector L-A	Kan	this study
pH300	TalDS (AB); 1HD 2HD 3HD 4HD 5NS(stop) in assembly vector A-B	Kan	this study
pH301	TalDS (BC); 1N* 2N*(40aa) 3HD 4HD 5NS(stop) in assembly vector B-C	Kan	this study
pH302	TalDS (CD); 1NS 2NN 3NN 4NI 5NG(stop) in assembly vector C-D	Kan	this study
pH303	TalDS (DR); 1NN 2NI 3N* 4NS 5N*(stop) in assembly vector D-R	Kan	this study
pH304	TalDV (AB); 1HD 2NN 3HD 4HD 5HD 6NS in assembly vector A-B	Kan	this study
pH305	TalDV (BC); 1HD 2N*(40aa) 3NI 4HD 5HD(stop) in assembly vector B-C	Kan	this study
pH306	TalDV (CD); 1NN 2NS 3NN 4NN 5NG 6NN in assembly vector C-D	Kan	this study
pH307	TalDV (DR); 1HD 2N* 3NS 4NS 5N*(stop) in assembly vector D-R	Kan	this study
multi repeat modules for the construction of *Xoc* TALEs (sorted by TALE class and assembly vector position)			
pH508	TalAF (LA1); 1NI 2NN 3NI 4NI 5NI 6NI in assembly vector L-A	Kan	this study
pH509	TalAF (LA2); 1NI 2NN 3NN 4NI 5NI 6NI in assembly vector L-A	Kan	this study
pH510	TalAF (AB1); 1N* 2NI 3NG 4NN 5NN 6NN in assembly vector A-B	Kan	this study
pH511	TalAF (AB2); 1HD 2NS 3HG 4NN 5NN 6NN in assembly vector A-B	Kan	this study
pH512	TalAF (BR1); 1NI 2NS 3NG 4NG(stop) in assembly vector B-R	Kan	this study
pH513	TalAF (BR2); 1NG 2NS 3NG 4HD(stop) in assembly vector B-R	Kan	this study
pH514	TalAF (BR3); 1NI 2NI 3NG 4HD(stop) in assembly vector B-R	Kan	this study
pH531	TalAJ (LA); 1NN 2HD 3NI 4NK 5HD 6HD in assembly vector L-A	Kan	this study
pH532	TalAJ (AB); 1HD 2NG 3NI 4NN 5HD 6HD in assembly vector A-B	Kan	this study
pH498	TalAK (LA1); 1HD 2HG 3HD 4HG 5N* 6NN in assembly vector L-A	Kan	this study
pH499	TalAK (LA2); 1HD 2HD 3HD 4HD 5HD(stop) in assembly vector L-A	Kan	this study
pH500	TalAK (AB1); 1NG 2HD 3NN 4HD 5NG 6NG in assembly vector A-B	Kan	this study
pH501	TalAK (AB2); 1NG 2HD 3NN 4HD 5NG 6HG in assembly vector A-B	Kan	this study
pH502	TalAK (BR); 1NN 2HD 3N* 4NG 5NG(stop) in assembly vector B-R	Kan	this study

pH407	TalAT (LA); 1NN 2HD 3HD 4HH 5NN 6NG in assembly vector L-A	Kan	this study
pH408	TalAT (AB1); 1NG 2NG 3HD 4NI 5NG 6HD in assembly vector A-B	Kan	this study
pH409	TalAT (AB2); 1NS 2NG 3HD 4NI 5NG 6HD in assembly vector A-B	Kan	this study
pH410	TalAT (BR); 1HH 2NG 3NG 4NG 5NG(stop) in assembly vector B-R	Kan	this study
pH411	TalAT (BC); 1HH 2NG 3NG(stop) in assembly vector B-C	Kan	this study
pH412	TalAT (CR); 1NG 2HA 3NG(stop) in assembly vector C-R	Kan	this study
pH533	TalAU (LA); 1HD 2HD 3NG 4NI 5HG 6HD in assembly vector L-A	Kan	this study
pH534	TalAU (AB1); 1HG 2NI 3NI 4NN 5HD 6HG in assembly vector A-B	Kan	this study
pH535	TalAU (AB2); 1HG 2NI 3HD 4NN 5HD 6HG in assembly vector A-B	Kan	this study
pH536	TalAU (BR); 1HG 2HG 3NG 4NI 5NG(stop) in assembly vector B-R	Kan	this study
pH382	TalAV (LA); 1NN 2NS 3NG 4NI 5HD 6HD in assembly vector L-A	Kan	this study
pH383	TalAV (AB); 1NG 2NN 3HD 4NI 5NG 6N* in assembly vector A-B	Kan	this study
pH384	TalAV (BC); 1N* 2NG 3HD 4HD(stop) in assembly vector B-C	Kan	this study
pH385	TalAV (CR1); 1NS 2NI 3NN 4NG 5NG(stop) in assembly vector C-R	Kan	this study
pH386	TalAV (CR2); 1NS 2NI 3NN 4NG 5N*(stop) in assembly vector C-R	Kan	this study
pH387	TalAX (LA); 1NI 2N* 3NI 4NI 5NN 6N* in assembly vector L-A	Kan	this study
pH388	TalAX (AB); 1NI 2NK 3HD 4HD 5HD 6NG in assembly vector A-B	Kan	this study
pH389	TalAX (BR1); 1NI 2NN 3HD 4NN 5HD 6NG in assembly vector B-R	Kan	this study
pH390	TalAX (BR2); 1NI 2NN 3HD 4N* 5HD 6NG in assembly vector B-R	Kan	this study
pH391	TalAY (LA1); 1HD 2HD 3NI 4HD 5HD 6NG in assembly vector L-A	Kan	this study
pH392	TalAY (LA2); 1HD 2HD 3HD 4HD 5HD 6NG in assembly vector L-A	Kan	this study
pH393	TalAY (AB); 1HD 2NN 3HD 4HD 5NG 6NG in assembly vector A-B	Kan	this study
pH394	TalAY (BR1); 1NN 2NG 3NG 4HD 5HD 6NG in assembly vector B-R	Kan	this study
pH395	TalAY (BR2); 1NN 2NG 3HD 4HD 5HD 6NG in assembly vector B-R	Kan	this study
pH446	TalAZ (LA); 1HH 2NN 3HD 4HD 5NG 6N* in assembly vector L-A	Kan	this study
pH447	TalAZ (AB); 1HD 2NN 3HD 4ND 5N* 6NG in assembly vector A-B	Kan	this study
pH448	TalAZ (BR); 1NS 2NS 3NN 4HD 5HD 6NG in assembly vector B-R	Kan	this study
pH492	TalBB (LA1); 1NS 2HG 3NS 4NG 5NS 6NN in assembly vector L-A	Kan	this study
pH493	TalBB (LA3); 1NI 2NG 3NI 4NG 5NS 6NN in assembly vector L-A	Kan	this study
pH494	TalBB (LA4); 1NN 2HG 3NI 4NG 5NI 6NN in assembly vector L-A	Kan	this study
pH495	TalBB (AB); 1NG 2NN 3NG 4HD 5NN 6HD in assembly vector A-B	Kan	this study
pH496	TalBB (BR1); 1HD 2NG(stop) in assembly vector B-R	Kan	this study
pH497	TalBB (BR2); 1HD 2N*(stop) in assembly vector B-R	Kan	this study
pH479	TalBC (LA1); 1NN 2NG 3NI 4 NN 5NG 6NN in assembly vector L-A	Kan	this study
pH480	TalBC (LA2); 1Nn 2HG 3NI 4NN 5NG 6NN in assembly vector L-A	Kan	this study
pH481	TalBC (LA3); 1NI 2HG 3NI 4NN 5NG 6NN in assembly vector L-A	Kan	this study
pH482	TalBC (LA4); 1NS 2HG 3NI 4NG 5NS 6NN in assembly vector L-A	Kan	this study
pH483	TalBC (AB); 1ND 2N* 3NG 4N* 5HN 6NN in assembly vector A-B	Kan	this study
pH484	TalBC (BC1); 1HD 2NN 3NI 4NN 5HD 6HD in assembly vector B-C	Kan	this study
pH486	TalBC (BC3); 1HD 2NS 3NI 4NN 5HD 6HD in assembly vector B-C	Kan	this study
pH487	TalBC (CD4); 1NG 2NG 3NG 4HD 5NG 6HD in assembly vector C-D	Kan	this study
pH488	TalBC (CR1); 1NG 2NG 3NG 4HD 5NG(stop) in assembly vector C-R	Kan	this study
pH489	TalBC (CR2); 1NG 2NG 3NG 4HD 5N*(stop) in assembly vector C-R	Kan	this study
pH490	TalBC (CR3); 1NG 2HD 3HG 4HD 5NG(stop) in assembly vector C-R	Kan	this study
pH491	TalBC (DR); 1NG(stop) in assembly vector D-R	Kan	this study
pH449	TalBD (LA); 1NN 2HD 3NI 4NI 5NG 6NN in assembly vector L-A	Kan	this study
pH450	TalBD (AB1); 1NN 2HD 3NN 4NI 5NI 6NN in assembly vector A-B	Kan	this study
pH451	TalBD (AB2); 1NN 2HD 3HH 4NI 5NI 6NN in assembly vector A-B	Kan	this study
pH452	TalBD (BR); 1NN 2NN 3NG 4NI 5HD 6NG in assembly vector B-R	Kan	this study
pH453	TalBE (LA); 1HD 2NN 3HD 4N* 5NS 6NN in assembly vector L-A	Kan	this study
pH454	TalBE (AB1); 1HG 2NI 3NI 4NS 5NI 6NG in assembly vector A-B	Kan	this study
pH455	TalBE (AB2); 1NG 2NI 3NI 4NG 5HY 6NG in assembly vector A-B	Kan	this study
pH456	TalBE (BC); 1HD 2NN 3HD 4HD 5NG 6HG in assembly vector B-C	Kan	this study
pH457	TalBE (CR); 1NG(stop) in assembly vector C-R	Kan	this study
pH503	TalBF (LA1); 1NN 2NN 3ND 4N* 5NS 6NS in assembly vector L-A	Kan	this study
pH504	TalBF (LA2); 1NN 2NN 3HD 4N* 5NS 6NS in assembly vector L-A	Kan	this study
pH505	TalBF (AB1); 1YG 2NI 3SN 4ND 5HD 6NG in assembly vector A-B	Kan	this study
pH506	TalBF (AB2); 1YG 2NI 3SN 4HD 5HD 6NG in assembly vector A-B	Kan	this study
pH507	TalBF (BR); 1HD 2NS 3NN 4HD 5NG(stop) in assembly vector B-R	Kan	this study
pH326	TalBG (LA); 1NI 2NN(stop) in assembly vector L-A	Kan	this study
pH327	TalBG (AB1); 1NI 2HD 3NN(30aa) 4NG 5HD 6NN in assembly vector A-B	Kan	this study
pH328	TalBG (BC); 1HD 2NG 3HD 4NG 5NG 6HD in assembly vector B-C	Kan	this study
pH329	TalBG (CR); 1HD 2NG(stop) in assembly vector C-R	Kan	this study
pH330	TalBG (AB2); 1NN 2NN 3NI(30aa) 4NG 5HD 6NN in assembly vector A-B	Kan	this study
pH458	TalBI (LA1); 1NN 2HD 3NI 4HD 5NN 6NG in assembly vector L-A	Kan	this study
pH459	TalBI (LA2); 1NN 2HD 3NI 4HD 5NS 6NG in assembly vector L-A	Kan	this study
pH460	TalBI (LA3); 1NN 2HD 3NI 4HD 5NI 6NG in assembly vector L-A	Kan	this study
pH461	TalBI (AB); 1NG 2HD 3NN 4HD 5N* 6NN in assembly vector A-B	Kan	this study
pH462	TalBI (BC); 1HD 2HD 3HD 4NG 5HD 6NN in assembly vector B-C	Kan	this study
pH463	TalBI (BR1); 1HD 2HD 3HD 4NG(stop) in assembly vector B-R	Kan	this study
pH464	TalBI (BR2); 1HD 2HD 3NN 4HD 5HD 6NG in assembly vector B-R	Kan	this study
pH465	TalBI (CR); 1HD 2NS 3NI 4NG(stop) in assembly vector C-R	Kan	this study
pH473	TalBL (LA1); 1NI 2HG 3N* 4NN 5N* 6HD in assembly vector L-A	Kan	this study

pH474	TalBL (LA2); 1NI 2NG 3N* 4NN 5N* 6HD in assembly vector L-A	Kan	this study
pH475	TalBL (AB1); 1HG 2HD 3NN 4HD 5NN 6HG in assembly vector A-B	Kan	this study
pH530	TalBL (AB2); 1NG 2NS 3N* 4HD(stop) in assembly vector B-R	Kan	this study
pH476	TalBL (BR1); 1HG 2NS 3NG 4HD(stop) in assembly vector B-R	Kan	this study
pH477	TalBL (BR2); 1NN 2NN 3NG 4HD(stop) in assembly vector B-R	Kan	this study
pH478	TalBL (BR3); 1NI 2HG 3N* 4NN 5N* 6HD in assembly vector L-A	Kan	this study
pH401	TalBN (LA); 1NI 2NN 3NI 4NI 5NG 6HD in assembly vector L-A	Kan	this study
pH402	TalBN (AB); 1NG 2NN 3NI 4NN 5NI 6NN in assembly vector A-B	Kan	this study
pH403	TalBN (BR1); 1HD(stop) in assembly vector B-R	Kan	this study
pH404	TalBN (BR2); 1NG(stop) in assembly vector B-R	Kan	this study
pH467	TalBO (LA2); 1NN 2HG 3HD 4NG 5HD 6HD in assembly vector L-A	Kan	this study
pH468	TalBO (LA3); 1NN 2NG 3HD 4NG 5HD 6HD in assembly vector L-A	Kan	this study
pH469	TalBO (AB); 1NI 2NH 3HG 4NI 5NN 6NN in assembly vector A-B	Kan	this study
pH470	TalBO (BC); 1HD 2NS 3NG(stop) in assembly vector B-C	Kan	this study
pH471	TalBO (CR1); 1NN 2HD 3NG 4NG 5NG 6NG in assembly vector C-R	Kan	this study
pH472	TalBO (CR2); 1NN 2HD 3NG 4NG 5N*(stop) in assembly vector C-R	Kan	this study
pH515	TalBR (LA); 1HD 2HD 3NN 4NN 5NG 6N* in assembly vector L-A	Kan	this study
pH516	TalBR (AB); 1HD 2NI 3NG 4HD 5NG 6NS in assembly vector A-B	Kan	this study
pH517	TalBR (BC1); 1ND 2HA 3ND 4N* 5ND 6NN in assembly vector B-C	Kan	this study
pH521	TalBR (CD1); 1HD 2NN 3NN 4HD 5HD 6N* in assembly vector C-D	Kan	this study
pH523	TalBR (DR); 1NN 2NG 3HD(stop) in assembly vector D-R	Kan	this study
pH438	TalBU (LA1); 1NI 2N* 3NI 4NG(stop) in assembly vector L-A	Kan	this study
pH439	TalBU (LA2); 1NI 2HD 3NI 4N* 5NI 6NG in assembly vector L-A	Kan	this study
pH440	TalBU (LA3); 1NS 2HD 3NI 4N* 5NI 6HG in assembly vector L-A	Kan	this study
pH441	TalBU (AB1); 1NI 2NI 3N* 4NG 5HD 6NN in assembly vector A-B	Kan	this study
pH442	TalBU (AB2); 1NI 2NI 3NG 4NG 5HD 6NN in assembly vector A-B	Kan	this study
pH443	TalBU (BC); 1NI 2NG 3HD 4NN 5NS 6NN in assembly vector B-C	Kan	this study
pH444	TalBU (CR1); 1NG 2NG(stop) in assembly vector C-R	Kan	this study
pH445	TalBU (CR2); 1NI 2NN(stop) in assembly vector C-R	Kan	this study
pH353	TalBV (LA); 1NN 2HD 3NI 4NI 5NN 6HA in assembly vector L-A	Kan	this study
pH354	TalBV (AB); 1NN 2NS(stop) in assembly vector A-B	Kan	this study
pH355	TalBV (BC1); 1NI 2NI 3HD 4HA 5HA 6HA in assembly vector B-C	Kan	this study
pH356	TalBV (BC2); 1NS 2NI 3HD 4HA 5HA 6HA in assembly vector B-C	Kan	this study
pH357	TalBV (BC3); 1NS 2NI(36v2) 3HD 4HA 5HA 6HA in assembly vector B-C	Kan	this study
pH358	TalBV (CR1); 1HD 2HD 3HD 4HA 5HD 6N* in assembly vector C-R	Kan	this study
pH359	TalBV (CR2); 1HD 2HD 3HD 4NI 5HD 6N* in assembly vector C-R	Kan	this study
pH431	TalBW (LA1); 1NI 2NG 3NI 4NG 5NI 6NV in assembly vector L-A	Kan	this study
pH432	TalBW (LA2); 1NI 2NG 3NS 4NG 5NN 6NG in assembly vector L-A	Kan	this study
pH433	TalBW (AB1); 1NN 2NG 3HD 4NN 5NN 6NG in assembly vector A-B	Kan	this study
pH434	TalBW (AB2); 1HH 2NG 3HD 4NN 5NN 6NG in assembly vector A-B	Kan	this study
pH435	TalBW (BC); 1HD 2NN 3NS 4NN 5HD 6HD in assembly vector B-C	Kan	this study
pH436	TalBW (CD); 1NG 2NA 3NN(stop) in assembly vector C-D	Kan	this study
pH437	TalBW (DR); 1HD 2HD 3HD 4HD 5N*(stop) in assembly vector D-R	Kan	this study
pH413	TalBY (LA1); 1HD 2HD 3NC 4HD 5NC 6NG in assembly vector L-A	Kan	this study
pH414	TalBY (LA2); 1HD 2HD 3HD 4HD 5HD 6NG in assembly vector L-A	Kan	this study
pH415	TalBY (AB); 1HD 2HH 3HD 4NI 5NG 6N* in assembly vector A-B	Kan	this study
pH416	TalBY (BC); 1NS 2N* 3HD 4HD 5NS 6NI in assembly vector B-C	Kan	this study
pH417	TalBY (CR); 1NG 2HD 3NG(stop) in assembly vector C-R	Kan	this study
pH360	TalCB (LA1); 1NS 2HD 3NI 4NS 5NN 6HA in assembly vector L-A	Kan	this study
pH361	TalCB (B-C): 1NS 2NI 3HD 4HD 5HD 6HA in assembly vector B-C	Kan	this study
pH362	TalCB (CR); 1HD 2NN 3HD 4NN 5HD 6N* in assembly vector C-R	Kan	this study
pH363	TalCB (LA2); 1NN 2HD 3NS 4NS 5NN 6HA in assembly vector L-A	Kan	this study
pH537	TalCD (LA); 1HD 2HG 3N* 4NN 5NG 6HD in assembly vector L-A	Kan	this study
pH538	TalCD (LA2); 1HD 2HG 3N* 4NN 5HG 6HD in assembly vector L-A	Kan	this study
pH539	TalCD (AB); 1NN 2HD 3NN 4NG 5NN 6HD in assembly vector A-B	Kan	this study
pH540	TalCD (BR1); 1N* 2NG 3NG(stop) in assembly vector B-R	Kan	this study
pH541	TalCD (BR2); 1NN 2HD 3NG(stop) in assembly vector B-R	Kan	this study
pH542	TalCE (LA); 1NN 2HD 3NI 4N* 5NI 6NI in assembly vector L-A	Kan	this study
pH543	TalCE (AB1); 1NK 2HD 3HD 4HD 5HG 6NI in assembly vector A-B	Kan	this study
pH544	TalCE (AB2); 1NK 2HD 3HD 4HD 5NG 6HD in assembly vector A-B	Kan	this study
pH545	TalCE (BR); 1NN 2HD 3HD(stop) in assembly vector B-R	Kan	this study
pH583	TalCF (LA1); 1NN 2NG 3NI 4N* 5NS 6NG in assembly vector L-A	Kan	this study
pH584	TalCF (LA2); 1NN 2NG 3NS 4HD 5NI 6NG in assembly vector L-A	Kan	this study
pH585	TalCF (AB1); 1NN 2NN 3HG 4ND 5ND 6HD in assembly vector A-B	Kan	this study
pH586	TalCF (AB2); 1NN 2NN 3HG 4ND 5ND 6HG in assembly vector A-B	Kan	this study
pH396	TalCG (LA1); 1HD 2NG 3N* 4NN 5NG 6NN in assembly vector L-A	Kan	this study
pH397	TalCG (LA2); 1HD 2HD 3NN 4NN 5NN(stop) in assembly vector L-A	Kan	this study
pH398	TalCG (AB); 1NI 2NN 3NI 4NN 5NN 6HD in assembly vector A-B	Kan	this study
pH399	TalCG (BR1); 1NG 2NG 3HD 4NG(stop) in assembly vector B-R	Kan	this study
pH400	TalCG (BR2); 1NG 2NG(stop) in assembly vector B-R	Kan	this study
pH418	TalCH (LA); 1HD 2HD 3NN 4NN 5NG 6N* in assembly vector L-A	Kan	this study
pH419	TalCH (AB); 1HD 2NI 3NG 4HD 5NG 6NS in assembly vector A-B	Kan	this study
pH420	TalCH (BC); 1HD 2HA 3HD 4N* 5HH 6NG in assembly vector B-C	Kan	this study

pH421	TalCH (CD); 1NI 2NN 3NN 4HD 5HD 6NG in assembly vector C-D	Kan	this study
pH422	TalCH (DR1); 1NN 2NG 3HD(stop) in assembly vector D-R	Kan	this study
pH423	TalCH (DR2); 1HD(stop) in assembly vector D-R	Kan	this study
pH546	TalCI (AB); 1HD 2NI 3NG 4HD 5NG 6NG in assembly vector A-B	Kan	this study
pH547	TalCI (BC1); 1NS 2ND 3HA 4HD 5N* 6N* in assembly vector B-C	Kan	this study
pH548	TalCI (BC2); 1NS 2ND 3HA 4HD 5N* 6HD in assembly vector B-C	Kan	this study
pH549	TalCI (CR); 1NN 2NI 3NN 4NN 5HD 6NG in assembly vector C-R	Kan	this study
pH550	TalCI (CD); 1NG 2NI 3NN 4NN 5HD 6HD in assembly vector C-D	Kan	this study
pH551	TalCI (DR); 1NG 2HD(stop) in assembly vector D-R	Kan	this study
pH552	TalCJ (LA); 1NI 2HG 3NI 4NG 5HN 6NG in assembly vector L-A	Kan	this study
pH553	TalCJ (AB); 1NN 2HD 3ND 4NG 5HD 6NI in assembly vector A-B	Kan	this study
pH554	TalCJ (BC); 1HG 2NG 3HD 4HH 5N* 6HH in assembly vector B-C	Kan	this study
pH555	TalCJ (CD); 1N* 2HD 3NN 4NG 5HD 6HD in assembly vector C-D	Kan	this study
pH556	TalCK (LA); 1NI 2HG 3N* 4NI 5SN 6HD in assembly vector L-A	Kan	this study
pH557	TalCK (AB1); 1HD 2HD 3NN 4NI 5NK 6HD in assembly vector A-B	Kan	this study
pH558	TalCK (AB2); 1HD 2HD 3NN 4HD 5NK 6HD in assembly vector A-B	Kan	this study
pH559	TalCK (BR1); 1NN 2NN 3HD 4NG(stop) in assembly vector B-R	Kan	this study
pH560	TalCK (BR2); 1HD 2NN 3HD 4NG(stop) in assembly vector B-R	Kan	this study
pH424	TalCL (LA1); 1NI 2NG 3NI 4N* 5HN 6NG in assembly vector L-A	Kan	this study
pH425	TalCL (LA2); 1NI 2NG 3NI 4NG 5HN 6NG in assembly vector L-A	Kan	this study
pH426	TalCL (AB); 1NN 2HD 3HD 4NG 5HD 6NI in assembly vector A-B	Kan	this study
pH427	TalCL (BC); 1NG 2NG 3HD 4NN 5HD 6HH in assembly vector B-C	Kan	this study
pH428	TalCL (CD); 1N* 2HD 3NN 4NN 5NG 6HD in assembly vector C-D	Kan	this study
pH429	TalCL (DR1); 1NN 2NN 3NG 4NN 5NI 6NG in assembly vector D-R	Kan	this study
pH430	TalCL (DR2); 1NN 2NN 3NG 4NN 5NI 6NV in assembly vector D-R	Kan	this study
pH561	TalCP (LA); 1HD 2HG 3HD 4NI 5NN 6ND in assembly vector L-A	Kan	this study
pH562	TalCP (AR); 1HD 2HD 3NN 4NN 5NI 6HD in assembly vector A-R	Kan	this study
pH563	TalCQ (LA); 1NN 2NI 3HN 4NG 5NN 6NN in assembly vector L-A	Kan	this study
pH564	TalCQ (AB); 1NG 2NI 3N* 4NI 5NG 6NN in assembly vector A-B	Kan	this study
pH565	TalCQ (BR); 1NN 2HG 3HD 4HD 5NG(stop) in assembly vector B-R	Kan	this study
pH566	TalCR (LA); 1NN 2NN 3HD 4NI 5NG 6HD in assembly vector L-A	Kan	this study
pH567	TalCR (AB); 1HD 2HD 3NG 4NI 5HD 6HD in assembly vector A-B	Kan	this study
pH568	TalCS (BC); 1HD 2NS 3NS 4NN 5NG 6NN in assembly vector B-C	Kan	this study
pH569	TalCS (CR); 1NG 2HD 3NN 4HD 5HD 6NG in assembly vector C-R	Kan	this study
overexpression constructs for aberrant repeat-carrying TALEs (sorted in order of appearance)			
pT048	NG NI NH HD NG NH HD NI NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH055, pH045, pCTH3	Spec	this study
pT299	NG NI NH HD NG NH HD(35aa v1) NI NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH274, pH123, pH124, pCTH3	Spec	this study
pT300	NG NI NH HD NG NH HD(35aa v2) NI NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH275, pH123, pH124, pCTH3	Spec	this study
pT090	NG NI NH HD NG NH HD(42aa) NI NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH097, pH045, pCTH3	Spec	this study
pT049	NG NI NH HD NG NH HD(40aa) NI NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH056, pH045, pCTH3	Spec	this study
pT351	NG NI NH HD NG NH HD(37aa v1) NI NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH333, pH123, pH124, pCTH3	Spec	this study
pT352	NG NI NH HD NG NH HD(37aa v2) NI NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH334, pH123, pH124, pCTH3	Spec	this study
pT294	NG NI NH HD NG NH HD(36aa v1) NI NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH268, pH123, pH124, pCTH3	Spec	this study
pT295	NG NI NH HD NG NH HD(36aa v2) NI NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH269, pH123, pH124, pCTH3	Spec	this study
pT130	NG NI NH HD NG NH HD(30aa) NI NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH210, pH123, pH124, pCTH3	Spec	this study
pT143	NG NI NH HD NG NH HD(28aa) NI NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH137, pH123, pH124, pCTH3	Spec	this study
pT115	NG NI NH HD NG NH NN NN NN NN NN NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH017, pH045, pCTH3	Spec	this study
pT144	NG NI NH HD NG NH NN NN NN(40aa) NN NN NN NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH022, pH045, pCTH3	Spec	this study
pT145	NG NI NH HD NG NH NN NN NN(40aa) NN(40aa) NN NN NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH029, pH045, pCTH3	Spec	this study
pT146	NG NI NH HD NG NH NN NN(40aa) NN(40aa) NN(40aa) NN NN NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH030, pH045, pCTH3	Spec	this study
pT147	NG NI NH HD NG NH NN NN(40aa) NN(40aa) NN(40aa) NN(40aa) NN NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH031, pH045, pCTH3	Spec	this study
pT148	NG NI NH HD NG NH NN(40aa) NN(40aa) NN(40aa) NN(40aa) NN(40aa) NN NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH063, pH045, pCTH3	Spec	this study
pT149	NG NI NH HD NG NH NN(40aa) NN(40aa) NN(40aa) NN(40aa) NN(40aa) NN(40aa) NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH073, pH045, pCTH3	Spec	this study
pT052	NG NI NH HD NG NH HD(40aa) NI(40aa) NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH059, pH045, pCTH3	Spec	this study

pT093	NG NI NH HD NG NH HD(42aa) NI(42aa) NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH100, pH045, pCTH3	Spec	this study
pT055	NG NI NH HD NG NH HD(40aa) NI(40aa) NN(40aa) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH062, pH045, pCTH3	Spec	this study
pT096	NG NI NH HD NG NH HD(42aa) NI(42aa) NN(42aa) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH103, pH045, pCTH3	Spec	this study
pT108	NG NI NH HD NG NH HD NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3; pH43, pH115, pH116, pCTH3	Spec	this study
pT162	NG NI NH HD NG NH HD(42aa) NN(42aa) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH154, pH116, pCTH3	Spec	this study
pT161	NG NI NH HD NG NH HD(40aa) NN(40aa) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH153, pH116, pCTH3	Spec	this study
pT323	NG NI NH HD NG NH HD(36aa v1) NN(36aa v1) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH290, pH045, pCTH3	Spec	this study
pT128	NG NI NH HD NG NH HD(30aa) NN(30aa) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH209, pH116, pCTH3	Spec	this study
pT233	NG NI NH HD NG NH HD(28aa) NN(28aa) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH213, pH045, pCTH3	Spec	this study
pT350	NG NI NH HD NG NH HD(37aa v2) NN(37aa v2) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH332, pH116, pCTH3	Spec	this study
pT111	NG NI NH HD NG NH HD(48aa) NI NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH119, pH045, pCTH3	Spec	this study
pT151	NG NI NH HD NG NH NN(40aa) NN NN NN NN NN NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH065, pH045, pCTH3	Spec	this study
pT152	NG NI NH HD NG NH NN(40aa) NN(40aa) NN NN NN NN NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH066, pH045, pCTH3	Spec	this study
pT153	NG NI NH HD NG NH NN(40aa) NN NN(40aa) NN NN NN NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH067, pH045, pCTH3	Spec	this study
pT154	NG NI NH HD NG NH NN(40aa) NN NN NN(40aa) NN NN NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH064, pH045, pCTH3	Spec	this study
pT155	NG NI NH HD NG NH NN(40aa) NN NN NN NN(40aa) NN NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH068, pH045, pCTH3	Spec	this study
pT156	NG NI NH HD NG NH NN(40aa) NN NN NN NN NN(40aa) NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH069, pH045, pCTH3	Spec	this study
pT181	NG NI NN HD NG NH HD NI NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH034, pH055, pH045, pCTH3	Spec	this study
pT163	NG NI NN HD(40aa) NG NH HD NI(40aa) NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH036, pH159, pH045, pCTH3	Spec	this study
pT180	NG NI NN HD NG NH HD NI(40aa) NN NG NI HD(40aa) NG NI NH HD NI NG in pSKA2; pNTH3, pH034, pH160, pH045, pCTH3	Spec	this study
pT197	NG NI NN HD(40aa) NG NH HD NI NN NG NI HD(40aa) NG NI NH HD NI NG in pSKA2; pNTH3, pH036, pH176, pH045, pCTH3	Spec	this study
pT164	NG NI NN HD(40aa) NG NH HD NI(40aa) NN NG NI HD(40aa) NG NI NH HD NI NG in pSKA2; pNTH3, pH036, pH160, pH045, pCTH3	Spec	this study
pT177	NG NI NN HD(40aa) NG NH HD NI NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH036, pH055, pH045	Spec	this study
pT179	NG NI NN HD NG NH HD NI NN NG NI HD(40aa) NG NI NH HD NI NG in pSKA2; pNTH3, pH034, pH176, pH045	Spec	this study
pT313	NG NI NN HD(40aa) NG NH HD NI NN NG NI NG NI NH HD NI NG in pSKA2; pNTH3, pH036, pH264, pH265, pH045, pCTH3	Spec	this study
pT314	NG NI NN NG NH HD NI NN NG NI HD(40aa) NG NI NH HD NI NG in pSKA2; pNTH3, pH262, pH176, pH265, pH045, pCTH3	Spec	this study
pT324	flexTALE; NS NG NI(36aa v1) NN NI NN(36aa v1) NN HD NI HD HD NS HD NI NI HD NG HD HD HD NG NG in pSKA2; pNTH3, pH291, pH295, pH296, pH294, pCTH3	Spec	this study
pT296	NG NI NH HD NG NH HD NN(36aa art1) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH270, pH45, pCHT3	Spec	this study
pT297	NG NI NH HD NG NH HD NN(36aa art2) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH271, pH45, pCHT3	Spec	this study
pT305	NG NI NH HD NG NH HD NN(36aa art3) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH272, pH45, pCHT3	Spec	this study
pT298	NG NI NH HD NG NH HD NN(36aa art4) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH273, pH45, pCHT3	Spec	this study
pT303	NG NI NH HD NG NH HD NN(36aa v1) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH43, pH278, pH116, pCTH3	Spec	this study
pT305	NG NI NH HD NG NH HD NN(36aa v1) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH43, pH279, pH116, pCTH3	Spec	this study
pT200	NG NI NH HD NG NH HD NN(40aa) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH194, pH45; pCTH3	Spec	this study
pT185	NG NI NH HD NG NH HD NN(40aa art1) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH179, pH45; pCTH3	Spec	this study
pT176	NG NI NH HD NG NH HD NN(40aa art2) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH175, pH45; pCTH3	Spec	this study
pT201	NG NI NH HD NG NH HD NN(42aa) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH195, pH45; pCTH3	Spec	this study

pT139	NG NI NH HD NG NH HD NN(42aa art1) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH148, pH45; pCTH3	Spec	this study
pT134	NG NI NH HD NG NH HD NN(42aa art2) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH143, pH45; pCTH3	Spec	this study
pT135	NG NI NH HD NG NH HD NN(7ins4) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH144, pH045, pCTH3	Spec	this study
pT133	NG NI NH HD NG NH HD NN(7ins5) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH142, pH045, pCTH3	Spec	this study
pT227	NG NI NH HD NG NH HD NN(7ins9) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH216, pH045, pCTH3	Spec	this study
pT239	NG NI NH HD NG NH HD NN(7ins10) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH221, pH045, pCTH3	Spec	this study
pT134	NG NI NH HD NG NH HD NN(7ins11) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH143, pH045, pCTH3	Spec	this study
pT183	NG NI NH HD NG NH HD NN(7ins13) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH177, pH045, pCTH3	Spec	this study
pT228	NG NI NH HD NG NH HD NN(7ins15) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH217, pH045, pCTH3	Spec	this study
pT229	NG NI NH HD NG NH HD NN(7ins17) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH218, pH045, pCTH3	Spec	this study
pT240	NG NI NH HD NG NH HD NN(7ins18) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH222, pH045, pCTH3	Spec	this study
pT241	NG NI NH HD NG NH HD NN(7ins19) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH224, pH045, pCTH3	Spec	this study
pT184	NG NI NH HD NG NH HD NN(7ins20) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH178, pH045, pCTH3	Spec	this study
pT251	NG NI NH HD NG NH HD NN(7ins22) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH231, pH045, pCTH3	Spec	this study
pT230	NG NI NH HD NG NH HD NN(7ins24) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH219, pH045, pCTH3	Spec	this study
pT242	NG NI NH HD NG NH HD NN(7ins25) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH225, pH045, pCTH3	Spec	this study
pT232	NG NI NH HD NG NH HD NN(7ins26) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH223, pH045, pCTH3	Spec	this study
pT243	NG NI NH HD NG NH HD NN(7ins28) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH226, pH045, pCTH3	Spec	this study
pT175	NG NI NH HD NG NH HD NN(7ins29) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH174, pH045, pCTH3	Spec	this study
pT231	NG NI NH HD NG NH HD NN(7ins31) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH220, pH045, pCTH3	Spec	this study
pT276	NG NI NH HD NG NH HD NN(1ins5) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH246, pH045, pCTH3	Spec	this study
pT277	NG NI NH HD NG NH HD NN(1ins9) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH247, pH045, pCTH3	Spec	this study
pT253	NG NI NH HD NG NH HD NN(1ins10) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH234, pH045, pCTH3	Spec	this study
pT278	NG NI NH HD NG NH HD NN(1ins13) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH248, pH045, pCTH3	Spec	this study
pT257	NG NI NH HD NG NH HD NN(1ins15) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH238, pH045, pCTH3	Spec	this study
pT279	NG NI NH HD NG NH HD NN(1ins17) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH249, pH045, pCTH3	Spec	this study
pT280	NG NI NH HD NG NH HD NN(1ins18) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH250, pH045, pCTH3	Spec	this study
pT292	NG NI NH HD NG NH HD NN(1ins19) NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH120, pH267, pH045, pCTH3	Spec	this study
pT078	NG NI NH HD NG NH HD link7aa NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH083, pH045, pCHT3	Spec	this study
pT099	NG NI NH HD NG NH HD link18aa NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH106, pH045, pCHT3	Spec	this study
pT123	NG NI NH HD NG NH HD L1/HA/R1 NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH132, pH045, pCTH3	Spec	this study
pT124	NG NI NH HD NG NH HD L1/GFP/R1 NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH133, pH045, pCTH3	Spec	this study
pT167	NG NI NH HD NG NH HD L1/GFP/R1 NI NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH043, pH164, pH123, pH124, pCHT3	Spec	this study
pT166	NG L1/GFP/R1 NI NH HD NG NH HD NI NN NG NI HD NG NI NH HD NI NG in pSKA2; pNTH3, pH162, pH163, pH123, pH124, pCTH3	Spec	this study
pT537	NG NI NH HD NG NH HD NI NN NG NI HD NG L1/GFP/R1 NI NH HD NI NG in pSKA2; pNTH3, pH043, pH055, pH577, pH124, pCTH3	Spec	this study
pT539	NG L1/GFP/R1 NI NH HD NG NH HD L1/GFP/R1 NI NN NG NI HD NG L1/GFP/R1 NI NH HD NI NG in pSKA2; pNTH3, pH162, pH580, pH581, pH582, pH579, pH587, pCTH3	Spec	this study
pT199	NG NI NH HD NG NH HD L1/GFP/R1 NI NN NG NI HD NG NI NH HD NI NG in pSKA4; pNTH3, pH043, pH164, pH123, pH124, pCHT3	Spec	this study

pT538	NG L1/GFP/R1 NI NH HD NG NH HD NI NN NG NI HD NG L1/GFP/R1 NI NH HD NI NG in pSKA2; pNTH3, pH162, pH163, pH371, pH578, pH491, pCTH3	Spec	this study
pGFP	GFP in pSKA2	Spec	Jana Streubel
***Xoo* TALE overexpression constructs (sorted by TALE class and destination vector)**			
pT329	artTal2 (K-74) in pSKA2; pNTH3, pH308, pH309, pH310, pH313, pH312, pCTH3	Spec	this study
pT332	artTal5 (MAI1) in pSKA2; pNTH3, pH315, pH316, pH317, pCTH3	Spec	this study
pT334	artTalC (BAI3) in pSKA2; pNTH3, pH320, pH321, pH322, pH323, pCTH3	Spec	this study
pT331	artAC2 in pSKA2; pNTH3,pH299, pH304, pH324, pH325, pCTH3	Spec	this study
pT330	artAC3 in pSKA2; pNTH3,pH299, pH304, pH301, pH302, pH303, pCTH3	Spec	this study
pT321	artAM6 in pSKA2; pNTH3, pH286, pH287, pH288, pH289, pCTH3	Spec	this study
pT322	artAM6 in pSKX1; pNTMXooAG4, pH286, pH287, pH288, pH289, pCTMXooAO3	Gent	this study
pT328	artBH in pSKA2; pNTH3, pH308, pH309, pH310, pH311, pH312, pCTH3	Spec	this study
pT319	artBK1 in pSKA2; pNTH3, pH280, pH282, pH283, pCTH3	Spec	this study
pT320	artBK1 in pSKX1; pNTMXooAG4, pH280, pH282, pH283, pCTMXooAO3	Gent	this study
pT315	artBK2 in pSKA2; pNTH3, pH280, pH281, pH284, pH285, pCTH3	Spec	this study
pT316	artBK2 in pSKX1; pNTMXooAG4, pH280, pH281, pH284, pH285, pCTMXooAO3	Gent	this study
pT317	artBK3 in pSKA2; pNTH3, pH280, pH281, pH283, pCTH3	Spec	this study
pT318	artBK3 in pSKX1; pNTMXooAG4, pH280, pH281, pH283, pCTMXooAO3	Gent	this study
pT326	artDS in pSKA2; pNTH3, pH299, pH300, pH301, pH302, pH303, pCTH3	Spec	this study
pT327	artDV in pSKA2; pNTH3, pH299, pH304, pH305, pH306, pH307, pCTH3	Spec	this study
***Xoc* TALE overexpression constructs (sorted by TALE class, class version and destination vector)**			
pnTAAD5	TalAD v2 in pSKA2	Spec	Mücke *et al*., 2019
pnTXAD5	TalAD v2 in pSKX1	Gent	Mücke *et al*., 2019
pT489	TalAF v1 in pSKA2; pNTH3, pH508, pH510, pH512, pCTH3	Spec	this study
pT560	TalAF v2 in pSKX1; pNTMAG4, pH508, pH510, pH513, pCTMAO3	Gent	this study
pT491	TalAF v3 in pSKA2; pNTH3, pH509, pH511, pH514, pCTH3	Spec	this study
pSM015	TalAF v3 in pSKX1	Gent	Mücke *et al*., 2019
pT503	TalAJ v1 in pSKA2; pNTH3, pH531, pH532, pH404, pCTH3	Spec	this study
pT519	TalAJ v1 in pSKX1; pNTMAG4, pH531, pH532, pH404, pCTMAO3	Gent	this study
pT484	TalAK v1 in pSKA2; pNTH3, pH498, pH500, pH502, pCTH3	Spec	this study
pT557	TalAK v1 in pSKX1; pNTMAG4, pH498, pH500, pH502, pCTMAO3	Gent	this study
pT485	TalAK v2 in pSKA2; pNTH3, pH499, pH501, pH502, pCTH3	Spec	this study
pT558	TalAK v2 in pSKX1; pNTMAG4, pH499, pH501, pH502, pCTMAO3	Gent	this study
pT403	TalAT v1 in pSKA2; pNTH3, pH407, pH408, pH410, pCTH3	Spec	this study
pT442	TalAT v1 in pSKX1; pNTMAG4, pH407, pH408, pH410, pCTMAO3	Gent	this study
pT404	TalAT v2 in pSKA2; pNTH3, pH407, pH409, pH410, pCTH3	Spec	this study
pT405	TalAT v3 in pSKA2; pNTH3, pH407, pH409, pH411, pCTH3	Spec	this study
pT504	TalAU v1 in pSKA2; pNTH3, pH533, pH534, pH536, pCTH3	Spec	this study
pT520	TalAU v1 in pSKX1; pNTMAG4, pH533, pH534, pH536, pCTMAO3	Gent	this study
pT505	TalAU v2 in pSKA2; pNTH3, pH533, pH535, pH536, pCTH3	Spec	this study
pT521	TalAU v2 in pSKX1; pNTMAG4, pH533, pH535, pH536, pCTMAO3	Gent	this study
pT394	TalAV v2 in pSKA2; pNTH3, pH382, pH383, pH384, pH386, pCTH3	Spec	this study
pT433	TalAV v2 in pSKX1; pNTMAG4, pH382, pH383, pH384, pH386, pCTMAO3	Gent	this study
pT395	TalAX v1 in pSKA2; pNTH3, pH387, pH388, pH389, pCTH3	Spec	this study
pT434	TalAX v1 in pSKX1; pNTMAG4, pH387, pH388, pH389, pCTMAO3	Gent	this study
pT396	TalAX v2 in pSKA2; pNTH3, pH387, pH388, pH390, pCTH3	Spec	this study
pT345	TalAX v2 in pSKX1; pNTMAG4, pH387, pH388, pH390, pCTMAO3	Gent	this study
pT397	TalAY v1 in pSKA2; pNTH3, pH391, pH393, pH394, pCTH3	Spec	this study
pT398	TalAY v2 in pSKA2; pNTH3, pH392, pH393, pH395, pCTH3	Spec	this study
pT437	TalAY v2 in pSKX1; pNTMAG4, pH392, pH393, pH395, pCTMAO3	Gent	this study
pT420	TalAZ v1 in pSKA2; pNTH3, pH446, pH447, pH448, pCTH3	Spec	this study
pT459	TalAZ v1 in pSKX1; pNTMAG4, pH446, pH447, pH448, pCTMAO3	Gent	this study
pT480	TalBB v1 in pSKA2; pNTH3, pH492, pH495, pH496, pCTH3	Spec	this study
pT481	TalBB v2 in pSKA2; pNTH3, pH482, pH495, pH496, pCTH3	Spec	this study
pT482	TalBB v3 in pSKA2; pNTH3, pH493, pH495, pH496, pCTH3	Spec	this study
pT555	TalBB v3 in pSKX1; pNTMAG4, pH493, pH495, pH496, pCTMAO3	Gent	this study
pT474	TalBC v1 in pSKA2; pNTH3, pH479, pH483, pH484, pH487, pH491, pCTH3	Spec	this study
pT475	TalBC v2 in pSKA2; pNTH3, pH479, pH483, pH484, pH488, pCTH3	Spec	this study
pT548	TalBC v2 in pSKX1; pNTMAG4, pH479, pH483, pH484, pH488, pCTMAO3	Gent	this study
pT476	TalBC v3 in pSKA2; pNTH3, pH480, pH483, pH484, pH489, pCTH3	Spec	this study
pT478	TalBC v5 in pSKA2; pNTH3, pH482, pH483, pH486, pH488, pCTH3	Spec	this study
pT479	TalBC v6 in pSKA2; pNTH3, pH482, pH483, pH486, pH490, pCTH3	Spec	this study
pT552	TalBC v6 in pSKX1; pNTMAG4, pH482, pH483, pH486, pH490, pCTMAO3	Gent	this study
pT421	TalBD v1 in pSKA2; pNTH3, pH449, pH450, pH452, pCTH3	Spec	this study
pT460	TalBD v1 in pSKX1; pNTMAG4, pH449, pH450, pH452, pCTMAO3	Gent	this study
pT422	TalBD v2 in pSKA2; pNTH3, pH449, pH451, pH452, pCTH3	Spec	this study
pT461	TalBD v2 in pSKX1; pNTMAG4, pH449, pH451, pH452, pCTMAO3	Gent	this study
pT423	TalBE v1 in pSKA2; pNTH3, pH453, pH454, pH456, pH457, pCTH3	Spec	this study
pT462	TalBE v1 in pSKX1; pNTMAG4, pH453, pH454, pH456, pH457, pCTMAO3	Gent	this study
pT424	TalBE v2 in pSKA2; pH453, pH455, pH456, pH457, pCTH3	Spec	this study
pT486	TalBF v1 in pSKA2; pNTH3, pH503, pH505, pH507, pCTH3	Spec	this study
pT487	TalBF v2 in pSKA2; pNTH3, pH503, pH506, pH507, pCTH3	Spec	this study
pT545	TalBF v2 in pSKX1; pNTMAG4, pH503, pH506, pH507, pCTMAO3	Gent	this study

pT488	TalBF v3 in pSKA2; pNTH3, pH504, pH506, pH507, pCTH3	Spec	this study
pT345	TalBG v1 in pSKA2; pNTH3, pH326, pH327, pH328, pH329, pCTH3	Spec	this study
pT346	TalBG v1 in pSKX1; pNTMAG4, pH326, pH327, pH328, pH329, pCTMAO3	Gent	this study
pT347	TalBG v2 in pSKA2; pNTH3, pH326, pH330, pH328, pH329, pCTH3	Spec	this study
pT425	TalBI v1 in pSKA2; pNTH3, pH458, pH461, pH462, pH465, pCTH3	Spec	this study
pT426	TalBI v2 in pSKA2; pNTH3, pH458, pH461, pH463, pCTH3	Spec	this study
pT465	TalBI v2 in pSKX1; pNTMAG4, pH458, pH461, pH463, pCTMAO3	Gent	this study
pT427	TalBI v3 in pSKA2; pNTH3, pH459, pH461, pH463, pCTH3	Spec	this study
pT428	TalBI v4 in pSKA2; pNTH3, pH460, pH461, pH464, pCTH3	Spec	this study
pT471	TalBL v1 in pSKA2; pNTH3, pH473, pH475, pH476, pCTH3	Spec	this study
pT496	TalBL v1 in pSKX1; pNTMAG4, pH473, pH475, pH476, pCTMAO3	Gent	this study
pT472	TalBL v2 in pSKA2; pNTH3, pH474, pH475, pH477, pCTH3	Spec	this study
pT473	TalBL v3 in pSKA2; pNTH3, pH473, pH530, pH478, pCTH3	Spec	this study
pT401	TalBN v1 in pSKA2; pNTH3, pH401, pH402, pH403, pCTH3	Spec	this study
pT440	TalBN v1 in pSKX1; pNTMAG4, pH401, pH402, pH403, pCTMAO3	Gent	this study
pT402	TalBN v2 in pSKA2; pNTH3, pH401, pH402, pH404, pCTH3	Spec	this study
pT441	TalBN v2 in pSKX1; pNTMAG4, pH401, pH402, pH404, pCTMAO3	Gent	this study
pT430	TalBO v2 in pSKA2; pNTH3, pH467, pH469, pH470, pH471, pCTH3	Spec	this study
pT469	TalBO v2 in pSKX1; pNTMAG4, pH467, pH469, pH470, pH471, pCTMAO3	Gent	this study
pT431	TalBO v3 in pSKA2; pNTH3, pH468, pH469, pH470, pH472, pCTH3	Spec	this study
pT492	TalBR v1 in pSKA2; pNTH3, pH515, pH516, pH517, pH521, pH523, pCTH3	Spec	this study
pT499	TalBR v1 in pSKX1; pNTMAG4, pH515, pH516, pH517, pH521, pH523, pCTMAO3	Gent	this study
pT416	TalBU v1 in pSKA2; pNTH3, pH438, pH441, pH443, pH444, pCTH3	Spec	this study
pT455	TalBU v1 in pSKX1; pNTMAG4, pH438, pH441, pH443, pH444, pCTMAO3	Gent	this study
pT417	TalBU v2 in pSKA2; pNTH3, pH439, pH441, pH443, pH444, pCTH3	Spec	this study
pT418	TalBU v3 in pSKA2; pNTH3, pH440, pH441, pH443, pH444, pCTH3	Spec	this study
pT458	TalBU v4 in pSKX1; pNTMAG4, pH440, pH442, pH443, pH445, pCTMAO3	Gent	this study
pT361	TalBV v1 in pSKA2; pNTH3, pH353, pH354, pH355, pH358, pCTH3	Spec	this study
pT378	TalBV v1 in pSKX1; pNTMAG4, pH353, pH354, pH355, pH358, pCTMAO3	Gent	this study
pT363	TalBV v2 in pSKA2; pNTH3, pH353, pH354, pH357, pH358, pCTH3	Spec	this study
pT362	TalBV v3 in pSKA2; pNTH3, pH353, pH354, pH356, pH358, pCTH3	Spec	this study
pT364	TalBV v4 in pSKA2; pNTH3, pH353, pH354, pH356, pH359, pCTH3	Spec	this study
pT413	TalBW v1 in pSKA2; pNTH3, pH431, pH433, pH435, pH436, pH437, pCTH3	Spec	this study
pT452	TalBW v1 in pSKX1; pNTMAG4, pH431, pH433, pH435, pH436, pH437, pCTMAO3	Gent	this study
pT414	TalBW v2 in pSKA2; pNTH3, pH431, pH434, pH435, pH436, pH437, pCTH3	Spec	this study
pT453	TalBW v2 in pSKX1; pNTMAG4, pH431, pH434, pH435, pH436, pH437, pCTMAO3	Gent	this study
pT415	TalBW v3 in pSKA2; pNTH3, pH432, pH434, pH435, pH436, pH437, pCTH3	Spec	this study
pT406	TalBY v1 in pSKA2; pNTH3, pH413, pH415, pH416, pH417, pCTH3	Spec	this study
pT445	TalBY v1 in pSKX1; pNTMAG4, pH413, pH415, pH416, pH417, pCTMAO3	Gent	this study
pT407	TalBY v2 in pSKA2; pNTH3, pH414, pH415, pH416, pH417, pCTH3	Spec	this study
pT382	TalCB v2 in pSKA2; pNTH3, pH360, pH354, pH361, pH362, pCTH3	Spec	this study
pT383	TalCB v2 in pSKX1; pNTMAG4, pH360, pH354, pH361, pH362, pCTMAO3	Gent	this study
pT506	TalCD v1 in pSKA2; pNTH3, pH537, pH539, pH540, pCTH3	Spec	this study
pT508	TalCE v1 in pSKA2; pNTH3, pH542, pH543, pH545, pCTH3	Spec	this study
pT524	TalCE v1 in pSKX1; pNTMAG4, pH542, pH543, pH545, pCTMAO3	Gent	this study
pT509	TalCE v2 in pSKA2; pNTH3, pH542, pH544, pH545, pCTH3	Spec	this study
pT540	TalCF v1 in pSKA2; pNTH3, pH583, pH585, pH404, pCTH3	Spec	this study
pT541	TalCF v2 in pSKA2; pNTH3, pH584, pH586, pH404, pCTH3	Spec	this study
pT543	TalCF v2 in pSKX1; pNTMAG4, pH584, pH586, pH404, pCTMAO3	Gent	this study
pT399	TalCG v1 in pSKA2; pNTH3, pH396, pH398, pH399, pCTH3	Spec	this study
pT438	TalCG v1 in pSKX1; pNTMAG4, pH396, pH398, pH399, pCTMAO3	Gent	this study
pT400	TalCG v2 in pSKA2; pNTH3, pH397, pH398, pH400, pCTH3	Spec	this study
pT439	TalCG v2 in pSKX1; pNTMAG4, pH397, pH398, pH400, pCTMAO3	Gent	this study
pT408	TalCH v1 in pSKA2; pNTH3, pH418, pH419, pH420, pH421, pH422, pCTH3	Spec	this study
pT409	TalCH v2 in pSKA2; pNTH3, pH418, pH419, pH420, pH421, pH423, pCTH3	Spec	this study
pT448	TalCH v2 in pSKX1; pNTMAG4, pH418, pH419, pH420, pH421, pH423, pCTMAO3	Gent	this study
pT510	TalCI v1 in pSKA2; pNTH3, pH418, pH546, pH547, pH549, pCTH3	Spec	this study
pT526	TalCI v1 in pSKX1; pNTMAG4, pH418, pH546, pH547, pH549, pCTMAO3	Gent	this study
pT511	TalCI v2 in pSKA2; pNTH3, pH418, pH546, pH548, pH550, pH551, pCTH3	Spec	this study
pT528	TalCJ v1 in pSKX1; pNTMAG4, pH552, pH553, pH554, pH555, pH430, pCTMAO3	Gent	this study
pT513	TalCK v1 in pSKA2; pNTH3, pH556, pH557, pH559, pCTH3	Spec	this study
pT529	TalCK v1 in pSKX1; pNTMAG4, pH556, pH557, pH559, pCTMAO3	Gent	this study
pT514	TalCK v2 in pSKA2; pNTH3, pH556, pH558, pH560, pCTH3	Spec	this study
pT530	TalCK v2 in pSKX1; pNTMAG4, pH556, pH558, pH560, pCTMAO3	Gent	this study
pT410	TalCL v1 in pSKA2; pNTH3, pH424, pH426, pH427, pH428, pH429, pCTH3	Spec	this study
pT449	TalCL v1 in pSKX1; pNTMAG4, pH424, pH426, pH427, pH428, pH429, pCTMAO3	Gent	this study
pT411	TalCL v2 in pSKA2; pNTH3, pH425, pH426, pH427, pH428, pH429, pCTH3	Spec	this study

pT412	TalCL v3 in pSKA2; pNTH3, pH425, pH426, pH427, pH428, pH430, pCTH3	Spec	this study
pT451	TalCL v3 in pSKX1; pNTMAG4, pH425, pH426, pH427, pH428, pH430, pCTMAO3	Gent	this study
pT515	TalCP v1 in pSKA2; pNTH3, pH561, pH562, pCTH3	Spec	this study
pT531	TalCP v1 in pSKX1; pNTMAG4, pH561, pH562, pCTMAO3	Gent	this study
pT516	TalCQ v1 in pSKA2; pNTH3, pH563, pH564, pH565, pCTH3	Spec	this study
pT532	TalCQ v1 in pSKX1; pNTMAG4, pH563, pH564, pH565, pCTMAO3	Gent	this study
pT517	TalCR v1 in pSKA2; pNTH3, pH566, pH567, pH404, pCTH3	Spec	this study
pT533	TalCR v1 in pSKX1; pNTMAG4, pH566, pH567, pH404, pCTMAO3	Gent	this study
pT518	TalCS v1 in pSKA2; pNTH3, pH482, pH483, pH568, pH569, pCTH3	Spec	this study
pT534	TalCS v1 in pSKX1; pNTMAG4, pH482, pH483, pH568, pH569, pCTMAO3	Gent	this study
reporter constructs for the analysis of aberrant repeat-carrying TALEs			
pR008	“op” box (TTAGCTGCAGTACTAGCAT)	Kan	this study
pR010	“op -1 p4” box (TTAGTGCAGTACTAGCAT)	Kan	this study
pR021	“op -1 p6” box (TTAGCTCAGTACTAGCAT)	Kan	this study
pR012	“op -1 p7” box (TTAGCTGAGTACTAGCAT)	Kan	this study
pR013	“op -1 p8” box (TTAGCTGCGTACTAGCAT)	Kan	this study
pR014	“op -1 p9” box (TTAGCTGCATACTAGCAT)	Kan	this study
pR049	“op -1 p12” box (TTAGCTGCAGTATAGCAT)	Kan	this study
pR050	“op -2 p4/8” box (TTAGTGCGTACTAGCAT)	Kan	this study
pR051	“op -2 p4/12” box (TTAGTGCAGTATAGCAT)	Kan	this study
pR023	“op -2 p6/7” box (TTAGCTAGTACTAGCAT)	Kan	this study
pR027	“op -2 p7/8” box (TTAGCTGGTACTAGCAT)	Kan	this study
pR029	“op -2 p8/9” box (TTAGCTGCTACTAGCAT)	Kan	this study
pR052	“op -2 p8/12” box (TTAGCTGCGTATAGCAT)	Kan	this study
pR053	“op -3 p4/8/12” box (TTAGTGCGTATAGCAT)	Kan	this study
pR030	“op -3 p7/8/9” box (TTAGCTGTACTAGCAT)	Kan	this study
pR038	“midG6 op” box (TTAGCTGGGGGGGTAGCAT)	Kan	this study
pR039	“midG6 op -1” box (TTAGCTGGGGGGTAGCAT)	Kan	this study
pR040	“midG6 op -2” box (TTAGCTGGGGGTAGCAT)	Kan	this study
pR045	“midG6 op -3” box (TTAGCTGGGGTAGCAT)	Kan	this study
pR046	“midG6 op -4” box (TTAGCTGGGTAGCAT)	Kan	this study
pR047	“midG6 op -5” box (TTAGCTGGTAGCAT)	Kan	this study
pR048	“midG6 op -6” box (TTAGCTGTAGCAT)	Kan	this study
pR072	“BK op” box (TATAGAGCACCAACAACTCCCTT)	Kan	this study
pR073	“BK op -1 p9” box (TATAGAGCACAACAACTCCCTT)	Kan	this study
pR074	“BK op -1 p12” box (TATAGAGCACCACAACTCCCTT)	Kan	this study
pR075	“BK op -2 p9/12” box (TATAGAGCACACAACTCCCTT)	Kan	this study
pR076	“BK op -4 p9-12” box (TATAGAGCACAACTCCCTT)	Kan	this study
pR056	“mm #1” box (TTACGAGCAGTACTAGCAT)	Kan	this study
pR057	“mm #2” box (TTAGCTGCAGATGTAGCAT)	Kan	this study
pR058	“mm #3” box (TTAGCTGCAGTACTTCGAT)	Kan	this study
pR077	1 kbp promoter fragment of *OsSWEET13* (Zhenshan)	Kan	this study
pR078	1 kbp promoter fragment of *OsSWEET13* (Nipponbare)	Kan	this study
pR079	1 kbp promoter fragment of *OsSWEET13* (Sadu Cho)	Kan	this study
pR080	1 kbp promoter fragment of *OsSWEET13* (IR24	Kan	this study
pOsSWEET14	1 kbp promoter fragment of *OsSWEET14*	Kan	Jana Streubel
pR095	1 kbp promoter fragment of *OsSWEET14 xa41(t)*	Kan	this study
reporter constructs for the analysis of known or putative TALE target genes (sorted by locus ID)			
pR202	1 kbp promoter fragment of *Os01g04190*	Kan	this study
pR474	1 kbp promoter fragment of *Os01g09470*	Kan	this study
pR147	1 kbp promoter fragment of *Os01g19870*	Kan	this study
pR424	1 kbp promoter fragment of *Os01g24710*	Kan	this study
pR318	1 kbp promoter fragment of *Os01g32439*	Kan	this study
pR336	1 kbp promoter fragment of *Os01g40290*	Kan	this study
pR123	1 kbp promoter fragment of *Os01g51040*	Kan	this study
pR396	1 kbp promoter fragment of *Os01g53920*	Kan	this study
pR204	1 kbp promoter fragment of *Os01g55600*	Kan	this study
pR304	1 kbp promoter fragment of *Os01g58330*	Kan	this study
pR420	1 kbp promoter fragment of *Os01g65590*	Kan	this study
pR190	1 kbp promoter fragment of *Os02g02190*	Kan	this study
pR131	1 kbp promoter fragment of *Os02g02620*	Kan	this study
pR302	1 kbp promoter fragment of *Os02g14720*	Kan	this study
pR416	1 kbp promoter fragment of *Os02g29230*	Kan	this study
pR232	1 kbp promoter fragment of *Os02g34970*	Kan	this study
pR344	1 kbp promoter fragment of *Os02g36850*	Kan	this study
pR428	1 kbp promoter fragment of *Os02g38386*	Kan	this study
pR220	1 kbp promoter fragment of *Os02g43760*	Kan	this study
pR226	1 kbp promoter fragment of *Os02g47660*	Kan	this study
pR216	1 kbp promoter fragment of *Os02g48570*	Kan	this study
pR442	1 kbp promoter fragment of *Os02g49710*	Kan	this study
pSM052	1 kbp promoter fragment of *Os02g51110*	Kan	Mücke *et al.*, 2019
pR296	1 kbp promoter fragment of *Os02g57860*	Kan	this study

pFNSI-1	1 kbp promoter fragment of *Os03g03034*	Kan	Claudia Schwietzer
pR472	1 kbp promoter fragment of *Os03g03350*	Kan	this study
pR228	1 kbp promoter fragment of *Os03g07540*	Kan	this study
pR276	1 kbp promoter fragment of *Os03g11050*	Kan	this study
pR456	1 kbp promoter fragment of *Os03g11980*	Kan	this study
pR222	1 kbp promoter fragment of *Os03g37840*	Kan	this study
pR484	1 kbp promoter fragment of *Os03g40100*	Kan	this study
pR386	1 kbp promoter fragment of *Os03g42130*	Kan	this study
pR380	1 kbp promoter fragment of *Os03g59790*	Kan	this study
pR214	1 kbp promoter fragment of *Os03g60720*	Kan	this study
pR392	1 kbp promoter fragment of *Os03g62010*	Kan	this study
pR482	1 kbp promoter fragment of *Os04g18830*	Kan	this study
pR412	1 kbp promoter fragment of *Os04g27060*	Kan	this study
pR422	1 kbp promoter fragment of *Os04g45730*	Kan	this study
pR121	1 kbp promoter fragment of *Os04g47890*	Kan	this study
pR402	1 kbp promoter fragment of *Os04g48160*	Kan	this study
pR230	1 kbp promoter fragment of *Os04g49194*	Kan	this study
pR286	1 kbp promoter fragment of *Os04g53210*	Kan	this study
pR127	1 kbp promoter fragment of *Os04g58560*	Kan	this study
pR234	1 kbp promoter fragment of *Os05g27590*	Kan	this study
pR446	1 kbp promoter fragment of *Os05g28800*	Kan	this study
pR139	1 kbp promoter fragment of *Os05g33390*	Kan	this study
pR486	1 kbp promoter fragment of *Os05g38320*	Kan	this study
pR210	1 kbp promoter fragment of *Os05g42150*	Kan	this study
pR450	1 kbp promoter fragment of *Os05g51690*	Kan	this study
pR358	1 kbp promoter fragment of *Os06g02130*	Kan	this study
pR194	1 kbp promoter fragment of *Os06g07260*	Kan	this study
pR362	1 kbp promoter fragment of *Os06g10670*	Kan	this study
pR414	1 kbp promoter fragment of *Os06g12410*	Kan	this study
pR306	1 kbp promoter fragment of *Os06g12530*	Kan	this study
pR320	1 kbp promoter fragment of *Os06g17020*	Kan	this study
pR342	1 kbp promoter fragment of *Os06g17910*	Kan	this study
pR400	1 kbp promoter fragment of *Os06g36270*	Kan	this study
pR224	1 kbp promoter fragment of *Os06g37080*	Kan	this study
pR470	1 kbp promoter fragment of *Os06g43650*	Kan	this study
pR348	1 kbp promoter fragment of *Os07g01490*	Kan	this study
pR125	1 kbp promoter fragment of *Os07g05720*	Kan	this study
pR384	1 kbp promoter fragment of *Os07g09470*	Kan	this study
pR141	1 kbp promoter fragment of *Os07g17280*	Kan	this study
pR376	1 kbp promoter fragment of *Os07g20290*	Kan	this study
pR430	1 kbp promoter fragment of *Os07g23470*	Kan	this study
pR260	1 kbp promoter fragment of *Os07g26520*	Kan	this study
pR466	1 kbp promoter fragment of *Os07g28990*	Kan	this study
pR133	1 kbp promoter fragment of *Os07g29310*	Kan	this study
pR390	1 kbp promoter fragment of *Os07g33740*	Kan	this study
pR262	1 kbp promoter fragment of *Os07g36430*	Kan	this study
pR200	1 kbp promoter fragment of *Os07g39280*	Kan	this study
pR438	1 kbp promoter fragment of *Os07g39590*	Kan	this study
pR334	1 kbp promoter fragment of *Os07g44890*	Kan	this study
pR135	1 kbp promoter fragment of *Os07g47790*	Kan	this study
pR278	1 kbp promoter fragment of *Os08g01260*	Kan	this study
pR454	1 kbp promoter fragment of *Os08g14940*	Kan	this study
pR378	1 kbp promoter fragment of *Os08g33120*	Kan	this study
pR458	1 kbp promoter fragment of *Os08g37432*	Kan	this study
pR398	1 kbp promoter fragment of *Os08g41950*	Kan	this study
pR292	1 kbp promoter fragment of *Os09g25200*	Kan	this study
pR294	1 kbp promoter fragment of *Os09g27090*	Kan	this study
pR186	1 kbp promoter fragment of *Os09g29100*	Kan	this study
pR290	1 kbp promoter fragment of *Os09g29130*	Kan	this study
pR368	1 kbp promoter fragment of *Os09g30120*	Kan	this study
pR270	1 kbp promoter fragment of *Os09g32100*	Kan	this study
pR196	1 kbp promoter fragment of *Os09g37350*	Kan	this study
pR198	1 kbp promoter fragment of *Os09g37470*	Kan	this study
pR366	1 kbp promoter fragment of *Os10g05910*	Kan	this study
pR284	1 kbp promoter fragment of *Os10g13820*	Kan	this study
pR137	1 kbp promoter fragment of *Os10g25550*	Kan	this study
pR436	1 kbp promoter fragment of *Os10g31380*	Kan	this study
pR448	1 kbp promoter fragment of *Os10g31640*	Kan	this study
pR212	1 kbp promoter fragment of *Os10g39840*	Kan	this study
pR308	1 kbp promoter fragment of *Os11g02440*	Kan	this study
pR188	1 kbp promoter fragment of *Os11g04130*	Kan	this study
pR356	1 kbp promoter fragment of *Os11g13440*	Kan	this study
pR406	1 kbp promoter fragment of *Os11g32650*	Kan	this study

pR346	1 kbp promoter fragment of *Os11g39480*	Kan	this study
pR350	1 kbp promoter fragment of *Os11g44170*	Kan	this study
pR145	1 kbp promoter fragment of *Os11g46070*	Kan	this study
pR322	1 kbp promoter fragment of *Os11g47350*	Kan	this study
pR452	1 kbp promoter fragment of *Os12g10080*	Kan	this study
pR476	1 kbp promoter fragment of *Os12g13810*	Kan	this study
pR352	1 kbp promoter fragment of *Os12g17340*	Kan	this study
pR298	1 kbp promoter fragment of *Os12g24320*	Kan	this study
pR364	1 kbp promoter fragment of *Os12g42040*	Kan	this study
pR206	1 kbp promoter fragment of *Os12g42200*	Kan	this study
pR143	1 kbp promoter fragment of *Os12g42970*	Kan	this study

For the assembly of TALEs, single repeat modules and assembly vectors from the "Golden TAL Technology" cloning kit were used (Geißler *et al*., 2011) as well as some of the new assembly vectors published in Mücke *et al*., 2019.

2.1.7 Buffers

Table 2.6: Used buffers and solutions.

buffer	components
agarose	1% (w/v) agarose; dissolved in 1x TAE buffer; 0.005% (v/v) Midori Green
blocking solution	5% (w/v) skin milk powder; 3% (w/v) BSA fraction V; dissolved in 1x TBST
towbin buffer	48 mM Tris; 39 mM glycine; 20% (v/v) methanol; 0.04% (w/v) SDS
ECL solution	100 mM Tris-HCl (pH8.5); 1.25 mM luminol (dissolved in DMSO); 0.225 mM p-coumaric acid (dissolved in DMSO); 0.3% (v/v) H_2O_2
GUS extraction buffer	50 mM sodium phosphate (pH 7.0); 10 mM EDTA (pH 8.0); 0.1% (w/v) SDS; 0.1% (v/v) Triton X-100; 10 mM β- mercaptoethanol
GUS staining solution	10 mM sodium phosphate (pH 7.0); 10 mM EDTA (pH 8.0); 1 mM potassium ferricyanide; 1 mM potassium ferrocyanide; 0.1% (v/v) Triton X-100; 0.1% (w/v) X-Gluc
Laemmli (4x)	250 mM Tris-HCl (pH 6.8); 8% (w/v) SDS; 40% (v/v) glycerol; 10% (v/v) β-Mercaptoenthanol; bromophenol blue
loading dye (5x)	15% Ficoll 400; orange G
separation gel (8%)	8% acrylamide (27% (v/v) from 30% acrylamide-bisacrylamide solution); 375 mM Tris-HCl (pH 8.8); 0.1% (w/v) SDS; 0.1% (v/v) TEMED; 0.06% (w/v) APS
stacking gel (5%)	5% acrylamide (17% (v/v) from 30% acrylamide-bisacrylamide solution); 130 mM Tris-HCl (pH 6.8); 0.1% (w/v) SDS; 0.1% (w/v) APS; 0.1% (v/v) TEMED
TAE (1x)	40 mM Tris; 20 mM acetic acid; 1 mM EDTA; pH 8.0
TANK (1x)	25 mM Tris; 250 mM glycine ; 0.1% (w/v) SDS
TBST (1x)	50 mM Tris-HCl (pH 8.0); 150 mM NaCl; 0.05% (v/v) Tween20

2.2 Methods

2.2.1 Chemical transformation of bacterial strains

Chemical competent cells (*E. coli,* ~50 µl) were mixed either with an assembly reaction (10-20 µl) or with plasmid DNA (1 µl) and incubated for ~20 minutes on ice. Afterwards, the mixture was heat-shocked for 40 seconds at 42°C using the Eppendorf ThermoMixer C Dry Block Heating and Cooling Shaker (Eppendorf). To allow for regeneration, 250 µl liquid LB media was added and the reaction was incubated for 1 h at 37°C. Finally, the reaction was plated on solid LB media containing the antibiotics necessary for the selection of positive colonies and incubated 12-24 h at 37°C.

2.2.2 Transformation of bacterial strains via electroporation

Electrocompetent cells (*A. tumefaciens,* ~50 µl; *X. oryzae,* ~100 µl) were mixed with 1-2 µl plasmid DNA and incubated for 5 minutes on ice. The mixture was transferred into a pre-

cooled electroporation cuvette (gap width 1 mm) and subjected to an electric impulse using the MicroPulser Electroporator (Bio-Rad Laboratories) on program Ec2 (voltage: 2500 V; capacity: 25 mF; resistance: 200 Ω; duration: 5 ms). To allow for regeneration, 250 µl media was added (YEB for *A. tumefaciens*, PSA for *X. oryzae*) and the reactions were incubated for 1-2 h at 28°C. Finally, the reaction was plated on solid media (YEB for *A. tumefaciens*, PSA for *X. oryzae*) containing the antibiotics necessary for the selection of positive colonies and incubated 1-2 (*A. tumefaciens*) or 3-4 (*X. oryzae*) days at 28°C.

2.2.3 Introduction of plasmids into *Xanthomonas* strain BAI3 Δ*talC*

TALE-expressing constructs were introduced into the *Xanthomonas* strain BAI3 Δ*talC* by conjugation using pRK2013 as a helper plasmid in triparental mating.

2.2.4 Storage of bacteria

To store bacteria, bacterial samples were mixed with media and glycerine to a final concentration of ~15% glycerine and stored at -80°C.

2.2.5 Isolation of plasmid DNA

The isolation of plasmid DNA was performed according to the manufacturer's specifications using the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific). A liquid culture of 5 ml was used for each preparation. The elution was usually done using 50 µl dH_2O.

2.2.6 Restriction digest

To analyse plasmid DNA, restriction digests were performed. Plasmid DNA (5-10 µl), restriction enzyme (0.5 µl) and a matching buffer (2 µl) were combined and water was used to adjust the total volume to 20 µl. Reactions were incubated for ~1 h at a temperature suitable for the used enzyme. Restriction enzymes were acquired from either Thermo Fisher Scientific or New England Biolabs.

2.2.7 Agarose gel electrophoresis

To evaluate the size of DNA fragments, agarose gel electrophoresis was performed. Agarose was dissolved in 1x TAE buffer and stained with ethidium bromide or Midori Green to visualize double stranded DNA. DNA samples were mixed with 5x loading dye prior to the analysis. Fragments were separated in 1x TAE buffer at 80-140 V using the Sub-Cell Modell 96 or 192 Cell (Bio-Rad Laboratories). The results were documented using the ChemiDoc Imaging System (Bio-Rad Laboratories). To assess the size of the DNA fragments, a DNA ladder was included on every agarose gel, most commonly, the 1 kb Plus DNA Ladder (Thermo Fischer Scientific) and the 2-Log DNA Ladder (New England Biolabs).

2.2.8 Gel extraction

The extraction of DNA fragments from an agarose gel was performed according to the manufacturer's specifications using the GeneJET Gel Extraction Kit (Thermo Fisher Scientific). Cut gel fragments were either brought into a vector for further cloning or directly used for sequencing. The elution was usually done using 20 µl dH_2O.

2.2.9 Sequencing of DNA fragments

Samples chosen for sequencing were sent to Microsynth Seqlab. For this 800-1200 ng DNA were combined with 3 µl of a primer and adjusted to a final volume of 12 µl.

2.2.10 Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) was used to amplify DNA fragments and to make alterations in a target sequence e.g. to add flanking recognition sites for type IIS restriction enzymes, thus allowing for its use in Golden Gate Cloning. To avoid unwanted nucleotide changes, amplifications were done using a proof-reading polymerase like the S7 Fusion (Thermo Fisher Scientific).

Reaction setup for a standard PCR using a proof reading polymerase:

component	volume
5x buffer (HF or GC)	10 µl
dNTPs (10 mM)	5 µl
primer F (10 µM)	2,5 µl
primer R (10 µM)	2,5 µl
template DNA (<10 ng)	1 µl
phusion DNA polymerase	0,5 µl
water	28,5 µl

Thermocycling protocol:

step		temperature	time	
1)	initial denaturation	98°C	30 seconds	
2a)	denaturation	98°C	10 seconds	35x
2b)	annealing	xx°C	10 seconds	
2c)	extension	72°C	yy seconds	
3)	final extension	72°C	10 minutes	
4)	hold	12°C		

xx°C = most commonly, the lowest melting temperature of the two primers was used as annealing temperature in the reaction

yy s = 15 seconds for 1 kbp on plasmid DNA
30 seconds for 1 kbp on genomic DNA

2.2.11 Blunt end cloning into pUC57

The blunt end cloning of fragments into pUC57 was used to generate new modules for Golden Gate Cloning. Flanking *Bpi*I or *Bsa*I recognition sites were added to a target sequence via PCR. After separating the PCR product via gel electrophoresis, the correct fragment was excised, purified and the elution was used for the cloning reaction. If the fragment is successfully inserted into the multiple cloning site (MCS) of the vector, its *lacZ* gene is disrupted (Fig. 2.1). Successful insertions can be selected for by their white colour using X-gal. Blunt end cloning reactions were performed using either *Sma*I or *EcoR*V, depending on the presence of internal restriction sites in the target fragment.

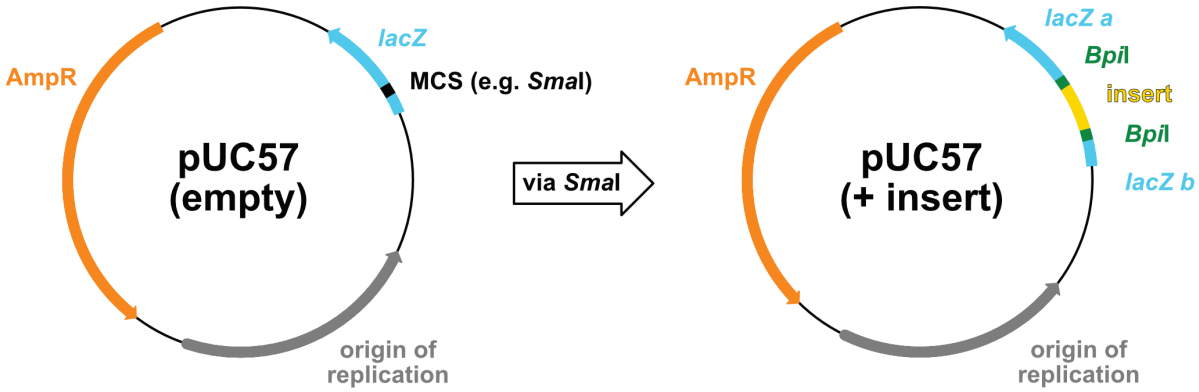


Figure 2.1: Blunt end cloning into pUC57. A PCR-amplified and purified DNA fragment is combined with the vector pUC57, a blunt end cutting enzyme and T4 DNA ligase, and thus integrated into the MCS of the vector. By having recognition sites for a type IIS restriction enzyme like *Bpi*I added in the initial PCR step, the resulting vector is suited for Golden Gate Cloning.

Reaction setup for a blunt end cut ligation into pUC57:

component	volume
eluted PCR fragment	10 µl
vector (pUC57) (50 ng/µl)	1 µl
ATP (10 mM)	2 µl
10x reaction buffer (e.g. Cut Smart)	2 µl
restriction enzyme (e.g. *Sma*I)	1 µl
T4 DNA Ligase (5u)	1 µl
water	3 µl

Thermocycling protocol:

step		temperature	time	
1a)	restriction	zz°C	10 minutes	} 10x
1b)	ligation	16°C	10 minutes	}
2)	final restriction	zz°C	20 minutes	
3)	inactivation	80°C	10 minutes	
4)	hold	12°C		

zz°C = 25°C for *Sma*I
37°C for *EcoR*V

2.2.12 Generation of single repeat modules of aberrant length

Aberrant repeat modules were generated by introducing the desired amino acids via PCR. For this, two oligonucleotides were placed next to each other, oriented in opposite directions. Both perfectly matched to the sequence of a 34aa standard repeat but one contained in its 5' region also the nucleotides encoding for the desired amino acid insertion. The oligonucleotides will result in the amplification of the complete plasmid, including the additional amino acids at one end. However, to yield a viable vector, the two ends of the fragment need to be ligated. Since the T4 DNA ligase requires phosphate groups at the free 5' ends of the fragment, those had to be added prior to a ligase reaction. For this, 8 µl of an eluted PCR fragment were combined with 1 µl polynucleotide-kinase (T4 PNK; New England Biolabs) and 1 µl PNK buffer and the mixture was incubated for 1 h at 37°C. After this phosphorylation, 1 µl T4 DNA ligase, 2 µl ligase buffer and 7 µl water were added to the reaction and the mixture was incubated for 1-2 h at room temperature, thereby ligating the fragments back into a circular and viable plasmid. In the next step, the reaction was treated with *Dpn*I. This endonuclease targets only methylated DNA, and thus the vector DNA that was used as template for the PCR is cleaved while the mutated PCR-derived version of the vector remains unaffected. Finally, the reaction was transformed into *E. coli*, single colonies were analysed for correct fragment patterns via restriction digest and the results were confirmed via sequencing of the modified region. To change the RVD and/or the position of the obtained repeat module, the necessary alterations were introduced via PCR as well.

2.2.13 Golden Gate Cloning

Golden Gate Cloning (GGC) is based on type IIS restriction enzymes and their ability to cut outside their own recognition site, thereby creating 4 bp overhangs (Engler *et al*., 2008). Since these overhangs can be specifically designed, it allows for the efficient and scar-free combination of multiple DNA fragments in any desired order. Moreover, since the final cloning products do not contain recognition sites for the used type IIS enzyme anymore, GGC allows the simultaneous use of restriction enzyme and T4 DNA ligase within the same reaction, a so called cut ligation.

2.2.13.1 Generation of artificial TALEs

Artificial TALEs were generated using the "Golden TAL Technology" cloning kit (Geißler *et al*., 2011). Artificial TALEs were assembled in two subsequent steps. The first step encompasses the assembly of multi repeat modules by combining up to six single repeat modules with selected RVDs in an assembly vector (Fig. 2.2). If less than six repeats are needed, special stop repeat modules can be used to connect an earlier repeat position with the vector.

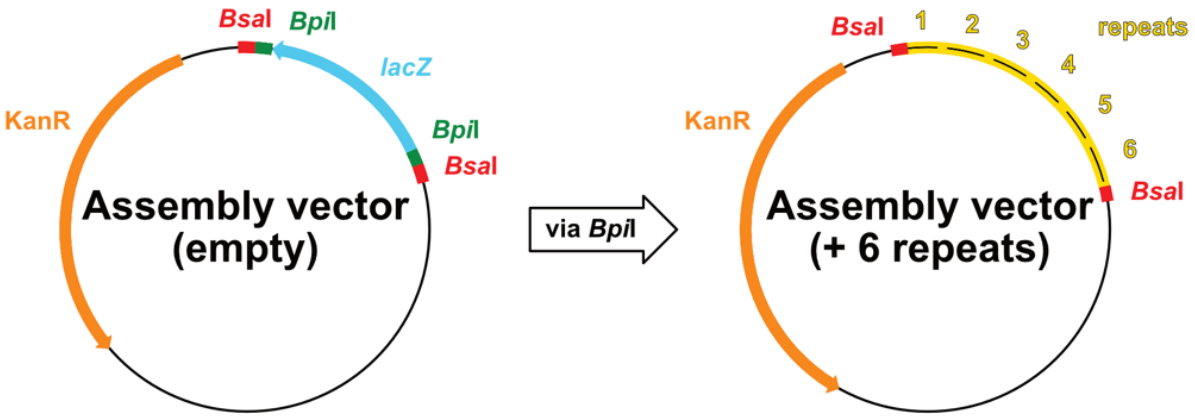


Figure 2.2: Generation of multi repeat modules. Up to six single repeat modules can combined in an assembly vector.

Reaction setup for the assembly of a multi-repeat module:

component	volume
10x buffer (buffer green)	2 µl
ATP (10 mM)	2 µl
*Bpi*I	1 µl
T4 DNA ligase (5u)	1 µl
assembly vector (50 ng/µl)	1 µl
single repeat module position 1 (50 ng/µl)	1 µl
single repeat module position 2 (50 ng/µl)	1 µl
single repeat module position 3 (50 ng/µl)	1 µl
single repeat module position 4 (50 ng/µl)	1 µl
single repeat module position 5 (50 ng/µl)	1 µl
single repeat module position 6 (50 ng/µl)	1 µl
water	7 µl

Thermocycling protocol:

step		temperature	time	
1a)	restriction	37°C	10 minutes	10x
1b)	ligation	16°C	10 minutes	
2)	final restriction	37°C	20 minutes	
3)	inactivation	80°C	10 minutes	
4)	hold	12°C		

In the second assembly step, up to six multi-repeat modules are combined with the N- and C-terminal regions of a TALE and a destination vector. Depending on the planned application, the N- and C-terminal regions as well as the final vector differ. For later analysis in *N. benthamiana*, the multi repeat modules are combined with the N- and C-terminal regions of Hax3 (Kay *et al*., 2005) and cloned into the vector pSKA2. For later analysis in rice, they are combined with the N-terminal region of TalAG3 and the C-terminal region of TalAO3 (Mücke *et al*., 2019) and assembled in the vector pSKX1 (Fig. 2.3).

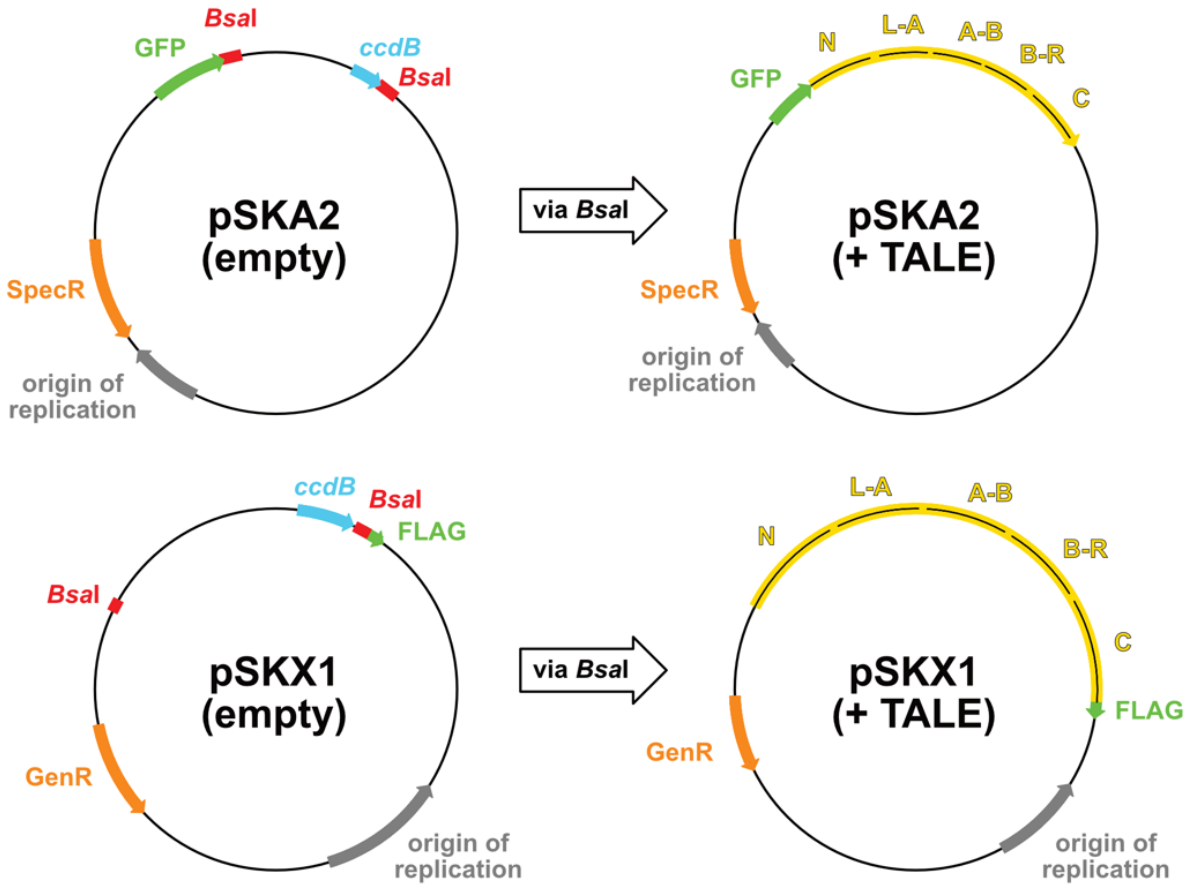


Figure 2.3: Cloning of TALE-expressing constructs. TALE-expressing constructs are generated by combining several multi repeat modules either with the N- and C-terminal regions of Hax3 in pSKA2 or with the N-terminal region of TalAG4 and the C-terminal region of TalAO3 in pSKX1. In this example, three-multi repeats modules are used (L-A, A-B and B-R), yielding a TALE with a total of 17.5 repeats. The constructs are then transformed into *A. tumefaciens* (pSKA2) or *X. oryza* (pSKX1) and are used for analysis in *N. benthamiana* or rice, respectively.

Reaction setup for the assembly of a TALE-expressing construct with 17.5 repeats:

component	volume
10x buffer (Cut Smart)	2 µl
ATP (10 mM)	2 µl
*Bsa*I	1 µl
T4 DNA ligase (5u)	1 µl
destination vector [*1] (50 ng/µl)	1 µl
N-terminal region of aTALE [*2] (50 ng/µl)	1 µl
C-terminal region of a TALE [*3] (50 ng/µl)	1 µl
multi repeat module L-A (50 ng/µl)	1 µl
multi repeat module A-B (50 ng/µl)	1 µl
multi repeat module B-R (50 ng/µl)	1 µl
water	8 µl

[*1] pSKA2 or pSKX1
[*2] N-terminal region of Hax3 or TalAG4
[*3] C-terminal region of Hax3 or TalAO3

The thermocycling protocol is identical to the one for the assembly of multi repeat modules.

2.2.13.2 Generation of GUS reporter constructs

GUS reporter constructs were generated by cloning either a TALE box and the *Bs4* minimal promoter or a 1000 bp promoter fragment (starting from the first ATG of the CDS) into the vector pGWB3GG, thus placing the insert in front of the *GUS* reporter gene (Fig. 2.4). To do so, flanking *Bsa*I sites/matching overhangs were added to the inserts via PCR. If not otherwise indicated, promoter fragments were amplified using genomic DNA from the rice cv. Nipponbare as template.

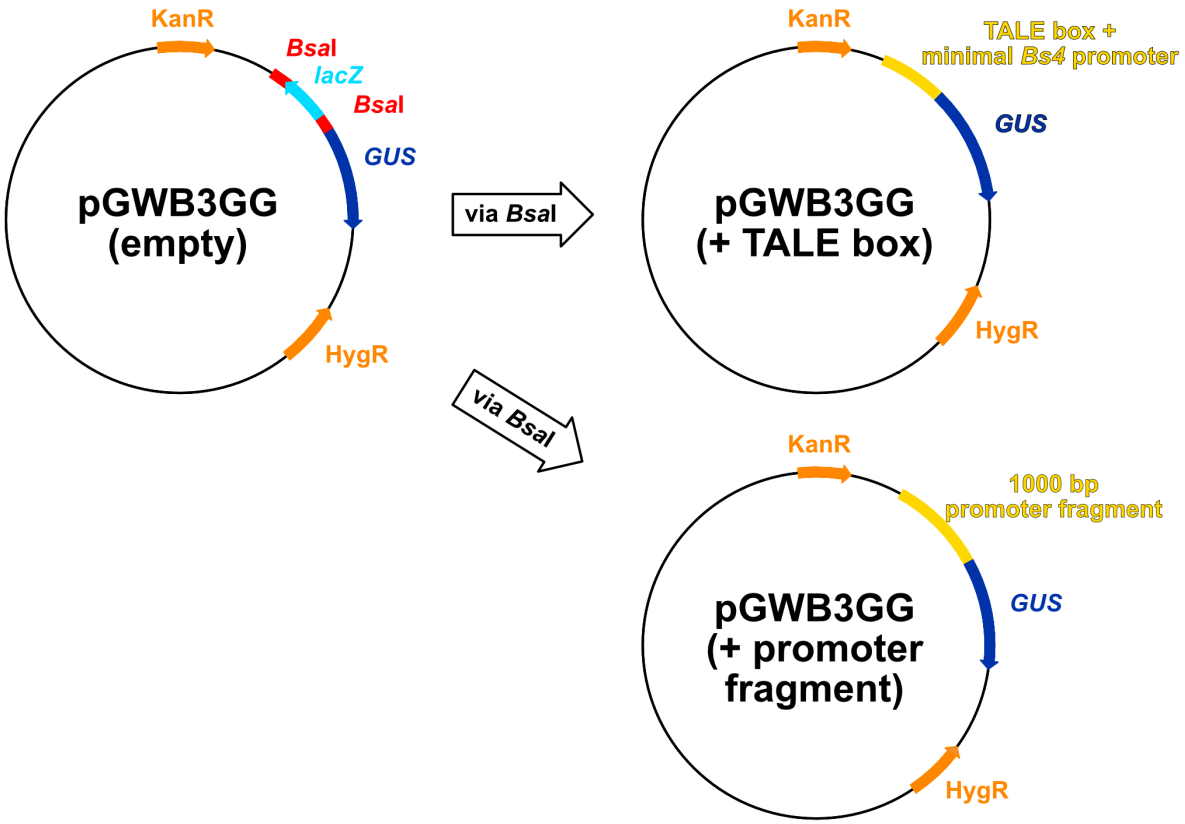


Figure 2.4: Cloning of GUS reporter constructs into pGWB3GG. In front of the *GUS* reporter gene in the vector pGWB3GG is placed either **(A)** a fusion between a TALE box and the *Bs4* minimal promoter or **(B)** a 1000 bp promoter fragment from a natural rice gene (beginning from the first ATG of the CDS).

Reaction setup for the assembly of a reporter construct:

component	volume
10x buffer (Cut Smart)	2 µl
ATP (10 mM)	2 µl
*Bsa*I	1 µl
T4 DNA ligase (5u)	1 µl
destination vector pGWB3GG (50 ng/µl)	1 µl
eluted PCR fragment [*1] (50 ng/µl)	1 µl
water	12 µl

[*1] TALE box + minimal *Bs4* promoter or a 1000 bp promoter fragment from a rice gene

The thermocycling protocol is identical to the one for the assembly of multi repeat modules.

2.2.14 *Agrobacterium*-mediated transient expression in *N. benthamiana*

A. tumefaciens strains were resuspended in *Agrobacterium* infiltration media (AIM) and adjusted to an optical density (OD_{600}) of 0.8. After combining a reporter- and a TALE-expressing strain in a 1:1 ratio, inoculation of *N. benthamiana* plants was performed using a needleless syringe. 48 h after inoculation samples were harvested. Samples were used either for analysing the GUS activity of TALEs in combination with reporter constructs, for analysing TALE protein expression in the plant or for analysing the *in planta* localisation and signal strength of TALEs via microscopy.

2.2.15 Microscopic analysis

For the microscopy analysis, TALE-expressing *A. tumefaciens* strains were inoculated into *N. benthamiana*. Two days after their inoculation, leaf samples were taken, placed in water and vacuum-infiltrated. The samples were then placed on a glass slide, covered with a droplet of water and analysed using a confocal laser scanning microscope. Pictures were made and analysed using the software ZEN (Carl Zeiss).

2.2.16 GUS reporter assay

For a GUS reporter assay, TALE- and reporter gene-expressing *A. tumefaciens* strains are mixed in a 1:1 ratio and inoculated into *N. benthamiana*. Both constructs integrate into the genome of plant cells but while the TALE constructs are constitutively expressed, the *GUS* gene of the reporter constructs remains (mostly) silent. However, if an adequate TALE binding site is present within a promoter, the TALE can bind and thereby induce the expression of the *GUS* gene. The strength of the induction can be visualized using a qualitative GUS assay or measured using a quantitative GUS assay (Figure 2.5).

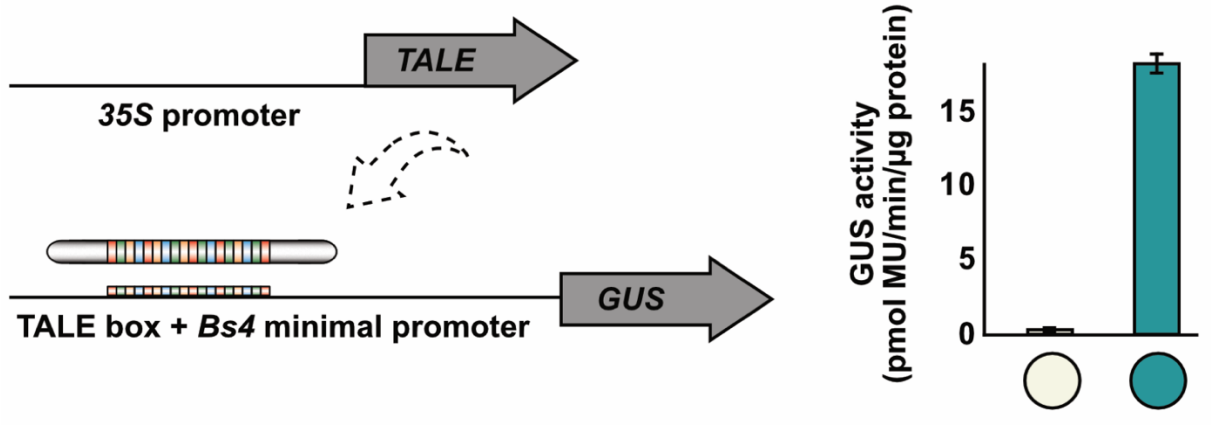


Figure 2.5: General principle of the GUS reporter assay used to analyse the activity of TALEs. If the constitutively expressed TALE is able to bind to the TALE box in the reporter construct, the otherwise silent *GUS* gene is induced and GUS activity can be detected.

2.2.16.1 Qualitative GUS assay

For the qualitative GUS assay, three leaf discs per construct (each one from an individual plant) were placed in GUS staining solution containing 0.1% X-Gluc. The leaf discs were briefly vacuum infiltrated and then incubated for ~12 h at 37°C. Afterwards the samples were destained by removing the staining solution and replacing it for 1 day with 96% and then for 1-2 days with 70% ethanol. In the end, the leaf discs were placed between two layers of acetate foil, dried and photo documented.

2.2.16.2 Quantitative GUS assay

For a quantitative GUS assay, two leaf discs per inoculation spot were pooled, homogenized in the TissueLyser II (Qiagen) for 40 seconds with 30 Hz and resolved using 300µl GUS extraction buffer. After a centrifugation step at 4°C and 13000 rpm for 10 minutes, 200 µl of the supernatant were transferred into a 96-well plate and the plate was centrifuged as well at 4°C and 13000 rpm for 10 minutes. Afterwards, the samples were mixed with 4-methyl-umbelliferyl-Beta-D-glucuronide (MUG) and incubated at 37°C for 60-90 minutes. For this, 10 µl per sample and 90 µl MUG-containing GUS extraction buffer were combined and briefly mixed. After the incubation period, the reactions were stopped by adding sodium carbonate (Na_2CO_3). For this, 10 µl per sample and 90µl Na_2CO_3 (0,2M) were combined and briefly mixed. The fluorescence of the reaction product 4-methylumbelliferone (4-MU) was measured in the Spark M10 multimode microplate reader (Tecan Group) using a black 96-well plate with translucent bottom. Finally, with the help of a Bradford assay, the total protein concentration of each sample was calculated and taken into account as well. To calculate the GUS activity, the following formula was used:

$$A_{GUS} = \frac{\Delta F \times 10\ \text{pmol MU}}{t\ [\text{min}] \times \text{protein}\ [\text{mg}] \times F_{10\ \text{pmol MU}}}$$

A_{GUS} = GUS activity in pmol 4-MU/min/µg protein
ΔF = fluorescent count of the sample minus the background fluorescence
t = reaction time in minutes
$F_{10\ \text{pmol MU}}$ = fluorescent count of the 10 µM MU standard

2.2.16.3 Bradford

To calculate the total protein concentration of a sample, a Bradford assay was performed. For this, 30 µl plant extract was combined with 770 µl water and 200 µl of the Bradford reagent Roti-Quant (Carl Roth) and the mix was incubated for 15-20 minutes at room temperature. After the incubation period, the absorption of the mixture was measured at

595 nm. The protein concentration of each sample was calculated by normalising the measured values against a standard curve obtained from several BSA (Molecular Biology Grade; New England Biolabs) samples with a defined protein concentration.

2.2.17 Preparation of *N. benthamiana* leaf material for analysis via SDS-PAGE

To detect GFP-tagged TALEs *in planta*, TALE-expressing *Agrobacterium* strains were inoculated into leaves of *N. benthamiana*. After two days, six small leaf discs were harvested and stored in liquid nitrogen. After homogenising the samples in the TissueLyser II (Qiagen) for 40 seconds with 30 Hz, the powdery remains of the leaf material were mixed with 100 µl 4x Laemmli and incubated for 10 minutes at 98°C. After centrifuging the samples at 14000 rpm for 5 minutes, the supernatants were analysed via SDS-PAGE or stored at 4°C until further use.

2.2.18 Preparation of *Xanthomonas* cell material for analysis via SDS-PAGE

To detect FLAG-tagged TALEs*, Xanthomonas* cultures with TALE-expressing constructs were cultivated in liquid media at 28°C overnight. The cultures were adjusted to an OD_{600} of 0.2 and 1 ml was harvested via centrifugation at full speed for 2 minutes. The supernatant was removed and the pellet resuspended in 70µl 4x Laemmli buffer. After denaturing the samples at 98°C for 10 minutes, they were either used immediately for analysis via SDS PAGE or stored at 4°C until needed.

2.2.19 SDS-PAGE

To separate protein samples by their molecular weight, they were subjected to a sulphate polyacrylamide gel electrophoresis (SDS-PAGE). For this, 15 µl per sample were loaded. Each gel consisted of two parts, a 5% stacking gel and an 8% separation gel. To asses the size of the protein bands after their separation, a protein ladder was included, usually the PageRuler Prestained Protein Ladder (Thermo Fisher Scientific). The electrophoresis was done in 1x TANK buffer with a constant 120 V and 25 mA per gel using the Mini-PROTEAN Tetra Vertical Electrophoresis Cell (Bio-Rad Laboratories).

2.2.20 Immunoblotting

Once separated via SDS-PAGE, the proteins were transferred from the gel to a nitrocellulose membrane via semi dry blotting using towbin buffer and the Trans-Blot Turbo Transfer System (Bio-Rad Laboratories). After the the blotting process, the membrane containing the proteins was placed in blocking solution and incubated for 1 hour at room temperature while shaking. Afterwards, the membrane was washed three times for 5 minutes with 1x TBST. Next, the membrane was incubated with primary antibody (dilution 1: 10000; in 1x TBST)

while shaking – either for several hours at room temperature or overnight at 4°C. Following this, the membrane was washed again three times for 5 minutes with 1x TBST before it was incubated with secondary antibody (dilution 1: 10000; in 1x TBST) for 1 hour at room temperature while shaking. Finally, the membrane was washed 3 more times for 5 minutes with 1x TBST and incubated for 1 minute with 6 ml ECL solution. Detection of signals was done using the ChemiDoc Imaging System (Bio-Rad Laboratories).

2.2.21 Inoculation of *Oryza sativa* with *Xanthomonas*

Xanthomonas was grown on solid media. After 2-3 days, the bacteria were harvested, resuspended in 10 mM $MgCl_2$ and adjusted to an OD_{600} of 0.5. The bacterial suspensions were inoculated into 3-5 weeks old rice plants using a needleless syringe. For this, 3-4 inoculation spots were placed within the first 5 cm of a leaf (excluding the first cm of the leaf tip). Disease symptoms (water-soaked lesions) were documented 4 days post inoculation.

2.2.22 Analysing *Xanthomonas* genomes and TALE sequences for the presence of aberrant repeats

All available completely sequenced *Xanthomonas* genomes as well as all TALE sequences were checked for the presence of repeat variants of aberrant length. For this analysis the "TALE prediction" and the "TALE class assignment" functions of the AnnoTALE tool v1.3 were used (Grau *et al*., 2016).

2.2.23 Prediction of TALE target genes

Prediction of TALE target genes was done using the "Predict and Intersect Targets" function of AnnoTALE v1.3 (Grau *et al*., 2016). As basis for the prediction, promoter sequences spanning the region between 300 bp upstream of the transcriptional start site and 200 bp downstream of the transcriptional start site (or the start codon, depending on what came first) of all rice genes were used. The Sequences were extracted from the genome of *Oryza sativa* cv. Nipponbare version 7.0 (available at http://rice.plantbiology.msu.edu).

3 Results

3.1 Rare repeat variants from *Xanthomonas* TALEs and their effect on DNA-binding

3.1.1 Naturally occurring aberrant repeats and their classification

The recent increase in fully sequenced *Xanthomonas* genomes offers the unique opportunity to analyse the distribution of aberrant repeats in TALEs in unprecedented detail as well as to look for and identify completely new aberrant repeat variants. In order to do so, all available TALE-sequences from *Xanthomonas* spp. were analysed for the presence of an aberrant repeat. For this, the AnnoTALE functions "TALE prediction" and "TALE analysis" were used (Grau *et al*., 2016). Every repeat deviating in length from a standard repeat of 34aa or from the commonly occurring length variants with 33aa or 35aa was classified as aberrant repeat. Within the 60 completely sequenced *Xanthomonas* genomes and the multiple individual TALE sequences that were analysed, 82 repeats of aberrant length were identified. These 82 aberrant repeats occurred within 80 TALEs (Appendix Table 3.1.1 and 3.1.2). With the exception of two TALEs found in the *X. axonopodis* pv. *citri* strain 29-1, all aberrant repeat-carrying TALEs were identified in Asian *X. oryzae* strains (*Xoo* or *Xoc*). Moreover, every single Asian *X. oryzae* strain sequenced so far carries at least one TALE with an aberrant repeat (Fig. 3.1.1 and 3.1.2). Thirty-two of the 80 aberrant repeat-carrying TALEs are annotated as pseudogenes. Since all but two of these 32 share similar small N- and C-terminal deletions as well as the aberrant repeat variant with 28aa, it seems likely that they have a shared evolutionary origin (Fig. 3.1.1-3.1.3). If the AnnoTALE nomenclature is taken into account (Grau *et al*., 2016) the 80 aberrant repeat-carrying TALEs can be categorized into 19 different TALE classes (Fig. 3.1.1-3.1.3; Appendix Table 3.1.1). The aberrant repeats themselves can be categorized into seven different length variants, with 42aa, 40aa, 37aa (version 1 and 2), 36aa (version 1 and 2), 30aa and 28aa (Fig. 3.1.3; Appendix Table 3.1.1 and 3.1.2).

For the repeat variants with 42aa, 40aa and 30aa, it was shown that they mediate a flexible recognition of a perfectly matching sequence (optimal box) and a -1 nucleotide frameshift sequence (Richter *et al*., 2014). In contrast, none of the other aberrant repeat variants was analysed for the ability to facilitate frameshift recognition and three of them have not been described at all (36aa v1, 37aa v1 and 2).

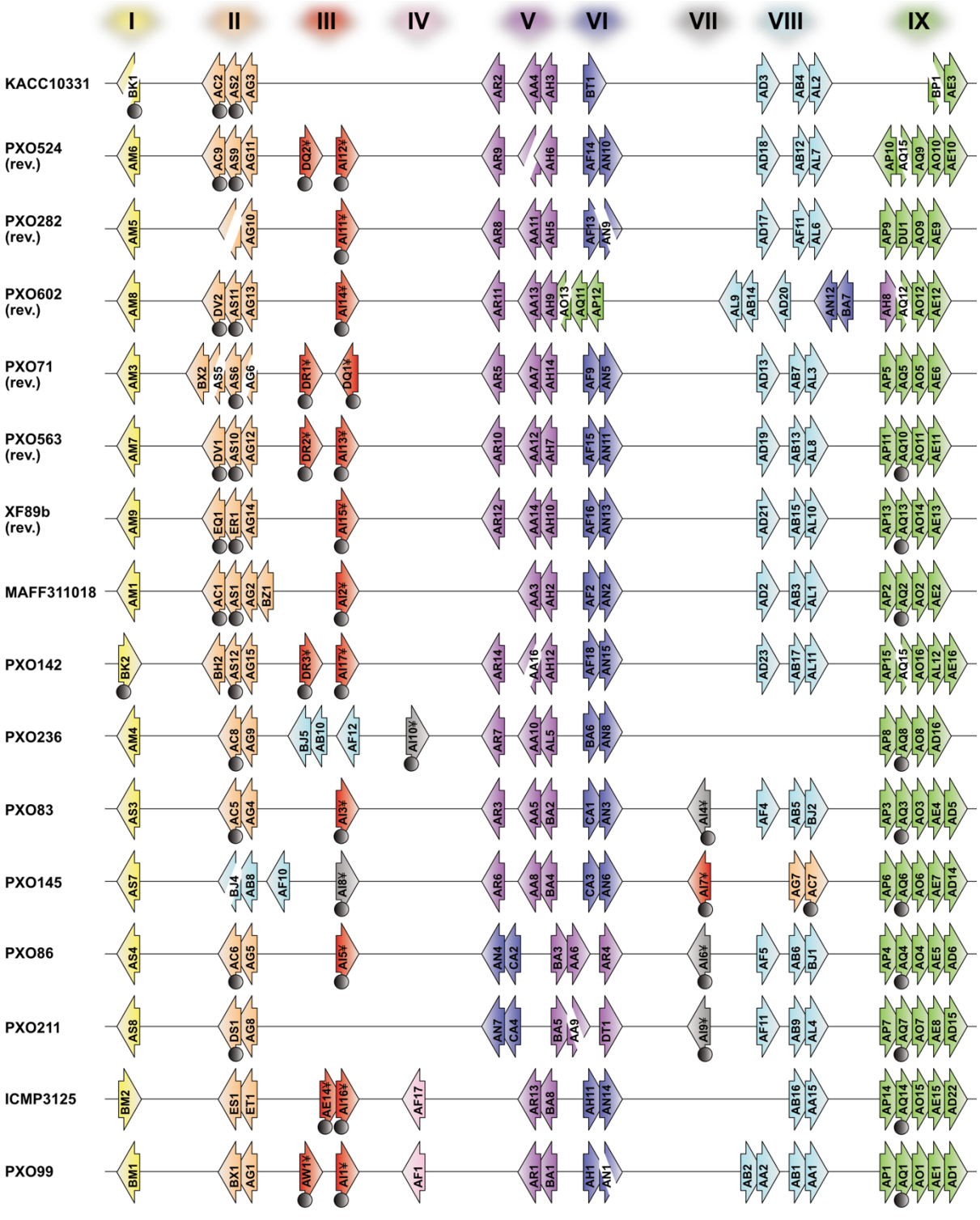


Figure 3.1.1: TALE cluster affiliation of all completely sequenced *Xoo* strains originating in Asia. *TALE* genes are indicated as arrows with regard to their orientation in the genome. All TALEs are named according to the AnnoTALE classification system. TALEs were assigned to TALE clusters depending on their genomic position and their adjacent genes. The genomic TALE clusters can be differentiated by colour. TALEs carrying an aberrant repeat are labeled with a grey dot. All TALEs that contain small N- and C-terminal deletions are indicated by the ¥ symbol while a broken arrow indicates fragmented TALE *genes* that contain large deletions or frameshifts. To increase clarity, several strains are shown in reverse orientation (rev.). The figure was modified according to Mücke *et al*., 2019. The TALE repertoire of African and Asian *Xoo* strains differs significantly and none of the so far sequenced *Xoo* strains from Africa contains an aberrant repeat. Therefore, only Asian *Xoo* strains were included in this overview.

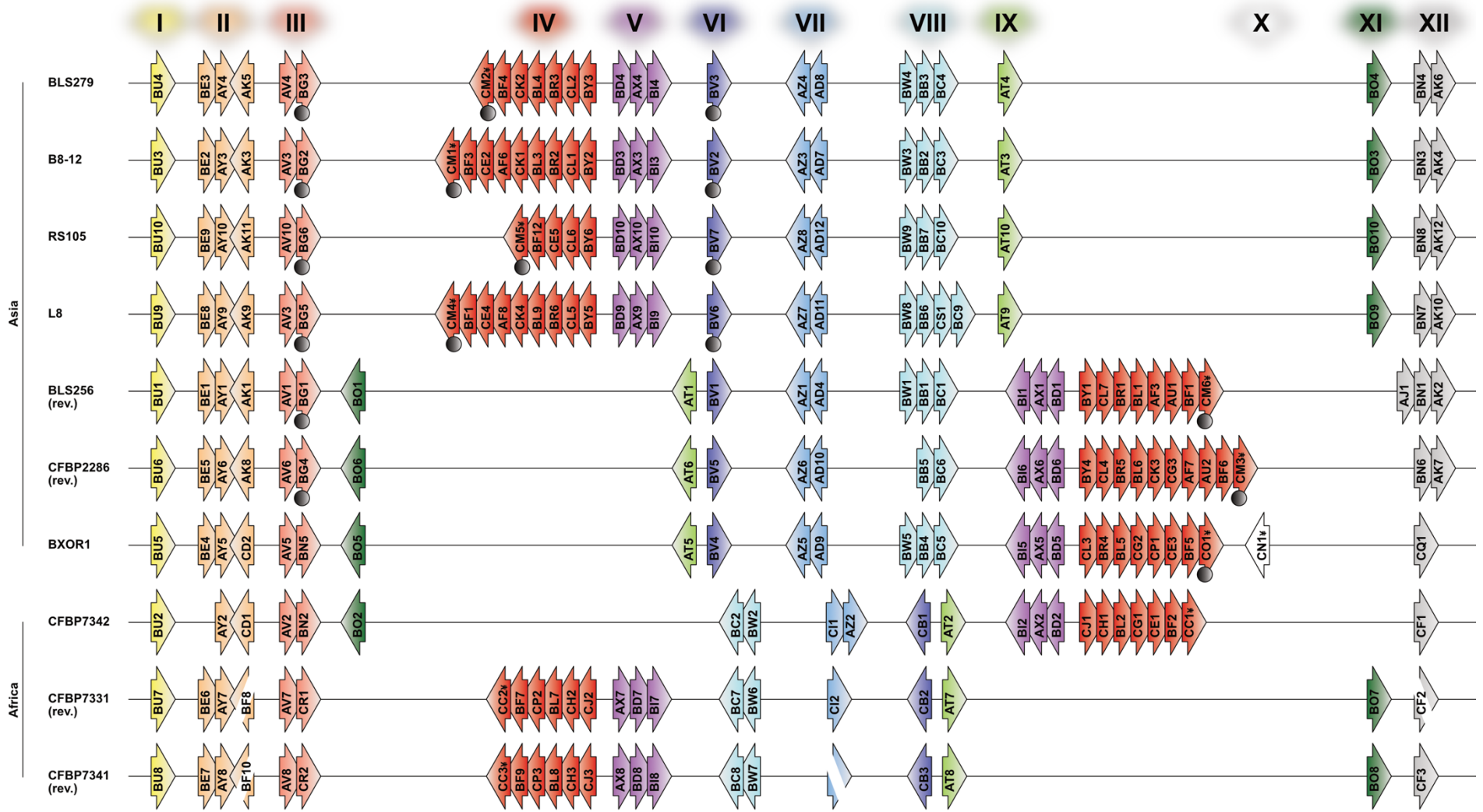


Figure 3.1.2: TALE cluster affiliation of all completely sequenced *Xoc* strains originating in Asia and Africa. *TALE* genes are indicated as arrows with regard to their orientation in the genome. TALEs are named according to the AnnoTALE classification system. Based on its genomic position and its adjacent genes, each TALE was assigned to one of 12 TALE clusters. The genomic TALE clusters can be differentiated by colour. TALEs carrying an aberrant repeat are labelled with a grey dot. TALEs with incomplete N- and C-terminal regions are labeled with a ¥ symbol while a broken arrow indicates fragmented *TALE* genes that contain large deletions or frameshifts. To increase clarity, several strains are shown in reverse orientation (rev.).

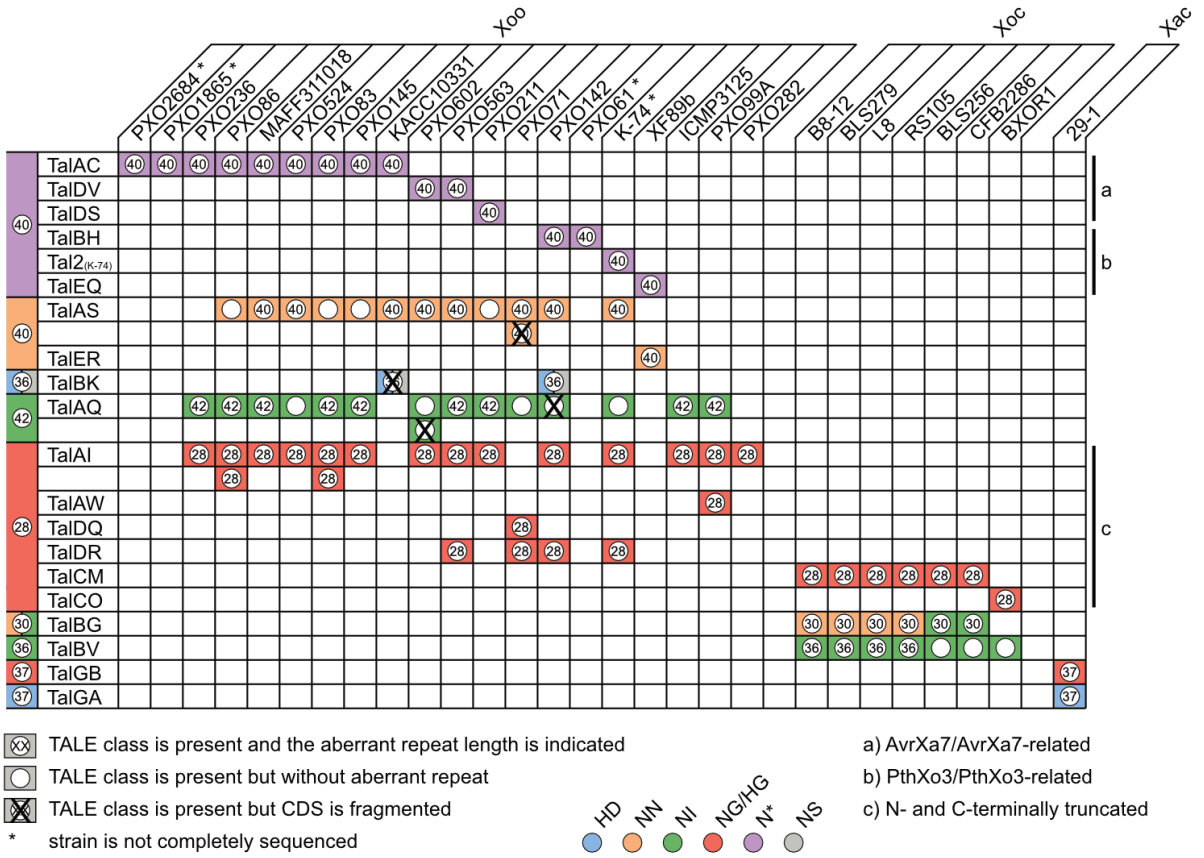


Figure 3.1.3: Overview of all TALE classes containing at least one member with an aberrant repeat. TALEs are named according to the AnnoTALE classification system. The TALEs from the strain K-74 are not yet included into AnnoTALE, however, four out of five of its aberrant repeat-carrying TALEs contain RVD sequences that are identical to members of known TALE classes and therefore were temporarily assigned into those. *Xanthomonas* spp., strain of origin and pathovar are indicated (*X. oryzae* pv. *oryzae* = *Xoo*; *X. oryzae* pv. *oryzicola* = *Xoc*; *X. axonopodis* pv. *citri* = *Xac*). The circled numbers indicate the length of each aberrant repeat while an empty circle indicates that no aberrant repeat is found in this particular class member. The different RVDs found within each of the aberrant repeats are colour-coded. All TALEs with the 28aa aberrant repeat have small N- and C-terminal truncations and are therefore regarded as pseudogenes. Other fragmented TALE genes with larger deletions or frameshift mutations are crossed out.

3.1.2 Not all aberrant repeats facilitate flexible frameshift recognition

To analyse the impact of the novel aberrant repeats on the DNA-binding mode of TALEs, a single aberrant repeat of each type was introduced into an artificial TALE. In addition to the aberrant repeats described in the previous chapter, two variants with 35aa were also included in the analysis. These repeats occur rather frequently in some *Xanthomomonas* spp. and they have been shown to confer normal base-pair recognition. However, they have never been analysed for the ability to confer frameshift recognition. To construct the artificial TALEs, the different aberrant repeat variants were converted into cloning modules that are fully compatible with the “Golden TAL Technology” cloning kit (Appendix Table 3.1.3 and 3.1.4; Geißler *et al.*, 2011). The constructed TALEs contained 17.5 repeats and an alternating DNA-binding specificity within the consecutive repeats to maximize the effect of a

frameshift mutation in the DNA target site. The same design was used previously in Richter *et al*., 2014. To ensure comparability, the aberrant repeats were placed at the same position (position 7) and contained the same RVD (HD) (Fig. 3.1.4 A-D). Each TALE was tested in combination with two different TALE target boxes in a GUS reporter assay. The first TALE-box was an optimal target sequence based on the TALE RVDs (op), the second one a frameshift derivative of this box with just a single nucleotide deleted at the position corresponding to the aberrant repeat (-1 p7). If a TALE successfully binds to a target box, the *GUS* reporter gene is induced and activity can be determined. For this, the TALE-boxes were fused to the minimal *Bs4* promoter and this fusion was than placed in front of the *β-Glucuronidase* (*GUS*) coding sequence.

The control TALE (34aa) without any aberrant repeat was able to activate transcription if combined with the optimal box but was unable to do so with the frameshift derivative (Fig. 3.1.4 E). Similar observations were made for the two repeats with 35aa, the variant 36aa v2 and the variant 37aa v1, making the latter two the first naturally occurring aberrant repeats

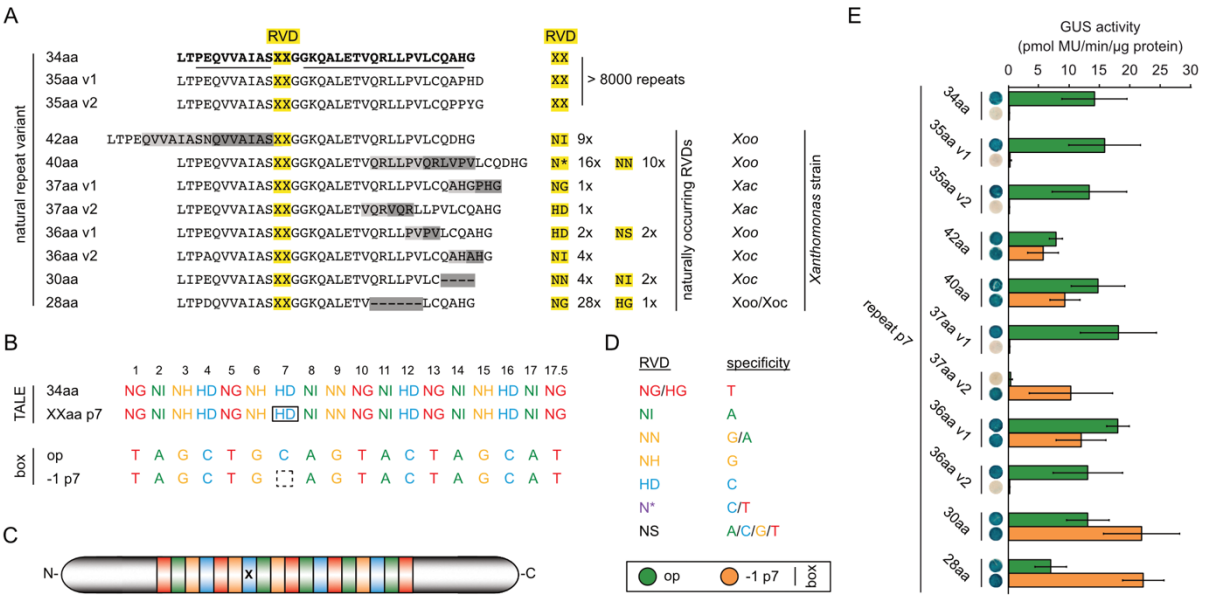


Figure 3.1.4: Naturally occurring aberrant repeat variants identified in TALEs from *Xanthomonas* spp. and their impact on the TALE-DNA interaction. (A) Amino acid alignment of repeats containing the known naturally occurring repeat length polymorphisms. The repeat variable di-residues (RVD) are shaded in yellow. A typical standard repeat of 34aa length is shown in bold, the helix-forming residues are underlined, duplicated or deleted aa positions are shaded in grey. Frequency of occurrence for each aberrant repeat variant is indicated as well as the *Xanthomonas* pathovar they occur (*Xoo* = *X. oryzae* pv. *oryzae*, *Xoc* = *X. oryzae* pv. *oryzicola*, *Xac* = *X. axonopodis* pv. *citri*). The two shown 35aa repeats are representative examples. **(B)** RVD composition of the generated TALEs. Position 7 (boxed) either contains a standard repeat with 34aa (top row) or one of the variants of aberrant length (XXaa) with 42aa, 40aa, 37aa, 36aa. 35aa, 30aa or 28aa, respectively. The reporter construct is a fusion between a TALE-box and the minimal *Bs4* promoter that was placed in front of the *GUS* reporter gene. The TALE-box is either perfectly matched to the RVDs, named optimal box (op), or a frameshift derivative of this box with a single nucleotide deletion placed at position 7 (-1 p7). **(C)** Schematic view of the generated TALEs. The repeats are shown as rectangles and their colours correspond to the different RVDs. The x represents the position of the aberrant repeat. **(D)** DNA-specificities of selected RVDs. **(E)** Results of the qualitative and quantitative GUS assay. Error bars represent standard deviation (n = 3). A representative leaf disc is shown. Similar results were obtained in at least two independent experiments.

that are unable to confer a frameshift recognition (Fig. 3.1.4 E). In contrast, the TALEs containing one of the three aberrant repeats known to confer frameshift recognition or the variant 36aa v1 were able to address both boxes (Fig. 3.1.4 E; 42aa, 40aa, 30aa and 36aa v1). The TALE containing the 28aa repeat was also able to address both boxes but showed a clear and reproducible preference for the frameshift box. Interestingly, the TALE containing the newly identified aberrant repeat 37aa v2 showed only activity on the frameshift box, rendering this variant the first and only natural aberrant repeat that is completely unable to participate in DNA-binding (Fig. 3.1.4 E).

These results allow for an additional categorization of naturally occurring aberrant repeats, one that is based on their functionality. The first category includes those repeats that are completely unaffected by their alteration, the second category contains the very flexible variants that are able to facilitate a facultative frameshift recognition while the members of the third group are completely unable to participate in DNA-recognition, likely by constantly looping out of the repeat array. Since some of the naturally occurring aberrant repeats show no effect on the TALE-DNA interaction, it can be concluded that a difference in length or amino acid composition alone does not necessarily lead to a repeat with the ability to confer frameshift recognition.

3.1.3 Aberrant repeats that do confer frameshift recognition prefer to loop out of the repeat array themselves

TALEs with an aberrant repeat of 42aa, 40aa or 30aa length showed the highest activity on a TALE-box with either no deletion or a deletion exactly corresponding to the position of the aberrant repeat (Richter *et al*., 2014). This observation lead to the conclusion that aberrant repeats are not forcing out one of their adjacent repeats but instead prefer to loop out themselves when challenged with a frameshift sequence. In order to investigate how the novel aberrant repeats with 37aa (v1 and v2), 36aa (v1 and v2) or 28aa behave, two additional frameshift boxes were generated, one containing a single nucleotide deletion at position 6 and one with the deletion at position 8 (Fig. 3.1.5 A-D). The control TALE with exclusively 34aa repeats and the TALEs with the different aberrant repeat variants were then compared on the optimal box and on the three different -1 frameshift derivatives containing a deletion at position 6, 7 or 8. All TALEs containing an aberrant repeat with the ability to facilitate a flexible frameshift recognition showed highest activity either on an optimal box or a box where the -1 nucleotide deletion was placed at the position exactly corresponding to the aberrant repeat (Fig. 3.1.5 F and G). The TALE with the 37aa v2 repeat variant could activate only the box with the -1 deletion at position 7, further supporting the idea that this repeat variant is looping out constitutively. The TALEs containing the aberrant repeats

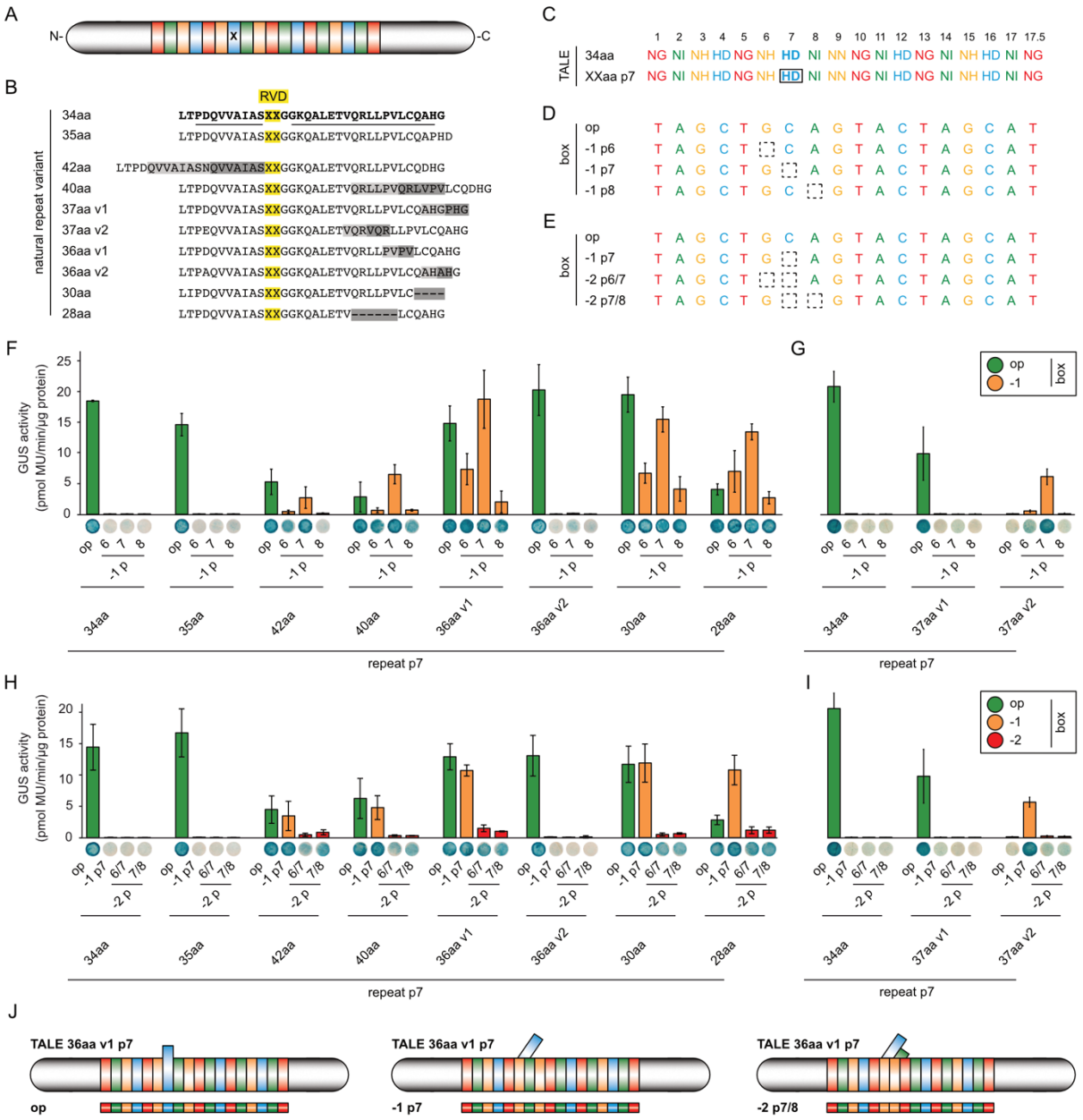


Figure 3.1.5: Aberrant repeats that can facilitate frameshift binding prefer a target box where the -1 nucleotide deletion is placed exactly corresponding to the position of the aberrant repeat. (A) Schematic view of the generated TALEs. The altered repeat is indicated by an x. **(B)** Amino acid alignment of repeats containing the known naturally occurring repeat length polymorphisms. **(C)** RVD composition of the generated TALEs. Position 7 (boxed) either contains a standard repeat with 34aa (top row) or one of the variants of aberrant length with 42aa, 40aa, 37aa, 36aa, 35aa, 30aa or 28aa, respectively. **(D)** The TALE-box is either an optimal box (op) or a frameshift derivative with a single nucleotide deletion at position 6, 7 or 8 (-1 p6, p7 or p8). **(E)** In an additional experiment, the same TALEs were tested in combination with an optimal box (op), a box with a single nucleotide deletion at position 7 (-1 p7) or a box with a -2 nucleotide deletion at the positions 6 and 7 (-2 p6/7) or 7 and 8 (-2 p7/8). **(F)** and **(G)** Results of the qualitative and quantitative GUS assay for the repeat variants on different -1 frameshift boxes. **(H)** and **(I)** Results of the qualitative and quantitative GUS assay for the repeat variants on the different -2 frameshift boxes. Error bars represent standard deviation (n = 3). Similar results were obtained in at least two independent experiments. A representative leaf disc is shown for each combination. **(J)** Model of a TALE with a functional aberrant repeat bound to an optimal box, a matching -1 and a -2 frameshift box.

37aa v1, 36aa v2 or 35aa, on the other hand, recognized only the optimal target sequence, and thus they behaved exactly like the TALE solely constructed with standard 34aa repeats (Fig. 3.1.5 F and G). These results suggest that aberrant repeats that can confer recognition of a frameshift sequence prefer to loop out of the repeat array themselves.

The loop out mechanism of aberrant repeats raised the question whether an aberrant repeat can also mediate the binding of a TALE to a target sequence containing a -2 nucleotide frameshift. This could be the case if the aberrant repeat causes disturbances in the stabilizing inter- and intra-repeat interactions so that one of its adjacent normal repeats is forced out as well. To investigate this, two new target boxes were generated, containing a -2 nucleotide deletion at the positions 6 and 7 or 7 and 8. These boxes were tested together with the TALEs containing one of the natural aberrant repeats at position 7 (Fig. 3.1.5 E). In contrast to a 34aa standard repeat, most aberrant repeats did facilitate the recognition of a -2 frameshift box, however, the observed activity was extremely low (Fig. 3.1.5 H and I). This result suggests nonetheless that aberrant repeats do have a weakening effect on the inter- and intra-repeat interactions of a TALE and that this effect can lead to a situation where a standard repeat of 34aa is forced out together with an aberrant repeat. This is also the first indication that a TALE is able to loop out multiple repeats simultaneously in order to fit to a specific target sequence (Fig. 3.1.5 J).

3.1.4 Multiple aberrant repeats in tandem can loop out at once

Aberrant repeats are rare and natural TALEs carrying an aberrant repeat contain usually only a single one of them. However, during the course of the present study, the first known exception was identified, class TalBK. Its members contain two aberrant repeats (36aa v1) placed at position 9 and 12 of the repeat array. This finding and the observation that an artificial TALE with two 40aa aberrant repeats in tandem was active (Richter *et al.*, 2014), raised the question, how many aberrant repeats a TALE can tolerate while remaining functional.

To address this question, several TALEs were generated with zero to six 40aa aberrant repeats in tandem. These TALEs were tested on an optimal box and on target sequences which had one to six nucleotides deleted (Fig. 3.1.6 A and B). To limit the number of necessary boxes to a minimum, while on the other hand gain full readout on how many aberrant repeats are tolerated and how many repeats can loop out, the RVD setup of the TALEs and the respective target sequences were simplified. The repeats in the central region of the TALEs (repeats 7 to 12) were substituted to NN repeats, which can bind to any of the Gs in the newly designed target sequences. At the same time, six repeats with mixed RVDs were kept as “specific anchors” in the beginning and the end of the TALEs to enforce

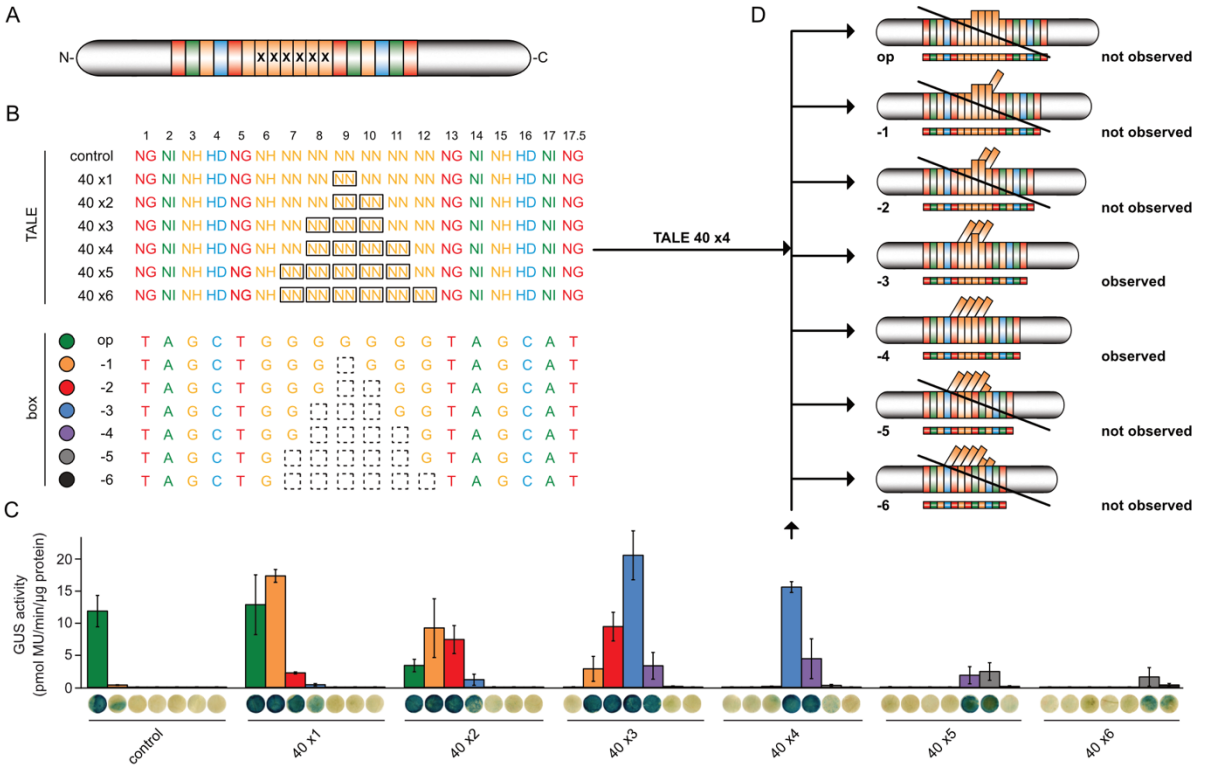


Figure 3.1.6: TALEs with multiple aberrant repeats in tandem are functional but require a target box with multiple nucleotides deleted. (A) Schematic view of the generated TALEs. The repeats altered in this experiment are indicated by an x. **(B)** RVD composition of the generated TALEs. The region from position 7 to 12 consists only of repeats carrying the RVD NN. These NN repeats contain a varying number of the 40aa aberrant repeat, ranging from 0 to 6 (normal, 40 x1 - 40 x6). Positions with an aberrant repeat are indicated by boxed RVDs. The target box was either an optimal box without any mismatches (op) or one of several frameshift variants with 1 to 6 nucleotides deleted (-1 to -6). Deleted nucleotide positions are boxed in dashed squares. **(C)** Results of the GUS assay for the artificial TALEs with varying numbers of aberrant repeats. Error bars represent standard deviation (n = 3). Similar results were obtained in at least two independent experiments. A representative leaf disc is shown for each combination. **(D)** To illustrate the possible binding mechanism of a TALE with multiple aberrant repeats, a schematic view of TALE 40 4x is shown. Activity for this TALE was observed only if it was combined with boxes carrying a deletion of 3 or 4 nucleotides (-3 or -4), thus demonstrating a direct correlation between the number of aberrant repeats and the number nucleotides deleted in the target box.

the exact positioning of both TALE ends (Fig. 3.1.6 A and B). This altered setup allowed testing of all TALEs on the same set of reporter boxes. The obtained results clarified that TALEs can not only tolerate the naturally observed one or two aberrant repeats but several more (Fig. 3.1.6 C). A TALE with as many as four aberrant repeats in tandem still yielded a very high activity on the -3 box, being comparable even to the control TALE on the optimal box (Fig. 3.1.6 C). Introducing more than four aberrant repeats, however, led to a drastic decrease in maximal TALE activity. Since TALEs with multiple aberrant repeats were only active if combined with boxes having multiple nucleotides deleted as well, it can be concluded that multiple aberrant repeats in tandem tend to loop out simultaneously (Fig. 3.1.6 C and D). The highest activities for the tested TALEs were observed when they were combined with boxes carrying as many nucleotide deletions or one less as aberrant repeats present. This suggests that the most preferable confirmation for a TALE with multiple

40aa aberrant repeats in tandem is obtained when all or all but one of the aberrant repeats loop out. These findings further support the idea that aberrant repeats weaken the repeat-to-repeat interaction of a TALE, thus allowing for their looping out in the first place. Finally, these results show that up to three repeats can loop out of a TALE at once without any penalty for its overall activity. However, this number is very likely influenced by the length and RVD composition of the TALE, and thus should not be considered as a constant.

3.1.5 The RVD setup of TALEs influences their ability to tolerate aberrant repeats

To investigate if the RVD composition of TALEs influences their ability to tolerate aberrant repeats, multiple aberrant repeats were incorporated into a new set of TALEs containing a mixed RVD setup. Tested on an optimal and various frameshift boxes, solid activity was observed for the TALEs containing none, one or two aberrant repeats while a drastically reduced activity was observed for the TALE containing three aberrant repeats (Fig. 3.1.7). This is in stark contrast to the results from the previous chapter where as many as four aberrant repeats in tandem were tolerated without a noticeable decrease in maximal TALE activity. Since the observed differences between the two TALE sets are most likely a direct consequence of the RVD changes at the positions 7 to 12, it can be concluded that the RVD composition of TALEs does affect their capability to tolerate multiple aberrant repeats. One possible explanation for this might lie in the reduced number of strong RVDs (HD and NN) found within the second set of TALEs (Streubel *et al*., 2013). A lower number of strong RVDs will result in TALEs that have a lower overall DNA affinity, and thus, most likely, also a decreased ability to cope with the partially disrupting effects of multiple aberrant repeats. If multiple aberrant repeats are to be placed within TALE, it might therefore be advisable to incorporate more strong RVDs than usually necessary.

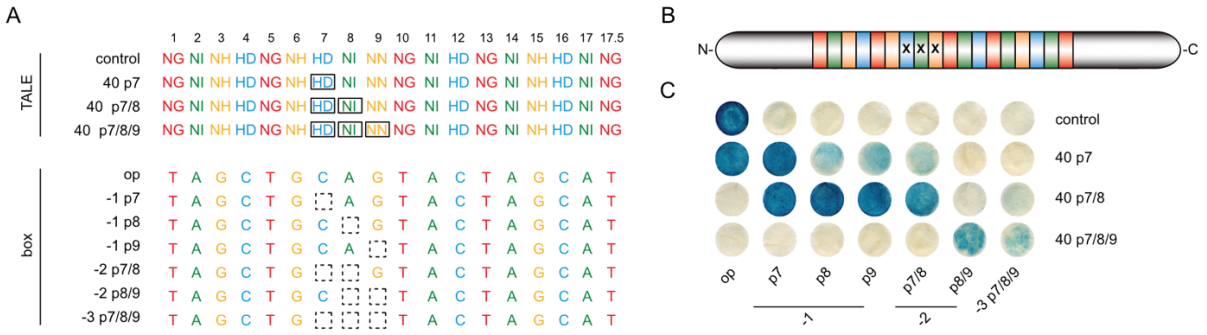


Figure 3.1.7: The RVD composition of TALEs affects their ability to tolerate multiple aberrant repeats. (A) RVD compositions of the generated TALEs. Position 7, 8 and 9 (boxed) contain either a standard repeat of 34aa length or an aberrant repeat with 40aa. The TALE-box is either an optimal box (op) or a frameshift variant with a single nucleotide deletion at position 7, 8 or 9, with two nucleotides deleted (p7/8 or p8/9) or with all three nucleotides deleted simultaneously. **(B)** Schematic view of the generated TALEs. The repeats altered in this experiment are indicated by an x. **(C)** Results of the qualitative GUS assay. A representative leaf disc is shown for each combination.

3.1.6 Neighborhood community – the nature of an aberrant repeat controls its impact on either one of the flanking repeats

The observation that aberrant repeats differ in their capabilities to confer frameshift recognition (see chapter 3.1.2) suggested that they also affect their adjacent repeats differently. To investigate this, two new TALEs were generated, both identical to the TALEs 40 p7/8 and 40 p7/8/9 from the previous chapter but instead of containing the 40aa aberrant repeat at positions 7 to 8 and 7 to 9, they contained the 42aa aberrant repeat at these positions. Comparing these four TALEs on different target boxes with various nucleotide deletions revealed that two TALEs with the same number and positioning of aberrant repeats in tandem as well as the same overall RVD setup can nonetheless address different target boxes – depending on which aberrant repeat version is present (Fig. 3.1.8).

To investigate this further, a new set of TALEs was generated, with each TALE containing one of the known aberrant repeat variants placed at positions 7 (HD) and 8 (NN) (Fig. 3.1.9). To distinguish which of the two aberrant repeats participates in DNA binding and which one loops out, RVDs with different nucleotide specificities were used at the two positions. The TALEs were tested on four different target boxes: one with no deletion (op), one with the

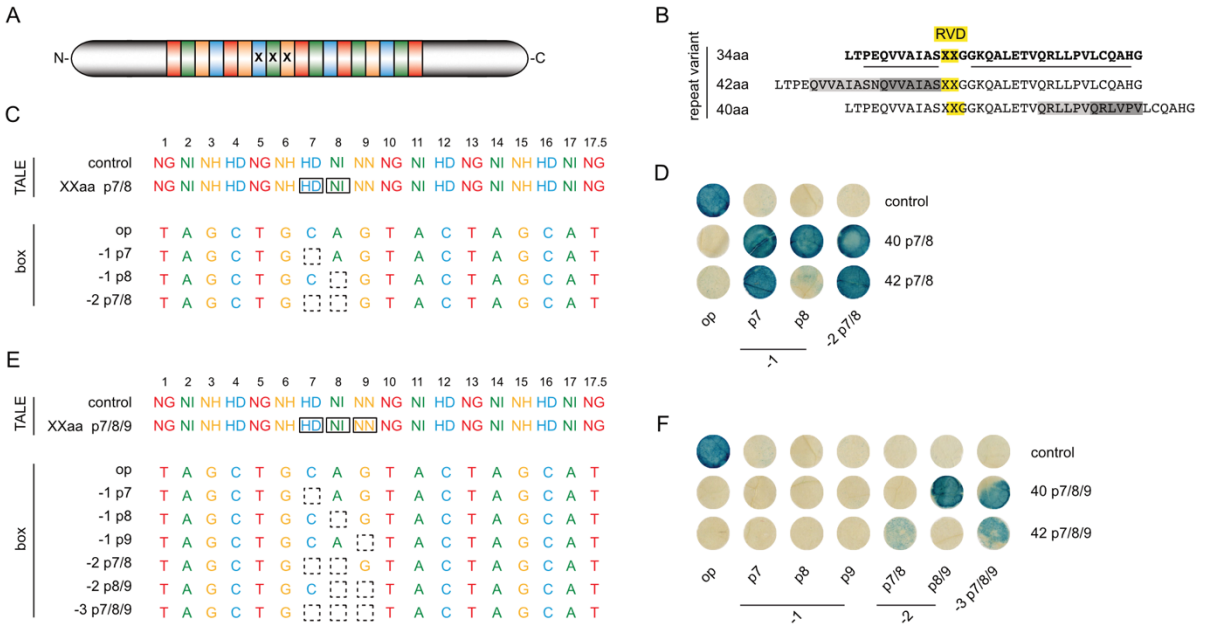


Figure 3.1.8: Placed in tandem, the 40aa and 42aa aberrant repeats result in TALEs that address different target box combinations. (A) Schematic view of the generated TALEs. The repeats altered in this experiment are indicated by an x. **(B)** Amino acid alignment of the three repeat variants used in this experiment. **(C)** RVD composition of the TALEs. Position 7 and 8 (boxed) contain either a standard repeat of 34aa length, the 40aa or 42aa aberrant repeat. The TALE-box is either an optimal box (op) or a frameshift variant with a single nucleotide deletion at position 7 (-1 p7), position 8 (-1 p8) or one with a deletion at both positions (-2 p7/8). **(D)** Results of the qualitative GUS assay for the TALEs from (C). **(E)** RVD compositions of the generated TALEs. Position 7, 8 and 9 (boxed) contain either a standard repeat of 34aa length, the 40aa or 42aa aberrant repeat. The TALE-box is either an optimal box (op) or a frameshift variant with a single nucleotide deletion at position 7, 8 or 9, with two nucleotides deleted (-2 p7/8 and -2 p8/9) or with all three nucleotides missing simultaneously (-3 p7/8/9). **(F)** Results of the qualitative GUS assay for the TALEs from (E). A representative leaf disc is shown for each combination. Similar results were obtained in at least two independent experiments.

nucleotide deleted that directly corresponds to the first aberrant repeat (-1 p7), one with the nucleotide deleted that corresponds to the second aberrant repeat (-1 p8) and one with both of these nucleotides deleted (-2). For the TALEs containing either two 28aa or two 37aa v2 repeats, activity was exclusively observed on the -2 box, indicating that in those two TALEs both aberrant repeats always loop out simultaneously (Fig. 3.1.9). In contrast, the TALE with the two 30aa repeats in tandem activated only the optimal and the -1 p8 box, indicating that the first aberrant repeat cannot loop out. The 42aa, 40aa, and 36aa v1 repeats all showed activity on -1 and -2 boxes, however, which of the two -1 boxes was activated depended on which of the aberrant repeat variants was present (Fig. 3.1.9). The TALEs with the two 40aa and the two 36aa v1 repeats looped out the second or both aberrant repeats, whereas the TALE with the two 42aa repeats looped out the first or both (Fig. 3.1.9).

These results support a model in which the alteration of an aberrant repeat negatively affects its interaction either to the repeat placed directly before or after. For example, the duplication in the 42aa repeat variant is within the first helix of the repeat, thus likely weakening the interaction to the repeat placed directly before. In contrast, the aberrant repeat with 40aa has its duplication in the second helix, thus suggesting a compromised interaction to the following repeat. Combining these two duplications within a single repeat should therefore weaken the interactions to both adjacent repeats, and hence might result in an aberrant repeat that constantly leaves the repeat array. To investigate this hypothesis, such an aberrant repeat was artificially constructed (48aa art) and placed at position 7 (HD) within a TALE with 17.5

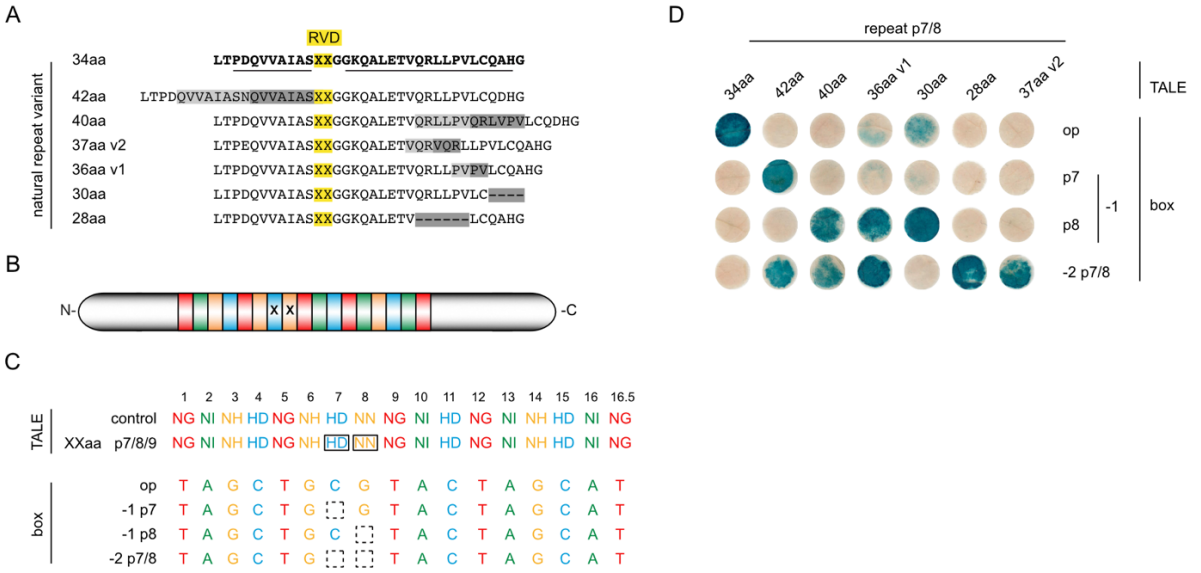


Figure 3.1.9: TALEs with different aberrant repeat variants in tandem recognize different target boxes. (A) Amino acid alignment of all aberrant repeats that can confer the recognition of a frameshift box. **(B)** Schematic view of the generated TALEs. The positions at which the aberrant repeats were inserted are indicated by an x. **(C)** RVD composition of the artificial TALEs. Position 7 and 8 (boxed) either contain a standard repeat with 34aa (top row) or one of the variants of aberrant length with 42aa, 40aa, 37aa, 36aa, 30aa or 28aa. The reporter construct consists of a TALE-box fused to the minimal *Bs4* promoter which is positioned in front of the *GUS* reporter gene. **(D)** Results of the qualitative GUS assay for the TALEs from (C). A representative leaf disc is shown for each combination. Similar results were obtained in at least two independent experiments.

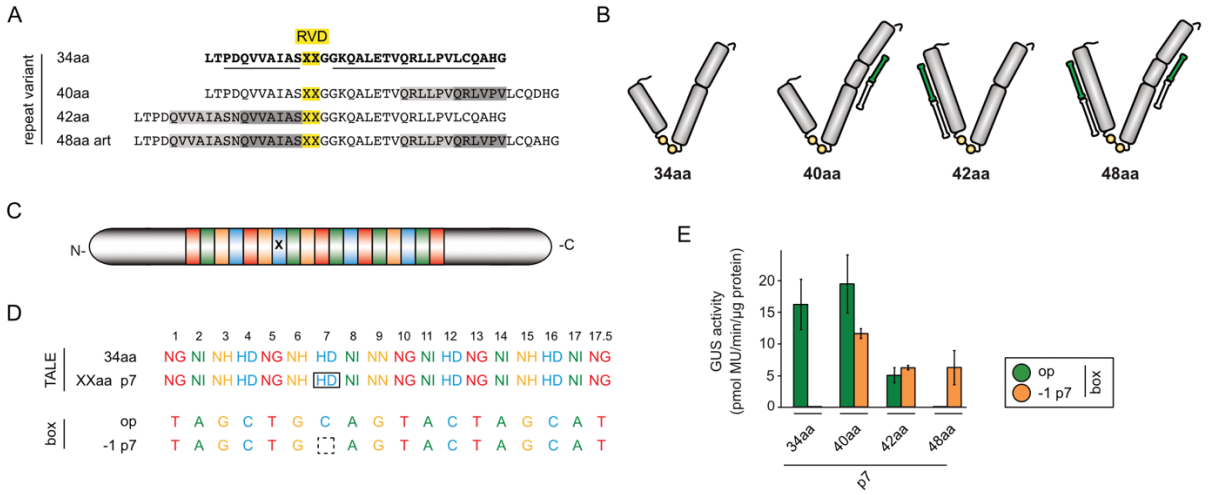


Figure 3.1.10: An artificial aberrant repeat that contains the duplications from both, the 40aa and the 42aa aberrant repeat, shows an obligatory loop out behaviour. (A) Amino acid alignment of a 34aa, the 40aa and the 42aa aberrant repeat as well as the artificially generated variant with 48aa length (48aa art). The 48aa repeat contains the duplications from the 40aa and 42aa repeat. **(B)** Cartoons for a standard repeat with 34aa, the two naturally occurring aberrant repeats with 40aa and 42aa as well as the artificially generated variant with 48aa. Shown are the two helices and the loop structure containing the RVD. All helices are elongated or shortened in accordance with the number of altered amino acids. Duplicated positions are shown green/white. **(C)** Schematic view of the generated TALEs. The altered repeat is indicated by an x. **(D)** RVD composition of the artificial TALEs. Position 7 (boxed) either contains a standard repeat with 34aa (top row) or one of the variants of aberrant length with 40aa, 42aa or 48aa. The TALE-box is either an optimal box (op) or a frameshift variant with a single nucleotide deleted at position 7 (-1 p7). **(E)** Results of the quantitative GUS assay. Error bars represent standard deviation (n = 3). Similar results were obtained in at least two independent experiments.

repeats. This TALE was tested for its activity on an optimal and a -1 frameshift box and compared to TALEs containing either no aberrant repeat, only the 40aa or only the 42aa repeat (Fig. 3.1.10). As shown previously, the control TALE activated only the optimal box while the two TALEs containing the 40aa and 42aa showed activity when combined with the optimal and the -1 frameshift box. The TALE containing the synthetic 48aa variant, however, showed only activity if combined with a -1 frameshift box, indicating a strong bias of this repeat variant to loop out of the repeat array constitutively (Fig. 3.1.10). Taken together, these data support a model in which the alteration within an aberrant repeat specifically affects the inter-repeat interactions to either one of its adjacent repeats, and thus governs its looping out behavior accordingly.

3.1.7 Two aberrant repeats only function independently at a distance

The flexible loop-out behavior of individual aberrant repeats raised the question how two aberrant repeats behave if they are separated by a varying number of standard repeats. To investigate this, the 17.5 repeat TALEs setup from chapter 3.1.4 was reused and a 40aa aberrant repeat was placed at position 7 and another one either at position 8, 9, 10, 11 or 12

(Fig. 3.1.11 A). These TALEs and a control TALE without any aberrant repeats were tested together on the previously introduced series of target boxes with 1 to 6 nucleotides deleted (Fig. 3.1.11 B). All tested TALEs were functional, but interesting differences in their behavior could be observed. Two aberrant repeats placed in direct vicinity preferentially looped out one or both aberrant repeats (Fig. 3.1.11 C). If one standard repeat of 34aa length was placed between the two aberrant repeats, the overall activity of the TALE dropped drastically on all tested boxes, indicating that this setup is very unfavorable (Fig. 3.1.11 C). Interestingly, if two standard repeats were placed between two aberrant repeats, the TALE was highly active again. Moreover, this TALE showed activity on a wide range of boxes, indicating that both aberrant repeats can either stay in (op), loop out separately (-1) or leave the repeat array simultaneously (-2) (Fig. 3.1.11 C and D). Strikingly, high activity for this TALE was also observed on the -3 and -4 boxes, a finding that can only be explained by a conformation

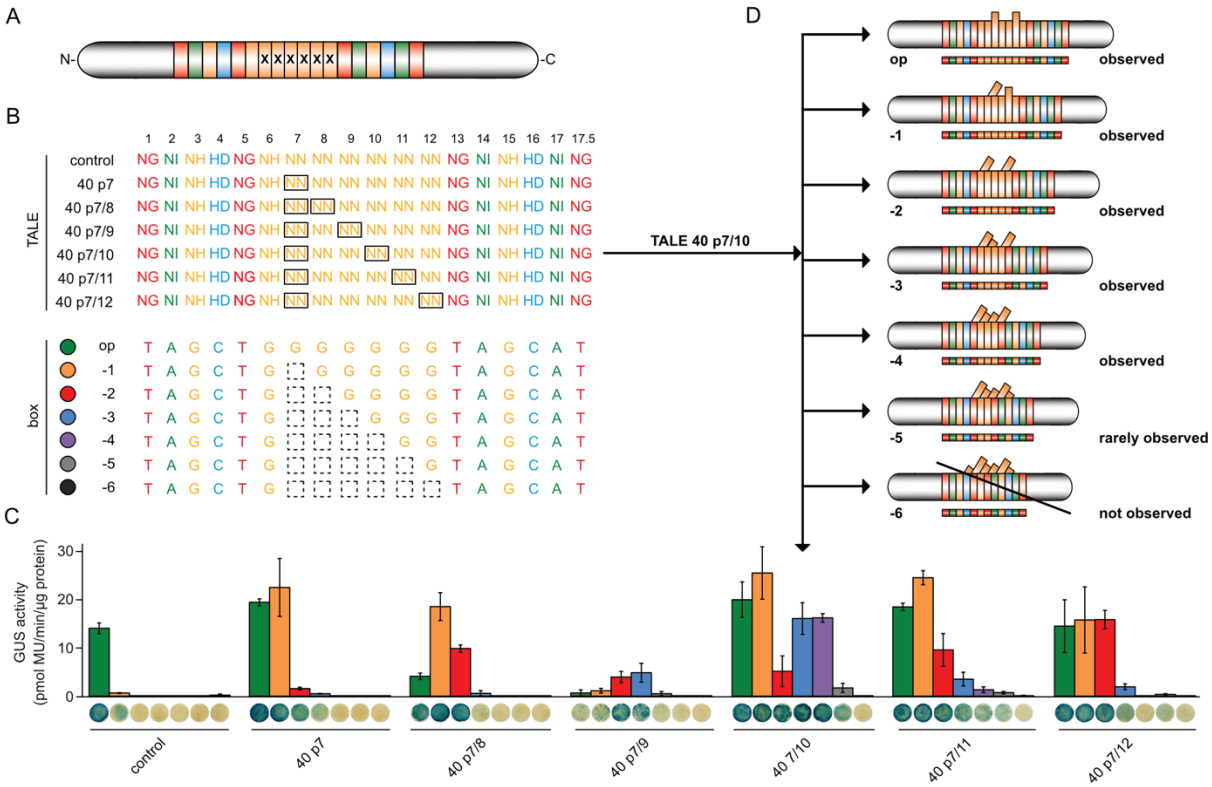


Figure 3.1.11: Two aberrant repeats need a minimal distance for functional independence. (A) Schematic view of the generated TALEs. Positions marked with an x either contain a 34aa standard repeat or the aberrant repeat variant with 40aa. **(B)** RVD composition of the generated TALEs. The region from position 7 to 12 contains only repeats with the RVD NN. Placed at position 7 is either a normal repeat or the aberrant repeat with 40aa. The aberrant repeat at position 7 was combined with a normal repeat or a second aberrant repeat at position 8, 9, 10, 11 or 12 (boxed RVDs). The TALE-box was either an optimal box (op) or a frameshift derivative with 1 to 6 nucleotides deleted (-1 to -6), deleted positions are boxed in dashed squares. **(C)** Results of the qualitative and quantitative GUS assay for the artificial TALEs described in (B). Error bars represent standard deviation (n = 3). Similar results were obtained in at least two independent experiments. **(D)** Schematic showing how TALE 40 p7/10 is able to bind to nearly all offered boxes. The models demonstrate the most likely condition of TALE 40 p7/10 contacting those boxes. Binding is achieved by either looping out none (op), one (-1) or both (-2) aberrant repeats simultaneously or by looping out the two aberrant repeats together with one or both of the normal repeats placed in between (-3 and -4).

with three or four repeats not partaking in DNA-binding. This indicates that the two aberrant repeats are able to drag out one or even both of the two standard repeats placed between them while they themselves leave the repeat array, thus leading to the extraordinary flexibility of the TALE (Fig. 3.1.11 D). It also demonstrates that a standard repeat of 34aa length can leave the repeat array without impairing the activity of a TALE. Interestingly, it reduced the overall number of inducible boxes when the number of standard repeats placed between the two aberrant repeats was increased to three or four. These TALEs had a good activity on the optimal box as well as on the -1 and -2 frameshift boxes, however, they displayed no or nearly no activity on boxes with 3, 4, 5 or 6 nucleotides deleted (Fig. 3.1.11 C). This suggests that the likelihood for a loop-out of standard repeats drops drastically, if three or more of them are placed between two aberrant repeats.

For two aberrant repeats to show an independent and flexible looping out behavior, it is therefore compulsory to position them with enough distance. In the particular setup shown here, three to four normal repeats were needed. Not doing so on the other hand will result in TALEs with an extraordinary flexibility towards a multitude of closely related but different frameshift sequences.

3.1.8 Synthetic networks – TALEs for multi-readout expression of logic genetic gates

The independent flexibility observed for multiple aberrant repeats placed within a single TALE suggested the possibility to use such TALEs as transcriptional switches in synthetic biology. To investigate this idea, a series of TALEs was designed to match different combinations of highly related target sequences. The TALEs contained 17.5 repeats of alternating specificity and combinations of two or three aberrant repeats at the positions 4, 8 and 12 (3.1.12 A-C). The TALEs were tested on a set of boxes, lacking either the base at position 4, 8, or 12, two of them in all possible combinations, or all three simultaneously (3.1.12 C). The TALEs with two aberrant repeats were only active if combined with the optimal box or those boxes in which the nucleotide deletions were placed exactly corresponding to the positions of the aberrant repeats in the respective TALE, either individually or in combination (3.1.12 D). In contrast to this, the TALE with aberrant repeats at all three positions showed only very weak activity (3.1.2 D; 40 p4/8/12).

These results indicate once more that factors such as the overall number of repeats, the number of strong RVDs and the general RVD setup of a TALE have to be considered carefully if multiple aberrant repeats are to be placed within a TALE. The results also demonstrate that the flexible frameshift recognition of aberrant repeats can be used to generate TALEs that specifically address individual subsets from a larger pool of highly related but distinct target boxes.

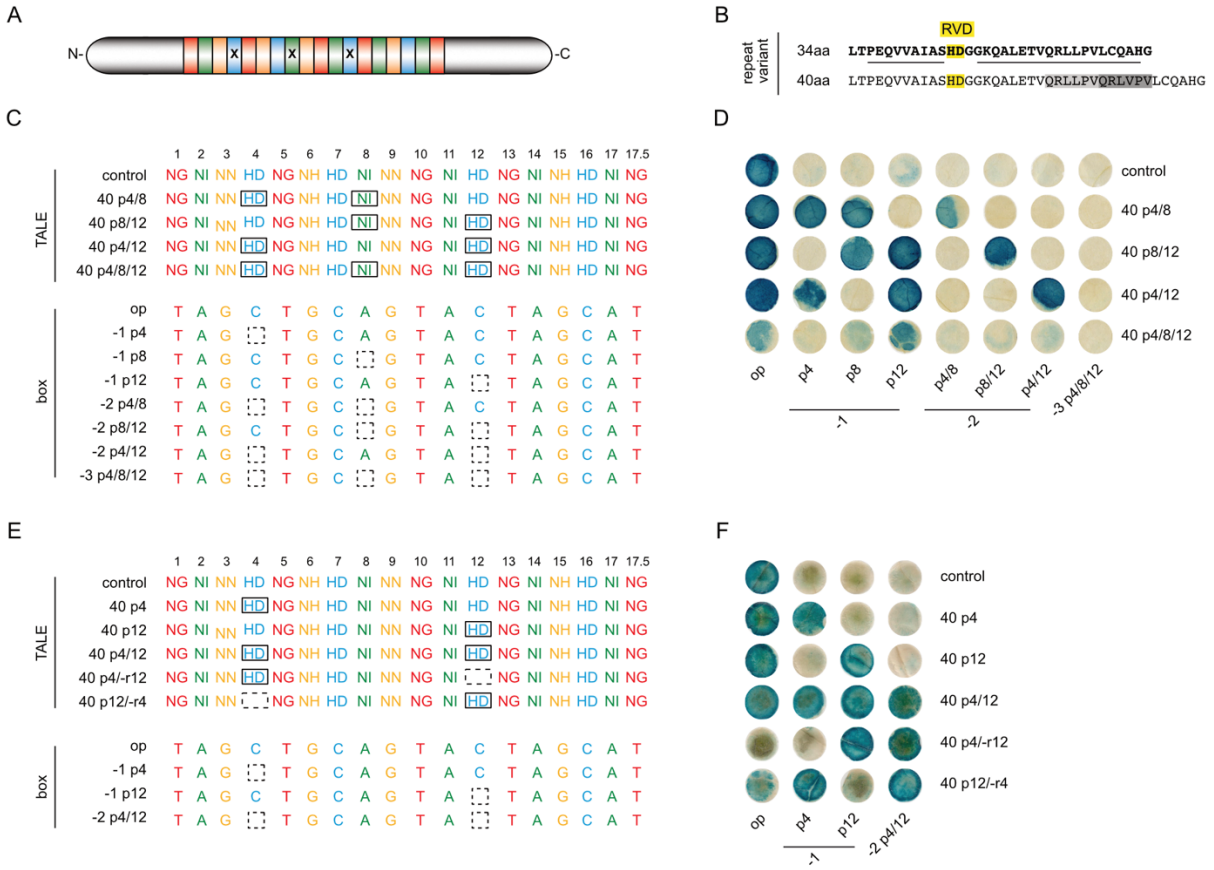


Figure 3.1.12: TALEs with two properly spaced aberrant repeats keep their specificity. (A) Schematic view of the generated TALEs. Positions marked with an x either contain a 34aa or 40aa repeat, all other positions contain 34aa repeats. **(B)** Amino acid alignment between a standard repeat of 34aa length and the 40aa aberrant repeat. **(C)** RVD composition of the artificial TALEs. Positions 4, 8 and 12 (boxed RVDs) contain either a standard repeat of 34aa length or the aberrant repeat with 40aa. The TALE-box was an optimal box (op) or a frameshift derivative with a nucleotide deletion at either position 4 (-1 p4), 8 (-1 p8) or 12 (-1 p12), or in all possible combinations (-2 p4/8, -2 p8/12, -2 p4/12 and -3 p4/8/12). Deleted positions are boxed in dashed squares. **(D)** Results of the qualitative GUS assay for the artificial TALEs shown in (C). A representative leaf disc is shown for each combination. **(E)** RVD composition of the artificial TALEs. Repeats 4 and 12 (boxed RVD) either contain a standard repeat of 34aa length, the 40aa aberrant repeat or are omitted. The TALE-box was either an optimal box (op), a frameshift variant with one nucleotide deleted at position 4 (-1 p4) or 12 (-1 p12) or one with both nucleotides deleted simultaneously (-2 p4/12). Deleted positions are indicated by dashed boxes. **(F)** Results of the qualitative GUS assay for the artificial TALEs shown in (E). A representative leaf disc is shown for each combination.

To further investigate this possibility, new TALEs were generated, containing either a single aberrant repeat at position 4 or 12, an aberrant repeat at both positions or no aberrant repeat at all. In addition, the repeat at position 4 or 12 was omitted in two TALEs (3.1.12 E). These TALEs were tested on an optimal box and on frameshift derivatives lacking the base at either position 4, 12, or both simultaneously (3.1.12 E). Depending on their composition, these TALEs allowed for the activation of individual boxes, as well as combinations of two or four of them (3.1.12 F).

These results confirm once more that it is possible to design TALEs that address several highly related but distinct target boxes individually as well as in various combinations. TALEs were already shown to be a versatile tool for the generation of synthetic genetic circuits

(Lebar and Jerala, 2016). However, using such a set of closely related, aberrant repeat-carrying TALEs as master switches while at the same time inserting the different target boxes as control elements into target promoters, could allow for an even more potent TALE-based regulation of genetic circuits with multi-readout expression.

3.1.9 Natural flexibility – TalBK2 contains two aberrant repeats and binds to multiple frameshift sequences

The existence of natural TALEs with two aberrant repeats prompted the question whether they are functional and capable of inducing gene expression. The two members of class TalBK are the first known examples for natural TALEs carrying multiple aberrant repeats and only one additional TALE with two aberrant repeats was identified – found within the recently sequenced but still unpublished genome of the *Xoo* strain PXO35. Due to a striking overall similarity (Appendix Table 3.1.1), this TALE was temporarily assigned to class TalBK, and thus termed tempTalBK3. All three TALEs share the aberrant repeat variant 36aa v1 at the positions 9 and 12, containing the RVDs HD and NS, respectively (Fig. 3.1.13 A and C).

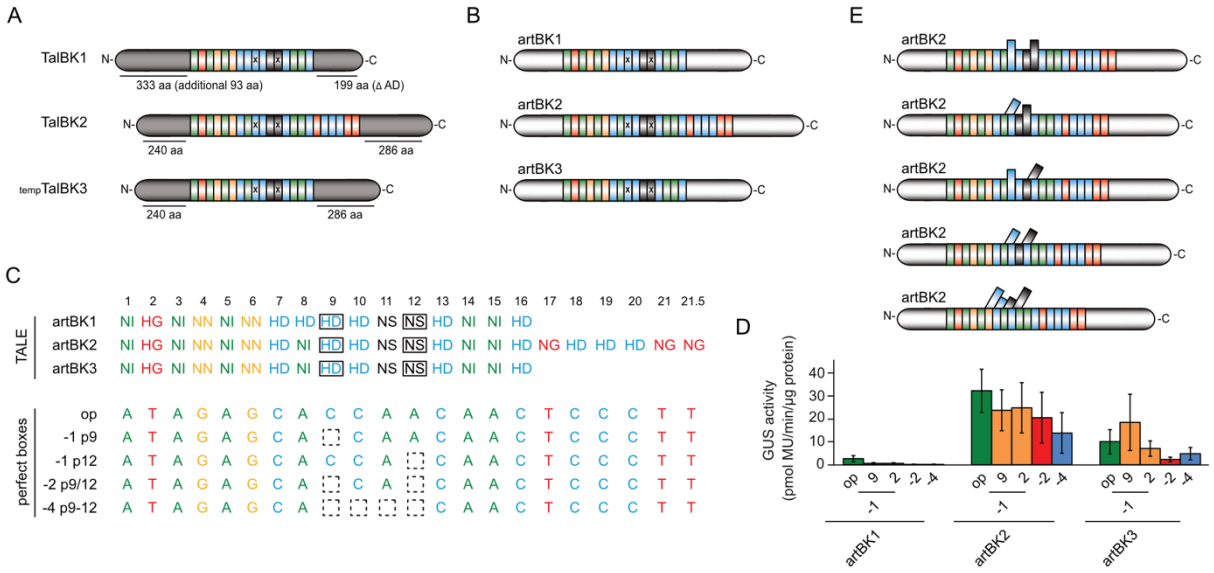


Figure 3.1.13: ArtBK2 is capable to address several different frameshift boxes. (A) Schematic view of the three members of the class TalBK in their natural form. TalBK1 lacks 87aa in its C-terminal region, including the activation domain, while its N-terminal region is 93aa longer than those found in TalBK2 and tempTalBK3. **(B)** Schematic view of the three artificially generated TALEs artBK1, artBK2 and artBK3. All three contain the same RVDs as their natural counterparts but are constructed using the N- and C-terminal regions from Hax3. **(C)** RVD composition of the artificial TALEs artBK1, artBK2 and artBK3. Positions 9 and 12 contain the 36aa v1 aberrant repeat (boxed positions). ArtBK1 and artBK3 are 6 repeats shorter than artBK2. ArtBK1 has an additional RVD change at position 8 (NI-> HD). The reporter constructs consist of a TALE-box fused to the minimal *Bs4* promoter which is positioned in front of the *GUS* reporter gene. The TALE-box was either a predicted optimal box (op) for TalBK2/artBK2 or derivatives of this box containing either one or multiple nucleotide deletions at the indicated positions (-1 p9, -1 p12, -2 p9/12 and -4 p9-12). **(D)** Results of the GUS assay for artBK1, artBK2 and artBK3. Error bars represent standard deviation (n = 3). Similar results were obtained in at least two independent experiments. **(E)** Schematic view of all likely binding confirmations of TalBK2/artBK2 by looping out either none, one or both aberrant repeats or a combination of the two aberrant repeats and the two repeats in between.

While their general RVD setup is very similar, TalBK2 is six repeats longer than its two class members. TalBK1 on the other hand contains the RVD HD at position 8 while its two class members share the RVD NI at this position. The CDS of TalBK1 is also N-terminally prolonged, containing 93 additional amino acids, and it has a frameshift mutation within its C-terminal region (Fig. 3.1.13 A). Since this frameshift mutation removes the activation domain of the protein, TalBK1 is most likely incapable of gene activation. The CDS of TalBK2 and tempTalBK3 hand contain N- and C-terminal regions of standard length, and thus they should yield, at least theoretically, functional proteins (Fig. 3.1.13 A).

To analyse the DNA-binding capabilities of the three class TalBK members, artificial TALEs containing the same RVDs were constructed and named artBK1-3 (Fig. 3.1.13 B). A set of different target boxes was generated based on the RVDs present in TalBK2. Since the aberrant repeat 36aa v1 was capable of conferring frameshift recognition (see chapter 3.1.2), this set included not only an optimal target box but also several different frameshift derivatives (Fig. 3.1.13 C). Testing the three TALEs on this set of reporter boxes revealed that TALE artBK2 is highly active on the optimal box as well as on frameshift derivatives carrying nucleotide deletions at either one or both positions corresponding to the aberrant repeats (Fig. 3.1.13 D and E). Intriguingly, artBK2 even showed activity when combined with a box missing the four nucleotides from positions 9 to 12, a box that can only be addressed if the two aberrant repeats loop out together with the two standard repeats found in-between (Fig. 3.1.13 C-E; Appendix Fig. 3.1.1 B). ArtBK3 was also able to induce all five boxes but to a much lesser extent. The last TALE, artBK1, activated only the optimal box and the observed activity was extremely low (Fig. 3.1.12 D and E). One reason for the lower activity of artBK1 and 3 might lie in their reduced number of repeats since it could decrease their overall DNA-binding affinity, and thus render them more vulnerable to the inter- and intra-repeat disturbing effects of the two aberrant repeats (Appendix Fig. 3.1.1 A and C). In addition, artBK1 is even further impaired by a mismatch at position 8.

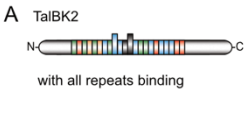
The results from this experiment indicate that at least two members of class TalBK, namely TalBK2 and tempTalBK3, are very likely active *in vivo*. TalBK2 might even display a target flexibility yet unseen in any other natural TALE.

3.1.10 An arms race – TALE classes TalBK and TalAM activate different *OsSWEET13* promoter variants

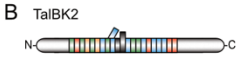
Demonstrating that two of the three TalBK class members are likely active *in vivo* prompted the question whether or not it is possible to identify their natural target gene(s). The “predict and intersect” function of AnnoTALE (Grau *et al*., 2016) was used to screen the promoterome of the rice cultivar Nipponbare for potential TalBK2 target sites. A target

prediction with all repeats participating in DNA binding yielded no promising candidates and, similarly, no good target genes were predicted when a simultaneous loop out of both aberrant repeats or a loop out of the two aberrant repeats and the two standard repeats in between was assumed (Fig. 3.1.14). However, when a confirmation was assumed in which only one of the aberrant repeats loops out, the well-known virulence target *OsSWEET13* ranked at the top position, moreover, with a very good score and a perfectly matching target site (Fig. 3.1.14).

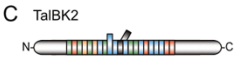
To evaluate this putative target experimentally, a 1 kbp fragment of the *OsSWEET13* promoter was amplified from the rice cv. Nipponbare, cloned in front of the *GUS* reporter gene and transformed into *A. tumefaciens.* The subsequent analysis in *N. benthamiana* confirmed that artBK2 is indeed able to activate the promoter of *OsSWEET13*, indicating that this gene is very likely the natural target of TalBK2 (Fig. 3.1.15). Interestingly, the binding site of TalBK2 within the *OsSWEET13* promoter coincides with the position of several single



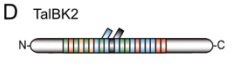
#	gene	function	position	score	target sequence/ mismatches
1	*Os02g47520.1*	expressed protein	193	-16.328	TATAGAGCATCAACAACACCATA
2	*Os03g18454.2*	gamma-interferon-inducible lysosomal reductase	308	-16.582	TAAAATGCACCCACACCTCCCCT
3	*Os06g48350.1*	CPuORF14 - conserved peptide uORF	83	-17.525	TCAATCACACCATCAACTCCCTT
4	*Os05g05580.1*	expressed protein	311	-17.731	CACAAACCACCACCACCTCCCAT
5	*Os10g37740.1*	CGMC_GSK.9 -CGMC includes CDA, MAPK, ...	264	-17.887	TCTCCACCACCACCACCTCCCCT



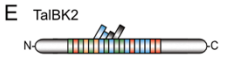
#	gene	function	position	score	target sequence/ mismatches
1	*Os12g29220.1*	nodulin MtN3 family protein (*OsSWEET13*)	267	-7.653	TATAAAGCACCACAACTCCCTT
2	*Os03g20090.1*	MYB family transcription factor	365	-15.959	TATAAACCCCGACACCTCCCTC
3	*Os06g24960.1*	retrotransposon protein, putative	215	-16.103	TTCACATCAAAACAACTCCCTT
4	*Os03g44360.1*	retrotransposon protein, putative	215	-16.103	TTCACATCAAAACAACTCCCTT
5	*Os11g23244.1*	retrotransposon protein, putative	215	-16.103	TTCACATCAAAACAACTCCCTT



#	gene	function	position	score	target sequence/ mismatches
1	*Os12g29220.1*	nodulin MtN3 family protein *OsSWEET13*	267	-6.250	TATAAAGCACCACAACTCCCTT
2	*Os02g03750.1*	polygalacturonase, putative, expressed	308	-14.889	CATAGAACACCACACCTCCATC
3	*Os04g54790.1*	ELMO/CED-12 family protein	283	-15.287	TAGAAACCACCACCACTCTCTT
4	*Os10g30440.1*	expressed protein	222	-15.649	TATAAAACACCACTCCTTCCTT
5	*Os01g51030.1*	expressed protein	265	-15.930	TATAAAACACCCAAACTCCCCG



#	gene	function	position	score	target sequence/ mismatches
1	*Os05g24230.1*	hypothetical protein	142	-15.177	TATAAACCACAAAACTACCTA
2	*Os01g42860.1*	inhibitor I family protein, putative ,expressed	269	-15.361	AATACACCACACAACTCCATC
3	*Os11g13710.1*	expressed protein	189	-15.648	TATAAACCACCAAACACCCTC
4	*Os01g05220.1*	expressed protein	210	-15.859	TCTCCAGCAACCAACTCCCTC
5	*Os11g05690.3*	amino acid permase family protein	44	-15.900	TCAACAGTACAAAACTCCCTT



#	gene	function	position	score	target sequence/ mismatches
1	*Os10g32050.1*	anykrin repeat domain containing protein	20	-9.587	TATAGAGTACAACTCCCTC
2	*Os06g23300.1*	expressed protein	225	-10.442	CATCGAGCACAACTCCCTC
3	*Os06g32960.1*	peroxidase precursor, putative	154	-11.881	TATAAAGCACACCTCCGTT
4	*Os09g19480.1*	retrotransposon, putative	67	-12.938	TATCCAGCCCAACTCCCTG
5	*Os04g46650.1*	expansin precursor, puative	15	-13.153	TAACCAACACAACTACCTT

Figure 3.1.14: Target gene prediction for TalBK2. Shown are the five best putative target genes predicted for TalBK2 under the condition that **(A)** all repeats participate in DNA binding, **(B)** the aberrant repeat at position 9 leaves the repeat array, **(C)** the aberrant repeat at position 12 leaves the repeat array, **(D)** both aberrant repeats (positions 9 and 12) leave the repeat array simultaneously or **(E)** all four repeats from position 9 to 12 leave the repeat array. A set of 1000 bp long sequences derived from all promoter regions found in the rice cv. Nipponbare was used as basis for the prediction. The prediction analysis was done using AnnoTALE (Grau *et al.*, 2016). Mismatched positions are coloured red.

nucleotide INDELs (Fig. 3.1.15 A; Richter *et al*., 2014; Zhou *et al*., 2015). Due to their extremely regular DNA-binding mechanisms, TALEs are usually very sensitive towards INDELs. As a result, even promoter variants with single nucleotide INDELs can confer resistance to TALEs by preventing their binding and thereby blocking the activation of key susceptibility genes (Chu *et al*., 2006; Yang *et al*., 2006; Liu *et al*.,2011; Richter *et al*., 2014; Zhou *et al*., 2015). However, the results from the previous chapter demonstrated that the two aberrant repeats found in TalBK2 facilitate the flexible recognition of multiple frameshift sequences and, consequently, might enable this TALE to address several of these naturally occurring promoter variants (Fig. 3.1.16). To investigate this, *OsSWEET13* promoter fragments were cloned from two rice cultivars known to contain differences in this particular region, Zhenshan and IR24. Additionally, *OsSWEET13* promoter fragments from five other rice cultivars were amplified and cloned and one novel variant was isolated, originating in the rice cv. Sadu Cho (Appendix Table 3.1.5). This new variant is nearly identical to the *OsSWEET13* promoter fragment from the cv. Nipponbare, however, it contains a single nucleotide polymorphism corresponding to the very first position of the TalBK2 target site that generates a mismatched RVD-nucleotide combination for repeat 1 (Fig. 3.1.15 A).

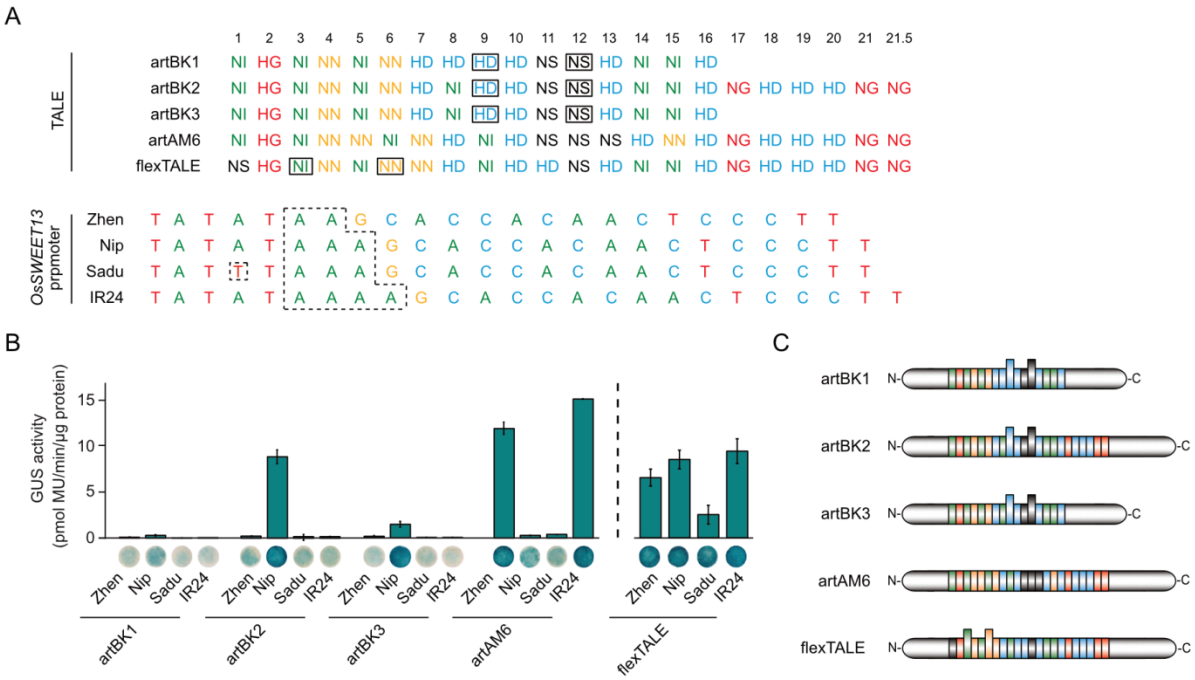


Figure 3.1.15: The TALEs artBK2, artBK3 and artAM6 bind and induce the promoter of the *OsSWEET13* gene. (A) RVD composition of artBK1, artBK2, artBK3, artAM6 and the flexTALE. Boxed positions contain the aberrant repeat variant 36aa v1. A 1 kbp fragment of the native *OsSWEET13* promoter was amplified and placed in front of the *GUS* reporter gene. Fragments from four different rice cultivars were analysed, representing the known variation within the specific part of the promoter containing the TALE binding sites. Those rice cultivars are Zhenshan (Zhen), Nipponbare (Nip), Sadu Cho (Sadu) and IR24 (IR24). Shown is only the region containing the putative binding sites for the analysed TALEs. **(B)** Results of the qualitative and quantitative GUS assay for artBK1, artBK2, artBK3, artAM6 and the flexTALE if combined with the four different *OsSWEET13* promoter variants. Error bars represent standard deviation (n = 3). Similar results were obtained in at least two independent experiments. A representative leaf disc is shown for each combination. **(C)** Schematic view of the TALEs used in this experiment. The N- and C-terminal regions are derived from Hax3, aberrant repeats are depicted elongated.

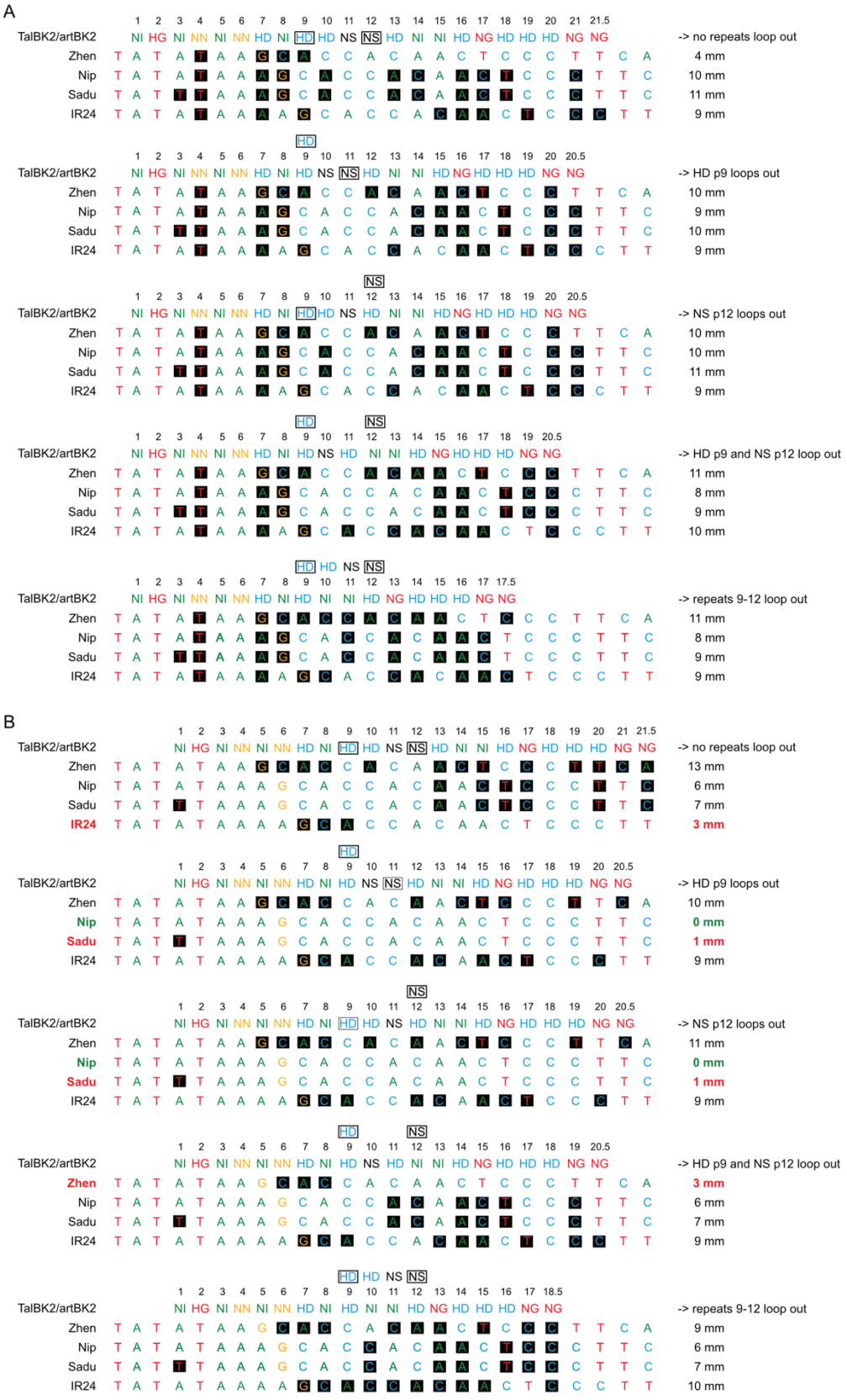


Figure 3.1.16: Binding possibilities for TalBK2/artBK2 in the four known *OsSWEET13* promoter variants. The TALE is shown to use **(A)** its first or **(B)** its second possible binding frame. Boxed positions contain the aberrant repeat variant 36aa v1. Mismatched positions within the putative binding site are shaded black. The lowest number of mismatches for each rice cultivar is indicated in bold. If the promoter fragment from this cultivar was activated in the GUS reporter assay, it is coloured in green, if not, it is coloured in red.

Surprisingly, the subsequent GUS reporter assay with these three promoter variants revealed that artBK2 can recognize only the promoter fragment derived from cv. Nipponbare (Fig. 3.1.15). Despite an increased frameshift tolerance provided by the presence of two aberrant repeats, artBK2 was unable to overcome the resistance mechanism in this promoter. One reason for this might lie in the fact that the alterations in the three other promoter variants correspond not exactly to the positions of the two aberrant repeats, and thus additional mismatches occur. Even using the best possible binding frame, the artBK2 binding sites in the two *OsSWEET13* promoters from Zhenshan and IR24 contain at least three mismatches (Fig. 3.1.16). While several natural TALEs with 17.5 repeats successfully address their targets with a similar number of mismatches (see chapter 3.2.3. and 3.2.4.), it is likely that the two aberrant repeats in TalBK2 reduce its overall DNA-affinity to some extent, and thus render this particular TALE unable to do so. Interestingly, despite containing only a single mismatched position, not even the promoter fragment from Sadu Cho was induced. However, this single mismatch corresponds to the very first repeat (Fig. 3.1.16) and mismatches within the first six repeats of a TALE were previously shown to affect its activating capabilities much more severe than mismatches in repeats at other positions (Meckler *et al*., 2013). It seems therefore likely that this single, prominently placed mismatch is sufficient to prevent artBK2 from activating this particular promoter fragment. ArtBK1 and artBK3 were also tested on the four promoter variants. ArtBK3 induced, like artBK2, only the variant from Nipponbare but it did so to a much lower extend (Fig. 3.1.15) while artBK3 showed no activity on any of tested reporter constructs (Fig. 3.1.15). The reason for their lower overall activity lies likely in their reduced number of repeats and, in case of artBK1, in an additional mismatch at position 8 (Appendix Fig. 3.1.2 and 3.1.3).

Interestingly, the same position in the *OsSWEET13* promoter is also targeted by another TALE class, TalAM (most prominent member: PthXo2; Richter *et al*., 2014; Zhou *et al*., 2015). In contrast to TalBK, members of TalAM do not contain any aberrant repeats, yet, they are reported to address *OsSWEET13* in two different rice cultivars, Zhenshan and IR24, by using two binding frames that are two nucleotides apart (Fig. 3.1.17). To investigate the ability of class TalAM to address the different *OsSWEET13* promoter variants, an artificial variant of a representative member, TalAM6, was constructed and termed artAM6. In accordance with published reports, the respective GUS analysis confirmed that artAM6 induces indeed the *OsSWEET13* promoter variants from the cultivars Zhenshan and IR24 but not the one from Nipponbare (Fig. 3.1.15). Interestingly, the TALE was also unable to induce the promoter fragment derived from the cv. Sadu Cho, indicating that the newly identified *OsSWEET13* allele from Sadu Cho could be a valuable resource for *Xoo* resistance in rice breeding.

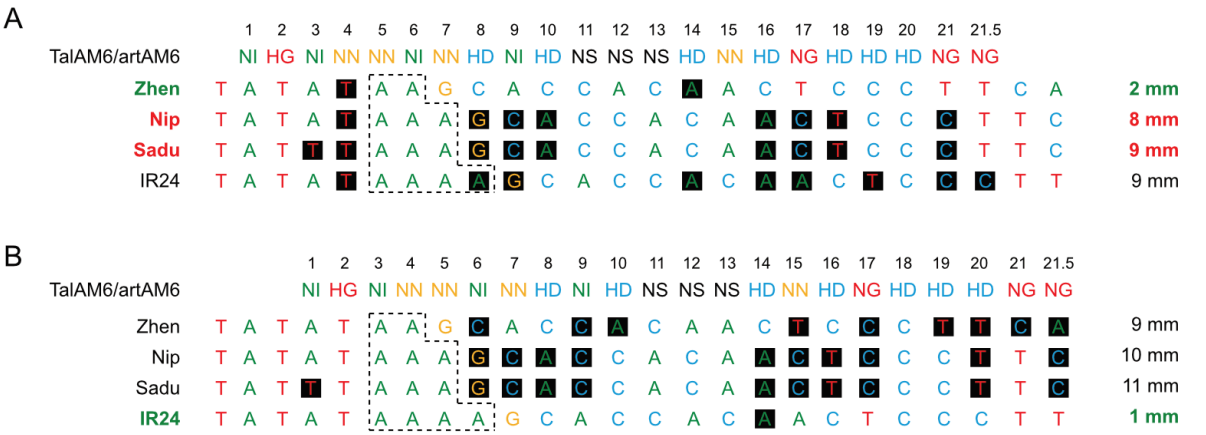


Figure 3.1.17: Binding possibilities for TalAM6/artAM6 in the four known *OsSWEET13* promoter variants. TalAM6/artAM6 is shown to use **(A)** its first or **(B)** its second possible binding frame. Mismatched positions within the target box are indicated by a black square. The lowest number of mismatches for each rice cultivar is indicated in bold. If the promoter fragment from this cultivar was activated in the GUS reporter assay, it is coloured in green, if not, it is coloured in red.

To analyse the impact of these TALEs in a more natural environment, artAM6 and artBK2 variants suitable for expression in *Xanthomonas* were generated and brought into the *Xoo* strain BAI3 Δ*talC*. This strain is a less virulent derivative of the *Xoo* strain BAI3 that lacks the major virulence factor TalC. It was demonstrated that the virulence of this strain can be restored by the introduction of an *OsSWEET*-inducing TALE (Yu *et al*., 2011; Streubel *et al*., 2013). The infection experiment confirmed the results obtained in the GUS assay, further supporting the hypothesis that *OsSWEET13* is indeed the natural target of TalBK2 but only in the cultivar Nipponbare (Fig. 3.1.18). The experiment also demonstrated that a BAI3 Δ*talC* strain containing artAM6 becomes virulent on the two rice cultivars Zhenshan and IR24 but not on Nipponbare or Sadu Cho (Fig. 3.1.18). Taken together these experiments demonstrate how the arms race between host and pathogen gives rise to promoter variations and TALE variants alike and how they can govern the pathogenicity and fate of individual *Xoo* strains.

In order to highlight the capabilities of TALEs as potent virulence factors, a highly flexible TALE (termed flexTALE) was designed to match and address all four known *OsSWEET13* promoter variants. For this, TalBK2 was used as the basis. Its two aberrant repeats (36aa v1) were moved to positions 3 and 6 and the repeat at position 1 was changed to contain an unspecific RVD (NS) (Fig. 3.1.15 and 3.1.19). In a subsequent GUS analysis, this flexTALE was active on all promoter variants (Fig. 3.1.15), demonstrating that TALEs with multiple aberrant repeats are indeed able to recognize several allelic variants of the same promoter. This proves that a combination of unspecific RVDs and properly positioned aberrant repeats can be used to generate TALEs with an unprecedented target flexibility.

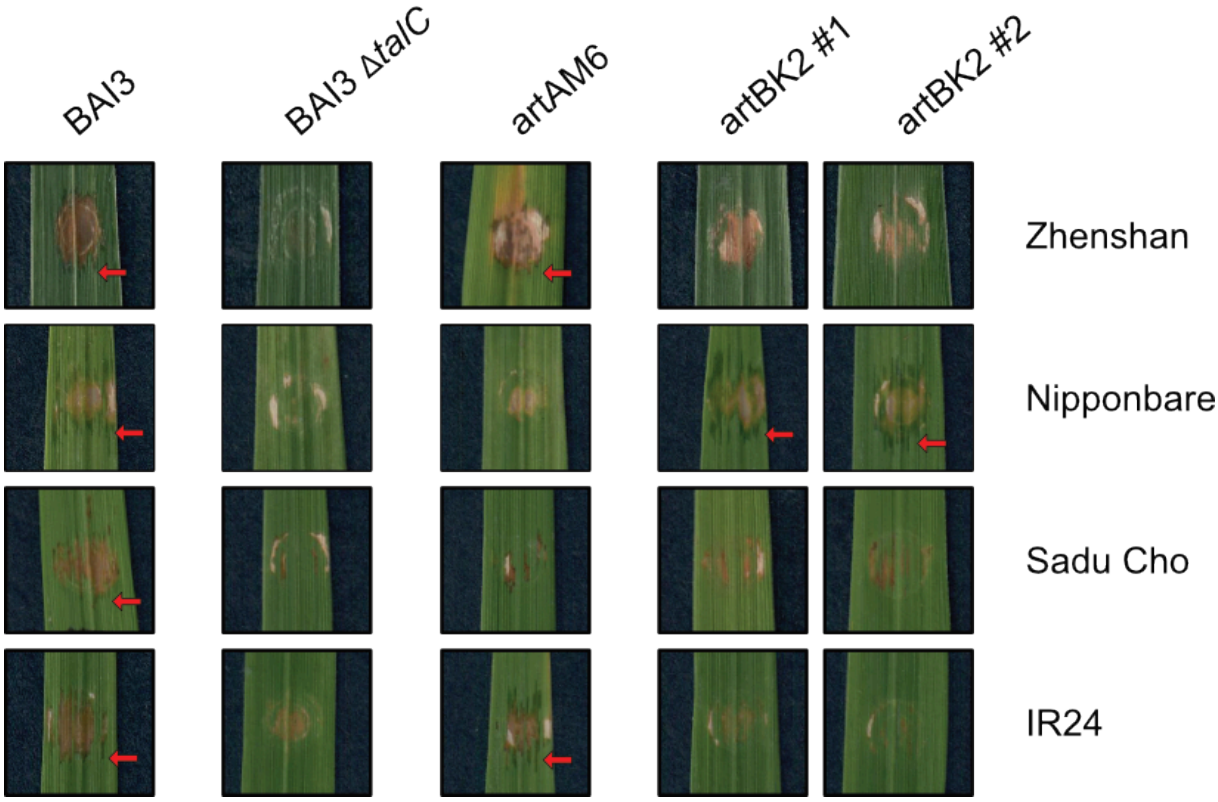


Figure 3.1.18: ArtBK2 and artAM6 lead to virulence on different rice cultivars. Disease phenotypes of the *Xoo* strain BAI3 and derivatives after inoculation on rice leaves. Leaves of 4 week-old rice plants were inoculated with the wildtype strain BAI3, the avirulent mutant strain BAI3 Δ*talC*, or the mutant strain containing either a construct expressing artAM6 or artBK2 (two different transformants). All strains were inoculated on the four rice cv. Zhenshan, Nipponbare, Sadu Cho and IR24. The positive control, the wild type strain BAI3, is not affected by the promoter variations in the *OsSSWET3* gene since its major TALE, TalC, activates a different *OsSWEET* gene, namely *OsSWEET14*. The virulence of the strains containing only TalAM6 or artBK2 on the other hand directly correlates with the presence of specific *OsSWEET13* promoter variants. The strain with TalAM6 can induce water soaking lesions only in Zhenshan and IR24 since only those two rice cultivars contain an adequate binding site for TalAM6 in their *OsSWEET13* promoter. The strain with TalBK2, in contrast, shows symptoms only in the cultivar Nipponbare, demonstrating that BK2 can address indeed only the *OsSWEET13* promoter allele in this cultivar. The experiment was performed only once.

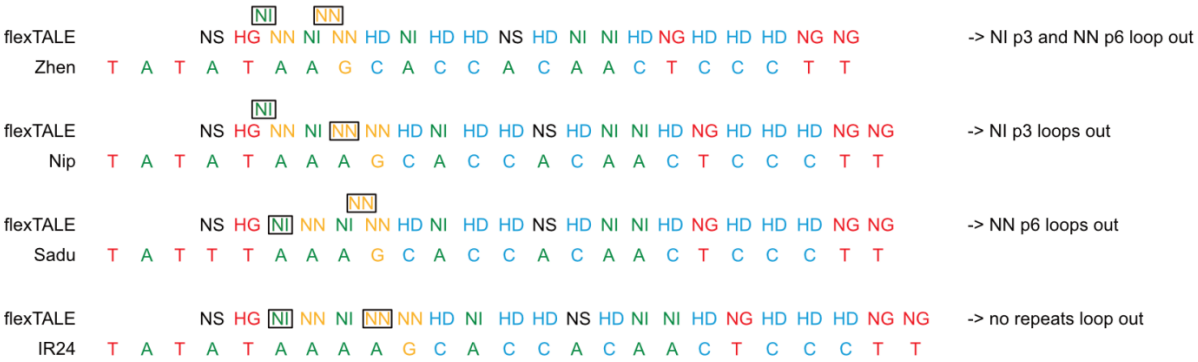


Figure 3.1.19: Binding possibilities for the flexTALE in the four known *OsSWEET13* promoter variants. By containing two properly positioned aberrant repeats and an unspecific RVD within the first repeat, the flexTALE is designed to address all four *OsSWEET*13 promoter variants without any mismatches. Boxed positions contain the aberrant repeat variant 36aa v1.

3.1.11 Many of the TALEs carrying the 40N* repeat target the *OsSWEET14* promoter

Four TALEs/TALE classes are known to target *OsSWEET14*: TalC and Tal5 from the African Xoo strains BAI3 and MAI1, respectively, and the two TALE classes TalAC and TalBH, which are both found in numerous *Xoo* strains from Asia (Yang and White, 2004; Antony *et al*., 2010; Yu *et al*., 2011; Streubel *et al*., 2013). Incidentally, all members found within the latter two TALE classes carry the 40N* aberrant repeat variant. However, this aberrant repeat variant is not limited to these two classes, it also occurs in four other TALE classes from Asian *Xoo*, namely TalDS from PXO211, TalDV from PXO563 and PXO602, TalEQ from XF89b and Tal2 from K-74 (Fig. 3.1.20). Interestingly, these TALEs are all located at similar genomic positions, they never occur within the same strain (Fig. 3.1.2) and they all have a striking resemblance to either TalAC or TalBH (Fig. 3.1.20). Consequently, a shared evolutionary origin seems likely. Some publications already speculated about the possibility

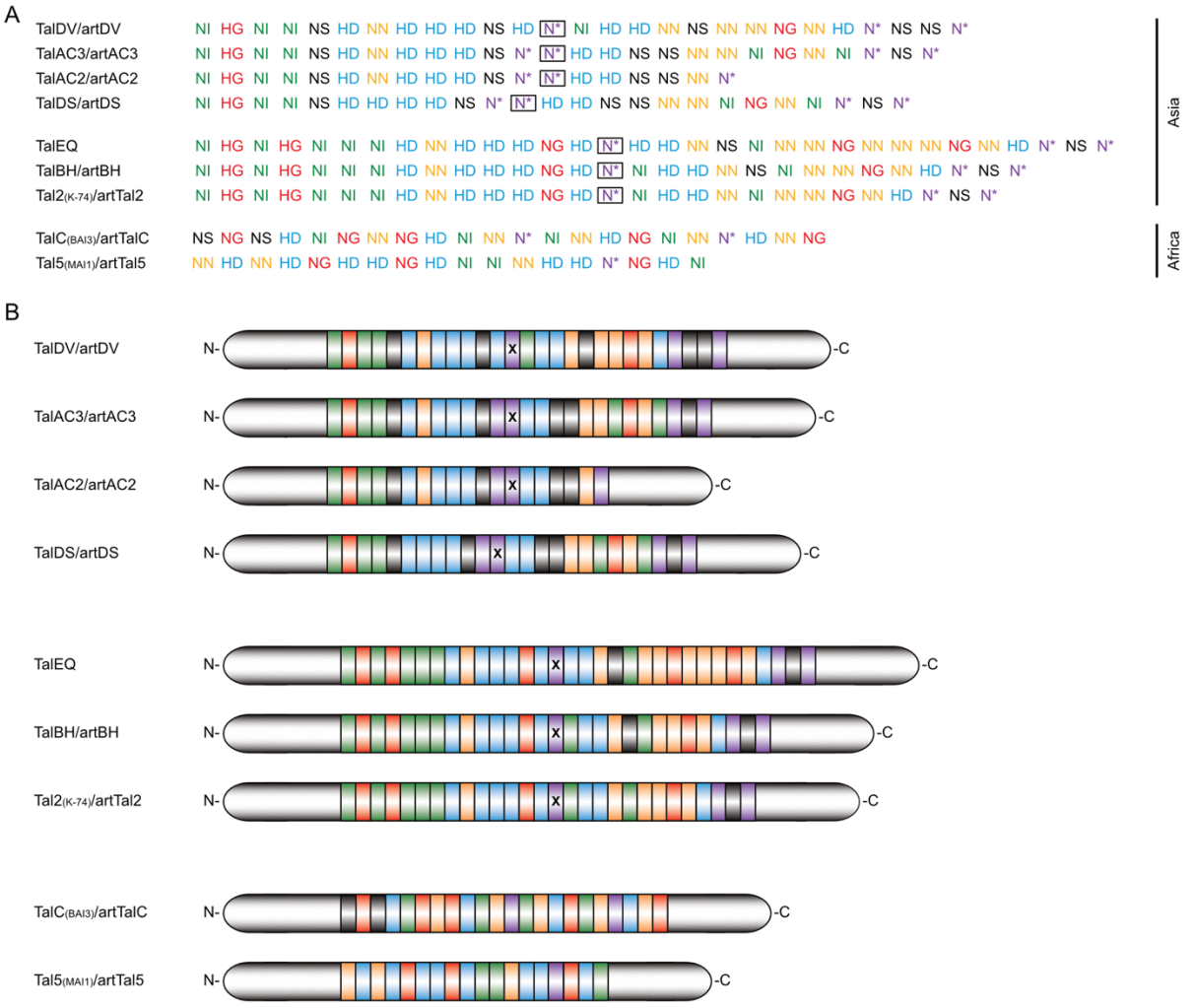


Figure 3.1.20: *Xoo* TALEs/TALE classes that carry the 40aa N* aberrant repeat and/or are known to induce *OsSWEET14*. **(A)** RVD composition and **(B)** schematic view of the four TALEs/TALE classes known to induce *OsSWEET14* expression (TalAC3, TalBH, TalC and Tal5) as well as the four TALE classes that carry the 40N* aberrant repeat, but which have not been analysed so far. The aberrant repeat is underlined or indicated by an x. For all TALEs not yet included into the AnnoTALE nomenclature, the published names and the strains they originate in are indicated.

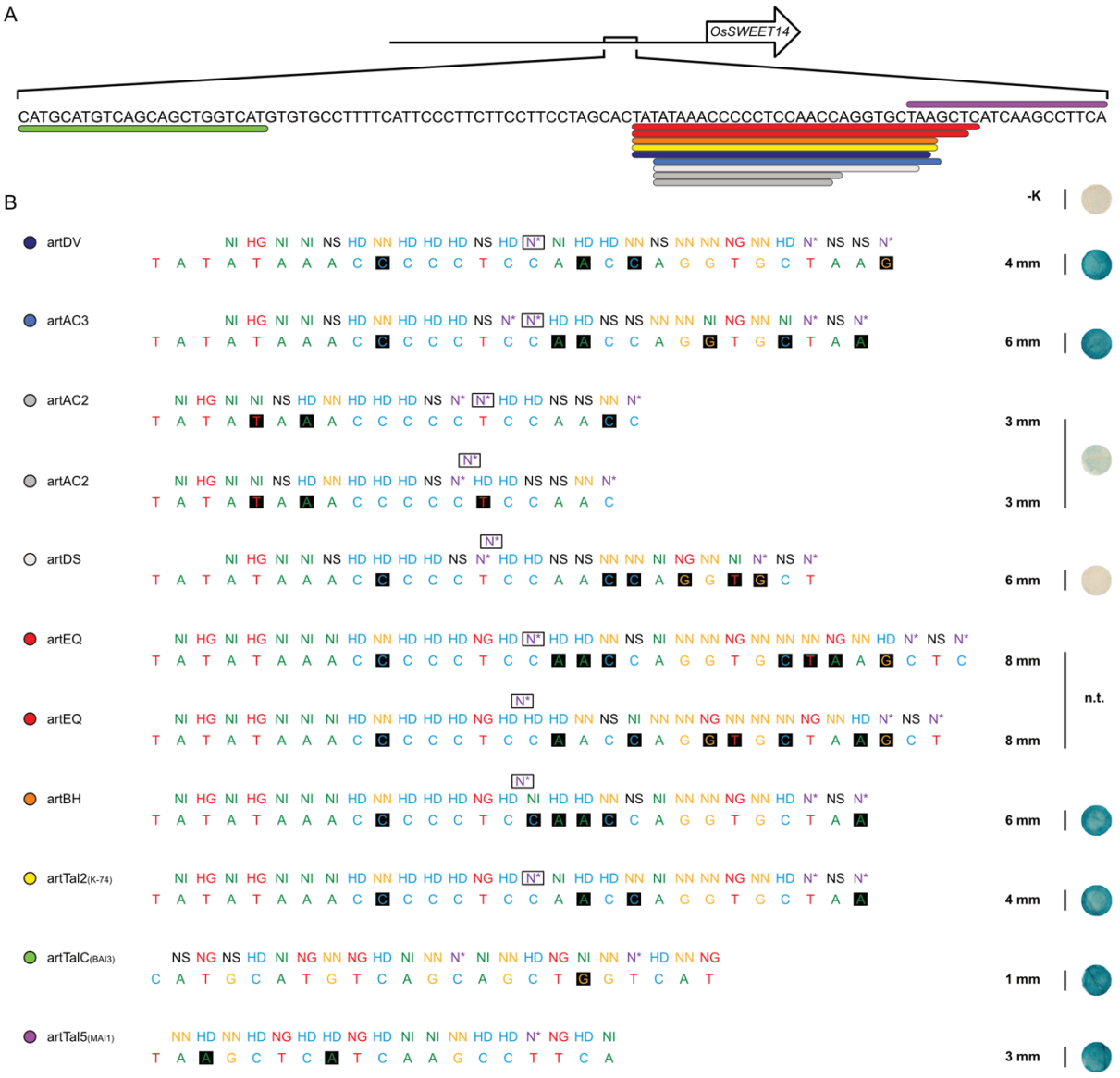


Figure 3.1.21: Best binding options in the *OsSWEET14* promoter for all TALEs/TALE classes that either carry the 40aa N* aberrant repeat and/or are known to induce *OsSWEET14*. (A) The *OsSWEET14* promoter is addressed by three largely independent TALE binding sites, two of which partially overlap. The colours correspond to the different TALEs/TALE classes. **(B)** RVD composition of all TALEs that carry the 40N* aberrant repeat variant and/or are known to address the *OsSWEET14* promoter. The best possible binding option yielding the least mismatches is indicated, for a full analysis see Appendix Fig. 3.1.4 and 3.1.5. Mismatched positions are shaded in black and their total number is indicated (mm). The results of the qualitative GUS analysis are indicated by a representative leaf disc. TalEQ was not tested (n.t.).

that these TALEs are also involved in *OsSWEET14* induction (Yu *et al*., 2015; Quibod *et al*., 2016) but, curiously, none of these studies actually analysed whether these TALEs do induce *OsSWEET14*.

In order to investigate their *OsSWEET14*-inducing capabilities experimentally, artificial variants were successfully generated for three of these four TALEs/TALE classes, namely artDS, artDV and artTal2. As positive control, artificial versions of the known *OsSWEET14*-inducing TALEs were generated as well; namely artAC3, artBH, artTalC and artTal5. The

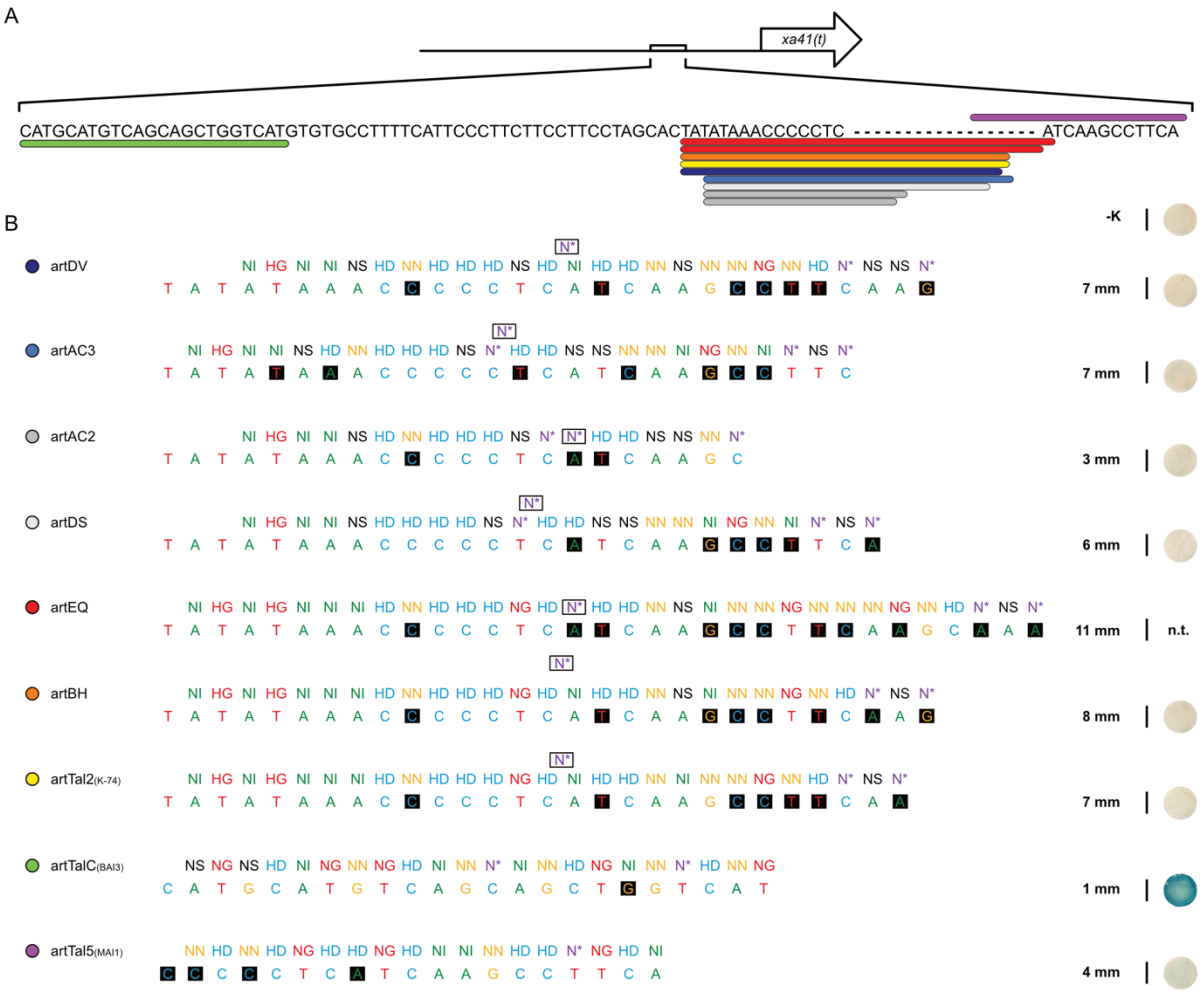


Figure 3.1.22: Best binding options in the *OsSWEET14* promoter variant *xa41(t)* for all TALEs/TALE classes that either carry the 40aa N* aberrant repeat and/or are known to induce *OsSWEET14*. (A) The 18 bp deletion found in the *xa41(t)* promoter sequence disrupts the binding site of all but one of the *OsSWEET14*-inducing TALEs/TALE classes. **(B)** RVD composition of all TALEs that carry the 40N* aberrant repeat variant and/or are known to address the *OsSWEET14* promoter. The best possible binding option yielding the least mismatches is indicated, for a full analysis see Appendix Fig 3.1.6 and 3.1.7. Mismatches are shaded in black and their total number is indicated (mm). The results of the qualitative GUS analysis are indicated by a representative leaf disc. TalEQ was not tested (n.t.). The results demonstrate that *xa41(t)* confers resistance against all but one of the TALEs/TALE classes that address *OsSWEET14*.

resulting constructs were transformed into *A. tumefaciens* and investigated in combination with a 1 kbp promoter fragment of the *OsSWEET14* gene in a GUS reporter assay. The four TALEs/TALE classes previously reported to address *OsSWEET14* caused indeed an induction of the reporter construct and likewise did artDV and artTal2, strongly suggesting that *OsSWEET14* is also their natural target gene (Fig. 3.1.21). However, *OsSWEET14* induction was not observed for artDS. *OsSWEET14* induction was also not observed for artAC2, an artificial version that was reconstructed based on the shorter TalAC variant occurring only in *Xoo* strain KACC10331 (Fig. 3.1.21). While the overall number of mismatches for TalAC2 is even lower than the number of mismatches found in the other TalAC class members, its repeat region is shorter, and thus it seems to lack the necessary

binding strength to successfully induce *OsSWEET14* (Fig. 3.1.21). TalDS on the other hand contains as many mismatched positions as for example TalAC3 or TalBH but its mismatches are predominantly found within the latter half of the TALE, and thus they might be too concentrated into a single region to allow for an efficient DNA-binding.

Interestingly, there is a second allelic variant of the *OsSWEET14* promoter, the resistance gene *xa41(t)* (Hutin *et al*., 2015b). This variant contains an 18 bp deletion that disrupts two of the three known TALE binding sites found within the *OsSWEET14* promoter, leaving only the binding site of TalC unaffected (Fig. 3.1.22 A; Hutin *et al*., 2015b). Analysing the above described TALEs with a promoter fragment containing the *xa41(t)* deletion showed that indeed only artTalC is able to induce this promoter, indicating that the putative binding sites of the other TALEs are altered too much for any of them to bind and activate to the promoter (Fig. 3.1.22 B). An alignment of the *xa41(t)* promoter with the four known *OsSWEET14*-inducing TALEs as well as with TalDS, TalDV and Tal2, confirms this observation, demonstrating that, with the exception of TalC, they all have a higher number of mismatches in the *xa41(t)* promoter than in the standard *OsSWEET14* allele (Fig. 3.1.22; Appendix Fig. 3.1.6 and 3.1.7). It remains therefore unknown whether TalDS and TalAC2 have evolved to address another *OsSWEET14* promoter allele that was not yet identified, whether they target a completely different gene or whether they simply do not have any biologically relevant target anymore.

3.1.12 A fragile balance – how a TALE repeat becomes aberrant

Not all naturally occurring aberrant repeats confer frameshift binding. This suggests that factors such as the position of an alteration and its specific amino acid composition influence how an aberrant repeat affects the TALE-DNA interaction. The natural 36aa v1 repeat (duplicated positions 27-28) loops out facultatively and allows standard as well as frameshift binding. The natural 36aa v2 repeat (duplicated positions 32-33) on the other hand is fixed within the array in a normal recognition mode. Based on this observation, 36aa repeat variants seemed a promising model system to further investigate which factors influence the behavior of an aberrant repeat. By introducing 2aa duplications at different positions of a TALE repeat, four artificial 36aa repeat variants were generated (Fig. 3.1.23 B; art1-art4). All six 36aa aberrant repeats were placed at position 8 of a TALE with 16.5 repeats using the RVD NN. The resulting TALEs were tested for their ability to bind to an optimal or a frameshift box (Fig. 3.1.23).

Two of the artificial aberrant repeats, 36aa art 1 and art4, were fully flexible, showing similar behavior as the natural variant 36aa v1 (Fig. 3.1.23). Surprisingly though, the two other artificial variants, 36aa art2 and art3, resulted in non-functional TALEs (Fig. 3.1.23). Since

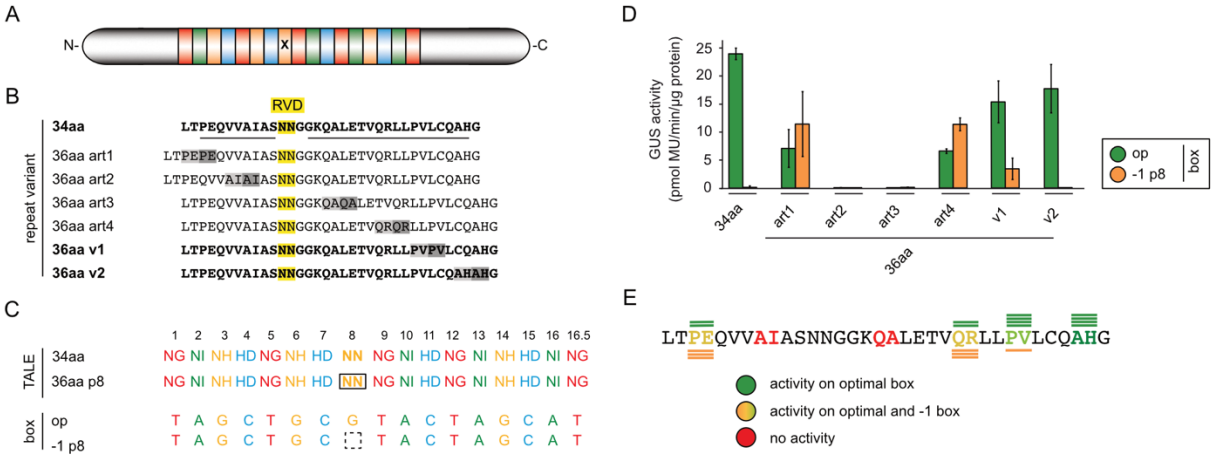


Figure 3.1.23: Mini-duplication scanning of 36aa repeat variants. (A) Schematic view of the generated TALEs. The x indicates the position of the aberrant repeat. **(B)** Amino acid alignment of the used repeat variants. The aberrant repeats labelled as artificial (36aa art1-4) were generated by introducing small, two amino acid duplications at different positions throughout a standard repeat (duplicated positions: 3-4, 8-9, 17-18, 23-24, 27-28, 32-33; termed 36aa art1-4, respectively). Duplicated aa positions are shaded in grey. **(C)** RVD composition of the generated TALEs. Position 8 (boxed) contains a standard repeat with 34aa (top row) or one of the variants with 36aa. The TALE-box is an optimal box (op) or a frameshift derivative with a single nucleotide deletion at position 8 (-1 p8). **(D)** Results of the quantitative GUS assay for the different 36aa repeat variants. Error bars represent standard deviation (n = 3). Similar results were obtained in two independent experiments. **(E)** Tolerance to small duplications throughout a TALE repeat. Duplications in the green part resulted in behaviour comparable to a 34aa repeat while duplications in the orange and light green parts resulted in facultative frameshift binding. Duplications in the red part of the repeat resulted in a complete loss of TALE activity. The strength of the depicted bars roughly corresponds to the observed activity on the respective boxes (green = optimal box; orange = -1 frameshift box).

both duplications in these two variants are placed relatively close to the RVD loop, it seems that alterations in this region are problematic for a TALE repeat. The results from this chapter strongly indicate that even a single repeat with an unfavorable amino acid composition can be sufficient to abolish the activity of a TALE completely. They also show that even minor alterations can influence whether an aberrant repeat has no impact on DNA binding, the ability to confer frameshift recognition or renders a TALE completely inactive, thus indicating an extremely fragile balance. This observation confirms that length differences alone are insufficient to explain why aberrant repeats behave differently and, consequently, that the composition of an alteration and/or its position must have an effect as well.

3.1.13 The position and the composition of repeat insertions control repeat behavior

To investigate how the specific amino acid composition of an aberrant repeat influences its effects on the TALE-DNA interaction, new artificial aberrant repeats were generated. However, this time not by introducing duplications of existing amino acids but by inserting completely new ones. For this, artificial variants of the two natural 40aa and 42aa aberrant repeats were generated. In these artificial aberrant repeats, each of the duplicated parts

present in the natural variants was replaced by a stretch of serine and glycines (SG) of identical length (Fig. 3.1.24). Chains of serines and glycines are often used as unstructured linkers to connect protein domains since they are very flexible and rarely affect protein function. These artificial aberrant repeats were placed at position 8 in a TALE with 16.5 repeats and the resulting constructs were tested for activity in a GUS assay by combining them with either an optimal or a -1 frameshift box. Striking differences could be observed between the natural repeats and their artificial derivatives, with the two TALEs containing the artificial 40aa variants showing an opposite behavior. While the variant with the first half of the duplication replaced strongly preferred the -1 frameshift box, the variant with the second half replaced preferred the optimal box (Fig. 3.1.24). Moreover, for the two artificial variants with 42aa, the differences were even more prominent. If the SG linker replaced the first half of the duplication, the TALE showed activity on both target boxes while replacing the second half yielded no activity at all (Fig. 3.1.24). These results further support the conclusion that the amino acid composition of an aberrant repeat influences how the aberrant repeat affects the TALE DNA-interaction.

In order to see if the exact positioning of an alteration within an aberrant repeat affects the DNA-binding behavior of a TALE as well, a 7aa SG-linker sequence (SSGGGGS) was introduced at different positions throughout a TALE repeat (Fig. 3.1.25 A). These repeats were inserted into a TALE at position 7 and the resulting TALEs were tested on an optimal (op) and a -1 nucleotide frameshift box with a deletion at position 7 (-1 p7) (Fig. 3.1.25 B and C). The positions close to the end of a TALE repeat (position 26-34) tolerated the insertion of the 7aa linker without any changes in the binding behavior, leading to activity only on the

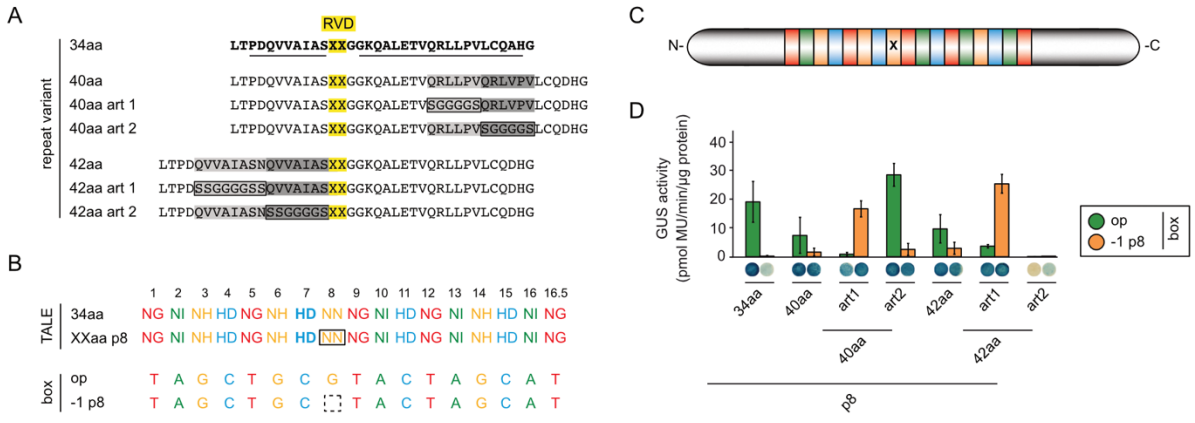


Figure 3.1.24: Serine/glycine linker swap experiments. (A) Amino acid alignment of the repeat variants used in this study. The repeat variable di-residues (RVD) are shaded yellow. A typical standard repeat of 34aa length is shown in bold, the helix-forming residues are underlined. Altered aa positions are shaded in grey. The aberrant repeats labelled as artificial were generated by swapping the indicated amino acids with a stretch of serines/glycines. **(B)** RVD composition of the generated TALEs. Position 8 (boxed) contains a standard repeat with 34aa or one of the aberrant repeat variants. The TALE-box is either an optimal box (op) or a frameshift derivative with a single nucleotide deletion at position 8 (-1 p8). **(C)** Schematic view of the generated TALEs. The x indicates the position of the aberrant repeat. **(D)** Results of the GUS assay for the different repeat variants. Error bars represent standard deviation (n = 3). Similar results were obtained in at least two independent experiments.

optimal box (Fig. 3.1.25 D-F). Insertions in the front (position 4-5) or middle (position 19-24) of the repeat on the other hand resulted in a compulsory loop out of the repeat (Fig. 3.1.25 F). Remarkably, only one of the artificial repeats, the one with an insertion at position 25, conferred a flexible binding to both, the optimal and the frameshift box. Interestingly, all linker insertions within the RVD loop or the surrounding residues (#9-18) resulted in a complete or

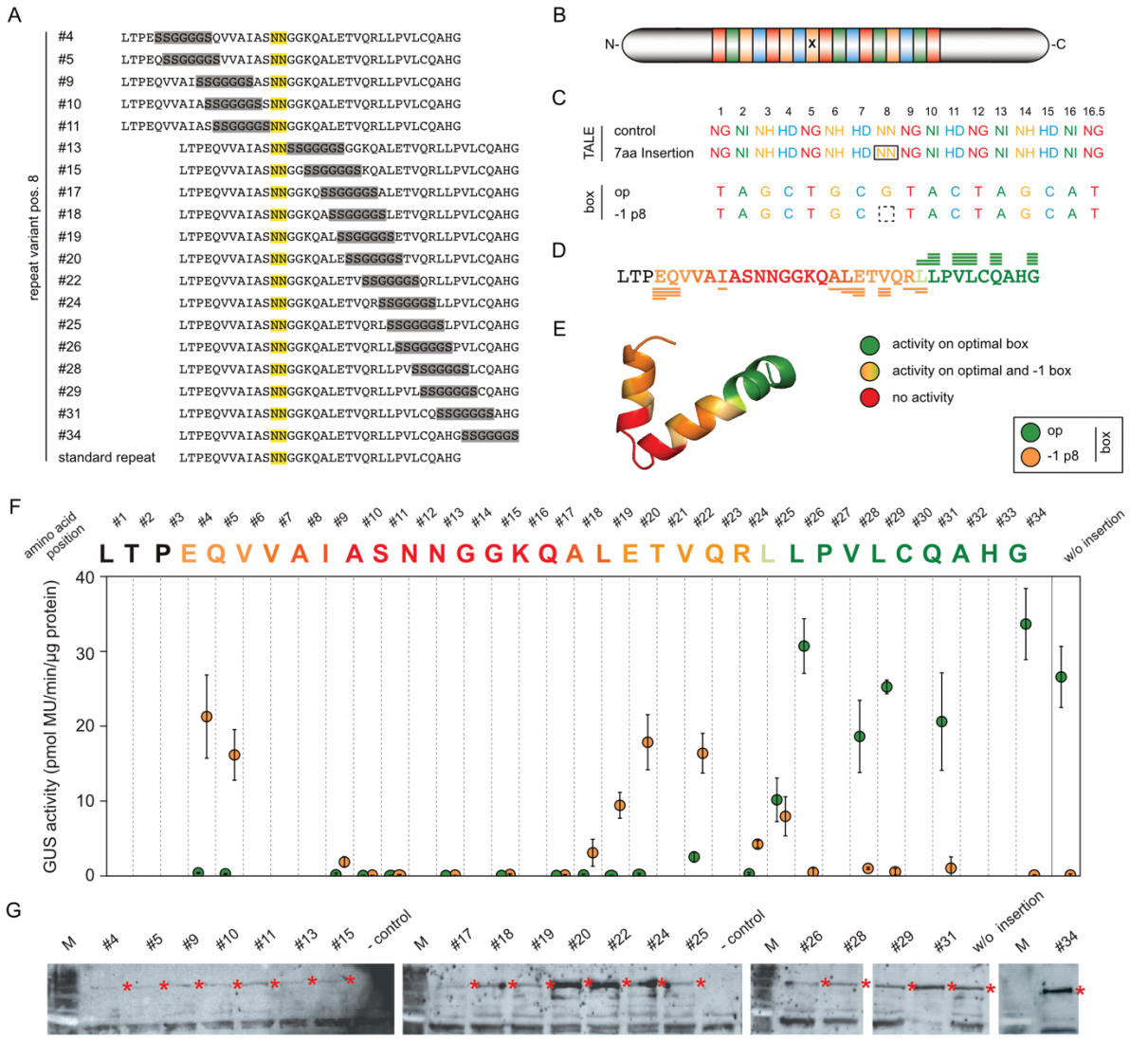


Figure 3.1.25: Insertion tolerance throughout a TALE repeat. (A) Amino acid alignment of the artificially generated repeat variants. A 7aa serine/glycine linker (SSGGGGS) was placed at different positions throughout a TALE repeat. The position of the insert is indicated. A standard repeat of 34aa length is shown in comparison. The repeat variable di-residues (RVD) are shaded in yellow, inserted aa in grey. **(B)** Schematic view of the generated TALEs. The altered repeat is marked with an x. **(C)** RVD composition of the artificial TALEs. The repeat at position 8 was replaced by one of the variants with a 7aa insertion but the RVD remained unchanged. All TALEs were tested on an optimal box (op) and a -1 frameshift derivative (-1 p8). **(D)** and **(E)** Tolerance of a TALE to a single repeat with insertions of a 7aa serine/glycine linker placed at different positions. Schematic is based on the data presented in (F). Insertions in the green part resulted in behaviour similar to a 34aa repeat (solid activity solemnly on the optimal box), insertions in the orange parts resulted in near-exclusive activity on the frameshift box, insertions into the light green part resulted in a facultative frameshift binding and insertions within the red part of the repeat resulted in a complete loss of TALE activity. (**D**) For amino acids coloured in black, no conclusions can be drawn with the presented data set. The bars underneath the letter representation roughly correspond to the strength of the measured GUS activity. **(F)** Results of the quantitative GUS assay. Error bars represent standard deviation (n = 3). Similar results were obtained in at least two independent experiments. **(G)** α-GFP Western Blot for all TALEs. Asterisks indicate correct heights of the TALEs.

near-complete loss of TALE function (Fig. 3.1.25 D-F). Western-blot analysis confirmed that this loss of function is attributed neither to a loss of protein expression nor to protein stability (Fig. 3.1.25 G). To investigate if the RVD loop region is similarly sensitive to smaller insertions, a single serine (S) was introduced at several positions within and around the RVD loop and the respective TALEs were tested for their GUS activity (Fig. 3.1.26). Remarkably, several of these TALEs also yielded no or nearly no activity (Fig. 3.1.26). This indicates that even one repeat with just a single additional amino acid inserted anywhere between positions 9-17 is sufficient to impair the activity of a TALE severely.

The results from this chapter confirmed that the exact position of an insertion as well as its specific composition are crucial factors that govern the behavior of an aberrant repeat. They showed that Insertions into either one of the helical regions of a TALE repeat negatively affect repeat-to-repeat interactions and cause the looping out of the aberrant repeat. The results also demonstrated that insertions within the RVD-loop or the adjacent amino acids have an even more severe impact, resulting in a non-functional TALE. Finally, the results showed that insertions that are placed into the C-terminal loop region that connects two repeats have mostly no negative impact on the DNA-binding behavior of the TALE, and thus represent a possibility for non-intrusive insertions into TALE repeats.

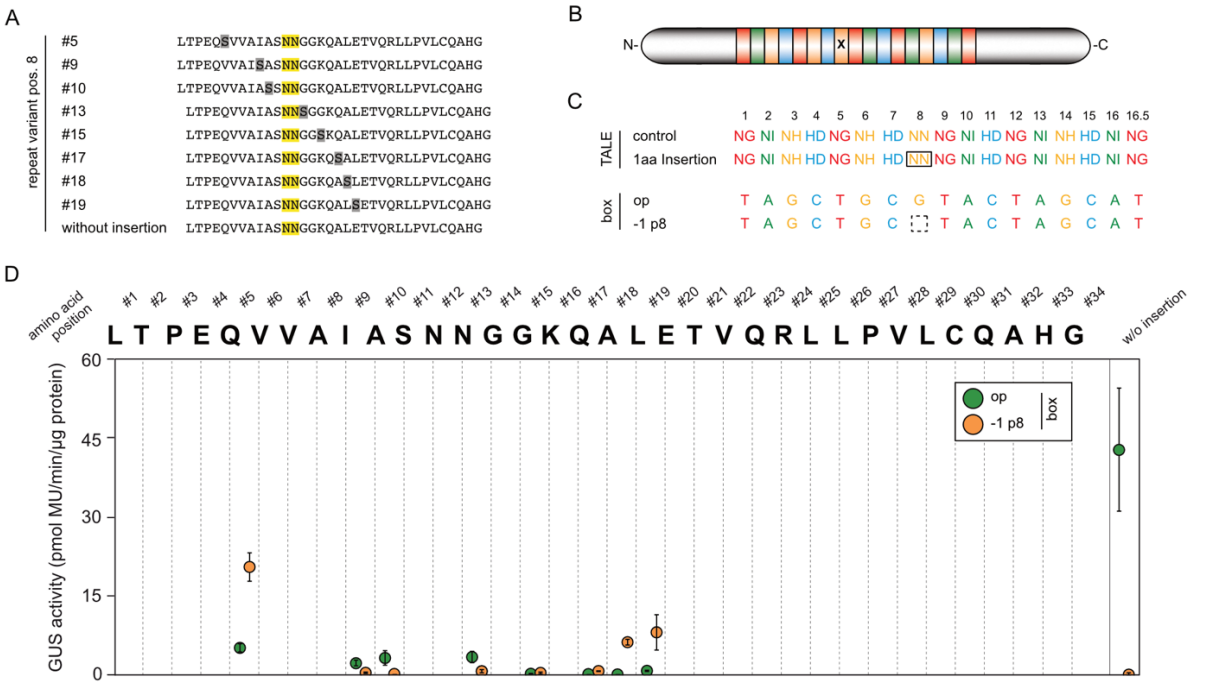


Figure 3.1.26: Single amino acid insertions into a TALE repeat. (A) Amino acid alignment of the artificially generated repeat variants. A single serine (S) was inserted at different positions throughout a TALE repeat, preferentially where the insertion of a 7aa SG linker resulted in a non-functional TALE. The position of the insert is indicated. A standard repeat without insertion is shown in comparison. The repeat variable di-residues (RVD) are shaded in yellow, inserted aa are shaded in grey. **(B)** Schematic view of the generated TALEs. The altered repeat is indicated by an x. **(C)** RVD composition of the artificial TALEs. The repeat at position 8 of the TALE contained one of the variants shown in (A). The RVD of the repeat remained unchanged. All TALEs were tested on an optimal box (op) and a -1 frameshift derivative (-1 p8). **(D)** Results of the quantitative GUS assay. Error bars represent standard deviation (n = 3). Similar results were obtained in two independent experiments.

3.1.14 Functional fusions – TALEs can tolerate the insertion of large domains between repeats

The previous chapter demonstrated that up to seven amino acids can be inserted into a TALE repeat without changing the DNA-binding behavior of the TALE – at least as long as the inserted amino acids are placed within a certain region of the repeat. This observation prompted the question whether it is possible to introduce even larger sequences or complete protein domains into or between TALE repeats. Since the 7aa linker was best tolerated if placed close to the end of a repeat, it seemed the most logical choice to place any further insertions simply between two repeats.

As a first test, a new linker was generated and placed between repeats 7 and 8, this time with a length of 18aa (SSGGGSGGSGSGGGSGSS). Similarly to the TALE with a 7aa linker, this TALE showed solid activity on the optimal box and no activity on a -1 or a +1 frameshift box (Fig. 3.1.27 A-C). In order to test whether it is possible to insert not only unstructured stretches of amino acids between two repeats but also functional domains, two different tags were selected. The first was the comparably small HA-tag (9aa) and the second the much larger, barrel-shaped GFP-tag (238aa). Both tags were placed between repeats 7 and 8 and, to prevent unnecessary strain on the neighboring repeats, combined with a 7aa SG linker on each side (Fig. 3.1.27 D and E). To determine whether such internally tagged TALEs are still able to bind to a specific target sequence in a manner comparable to a standard TALE, the two generated constructs were tested in a GUS assay. For this, they were combined with an optimal box (op) or a -1 frameshift box containing a deletion either at position 7 (-1 p7) or position 8 (-1 p8). Both internally tagged proteins induced the expression of the reporter construct with the optimal box in a comparable manner to the control TALE without any insertions (Fig. 3.1.27 D-F), thus indicating that neither of the two internal tags interfered much with the DNA-binding or gene-inducing capability of the TALE. This suggests that both tags loop out of the array, thus behaving somewhat comparable to an aberrant repeat. Both constructs showed also a very low but reproducible activity on the -1 frameshift box, indicating that sometimes one of the standard repeats adjacent to the internal tag loops out as well (Fig. 3.1.27 F). However, this activity seems neglectable in comparison to the solid activity on the optimal box observed for both constructs.

To clarify if all parts of an internally tagged TALE are bound to its DNA target sequence, three new boxes were designed, each one containing three mismatches either before (#3-5) or after (#10-12, #14-16) the inserted domain (Fig. 3.1.28 A). The subsequent analysis of these three boxes in combination with the internally GFP-tagged TALE demonstrated that only the optimal box was bound and activated, thus confirming that all or at least most repeats of the TALE are required to contact the DNA target sequence (Fig. 3.1.28 B). Further analysis also demonstrated that it is possible to place the internal GFP-tag at different

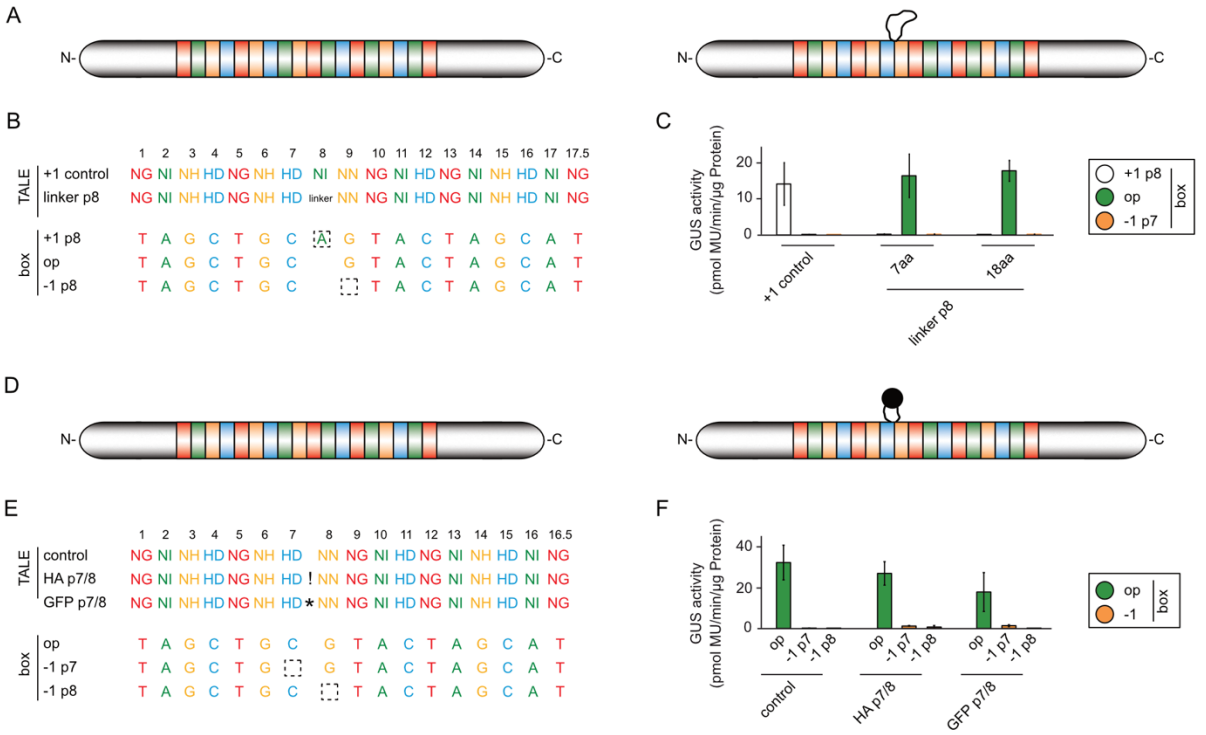


Figure 3.1.27: The region between two repeats is suited for the insertion of additional amino acids or functional domains. (A) Schematic view and **(B)** RVD composition of the generated TALEs. Short SG linker sequences with 7aa or 18aa were placed between two repeats and the resulting TALEs combined with an optimal box (op), a -1 frameshift box lacking one nucleotide close to the insertion (-1 p8) or a +1 box containing an additional nucleotide at the position of the insertion (+1 p8). **(C)** Results of the quantitative GUS assay. Error bars represent standard deviation (n = 3). Similar results were obtained in at least two independent experiments. **(D)** Schematic view and **(E)** RVD composition of the generated TALEs. An HA- or a GFP-tag was inserted between two repeats and the resulting TALEs were combined with an optimal box (op) or -1 frameshift derivatives lacking one nucleotide either at position 7 (-1 p7) or 8 (-1 p8). **(F)** Results of the quantitative GUS assay. Error bars represent standard deviation (n = 3). Similar results were obtained in at least two independent experiments.

positions throughout the repeat array (positions p1/2, p7/8 and p13/14) with similar results (Fig. 3.1.28 C and D). However, inserting three GFPs simultaneously at these three positions yielded a TALE with an extremely low transcriptional activity, suggesting that this TALE is unable to interact properly with its target sequences (Fig. 3.1.28 C and D). Laser scanning microscopy confirmed that a GFP domain placed within the repeat array of a TALE is functional. However, compared to a TALE with an N-terminal GFP tag, the observed signal intensity was slightly reduced (Fig. 3.1.28 E). Interestingly, increasing the number of internal GFP tags increased also the signal strength, reaching a peak with the TALE containing one N-terminal and three internal GFP tags (Fig. 3.1.28 E). This suggests that even multiple inter-repeat GFPs are correctly folded and functional – despite the observation that the TALE itself is drastically impaired in its gene activating capabilities (Fig. 3.1.28 D and E).

The results from this chapter demonstrated that it is possible to insert functional domains between individual TALE repeats without interfering with the specific DNA recognition of the TALE *in vivo* while at the same time maintaining functionality of the domain itself. It is likely

that some inserts, especially larger ones with a more complex or less favorable 3D structure, are not suited for an internal insertion. However, varying the size and/or the composition of the linker sequences might help to identify variants that exhibit less strain on the folding and structure of a TALE, thus improving its tolerance to insertions.

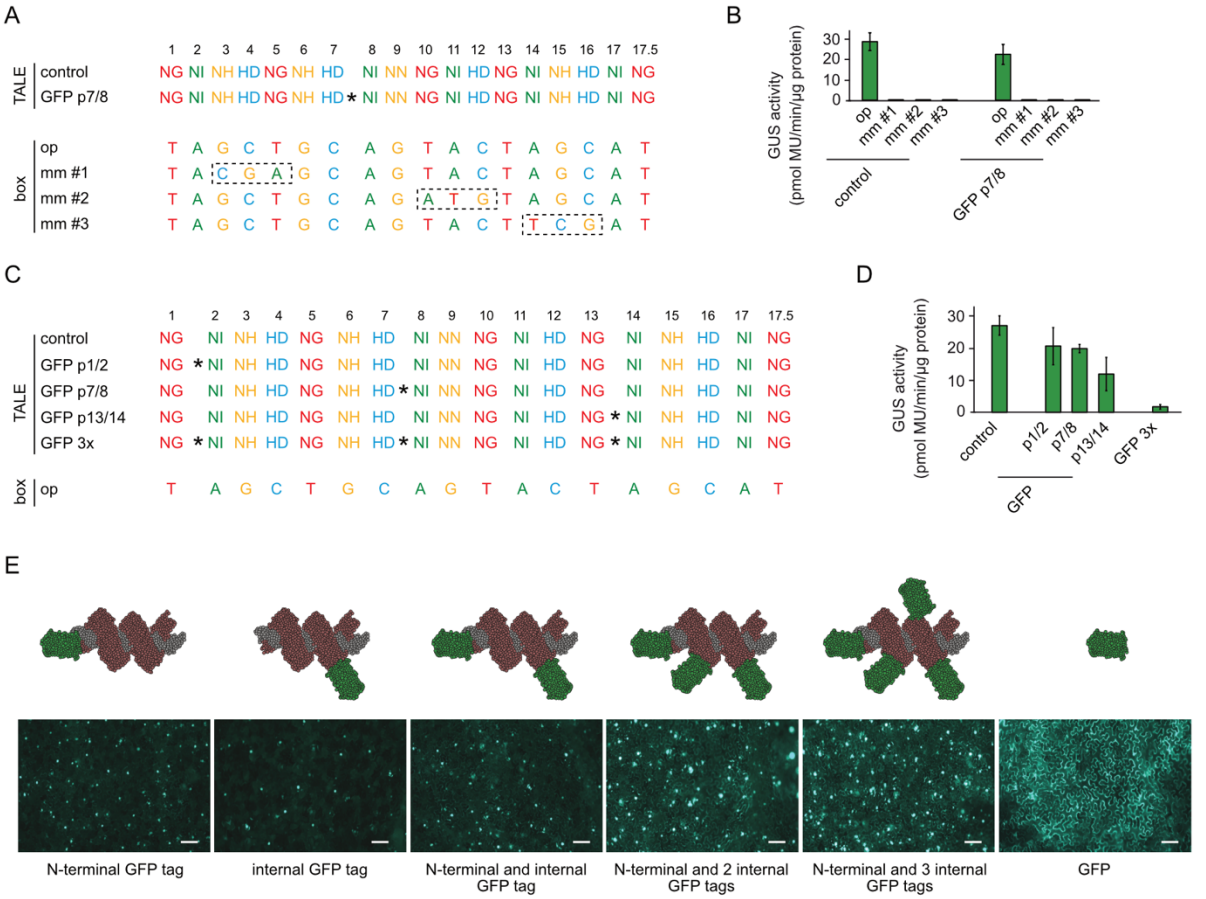


Figure 3.1.28: A GFP-tag placed between two TALE repeats remains functional. (A) RVD composition of the generated TALEs. A TALE with an internally fused GFP was combined with three derivatives of an optimal box (op), each yielding three different consecutive mismatches if paired with the TALE. The GFP was placed between repeats 7/8 and is represented by an asterisk. **(B)** Results of the quantitative GUS assay. Error bars represent standard deviation (n = 3). Similar results were obtained in at least two independent experiments. **(C)** RVD composition of the generated TALEs. A GFP tag was inserted at different positions of a TALE repeat array (p1/2, p7/8 or p13/14) or at all three positions simultaneously (GFP 3x). An asterisk represents the position of the GFP. **(D)** Results of the quantitative GUS assay. Error bars represent standard deviation (n = 3). Similar results were obtained in at least two independent experiments. **(E)** Microscopy images and 3D-cartoons of GFP or GFP-tagged TALEs. The TALEs contained a varying number of GFP tags that were placed either N-terminally, internally or both, with the number of internal GFP tags ranging from one to three. The white bar represents 100µm. The microscopy analysis was done only once.

3.2 TALEs from *Xanthomonas oryzae* pv. *oryzicola* and their target genes in rice

Xanthomonas oryzae pv. *oryzicola* (*Xoc*) and *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) are the two causal agents of bacterial leaf streak (BLS) and bacterial blight (BB) in rice. While *Xoo* is undoubtedly the more severe threat, more than 40 genes conferring resistance against BB were identified so far. In contrast, not a single resistance gene was identified in the *Xoc*-rice pathosystem and neither have the resistance genes against BB been shown functional against BLS. To keep BLS under control and prevent *Xoc* from emerging as a more serious threat, it is therefore crucial to characterize the molecular mechanisms behind its virulence in more detail and to intensify the search for ways that confer resistance against this pathogen.

3.2.1 Classification of *Xoc* TALEs using AnnoTALE and the assignment of genomic TALE clusters in the 10 completely sequenced *Xoc* strains

Classifying TALEs according to a unified nomenclature and defining genomic TALE clusters immensely increases the overview and clarity when analysing TALEs and their target genes, and thus such an overview is desireable for the *TALE* genes occurring in *Xoc*. However, while some studies sorted *Xoc TALE* genes into clusters, they neither attempted to define these clusters clearly nor did they try to unify the naming of these clusters within the different *Xoc* strains (Cernadas *et al*., 2014; Wilkins *et al*., 2015). Consequently, the present study attempted to change this.

To assign genomic TALE clusters, it was taken into account whether a TALE (or a group of TALEs) occurs always in a similar genomic context, shares the same adjacent genes and is comprised of the same TALE classes. It was possible to group the 43 different TALE classes from the 10 completely sequenced *Xoc* strains into 12 genomic TALE clusters, labeled I-XII (Fig. 3.2.1; Table 3.2.1 and 3.2.2). The number of *TALE* genes in these groups ranges from one in the smallest (cluster I, VI, IX X and XI) to ten in the largest cluster (cluster IV). As observed before in *Xoo*, the number of *Xoc TALE* genes within a cluster, their order, their

Table 3.2.1: Completely sequenced *Xoc* strains from Asia and Africa.

strain	origin	# of TALEs	year of isolation	genome size (Mbp)	GC content (%)	reference	accession number
B8-12	China	28	2007	4794	64.6	Wilkins (2015)	CP011955.1
BLS256	Philippines	28	1984	4831	63.9	Bogdanove (2011)	CP003057.2
BLS79	Philippines	26	1984	4790	67.5	Wilkins (2015)	CP011956.1
BXOR1	India	27	1996	4692	64.1	Wilkins (2015)	CP011957.1
CFBP2286	Malaysia	28	1964	4969	62.6	Wilkins (2015)	CP011962.1
CFBP7331	Mali	20	2003	5008	63.9	Wilkins (2015)	CP011958.1
CFBP7341	Burkina	20	2009	517	63.9	Wilkins (2015)	CP011959.1
CFBP7342	Burkina	23	2009	5080	64.0	Booher (2015)	CP007221.1
L8	China	29	<1995	4796	64.6	Wilkins (2015)	CP011960.1
RS105	China	24	1992	4779	64.6	Wilkins (2015)	CP011961.1

orientation and their class affiliation can vary due to deletions, duplications, inversions or recombination events (Wilkins *et al*., 2015; Erkes *et al*., 2017). However, despite these extensive genomic rearrangements, 11 out of the 12 assigned TALE clusters can be found in all 10 analysed *Xoc* strains and in nearly all cases with the adjacent genes that were used to define the clusters being preserved as well (Table 3.2.2; Appendix Table 3.2.1). The only exception is cluster X but it contains just a single *TALE* gene, TalCN, and occurs only in one *Xoc* strain, BXOR1. The presence of a TALE class is usually limited to a single corresponding genomic TALE cluster, a fact that strongly supports the solidity of the AnnoTALE-based classification system and of the newly assigned genomic TALE clusters alike. However, three exceptions do exist. The first one concerns TalBN, whose gene moved from cluster XII to cluster III in the two strains BXOR1 and CFBP7342 – in case of BXOR1 even without any changes in its RVD sequence (Fig. 3.2.1). The second one is class TalBF,

Table 3.2.2: Definition of genomic TALE clusters in *Xoc* in reference to the type strain BLS256[1].

cluster	genes used to define TALE cluster border (locus tag / protein ID) [2]	
I	Esterase-lipase-thioesterase family protein CDS (*XOC_4538* / AEQ98595.1)	tRNA uridine 5-carboxymethylaminomethyl modification enzyme MnmG CDS (*XOC_4523* / AEQ98581.1)
II	NAD-binding domain 4 protein CDS (*XOC_4410* / AEQ98476.1)	Prolyl oligopeptidase-like protein CDS (*XOC_4374* / AEQ98445.1)
III	beta-N-acetylglucosaminidase CDS (*XOC_4258* / AEQ98331.1)	sperimidine synthase CDS (*XOC_4242* / AEQ98317.1)
IV	Chemotaxis protein methyltransferase CDS (*XOC_1589* / AEQ95757.1)	gpmA CDS (*XOC_1554* / AEQ95734.1)
V	Integral membrane protein CDS (*XOC_1719* / AEQ95883.1)	Glutathione reductase CDS (*XOC_1702* / AEQ95870.1)
VI	LysR family protein CDS (*XOC_2870* / AEQ96975.1)	Methyl-accepting chemotaxis protein CDS (*XOC_2854* / AEQ96961.1)
VII	Efflux pump membrane transporter CDS (*XOC_2507* / AEQ96626.1)	Serine protease CDS (*XOC_2475* / AEQ96605.1)
VIII	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase CDS (*XOC_2015* / AEQ96162.1)	Rhs element Vgr protein CDS (*XOC_2004* / AEQ96155.1)
IX	UvrABC system protein B CDS (*XOC_2896* / AEQ96999.1)	Putative acyltransferase CDS (*XOC_2890* / AEQ96994.1)
X	n.a. [3]	n.a. [3]
XI	Putative NADH dehydrogenase/NAD(P)H nitroreductase CDS (*XOC_4041* / AEQ98129.1)	ABC transporter ATP-binding protein CDS (*XOC_4024* / AEQ98112.1)
XII	short chain dehydrogenase CDS (*XOC_0479* / AEQ94699.1)	Thioredoxin CDS (*XOC_0457* / AEQ94692.1)

[1] See Appendix Table 3.2.1 for locus tags and protein IDs in all 10 *Xoc* strains.

[2] Sometimes it was not the first adjacent gene that was used to define the cluster borders but the first that was conserved in all 10 strains. Hypothetical proteins and mobile genetic elements were not used as border genes.

[3] The cluster is not present in BLS256.

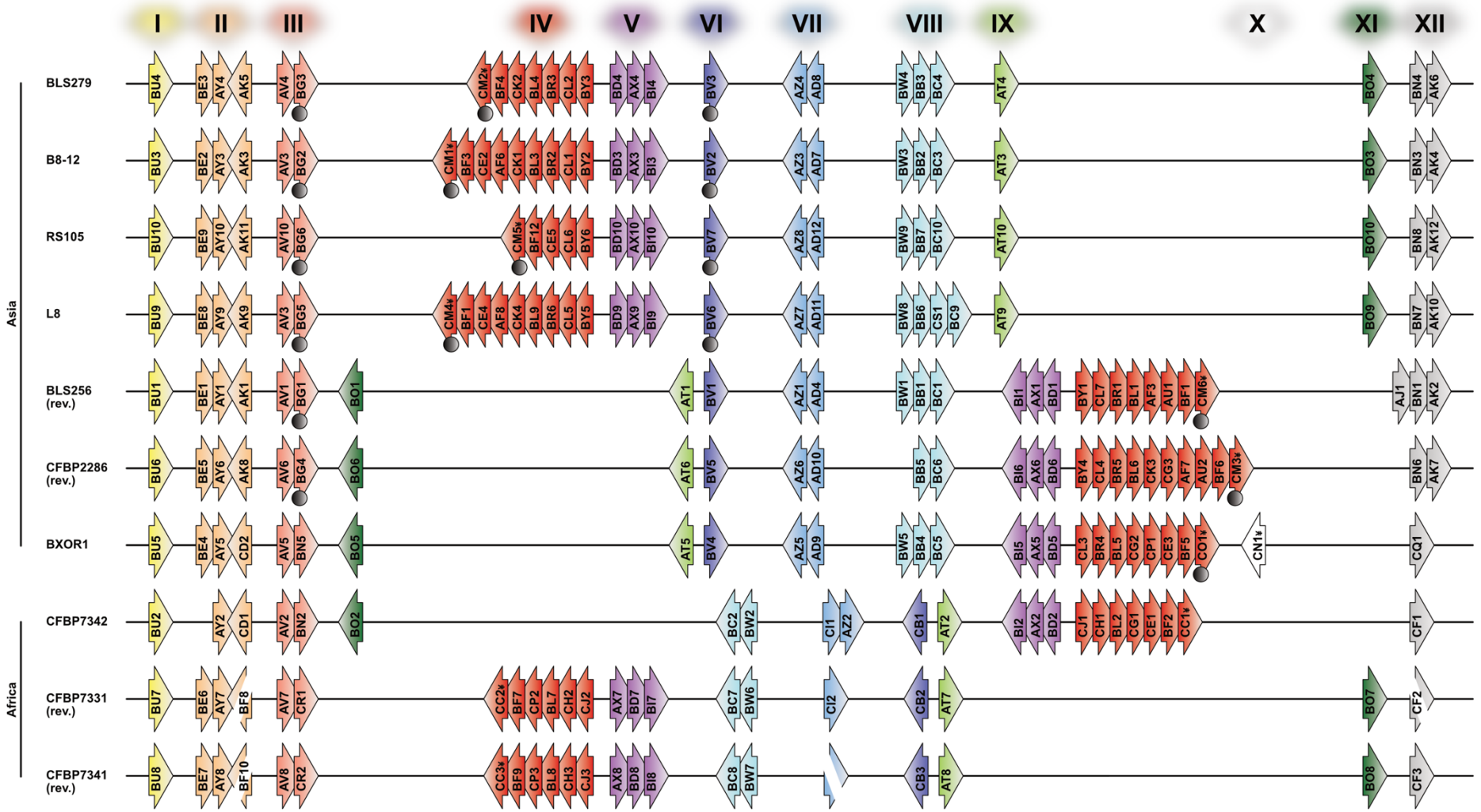


Figure 3.2.1: TALE cluster affiliation in the 10 completely sequenced *Xoc* strains. *TALE* genes are indicated as arrows with regard to their orientation in the genome. All TALEs are named according to the AnnoTALE classification system. All TALEs were further categorized into 12 genomic clusters (I-XII). This categorization was based on their genomic positions and their adjacent genes (Table 3.2.2). The genomic TALE clusters can be differentiated by colour. TALEs carrying an aberrant repeat are labelled with a grey dot. TALEs with truncated N- and C-terminal regions are labelled with the ¥ symbol, all other fragmented *TALE* genes are indicated by an incomplete arrow. For a better overview BLS256, CFBP2286, CFBP7331 and CFBP7341 are shown in reverse orientation (rev.).

which was duplicated in the two strains CFBP7331 and CFBP734, and thus occurs in these strains once in its original position in cluster IV and once in form of a fragmented pseudogene in cluster II (Fig. 3.2.1). The third exception is class TalAK, which occurs in six strains twice, in cluster II and XII (Fig. 3.2.1).

A previous study analysing *Xoc* TALEs and their target genes concluded that only five TALEs are found in all ten completely sequenced *Xoc* strains (Wilkins *et al*., 2015). However, the authors allowed only a maximum of two RVDs to differ. By sorting them not simply based on differences in their RVDs but rather their specificities, the more elaborate algorythm used by AnnoTALE identified 10 TALE classes that occur in all 10 *Xoc* strains (Fig. 3.2.2; Grau *et al*., 2016). Comparing their RVD sequences and their optimal target sequences confirmed a general similarity between the different members of each class while their locations within the newly assigned genomic TALE clusters supported this categorization further (Fig. 3.2.1; Appendix Table 3.2.2 and 3.2.3). The five TALE classes previously identified as conserved in all 10 completely sequenced *Xoc* strains are TalAY, TalAV, TalBF, TalAX and TalBD (Wilkins *et al*., 2015), while the five novel classes are TalBU, TalBI, TalBC, TalAT and TalBO (Fig. 3.2.2). Three of these conserved TALE classes occur alone, each within a separate TALE cluster, additional four are located in TALE clusters with 2, 3 or 9 other, less conserved *TALE* genes while the last three all occur together in cluster V, rendering it also the most conserved of all the TALE clusters found in *Xoc* (Fig. 3.2.1). Each of the 10 completely sequenced *Xoc* strains also contains at least one TALE with small N- and C-terminal deletions and, frequentely but not always, the aberrant repeat variant with 28aa. While the high degree of variability in their repeat regions caused these putative pseudogenes to be placed into several different TALE classes, all but one occur at the same genomic position, the end of cluster IV and thus, a shared evolutionary origin seems likely (Fig. 3.2.1 and 3.2.2). Particularly interesting are also the two TALE classes TalAD and TalAF since they are the

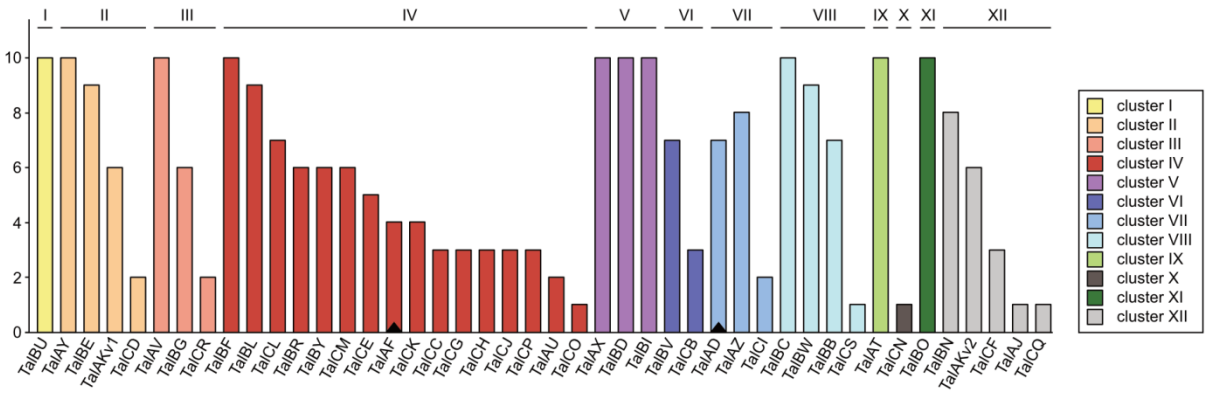


Figure 3.2.2: TALE class abundance and cluster affiliation in the 10 completely sequenced *Xoc* strains. TALEs are named and grouped according to the AnnoTALE classification system. The different colours represent the newly assigned genomic TALE clusters I-XII. TALE classes are sorted based on their cluster affiliation and their abundance within each cluster. TalBN, which occurs at two different clusters (6x cluster XI and 2x cluster III), is affiliated to its most frequent position, cluster XI. The small triangles indicate the two TALE classes that are shared between *Xoc* and *Xoo* (TalAF and TalAD).

only ones shared between strains from *Xoo* and *Xoc* (Fig. 3.2.1 and 3.2.2). Despite the fact that the conservation of the latter two affects neither all *Xoo* nor all *Xoc* strains, their occurrence in strains of both pathovars suggests an important role in virulence and thus, analysing them in more detail might help to uncover yet unknown infection strategies that are shared by both pathogens.

To evaluate the AnnoTALE class building and naming system, a phylogenetic tree was generated using an alignment of the predicted optimal target sequence of each *Xoc* TALE with a unique RVD sequence (Fig. 3.2.3). For this, all TALE classes were sub-categorized into versions, with each version representing a unique RVD sequence (Appendix Fig. 3.2.1; Appendix Table 3.2.3). A list of the used optimal target sequences as predicted by AnnoTALE can be found in the Appendix (Appendix Table 3.2.2). For this analysis, only putatively functional TALEs were taken into account, therefore excluding all TALEs with fragmented sequences. The results of this analysis largely support the TALE classes assigned by AnnoTALE, with the overwhelming majority of TALEs showing closest similarity to all other members of the same class (Fig. 3.2.3). This analysis also identified a few TALE classes that overlap throughout most of their predicted optimal target sequences, indicating that they might share a plant target gene. A direct comparison between the RVDs and the possible optimal binding sequences of these TALE classes highlighted five combinations in particular that could address the same target sequences with only very few mismatched positions. These are: TalBV and CB, both the only TALE classes occurring in cluster VI, TalBR and CH, both occurring in cluster IV, and TalCJ and CL, both occurring in cluster IV, (Fig. 3.2.4 A, B and C). The optimal sequences of the two other combinations, TalCF and CQ and TalCK and CP, show slightly more differences, yet, they are still similar enough for the possibility of a shared target gene or different alleles of one gene (Fig. 3.2.4 D and E). These observations are largely supported by a follow-up study analysing TALE evolution in *X. oryzae* strains (Erkes *et al*., 2017). With the exception of the combination TalCF and TalCQ, their analysis of *Xoc* TALEs allows for similar conclusions – despite the fact that they did not use the optimal target sequences of every unique member as basis for their calculations but the RVD similarities between the different TALE classes in general (Erkes *et al*., 2017). When analysing potential target genes, it is therefore advisable to keep in mind that some of these combinations might share a target.

The analysis also identified a single case in which the predicted optimal target sequence of a TALE shows more similarities to members from a different TALE class than to the other members of its own class. The predicted optimal target sequences of the two class TalAK versions align only partially, while TalAK v1 and TalCD v1 could bind an identical target sequence with just a single mismatch (Fig. 3.2.4 F, G and H). Additionally, the two versions of TalAK are always present simultaneously and occur at different genomic clusters while

TalAK v1 and TalCD are located at the same genomic TALE cluster but never within the same strain (Fig. 3.2.1). Consequently, TalAK v1 and TalAK v2 are from now on regarded as two separate TALE classes, at least for the purpose of this study.

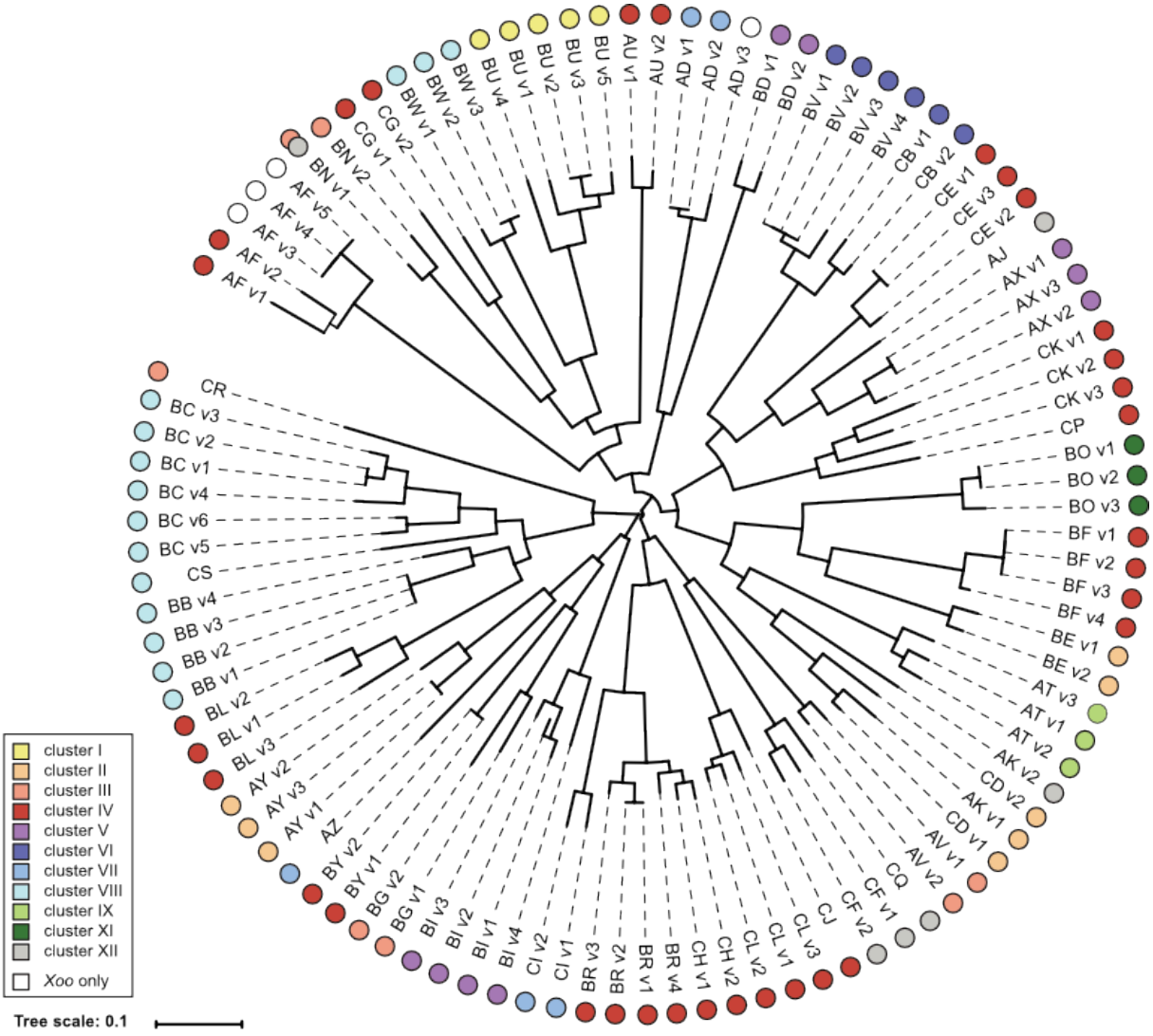


Figure 3.2.3: Phylogenetic Tree of all unique RVD sequence from *Xoc* TALEs. The tree is based on an alignment of the predicted optimal target sequences of all 101 unique RVD sequences occurring in *Xoc* TALEs. These optimal target sequences were predicted using AnnoTALE (for a list of the used sequences, see Appendix Table 3.2.2). The alignment was generated using Clustal Omega, the visualization was done using iTol. All TALEs were categorized and named according to AnnoTALE. All class members were subcategorized into different versions, with each version representing a unique RVD sequence shared by all other class members with the same version number (for a list of all unique RVD sequences, see Appendix Table 3.2.3). Cluster affiliations are indicated by coloured dots next to the TALE class versions. Only TALEs with complete N- and C-terminal regions were taken into account. The tree also contains three TALE class versions found only in *Xoo*, namely TalAF v1, v2 and AD v3. However, since they belong to a class that is shared by the two pathovars they were also included into this analysis.

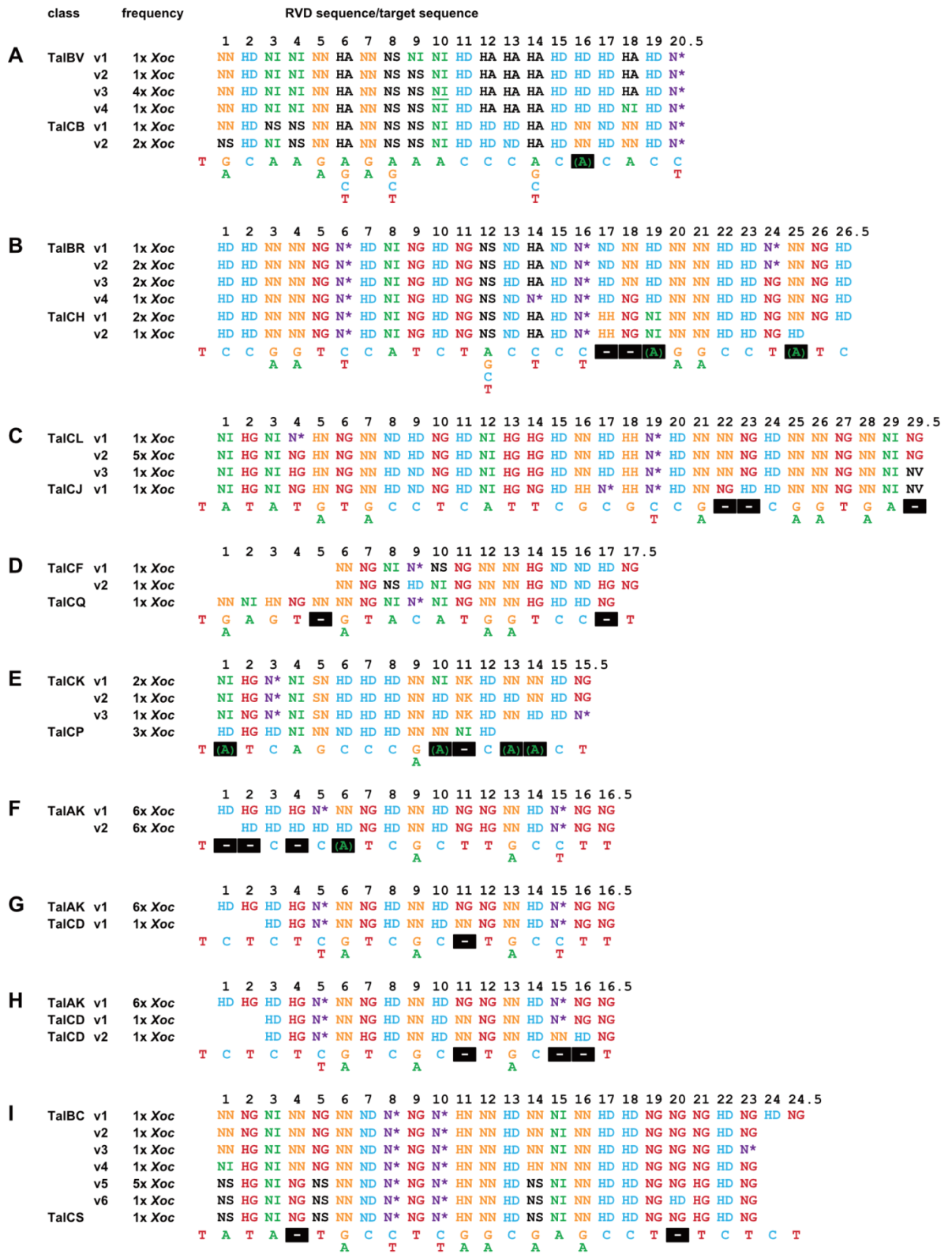


Figure 3.2.4: Direct comparison of all *Xoc* TALE classes whose predicted optimal target sequences were identified as highly similar. Shown are the RVD sequences of the selected TALE classes as well as the shared optimal binding sequences for the different combinations. Each position shows only those nucleotides that can be tolerated by all indicated TALEs. Positions that are marked as empty (-) cannot be addressed by all TALEs in question without some of them mispairing. Comparison of TALE classes: **(A)** TalBV and TalCB, **(B)** TalBR and TalCH, **(C)** TalCL and TalCJ, **(D)** TalCF and TalCQ, **(E)** TalCK and TalCP, **(F)** TalAK v1 and v2, **(G)** TalAK v1 and TalCD v1, **(H)** TalAK v1 and TalCD.

3.2.2 Construction of artificial *Xoc* TALEs

In order to analyse the naturally occurring *Xoc* TALEs, a set of artificial *Xoc* TALEs was constructed. By containing the same RVDs as their naturally occurring counterparts, these artificial *Xoc* TALEs contain identical DNA-binding specificities, and thus they should be well suited for the analysis and confirmation of putative new target genes. This approach is as reliable as the subcloning of individual TALEs from *Xoc* strains and if an adequate TALE repeat library is avaliable, it is easier and faster accomplished.

The 10 completely sequenced *Xoc* strains contain 43 different TALE classes with a total of 108 unique RVD sequences (Fig. 3.2.5; Appendix Table 3.2.3). Originally, it was planned to reconstruct the RVD sequence of only one member from each of the 40 putatively functional TALE classes found in *Xoc*. However, most TALE classes exhibit differences in the RVD sequences of their individual members, and thus this variation should be taken into account as well. This could help to identify especially those targets that are not only addressed by just a few but by all members of a TALE class. Consequentely, multiple versions were constructed for nearly all TALE classes, totaling 80 TALEs (Fig. 3.2.5). For this, their RVD sequences were combined with the N- and C-terminal regions from the TALE Hax3 (Kay *et al*., 2005) and cloned into the binary vector pSKA2. After transformation into *Agrobacterium tumefaciens* they were ready for subsequent analysis in *N. benthamiana*. To reconstruct the RVD sequences of the naturally occurring *Xoc* TALEs perfectly, several novel repeat modules were generated and implementated into the Golden TAL Technology cloning kit (Appendix Table 3.2.4). With the help of these novel repeat modules, it was possible to

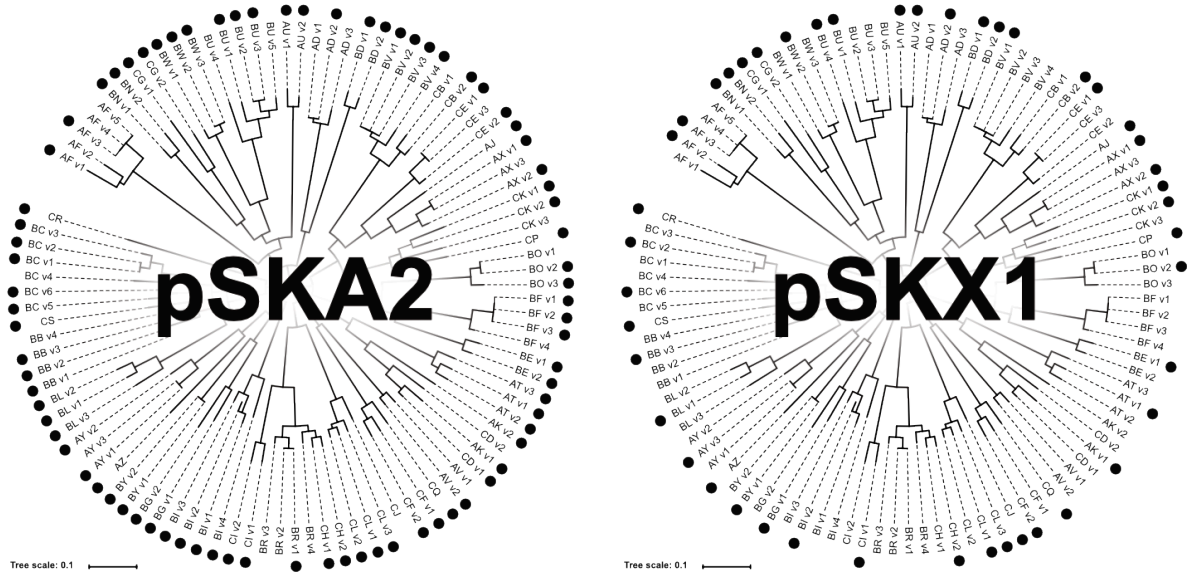


Figure 3.2.5: Artificial *Xoc* TALEs constructed in this study. All TALE class versions that were artificially reconstructed in this study are indicated by a black dot. TALEs were either assembled using the vector pSKA2 and transformed into *A. tumefaciens* strain GV3101 or using the vector pSKX1 and transformed into *Xo*-US strain Roth X1-8, thus, allowing for a flexible analysis of these TALEs in either *Nicotiana benthamiana* or *Oryza sativa*, respectively. For a detailed list of the used RVD sequences, see Appendix Table 3.2.3.

reconstruct the vast majority of the artificial *Xoc* TALEs perfectly. Nonetheless, nine of the 80 TALEs differ in one or several RVDs from the naturally occurring RVD setup (Appendix Table 3.2.3, RVDs highlighted in yellow). However, the only two RVD combinations affected are NG/HG and HD/ND and both were shown to confer identical nucleotide specificities (Yang *et al*., 2014; Miller *et al*., 2015; Zhou *et al*., 2015).

To analyse their effects and target genes also directly in rice, the RVD sequences of 48 *Xoc* TALEs were additionally combined with the N- and C-terminal region of TalAG4 and TalAO3, respectively, and assembled into the vector pSKX1 (Fig. 3.2.5; Appendix Table 3.2.3). Both, TalAG4 and TalAO3, originate in the *Xoo* strain PXO83, and thus their N- and C-terminal regions do not perfectly represent the ones occurring in *Xoc*. However, a recent publication by our group demonstrated that artificial TALEs constructed with these two parts are functional in rice (Mücke *et al*., 2019). The succesful transformation of 43 TALE constructs into an *X. oryzae* strain without any TALEs, Roth X1-8, was confirmed via α-FLAG western blot (Fig. 3.2.6). Inoculating them into rice will allow for the analysis of expression changes in putative TALE target genes and might even help to identify putative phenotypic effects of individual TALEs.

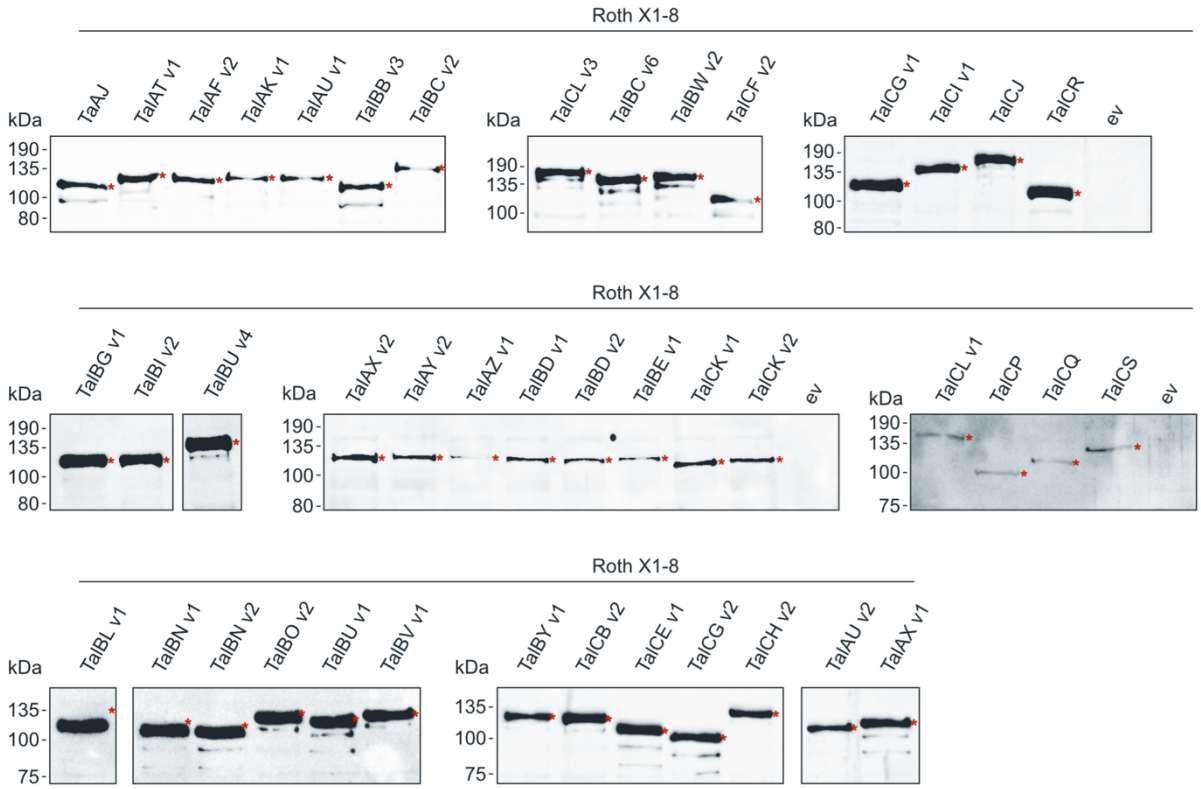


Figure 3.2.6: Western blot analysis of the artificial *Xoc* TALEs after transformation into *Xo*-US strain Roth X1-8. The pSKX1-based TALE expression constructs were transformed into *Xo*-US strain Roth X1-8. Single colonies were picked and grown in liquid media. Their whole protein content was separated via SDS PAGE. After the transfer onto nitrocellulose membranes, the TALE proteins were visualized using an α-FLAG primary antibody from mouse and an horseradish peroxidase-coupled secondary antibody against mouse.A red asterisk indicates the correct size of a TALE. ev = empty vector.

3.2.3 Activity of *Xoc* TALEs at the promoter regions of their published target genes

So far, 18 different *Xoc* TALE targets have been published (Cernadas *et al*., 2014; Wilkins *et al*., 2015). These 18 rice genes are upregulated by 15 different TALE classes, with four classes, namely TalBF, TalBV, TalBI and TalAX, addressing multiple targets (two target genes for TalBF, TalBI and TalBV and three for TalAX). Additionally, the two class combinations TalBV/CB and TalBR/CH each share a target gene (Appendix Table 3.2.5). The latter observation is in accordance with the phylogenetic analysis of the aligned optimal target sequences of these four classes, and thus gives further support for the validity of the approach used in chapter 3.2.1. Interestingly, in four cases a target gene was upregulated after inoculation with a *Xoc* strain despite the presence of a suitable TALE (Wilkins *et al*., 2015). By comparing the published expression pattern of these genes after *Xoc* infection with the the presence or absence of *Xoc* TALE classes in general and individual member TALEs in particular it was possible to identify the responsible TALEs, namely TalBC v1, TalBI v1 and v4 and TalCH v2 (Appendix Table 3.2.5; marked with asterisks). The most likely explanation for them being overlooked in the previous studies lies in the RVD differences between them and their class members, thus proving once more the value of the AnnoTALE classification system. Overall, the published expression pattern of the known *Xoc* TALE target genes correlates well with the presence of individual AnnoTALE class members, yet, three exceptions were found, namely TalAK, TalBI and TalBY. A single version from each of these three classes was unable to activate the target gene. However, in all three cases, the failure to induce the target gene could be explained by an increased number of mismatched positions for the affected TALE versions (Appendix Table 3.2.5; Appendix Fig. 3.2.2).

One objective of the present study was the identification and evaluation of new putative *Xoc* TALE target genes. This was be done by analysing whether a direct interaction between a TALE and the promoter of a putative target gene takes place using GUS reporter assays. Interestingly, the confirmation for a direct interaction between target gene promoter and TALEs is still missing for most of the published target genes. To confirm that the published target genes are indeed directly induced by their respective TALEs, a 1 kbp fragment of each target promoter was PCR-amplified from the rice cultivar Nipponbare, cloned in front of the GUS reporter gene and transformed into *A. tumefaciens*. The resulting *Agrobacterium* strains were then either mixed with a second strain that encodes for one of the artificial *Xoc* TALEs reported to induce their expression or a negative control. If a promoter fragment contains an accessible target site for a TALE, the TALE should be able to bind and activate its expression.

Because they were already studied extensively, two of the 18 reported *Xoc* TALE target genes, *Os01g52130* and *Os06g46500*, were not analysed (Cernadas *et al*., 2014; Wilkins *et al*., 2015). From the remaining 16 reporter constructs, twelve were activated if combined with

their respective TALE classes, thus confirming a direct interaction between the promoter of the target gene and the TALE class (Fig. 3.2.7). Nine of these reporter constructs were induced by every tested TALE class member while three were only induced by some class members (Fig. 3.2.7). In these three cases, namely *Os07g06970*, *Os02g47660* and *Os03g07540,* a lack of induction was observed only for one class version, namely TalAK v1, TalBI v1 and TalBI v1, respectively, and in all three cases the expression pattern of the original RNA-Seq analysis also suggested that they are not capable of inducing this target (Appendix Table 3.2.5). In all three cases, the affected TALE versions contain also more mismatches in their predicted TALE binding sites than the other versions from the same class (Appendix Fig. 3.2.2). Finally, four reporter constructs did not allow for a confirmation of a direct TALE/target gene interaction. The first of them is class TalAD and its published target *Os02g15290*. Even inoculated alone, the TalAD-expression strain caused visible changes in the infected *N. benthamiana* leaf tissue, causing cell death-like symptoms that interefered with the experiment, and thus allowed only for inconclusive results (data not

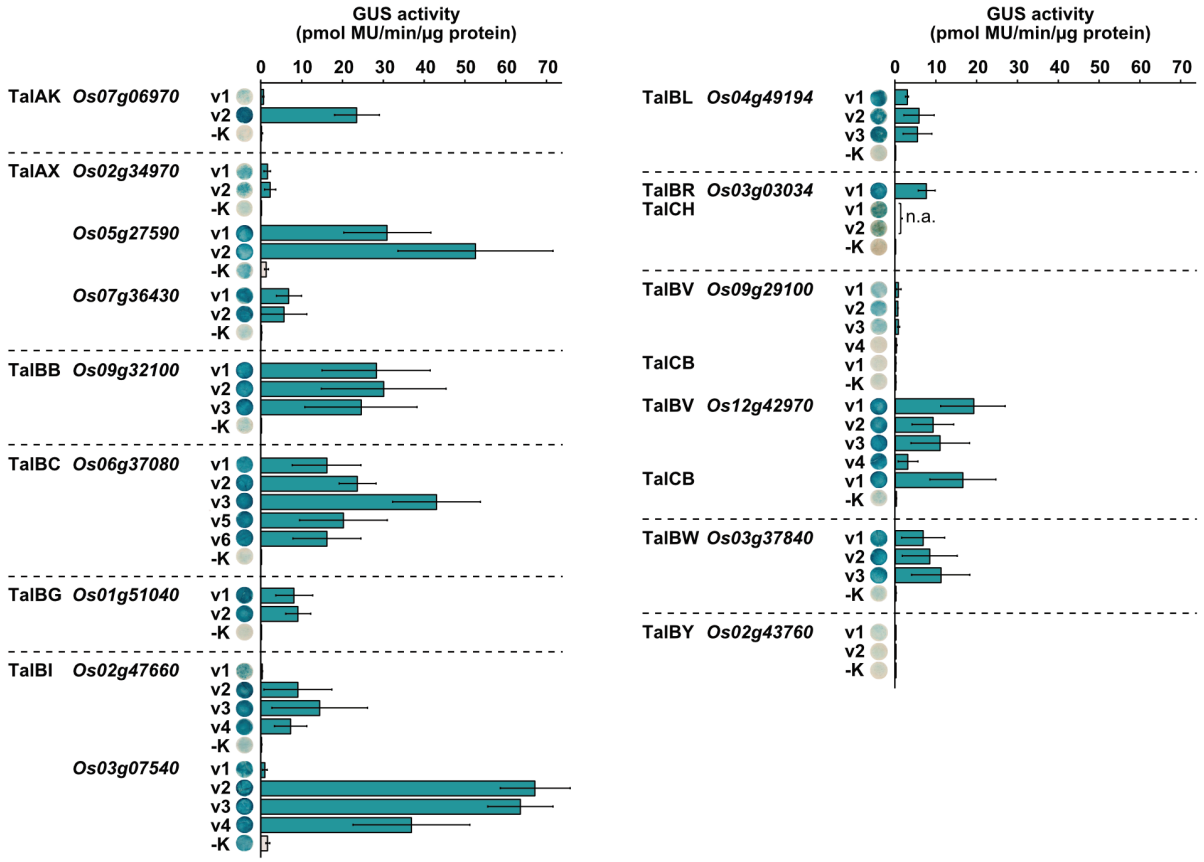


Figure 3.2.7: Results of the qualitative and quantitative GUS reporter assays for the published *Xoc* TALE targets and their respective TALEs. For each gene, a promoter fragment was amplified from the rice cultivar Nipponbare and cloned in front of a *GUS* reporter gene. The fragment encompassed always the first 1000 bp upstream of the first ATG of the CDS. The resulting reporter constructs were either inculated in combination with a TALE expression construct or without a TALE as a negative control (-K). A representative leaf disc from the qualitative GUS staining is shown. Error bars represent standard deviation (n = 3). Similar results were obtained in at least two independent experiments. Quantiative GUS values for two samples are not available (n.a.).

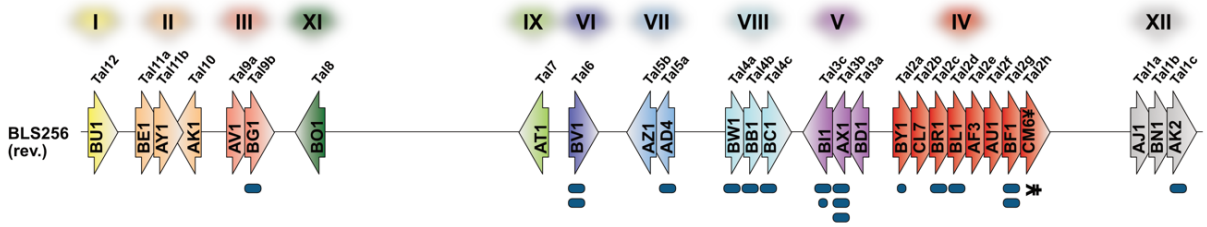


Figure 3.2.8: TALE classes in *Xoc* strain BLS256 and whether or not they have a published target gene. TALEs are sorted into genomic TALE clusters and are coloured accordingly. Each blue symbol represents one gene that was shown to be activated by this specific TALE class member. An oval represents a target gene that is addressed by all members of the TALE class, while a dot represents a target gene that is activated only by some of the class members (including the version present in BLS256). The ¥ symbol marks a TALE with truncated N- and C-terminal region. BLS is shown in reverse orientation (rev.).

shown). The remaining three promoter fragments were not upregulated at all, namely Os02g43760, the published target of class TalBY, *Os09g29100*, which is one of the reported targets of class TalBV and *Os02g34970,* which was reported as one of several targets for class TalAX (Fig. 3.2.7). This could mean that these four genes are indirect targets whose promoters do not contain a TALE binding site. However, it is also possible, even more likely, that the chosen 1 kbp promoter fragments are somehow flawed or incomplete or that they lack essential rice-specific factors needed for their functionality, and thus they might simply not be suited for an analysis in *N. benthamiana*.

In spite of the 18 published *Xoc* TALE target genes in rice, many *Xoc* TALEs remain without a target. From the 28 TALE classes/TALEs found in BLS256, the first and best-studied among the 10 so far sequenced *Xoc* strains, 15 are still without a known target gene (Fig. 3.2.8). Moreover, if all 10 completely sequenced *Xoc* strains are taken together, the number

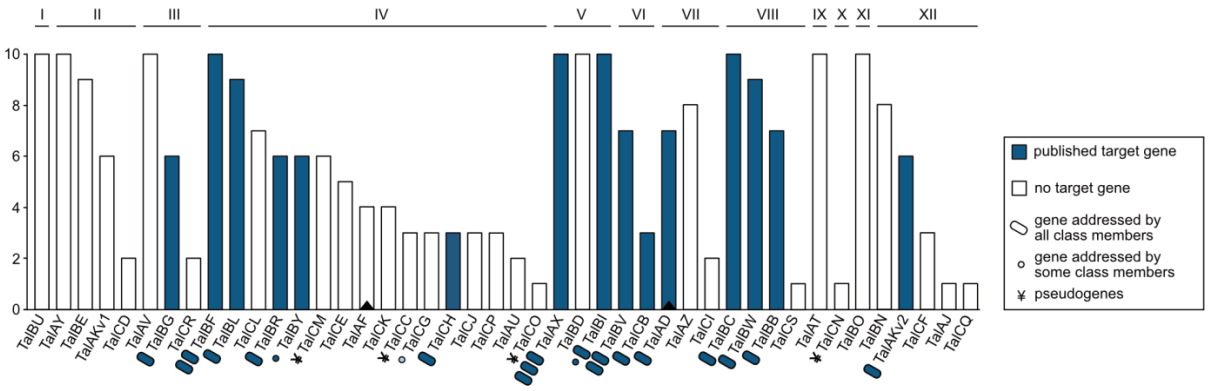


Figure 3.2.9: Overview showing the abundance and distribution of the known *Xoc* TALE classes as well as their published target genes. *Xoc* TALE classes are sorted according to their cluster affiliation and their abundance within each genomic TALE cluster. White Bars represent TALE classes without a target gene, those coloured in dark blue indicate a TALE class with at least one published target gene. Each symbol represents one target gene, with an oval standing for a target gene that is addressed by all members of the TALE class while a dot stands for a target that is only activated by some class members. TALE classes labelled with the ¥ symbol contain only TALEs with incomplete N- and C-terminal regions, and thus they are likely unable to bind to DNA. The small triangles indicate the two TALE classes that are shared between *Xoc* and *Xoo* (TalAF and TalAD).

of TALE classes without target gene rises even further, totaling 25 out of 40 (Fig. 3.2.9). Consequentely, additional effort should be put into the identification and confirmation of genes that are specificially induced by those TALEs. Identifying their targets and analysing their function will be vital for our in-depth understanding of the *Xoc* infection mechanisms and should shed light on the many still unknown processes controlled by *Xoc* TALEs.

3.2.4 Identification of novel *Xoc* TALE target genes in rice

3.2.4.1 Evaluation of previously predicted putative *Xoc* TALE target genes

An RNA-Seq analysis of rice after inoculation with *Xoc* strains carrying a known TALE repertoire is a promising method to identify TALE target genes. By combining the patterns observed in the expression data with the presence or absence of individual TALEs/TALE classes and bioinformatical predictions of putative TALE binding sites in promoter regions, conclusions can be drawn with regard to which gene could be regulated as a direct response to a specific TALE. Such an analysis was already done with all 10 completely sequenced *Xoc* strains but it yielded only targets for about half of the known *Xoc* TALE classes (Wilkins *et al*., 2015). However, by re-evaluating this dataset, using other parameters and a different prediction program, Erkes and colleagues identified nearly 30 new putative *Xoc* TALE target genes that were not discovered in the original study (Erkes *et al*., 2017). These putative targets are a valuable resource since they are, so far, only predictions and were not yet analysed experimentally in any way. Consequentely, their analysis could allow for a fast and straightforward confirmation of multiple new *Xoc* TALE targets.

After briefly evaluating them for their plausibility, roughly half of the initially 30 putative new target genes were discarded, some for not having an initial thymine preceding the predicted TALE binding site, some for containing a strong mismatch at the position of the very first repeat and some for showing extreme differences between strains that share identical TALE versions. This left 16 genes (Appendix Table 3.2.6), from which 10 yielded a promoter fragment via PCR. These 10 were used for the generation of reporter constructs and the subsequent analysis via GUS reporter assay. Two of the 10 reporter constructs, those derived from *Os01g32439* and *Os06g17020*, showed no TALE-dependent induction. A third one, derived from *Os07g26520*, showed a strong basal activity and yielded no additional induction in response to its respective TALEs (Fig. 3.2.10). The remaining seven constructs showed a strong TALE-dependent upregulation, and thus these genes can be regarded as newly confirmed target genes.

The first of these, *Os02g51110*, encodes for the aquaporin OsNIP2;1 and its promoter fragment was strongly induced in combination with a member of class TalAV (Fig. 3.2.10). The second one is the indole-3-acetic acid-amido synthetase OsGH3;4 which is encoded by

the gene *Os05g42150*. Its promoter fragment was activated by all three members of class TalCL (Fig. 3.2.10). In a previous study, *Os05g42150* has already been identified as the single most significantly induced rice gene in response to the *Xoc* strain BLS256 but since no adequate TALE binding element was found within its promoter at that time, it was disregarded as a possible TALE target gene (Cernadas *et al*., 2014). The reason for this lies very likely in the fact that the putative TalCL binding site in its promoter contains 8 mismatches, a number that disqualifies any gene as potential TALE target, at least under normal conditions (Fig. 3.2.10). However, TalCL members contain an unusually large number of repeats (29.5) as well as many strong RVDs (up to 13), two factors that are likely to render them more tolerant against mismatches than most other TALEs. In fact, a brief target prediction using AnnoTALE revealed that even the best putative target genes for TalCL within the Nipponbare genome contain binding elements with at least 8 mismatcheds (data not shown), thus indicating that the TalCL binding site found within the *Os05g42150* promoter is actually among the best of all the potential target gene candidates. A second gene with a similarly high number of mismatches was also confirmed as target for TalCL, *Os10g39840*, and it encodes for the glycosyl hydrolase OsXTH20. Curiously, its expression

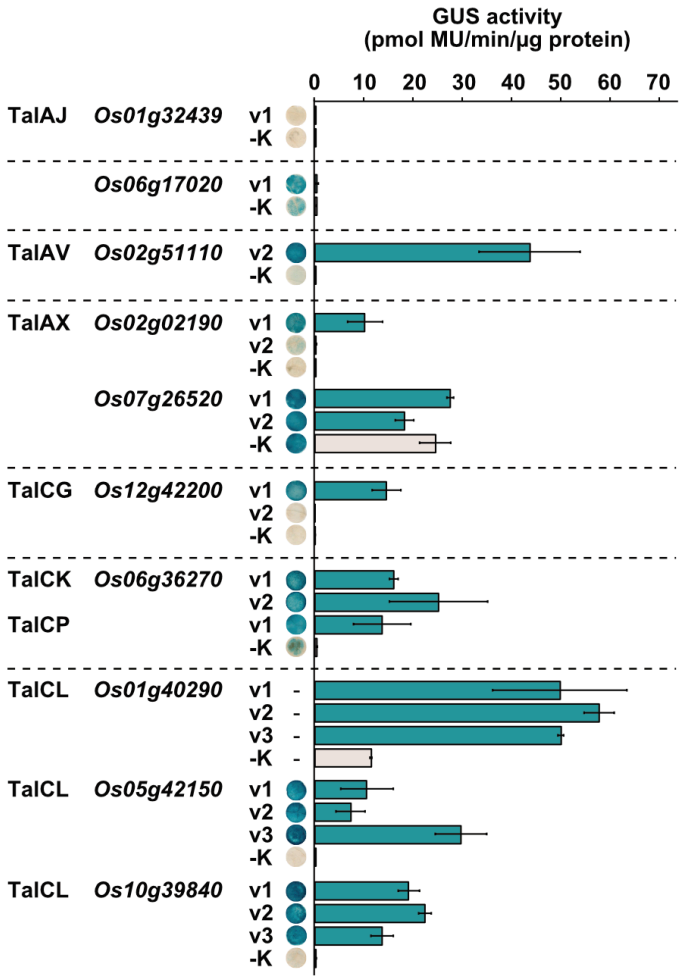


Figure 3.2.10: Results of the quantitative and qualitative GUS reporter assay for those putative *Xoc* TALE target genes that were selected for experimental analysis. For each gene, a promoter fragment was amplified from the rice cultivar Nipponbare and cloned in front of a GUS reporter gene. The fragment encompassed always the first 1000 bp upstream of the first ATG of the CDS. The resulting reporter constructs were either inculated in combination with a TALE expression construct or without a TALE as a negative control (-K). A representative leaf disc from the qualitative GUS staining is shown. Error bars represent standard deviation (n = 3). Similar results were obtained in at least two independent experiments.

pattern in the original RNA-Seq experiment suggested a correlation for only two of the three versions of TalCL (v1 and v2; Appendix Table 3.2.6). However, while TalCL v3 contains indeed an additional mismatch, this mismatch is placed at the very end of the TALE binding site and is therefore not expected to be of major impact (Fig. 3.2.11). Consequently, all three members of TalCL induced the reporter construct with this promoter fragment (Fig. 3.2.10). With *Os01g40290*, it was possible to confirm even a third target gene for class TalCL. While its reporter construct showed a rather strong GUS activity even in the absence of any TALE, it yielded an even stronger response if combined with the members of TalCL (Fig. 3.2.10). Interestingly, TalCL is also one of the TALE classes whose predicted optimal target sequence suggested the possibility for a shared target gene with a second class, in this case TalCJ (chapter 3.2.1). Yet, upon closer inspection does neither the expression data from the RNA-Seq experiment (Appendix Table 3.2.6) nor the number of mismatches in the putative TalCJ binding sites suggest that the members of TalCJ are able to activate *Os05g42150* or *Os10g39840* (Fig. 3.2.11). However, TalCJ and TalCL do share a similar mismatch count for *Os01g40290* (Fig. 3.2.11) and its RNA-Seq expression pattern suggests that it is also highly expressed in those *Xoc* strains that lack TalCL but carry a member of TalCJ (Appendix Table 3.2.6). Unfortunately, the cloning of TalCJ into pSKA2 was not successful, and thus it was not possible to analyse its effects on the promoter of *Os01g40290* experimentally.

The fifth newly confirmed *Xoc* TALE target gene is the putative receptor-like kinase *Os06g36270*. While Erkes and colleagues suggest a correlation between its upregulation and the presence of class TalCK members, it was also induced in three strains that lack TalCK and no suitable other TALEs were identified to explain this observation (Appendix Table 3.2.6; Erkes *et al*., 2017). However, a large overlap in the predicted optimal target sites for TalCK and TalCP was found earlier (chapter 3.2.1) and a closer inspection of the promoter region of *Os06g36270* confirmed that the binding site of the TalCK members might in fact allow also for the binding of TalCP (Fig. 3.2.11). This was confirmed via GUS assay, demonstrating that the promoter of *Os06g36270* is indeed responding to the presence of members from TalCK and TalCP (Fig. 3.2.10).

The sixth newly confirmed *Xoc* TALE target gene, *Os12g42200*, encodes for a putative cation-hydrogen exchanger. For this gene, the published RNA-Seq data and the mismatch analysis indicated that only one of the two members of TalCG, namely TalCG v1, would be able to induce the reporter construct (Fig. 3.2.11; Appendix Table 3.2.6), and indeed, only TalCG v1 was found to activate this particular reporter construct (Fig. 3.2.10).

The seventh and last newly confirmed target gene is *Os02g02190*. It encodes for the nitrate transporter OsNRT2.2 and it was predicted to contain a binding element for TalAX v1 but not v2 or v3 (Fig. 3.2.11; Appendix Table 3.2.6). This prediction was confirmed as well by

demonstrating that the respective GUS reporter construct can be induced by TalAX v1 but not by v2 (Fig. 3.2.10).

Experimentally evaluating the results from a previously published TALE target prediction allowed for the confirmation of seven new target genes, six of which are induced by *Xoc* TALE classes that were previously without known target (Fig. 3.2.12). Given the comparably high rate of success with these 10 genes, it might be worthwile to repeat the cloning attempts for the remaining six promoter fragments from the initial list of 16 and analyse them as well (Appendix Table 3.2.6; Appendix Fig. 3.2.3).

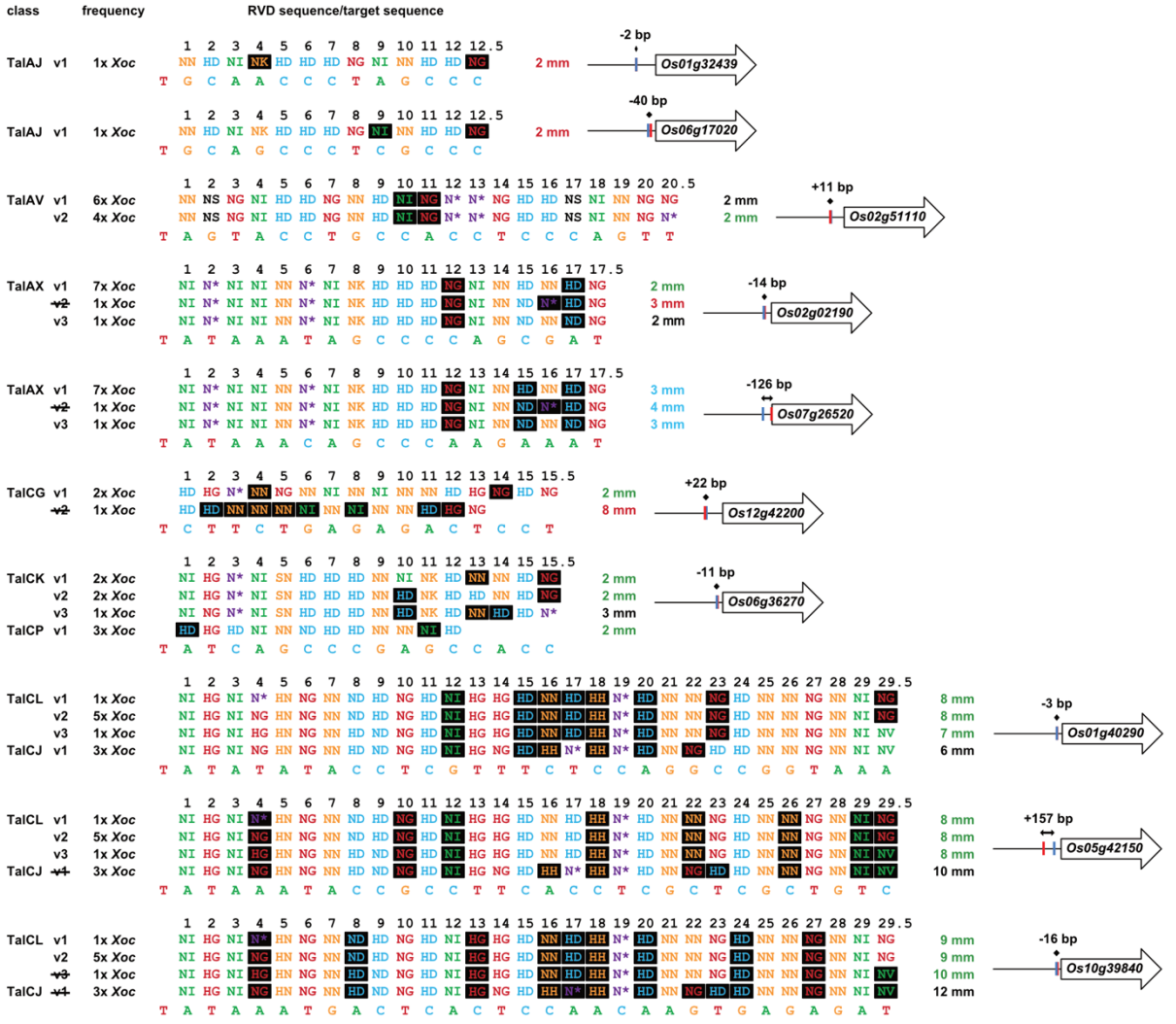


Figure 3.2.11: Predicted TALE binding sites within the promoter regions of those putative TALE target genes that were selected for experimental validation. TALEs are depicted by their RVDs and they are aligned to the predicted EBEs found in the promoter regions of their putative target genes. RVDs are shown for all unique TALE class versions. Class versions whose presence did not yield an induction of the target gene in the original RNA-Seq experiment are crossed out (RNA-Seq originally by Wilkins *et al.*, 2015; re-analysis and target predictions by Erkes *et al.*, 2017). The frequency of occurrence within the 10 completely sequenced *Xoc* strains is indicated for each TALE class version. Mismatched RVDs are boxed black. The number of predicted mismatches (mm) is shown next to the RVD sequence and its colour indicates the activity of this version in the GUS reporter assay. Green indicates solid activity, red no activity, blue a strong autoactivity of the reporter construct alone and black that this version was not tested. The putative target genes are depicted as arrows and a schematic of the first 1000 bp of each promoter is shown (beginning upstream of the first ATG). The relative distance between the TALE binding site (blue) and the annotated transcriptional start site (red) is indicated in base pairs (bp).

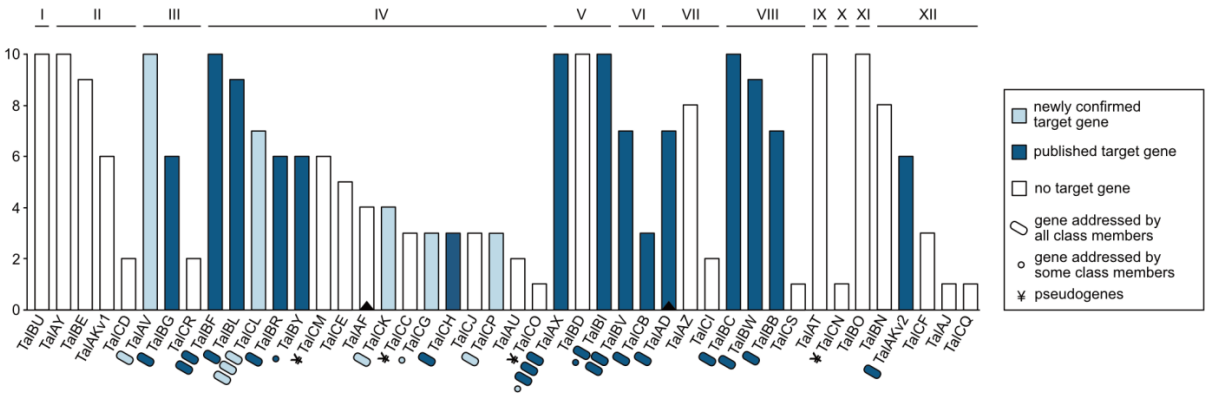


Figure 3.2.12: Overview showing the abundance and distribution of the known *Xoc* TALE classes, their published target genes as well as the newly confirmed ones. *Xoc* TALE classes are sorted according to their cluster affiliation and their abundance within each genomic TALE cluster. White Bars represent TALE classes without a target gene, those coloured in dark blue a TALE class with at least one published target gene while those in a light blue colour indicate TALE classes for which a new target gene was identified. Each symbol represents one target gene, with an oval standing for a target gene that is addressed by all members of the TALE class and a dot representing a target that is only activated by some class members. TALE classes labelled with the ¥ symbol contain only TALEs with incomplete N- and C-terminal regions and thus they are likely unable to bind to DNA. The small triangles indicate the two TALE classes that are shared between *Xoc* and *Xoo* (TalAF and TalAD).

3.2.4.2 Analysis of novel putative *Xoc* TALE targets identified via comparative mismatch analysis

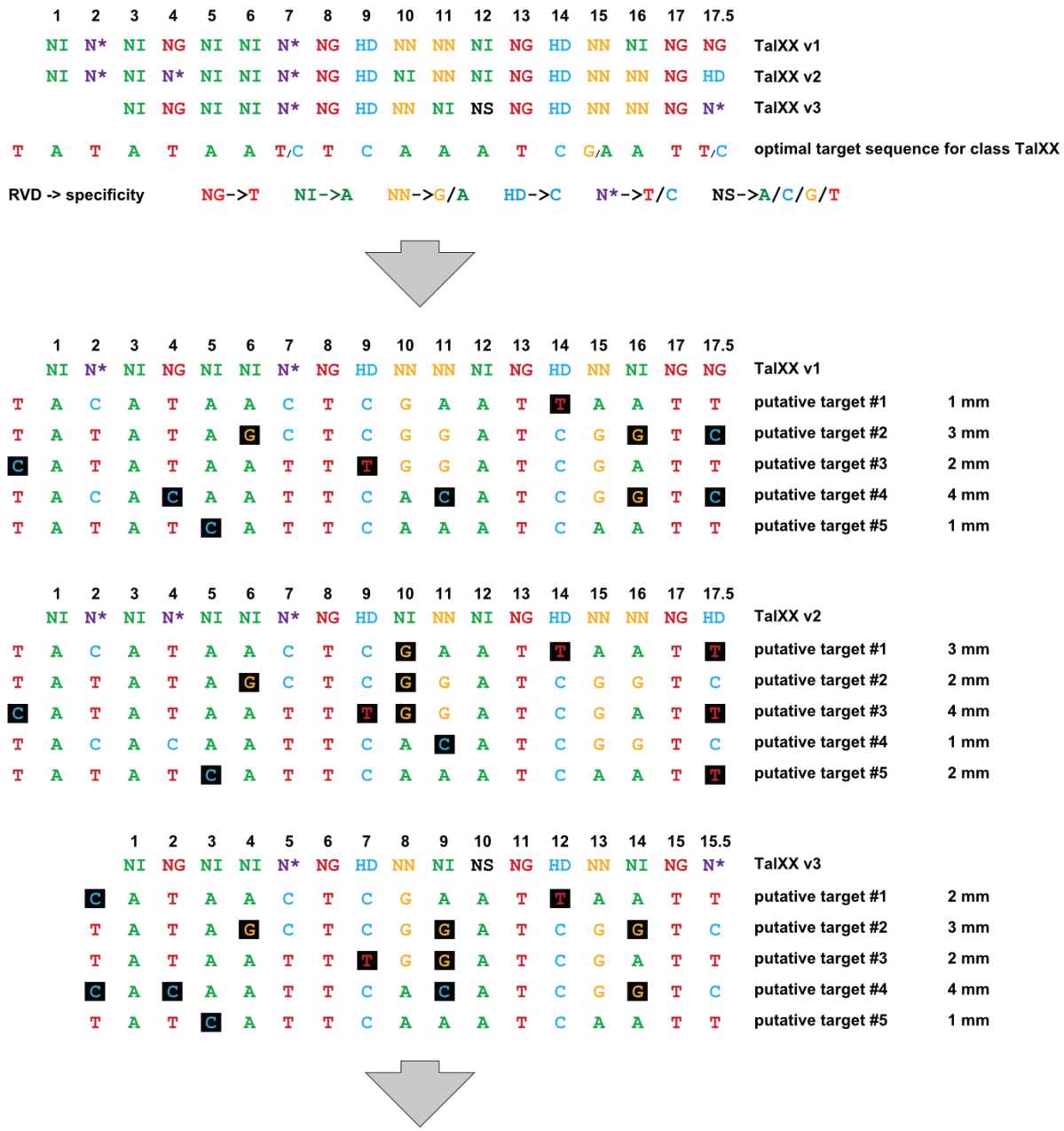
Based on the predictions published in Erkes *et al*., 2017, it was possible to confirm seven new *Xoc* TALE target genes. Nonetheless, many *Xoc* TALE classes remain without known function, among them extremely conserved and very rare ones alike. However, since the DNA-binding mechanism of TALEs is extremely regular, it is possible to use their RVD sequences to predict optimal target sequences. With this, it is possible to search the plant host genome for promising novel candidate TALE binding sites even without relying on gene expression data. Unfortunately, several factors complicate TALE target predictions and reduce their reliability. One factor is the frequent occurrence of RVDs that tolerate multiple nucleotides. Problematic are also very long TALEs and those with multiple strong RVDs since they likely allow for a high number of mismatches. To obtain more reliable results, scientists addressed this problem bioinformatically and designed programs that are not only able to predict potential TALE binding sites within the genome of the targeted host plant but also to evaluate the likelihood of them being a real target based on the number and position of the mismatches. To name a few them: AnnoTALE (Grau *et al*., 2016), TALgetter (Grau *et al*., 2013), TALE-NT 2.0 (Doyle *et al*., 2012) or Talvez (Pérez-Quintero *et al*., 2013).

While AnnoTALE was used in the present study, all predictions were manually re-evaluated. This was done to ensure that only the most promising putative TALE target genes were selected for experimental analysis. For example, the mismatched positions for some rarely occurring RVDs like HH, NC and QD had to be corrected. These RVDs were experimentally

shown to preferentially accept G, A and C (Yang *et al*., 2014; Miller *et al*., 2015; Grau *et al*., 2016) but are incorporated into AnnoTALE as targeting C, C and A, respectively. Affected were also the RVDs NS and HA, which are predicted by AnnoTALE to match only to A and C, respectively, but which were shown to tolerate all four nucleotides reasonably well (Yang *et al*., 2014; Miller *et al*., 2015). A last example is the frequently occurring RVD NN. AnnoTALE regards only the nucleotide G as a perfect match since NN functions only in this pairing as a strong RVD and thus, contributes positively to the general DNA-binding affinity of the TALE (Streubel *et al*., 2012). However, studies could demonstrate that NN is capable of accepting both purines, G and A, without any aparant penalty (Streubel *et al*., 2012; Yang *et al*., 2014; Miller *et al*., 2015). Manually evaluating and correcting the TALE target predictions with regard to these criteria should further improve the reliability of the obtained results and thus, hopefully, also the number of newly confirmed targets.

So far, only a few *X. oryzae* TALE classes were shown to address multiple target genes, namely TalAX, TalBI, TalBV (Cernadas *et al*., 2014) and TalCL (see previous chapter), but it is widely assumed that most TALEs induce more than one gene. However, given the still small number of TALE target genes shown to be involved in virulence, many scientists also assume that, in most cases, only one target per TALE/TALE class actually affects the virulence and survival of the pathogen in a positive way. To identify those targets that have the highest probability to possess a biological relevant role, it seemed reasonable to focus on those genes that have a putative binding site that is likely addressed by all or at least most members of a TALE class. A gene that is addressed by multiple members of a TALE class, despite them having differences in their RVD composition/DNA specificity, could indicate an underlying evolutionary pressure to keep this gene as a TALE target. Consequentely, such a target might more likely possess a benefical function for the virulence of *Xoc*. Therefore, the manual evaluation was also used to select preferentially those target genes that show the smallest number of mismatches for all individual class versions combined or those whose mismatches should have the least severe effects. For an illustration of this selection process, see Fig. 3.2.13. Since an annotated function, even a putative one, should increase the chance to identify the effects caused by the TALE-dependent overexpression of a gene, putative target genes with unknown function (expressed or hypothetical proteins) as well as genes related to mobile genetic elements (like transposons) were mostly discarded. However, in thoses cases in which such a gene contained an extremely well suited, putative TALE binding site, it was included into the analysis nonetheless.

Taken together, around 100 putative TALE target genes were selected for experimental validation. Around 85 of them yielded a promoter fragment, and thus allowed for the successfull generation of a reporter construct. In a qualitative GUS assay, roughly half of them yielded either no GUS activity at all or the observed GUS activity of the reporter



	TalXX v1	TalXX v2	TalXX v3	selected for experimental evaluation
putative target #1	1 mm	3 mm	2 mm (T0)	no
putative target #2	3 mm	2 mm	3 mm	no
putative target #3	2 mm (T0)	4 mm (T0)	2 mm	no
putative target #4	4 mm	1 mm	4 mm (T0)	no
putative target #5	1 mm	2 mm	1 mm	yes

Figure 3.2.13: Schematic showing the manual evaluation process used to select those newly predicted putative target genes with the highest chance of being addressed by most or all versions of a TALE class. Individual predictions were performed for each unique class version using AnnoTALE (Grau *et al.*, 2016). The top 10-50 genes from each prediction were manually evaluated. Those target genes with the least amount of mismatched positions were then checked for their mismatches with all versions other this particual TALE class. Only those putative targets that had a low number of mismatches for all class members were considered for experimental validation. Mismatched positions are boxed black. TALEs are shown as RVDs. Note: TALE class TalXX is a hypothetical example made up to illustrate the manual selection process.

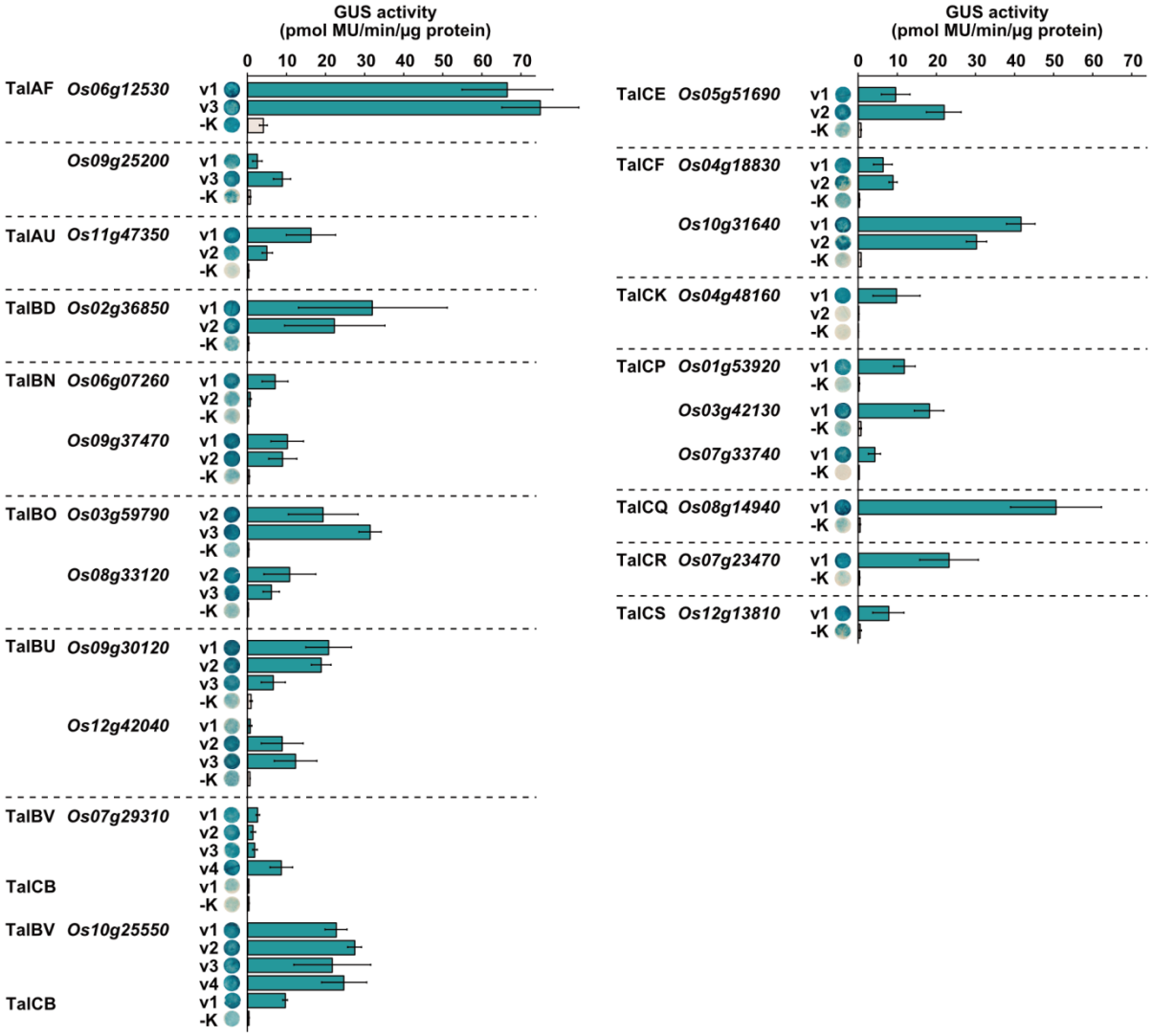
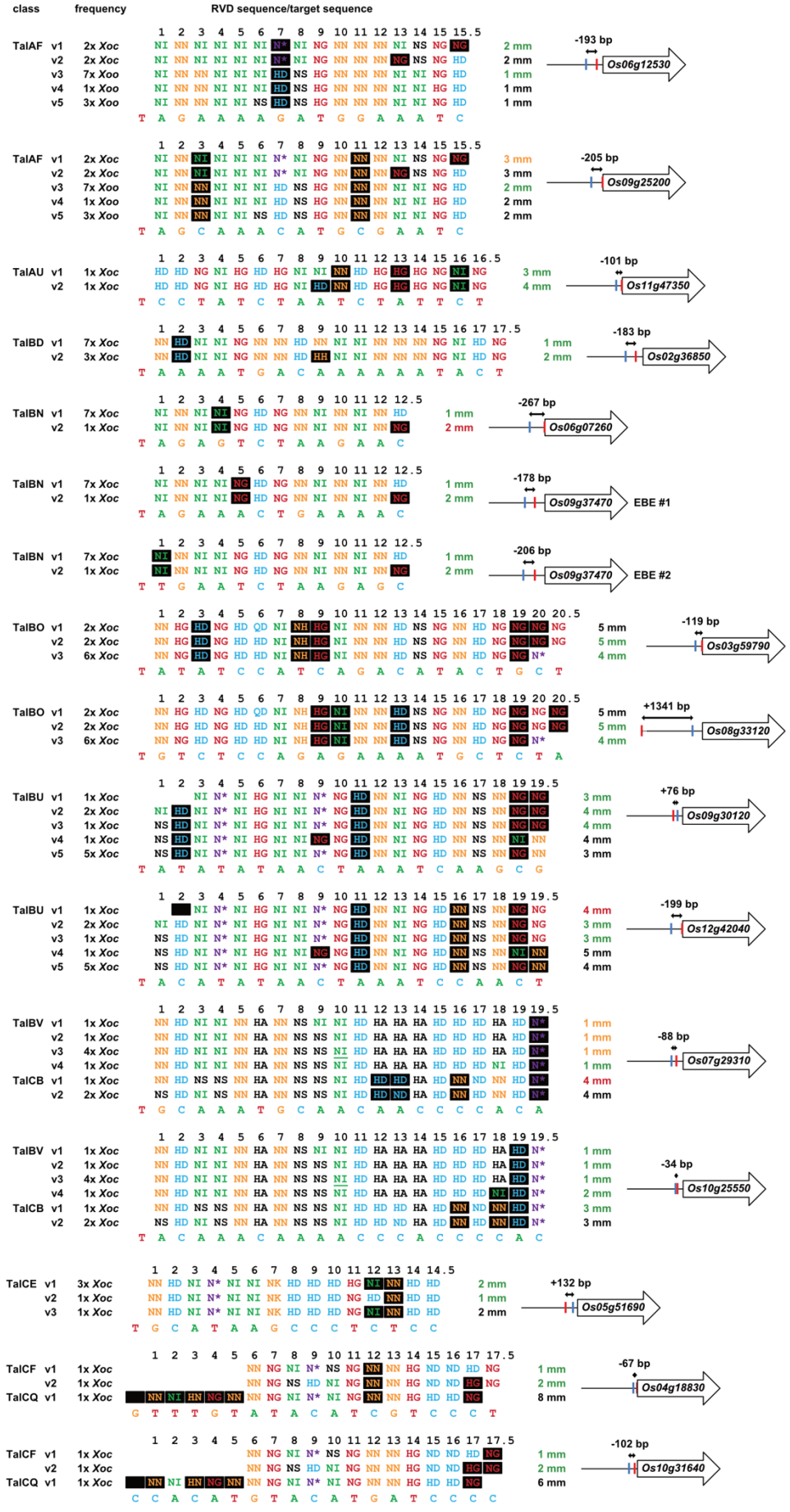


Figure 3.2.14: Results of the qualitative and quantitative GUS reporter assays for all newly identified rice genes whose promoters showed a clear TALE-dependent induction. For each gene, a promoter fragment was amplified from the rice cultivar Nipponbare and cloned in front of a *GUS* reporter gene. The fragment encompassed always the first 1000 bp upstream of the first ATG of the CDS. The resulting reporter constructs were either inculated in combination with a TALE expression construct or without a TALE as a negative control (-K). A representative leaf disc from the qualitative GUS staining is shown. Error bars represent standard deviaton (n = 3). Similar results were obtained in at least two independent experiments.

construct was not further increased if they were combined with their respective TALEs (Appendix Fig. 3.2.4). The other half showed at least a minimal response to the presence of a particular TALE, and thus those constructs were inoculated again and subjected to a quantitive GUS analysis. Twelve of them were induced by the presence of a particular *Xoc* TALE but to such a low extent that they were considered as relevant (Appendix Fig. 3.2.5). Whether their induction is stronger in the natural rice-*Xanthomonas* pathosystem will have to remain the subject of further studies. This left 22 reporter constructs with a strong response to a specific *Xoc* TALE or TALE class (Fig. 3.2.14). Those 22 reporter constructs were activated by 15 different TALE classes. From these, 12 had previously no target gene.


class
frequency
RVD sequence/target sequence
TalAF v1 2x Xoc
v2 2x Xoc
v3 7x Xoo
v4 1x Xoo
v5 3x Xoo
2 mm
2 mm
1 mm
1 mm
-193 bp
Os06g12530
TalAF v1 2x Xoc
v2 2x Xoc
v3 7x Xoo
v4 1x Xoo
v5 3x Xoo
3 mm
3 mm
2 mm
2 mm
2 mm
-205 bp
Os09g25200
TalAU v1 1x Xoc
v2 1x Xoc
3 mm
4 mm
-101 bp
Os11g47350
TalBD v1 7x Xoc
v2 3x Xoc
1 mm
2 mm
-183 bp
Os02g36850
TalBN v1 7x Xoc
v2 1x Xoc
1 mm
2 mm
-267 bp
Os06g07260
TalBN v1 7x Xoc
v2 1x Xoc
1 mm
2 mm
-178 bp
Os09g37470
EBE #1
TalBN v1 7x Xoc
v2 1x Xoc
1 mm
2 mm
-206 bp
Os09g37470
EBE #2
TalBO v1 2x Xoc
v2 2x Xoc
v3 6x Xoc
5 mm
5 mm
4 mm
-119 bp
Os03g59790
TalBO v1 2x Xoc
v2 2x Xoc
v3 6x Xoc
5 mm
5 mm
4 mm
+1341 bp
Os08g33120
TalBU v1 1x Xoc
v2 2x Xoc
v3 1x Xoc
v4 1x Xoc
v5 5x Xoc
3 mm
4 mm
4 mm
4 mm
3 mm
+76 bp
Os09g30120
TalBU v1 1x Xoc
v2 2x Xoc
v3 1x Xoc
v4 1x Xoc
v5 5x Xoc
4 mm
3 mm
3 mm
5 mm
4 mm
-199 bp
Os12g42040
TalBV v1 1x Xoc
v2 1x Xoc
v3 4x Xoc
v4 1x Xoc
TalCB v1 1x Xoc
v2 2x Xoc
1 mm
1 mm
1 mm
1 mm
4 mm
4 mm
-88 bp
Os07g29310
TalBV v1 1x Xoc
v2 1x Xoc
v3 4x Xoc
v4 1x Xoc
TalCB v1 1x Xoc
v2 2x Xoc
1 mm
1 mm
1 mm
2 mm
3 mm
3 mm
-34 bp
Os10g25550
TalCE v1 3x Xoc
v2 1x Xoc
v3 1x Xoc
2 mm
1 mm
2 mm
+132 bp
Os05g51690
TalCF v1 1x Xoc
v2 1x Xoc
TalCQ v1 1x Xoc
1 mm
2 mm
8 mm
-67 bp
Os04g18830
TalCF v1 1x Xoc
v2 1x Xoc
TalCQ v1 1x Xoc
1 mm
2 mm
6 mm
-102 bp
Os10g31640

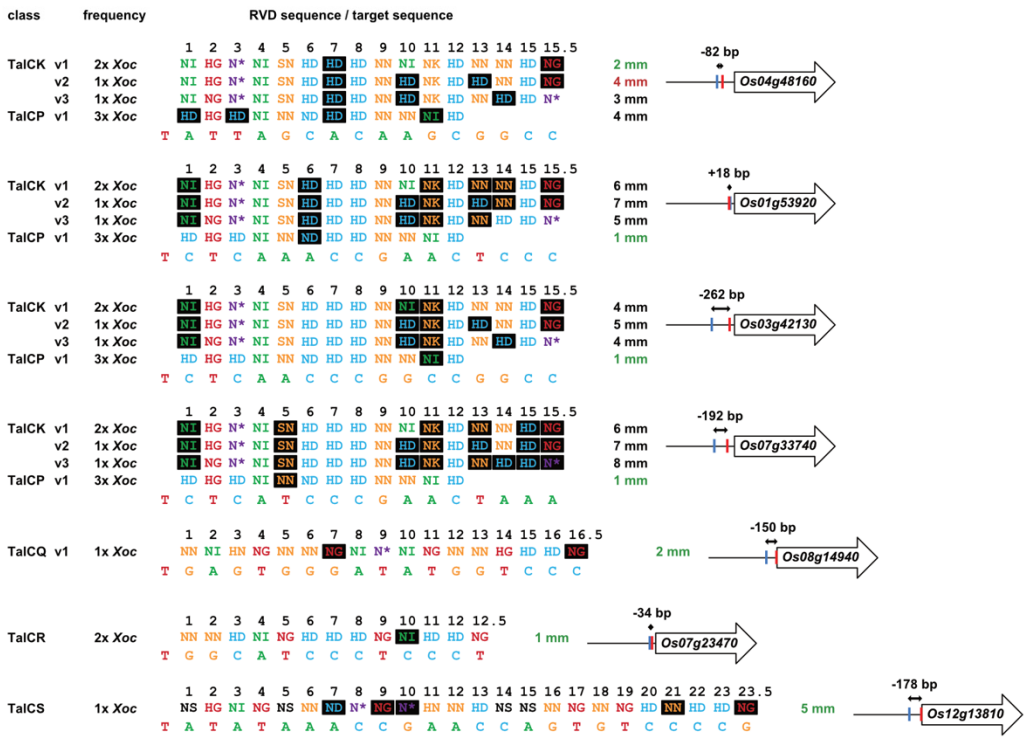


Figure 3.2.15: Predicted TALE binding sites within the promoter regions of all those newly identified rice genes whose promoters showed a clear TALE-dependent induction. TALEs are depicted by their RVDs and are aligned to the predicted TALE binding sites found in the promoter regions of the putative target genes. RVDs are shown for all unique TALE class versions. The frequency of occurrence within the 10 completely sequenced *Xoc* strains is indicated for each TALE class version. Mismatched RVDs are boxed black. The number of predicted mismatches (mm) is shown next to the RVD sequence and its colour indicates the activity of this version in the GUS reporter assay. Green indicates solid activity, orange a weak activity (of less than 5pmol MU/min/µg protein), red no activity and black that this version was not tested. The genes are depicted as arrows and a schematic of the first 1000 bp of each promoter is shown (beginning upstream of the first ATG). The relative distance between the TALE binding site (blue) and the annotated transcriptional start site (red) is indicated in base pairs (bp).

Six of the tested TALE classes were able to address two different reporter constructs, namely TalAF, TalBN, TalBO, TalBU, TalBV and TalCF. One class, TalCP, activated even three different ones, thus further supporting the hypothesis that many natural TALEs likely induce not only one but multiple genes. Interestingly, 19 out of the 22 genes have a promoter that was induced by all tested members of a TALE class. While in some cases not all possible versions of a TALE class were analysed, those class versions that remain untested show comparable mismatches in their predicted binding element, thus suggesting that these targets will be addressed by them as well (Fig. 3.2.15). This result strongly supports the reliability of the AnnoTALE class system since it demonstrates that different members from the same TALE class often share at least one target gene. These results also support the chosen selection process, indicating that it was not only suited to identify genes whose promoter regions contain a TALE binding site but also to select for genes that are addressed by all or most members of a TALE class.

By combining bioinformatic predictions with a manual evaluation process, it was possible to identify 22 novel rice genes whose promoter regions showed a strong TALE-dependent induction in a GUS assay. While this proves that their promoters contain a suitable binding site for one or several *Xoc* TALEs, it remains to be seen whether these genes are also induced in rice and, if this is the case, how strongly. Until then, these 22 genes cannot be regarded as fully confirmed target genes. However, the fact that their reporter constructs yielded GUS values comparable to those derived from the published *Xoc* TALE targets, suggests that they should be induced *in planta* as well. Their addition to the list of TALE target genes would lower the number of *Xoc* TALE classes without a target gene significantly, reducing them to an overall of only 9 (Fig. 3.2.16; Table 3.2.3).

Curiously, despite analysing multiple putative target genes for most TALE classes, sometimes as many as eigth different ones, some TALE classes still remain without even a single promising putative target gene and unfortunately, with TalAT, TalAY, TalAZ and TalBE, four of them belong to the most conserved TALE classes found in *Xoc* (Fig. 3.2.16). However, it is important to note that not all target genes with a promising mismatch pattern could be tested since it was not possible to amplify and clone all initially selected promoter fragments. Intensifying the effort to clone these promoter fragments or analysing these genes with an alternative method should therefore be a priority. It might even be worthwhile to repeat the prediction and evaluation process for these classes using another prediction program since it might allow for the identification of new promising target gene candidates that were overlooked.

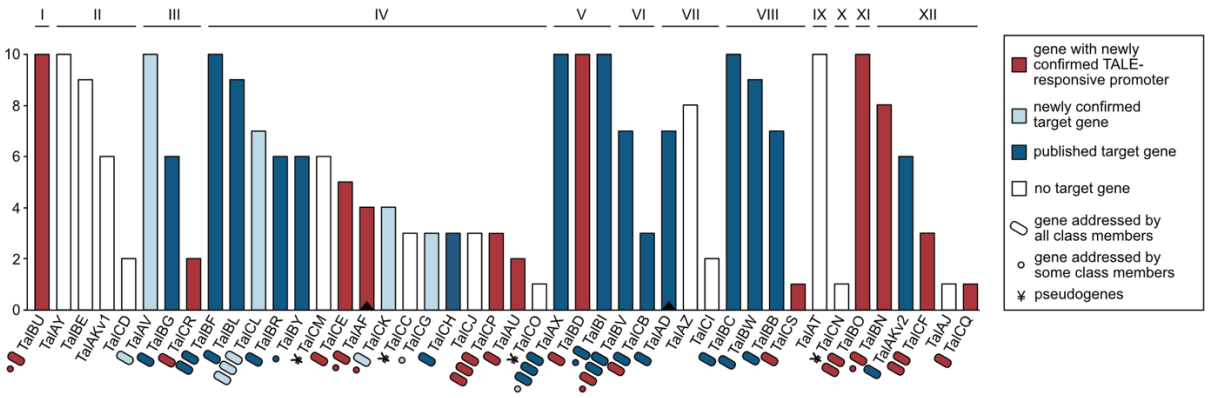


Figure 3.2.16: Updated overview showing the abundance and distribution of all *Xoc* TALE classes and their target genes. TALE classes are sorted by abundancy and alphabet. A white bar indicates a TALE class still without target gene, dark blue colour signifies a TALE class with at least one published target gene while a light blue colour indicates TALE classes with at least one newly confirmed target gene that was predicted by Erkes *et al*., 2017. Bars shown in red represent TALE classes with at least one gene that was predicted via comparative mismatch analysis and confirmed via GUS assay. Each symbol represents one target gene, with an oval standing for a target gene that is addressed by all members of the TALE class while a dot stands for a target that is only activated by some class members. TALE classes labelled with the ¥ symbol contain only TALEs with incomplete N- and C-terminal regions that are likely unable to bind to DNA. The small triangles indicate the two TALE classes that are shared between *Xoc* and *Xoo* (TalAF and TalAD).

Table 3.2.3: Updated list of confirmed and partially confirmed *Xoc* TALE targets genes.

TALE class and versions [1]	locus ID [2]	annotated function	reference
AD v1 v2	*Os02g15290*	OsVQ7 (VQ domain containing protein)	(1) (2)
AF v1 **v2** v3 **v4 v5**	*Os06g12530*	ROX3 (CS-domain containing protein)	this study
AF v1 **v2** v3 **v4 v5**	*Os09g25200*	protein binding protein	this study
AK v2	*Os07g06970*	OsHEN1 (metyhtransferase)	(1) (2)
AV v1 v2	*Os02g51110*	OsNIP2;1 (aquaporin)	(3) (4) and this study
AU v1 v2	*Os11g47350*	Beta-D-Xylosidase	this study
AX v1 **v2** v3 v4	*Os02g02190*	OsNRT2.2 (nitrate transporter)	(3) and this study
AX v1 v2 v3 v4	*Os02g34970*	no apical meristem protein	(1) (2)
AX v1 v2 v3 v4	*Os05g27590*	wound-induced protein WI12 like	(1) (2)
AX v1 v2 v3 v4	*Os07g36430*	expressed protein	(1) (2)
BB v1 v2 v3 v4	*Os09g32100*	expressed protein	(1) (2)
BC v1 v2 v3 v4 v5 v6	*Os06g37080*	OsAAO1 (L-ascorbate oxidase)	(1) (2)
BD v1 v2	*Os02g36850*	oxygen evolving enhancer protein 3	this study
BF v1 v2 v3	*Os01g52130*	OsSULTR3;6 (sulfate transporter)	(1) (2)
BF v1 v2 v3	*Os06g46500*	monocopper oxidase	(1) (2)
BG v1 v2	*Os01g51040*	16k transmembrane protein	(1) (2)
BI **v1** v2 v3 v4	*Os02g47660*	OsbHLH079 (basic helix-loop-helix TF)	(1) (2)
BI v1 v2 v3 v4	*Os03g07540*	OsbHLH153 (basic helix-loop-helix TF)	(1) (2)
BL v1 v2 v3	*Os04g49194*	OsDOX-2	(1) (2) (4)
BN v1 **v2**	*Os06g07260*	OsWAK62 (wall-associated kinase)	this study
BN v1 v2	*Os09g37470*	OsSAUR52 (small auxin upregulated RNA)	this study
BO v1 v2 v3	*Os03g59790*	EF-hand family protein	this study
BO v1 v2 v3	*Os08g33120*	RNA recognition motif protein	this study
BR v1 v2 v3 v4 **CH** v1 v2	*Os03g03034*	OsDOX-1	(1) (2) (4)
BU v1 v2 v3 **v4 v5**	*Os09g30120*	cellulose synthase-like protein E1	this study
BU **v1** v2 v3 **v4 v5**	*Os12g42040*	OsWAK126 (wall-associated kinase)	this study
BV v1 v2 v3 v4	*Os07g29310*	OsSAUR30 (small auxin upregulated RNA)	this study
BV v1 v2 v3 v4	*Os09g29100*	Cyclin-4D-1 (cyclin)	(1) (2)
BV v1 v2 v3 v4 **CB** v1 **v2**	*Os10g25550*	OsRLCK298 (tyrosine protein kinase)	this study
BV v1 v2 v3 v4 **CB** v1 v2	*Os12g42970*	GATA zinc finger	(1) (2)
BW v1 v2 v3	*Os03g37840*	OsHAK16 (potassium transporter)	(1) (2)
BY v1 **v2**	*Os02g43760*	ubiquitin carboxyl-terminal hydrolase	(1) (2)
CE v1 v2 **v3**	*Os05g51690*	NRR/CRCT (CCT motif family protein)	this study
CF v1 v2	*Os04g18830*	MATH domain containing protein	this study
CF v1 v2	*Os10g31640*	glycine-rich cell wall structural protein 2	this study
CG v1 **v2**	*Os12g42200*	cation-hydrogen exchanger	(3) and this study
CK v1 **v2** v3	*Os04g48160*	IQ calmodulin-binding motif protein	this study
CK v1 v2 v3 **CP** v1	*Os06g36270*	receptor-like kinase 5	(3) and this study
CL v1 v2 v3	*Os01g40290*	expressed protein	(3) and this study
CL v1 v2 v3	*Os05g42150*	OsGH3.4 (indole-3-acetic acid-amido synthetase)	(3) and this study
CL v1 v2 v3	*Os10g39840*	OsXTH20 (glycosyl hydrolase family 16)	(3) and this study
CP v1	*Os01g53920*	receptor-like kinase 5	this study
CP v1	*Os03g42130*	gibberellin 20 oxidase 2	this study
CP v1	*Os07g33740*	vrga1 resistance protein	this study
CQ v1	*Os08g14940*	LRR-XII receptor kinase	this study
CR v1	*Os07g23470*	thaumatin like protein	this study
CS v1	*Os12g13810*	OsNOMT (naringenin methyltransferase)	this study

[1] The TALE versions shown to activate the respective target gene (via published expression data or via GUS analysis) are shown in green, those coloured in orange activated the target gene only weakly, those in red were unable to activate the target gene and those in black were not yet analysed in any detail.

[2] Previously published target genes are shows in black while those that were previously predicted but which were confirmed for the first time in the present study are shown in blue. Shown in red are the partially confirmed new target genes that were identified solely based on the comparative mismatch analysis. Their promoters were specifically induced by one or several *Xoc* TALEs but they were not yet analysed for a TALE-dependent induction in rice.

(1) Cernadas *et al*., 2014; (2) Wilkins *et al*., 2015; (3) Erkes *et al*., 2017; (4) Mücke *et al*., 2019

4 Discussion

4.1 Aberrant repeat-carrying TALEs from *X. oryzae* – functions and target genes

When the present work was started, little was known about the specific functions of aberrant repeat carrying TALEs or their target genes in general. Most prominent among the few known examples were the two *Xoo* TALE classes TalAC (AvrXa7) and TalBH (PthXo3), both containing the 40N* aberrant repeat variant. Both TALE classes were shown to target the major susceptibility gene *OsSWEET14* (Yang and White, 2004; Antony *et al.*, 2010). Only two additional aberrant repeat-carrying TALE classes from *Xoc* had a confirmed target gene and for both phenotypic effects were not described. The first is class TalBG, whose members induce the expression of a 16k transmembrane protein (Cernadas *et al*., 2014; Wilkins *et al*., 2015) that shows similarities to the anoctamin family of Ca^{2+} activated ion channel proteins. The second class is TalBV, whose members were shown to target two genes, a GATA ZF TF and a cyclin (Cernadas *et al*., 2014; Wilkins *et al*., 2015).

Recently, our group was able to add another example to this list by showing that the the members of class TalAQ from *Xoo* induce *OsDOX-1* (Mücke *et al*., 2019). The present work contributed to the extension of this list, demonstrating that three other aberrant repeat-carrying TALEs address clade III *OsSWEETs*. By showing that not all naturally occurring aberrant repeats facilitate a loop out, the present study also highlighted the necessity to take into account the exact type of aberrant repeat found within a TALE when looking for potential target gene candidates. Consequently, this knowledge should increase the reliability of target gene predictions for aberrant repeat-carrying TALEs.

4.1.1 Class TalBK members with the novel 36aa v1 aberrant repeat address the *OsSWEET13* allele from Nipponbare

TalBK2 was the first naturally occurring TALE found to carry two aberrant repeats and it addresses the *OsSWEET13* promoter allele from Nipponbare by looping out one of its two aberrant repeats. The promoter of the *OsSWEET13* gene was known to have three different allelic variants, two of which (Zhenshan and IR24) are targeted by TALEs from class TalAM (PthXo2-like) (Zhou *et al*., 2015). By demonstrating that TalBK2 targets the allelic variant from cv. Nipponbare, the present work gave further support to the hypothesis that the on-going arms-race between rice and *Xanthomonas* yields constantly new promoter variants that prevent activation by the pathogen as well as new TALE derivatives and mechanisms to keep target gene induction. The novel allelic variant of the *OsSWEET13* promoter found in the rice cv. Sadu Cho highlights this point even further, especially since neither TalAM nor TalBK were able to bind it. Incidentally, in a recent study Zaka and colleagues screened for

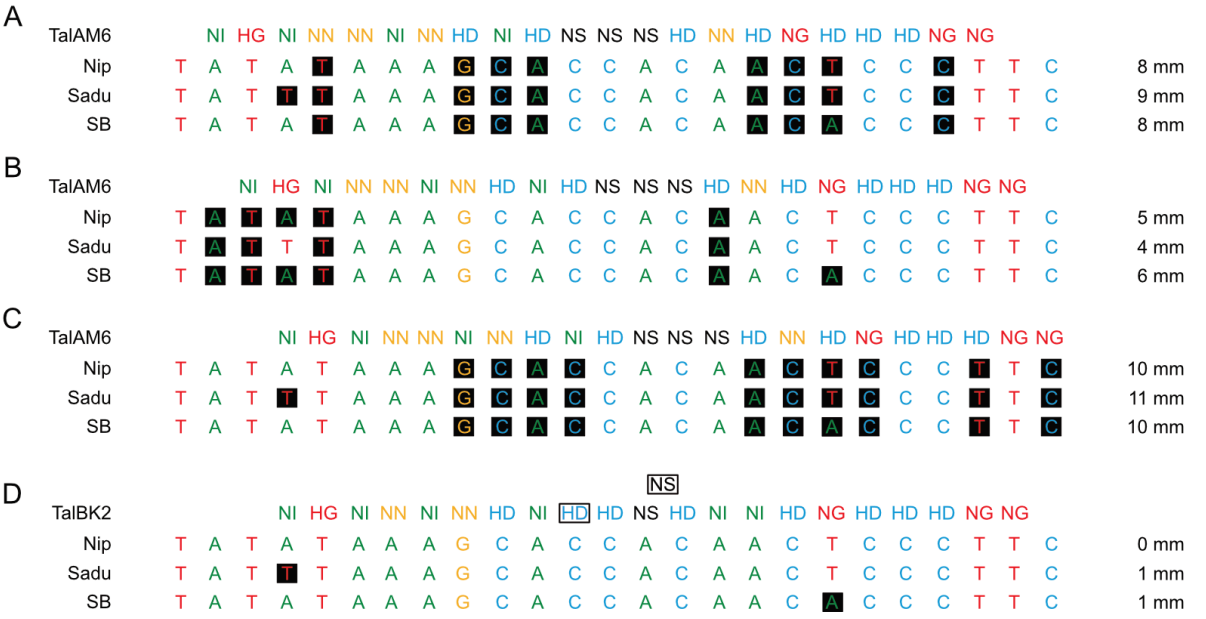


Figure 4.1: Binding possibilities for TalAM6 and TalBK2 in the *OsSWEET13* promoter variants from Nipponbare (Nip), Sadu Cho (Sadu) and Super Basmati (SB). (A) (B) and **(C)** show the three best binding possibilities for TalAM6 in those three promoter variants. **(D)** Best binding possibility for TalBK2 in those three promoter variants. Mismatched positions are boxed black and their number is indicated for each each (mm).

allelic variants of all known TALE susceptibility targets in rice and they also identified a new allele of the *OsSWEET13* promoter in the rice cv. Super Basmati (SB; Zaka *et al*., 2018). The authors state that TalAM is unable to activate this new promoter variant but since this new allele differs from the not-induced variant from cv. Nipponbare only by a single position, this is unsurprising (Fig. 4.1). However, it will be interesting to see if TalBK is able to address the *OsSWEET13* promoter from cv. SB since its putative binding site there contains only a single mismatch (Fig. 4.1). It is true that the artTalBK2 failed to address the *OsSWEET13* promoter fragment from Sadu Cho, an allelic variant that carries also just a single mismatch. Yet, this mismatch is placed at position 1 of the target site, and thus in an extremely sensitive region. In contrast, the mismatch in the promoter from the cv. SB is placed near the very end of the TALE binding site and therefore less likely to drastically reduce the TALEs activity (Fig. 4.1). Until this is evaluated experimentally, the allelic variant from Sadu Cho remains the only *OsSWEET13* promoter allele that is not induced by a natural TALE from *Xanthomonas*.

4.1.2 TALEs with the 40aa N* repeat and the induction of *OsSWEET14*

So far, only members from two TALE classes that carry the 40N* aberrant repeat were analysed in detail and both, TalAC and TalBH, were demonstrated to activate the major susceptibility target *OsSWEET14* (Yang and White, 2004; Antony *et al*., 2010). However, with TalDS, TalDV, TalEQ and Tal2 from K-74, four additional *Xoo* TALEs/TALE classes

exist in which the same aberrant repeat variant can be found and, interestingly, they all share a striking similarity in their RVD compositions with either TalAC or TalBH (Yu *et al*., 2015; Quibod *et al*., 2016). Since it is likely that the 40N* repeat emerged only once in evolution, it seems reasonable to assume that all TALEs carrying this variant are related to each other, and thus might share the same target gene. By analysing artificially generated versions of some of these TALEs, the present study was able to demonstrate that TalDV and Tal2 from K-74 do also target *OsSWEET14* while TalDS and a shortened variant of TalAC occurring in KACC10331 are unable to do so.

The observation that the TALEs that carry the 40N* repeat show such a strong variation in their RVD sequence and repeat number might be the result of an evolutionary-driven adaptation process e.g. to changing promoter variants. However, only two other *OsSWEET14* promoter alleles are known. The first, *xa41(t)*, is reported to confer resistance against TalAC2 and Tal5 but not against TalC (Hutin *et al*., 2015b). It also seems to be inaccessibly for most or all of the 40N* carrying TALEs. The second variant was discovered more recently. It was derived from the rice cultivar Ejali. This variant is also reported to confer resistance against the major virulence factor TalAC2 (Zaka *et al*., 2018). However, compared to the most common variant of the *OsSWEET14* promoter, it varies just at a single nucleotide position placed approximately in the middle of the TALE binding site, changing a C to a G (Fig. 4.2). It is true that this introduces an additional mismatch into the putative binding site of TalAC2 and many of the other 40N* carrying TALE classes. However, a single mismatch, especially one placed in such a central position, seems unlikely to impair the activity of such long TALEs. A study with artificial TALEs could show that longer TALEs are less affected by mismatches (Mali *et al*., 2013) while some of the longest naturally occurring TALEs, the members of class TalCL from *Xoc*, induce rice genes despite having as many as 8 mismatches in their binding sites (see chapter 3.2.4.1). Nonetheless, the existence of these two less common promoter variants could at least partially explain the observed variation in the 40N* carrying TALEs. Aberrant repeat carrying TALEs might also be much more tolerant towards alterations in their repeat number than normal TALEs since the flexibility given by an aberrant repeat's ability to loop out when required should help in compensating for duplicated or lost repeats, at least to some extent. Since such mutations would render an aberrant repeat-carrying TALE not automatically unable to activate its virulence target, they could be

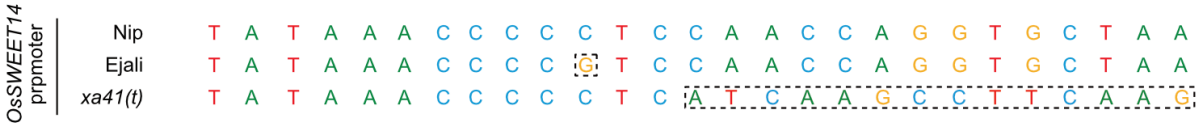


Figure 4.2: The new allele of the *OsSWEET14* promoter from the rice cv. Ejali. The most frequently found variant of the *OsSWEET14* promoter is represented by the sequence from the cv. Nipponbare. The displayed region of the promoter is targeted by several TALEs, among them class TalAM (AvrXa7) and class TalBH (PthXo3).

more often kept than in a TALE without aberrant repeat. It might also be possible that the differences between the 40N* repeat-carrying TALEs are the result of an evolutionary pressure selecting against the on-going activation of an unknown target gene with negative effects on *Xoo*. Interestingly, many of the sequenced *Xoo* strains carry two different *OsSWEET*-activating TALEs. One could therefore speculate that this serves to ensure virulence on different rice varieties or to have an alternative if the induction of one of these two major virulence targets becomes undesirable e.g. due to the emergence of a new resistance gene with a similar binding site in its promoter.

Interestingly, there are two examples in which a *Xoo* strain does not induce an *OsSWEET* gene but remains highly virulent. The first is the *Xoo* strain PXO211. This strain is reported to be fully virulent on several different rice cultivars (Quibod *et al*., 2016), yet, it contains no *OsSWEET*-inducing TALE at all. It does contain TalDS, this TALE, however, seems unable to address any *OsSWEET14* promoter alleles. The second example is from a more recent study that sequenced two *Xoo* strains from India (Carpenter *et al*., 2018). The authors found that one of their strains, IX-280, induces *OsSWEET11* in IR24 but not in IRBB13. Remarkably, neither was any of the other clade III *OsSWEETs* induced in IRBB13 but still the strain showed a very strong virulence phenotype on both rice lines (Carpenter *et al*., 2018). This observation led the authors to speculate that recombination events resulted in a TALE that targets a new and yet unknown virulence target that is as potent as the *OsSWEETs*. Given these results, it seems therefore possible that some 40N* repeat carrying TALEs have acquired a completely new potent virulence target.

4.1.3 TALEs with the 28aa repeat – potent suppressors of a plant resistance mechanisms aimed against TALEs

Two recent studies discovered an unexpected function for the 28aa aberrant repeat-carrying TALEs. These TALEs all share similar N- and C-terminal truncations that render them unable to activate gene expression, and thus they were previously believed to be non-functional pseudogenes. Yet, these specific characteristics allow them to suppress a potent resistance mechanism that recognizes TALEs in general (Read *et al*., 2016). Accordingly, they were named truncated (truncTALEs; Read *et al*., 2016) or interfering TALEs (iTALEs; Ji *et al*., 2016). This resistance mechanism was demonstrated to work completely independent from the RVD sequence of the TALE or its capability to activate gene expression, and thus it was proposed that it is based on an interaction between these TALEs and a specific R-protein that directly recognizes TALEs (Ji *et al*., 2016; Read *at al*., 2016). Since the two responsible rice *R*-genes, named *Xa1* and *Xo1*, are not yet cloned, the exact mechanisms remain unknown. However, it is assumed that the trunc/iTALEs bind and keep being bound to these R-proteins without triggering their resistance effects, thus blocking them from recognizing

any other TALEs (Ji *et al*., 2016; Read *at al*., 2016). This proposed mechanism shows remarkable similarities to the model of the Bs4-mediated recognition of AvrBs4 in tomato, however, since Bs4 failed to recognize all TALEs its mechanism seems to be different (Schornack *et al*., 2004).

It is noteworthy that not all of the identified trunc/iTALEs were found to carry an aberrant repeat (Ji *et al*., 2016; Read *et al*., 2016) and that experiments with artificial TALE derivatives demonstrated that the presence of the 28aa repeat is not required for the resistance suppressing function of the truncTALEs (Read *et al*., 2016). In DNA binding experiments with synthetic TALEs containing the 28aa repeat, it was also observed that the 28aa aberrant repeat confers an obligatory frameshift recognition (Read *et al*., 2016). Curiously, this observation is in contrast to the results shown in chapter 3.1. In the present work, the 28aa aberrant repeat displayed a strong preference for a -1 frameshift box but it was nonetheless able to address the optimal box. However, with the 37aa repeat v2, an aberrant repeat was identified that is indeed unable to participate in DNA binding, presumably doing so by constantly looping out.

4.2 The near-exclusive occurrence of aberrant repeats Asian *X. oryzae*

With the exception of the two examples found in the *X. axonopodis* pv. *citri* strain 29-1, all TALE repeats of aberrant length were identified in *X. oryzae* strains from Asia. It seems tempting to account this imbalanced occurrence to the fact that many *X. oryzae* strains, especially those from Asia, carry an unusually large number of TALEs and therefore simply have a higher chance of containing an aberrant repeat. In addition, among the completely sequenced *Xanthomonas* spp. genomes, those from rice-infecting *X. oryzae* strains are strongly overrepresented, and thus the probability to find an aberrant repeat in *X. oryzae* strains is also increased due to the larger sample size. However, African *Xoc* strains carry between 20-23 TALEs and therefore nearly as many as those isolated in Asia, yet, not a single aberrant repeat was found within the three sequenced *Xoc* strains from Africa. At the same time, all *X. oryzae* strains from Asia that have been sequenced so far, *Xoo* and *Xoc* alike, contain at least one aberrant repeat-carrying TALE, thus suggesting that aberrant repeats might even be ubiquitous in Asian *X. oryzae* strains. While it is reasonable to assume that a greater number of TALEs increases the probability of an aberrant repeat to emerge in course of evolution and that a higher number of available TALE sequences increases the chances to identify one, these factors alone are not sufficient to explain the observed pattern.

Interestingly, a recent publication provided a new viewpoint that could help to understand this phenomenon. The authors of this study hypothesized that the difference in the diversity and number of TALEs between Asian and African *X. oryzae* strains might be a direct

consequence of the differences in cultivation, breeding and farming of rice in these two regions (Schandry *et al*., 2018). While Asia has a long history of intense rice cultivation, involving a frequent switching between steadily improved rice cultivars and hybrid lines containing a wide range of different resistance genes, its farming history in Africa is comparatively short (Schandry *et al*., 2018). The authors reason that the *X. oryzae* strains in Asia were therefore forced to rapidly adapt to host plants with frequently changing characteristics, thus speeding up the evolution of their TALE repertoires (Schandry *et al*., 2018). This argument is supported by a study that found that the planting of resistant rice lines exerts a selective pressure on a *Xoo* population, ultimately leading to the emergence of strains that are able to overcome the resistance (Vera Cruz *et al*., 2000). This result is also in accordance with the observation that the recombination frequency in many bacteria increases under stress (Bjedov *et al*., 2003). Additional support for this argument comes from the observation that most of the so far known TALE-based mechanisms to counter plant resistances were also identified in *X. oryzae* strains originating from Asia (Schandry *et al*., 2018). However, for the numerous TALEs in the *Xoc* strains from Africa this explanation cannot apply but the authors hypothesized they might have evolved more recently and from an Asian ancestor with an already expanded TALE repertoire (Schandry *et al*., 2018).

Their hypothesis might nonetheless be suited to explain, at least in part, why aberrant repeats are so frequently found in Asian but not at all in African *X. oryzae* strains. The increased selection pressure for the pathogen caused by the intense rice cultivation in Asia could have sped up the processes that lead to the emergence of an aberrant repeat even more. TALEs are already well suited for fast, stress-driven adaptations since even a single mutation in one RVD can change the specificity of the protein. Moreover, *TALE* genes and genomic TALE clusters are often flanked by inverted repeats that are believed to confer genetic mobility e.g. as part of a Tn3-like transposon (Ferreira *et al*., 2015; Erkes *et al*., 2017). These repeats are presumably the cause for the frequently observed inversion and recombination events that affect *TALE* genes or even complete TALE clusters (Erkes et al., 2017). Additionally, *TALE* genes are extremely repetitive and thus, they are much more likely to cause problems during DNA replication than non-repetitive sequences. This often results in strand slipping and recombination events, causing the duplication and deletion of repeat units or the fusion TALEs (Erkes et al., 2017; Schandry et al., 2016). It is also assumed that strand slipping events are the cause for the emergence of (small) repetitive sequences in general, and thus such events seem also the most likely explanation for the duplications that lead to the emergence of aberrant repeats in TALEs.

Interestingly, several aberrant repeat-carrying TALE classes have members that contain an aberrant repeat as well as members that do not carry an aberrant repeat but just a normal 34aa repeat of standard length at the same position. This suggests that several of theses

aberrant repeats emerged rather recently. This is the case for TalBV, TalAQ and TalAS (Appendix Table 3.1.1 and 3.1.2).

4.3 Naturally occurring length polymorphisms in repetitive proteins

The discovery that some rarely occurring repeat variants of aberrant length affect the otherwise so very regular TALE-DNA interaction in such a drastic manner prompted the question if similar alterations can be found in other repetitive proteins as well and if this is the case, how they affect protein function. TALEs match many of the characteristics typically associated with the superfamily of alpha solenoid proteins (Mak *et al*., 2012; Deng *et al*., 2014). An alpha solenoid repeat consists of two antiparallel helices that are interconnected either by a loop or by a third helix and in their entirety alpha solenoid proteins often take on the shape of a superhelix (Fournier *et al*., 2013). Repeat types known to form alpha solenoid structures are extremely diverse and many of them show a much higher degree of degeneration than TALE repeats. As a result, individual repeats from the same protein and from the same repeat type can vary substantially in length or amino acid sequence, and thus even defining a repeat of aberrant length can be difficult in those cases. Examples for this are HEAT repeats, which vary between 30aa and 40aa (Yoshimura and Hirano, 2016), or leucine-rich repeats (LRR), which were found to vary between 22aa and 28aa (Bella *et al*., 2008). Nonetheless, sometimes individual repeat units are clearly separated by the insertion of additional amino acids that do not belong to the repeat consensus or shortened by one or two amino acids, examples for this can be found in ankyrin repeats (Mosavi *et al*., 2004; Kozlov *et al*., 2018) as well as in tetratricopeptide repeats (TPRs) (Das *et al*., 1998). Yet, since these repeat types are known to facilitate protein-protein interactions, the resulting repeat regions function as one unit, and thus it is difficult to predict or evaluate the effects of individual repeats.

However, there are two examples found among the alpha solenoid repeats that are remarkably similar to TALEs: pumilio (PUFs) and pentatricopeptide repeats (PPRs) (Fig. 4.3). Like TALEs, both facilitate the binding to nucleotides by connecting in a one-repeat-to-one-nucleotide manner to their targets (Miranda *et al*., 2018) but in contrast to TALEs, PUF and PPR repeats facilitate binding towards single stranded RNA (Filipovska and Rackham, 2012). Moreover, their nucleotide specificity is not conferred by a single amino acid but by at least three (Fig. 4.3; Wang *et al*., 2013; Miranda *et al*., 2018), thus rendering predictions towards their specificity somewhat difficult. Nonetheless, studies could demonstrate that it is possible to alter their sequence specificity (Hall, 2016; Miranda *et al*., 2018). PUF repeats are, however, not freely combinable, and thus their use is limited to alterations of the natural

Figure 4.3: Schematic comparison of TALE, PUF und PPR proteins and their repeats. (Left) Schematic representation of typical TALE, PUF and PPR proteins. The repeats and other crucial motifs are indicated. NCR = non canonical repeats; T3SS-S = signal for T3SS-dependent translocation; NLS = nuclear localization signal; AAD = acidic activation domain; MTS = mitochondrial targeting sequence. **(Right)** Schematic representation showing the secondary structure of a typical repeat unit from a TALE, PUF and PPR protein. Amino acids involved in the specific recognition of nucleotides are indicated by black dots. The figure is based on Filipovska and Rackham, 2012.

occurring proteins, allowing them to bind only RNA sequences of either 4, 8-9 or 13 nucleotides (Wang *et al*., 2018). Each PUF repeat also consists of three helices, two larger ones in antiparallel orientation interconnected by the third (Fig. 4.3; Wang *et al*., 2018). PPRs on the other hand were found to contain up to 29 repeats (Barkan and Small, 2014) and each PPR repeat contains two helices that are connected by a loop (Fig. 4.3; Miranda *et al*., 2018). Considering the similarities between these three repeat types, it seems imaginable that PUFs and PPRs would be as drastically affected by an aberrant as TALEs are.

PUF repeats contain usually 36 amino acids but examples with prolonged repeat units exist (White *et al*., 2001). These additional amino acids, however, are usually situated at the end of a repeat. By taking into account the length and composition of the published RNA targets, it can be concluded that these prolonged repeats seem to have no effect on the binding behaviour of their proteins (Wang *et al*., 2018). This would also be in accordance with the results from the present study that showed that artificially prolonged TALE repeats behave more or less normally if it the additional amino acids are placed near the very end or between two repeats. However, there are examples in which a PUF repeat alters the binding properties of the protein. Most PUF proteins recognize a RNA target of 8 nucleotides but at least two examples were found to facilitate the binding of 9 nucleotides and doing so despite the presence of only 8 repeat units. These are Puf4 from *S. cerevisiae* (Gerber *et al*., 2004; Valley *et al*., 2012) and FBF from *C. elegans* (Bernstein *et al*., 2005; Opperman *et al*., 2005). In both proteins, additional 16 amino acids are found between two repeats and extensive analysis with the FBF protein could show that this insertion is - while not sufficient on its own - of crucial importance for the observed change in the target length (Opperman *et al*., 2005; Wang *et al*., 2009). It was also demonstrated that the additional nucleotide in the target site corresponds to the position of the prolonged repeat (Opperman *et al*., 2005; Wang *et al*.,

2009; Lu and Hall, 2011). Curiously, this behaviour is contrary to the effects caused by an aberrant repeat in a TALE. Instead of allowing for the flexible loop out of the aberrant repeat, and thus a tolerance for a target site which lacks one nucleotide (as seen in TALEs), PUF repeats of aberrant length allow for an additional nucleotide. The prevalent opinion as to why and how this is happening assumes that the additional nucleotide in the RNA-target is flipped out (Opperman *et al*., 2005; Valley *et al*., 2012). This seems a valid possibility since a single stranded RNA target could have indeed a greater flexibility to do so than a double stranded DNA molecule. However, the additional amino acids could also function as spacer, and thus enforce a certain distance between the two affected repeats. The fact that not all possible nucleotides are accepted at this skipped position, with guanine being not tolerated, seems to further support this idea since it indicates that the additional extracyclic amino group of the guanine is the cause (Opperman *et al*., 2005).

PPR repeats on the other hand occur in at least three different repeat types that can be distinguished by amino acid sequence pattern and length, with the three PPR repeat motifs being P, L and S, with 35aa, 35-36aa and 31-32aa, respectively (Barkan and Small, 2014; Manna, 2015; Cheng *et al*., 2016). While they all show an identical binding mechanism, with the same amino acid positions being responsible for the connection to the RNA, they seem to display differences in their nucleotide specifies that depend on the PPR repeat type they are situated in (Barkan and Small, 2014). Furthermore, there are at least two examples reported in which a PPR repeat contains an insertion. The two proteins Rfo and PPR-A both contain the same 11 additional amino acids placed at the very end of repeat 13 (Qin *et al*., 2014). However, these proteins are reported to confer binding to a RNA sequence whose length corresponds exactly to the number of PPR repeats, thus indicating that the additional amino acids have no unusual effect on the function of the proteins. Yet, given the striking resemblance between TALE and PPR repeats, it is very well possible that a more profound alteration (e.g. a longer insertion or one closer to a crucial residue) could also have a more drastic effect. However, whether such a repeat would indeed cause a facultative or obligatory recognition of a frameshift target or whether it would instead interfere with the normal structure and impair protein activity remains a subject of speculation until a natural example is found or artificial variants are generated and analysed.

4.4 Artificially shortened TALE repeats and the recognition of methylated cytosine derivatives

5-methyl cytosine (5mC) is an epigenetic modification of cytosine that is found in most eukaryotes and that is well known to negatively affect the transcriptional activity of genes (Law and Jacobsen, 2010). Interestingly, the TALE RVD HD, which is most commonly used to recognize cytosine, is incapable to address 5mC, presumably as a result of sterical

clashes caused by the presence of an additional side chain in the methylated cytosine derivative (Deng *et al*., 2012b). However, RVDs with smaller side chains (e.g. NG) or those lacking the amino acid at position 13 completely (e.g. N*) were found to recognize 5mC (Deng *et al*., 2012b; Valton *et al*., 2012). More recent studies aimed to expand this knowledge by investigating whether there are also TALE repeats able to recognize 4-methyl cytosine (4mC) or the oxidized cytosine derivatives 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). In the course of these studies, artificial aberrant repeats with size reduced loop regions were generated by deleting the amino acids at position 13 and 14 (e.g. H**, N**, Q**; Maurer *et al*., 2016; Rathi *et al*., 2016 and 2018). While these shortened repeats seem to tolerate cytosine as good as its derivatives, they yielded TALEs with a significantly reduced overall activity (Maurer *et al*., 2016; Rathi *et al*., 2016). While not directly comparable, it is worth mentioning once more that the insertion of amino acids within or close to the loop region of a TALE repeat had similar but even more drastic negative effects on the functionality TALEs, often abolishing TALE activity completely (chapters 3.1.12 and 3.1.13). Taken together, these results give further support that the loop region of TALE repeats plays a crucial role in the overall functionality and activity of a TALE.

4.5 Crucial repeat residues and how they are affected in repeats of aberrant length

The present study demonstrated that not all aberrant repeats behave identical. Instead, they can be grouped into four functional categories, namely standard DNA binding, facultative frameshift binding, obligatory frameshift binding and disruption of TALE function. Into which category an aberrant repeat fits, seems largely based on where the alteration within the repeat is located and which specific amino acids are affected. Taking into account which structural motifs and important residues of a TALE repeat are altered in an aberrant repeat, could help to understand this phenomenon in more detail. Based on the crystal structures of TALE proteins, it is possible to separate a TALE repeat into smaller motifs and to define the most crucial amino acids. Important are the residues from position 12 to 17, which are responsible for the stabilization of the RVD loop (position 12), the base specificity of the repeat (position 13) and the connection to the DNA phosphate backbone either via water-mediated (positions 14-15) or via direct and water-mediated H-bonds (positions 16-17) (Fig. 4.4; Deng *et al*., 2012a and 2014; Mak *et al*., 2012). Crucial is also the proline at position 27 of most TALE repeats since it generates the kink that allows the sequential stacking of the repeats, thus preventing sterical hindrances (Fig. 4.4; Deng *et al*., 2012a and 2014; Mak *et al*., 2012). In addition to those, each TALE repeat has also ten mostly hydrophobic residues that form the inter- and intra-repeat connections, and thus keep the TALE 3D structure in place. They are mainly found within a distance of 5–10aa around the RVD loop (Fig. 4.4;

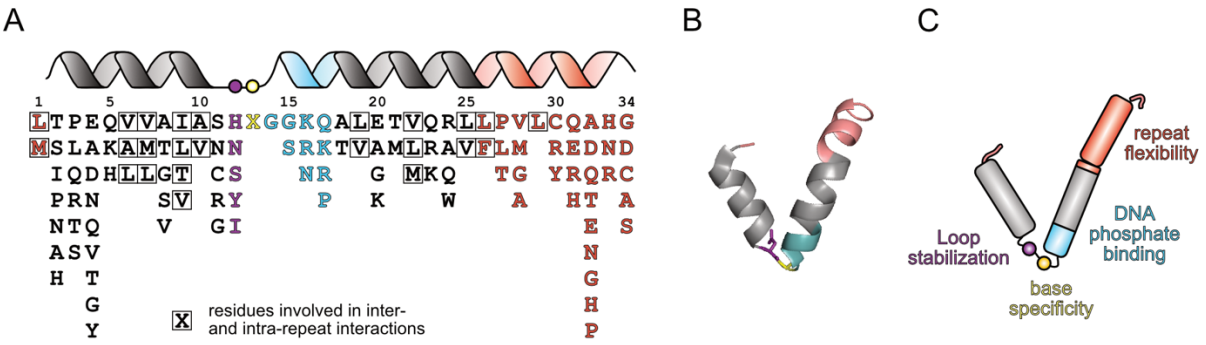


Figure 4.4: Annotated motifs and crucial amino acids within a TALE repeat. (**A**) Amino acid abundance found within >2000 TALE repeats as depicted by Boch and Bonas, 2010. The amino acids are sorted according to their frequency of occurrence. Aligned above is the secondary structure of a TALE repeat. The amino acid that actually governs base specificity is coloured in yellow (position 13), the amino acid responsible for stabilizing the loop structure is coloured in purple (position 12). The four amino acids that mediate interactions with the negatively charged phosphate backbone of the DNA are highlighted in blue (positions 14-17) while those amino acids that are involved in intra- and inter-repeat interactions are boxed. The very first amino acid and those within the second half of the second helix belong to the region of repeat flexibility and are coloured in red (positions 1 and 26-34). (**B**) 3D structure and (**C**) schematic view of a TALE repeat subdivided into its different functional parts. The functional motifs and crucial residues are depicted according to Deng *et al*., 2014.

boxed residues). Furthermore, the last 10aa of each repeat were defined as the region of repeat flexibility. This region contains only very few of the hydrophobic residues and is crucial for the structural plasticity of the repeats and the TALE protein as a whole (Fig. 4.4; Deng *et al*., 2014). The large number of motifs and crucial residues within TALE repeats raises the question how they are affected in repeats of aberrant length and if it is possible to explain the observed differences in their ability to facilitate frameshift recognition with changes in these motifs and residues.

Taking into account the different motifs and crucial residues of a TALE repeat while re-analysing the experimental data, allows for several conclusions. The first is that the position of a repeat alteration strongly affects its impact on repeat function. For example, all repeat variants that rendered their TALEs inactive have in common that their alteration is placed within the RVD loop or the adjacent amino acids (Fig. 4.5; no dot). Considering the extremely regular structure of TALEs (Deng *et al*., 2012a; Mak *et al*., 2012) one can assume that the stacking of their repeats must be a delicate process, and thus it is understandable that even a single, sufficiently misfolded or misplaced repeat can be enough to severely disturb the function of the protein. This observation highlights once more the crucial importance of the RVD loop, indicating that it is not only essential in connecting the repeats to the nucleotides of the TALE target sequence (Deng *et al*., 2012a; Mak *et al*., 2012), and thus governing the specificity of the TALE (Boch *et al*., 2009; Moscou and Bogdanove, 2009), it also helps facilitating and stabilizing the general structure of a TALE repeat. One possible explanation for the lack of TALE activity in these repeats could be that insertions in or around the RVD loop might disturb the positioning and the necessary proximity of the two repeat helices and

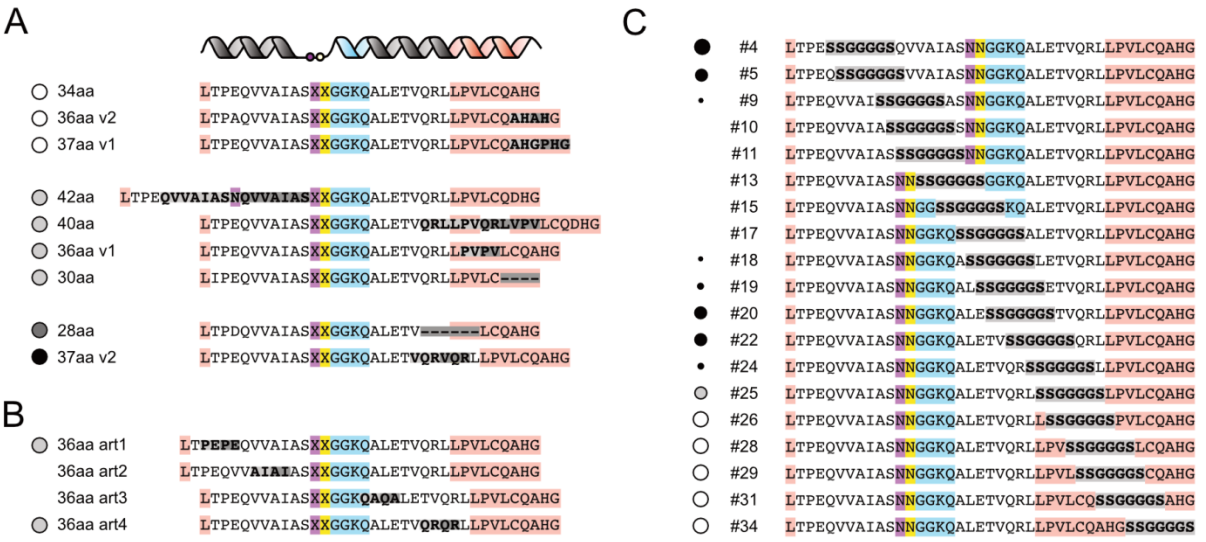


Figure 4.5: Functionally critical residues in a TALE repeat and how they are affected in aberrant repeats. (A) Naturally occurring aberrant repeats. **(B)** Artificially generated aberrant repeat variants with duplications. **(C)** Artificially generated aberrant repeat variants with insertions. The dots in front of every repeat variant correspond to the GUS assay results. The colour represents the preferred binding mode. White = only standard DNA-binding. Light grey = facultative frameshift binding. Dark grey = strong tendency towards frameshift binding. Black = obligatory frameshift binding. Dot size roughly corresponds to the observed activity. No dot means no activity.

thereby block the establishment of proper inter- and intra-repeat interactions. The position of the repeat alteration is also crucial in those aberrant repeats that have no effect on the DNA binding behaviour of a TALE. All of them share that they contain several additional amino acids (duplications or linker insertions) within the last few amino acids of the repeat (Fig. 4.5; white dot). In this region, linker insertions were tolerated particularly well since they did not even convey frameshift binding when they were placed close to the beginning of the region of repeat flexibility, thus designating this region as well-suited for internal insertions of foreign functional domains. However, the duplication of amino acids in the beginning of the region of repeat flexibility did lead to frameshift binding in two cases (Fig. 4.5; 40aa and 36aa v1), thus proving again that the exact amino acid sequence of an insertion is important. Having a closer look at the nature of the duplicated amino acids in those two repeats does offer one possible explanation for this observation. Both repeats share a duplication of the kink-inducing proline, and thus they should contain a second kink in their second helix that is likely to disrupt the sequential stacking of the TALE repeats. In order to avoid such complications when inserting additional amino acids into a TALE repeat, it is therefore recommended to place the desired insertion at the very end of a repeat or, even better, between two repeats and to include a small linker sequence before and after the desired functional domain.

A second conclusion that can be drawn is that it does make a difference if a repeat-specific amino acid sequence is duplicated or if a stretch of TALE-unrelated amino acids is inserted.

Amino acid duplications resulted frequently in aberrant repeats that confer facultative frameshift binding (Fig. 4.5 A and B; grey dots) while linker insertions at similar positions more often resulted in aberrant repeats that looped out constitutively (Fig. 4.5 C; black dots) or had no effect at all (Fig. 4.5; white dots; region of repeat flexibility). One possibly explanation for this could be that a duplication is more likely to yield amino acids that are able to interact with the native amino acids of the TALE repeats than a duplication. Intriguingly, among all the artificial aberrant repeats that were generated by linker insertions, only one was found to confer facultative frameshift binding (Fig. 4.5 C, #25). Curiously, this one variant has the insertion placed exactly at the border that connects the repeat region of flexibility with the region containing most of the hydrophobic residues. This, once more, suggests that the region of repeat flexibility is not an arbitrary motif but an integral part of a TALE repeat. However, whether the region of repeat flexibility requires a specific sequence or whether any sufficiently-flexible, non-intrusive sequence with a similar length could fulfil the same function remains the subject of future studies.

From the obtained data it is also possible to assume that it does make a difference if the amino acids at a particular position are deleted (Fig. 4.5 A; 30aa and 28aa) or if they are disrupted by an insertion (Fig. 4.5 A; 36aa v2 and 37aa v1 or 40aa and 36aa art4, respectively). In several cases, deletions had a more disruptive effect on the TALE. One could therefore assume that deletions are more likely to have severe effects than amino acid duplications or insertions because they do not only disturb the surrounding residues and bring them out of place, they also remove several, potentially crucial amino acids completely. However, since the 37aa v2 repeat is actually more prone to an obligatory frameshift binding than even the 28aa repeat, it can be concluded that amino acid insertions can be as disruptive as deletions, and thus evaluating each repeat individually remains important (Fig. 4.5 A, 28aa and 37aa v2). Interestingly, the binding behaviour of the 37aa v2 repeat fits better to those of the repeats with arbitrary insertions (compare Fig. 4.5 B, 37aa v2 and C, #22 - #24), suggesting that the amino acid sequence resulting from this specific duplication might just be particularly unfavourable for the repeat structure.

4.6 A model for the loop out mechanism of different aberrant repeats

As described in the previous chapter, aberrant repeats can be grouped into four different functional categories. However, even within these categories functional differences exist. This becomes obvious if, for example, the 42aa and the 40aa aberrant repeats are compared. Both are the result of a duplication, but while the 42aa aberrant repeat has eight additional amino acids in the N-terminal helix, the 40aa aberrant repeat has six additional amino acids in its C-terminal helix. Interestingly, it seems irrelevant in which of the two helices the amino acid alteration is placed since both confer a facultative frameshift binding

and behave more or less identical, at least if only one of them is placed in a TALE (Fig. 4.6 A). However, as soon as two aberrant repeats were placed in tandem, the TALEs recognized different target boxes, depending on which of the two aberrant repeat versions was used. Both TALEs were able to recognize a -1 and a -2 target box while they failed to activate the optimal target. This indicates that two 42aa as well as two 40aa aberrant repeats in tandem are unable to participate simultaneously in DNA-binding and that, preferentially, one or both of them loop out to confer the recognition of a -1 or -2 frameshift box, respectively (Fig. 4.6 B). Interestingly, the two TALEs did recognize different -1 boxes. The TALE with two 42aa aberrant repeats activated only the reporter with the -1 p1 box, the one where the nucleotide corresponding to the first aberrant repeat was missing. The TALE with two 40aa aberrant repeats on the other hand recognized the -1 p2 box, the box where the nucleotide

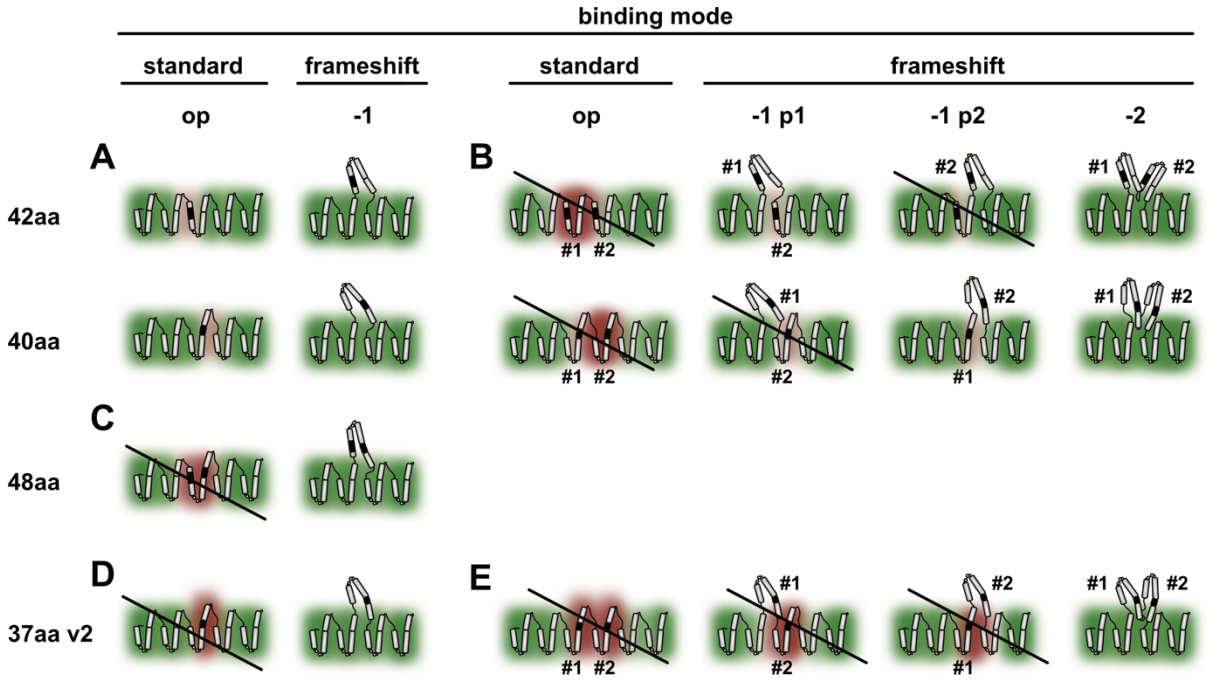


Figure 4.6: Model for the loop out behaviour of different aberrant repeats. (A) The two naturally occurring aberrant repeats with 42aa and 40aa confer a facultative frameshift binding. Both variants likely disrupt the inter- and intra-repeat interactions to the repeat closest to their altered helix (42aa, N-terminal helix; 40aa, C-terminal helix). **(B)** TALEs with the 42aa as well as with the 40aa aberrant repeat in tandem cannot recognize an optimal box but activate -1 and a -2 frameshift boxes by looping out one or both aberrant repeats. However, if two 42aa aberrant repeats are placed in tandem, the first one is likely to have more disturbances in its inter- and intra-repeat interactions, and so this one loops out. If two 40aa aberrant repeats are placed in tandem, the second one is likely to have more disturbances in its inter- and intra-repeat interactions, and thus the second aberrant repeat loops out. **(C)** The artificial repeat variant with 48aa confers an obligatory frameshift binding. Since this variant contains the duplications from both, the 42aa and the 40aa variant, it likely forms impaired inter- and intra-repeat interactions with both helices. This likely prevents it from properly connecting to the adjacent repeats on either side and causes its obligatory loop out behaviour. **(D)** The natural aberrant repeat variant 37aa v2 confers also an obligatory frameshift binding. This implies that its alteration is, despite affecting only one helix, similarly disruptive as the two alterations in the artificial 48aa variant. **(E)** A TALE with two 37aa v2 aberrant repeats placed in tandem induces only the -2 box, indicating that both aberrant repeats loop out constantly.

corresponding to the second aberrant repeat was missing. This observation correlates remarkably well with the position of the alteration in the respective aberrant repeat variant. The 42aa aberrant repeat is impaired in its first helix. Consequently, if two of these repeats are placed next to each other, the first one has disturbed inter- and intra-repeat connections on both sides, and thus it is more prone to being ejected. The 40aa aberrant repeat on the other hand has the alteration in the second helix, and thus, if placed in tandem, the second of the two repeats has disturbed interactions on both sides and gets ejected preferentially. This once more supports a model in which an aberrant repeat variant with disturbed connections on both sides has to leave the repeat array. This assumption is further supported by the observation that an artificial aberrant repeat with 48aa, which contains the duplications from both, the 42aa as well as the 40aa aberrant repeat, conferred an obligatory frameshift binding (Fig. 4.6 C). Interestingly, the 37aa v2 repeat also conferred an obligatory frameshift – despite being altered only in one helix. One must therefore assume that its alteration is more severe than those found in the 40aa and 42aa repeats since it is on its own sufficient to disrupt the inter- and intra-repeat interactions in a manner comparable to the artificial variant with two altered helices (Fig. 4.6 D). Placing this repeat variant in tandem resulted in a TALE that was able to recognize only a single target box, the one where both nucleotides corresponding to the aberrant repeats were missing (Fig. 4.6 E).

4.7 How aberrant repeats affect a TALEs search for and binding to DNA target sequences

TALEs exist in at least two different states: a relaxed DNA-free form that is also adopted while searching for potential target sites and a more condensed state when actually bound to a DNA target (Deng *et al*., 2012a; Cuculis *et al*., 2015 and 2016). In search mode, a TALE stays loosely wrapped around the DNA and moves along in a fast, one dimensional, non-rotational and unbiased fashion (Cuculis *et al*., 2015 and 2016). Once it encounters a stretch of matching nucleotides, it switches to a partially compressed state. This brings the repeats closer to sync with the DNA double helix and strengthens the connection to the DNA backbone, which enables the TALE to check more nucleotides of the putative target site (Cuculis *et al*., 2015 and 2016). If enough nucleotides match, the TALE enters the fully condensed binding mode and initiates gene expression, if not, the TALE relaxes again and resumes its search process (Cuculis *et al*., 2016). Curiously, it remains unclear, whether the initial recognition and the subsequent condensation are required to start from the N-terminus and the first repeats or whether it can happen from any position within the repeat region (Becker and Boch, 2016). The non-rotational search mode seems to suggest that the condensation can start from any point in the repeat region since the TALE would otherwise

be out of sync with its potential DNA target sites for most of the time, and thus drastically reduce its chances to successfully detect a target sequence. However, the N-terminus of a TALE is of crucial importance for its initial binding towards DNA (Cuculis *et al*., 2015 and 2016) and mismatches within the first repeats were shown to be much more severe in their effect on TALE activity than those within later repeats (Meckler *et al*., 2013), therefore indicating that the initial condensation has to start from the first repeats.

But how could the presence of an aberrant repeat with the ability to facilitate a flexible frameshift recognition affect this search and the subsequent binding process? If combined with an optimal target box, such an aberrant repeat can participate in DNA binding like any other repeat, thus indicating that it causes only minor inter- and intra-repeat disturbances. Upon encountering a matching frameshift box, however, the aberrant repeat can confer frameshift binding, presumably by leaving the repeat array. There are at least four possibilities to explain how this might happen. First, it could be that a population of TALE proteins is separated into two fixed folding states, one where the aberrant repeat is always within the DNA array and participating in DNA-binding and one where the aberrant repeat is always looped out, and thus conferring frameshift binding. Second, the aberrant repeat could be able to freely switch between being looped out and being properly integrated in the repeat array. This, however, would inevitably result in situations where the aberrant repeat is looped out during search mode and thus, the TALE would miss target sites. In a third alternative, the weak inter- and intra-repeat interactions of the aberrant repeat cause it to leave the repeat array as soon as it encounters a nucleotide. In this model, the collective DNA-binding strength of the TALE and its normal repeats would force the aberrant repeat back into the repeat array when an optimal target site requires its participation and only then. In the fourth and last alternative, the aberrant repeat remains in a normal conformation within the repeat array until it encounters a matching frameshift box. In this model, the impaired inter- and intra-repeat interactions of the aberrant repeat would result in it either leaving the repeat array by itself or it being pushed out from the repeat array by the next repeat in line while it establishes a connection to the nucleotide that the aberrant repeat failed to bind. However, until further experimental data is available, the exact mechanism remains the subject of speculation and so does which of these models, if any, is the most accurate.

Interestingly, one can assume that the search and loop out process between a TALE that contains only one and a TALE that contains multiple aberrant repeats in tandem does not differ significantly. The only real difference is likely that instead of a single aberrant repeat, multiple ones are ejected. At this point, it is important to remember that TALEs with multiple aberrant repeats were only able to recognize frameshift boxes that had also multiple nucleotides deleted. A TALE with two aberrant repeats for example was only functional if it was combined with a target box that had a deletion of 1 or 2 nucleotides at the respective

positions while a TALE with 4 aberrant repeats recognized primarily a target box with either 3 or 4 nucleotides deleted. This means, in TALEs with multiple aberrant repeats all or all but one aberrant repeat leave the repeat array. This strongly implies that the inter- and intra-repeat interactions between two neighbouring aberrant repeats are too weak to keep them both stabilized within the repeat array, and thus usually one, sometimes both get ejected. Moreover, an aberrant repeat that is flanked by an aberrant repeat on both sides cannot partake in DNA-binding and has to loop out. It seems likely that such an aberrant repeat is unable to stay in the repeat array because not only its immediate inter- and intra-repeat interactions are impaired but also those of its neighbouring repeats, thus further destabilizing its anchoring in the repeat array.

4.8 Novel *Xoc* TALE targets

In the course of this work, it was possible to verify seven novel *Xoc* TALE target genes that were previously postulated but for which experimental data was not available. Moreover, six of them are induced by TALE classes that were previously without a target, and thus the verification of these target genes in particular should be a valuable foundation for future studies trying to identify the natural function of these TALEs. Using the library of artificial *Xoc* TALEs generated in this study, it was also possible to identify 22 new rice genes whose promoters were specifically induced by one or multiple *Xoc* TALEs. While their TALE-dependent induction has not yet been confirmed in rice, the strong TALE-specific response of their respective reporter constructs clearly demonstrated the presence of suitable TALE binding sites within their promoters. Since these 22 genes were previously not even considered as putative TALE targets, their further analysis might offer completely new insights into the natural function of their respective *Xoc* TALE classes. The putative functions of the newly identified target genes are very diverse, ranging from transporter proteins, transcription factors, enzymes, kinases and proteins involved in the regulation of root growth to proteins that interact with RNAs or interfere with phytohormone signalling and those that regulate cell-wall structure and composition. To get an idea about how their TALE-dependent induction could affect the *Xanthomonas*-rice pathosystem, the relevant primary literature was analysed with regard to their potential natural functions and the most interesting ones are briefly discussed in the following chapters.

4.8.1 OsLsi1 – a silicon-transporting aquaporin

One of the novel TALE target genes is of particular interest since it is not only targeted by the highly conserved class TalAV from *Xoc* but, as shown in Mücke *et al*., 2019, also by the frequently occurring *Xoo* TALE class TalAL. It encodes for an aquaporin with the systematic

name of OsNIP2;1 and it was initially discovered due to its involvement in silicon (Si) transport and hence named OsLsi1 (Low silicon rice 1) (Ma *et al*., 2006; Mitani *et al*., 2008). OsLsi1 localizes to the distal side of the plasma membrane of the exo- and endodermis in rice roots. There it facilitates the uptake of silicic acid and its transport through the casparian strip (Ma *et al*., 2006; Yamaji and Ma, 2007; Ma *et al*., 2011). *OsLsi1* mutant plants accumulated less silicon, were more susceptible to diseases and showed a drastically decreased grain yield (Ma *et al*., 2006). Consequently, silicon availability was shown to affect the growth, yield and drought tolerance of rice (Savant *et al*., 1996; Epstein, 1999; Ma and Yamaji, 2006; Currie and Perry, 2007). Interestingly, it is also reported to influence the resistance of various plants against a wide range of pathogens – in rice for example against rice blast (Rodrigues *et al*., 2003a; Sun *et al*., 2010), brown spot (Ning *et al*., 2014), sheath blight (Rodrigues *et al*., 2003b) and bacterial blight (Song *et al*., 2016). Remarkably, the exact mechanisms behind these Si-mediated resistance effects remain mostly elusive. One hypothesis suggests that the Si-depositions regularly found in the leaves of Si-treated plants act as a physical barrier that prevents pathogens from penetrating the cell efficiently (Belanger *et al*., 1995; Kamenidou *et al*., 2009). A second hypothesis accounts the positive effects of Si to the accumulation of antimicrobial and antioxidative compounds like phenolics and phytoalexins, so observed in Si-treated and pathogen-challenged rice, cucumber and wheat (Fawe *et al*., 1998; Rodrigues *et al*. 2003a; Song *et al*., 2016). One study even found that the expression of several defence related genes increased if pathogen-challenged rice was provided with Si prior to the infection, causing the authors to speculate that Si might act as messenger molecule that directly regulates plant defence in general (Song *et al*., 2016). Overexpressing *OsLsi1* with a TALE might therefore be a way for *Xanthomonas* to counter some of these effects. It could for example change the levels of available Si at the immediate infection site, and thus alter Si-mediated defence responses. Since OsLsi1 was also shown to facilitate the uptake of other substances like selenite (Zhao *et al*., 2010), arsenite (Bienert *et al*., 2008; Ma *et al*., 2008) and methylated arsenic species (Li *et al*., 2009), its overexpression might also be used by the pathogen to remove one of these substances from the apoplast in order to avoid toxic effects. Moreover, all aquaporins exhibit water permeability. This means that they transport water in response to the osmotic pressure (Borgnia *et al*., 1999; Agre and Kozono, 2003). Therefore, it is even imaginable that a TALE-induced accumulation of OsLsi1 could generate or maintain a water flow that allows for an easier spreading of the pathogen, either into or within the plant. Possibly, a local increase in aquaporine abundancy might even generate an entrance point for the bacteria by causing a water flow into guard cells, and thus induce their swelling and the opening of the associated stomata.

4.8.2 The three novel target genes for class TalCL

Their annotated functions suggest that two of the newly confirmed target genes for class TalCL are involved in the regulation of cell-wall structure and cell growth in rice.

The first of these genes encodes for OsGH3.4, an indole-3-acetic acid-amido synthetase (GH3*;* Gretchen Hagen 3). GH3 proteins modify the function and activity of various phytohormones by catalysing the formation of amido conjugates, and thus they are often crucial factors influencing the interaction between plants and pathogens (Staswick and Tiryaki, 2004; Staswick *et al*., 2005; Nobuta *et al*., 2007; Hui *et al*., 2019). From the thirteen GH3 proteins in rice, four were shown to conjugation jasmonic acid (JA) with Ile (Hui *et al*., 2019) and three to facilitate preferentially the conjugation of indole-3-acetic acid (IAA) to Asp (Ding *et al*., 2008; Zhang *et al*. 2009; Fu *et al*. 2011). Interestingly, those two conjugations have antithetical effects on their respective phytohormones, with IAA-Asp causing IAA to be degraded (Staswick *et al*., 2005) while JA-Ile converts JA into its biologically active form (Katsir *et al*., 2008; Hui *et al*., 2019). Eight of the 13 GH3 family members in rice were already subjected to overexpression analysis, namely OsGH3.1, OsGH3.2, OsGH3.3, OsGH3.5, OsGH3.6, OsGH3.8, OsGH3.12 and OsGH3.13. Remarkably, in all eight cases, an overexpression led to an enhanced resistance of the affected rice plants against *Xoo, Xoc* and/or *Magnaporthe oryzae (*Zhang *et al*., 2007 and 2009; Ding *et al*., 2008; Domingo *et al*., 2009; Fu *et al*. 2011; Hui *et al*., 2019). It was even shown that the two OsGH3 proteins OsGH3.2 and OsGH3.8 are induced upon infection with *Xoo*, further supporting the idea that at least some OsGH3 proteins are directly involved in the regulation of plant resistance (Ding *et al*., 2008; Fu *et al*., 2011). The results from the IAA-conjugating OsGH3 proteins in particular are also in accordance with the observation that several plant pathogens, among them *X. oryzae*, cause an accumulation of IAA at their immediate infection site - doing so presumably to stimulate expansin expression and cell-wall loosening (Fu *et al*., 2011). Consequently, suppression of auxin accumulation and auxin signalling is well known to accompany *R*-gene mediated resistance (Ding *et al*., 2008; Fu *et al*., 2011).

Since all so far analysed OsGH3 proteins are directly involved in phytohormone homeostasis and the regulation of the defence responses, it seems very likely that the TalCL target *OsGH3.4* has a similarly crucial role. However, all of the examples given above suggest that its overexpression will have a resistance enhancing, and thus a negative effect for *Xanthomonas*. Yet, two studies from other plants indicate that an opposite effect is also possible. The first of these studies found an increased susceptibility to *Botrytis cinerea* and *Pseudomonas syringae* after overexpressing *AtGH3.2* in Arabidopsis (González-Lamothe *et al*., 2012). The second one observed a strong induction of *SlGH3.4* in tomato in response to the initiation and the successful establishment of an arbuscular symbiosis with the fungus *Rhizophagus irregularis* (Liao *et al*., 2014). It remains therefore a possibility that the TALE-

induced overexpression of *OsGH3.4* is beneficial for the pathogen. This could for example be the case if OsGH3.4 facilitates the conjugation of other than the above mentioned amino acids to IAA or JA, thus altering their tightly controlled homeostasis into the opposite direction. Alternatively, OsGH3.4 could use another phytohormone as substrate. By facilitating for example the formation of a salicylic acid (SA) amino acid conjugate that leads to a less active version of the hormone or cause its degradation, a GH3 protein might cause a suppression of the SA-mediated defence responses. Theoretically, if a high abundance of OsGH3;4 has indeed negative effects on *Xanthomonas*, its TALE-induced overexpression might even be used to flood affected plant cells with incomplete transcripts that act as silencing molecules, and thus reduce the overall presence of active proteins. At last, *OsGH3;4* could also be an unintentional target whose effects are either neglectable or outweighed by the positive effects of another TalCL target gene.

A possible candidate for the latter scenario could be the second newly identified target gene of class TalCL. This target, *Os10g39840,* encodes for the Xyloglucan endotransglucosylase/hydrolase (XTH) OsXTH20. XTHs can possess one or both of the following two catalytic activities: They can act as xyloglucan endohydrolases (XEH), causing the irreversible shortening of xyloglucan or other glucan derivatives or they can act as xyloglucan endotransglycosylase (XET), resulting in the cleavage and re-ligation of xyloglucan chains (Eklöf and Brumer, 2010). However, OsXTH20 in particular was shown to possess only XEH activity and further to exclusively act on xyloglucan (Hara *et al*., 2014). This is remarkable since Xyloglucan – while being a major compound of the plant cell wall in dicotyledones – is substantially less common in the cell wall of monocotyledons like rice (Carpita and Gibeaut, 1993; Vogel, 2008; Hsieh and Harris, 2009) where it is mostly substituted by arabinoxylans and mixed-linkage glucans (MLG) (Hara *et al*., 2014). Yet, it is possible that XTHs like OsXTH20 fulfil even in monocotyledons important functions in the regulation of cell expansion and cell wall properties, for example by acting primarily in strategically important key cells or in tissues that are enriched in xyloglucan (Hara *et al*., 2014). This line of thought seems somewhat supported by the observation that the XTH protein family in rice, a monocotyledonous plant with a low xyloglucan level, is with its 30 members as large as that of a dicotyledonous plant like Arabidposis (Yokoyama *et al*., 2004). Moreover, many *OsXTHs* were found to be expressed tissue-specific, *OsXTH19* for example in the dividing and elongating zone of shoot and internode, thus further hinting towards them having a relevant role in loosening rigid cell wall structures in particularly relevant key positions (Yokoyama *et al*., 2004). However, OsXTH19 is, so far, the only XTH from rice analysed in any detail and silencing its gene had no obvious phenotypic effects (Hara *et al*., 2014). It nonetheless remains a possibility that the TALE-induced overexpression of

OsXTH20 causes cell wall loosening, and thus beneficial effects for *Xanthomonas*, especially if occurring primarily at the immediate infection site.

Interestingly, it was possible to confirm yet another target gene for class TalCL, namely *Os01g40290.* In the original RNA-Seq analysis, this gene was found to be induced in response to every single one of the 10 *Xoc* strains (Wilkins *et al*., 2015). Since TalCL occurs only in seven of these ten, it is likely that *Os01g40290* is addressed by a second *Xoc* TALE class. Due to the large overlap in their RVD sequence and their predicted optimal target sequences, the most promising candidate for this would be class TalCJ. Interestingly, *Os01g40290* was also one of only five rice genes found to be highly induced in response to *Xoo* (PXO99) and *Xoc* (BLS256) (Cernadas *et al*., 2014) and two publications predicted it as the top target candidate for class TalAA from *Xoo* (Cernadas *et al*., 2014; Yu *et al*., 2015) – a comparatively conserved *Xoo* TALE class that remained, so far, without any confirmed target gene. Intriguingly, the predicted TALE binding sites of all three TALE classes cover exactly the same region (Fig. 4.7), thus further supporting the idea that *Os01g40290* could be another of the rare TALE target genes shared between *Xoo* and *Xoc*. Unfortunately, it is simply annotated as an expressed protein and a blast analysis using its 126aa long protein sequence yielded no similarities to any protein with known function, and thus its putative function remains enigmatic. However, there is one study hinting towards an involvement of *Os01g40290* in plant pathogenicity. In this study, it showed the highest fold change in a microarray analysis investigating resistance against gall midge in rice (Agarrwal *et al*., 2016). Given the possibility that *Os01g40290* could be targeted not only by one but three different TALE classes, moreover from *Xoo* and *Xoc*, any future study aiming to identify or evaluate TALE targets from *X. oryzae* should include an experiment that clarifies if it is induced by TalAA and/or TalCJ and, if it is, what function its induction might have.

Figure 4.7: Putative TALE binding sites within the promoter of *Os01g40290.* Shown are the RVD sequences of all unique versions of the TALE classes TalC and, TalCJ from *Xoc* and TalAA from *Xoo*. All three classes have a putative binding site in the promoter of this gene and all three cover the same region. Mismatched RVDs are boxed. The number of predicted mismatches (mm) is shown next to the RVD sequence The genes is indicated by its locus identifier and a schematic of the first 1000 bp of its promoter is shown (beginning upstream of the first ATG). The relative distance between TALE EBE (blue) and the annotated transcriptional start site (red) is indicated.

4.8.3 Leucine-rich repeat receptor-like kinases

The next newly confirmed *Xoc* TALE target gene, *Os06g36270*, encodes for a leucine-rich repeat receptor-like kinase (LRR-RLK). Interestingly, *Os06g36270* was confirmed as target of TalCP and TalCK members, while a second, very closely related LRR-RLK with the ID *Os01g53920* was identified as a putative candidate target of TalCP but not TalCK. RLKs are among the largest protein families found in plants (Tör *et al*., 2009) and LRR-RLKs in particular are often the largest subfamily, this is for example the case in rice (Shiu *et al*., 2004). Due to this fact, many LRR-RLKs remain uncharacterized. However, all LRR-RLKs typically share the following three functional domains: a LRR-containing, extracellular domain that is responsible for signal detection, a transmembrane domain needed for the connection to the cell membrane and an intracellular kinase domain necessary to relay the perceived information (Gou *et al*., 2010). Yet, despite these shared features, LRR-RLKs are grouped into several different subfamilies, many of which seem to fulfil distinct biological roles. Subfamily III members for example were shown to be primarily involved in regulating developmental processes while those from subfamily I and XII were shown to be involved mainly in defence processes (Liu *et al*., 2017). Especially in subfamily XII, three noteworthy examples are known to promote resistance against pathogens: the flg22 receptor FLS2 from Arabidopsis (Gómez-Gómez and Boller, 2002), EFR, the receptor that allows many plants to recognize the bacterial elongation factor Tu (Zipfel *et al*. 2006), and the well-known BB resistance protein XA21 (*Os11g35500*) from rice (Pruitt *et al*., 2015). Unfortunately, neither the LRR-RLK targeted by TalCK/TalCP nor the one addressed by TalCP have been analysed experimentally yet. Furthermore, both targets belong to subfamily XI, a subfamily from which, so far, only very few examples have been characterized at all. However, of those, two affect the resistance response of rice against pathogens. The first one is OsBRR1, which was shown to be involved in the resistance response to blast fungus (Peng *et al*., 2009). The second one is XIK1, which is an essential component of the *Xa21*-mediated resistance against *Xoo* (Hu *et al*., 2015). While these examples suggest that the TALE-dependent overexpression of a LRR-RLK is likely to enhance the plants resistance against pathogens, an opposite effect is also possible. The two LRR-RLKs PSKR1 and PSY1R from Arabidopsis, for example, were initially discovered due to their involvement in the regulation of growth and development but later they were also found to cause a general suppression of pattern-triggered immunity when overexpressed (Mosher *et al*., 2013). Another LRR-RLK from Arabidopsis, termed Impaired Oomycete Susceptibility 1 (IOS1), was shown to positively contribute to the severity of downy mildew disease (Hok *et al*., 2011). A last example, while not belonging to the LRR-RLK, is the symbiosis RLK SYMRK. This protein was shown to be an essential component during the establishment of symbiosis in legumes and other actinorhizal plants (Markmann *et al*., 2008; Antolin-Llovera *et al*., 2014).

Taken together, these examples clearly demonstrate that the TALE-dependent upregulation of the two LRR-RLKs *Os06g36270* and *Os01g53920* could have either a positive or a negative effect on the virulence of *Xanthomonas*, and thus that further experimental analysis towards their natural function is needed.

Interestingly, there is a third LRR-RLK that was also shown to contain a TALE binding site in its promoter, *Os08g14940*, one of the putative new targets for TalCQ. It encodes for a LRR-RLK from subfamily XII, and thus it belongs to the same subfamily as *XA21* (Liu *et al*., 2017). Intriguingly, results from domain swapping experiments indicate that the LRR domain of LRR-RLKs is the major determinant for the specific recognition of *Xoo* races (Zhao *et al*., 2009). It is therefore tempting to speculate that the different members of this particular subfamily have similar functions and are simply activated by different elicitor molecules. While further experimental date is needed, this could indicate that TalCQ involuntarily induces a resistance enhancing target gene.

4.8.4 The cation/H^+ exchanger OsCHX15

Os12g42200 encodes for the cation/H^+ exchanger OsCHX15 and is induced by one of the two members of TalCG. While nothing is known about the function of OsCHX15 in particular, a close relative in rice, OsCHX14, as well as its closest homologue from Arabidopsis, AtCHX19, were both found to regulate K^+ homeostasis by facilitating a proton coupled transport of the cation (Sze *et al*., 2004; Chen *et al*., 2016). Knockout experiments also suggest an involvement of AtCHX19 and several other members of the AtCHX family in male fertility and embryo development, presumably by impacting cell wall and plasma membrane remodelling (Padmanaban *et al*., 2017). Such a function in cell wall remodelling would render OsCHX15 a valuable TALE target since cell wall integrity is often a crucial factor affecting the outcome of a plant-pathogen interaction. However, the functional diversity of CHXs is not yet exhausted and other examples with similarly interesting functions exist. One in particular is worth mentioning here, namely CHX20 from Arabidopsis. It was shown to be a key factor regulating the osmolality of guard cells, thus being directly involved in the opening and closing of stomata (Padmanaban *et al*., 2007). Targeting a CHX to force a local opening of stomata or to prevent the plant from closing them in course of a resistance reaction would be of immense benefit for a pathogen like *Xoc* since it would help to establish and secure an entry point into the plant. However, even altering the osmolality of normal cells could be beneficial for the pathogen since it might cause an abnormal enlargement of the affected cells. Such hypertrophies are a frequently occurring symptom in numerous plant-pathogen interactions and, with AvrBs3 from *Xanthomonas campestris* pv. *vesicatoria*, there is even one example known in which a TALE induces a hypertrophy (Marois *et al*., 2002; Kay *et al*.,

2007). However, while it is assumed that this TALE-induced hypertrophy helps to spread the bacteria, this has, so far, not been confirmed experimentally.

4.8.5 ROX3 – a negative regulator of the *Xa21*-mediated resistance reaction

Another newly identified gene with a TALE-responsive promoter is *ROX3*. It was shown to be addressed by at least two different versions of TalAF, one originating in *Xoo* and one in *Xoc*. ROX3 was identified as negative regulator of the XA21-mediated immunity (ROX) (Lee *et al*., 2011). It was shown to interact with at least two different proteins involved in the XA21-mediated resistance. and its overexpression in rice plants homozygous for the *Xa21* resistance gene resulted in an enhanced susceptibility to *Xoo* (Lee *et al*., 2011). ROX3 belongs to the family of nuclear migration proteins. Members of this protein family are ubiquitous in eukaryotes and several of them were even shown to be crucial for cell division by affecting formation of the mitotic spindles and chromosome separation during mitosis (Riera and Lazo, 2009). Interestingly, this function shows a remarkable correlation to a recently discovered mechanism to suppress *R*-gene-mediated resistance in rice. By inducing the expression of the cell-cycle regulator Cyclin-D4-1 with a TALE, several *Xoc* strains seem able to suppress the *Xa7*-mediated resistance despite the presence of the adequate elicitor (Cai *et al*., 2017).

4.8.6 GA20ox5 and OsSAURs –TALE targets affecting phytohormone balance

Another gene with a newly confirmed TALE-binding site in its promoter is *GA20ox5*. It encodes for the putative gibberellin 20 oxidase 5 (Han and Zhu *et al*., 2011). Ga20ox proteins were shown to catalyse several crucial steps in the biosynthesis of gibberellic acid (GA) (Salazar-Cerezo *et al*., 2018). Consequently, silencing of OsGA20ox1 and OsGA20ox2, two other members of the OsGA20ox protein family in rice, caused semi-dwarfism and severely reduced GA content (Ashikari *et al*., 2002; Spielmeyer *et al*., 2002) while overexpressing OsGA20ox1 increased plant size and the levels of bioactive GA (Oikawa *et al*., 2004). GA is also well known to negatively affect plant defence in general (Robert-Seilaniantz *et al*., 2007; Nagel and Peters, 2017). Consequently, the overexpression of *OsGA20ox3*, yet another member of the OsGA20 protein family, caused a higher susceptibility against the two rice pathogens *Magnaporthe oryzae* and *Xoo* while transgenic rice silenced for this gene showed an enhanced resistance and an increased expression of defence-related genes (Qin *et al*., 2012). Interestingly, multiple plant pathogenic bacteria, among them *X. translucens* and *Xoc*, were shown to possess an operon containing several genes for the biosynthesis of GA (Nagel *et al*., 2017), with one study demonstrating even the

production of a GA-precursor by *Xoc* (Lu *et al*., 2015). It is hypothesized that the production and secretion of GA by these pathogens antagonizes and suppresses the JA-mediated defence response associated with wounds, thereby facilitating the initial colonisation of and the spreading within wounded tissue (Nagel and Peters, 2017). To additionally Hijack the GA production of the host might lead to an even greater increase in the local GA levels, thus supporting the pathogens capabilities to suppress plant defences even further.

With *OsSAUR30* and *OsSAUR52*, two genes with newly confirmed TALE binding sites in their promoters are annotated as small auxin upregulated RNAs. Despite their name, *SAURs* encode actually for small, short-lived proteins and they are also not only responsive to auxin but can be regulated by a number of other phytohormones as well, among them brassinosteroids (BR), gibberellins (GA), jasmonate, abscisic acid (ABA) (Ren and Gray, 2015). Consequently, different SAURs were found to localize not only to different cellular compartments, they also exhibit distinct expression patterns that vary between tissues, developmental stages and in response to light, cold, heat or drought (Jain *et al*., 2006; Du *et al*., 2013; Stortenbeker and Bemer, 2019). Many SAURs have been studied in Arabidopsis where they were shown to act as positive and negative regulators for a wide range of processes, among them auxin transport, root growth and cell expansion (Ren and Gray, 2015). While the two potential TALE targets *OsSAUR30* and *OsSAUR52* from rice were not yet characterized, three other SAURs from rice have been analysed and all three negatively regulate auxin synthesis and transport. Since many pathogens, among them *Xoo*, cause an accumulation of auxin at their infection site (Ding *et al*., 2008; Fu and Wang, 2011), an overexpression of any of these three SAURs should result in potentially negative effects for the pathogen. This line of thought is confirmed by the observation that a transgenic rice line deficient for one of these three genes, *OsSAUR51*, shows an increased susceptibility to *Xoo* (Aoki *et al*., 2016). However, given the wide range of possible SAUR functions in general and their involvement in phytohormone signalling in particular, it is possible that the TALE-induced expression of one or both of the above-mentioned members of this protein family has an opposite effect, and thus is of benefit for the pathogen. Interestingly, SAURs were also found to be induced in response to TALEs from two other *Xanthomonas* species. The first example is from pepper. There, five different *SAURs* were induced in response to the TALE AvrBs3. However, their induction is most likely an indirect response caused by the overexpression of the original TALE target, the TF UPA20 (Marois *et al*., 2002). Since *SAUR* transcripts were also shown to accumulate shortly before cell enlargement in soybean (McClure and Guilfoyle, 1989), it is reasonable to assume that the induction of the five SAURs in pepper might nonetheless contribute towards the typical hypertrophic phenotype caused by AvrBs3 (Kay *et al*., 2007). Another example is from tomato, where two *SAURs* were among the genes found to be upregulated when the plants were subjected to the

presence of TALE AvrHah1 (Schwartz *et al*., 2017). Inoculation or overexpression of AvrHah1 leads to water uptake and water-soaking lesions. While this effect was eventually found to be the result of two bHLH TFs that are directly addressed by TALE AvrHah1, the study did unfortunately not address whether the two SAURs are impacting the observed phenotype in any way or if they cause any additional effects (Schwartz *et al*., 2017).

4.9 Common ground – TALE target genes shared between *Xoc* and *Xoo*

Only two TALE classes occur in strains from *Xoo* and *Xoc*. Therefore, their target genes might be of particular relevance for the virulence of *Xanthomonas*. The first one is TalAF and its previously discussed candidate target *ROX3* (Fig. 4.8). The second one is TalAD. Its members are reported to activate two genes, one encoding for the VQ-domain containing protein OsVQ7 (Fig. 4.8; Cernadas *et al*., 2014; Wilkins *et al*., 2015) and the other one for the F-box protein FBX109 (Mücke *et al*., 2019). While there is no information available about OsVQ7 in particular, VQ proteins in general are known to interact with WRKY TFs and in doing so, they regulate numerous processes that affect photomorphogenesis, the development of seeds and the responses to abiotic and biotic stresses (Cheng *et al*., 2012; Kwon and Hwang, 2014; Jing and Lin, 2015). There are also no specific information about FBX109 available but like all F-box proteins it is most likely involved in the protein degradation machinery of the cell. Interestingly, XopI and XopL, two non-TALE effector proteins from *Xanthomonas* spp., are known to hijack the protein degradation machinery by either mimicking an ubiquitin ligase (Singer *et al*., 2013) or an F-box protein (Üstün and Börnke, 2014). Using a TALE to induce an F-box protein as an alternative way to compromise the protein degradation machinery of the cell seems therefore at least theoretically a promising approach for the bacterium. Curiously, despite their occurrence in both pathovars a significant decrease in virulence was neither observed for a TALE-mutant of the *Xoc* strain BL256 lacking TalAD nor for the one lacking TalAF (M38 = ΔTalAD; M2 = ΔTalAF; Cernadas *et al*., 2014).

When the present work was started, there was only a single rice gene known to be targeted by two different TALEs from the two *X. oryzae* pathovars, the methyltransferase-encoding gene *OsHEN1* (Fig. 4.8; Moscou and Bogdanove, 2009; Cernadas *et al*., 2014; Wilkins *et al*., 2015). Curiously, despite this special status, experiments in the *Xoc* strain BLS256 demonstrated that the deletion of the *OsHEN1*-activating TALE has also no major effect on the virulence of the pathogen, leaving the question unanswered why its activation is conserved (M87 = ΔTalAKv2; Cernadas *et al*., 2014). Recently, the existence of yet another target shared by different TALEs from *Xoo* and *Xoc* was confirmed, the TF-encoding gene *OsERF#123* (Wang *et al*., 2017; Doucouré *et al*., 2018; Tran *et al*., 2018). This gene was

shown to be addressed by several variants of TalB (*Xoo* from Africa) as well as by one variant of TalBI (*Xoc*) (Wang *et al*., 2017; Doucouré *et al*., 2018; Tran *et al*., 2018). *OsERF#123* also contains two distinct binding sites for class TalBI, one in forward and one in reverse orientation and both are reported to be essential for a successful induction (Fig. 4.8; Wang *et al*., 2017). Curiously, Tran and colleagues reported a virulence enhancing effect caused by the TALE-induced expression of *OsERF#123*, while and colleagues did observe no such effect. Due to this discrepancy and due to the fact that its downstream-activated targets are still unknown, the exact contribution of OsERF#123 to virulence remains uncertain (Doucouré *et al*., 2018; Tran *et al*., 2018). The present work contributed to the identification of two new TALE targets shared between *Xoo* and *Xoc*. The first is the silicon transporter OsLsi1 which is addressed by class TalAV from *Xoc* and by class TalAL from *Xoo* (Fig. 4.8; Mücke *et al*., 2019). In addition to *OsERF#123* and *OsSWEET14*, *OsLsi1* is one of only three genes whose promoters contain at least two completely distinct TALE binding sites (Fig. 4.8; Hutin *et al*., 2015a; Tran *et al*., 2018). Several TALE binding sites within the promoter of a single target gene strongly implies that the gene was targeted multiple times during evolution, and thus fulfils an important role for the virulence of the pathogen. Yet of these three genes, *OsSWEET14* is the only one with a significant effect on the virulence of *X. oryzae* (Antony *et al*., 2010*).* For a *Xoc* strain deficient for the *OsLsi1*-inducing TALE, no relevant decrease in virulence was observed (M80 = ΔTalAV; Cernadas *et al*., 2014). Since it is possible that the numerous TALEs remaining in this mutant strain were able to compensate for the loss of TalAV, it seems advisable to analyse the two *OsLsi1* inducing TALE classes again in a context with less or no other TALEs present. The subsequent search for a phenotype should also take into account what is known about the function of OsLsi1 since it might give valuable insights on how its TALE-dependent overexpression might benefit *Xanthomonas* and what specific effects to expect. The second new gene found to be induced by TALEs from both *X. oryzae* pathovars is *OsDOX-1*. While the activation of *OsDOX-1* by members of TalBR and TalCH from *Xoc* was already published (Cernadas *et al*., 2014; Wilkins *et al*., 2015), our lab could recently demonstrate that it is also induced by members of class TalAQ from *Xoo* (Mücke *et al*., 2019). Despite numerous differences in their RVD sequences, all three TALE classes are using the same binding site to activate *OsDOX-1* (Fig. 4.8). Curiously, yet another published *Xoc* TALE target, the one addressed by TalBL (Cernadas *et al*., 2014; Wilkins *et al*., 2015), is in fact a closely related homologue of *OsDOX-1* and might therefore have not only a similar function but affect also the same biological pathway. Remarkably, this connection was completely overlooked in previous studies. Consequently, the re-naming of this target to *OsDOX-2* was proposed (Fig. 4.8; Mücke *et al*., 2019). While these observations strongly suggest that the affected biological pathway is of high relevance for *X. oryzae*, no visible effect on virulence was found

due to its TALE-dependent induction - at least in case of BLS256 mutant strains that lack either one of the two *OsDOX*-inducing TALEs (M35 = ΔTalBL and M79 = ΔTalBR; Cernadas *et al*., 2014). However, since Cernadas and colleagues were not aware that the two different TALE classes induce homologous genes, a *Xanthomonas* strain deficient for both *TALE* genes was not generated. Since both proteins could function in a redundant manner, no definite conclusions can be drawn. Both genes, *OsDOX-1* and *2*, encode for 2-oxoglutarate dioxigenases (Falcone Ferreyra *et al*., 2015). Unfortunately, the existing literature about their biochemical function is inconclusive, with one group reporting that OsDOX-1 uses naringenin as a substrate (Kim *et al*., 2008) while another group measured no significant enzymatic activity with this particular substrate (Lam *et al*., 2014). However, it is noteworthy that OsDOX-1 and OsDOX-2 are the two rice proteins that are most closely related to the downy mildew resistant 6 (DMR6) protein from Arabidopsis (Falcone Ferreyra *et al*., 2015). Since different groups published also contradictory results towards the biochemical function of AtDMR6, stating that it can use either SA or flavones as a substrate, its exact enzymatic

Figure 4.8: TALE target genes shared between *Xoo* and *Xoc*. Shown are all rice genes known to be simultaneously addressed by TALEs from *Xoo* and *Xoc*. Each target gene is represented by 500bp upstream of its ATG. The positions of the different TALE binding sites are indicated. Red TALEs/TALE classes occur only in *Xoc*, those in green occur only in *Xoo* and those sown in mixed colour occur in both pathovars. In two cases, different TALEs address a single gene via at least two completely different binding sites, namely *OsERF#123* and *OsLsi1*.

function remains also unclear (Falcone Ferreyra *et al*., 2015; Zhang *et al*., 2017). However, *AtDMR6* was shown to be specifically upregulated during pathogen infection and disrupting its coding sequence caused an accumulation of SA that lead to a broad-spectrum resistance against multiple pathogens (van Damme *et al*., 2008; Zeilmaker *et al*., 2015; Zhang *et al*., 2017). Mutating the *dmr6* ortholog in tomato similarly conferred a broad-spectrum resistance against several pathogens, among them *Xanthomonas* spp. (de Toledo Thomazella *et al*., 2016). Given these observations and the close phylogenetic relationship of OsDOX-1 and 2, it is possible that a similarly important function is fulfilled by a combination of both proteins. Altering the abundance of enzymes involved in SA or flavone homeostasis appears to be a promising approach for *Xanthomonas* since both substances are known to be strongly involved in resistance against pathogens. The SA pathway in particular was just recently shown again to be crucial for the resistance against *Xoo* (Zhang *et al*., 2018).

Taken together these examples highlight that *Xoo* and *Xoc* share more TALE-dependent infection strategies than previously assumed. Analysing them in particular could help to unravel the most fundamental mechanisms utilized by *X. oryzae*.

4.10 The on-going search for phenotypic effects

Many *X. oryzae* strains, especially from the pathovar *oryzicola*, contain a high number of TALEs. Remarkably, only few of these TALEs were shown to affect the virulence of the pathogen in a visible manner. It is therefore possible that some of them contribute only in a minor way to virulence or that they function in a redundant fashion. It was even suggested that most *X. oryzae* TALEs fulfil no immediate function in the infection and represent simply a source for recombination and evolutionary-driven selection processes (Schandry *et al*., 2018). However, it is likewise possible that the experimental approaches usually used to determine the virulence of a *X. oryzae* strain are simply not suited to visualize the effects of all TALEs and their target genes - even if they are relevant under natural conditions. Keeping this in mind, it might be worthwhile to analyse the *Xoc* TALEs once more for a phenotypic effect but this time in a context with a reduced number of TALEs, preferentially one by one, thus avoiding effects caused by functionally redundant TALEs completely. For this, the single TALE-expressing *Xo* strains generated in the present study are well suited. It is also imaginable that some TALEs fulfil a function that becomes only visible if specific requirements are fulfilled, e.g. a certain genetic background of the host plant or the presence of an additional TALE. The authors of a recent publication found exactly such an example by showing that a *Xoc* TALE target gene originally reported to have no effect on virulence, Cyclin-D4-1, exhibits in fact a resistance supressing function but doing so only under very specific conditions (Cai *et al*., 2017). The effect was only observed in a rice line carrying the resistance gene *Xa7* and only when the elicitor for the resistance gene, the TALE AvrXa7,

was present as well (Cai *et al*., 2017). It is therefore tempting to speculate that some of the other TALEs without obvious virulence effect are involved in similar but yet not discovered mechanisms. Analysing the previously mentioned single TALE-expressing strains in combination with rice lines carrying known resistance genes and in the presence of their elicitors could be used to investigate this idea.

Another valuable approach could be the generation and analysis of rice lines that are defective in individual TALE target genes. Some of these plants might show an altered phenotype when grown under normal conditions, and thus their subsequent analysis could help to understand the natural function of the affected gene. Plants that show no obvious phenotype on the other hand could be used for inoculation experiments. Comparing the effects of *X. oryzae* strains with known differences in their TALE repertoires on these plants and on unaltered wild type plants, might allow to identify additional TALE target genes that actually affect virulence. While this might seem like an obvious approach, it has, so far, only been used for a few selected TALE targets, those with a major effect on susceptibility like for example *CsLOB1* from citrus (Jia *et al*., 2017). Additionally, it might be worthwhile to generate plants that contain only small deletions within their TALE binding sites. Having altered multiple TALE binding sites at once, could even clarify if and how the failure to induce several TALE target genes affects the virulence of the a *Xanthomomas* strain while at the same time being less intrusive than a simultaneous knockout of multiple genes. Again, this approach was already successfully used to create pathogen-resistant plants, however, only with TALE targets known to affect the pathogenicity of *Xanthomonas* in a major way e.g. the OsSWEET sugar transporters from rice (Li *et al*., 2012; Blanvillain-Baufumé *et al*., 2017). Due to their scientific value, the generation of such rice lines, ones with knocked out TALE target genes as well as ones with small deletions in the TALE binding site, was started in this study and preliminary results indicate successful editing events in several TALE target genes and binding sites (data not shown).

Interestingly, a recent study found that two target genes addressed by the same TALE/TALE class, namely *OsTFX1* and *OsERF#123*, can contribute to virulence (Tran *et al*., 2018). There are also results indicating that the most conserved TALE target gene is not necessarily the one with the only or even the most drastic effects on pathogenicity. The induction of Cyclin-D4-1 by the members of TalBV, for example, was shown to supress the function of the resistance gene *Xa7*. Yet, a second target gene, which encodes for a GATA zinc finger protein and which is induced by members of TalBV and TalCB, is reported to have no phenotype (Cai *et al*., 2017). While more conserved target genes are still more likely to carry a relevant function in virulence, these two studies highlight the important to identify and analyse as many target genes as possible, even those that are addressed by only some TALEs/TALE class members.

References

Abe, M., Yoshikawa, T., Nosaka, M., Sakakibara, H., Sato, Y., Nagato, Y., and Itoh, J.-i. (2010). *WAVY LEAF1*, an ortholog of Arabidopsis *HEN1*, regulates shoot development by maintaining microRNA and trans-acting small interfering RNA accumulation in rice. Plant Physiology *154*, 1335-1346.

Agarrwal, R., Padmakumari, A.P., Bentur, J.S., and Nair, S. (2016). Metabolic and transcriptomic changes induced in host during hypersensitive response mediated resistance in rice against the Asian rice gall midge. Rice *9*, 5.

Agre, P., and Kozono, D. (2003). Aquaporin water channels: molecular mechanisms for human diseases. FEBS Letters *555*, 72-78.

Al-Saadi, A., Reddy, J.D., Duan, Y.P., Brunings, A.M., Yuan, Q., and Gabriel, D.W. (2007). All five host-range variants of *Xanthomonas citri* carry one *pthA* homolog with 17.5 repeats that determines pathogenicity on citrus, but none determine host-range variation. Molecular Plant-Microbe Interactions *20*, 934-943.

Alfano, J.R., and Collmer, A. (2004). TYPE III SECRETION SYSTEM EFFECTOR PROTEINS: double agents in bacterial disease and plant defense. Annual Review of Phytopathology *42*, 385-414.

Antolín-Llovera, M., Ried, Martina K., and Parniske, M. (2014). Cleavage of the SYMBIOSIS RECEPTOR-LIKE KINASE ectodomain promotes complex formation with Nod Factor Receptor 5. Current Biology *24*, 422-427.

Antony, G., Zhou, J., Huang, S., Li, T., Liu, B., White, F., and Yang, B. (2010). Rice *xa13* recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene *Os-11N3*. The Plant Cell *22*, 3864-3876.

Aoki, H., Onishi, A., Miyashita, M., Miyagawa, H., Yatou, O., and Saito, K. (2016). Involvement of the rice *OsSAUR51* gene in the auxin-related field resistance mechanism against bacterial blight disease. Japan Agricultural Research Quarterly: JARQ *50*, 219-227.

Ashikari, M., Sasaki, A., Ueguchi-Tanaka, M., Itoh, H., Nishimura, A., Datta, S., Ishiyama, K., Saito, T., Kobayashi, M., Khush, G.S*., et al.* (2002). Loss-of-function of a rice gibberellin biosynthetic gene, *GA20 oxidase* (*GA20ox-2*), led to the rice 'Green Revolution'. Breeding Science *52*, 143-150.

Bai, J., Choi, S.-H., Ponciano, G., Leung, H., and Leach, J.E. (2000). *Xanthomonas oryzae* pv. *oryzae* avirulence genes contribute differently and specifically to pathogen aggressiveness. Molecular Plant-Microbe Interactions *13*, 1322-1329.

Baker, F.R., Leach, K.A., and Braun, D.M. (2012). SWEET as sugar: new sucrose effluxers in plants. Molecular Plant *5*, 766-768.

Barkan, A., and Small, I. (2014). Pentatricopeptide repeat proteins in plants. Annual Review of Plant Biology *65*, 415-442.

Bélanger, R.R., Bowen, P.A., Ehret, D.L., and Menzies, J.G. (1995). Soluble silicon: its role in crop and disease management of greenhouse crops. Plant Disease *79*, 329-336.

Bella, J., Hindle, K.L., McEwan, P.A., and Lovell, S.C. (2008). The leucine-rich repeat structure. Cellular and Molecular Life Sciences *65*, 2307-2333.

Bernstein, D., Hook, B., Hajarnavis, A., Opperman, L., and Wickens, M. (2005). Binding specificity and mRNA targets of a *C. elegans* PUF protein, FBF-1. RNA *11*, 447-458.

Bienert, G.P., Thorsen, M., Schüssler, M.D., Nilsson, H.R., Wagner, A., Tamás, M.J., and Jahn, T.P. (2008). A subgroup of plant aquaporins facilitate the bi-directional diffusion of As(OH)3 and Sb(OH)3 across membranes. BMC Biology *6*, 26.

Bimolata, W., Kumar, A., M, S.K.R., Sundaram, R.M., Laha, G.S., Qureshi, I.A., and Ghazi, I.A. (2015). Nucleotide diversity analysis of three major bacterial blight resistance genes in rice. PLOS ONE *10*, e0120186.

Bjedov, I., Tenaillon, O., Gérard, B., Souza, V., Denamur, E., Radman, M., Taddei, F., and Matic, I. (2003). Stress-induced mutagenesis in bacteria. Science *300*, 1404-1409.

Blanvillain-Baufumé, S., Reschke, M., Solé, M., Auguy, F., Doucoure, H., Szurek, B., Meynard, D., Portefaix, M., Cunnac, S., Guiderdoni, E.*, et al.* (2017). Targeted promoter editing for rice resistance to *Xanthomonas oryzae* pv. *oryzae* reveals differential activities for *SWEET14*-inducing TAL effectors. Plant Biotechnology Journal *15*, 306-317.

Boch, J., and Bonas, U. (2010). *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. Annual Review of Phytopathology *48*, 419-436.

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.

Borgnia, M., Nielsen, S., Engel, A., and Agre, P. (1999). Cellular and molecular biology of the aquaporin water channels. Annual Review of Biochemistry *68*, 425-458.

Cai, L., Cao, Y., Xu, Z., Ma, W., Zakria, M., Zou, L., Cheng, Z., and Chen, G. (2017). A transcription activator-like effector Tal7 of *Xanthomonas oryzae* pv. *oryzicola* activates rice gene *Os09g29100* to suppress rice immunity. Scientific Reports *7*, 5089.

Carpenter, S.C.D., Mishra, P., Ghoshal, C., Dash, P.K., Wang, L., Midha, S., Laha, G.S., Lore, J.S., Kositratana, W., Singh, N.K.*, et al.* (2018). A Strain of an emerging indian *Xanthomonas oryzae* pv. *oryzae* pathotype defeats the rice bacterial blight resistance gene *xa13* without inducing a clade III *SWEET* gene and is nearly identical to a recent thai isolate. Frontiers in Microbiology *9*, 2703.

Carpita, N.C., and Gibeaut, D.M. (1993). Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. The Plant Journal *3*, 1-30.

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

Cernadas, R.A., Doyle, E.L., Niño-Liu, D.O., Wilkins, K.E., Bancroft, T., Wang, L., Schmidt, C.L., Caldo, R., Yang, B., White, F.F.*, et al.* (2014). Code-assisted discovery of TAL effector targets in bacterial leaf streak of rice reveals contrast with bacterial blight and a novel susceptibility gene. PLOS Pathogens *10*, e1003972.

Chen, L.-Q. (2014). SWEET sugar transporters for phloem transport and pathogen nutrition. New Phytologist *201*, 1150-1155.

Chen, L.-Q., Qu, X.-Q., Hou, B.-H., Sosso, D., Osorio, S., Fernie, A.R., and Frommer, W.B. (2012). Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. Science *335*, 207-211.

Chen, L.Q., Hou, B.H., Lalonde, S., Takanaga, H., Hartung, M.L., Qu, X.Q., Guo, W.J., Kim, J.G., Underwood, W., Chaudhuri, B.*, et al.* (2010). Sugar transporters for intercellular exchange and nutrition of pathogens. Nature *468*, 527-532.

Chen, X., Liu, J., Cheng, Y., and Jia, D. (2002). *HEN1* functions pleiotropically in Arabidopsis development and acts in C function in the flower. Development *129*, 1085-1094.

Chen, Y., Ma, J., Miller, A.J., Luo, B., Wang, M., Zhu, Z., and Ouwerkerk, P.B.F. (2016). OsCHX14 is involved in the K^+ homeostasis in rice (*Oryza sativa*) flowers. Plant and Cell Physiology *57*, 1530-1543.

Cheng, S., Gutmann, B., Zhong, X., Ye, Y., Fisher, M.F., Bai, F., Castleden, I., Song, Y., Song, B., Huang, J.*, et al.* (2016). Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. The Plant Journal *85*, 532-547.

Cheng, Y., Zhou, Y., Yang, Y., Chi, Y.-J., Zhou, J., Chen, J.-Y., Wang, F., Fan, B., Shi, K., Zhou, Y.-H.*, et al.* (2012). Structural and functional analysis of VQ motif-containing proteins in Arabidopsis as interacting proteins of WRKY transcription factors. Plant Physiology *159*, 810-825.

Chong, J., Piron, M.-C., Meyer, S., Merdinoglu, D., Bertsch, C., and Mestre, P. (2014). The SWEET family of sugar transporters in grapevine: VvSWEET4 is involved in the interaction with *Botrytis cinerea*. Journal of Experimental Botany *65*, 6589-6601.

Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J., and Voytas, D.F. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. Genetics *186*, 757-761.

Chu, Z., Yuan, M., Yao, J., Ge, X., Yuan, B., Xu, C., Li, X., Fu, B., Li, Z., Bennetzen, J.L.*, et al.* (2006). Promoter mutations of an essential gene for pollen development result in disease resistance in rice. Genes & Development *20*, 1250-1255.

Cohn, M., Bart, R.S., Shybut, M., Dahlbeck, D., Gomez, M., Morbitzer, R., Hou, B.-H., Frommer, W.B., Lahaye, T., and Staskawicz, B.J. (2014). *Xanthomonas axonopodis* virulence is promoted by a transcription activator-like effector–mediated induction of a SWEET sugar transporter in Cassava. Molecular Plant-Microbe Interactions *27*, 1186-1198.

Cong, L., Zhou, R., Kuo, Y.-c., Cunniff, M., and Zhang, F. (2012). Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. Nature Communications *3*, 968.

Cox, K.L., Meng, F., Wilkins, K.E., Li, F., Wang, P., Booher, N.J., Carpenter, S.C.D., Chen, L.-Q., Zheng, H., Gao, X.*, et al.* (2017). TAL effector driven induction of a *SWEET* gene confers susceptibility to bacterial blight of cotton. Nature Communications *8*, 15588.

Cuculis, L., Abil, Z., Zhao, H., and Schroeder, C.M. (2015). Direct observation of TALE protein dynamics reveals a two-state search mechanism. Nature Communications *6*, 7277.

Currie, H.A., and Perry, C.C. (2007). Silica in plants: biological, biochemical and chemical studies. Annals of Botany *100*, 1383-1389.

Das, A.K., Cohen, P.W., and Barford, D. (1998). The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. The EMBO Journal *17*, 1192-1199.

de Toledo Thomazella, D.P., Brail, Q., Dahlbeck, D., and Staskawicz, B.J. (2016). CRISPR-Cas9 mediated mutagenesis of a *DMR6* ortholog in tomato confers broad-spectrum disease resistance. bioRxiv, 064824.

Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.-K., Shi, Y., and Yan, N. (2012a). Structural basis for sequence-specific recognition of DNA by TAL effectors. Science *335*, 720-723.

Deng, D., Yan, C., Wu, J., Pan, X., and Yan, N. (2014). Revisiting the TALE repeat. Protein & Cell *5*, 297-306.

Deng, D., Yin, P., Yan, C., Pan, X., Gong, X., Qi, S., Xie, T., Mahfouz, M., Zhu, J.-K., Yan, N.*, et al.* (2012b). Recognition of methylated DNA by TAL effectors. Cell Research *22*, 1502-1504.

Ding, X., Cao, Y., Huang, L., Zhao, J., Xu, C., Li, X., and Wang, S. (2008). Activation of the indole-3-acetic acid-amido synthetase GH3-8 suppresses expansin expression and promotes salicylate- and jasmonate-independent basal immunity in rice. The Plant Cell *20*, 228-240.

Domingo, C., Andrés, F., Tharreau, D., Iglesias, D.J., and Talón, M. (2009). Constitutive expression of *OsGH3.1* reduces auxin content and enhances defense response and resistance to a fungal pathogen in rice. Molecular Plant-Microbe Interactions *22*, 201-210.

Doucouré, H., Pérez-Quintero, A.L., Reshetnyak, G., Tekete, C., Auguy, F., Thomas, E., Koebnik, R., Szurek, B., Koita, O., Verdier, V.*, et al.* (2018). Functional and genome sequence-driven characterization of *tal* effector gene repertoires reveals novel variants with altered specificities in closely related malian *Xanthomonas oryzae* pv. *oryzae* strains. Frontiers in Microbiology *9*, 1657.

Doyle, E.L., Booher, N.J., Standage, D.S., Voytas, D.F., Brendel, V.P., VanDyk, J.K., and Bogdanove, A.J. (2012). TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. Nucleic Acids Research *40*, W117-W122.

Doyle, E.L., Stoddard, B.L., Voytas, D.F., and Bogdanove, A.J. (2013). TAL effectors: highly adaptable phytobacterial virulence factors and readily engineered DNA targeting proteins. Trends in Cell Biology *23*, 390-398.

Du, H., Liu, H., and Xiong, L. (2013). Endogenous auxin and jasmonic acid levels are differentially modulated by abiotic stresses in rice. Frontiers in Plant Science *4*, 397.

Duan, Y.P., Castañeda, A., Zhao, G., Erdos, G., and Gabriel, D.W. (1999). Expression of a single, host-specific, bacterial pathogenicity gene in plant cells elicits division, enlargement, and cell death. Molecular Plant-Microbe Interactions *12*, 556-560.

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

Eklöf, J.M., and Brumer, H. (2010). The *XTH* gene family: an update on enzyme structure, function, and phylogeny in xyloglucan remodeling. Plant Physiology *153*, 456-466.

Epstein, E. (1999). SILICON. Annual Review of Plant Physiology and Plant Molecular Biology *50*, 641-664.

Erkes, A., Reschke, M., Boch, J., and Grau, J. (2017). Evolution of transcription activator-like effectors in *Xanthomonas oryzae*. Genome Biology and Evolution *9*, 1599-1615.

Falcone Ferreyra, M.L., Emiliani, J., Rodriguez, E.J., Campos-Bermudez, V.A., Grotewold, E., and Casati, P. (2015). The identification of maize and Arabidopsis type I FLAVONE SYNTHASEs links flavones with hormones and biotic interactions. Plant Physiology *169*, 1090-1107.

Fawe, A., Abou-Zaid, M., Menzies, J.G., and Bélanger, R.R. (1998). Silicon-mediated accumulation of flavonoid phytoalexins in cucumber. Phytopathology *88*, 396-401.

Ferreira, R.M., de Oliveira, A.C.P., Moreira, L.M., Belasque, J., Gourbeyre, E., Siguier, P., Ferro, M.I.T., Ferro, J.A., Chandler, M., and Varani, A.M. (2015). A TALE of transposition: Tn*3*-like transposons play a major role in the spread of pathogenicity determinants of *Xanthomonas citri* and other xanthomonads. mBio *6*, e02505-14.

Filipovska, A., and Rackham, O. (2012). Modular recognition of nucleic acids by PUF, TALE and PPR proteins. Molecular BioSystems *8*, 699-708.

Fournier, D., Palidwor, G.A., Shcherbinin, S., Szengel, A., Schaefer, M.H., Perez-Iratxeta, C., and Andrade-Navarro, M.A. (2013). Functional and genomic analyses of alpha-solenoid proteins. PLOS ONE *8*, e79894.

Fu, J., Liu, H., Li, Y., Yu, H., Li, X., Xiao, J., and Wang, S. (2011). Manipulating broad-spectrum disease resistance by suppressing pathogen-induced auxin accumulation in rice. Plant Physiology *155*, 589-602.

Fu, J., and Wang, S. (2011). Insights into auxin signaling in plant-pathogen interactions. Frontiers in Plant Science *2*, 74.

Gao, H., Wu, X., Chai, J., and Han, Z. (2012). Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. Cell Research *22*, 1716-1720.

Geiβler, R., Scholze, H., Hahn, S., Streubel, J., Bonas, U., Behrens, S.-E., and Boch, J. (2011). Transcriptional activators of human genes with programmable DNA-specificity. PLOS ONE *6*, e19509.

Gerber, A.P., Herschlag, D., and Brown, P.O. (2004). Extensive association of functionally and cytotopically related mRNAs with PUF family RNA-binding proteins in yeast. PLOS Biology *2*, e79.

Gómez-Gómez, L., and Boller, T. (2002). Flagellin perception: a paradigm for innate immunity. Trends in Plant Science *7*, 251-256.

González-Lamothe, R., El Oirdi, M., Brisson, N., and Bouarab, K. (2012). The conjugated auxin indole-3-acetic acid-aspartic acid promotes plant disease development. The Plant Cell *24*, 762-777.

Gonzalez, C., Szurek, B., Manceau, C., Mathieu, T., Séré, Y., and Verdier, V. (2007). Molecular and pathotypic characterization of new *Xanthomonas oryzae* strains from West Africa. Molecular Plant-Microbe Interactions *20*, 534-546.

Gou, X., He, K., Yang, H., Yuan, T., Lin, H., Clouse, S.D., and Li, J. (2010). Genome-wide cloning and sequence analysis of leucine-rich repeat receptor-like protein kinase genes in *Arabidopsis thaliana*. BMC Genomics *11*, 19.

Grau, J., Reschke, M., Erkes, A., Streubel, J., Morgan, R.D., Wilson, G.G., Koebnik, R., and Boch, J. (2016). AnnoTALE: bioinformatics tools for identification, annotation, and nomenclature of TALEs from *Xanthomonas* genomic sequences. Scientific Reports *6*, 21077.

Grau, J., Wolf, A., Reschke, M., Bonas, U., Posch, S., and Boch, J. (2013). Computational predictions provide insights into the biology of TAL effector target sites. PLOS Computational Biology *9*, e1002962.

Gu, K., Tian, D., Qiu, C., and Yin, Z. (2009). Transcription activator-like type III effector AvrXa27 depends on OsTFIIAγ5 for the activation of *Xa27* transcription in rice that triggers disease resistance to *Xanthomonas oryzae* pv. *oryzae*. Molecular Plant Pathology *10*, 829-835.

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.

Guo, A. and Leach, J.E. (1989). Examination of rice hydathode water pores exposed to *Xanthomonas campestris* pv.*oryzae*. Phytopathology *79,* 433–436.

Gürlebeck, D., Thieme, F., and Bonas, U. (2006). Type III effector proteins from the plant pathogen *Xanthomonas* and their role in the interaction with the host plant. Journal of Plant Physiology *163*, 233-255.

Hall, T.M.T. (2016). De-coding and re-coding RNA recognition by PUF and PPR repeat proteins. Current Opinion in Structural Biology *36*, 116-121.

Han, F., and Zhu, B. (2011). Evolutionary analysis of three gibberellin oxidase genesin rice, Arabidopsis, and soybean. Gene *473*, 23-35.

Hara, Y., Yokoyama, R., Osakabe, K., Toki, S., and Nishitani, K. (2014). Function of xyloglucan endotransglucosylase/hydrolases in rice. Annals of Botany *114*, 1309-1318.

Hok, S., Danchin, E.G.J., Allasia, V., PanabiÈRes, F., Attard, A., and Keller, H. (2011). An Arabidopsis (malectin-like) leucine-rich repeat receptor-like kinase contributes to downy mildew disease. Plant, Cell & Environment *34*, 1944-1957.

Hsieh, Y.S.Y., and Harris, P.J. (2009). Xyloglucans of monocotyledons have diverse structures. Molecular Plant *2*, 943-965.

Hu, H., Wang, J., Shi, C., Yuan, C., Peng, C., Yin, J., Li, W., He, M., Wang, J., Ma, B.*, et al.* (2015). A receptor like kinase gene with expressional responsiveness on *Xanthomonas oryzae* pv. *oryzae* is essential for *Xa21*-mediated disease resistance. Rice *8*, 1.

Hu, Y., Zhang, J., Jia, H., Sosso, D., Li, T., Frommer, W.B., Yang, B., White, F.F., Wang, N., and Jones, J.B. (2014). *Lateral organ boundaries 1* is a disease susceptibility gene for citrus bacterial canker disease. Proceedings of the National Academy of Sciences of the United States of America *111*, E521-E529.

Huang, N., Angeles, E.R., Domingo, J., Magpantay, G., Singh, S., Zhang, G., Kumaravadivel, N., Bennett, J., and Khush, G.S. (1997). Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. Theoretical and Applied Genetics *95*, 313-320.

Hui, S., Hao, M., Liu, H., Xiao, J., Li, X., Yuan, M., and Wang, S. (2019). The group I GH3 family genes encoding JA-Ile synthetase act as positive regulator in the resistance of rice to *Xanthomonas oryzae* pv. *oryzae*. Biochemical and Biophysical Research Communications *508*, 1062-1066.

Hummel, A.W., Doyle, E.L., and Bogdanove, A.J. (2012). Addition of transcription activator-like effector binding sites to a pathogen strain-specific rice bacterial blight resistance gene makes it effective against additional strains and against bacterial leaf streak. New Phytologist *195*, 883-893.

Hutin, M., Pérez-Quintero, A.L., Lopez, C., and Szurek, B. (2015a). MorTAL Kombat: the story of defense against TAL effectors through loss-of-susceptibility. Frontiers in Plant Science *6*, 535.

Hutin, M., Sabot, F., Ghesquière, A., Koebnik, R., and Szurek, B. (2015b). A knowledge-based molecular screen uncovers a broad-spectrum *OsSWEET14* resistance allele to bacterial blight from wild rice. The Plant Journal *84*, 694-703.

Iyer, A.S., and McCouch, S.R. (2004). The rice bacterial blight resistance gene *xa5* encodes a novel form of disease resistance. Molecular Plant-Microbe Interactions *17*, 1348-1354.

Jain, M., Tyagi, A.K., and Khurana, J.P. (2006). Genome-wide analysis, evolutionary expansion, and expression of early auxin-responsive *SAUR* gene family in rice (*Oryza sativa*). Genomics *88*, 360-371.

Jalan, N., Kumar, D., Andrade, M.O., Yu, F., Jones, J.B., Graham, J.H., White, F.F., Setubal, J.C., and Wang, N. (2013). Comparative genomic and transcriptome analyses of pathotypes of *Xanthomonas citri* subsp. *citri* provide insights into mechanisms of bacterial virulence and host range. BMC Genomics *14*, 551.

Jeger, M., Candresse, T., Chatzivassiliou, E., Dehnen-Schmutz, K., Gilioli, G., Grégoire, J.-C., Jaques Miret, J.A., MacLeod, A., Navajas Navarro, M., Niere, B*., et al.* (2018). Pest categorisation of *Xanthomonas oryzae* pathovars oryzae and *oryzicola*. EFSA Journal *16*, e05109.

Ji, Z., Ji, C., Liu, B., Zou, L., Chen, G., and Yang, B. (2016). Interfering TAL effectors of *Xanthomonas oryzae* neutralize *R*-gene-mediated plant disease resistance. Nature Communications *7*, 13435.

Jia, H., Zhang, Y., Orbović, V., Xu, J., White, F.F., Jones, J.B., and Wang, N. (2017). Genome editing of the disease susceptibility gene *CsLOB1* in citrus confers resistance to citrus canker. Plant Biotechnology Journal *15*, 817-823.

Jing, Y., and Lin, R. (2015). The VQ motif-containing protein family of plant-specific transcriptional regulators. Plant Physiology *169*, 371-378.

Jones, R.K., Barnes, L.W., Gonzalez, C.F., Leach, J.E., Alvarez, A.M., and Benedict, A.A. (1989). Identification of low-virulence strains of *Xanthomonas campestris* pv. *oryzae* from rice in the United States. Phytopathology *79*, 984-990.

Kamenidou, S., Cavins, T.J., and Marek, S. (2009). Evaluation of silicon as a nutritional supplement for greenhouse zinnia production. Scientia Horticulturae *119*, 297-301.

Katsir, L., Chung, H.S., Koo, A.J.K., and Howe, G.A. (2008). Jasmonate signaling: a conserved mechanism of hormone sensing. Current Opinion in Plant Biology *11*, 428-435.

Kay, S., Boch, J., and Bonas, U. (2005). Characterization of AvrBs3-like effectors from a *Brassicaceae* pathogen reveals virulence and avirulence activities and a protein with a novel repeat architecture. Molecular Plant-Microbe Interactions *18*, 838-848.

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

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. The Plant Journal *59*, 859-871.

Kim, J.H., Cheon, Y.M., Kim, B.G., and Ahn, J.H. (2008). Analysis of flavonoids and characterization of the *OsFNS* gene involved in flavone biosynthesis in Rice. Journal of Plant Biology *51*, 97.

Koncz, C., and Schell, J. (1986). The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. Molecular and General Genetics MGG *204*, 383-396.

Konermann, S., Brigham, M.D., Trevino, A.E., Hsu, P.D., Heidenreich, M., Le, C., Platt, R.J., Scott, D.A., Church, G.M., and Zhang, F. (2013). Optical control of mammalian endogenous transcription and epigenetic states. Nature *500*, 472-476.

Kozlov, G., Wong, K., Wang, W., Skubák, P., Muñoz-Escobar, J., Liu, Y., Siddiqui, N., Pannu, N.S., and Gehring, K. (2018). Ankyrin repeats as a dimerization module. Biochemical and Biophysical Research Communications *495*, 1002-1007.

Kwon, S.I., and Hwang, D.J. (2014). Plant-specific VQ-domain proteins as interaction partners of WRKY transcription factors. Plant Genetic Resources *12*, S121-S124.

Lam, P.Y., Zhu, F.-Y., Chan, W.L., Liu, H., and Lo, C. (2014). Cytochrome P450 93G1 is a flavone synthase II that channels flavanones to the biosynthesis of tricin *O*-linked conjugates in rice. Plant Physiology *165*, 1315-1327.

Larebeke, N.V., Engler, G., Holsters, M., Den Elsacker, S.V., Zaenen, I., Schilperoort, R.A., and Schell, J. (1974). Large plasmid in *Agrobacterium tumefaciens* essential for crown gall-inducing ability. Nature *252*, 169-170.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet *11*, 204-220.

Leach, J.E. and White, F.W. (1991). Molecular probes for disease diagnosis and monitoring. Rice Biotechnology, 281-307.

Lebar, T., and Jerala, R. (2016). Benchmarking of TALE- and CRISPR/dCas9-based transcriptional regulators in mammalian cells for the construction of synthetic genetic circuits. ACS Synthetic Biology *5*, 1050-1058.

Lee, B.-M., Park, Y.-J., Park, D.-S., Kang, H.-W., Kim, J.-G., Song, E.-S., Park, I.-C., Yoon, U.-H., Hahn, J.-H., Koo, B.-S.*, et al.* (2005). The genome sequence of *Xanthomonas oryzae* pathovar *oryzae* KACC10331, the bacterial blight pathogen of rice. Nucleic Acids Research *33*, 577-586.

Lee, I., Seo, Y.-S., Coltrane, D., Hwang, S., Oh, T., Marcotte, E.M., and Ronald, P.C. (2011). Genetic dissection of the biotic stress response using a genome-scale gene network for rice. Proceedings of the National Academy of Sciences of the United States of America *108*, 18548-18553.

Li, R.-Y., Ago, Y., Liu, W.-J., Mitani, N., Feldmann, J., McGrath, S.P., Ma, J.F., and Zhao, F.-J. (2009). The rice aquaporin Lsi1 mediates uptake of methylated arsenic species. Plant Physiology *150*, 2071-2080.

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

Li, T., Liu, B., Spalding, M.H., Weeks, D.P., and Yang, B. (2012). High-efficiency TALEN-based gene editing produces disease-resistant rice. Nature Biotechnology *30*, 390-392.

Liao, D., Chen, X., Chen, A., Wang, H., Liu, J., Liu, J., Gu, M., Sun, S., and Xu, G. (2015). The characterization of six auxin-induced tomato *GH3* genes uncovers a member, *SlGH3.4*, strongly responsive to arbuscular mycorrhizal symbiosis. Plant and Cell Physiology *56*, 674-687.

Liu, P.-L., Du, L., Huang, Y., Gao, S.-M., and Yu, M. (2017). Origin and diversification of leucine-rich repeat receptor-like protein kinase (*LRR-RLK*) genes in plants. BMC Evolutionary Biology *17*, 47.

Liu, Q., Yuan, M., Zhou, Y., Li, X., Xiao, J., and Wang, S. (2011). A paralog of the MtN3/saliva family recessively confers race-specific resistance to *Xanthomonas oryzae* in rice. Plant, Cell & Environment *34*, 1958-1969.

Liu, W., Liu, J., Triplett, L., Leach, J.E., and Wang, G.-L. (2014). Novel insights into rice innate immunity against bacterial and fungal pathogens. Annual Review of Phytopathology *52*, 213-241.

Lu, G., and Hall, T.M.T. (2011). Alternate modes of cognate RNA recognition by human PUMILIO proteins. Structure *19*, 361-367.

Lu, X., Hershey, D.M., Wang, L., Bogdanove, A.J., and Peters, R.J. (2015). An *ent*-kaurene-derived diterpenoid virulence factor from *Xanthomonas oryzae* pv. *oryzicola*. New Phytologist *206*, 295-302.

Ma, H., Reyes-Gutierrez, P., and Pederson, T. (2013). Visualization of repetitive DNA sequences in human chromosomes with transcription activator-like effectors. Proceedings of the National Academy of Sciences of the United States of America *110*, 21048-21053.

Ma, J.F., Tamai, K., Yamaji, N., Mitani, N., Konishi, S., Katsuhara, M., Ishiguro, M., Murata, Y., and Yano, M. (2006). A silicon transporter in rice. Nature *440*, 688.

Ma, J.F., and Yamaji, N. (2006). Silicon uptake and accumulation in higher plants. Trends in Plant Science *11*, 392-397.

Ma, J.F., Yamaji, N., and Mitani-Ueno, N. (2011). Transport of silicon from roots to panicles in plants. Proceedings of the Japan Academy, Series B, Physical and Biological Sciences *87*, 377-385.

Ma, J.F., Yamaji, N., Mitani, N., Xu, X.-Y., Su, Y.-H., McGrath, S.P., and Zhao, F.-J. (2008). Transporters of arsenite in rice and their role in arsenic accumulation in rice grain. Proceedings of the National Academy of Sciences of the United States of America *105*, 9931-9935.

Ma, W., Zou, L., Zhiyuan, J., Xiameng, X., Zhengyin, X., Yang, Y., Alfano, J.R., and Chen, G. (2018). *Xanthomonas oryzae* pv. *oryzae* TALE proteins recruit OsTFIIAγ1 to compensate for the absence of OsTFIIAγ5 in bacterial blight in rice. Molecular Plant Pathology *19*, 2248-2262.

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

Mak, A.N.-S., Bradley, P., Cernadas, R.A., Bogdanove, A.J., and Stoddard, B.L. (2012). The crystal structure of TAL effector PthXo1 bound to its DNA target. Science *335*, 716-719.

Makino, S., Sugio, A., White, F., and Bogdanove, A.J. (2006). Inhibition of resistance gene-mediated defense in rice by *Xanthomonas oryzae* pv. *oryzicola*. Molecular Plant-Microbe Interactions *19*, 240-249.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church G.M. (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nature Biotechnology *31*, 833-838.

Manna, S. (2015). An overview of pentatricopeptide repeat proteins and their applications. Biochimie *113*, 93-99.

Markmann, K., Giczey, G., and Parniske, M. (2008). Functional adaptation of a plant receptor-kinase paved the way for the evolution of intracellular root symbioses with bacteria. PLOS Biology *6*, e68.

Marois, E., Van den Ackerveken, G., and Bonas, U. (2002). The *Xanthomonas* type III effector protein AvrBs3 modulates plant gene expression and induces cell hypertrophy in the susceptible host. Molecular Plant-Microbe Interactions *15*, 637-646.

Maurer, S., Giess, M., Koch, O., and Summerer, D. (2016). Interrogating key positions of size-reduced TALE repeats reveals a programmable sensor of 5-Carboxylcytosine. ACS Chemical Biology *11*, 3294-3299.

McClure, B., and Guilfoyle, T. (1989). Rapid redistribution of auxin-regulated RNAs during gravitropism. Science *243*, 91-93.

Meckler, J.F., Bhakta, M.S., Kim, M.-S., Ovadia, R., Habrian, C.H., Zykovich, A., Yu, A., Lockwood, S.H., Morbitzer, R., Elsäesser, J.*, et al.* (2013). Quantitative analysis of TALE–DNA interactions suggests polarity effects. Nucleic Acids Research *41*, 4118-4128.

Mendenhall, E.M., Williamson, K.E., Reyon, D., Zou, J.Y., Ram, O., Joung, J.K., and Bernstein, B.E. (2013). Locus-specific editing of histone modifications at endogenous enhancers. Nature Biotechnology *31*, 1133-1136.

Mew, T.W., Alvarez, A.M., Leach, J.E., and Swings, J. (1993). Focus on bacterial blight of rice. Plant Disease *77*, 5-12.

Mew, T.W., Mew, I.C., and Haung, J.S. (1984). Scanning of microscopy of virulent and avirulent strains of *Xanthomonas campestris* pv. *campestris* on rice leaves. Phytopathology *74*, 635-641.

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. Nature Biotechnology *29*, 143-148.

Miller, J.C., Zhang, L., Xia, D.F., Campo, J.J., Ankoudinova, I.V., Guschin, D.Y., Babiarz, J.E., Meng, X., Hinkley, S.J., Lam, S.C.*, et al.* (2015). Improved specificity of TALE-based genome editing using an expanded RVD repertoire. Nature Methods *12*, 465-471.

Miranda, R.G., McDermott, J.J., and Barkan, A. (2018). RNA-binding specificity landscapes of designer pentatricopeptide repeat proteins elucidate principles of PPR-RNA interactions. Nucleic Acids Research *46*, 2613-2623.

Mitani, N., Yamaji, N., and Ma, J.F. (2008). Characterization of substrate specificity of a rice silicon transporter, Lsi1. Pflügers Archiv - European Journal of Physiology *456*, 679-686.

Moore, R., Chandrahas, A., and Bleris, L. (2014). Transcription activator-like effectors: a toolkit for synthetic biology. ACS Synthetic Biology *3*, 708-716.

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

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. Proceedings of the National Academy of Sciences of the United States of America *107*, 21617-21622.

Mosavi, L.K., Cammett, T.J., Desrosiers, D.C., and Peng, Z.-Y. (2004). The ankyrin repeat as molecular architecture for protein recognition. Protein Science *13*, 1435-1448.

Moscou, M.J., and Bogdanove, A.J. (2009). A simple cipher governs DNA recognition by TAL effectors. Science *326*, 1501.

Mosher, S., Seybold, H., Rodriguez, P., Stahl, M., Davies, K.A., Dayaratne, S., Morillo, S.A., Wierzba, M., Favery, B., Keller, H.*, et al.* (2013). The tyrosine-sulfated peptide receptors PSKR1 and PSY1R modify the immunity of Arabidopsis to biotrophic and necrotrophic pathogens in an antagonistic manner. The Plant Journal *73*, 469-482.

Mücke, S., Reschke, M., Erkes, A., Schwietzer, C.-A., Becker, S., Streubel, J., Morgan, R.D., Wilson, G.G., Grau, J., and Boch, J. (2019). Transcriptional reprogramming of rice cells by *Xanthomonas oryzae* TALEs. Frontiers in Plant Science *10*, 162.

Nagel, R., and Peters, R.J. (2017). Investigating the phylogenetic range of gibberellin biosynthesis in bacteria. Molecular Plant-Microbe Interactions *30*, 343-349.

Nagel, R., Turrini, P.C.G., Nett, R.S., Leach, J.E., Verdier, V., Van Sluys, M.-A., and Peters, R.J. (2017). An operon for production of bioactive gibberellin A4 phytohormone with wide distribution in the bacterial rice leaf streak pathogen *Xanthomonas oryzae* pv. *oryzicola*. New Phytologist *214*, 1260-1266.

Ning, D., Song, A., Fan, F., Li, Z., and Liang, Y. (2014). Effects of slag-based silicon fertilizer on rice growth and brown-spot resistance. PLOS ONE *9*, e102681.

Niño-Liu, D.O., Ronald, P.C., and Bogdanove, A.J. (2006). *Xanthomonas oryzae* pathovars: Model pathogens of a model crop. Mol Plant Pathol *7*, 303-324.

Nobuta, K., Okrent, R.A., Stoutemyer, M., Rodibaugh, N., Kempema, L., Wildermuth, M.C., and Innes, R.W. (2007). The GH3 acyl adenylase family member PBS3 regulates salicylic acid-dependent defense responses in Arabidopsis. Plant Physiology *144*, 1144-1156.

Oikawa, T., Koshioka, M., Kojima, K., Yoshida, H., and Kawata, M. (2004). A role of *OsGA20ox1*, encoding an isoform of gibberellin 20-oxidase, for regulation of plant stature in rice. Plant Molecular Biology *55*, 687-700.

Opperman, L., Hook, B., DeFino, M., Bernstein, D.S., and Wickens, M. (2005). A single spacer nucleotide determines the specificities of two mRNA regulatory proteins. Nature Structural & Molecular Biology *12*, 945-951.

Padmanaban, S., Chanroj, S., Kwak, J.M., Li, X., Ward, J.M., and Sze, H. (2007). Participation of endomembrane cation/H^+ exchanger AtCHX20 in osmoregulation of guard cells. Plant Physiology *144*, 82-93.

Padmanaban, S., Czerny, D.D., Levin, K.A., Leydon, A.R., Su, R.T., Maugel, T.K., Zou, Y., Chanroj, S., Cheung, A.Y., Johnson, M.A.*, et al.* (2017). Transporters involved in pH and K^+ homeostasis affect pollen wall formation, male fertility, and embryo development. Journal of Experimental Botany *68*, 3165-3178.

Peng, H., Zhang, Q., Li, Y., Lei, C., Zhai, Y., Sun, X., Sun, D., Sun, Y., and Lu, T. (2009). A putative leucine-rich repeat receptor kinase, OsBRR1, is involved in rice blast resistance. Planta *230*, 377-385.

Pereira, A.L.A., Carazzolle, M.F., Abe, V.Y., de Oliveira, M.L.P., Domingues, M.N., Silva, J.C., Cernadas, R.A., and Benedetti, C.E. (2014). Identification of putative TAL effector targets of the citrus canker pathogens shows functional convergence underlying disease development and defense response. BMC Genomics *15*, 157.

Pérez-Quintero, A.L., Lamy, L., Gordon, J.L., Escalon, A., Cunnac, S., Szurek, B., and Gagnevin, L. (2015). QueTAL: a suite of tools to classify and compare TAL effectors functionally and phylogenetically. Frontiers in Plant Science *6*, 545.

Pérez-Quintero, A.L., Rodriguez-R, L.M., Dereeper, A., López, C., Koebnik, R., Szurek, B., and Cunnac, S. (2013). An improved method for TAL effectors DNA-binding sites prediction reveals functional convergence in TAL repertoires of *Xanthomonas oryzae* strains. PLOS ONE *8*, e68464.

Politz, M.C., Copeland, M.F., and Pfleger, B.F. (2013). Artificial repressors for controlling gene expression in bacteria. Chemical Communications *49*, 4325-4327.

Pruitt, R.N., Schwessinger, B., Joe, A., Thomas, N., Liu, F., Albert, M., Robinson, M.R., Chan, L.J.G., Luu, D.D., Chen, H.*, et al.* (2015). The rice immune receptor XA21 recognizes a tyrosine-sulfated protein from a Gram-negative bacterium. Science Advances *1*, e1500245.

Qin, X., Liu, J.H., Zhao, W.S., Chen, X.J., Guo, Z.J., and Peng, Y.L. (2012). Gibberellin 20-oxidase gene *OsGA20ox3* regulates plant stature and disease development in rice. Molecular Plant-Microbe Interactions *26*, 227-239.

Qin, X., Warguchuk, R., Arnal, N., Gaborieau, L., Mireau, H., and Brown, G.G. (2014). *In vivo* functional analysis of a nuclear restorer PPR protein. BMC Plant Biology *14*, 313.

Quibod, I.L., Perez-Quintero, A., Booher, N.J., Dossa, G.S., Grande, G., Szurek, B., Vera Cruz, C., Bogdanove, A.J., and Oliva, R. (2016). Effector diversification contributes to *Xanthomonas oryzae* pv. *oryzae* phenotypic adaptation in a semi-isolated environment. Scientific Reports *6*, 34137.

Rademaker, J.L., Hoste, B., Louws, F.J., Kersters, K., Swings, J., Vauterin, L., Vauterin, P., and de Bruijn, F.J. (2000). Comparison of AFLP and rep-PCR genomic fingerprinting with DNA-DNA homology studies: *Xanthomonas* as a model system. International Journal of Systematic and Evolutionary Microbiology *50*, 665-677.

Rathi, P., Maurer, S., Kubik, G., and Summerer, D. (2016). Isolation of human genomic DNA sequences with expanded nucleobase selectivity. Journal of the American Chemical Society *138*, 9910-9918.

Rathi, P., Maurer, S., and Summerer, D. (2018). Selective recognition of *N*4-methylcytosine in DNA by engineered transcription-activator-like effectors. Philosophical Transactions of the Royal Society B: Biological Sciences *373*, 20170078.

Raymundo, A.K., Briones, A.M. Jr., Ardales, E.Y., Perez, M.T., Fernandez, L.C., Leach, J.F., Mew, T.W., Ynalvez, M.A., McLaren, C.G., and Nelson, R.J. (1999). Analysis of DNA polymorphism and virulence in Philippine strains of *Xanthomonas oryzae* pv. *oryzicola*. Plant Disease *83*, 434-440.

Read, A.C., Rinaldi, F.C., Hutin, M., He, Y.-Q., Triplett, L.R., and Bogdanove, A.J. (2016). Suppression of *Xo1*-mediated disease resistance in rice by a truncated, non-DNA-binding TAL effector of *Xanthomonas oryzae*. Frontiers in Plant Science *7*, 1516.

Reddy, A.K.P., Catyal, J.C., Rouse, D.I., and MacKenzie, D.R. (1979). Relationship between nitrogen fertilization, bacterial leaf blight severity, and yield of rice. Phytophathology *69*, 970-973.

Ren, H., and Gray, William M. (2015). SAUR proteins as effectors of hormonal and environmental signals in plant growth. Molecular Plant *8*, 1153-1164.

Ren, R., Deng, L., Xue, Y., Suzuki, K., Zhang, W., Yu, Y., Wu, J., Sun, L., Gong, X., Luan, H.*, et al.* (2017). Visualization of aging-associated chromatin alterations with an engineered TALE system. Cell Research *27*, 483-504.

Richter, A., Streubel, J., Blücher, C., Szurek, B., Reschke, M., Grau, J., and Boch, J. (2014). A TAL effector repeat architecture for frameshift binding. Nature Communications *5*, 3447.

Riera, J., and Lazo, P.S. (2009). The mammalian *NudC*-like genes: a family with functions other than regulating nuclear distribution. Cellular and Molecular Life Sciences *66*, 2383-2390.

Robert-Seilaniantz, A., Navarro, L., Bari, R., and Jones, J.D.G. (2007). Pathological hormone imbalances. Current Opinion in Plant Biology *10*, 372-379.

Rodrigues, F.Á., Benhamou, N., Datnoff, L.E., Jones, J.B., and Bélanger, R.R. (2003a). Ultrastructural and cytochemical aspects of silicon-mediated rice blast resistance. Phytopathology *93*, 535-546.

Rodrigues, F.Á., Vale, F.X.R., Datnoff, L.E., Prabhu, A.S., and Korndörfer, G.H. (2003b). Effect of rice growth stages and silicon on sheath blight development. Phytopathology *93*, 256-261.

Römer, P., Recht, S., and Lahaye, T. (2009a). A single plant resistance gene promoter engineered to recognize multiple TAL effectors from disparate pathogens. Proceedings of the National Academy of Sciences of the United States of America *106*, 20526-20531.

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

Rouxel, T. and Balesdent, M. H. (2010). Avirulence genes. Encyclopedia of Life Sciences, doi: 10.1002/9780470015902.a0021267.

Salazar-Cerezo, S., Martínez-Montiel, N., García-Sánchez, J., Pérez-y-Terrón, R., and Martínez-Contreras, R.D. (2018). Gibberellin biosynthesis and metabolism: A convergent route for plants, fungi and bacteria. Microbiological Research *208*, 85-98.

Salzberg, S.L., Sommer, D.D., Schatz, M.C., Phillippy, A.M., Rabinowicz, P.D., Tsuge, S., Furutani, A., Ochiai, H., Delcher, A.L., Kelley, D.*, et al.* (2008). Genome sequence and rapid evolution of the rice pathogen *Xanthomonas oryzae* pv. *oryzae* PXO99(A). BMC Genomics *9*, 204-204.

Savant, N.K., Snyder, G.H., and Datnoff, L.E. (1996). Silicon management and sustainable rice production. In Advances in Agronomy *58*, 151-199.

Schandry, N., Jacobs, J.M., Szurek, B., and Perez-Quintero, A.L. (2018). A cautionary TALE: how plant breeding may have favoured expanded TALE repertoires in *Xanthomonas*. Molecular Plant Pathology *19*, 1297-1301.

Scharenberg, A.M., Duchateau, P., and Smith, J. (2013). Genome engineering with TAL-effector nucleases and alternative modular nuclease technologies. Current Gene Therapy *13*, 291-303.

Schornack, S., Ballvora, A., Gürlebeck, D., Peart, J., Ganal, M., Baker, B., Bonas, U., and Lahaye, T. (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. The Plant Journal *37*, 46-60.

Schornack, S., Minsavage, G.V., Stall, R.E., Jones, J.B., and Lahaye, T. (2008). Characterization of AvrHah1, a novel AvrBs3-like effector from *Xanthomonas gardneri* with virulence and avirulence activity. New Phytologist *179*, 546-556.

Schornack, S., Moscou, M.J., Ward, E.R., and Horvath, D.M. (2013). Engineering plant disease resistance based on TAL effectors. Annual Review of Phytopathology *51*, 383-406.

Schwartz, A.R., Morbitzer, R., Lahaye, T., and Staskawicz, B.J. (2017). TALE-induced bHLH transcription factors that activate a pectate lyase contribute to water soaking in bacterial spot of tomato. Proceedings of the National Academy of Sciences of the United States of America *114*, E897-E903.

Shiu, S.-H., Karlowski, W.M., Pan, R., Tzeng, Y.-H., Mayer, K.F.X., and Li, W.-H. (2004). Comparative analysis of the receptor-like kinase family in Arabidopsis and rice. The Plant Cell *16*, 1220-1234.

Siemens, J., Keller, I., Sarx, J., Kunz, S., Schuller, A., Nagel, W., Schmülling, T., Parniske, M., and Ludwig-Müller, J. (2006). Transcriptome analysis of Arabidopsis clubroots indicate a key role for cytokinins in disease development. Molecular Plant-Microbe Interactions *19*, 480-494.

Singer, A.U., Schulze, S., Skarina, T., Xu, X., Cui, H., Eschen-Lippold, L., Egler, M., Srikumar, T., Raught, B., Lee, J.*, et al.* (2013). A pathogen type III effector with a novel E3 ubiquitin ligase architecture. PLOS Pathogens *9*, e1003121.

Song, A., Xue, G., Cui, P., Fan, F., Liu, H., Yin, C., Sun, W., and Liang, Y. (2016). The role of silicon in enhancing resistance to bacterial blight of hydroponic- and soil-cultured rice. Scientific Reports *6*, 24640.

Sonnewald, U. (2011). SWEETs – the missing sugar efflux carriers. Frontiers in Plant Science *2*, 7.

Spielmeyer, W., Ellis, M.H., and Chandler, P.M. (2002). Semidwarf (*sd-1*), "green revolution" rice, contains a defective gibberellin 20-oxidase gene. Proceedings of the National Academy of Sciences of the United States of America *99*, 9043-9048.

Staswick, P.E., Serban, B., Rowe, M., Tiryaki, I., Maldonado, M.T., Maldonado, M.C., and Suza, W. (2005). Characterization of an Arabidopsis enzyme family that conjugates amino acids to indole-3-acetic acid. The Plant Cell *17*, 616-627.

Staswick, P.E., and Tiryaki, I. (2004). The oxylipin signal jasmonic acid is activated by an enzyme that conjugates it to isoleucine in Arabidopsis. The Plant Cell *16*, 2117-2127.

Stortenbeker, N., and Bemer, M. (2019). The *SAUR* gene family: the plant's toolbox for adaptation of growth and development. Journal of Experimental Botany *70*, 17-27.

Streubel, J., Blücher, C., Landgraf, A., and Boch, J. (2012). TAL effector RVD specificities and efficiencies. Nature Biotechnology *30*, 593-595.

Streubel, J., Pesce, C., Hutin, M., Koebnik, R., Boch, J., and Szurek, B. (2013). Five phylogenetically close rice *SWEET* genes confer TAL effector-mediated susceptibility to *Xanthomonas oryzae* pv. *oryzae*. New Phytologist *200*, 808-819.

Sugio, A., Yang, B., Zhu, T., and White, F.F. (2007). Two type III effector genes of *Xanthomonas oryzae* pv. *oryzae* control the induction of the host genes *OsTFIIAγ1* and *OsTFX1* during bacterial blight of rice. Proceedings of the National Academy of Sciences of the United States of America *104*, 10720-10725.

Suh, J.-P., Jeung, J.-U., Noh, T.-H., Cho, Y.-C., Park, S.-H., Park, H.-S., Shin, M.-S., Kim, C.-K., and Jena, K.K. (2013). Development of breeding lines with three pyramided resistance genes that confer broad-spectrum bacterial blight resistance and their molecular analysis in rice. Rice *6*, 5.

Sun, W., Zhang, J., Fan, Q., Xue, G., Li, Z., and Liang, Y. (2010). Silicon-enhanced resistance to rice blast is attributed to silicon-mediated defence resistance and its role as physical barrier. European Journal of Plant Pathology *128*, 39-49.

Swings, J., Van Den Mooter, M., Vauterin, L., Hoste, B., Gillis, M., Mew, T.W., and Kersters, K. (1990). Reclassification of the causal agents of bacterial blight (*Xanthomonas campestris* pv. *oryzae*) and bacterial leaf streak (*Xanthomonas campestris* pv. *oryzicola*) of rice as pathovars of *Xanthomonas oryzae* (ex Ishiyama 1922) sp. nov., nom. rev. International Journal of Systematic and Evolutionary Microbiology *40*, 309-311.

Sze, H., Padmanaban, S., Cellier, F., Honys, D., Cheng, N.-H., Bock, K.W., Conéjéro, G., Li, X., Twell, D., Ward, J.M.*, et al.* (2004). Expression patterns of a novel *AtCHX* gene family highlight potential roles in osmotic adjustment and K^+ homeostasis in pollen development. Plant Physiology *136*, 2532-2547.

Szurek, B., Marois, E., Bonas, U., and Van den Ackerveken, G. (2001). Eukaryotic features of the *Xanthomonas* type III effector AvrBs3: protein domains involved in transcriptional activation and the interaction with nuclear import receptors from pepper. The Plant Journal *26*, 523-534.

Tabei, H. (1967). Anatomical studies of rice plant affected with bacterial leaf blight. Annals of the Phytopathological Society of Japan *33*, 12-16.

Thomas, M.C., and Chiang, C.-M. (2006). The general transcription machinery and general cofactors. Critical Reviews in Biochemistry and Molecular Biology *41*, 105-178.

Tian, D., Wang, J., Zeng, X., Gu, K., Qiu, C., Yang, X., Zhou, Z., Goh, M., Luo, Y., Murata-Hori, M.*, et al.* (2014). The rice TAL effector-dependent resistance protein XA10 triggers cell death and calcium depletion in the endoplasmic reticulum. Plant Cell *26*, 497-515.

Tkaczuk, K.L., Obarska, A., and Bujnicki, J.M. (2006). Molecular phylogenetics and comparative modeling of HEN1, a methyltransferase involved in plant microRNA biogenesis. BMC Evolutionary Biology *6*, 6.

Tör, M., Lotze, M.T., and Holton, N. (2009). Receptor-mediated signalling in plants: molecular patterns and programmes. Journal of Experimental Botany *60*, 3645-3654.

Tran, T.T., Pérez-Quintero, A.L., Wonni, I., Carpenter, S.C.D., Yu, Y., Wang, L., Leach, J.E., Verdier, V., Cunnac, S., Bogdanove, A.J.*, et al.* (2018). Functional analysis of African *Xanthomonas oryzae* pv. *oryzae* TALomes reveals a new susceptibility gene in bacterial leaf blight of rice. PLOS Pathogens *14*, e1007092.

Triplett, L.R., Hamilton, J.P., Buell, C.R., Tisserat, N.A., Verdier, V., Zink, F., and Leach, J.E. (2011). Genomic analysis of *Xanthomonas oryzae* isolates from rice grown in the United States reveals substantial divergence from known *X. oryzae* pathovars. Applied and Environmental Microbiology *77*, 3930-3937.

Tsuji, S., Futaki, S., and Imanishi, M. (2013). Creating a TALE protein with unbiased 5′-T binding. Biochemical and Biophysical Research Communications *441*, 262-265.

Tsuji, S., Futaki, S., and Imanishi, M. (2016). Sequence-specific recognition of methylated DNA by an engineered transcription activator-like effector protein. Chemical Communications *52*, 14238-14241.

Üstün, S., and Börnke, F. (2014). Interactions of *Xanthomonas* type-III effector proteins with the plant ubiquitin and ubiquitin-like pathways. Frontiers in Plant Science *5*, 736.

Valley, C.T., Porter, D.F., Qiu, C., Campbell, Z.T., Hall, T.M.T., and Wickens, M. (2012). Patterns and plasticity in RNA-protein interactions enable recruitment of multiple proteins through a single site. Proceedings of the National Academy of Sciences of the United States of America *109*, 6054-6059.

Valton, J., Dupuy, A., Daboussi, F., Thomas, S., Marechal, A., Macmaster, R., Melliand, K., Juillerat, A., and Duchateau, P. (2012). Overcoming TALE DNA binding domain sensitivity to cytosine methylation. Journal of Biological Chemistry *287*, 38427-38432.

Van Damme, M., Huibers, R.P., Elberse, J., and Van den Ackerveken, G. (2008). Arabidopsis *DMR6* encodes a putative 2OG-Fe(II) oxygenase that is defense-associated but required for susceptibility to downy mildew. The Plant Journal *54*, 785-793.

Van den Ackerveken, G., Marois, E., and Bonas, U. (1996). Recognition of the bacterial avirulence protein AvrBs3 occurs inside the host plant cell. Cell *87*, 1307-1316.

Vera Cruz, C.M., Bai, J., Oña, I., Leung, H., Nelson, R.J., Mew, T.-W., and Leach, J.E. (2000). Predicting durability of a disease resistance gene based on an assessment of the fitness loss and epidemiological consequences of avirulence gene mutation. Proceedings of the National Academy of Sciences of the United States of America *97*, 13500-13505.

Vera Cruz, C.M., Cottyn, B., Nguyen, M.H., Lang, J., Verdier, V., Mew, T.W., and Leach J.E. (2017). Detection of *Xanthomonas oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola* in rice seeds. Detection of Plant-Pathogenic Bacteria in Seed and Other Planting Material *2nd Edition*, 45-55.

Vera Cruz, C.M., Gosselé, F., Kersters, K., Segers, P., Van Den Mooter, M., Swings, J., and De Ley, J. (1984). Differentiation between *Xanthomonas campestris* pv. *oryzae*, *Xanthomonas campestris* pv. *oryzicola* and the bacterial 'brown blotch' pathogen on rice by numerical analysis of phenotypic features and protein gel electrophoregrams. Microbiology *130*, 2983-2999.

Verdier, V., Triplett, L.R., Hummel, A.W., Corral, R., Cernadas, R.A., Schmidt, C.L., Bogdanove, A.J., and Leach, J.E. (2012). Transcription activator-like (TAL) effectors targeting *OsSWEET* genes enhance virulence on diverse rice (*Oryza sativa*) varieties when expressed individually in a TAL effector-deficient strain of *Xanthomonas oryzae*. New Phytologist *196*, 1197-1207.

Vogel, J. (2008). Unique aspects of the grass cell wall. Current Opinion in Plant Biology *11*, 301-307.

Wang, C., Fan, Y., Zheng, C., Qin, T., Zhang, X., and Zhao, K. (2014). High-resolution genetic mapping of rice bacterial blight resistance gene *Xa23*. Molecular Genetics and Genomics *289*, 745-753.

Wang, C., Zhang, X., Fan, Y., Gao, Y., Zhu, Q., Zheng, C., Qin, T., Li, Y., Che, J., Zhang, M*., et al.* (2015). XA23 is an executor R protein and confers broad-spectrum disease resistance in rice. Mol Plant *8*, 290-302.

Wang, J., Tian, D., Gu, K., Yang, X., Wang, L., Zeng, X., and Yin, Z. (2017a). Induction of *Xa10*-like genes in rice cultivar Nipponbare confers disease resistance to rice bacterial blight. Molecular Plant-Microbe Interactions *30*, 466-477.

Wang, L., Rinaldi, F.C., Singh, P., Doyle, E.L., Dubrow, Z.E., Tran, T.T., Pérez-Quintero, A.L., Szurek, B., and Bogdanove, A.J. (2017b). TAL effectors drive transcription bidirectionally in plants. Molecular Plant *10*, 285-296.

Wang, M., Ogé, L., Perez-Garcia, M.-D., Hamama, L., and Sakr, S. (2018). The PUF protein family: overview on PUF RNA targets, biological functions, and post transcriptional regulation. International Journal of Molecular Sciences *19*, 410.

Wang, Y., Opperman, L., Wickens, M., and Hall, T.M.T. (2009). Structural basis for specific recognition of multiple mRNA targets by a PUF regulatory protein. Proceedings of the National Academy of Sciences of the United States of America *106*, 20186-20191.

Wang, Y., Wang, Z., and Tanaka Hall, T.M. (2013). Engineered proteins with Pumilio/*fem-3* mRNA binding factor scaffold to manipulate RNA metabolism. The FEBS journal *280*, 3755-3767.

Weber, E., Gruetzner, R., Werner, S., Engler, C., and Marillonnet, S. (2011). Assembly of designer TAL effectors by Golden Gate cloning. PLOS ONE *6*, e19722.

White, E.K., Moore-Jarrett, T., and Ruley, H.E. (2001). PUM2, a novel murine puf protein, and its consensus RNA-binding site. RNA *7*, 1855-1866.

Wicky, B.I.M., Stenta, M., and Dal Peraro, M. (2013). TAL effectors specificity stems from negative discrimination. PLOS ONE *8*, e80261.

Wilkins, K.E., Booher, N.J., Wang, L., and Bogdanove, A.J. (2015). TAL effectors and activation of predicted host targets distinguish Asian from African strains of the rice pathogen *Xanthomonas oryzae* pv. *oryzicola* while strict conservation suggests universal importance of five TAL effectors. Frontiers in Plant Science *6*, 536.

Xie G, Wang H., Chen, G.A., and Ye, Y.W. (1991). A study of the ingress of *Xanthomonas oryzae* pv. *oryzicola* to rice plant. Zhiwu Jianyi *5*, 1-4.

Yamaji, N., and Ma, J.F. (2007). Spatial distribution and temporal variation of the rice silicon transporter Lsi1. Plant Physiology *143*, 1306-1313.

Yang, B., Sugio, A., and White, F.F. (2006). *Os8N3* is a host disease-susceptibility gene for bacterial blight of rice. Proceedings of the National Academy of Sciences of the United States of America *103*, 10503-10508.

Yang, B., and White, F.F. (2004). Diverse members of the AvrBs3/PthA family of type III effectors are major virulence determinants in bacterial blight disease of rice. Molecular Plant-Microbe Interactions *17*, 1192-1200.

Yang, J., Zhang, Y., Yuan, P., Zhou, Y., Cai, C., Ren, Q., Wen, D., Chu, C., Qi, H., and Wei, W. (2014). Complete decoding of TAL effectors for DNA recognition. Cell Research *24*, 628-631.

Yokoyama, R., Rose, J.K.C., and Nishitani, K. (2004). A surprising diversity and abundance of xyloglucan endotransglucosylase/hydrolases in rice. Classification and expression analysis. Plant Physiology *134*, 1088-1099.

Yoshimura, S.H., and Hirano, T. (2016). HEAT repeats – versatile arrays of amphiphilic helices working in crowded environments? Journal of Cell Science *129*, 3963-3970.

Yu, Y.-H., Lu, Y., He, Y.-Q., Huang, S., and Tang, J.-L. (2015). Rapid and efficient genome-wide characterization of *Xanthomonas* TAL effector genes. Scientific Reports *5*, 13162.

Yu, Y., Streubel, J., Balzergue, S., Champion, A., Boch, J., Koebnik, R., Feng, J., Verdier, V., and Szurek, B. (2011). Colonization of rice leaf blades by an African strain of *Xanthomonas oryzae* pv. *oryzae* depends on a new TAL effector that induces the rice nodulin-3 *Os11N3* gene. Molecular Plant-Microbe Interactions *24*, 1102-1113.

Yuan, M., Ke, Y., Huang, R., Ma, L., Yang, Z., Chu, Z., Xiao, J., Li, X., and Wang, S. (2016). A host basal transcription factor is a key component for infection of rice by TALE-carrying bacteria. eLife *5*, e19605.

Zaka, A., Grande, G., Coronejo, T., Quibod, I.L., Chen, C.-W., Chang, S.-J., Szurek, B., Arif, M., Cruz, C.V., and Oliva, R. (2018). Natural variations in the promoter of *OsSWEET13* and *OsSWEET14* expand the range of resistance against *Xanthomonas oryzae* pv. *oryzae*. PLOS ONE *13*, e0203711.

Zeilmaker, T., Ludwig, N.R., Elberse, J., Seidl, M.F., Berke, L., Van Doorn, A., Schuurink, R.C., Snel, B., and Van den Ackerveken, G. (2015). DOWNY MILDEW RESISTANT 6 and DMR6-LIKE OXYGENASE 1 are partially redundant but distinct suppressors of immunity in Arabidopsis. The Plant Journal *81*, 210-222.

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G., and Arlotta, P. (2011a). Programmable sequence-specific transcriptional regulation of mammalian genome using designer TAL effectors. Nature Biotechnology *29*, 149-153.

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

Zhang, J., Yin, Z., and White, F. (2015). TAL effectors and the executor *R* genes. Frontiers in Plant Science *6*, 641.

Zhang, S.-W., Li, C.-H., Cao, J., Zhang, Y.-C., Zhang, S.-Q., Xia, Y.-F., Sun, D.-Y., and Sun, Y. (2009). Altered architecture and enhanced drought tolerance in rice via the down-regulation of indole-3-acetic acid by *TLD1/OsGH3.13* activation. Plant Physiology *151*, 1889-1901.

Zhang, X.-B., Feng, B.-H., Wang, H.-M., Xu, X., Shi, Y.-F., He, Y., Chen, Z., Sathe, A.P., Shi, L., and Wu, J.-L. (2018). A substitution mutation in *OsPELOTA* confers bacterial blight resistance by activating the salicylic acid pathway. Journal of Integrative Plant Biology *60*, 160-172.

Zhang, Y., Zhao, L., Zhao, J., Li, Y., Wang, J., Guo, R., Gan, S., Liu, C.-J., and Zhang, K. (2017). *S5H/DMR6* encodes a salicylic acid 5-hydroxylase that fine-tunes salicylic acid homeostasis. Plant Physiology *175*, 1082-1093.

Zhang, Z., Li, Q., Li, Z., Staswick, P.E., Wang, M., Zhu, Y., and He, Z. (2007). Dual regulation role of GH3.5 in salicylic acid and auxin signaling during Arabidopsis-*Pseudomonas syringae* interaction. Plant Physiology *145*, 450-464.

Zhao, B., Lin, X., Poland, J., Trick, H., Leach, J., and Hulbert, S. (2005). A maize resistance gene functions against bacterial streak disease in rice. Proceedings of the National Academy of Sciences of the United States of America *102*, 15383-15388.

Zhao, J., Fu, J., Li, X., Xu, C., and Wang, S. (2009). Dissection of the factors affecting development-controlled and race-specific disease resistance conferred by leucine-rich repeat receptor kinase-type *R* genes in rice. Theoretical and Applied Genetics *119*, 231-239.

Zhao, X.Q., Mitani, N., Yamaji, N., Shen, R.F., and Ma, J.F. (2010). Involvement of silicon influx transporter OsNIP2;1 in selenite uptake in rice. Plant Physiology *153*, 1871-1877.

Zhou, J., Peng, Z., Long, J., Sosso, D., Liu, B., Eom, J.-S., Huang, S., Liu, S., Vera Cruz, C., Frommer, W.B.*, et al.* (2015). Gene targeting by the TAL effector PthXo2 reveals cryptic resistance gene for bacterial blight of rice. The Plant Journal *82*, 632-643.

Zhu, W., Yang, B., Chittoor, J.M., Johnson, L.B., and White, F.F. (1998). AvrXa10 contains an acidic transcriptional activation domain in the functionally conserved C terminus. Molecular Plant-Microbe Interactions *11*, 824-832.

Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J.D.G., Boller, T., and Felix, G. (2006). Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts *Agrobacterium*-mediated transformation. Cell *125*, 749-760.

Appendix

Appendix Table 3.1.1: All unique RVD sequences of all TALE classes that have at least one member with an aberrant repeat.

class [1]	aberrant repeat	RVD sequence [2, 3]	strain [4]
TalDS	40aa	NI HG NI NI NS HD HD HD HD NS N* N* HD HD NS NS NN NN NI NG NN NI N* NS N*	PXO211
TalAC	40aa	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*	PXO86, PXO524, PXO2684, PXO83, PXO1865, PXO236, MAFF311018, PXO145
TalAC	40aa	NI HG NI NI NS HD NN HD HD HD NS N* N* HD HD NS NS NN N*	KACC10331
TalDV	40aa	NI HG NI NI NS HD NN HD HD HD NS HD N* NI HD HD NN NS NN NN NG NN HD N* NS NS N*	PXO602, PXO563
Tal2$_{(K-74)}$	40aa	NI HG NI HG NI NI NI HD NN HD HD HD NG HD N* NI HD HD NN NI NN NN NG NN HD N* NS N*	K-74
TalBH	40aa	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*	PXO142, PXO61
TalEQ	40aa	NI HG NI HG NI NI NI HD NN HD HD HD NG HD N* HD HD NN NS NI NN NN NG NN NN NN NG NN HD N* NS N*	XF89b
TalAS	34aa	NI HG NI NI HG HD NN HD HD HD NI NI NN NI HD HD HD HG NN NN HD NS NN HD N* NS N*	PXO211, PXO86, PXO83, PXO145
TalAS	40aa	NI HG NI NI HG HD NN HD HD HD NI NI NN NI HD HD HD HG NN NN HD NS NN HD NG NS N*	PXO563, PXO602, MAFF311018, PXO142, PXO524, K74
TalAS	40aa	NI HG NI NS HG HD NN HD HD HD NI NI NN NI HD HD HD HG NN NN HD NS NN HD NG NS N*	PXO71
TalAS	40aa	NN HD HD HD NI NI NN NI HD HD HD HG NN NN HD NS NN HD NG NS N*	PXO71 (pseudogene)
TalAS	40aa	NI HG NI NI NS HD NN HD HG NI NI HG HD NN HD HD HD NI NI NN NI HD HD HD HG NN NN HD NS NN HD N* NS N*	KACC10331
TalER	40aa	NI HG NI NI HG HD NN HD HD HD NI NI NN NI HD NG NI NN NN HD NS NN HD NG NS N*	XF89b
TalAQ	34aa	HD HD NN NN NI NG HD S* HG HD NG N* NG HD	PXO142 (pseudogene)
TalAQ	34aa	HD HD NN NN NI NG HD S* HG HD NG N* NG HD HD N* NI NI NN HD HI ND HD NG NN HG N*	PXO602, PXO71, PXO524, K74
TalAQ	34aa	HG HD NG N* NG HD HD N* NI NI NN HD HI ND HD NG NN HG N*	PXO602 (pseudogene)
TalAQ	42aa	HD HD NN NN NG NG HD NS HG HD NG N* HD HD HD N* NN NI NN HD HI ND HD HG NN HG N*	PXO99A, ICMP3125
TalAQ	42aa	HD HD NN NN NS NG HD S* HG HD NG N* HD HD HD N* NN NI NN HD HM ND HD HG AN HG N*	PXO145
TalAQ	42aa	HD HD NN NN NS NG HD S* HG HD NG N* HD HD HD N* NN NI NN HD HI ND HD HG NN HG N*	PXO563, MAFF311018, PXO236, PXO86, PXO83, PXO211
TalGB	37aa v1	NI HD NI HD NI NG HD NG NG NG NI NI NI NI HD NG	29-1 (*Xac*)
TalGA	37aa v2	NI HD NI HD NI NG N* NI N* NI HD HD NI HD NG	29-1 (*Xac*)

TalBV	36aa v1	NN HD NI NI NN HA NN NS NI NI HD HA HA HA HD HD HD HA HD N*	CFBP2286
TalBV	36aa v1	NN HD NI NI NN HA NN NS NS NI HD HA HA HA HD HD HD HA HD N*	BLS256
TalBV	36aa v1	NN HD NI NI NN HA NN NS NS NI HD HA HA HA HD HD HD NI HD N*	BXOR1
TalBV	36aa v1	NN HD NI NI NN HA NN NS NS NI HD HA HA HA HD HD HD HA HD N*	L8, RS105, B8-12, BLS279
TalBK	36aa v2	NI HG NI NN NI NN HD NI HD HD NS NS HD NI NI HD NG HD HD HD NG NG	PXO142
TalBK	36aa v2	NI HG NI NN NI NN HD HD HD HD NS NS HD NI NI HD	KACC10331 (pseudogene)
Temp TalBK	36aa v2	NI HG NI NN NI NN HD NI HD HD NS NS HD NI NI HD	PXO35
TalBG	30aa	NI NN NI HD NN NG HD NN HD HG HD HG HG HD HD NG	RS105, L8, B8-12, BLS279,
TalBG	30aa	NI NN HN NN NI NG HD NN HD HG HD HG HG HD HD NG	CFBP2286, BLS256
TalAI	28aa	NS HD NG NG NG NG HD HD HD HD NN HD NG HD NI HD NN N*	K74, PXO563, PXO602, PXO142
TalAI	28aa	NS HD NG NG NG NG NG HD HD HD NN HD NG HD NI HD NN N*	PXO145, PXO83, PXO86
TalAI	28aa	NS NG NG NG NG NG HD HD HD HD NN HD NG HD HD HD NN H*	PXO282
TalAI	28aa	NS HD NG NG NG NG HD HD HD HD NN HD NG HD HD HD NN H*	MAFF311018
TalAI	28aa	NS HD NG NG NG NG NG HD HD HD NN HD NG HD HD HD HD N*	PXO99A
TalAI	28aa	NS HD NG NG NG NG NG HD HD HD NN HD NG HD HD HD HD H*	PXO524, ICMP3125
TalAI	28aa	NS NN NG NG NG NG NG HD HD HD NN HD NG HD HD HD HD H*	PXO236
TalAI	28aa	NS NG NG NG NG NG NG HD HD HD NN HD NG HD HD HD HD H*	PXO86, PXO145, PXO83
TalAI	28aa	NS NG NG NG NG NG HD HD HD NN HD NG HD HD HD HD H*	PXO211
TalCM	28aa	NS ND HG HG HG NG HG HG HD HD HD NN NN HD HG HH H* H* NN HD H*	BLS256, CFBP2286
TalCM	28aa	NI ND HG HG HG NG HG HG HD HD HD NN NN HD HG HH H* H* NN HD H*	BLS279, L8
TalCM	28aa	NI ND HG HG HG HG NG HG HG HD HD HD HD NN NN HD HG HH H* H* NN HD H*	B8-12, RS105
TalDR	28aa	NS HD NG NG NG NG HD HD HD HD NN HD HD HD HD NN H*	PXO71, PXO563, PXO142, K-74
TalAW	28aa	NS HD NG NG NG HD HD NG HD NN NG HD NN HD NG HD NI N*	PXO99A
TalDQ	28aa	NS HD NG NG NG NG HD HD HD NN HD NG HD NI HD N*	PXO71
TalCO	28aa	NI HG HG HG HG NG HD HD HD HD HD NG N* NI NN HD HD H*	BXOR1

[1] TALEs are sorted into classes according to the AnnoTALE nomenclature (Grau *et al.*, 2016).
[2] TALEs are aligned to the position of the aberrant repeat. Aberrant repeats are underlined.
[3] Only unique RVD sequences are shown. The strains in which they occur are indicated.
[4] If not otherwise indicated, strains belong to either *X. oryzae* pv. *oryzae* (*Xoo*) or pv. *oryzicola* (*Xoc*), the two exceptions listed above belong to *X. axonopodis* pv. *citri* (*Xac*).

Appendix Table 3.1.2: Unique nucleotide and amino acid sequences of all known naturally occurring aberrant repeats.

length / class	nucleotide and amino acid sequence [1, 2, 3]	strain
34aa [4] (AQ/DU)	CTGACCCCGGAC----------CAGGTGGTGGCCATCGCCAGTAATATTGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGG L T P D Q V V A I A S N I G G K Q A L E T V Q R L L P V L C Q D H G	PXO71, PXO524, PXO602, K-74
42aa (AQ)	CTGACCCCGGACCAGGTCGTGGCCATCGCCAGTAATCAGGTGGTGGCCATCGCCAGTAATATTGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGC L T P D Q V V A I A S N Q V V A I A S N I G G K Q A L E T V Q R L L P V L C Q D H G	PXO99A, MAFF311018, PXO83, PXO86, PXO211, PXO236, PXO563, ICMP3125
42aa (AQ)	CTGACCCCGGACCAGAGCGTGGCCATCGCCAGTAATCGGGTGGTGGCCATCGCCAGTAATATTGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGC L T P D Q S V A I A S N R V V A I A S N I G G K Q A L E T V Q R L L P V L C Q D H G	PXO145
40aa (AC)	CTGACCCCGGACCATGTTGTGGCTATCGCCAGCAAT---GGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTACAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGC L T P D H V V A I A S N * G G K Q A L E T V Q R L L P V Q R L L P V L C Q D H G	KACC10331
40aa (AC/DS/ DV)	CTGACCCCGGACCAGGTCGTGGCCATCGCCAGCAAT---GGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTACAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGC L T P D Q V V A I A S N * G G K Q A L E T V Q R L L P V Q R L L P V L C Q D H G	MAFF311018, PXO1865, PXO2684, PXO83, PXO86, PXO145, PXO236, PXO524, PXO211, PXO563, PXO602
40aa (BH/DY)	CTGACCCCGGACCAGGTCGTGGCCATCGCCAGCAAT---GGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCTGTACAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGC L T P D Q V V A I A S N * G G K Q A L E T V Q R L L P V Q R L L P V L C Q D H G	PXO61, K-74, PXO142
40aa (EQ)	CTGACCCCGGACCAGGTCGTGGCCATCGCCAGCAAT---GGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCTGTACAGCGGCTGTTGCCGGTGCTGTGCCAGGCCAATGGC L T P D Q V V A I A S N * G G K Q A L E T V Q R L L P V Q R L L P V L C Q A N G	XF89b
34aa [4] (AS)	CTGACCCCGGACCAGGTGGTGGCCATCGCCAACAATAACGGCGGCAAGCAGGCGCTGGAGACGGTG----------CAGCGGCTGGTGCCGGTGCTGTGCCAGGACCATGGC L T P D Q V V A I A N N N G G K Q A L E T V Q R L V P V L C Q D H G	PXO211, PXO145, PXO83, PXO86
40aa (AS/ER)	CTGACCCCGGACCAGGTGGTGGCCATCGCCAACAATAACGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTACAGCGGCTGGTGCCGGTGCTGTGCCAGGACCATGGC L T P D Q V V A I A N N N G G K Q A L E T V Q R L L P V Q R L V P V L C Q D H G	PXO142, PXO602, PXO563, PXO524, , KACC10331, MAFF311018, K-74, PXO71, FX89b
37aa v1 (GB)	CTGACCCCGGACCAGGTGGTGGCCATCGCCAGCAATGGCGGTGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCCCCATGGC L T P D 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 P H G	29-1 (*Xac*)
37aa v2 (GA)	CTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGC L T P E Q V V A I A S H D G G K Q A L E T V Q R V Q R L L P V L C Q A H G	29-1 (*Xac*)
36aa v1 (BK)	CTGACCCCGGCCCAGGTCGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGCCTGTTGCCGGTGCCGGTGCTGTGCCAGGCCCATGGC L T P A Q V V A I A S H D G G K Q A L E T V Q R L L P V P V L C Q A H G	KACC10331, PXO142, PXO35

36aa v1 (BK)	CTGACCCCGGACCAGGTGGTGGCCATCGCCAGCAATAGTGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGCCTGTTG**CCGGTGCCGGTG**CTGTGCCAGGCCCATGGC L T P D Q V V A I A S N S G G K Q A L E T V Q R L L **P V P V** L C Q A H G	KACC10331, PXO142, PXO35
34aa [4] (BV)	CTGACCCCGGCGCAGGTGGTGGCCATCGCCAGCAATATTGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTTTGCCAG**GCCCAT**------GGC L T P A Q V V A I A S N I G G K Q A L E T V Q R L L P V L C Q **A H** G	BXOR1
34aa [4] (BV)	CTGACCCCGGCGCAGGTGGTGGCCATCGCCAGCAATATTGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAG**GCCCAT**------GGC L T P A Q V V A I A S N I G G K Q A L E T V Q R L L P V L C Q **A H** G	BLS256, CFBP2286
36aa (BV)	CTGACCCCGGCGCAGGTGGTGGCCATCGCCAGCAATATTGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAG**GCCCATGCCCAT**GGC L T P A Q V V A I A S N I G G K Q A L E T V Q R L L P V L C Q **A H A H** G	B8-12, BLS279, L8, RS105
30aa (BG)	CTGATCCCGGACCAGGTGGTGGCCATCGCCAGCAATATTGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGC---------- L I P D Q V V A I A S N I G G K Q A L E T V Q R L L P V L C	BLS256, CFBP2286
30aa (BG)	CTGACCCCAGACCAGGTGGTGGCCATCGCCAGCAATAACGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGC---------- L T P D Q V V A I A S N N G G K Q A L E T V Q R L L P V L C	B8-12, BLS279, L8, RS105
28aa (CM)	CTGACCCCGGACCAGGTGGTGGCCATCGCCAGCAATGGCGGCGGCAAGCAGGCGCTGGAGACG----------GTACTGCGCCAAGCCCATGGC L T P D Q V V A I A S N G G G K Q A L E T V L R Q A H G	B8-12, BLS279, CFBP2286, L8, RS105, BLS256
28aa (AW)	CTGACCCCGGACCAGGTGGTGGCCATCGCCAGCAATGGCGGCGGCAAGCAGGCGCTGGAGACG----------GTACTGTGTCAGGCCCATGGC L T P D Q V V A I A S N G G G K Q A L E T V L C Q A H G	PXO99A
28aa (CO)	CTGACCCCGGACCAGGTGGTGGCCATCGCCAGCCATGGCGGCGGCAAGCAGGCGCTGGAGACG----------GTGCTGTGCCAGGCCCATGGC L T P D Q V V A I A S H G G G K Q A L E T V L C Q A H G	BXOR1
28aa (DQ)	CTGACCCCGGACCAGGTGGTGGCCATCGCCAGCAATGGCGGCGGCAAGCAGGCGCTGGAGACG----------GTACTGTGCCAGGCCCATGGC L T P D Q V V A I A S N G G G K Q A L E T V L C Q A H G	PXO71
28aa (DR)	CTGACCCCGGACCAGGTGGTGGCCATCGCCAGCAATGGCGGCGGCAAGCAGGCGCTGGAGACG----------GTACTGTGCCAGGCCCATGGC L T P D Q V V A I A S N G G G K Q A L E T V L C Q A H G	PXO71, PXO563, PXO142, K-74
28aa (AI)	CTGACCCCGGACCAGGTGGTGGCCATCGCCAGCAATGGCGGCGGCAAGCAGGCGCTGGAGACG----------GTACTGTGCCAGGCCCATGGC L T P D Q V V A I A S N G G G K Q A L E T V L C Q A H G	PXO99, MAFF311018, PXO83, PXO86, PXO145, PXO211, PXO236, PXO282, PXO524, PXO563, PXO602, ICMP3125, PXO142, K-74

[1] Nucleotides encoding for an RVD and the RVDs themselves are highlighted in yellow.

[2] Naturally occurring polymorphisms in the nucleotide and amino acid sequences of the different aberrant repeat types are highlighted in green.

[3] Regions affected by the length alterations are indicated in bold, normal and dashed lines highlight regions that are duplicated/deleted.

[4] If a TALE class has members without the aberrant repeat present, the 34aa repeat from the corresponding position is listed as well.

Appendix Table 3.1.3: General architecture of the generated aberrant repeat modules.

aberrant repeat module	general nucleotide and amino acid sequence of the generated aberrant repeat modules and their overhangs [1]
p1 (40aa)	GAAGACTTTTACCCCGGAGCAGGTGGTGGCCATCGCCAGCNNNNNNGGCGGCAAGCAGGCGCTGGAGACGGTG**CAGCGGCTGTTGCCGGTACAGCGGCTGGTGCCGGTG**CTGTGCCAGGCCCATGGCCTGAAGTCTTC T P E Q V V A I A S X X G G K Q A L E T V **Q R L L P V Q R L V P V** L C Q A H G L
p2 (40aa)	GAAGACTTCTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCNNNNNNGGCGGCAAGCAGGCGCTGGAGACGGTG**CAGCGGCTGTTGCCGGTACAGCGGCTGGTGCCGGTG**CTGTGCCAGGACCATGGCCTGACACCGAAGTCTTC L T P E Q V V A I A S X X G G K Q A L E T V **Q R L L P V Q R L V P V** L C Q D H G L T P
p3 (40aa)	GAAGACTTACCGGAGCAGGTGGTGGCCATCGCCAGCNNNNNNGGCGGCAAGCAGGCGCTGGAGACGGTG**CAGCGGCTGTTGCCGGTACAGCGGCTGGTGCCGGTG**CTGTGCCAGGCCCATGGCCTCACCAAGTCTTC P E Q V V A I A S X X G G K Q A L E T V **Q R L L P V Q R L V P V** L C Q A H G L T
p4 (40aa)	GAAGACTTCACCCGGAGCAGGTGGTGGCCATCGCCAGCNNNNNNGGCGGCAAGCAGGCGCTGGAGACGGTG**CAGCGGCTGTTGCCGGTACAGCGGCTGGTGCCGGTG**CTGTGCCAGGCCCATGGCCTGACTCAAGTCTTC T P E Q V V A I A S X X G G K Q A L E T V **Q R L L P V Q R L V P V** L C Q A H G L T
p5 (40aa)	TGAAGACTTACTCCGGAGCAGGTGGTGGCCATCGCCAGCNNNNNNGGCGGCAAGCAGGCGCTGGAGACGGTG**CAGCGGCTGTTGCCGGTACAGCGGCTGGTGCCGGTG**CTGTGCCAGGCCCATAAGTCTTC T P E Q V V A I A S X X G G K Q A L E T V **Q R L L P V Q R L V P V** L C Q A H
p1 (42aa)	GAAGACTTTTACCCCGGAG**CAGGTGGTGGCCATCGCCAGTAATCAGGTGGTGGCCATCGCCAGC**NNNNNNGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGAAGTCTTC T P E **Q V V A I A S N Q V V A I A S** X X G G K Q A L E T V Q R L L P V L C Q A H G L
p2 (42aa)	GAAGACTTCTGACCCCGGAC**CAGGTGGTGGCCATCGCCAGTAATCAGGTGGTGGCCATCGCCAGC**NNNNNNGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACACCGAAGTCTTC L T P D **Q V V A I A S N Q V V A I A S** X X G G K Q A L E T V Q R L L P V L C Q D H G L T P
p3 (42aa)	GAAGACTTACCGGAC**CAGGTGGTGGCCATCGCCAGTAATCAGGTGGTGGCCATCGCCAGC**NNNNNNGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTCACCAAGTCTTC P D **Q V V A I A S N Q V V A I A S** X X G G K Q A L E T V Q R L L P V L C Q D H G L T
p4 (42aa)	GAAGACTTCACCCCGGAG**CAGGTGGTGGCCATCGCCAGTAATCAGGTGGTGGCCATCGCCAGC**NNNNNNGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGACTCAAGTCTTC T P E **Q V V A I A S N Q V V A I A S** X X G G K Q A L E T V Q R L L P V L C Q A H G L T
p5 (42aa)	GAAGACTTACTCCGGAG**CAGGTGGTGGCCATCGCCAGTAATCAGGTGGTGGCCATCGCCAGC**NNNNNNGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCATAAGTCTTC T P E **Q V V A I A S N Q V V A I A S** X X G G K Q A L E T V Q R L L P V L C Q A H
p6 (42aa)	GAAGACTTCCATGGCCTGACCCCGGAG**CAGGTGGTGGCCATCGCCAGTAATCAAGTCGTCGCGATTGCCAGT**NNNNNNGGCGGCAAGCAGGCGCTGGAAAGTCTTC H G L T P E **Q V V A I A S N Q V V A I A S** X X G G K Q A L E
p1 (30aa)	GAAGACTTTTACCCCGGACCAGGTGGTGGCCATCGCCAGCNNNNNNGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGC----------CTGAAAGTCTTC T P D Q V V A I A S X X G G K Q A L E T V Q R L L P V L C L
p2 (30aa)	GAAGACTTCTGATCCCGGACCAGGTGGTGGCCATCGCCAGCNNNNNNGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGC----------CTGATACCGAAGTCTTC L I P D Q V V A I A S X X G G K Q A L E T V Q R L L P V L C L I P
p3 (30aa)	GAAGACTTACCGGACCAGGTGGTGGCCATCGCCAGCNNNNNNGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGC----------CTCACCAAGTCTTC P D Q V V A I A S X X G G K Q A L E T V Q R L L P V L C L T
p4 (30aa)	GAAGACTTCACCCGGACCAGGTGGTGGCCATCGCCAGCNNNNNNGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGC----------CTGACTCAAGTCTTC T P D Q V V A I A S X X G G K Q A L E T V Q R L L P V L C L T
p1 (28aa)	GAAGACTTTTACCCCGGAGCAGGTGGTGGCCATCGCCAGCNNNNNNGGTGGCAAGCAGGCGCTGGAGACG----------GTGCTGTGCCAGGCCCATGGCCTGAAGTCTTC T P E Q V V A I A S X X G G K Q A L E T V L C Q A H G L
p2 (28aa)	GAAGACTTCTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCNNNNNNGGCGGCAAGCAGGCGCTGGAGACG----------GTGCTGTGCCAGGCCCATGGCCTGACACCGAAGTCTTC L T P E Q V V A I A S X X G G K Q A L E T V L C Q A H G L T P

p3 (28aa)	GAAGACTTACCGGAGCAGGTGGTGGCCATCGCCAGCNNNNNNGGCGGCAAGCAGGCGCTGGAGACG----------GTGCTGTGCCAGGCCCATGGCCTCACCAAGTCTTC P E Q V V A I A S X X G G K Q A L E T V L C Q A H G L T
p1 (36aa v1)	GAAGACTTTTACCCCGGAGCAGGTGGTGGCCATCGCCAGCNNNNNNGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTG**CCGGTGCCGGTG**CTGTGCCAGGCCCATGGCCTGAAAGTCTTC T P E Q V V A I A S X X G G K Q A L E T V Q R L L **P V P V** L C Q A H G L
p2 (36aa v1)	GAAGACTTCTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCNNNNNNGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTG**CCGGTGCCGGTG**CTGTGCCAGGCCCATGGCCTGACACCGAAGTCTTC L T P E Q V V A I A S X X G G K Q A L E T V Q R L L **P V P V** L C Q A H G L T P
p3 (36aa v1)	GAAGACTTACCGGAGCAGGTGGTGGCCATCGCCAGCNNNNNNGGTGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTG**CCGGTGCCGGTG**CTGTGCCAGGCCCATGGCCTCACCAAGTCTTC L P E Q V V A I A S X X G G K Q A L E T V Q R L L **P V P V** L C Q A H G L T
p2 (36aa v2)	GAAGACTTCTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCNNNNNNGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAG**GCCCATGCCCAT**GGCCTGACACCGAAGTCTTC L T P E Q V V A I A S X X G G K Q A L E T V Q R L L P V L C Q **A H A H** G L T P
p2 (35aa v1)	GAAGACTTCTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCNNNNNNGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGCCCCGCATGATCTGACACCGAAGTCTTC L T P E Q V V A I A S X X G G K Q A L E T V Q R L L P V L C Q A P H D L T P
p2 (35aa v2)	GAAGACTTCTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCNNNNNNGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGCCGCCGTACGGCCTGACACCGAAGTCTTC L T P E Q V V A I A S X X G G K Q A L E T V Q R L L P V L C Q P P Y G L T P
p1 (37aa v1)	GAAGACTTTTACCCCGGAGCAGGTGGTGGCCATCGCCAGCNNNNNNGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAG**GCCCATGGCCCCCATGGC**CTGAAAGTCTTC T P E Q V V A I A S X X G G K Q A L E T V Q R L L P V L C Q **A H G P H G** L
p2 (37aa v1)	GAAGACTTCTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCNNNNNNGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAG**GCCCATGGCCCCCATGGC**CTGACACCGAAGTCTTC L T P E Q V V A I A S X X G G K Q A L E T V Q R L L P V L C Q **A H G P H G** L T P
p3 (37aa v1)	GAAGACTTACCGGAGCAGGTGGTGGCCATCGCCAGCNNNNNNGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAG**GCCCATGGCCCCCATGGC**CTCACCAAGTCTTC P E Q V V A I A S X X G G K Q A L E T V Q R L L P V L C Q **A H G P H G** L T
p4 (37aa v1)	GAAGACTTCACCCCGGAGCAGGTGGTGGCCATCGCCAGCNNNNNNGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAG**GCCCATGGCCCCCATGGC**CTGACTCAAGTCTTC T P E Q V V A I A S X X G G K Q A L E T V Q R L L P V L C Q **A H G P H G** L T
p1 (37aa v2)	GAAGACTTTTACCCCGGAGCAGGTGGTGGCCATCGCCAGCNNNNNNGGTGGCAAGCAGGCGCTGGAGACG**GTGCAGCGGGTGCAGCGG**CTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGAAAGTCTTC T P E Q V V A I A S X X G G K Q A L E T **V Q R V Q R** L L P V L C Q A H G L
p2 (37aa v2)	GAAGACTTCTGACCCCGGAGCAGGTGGTGGCCATCGCCAGCNNNNNNGGCGGCAAGCAGGCGCTGGAGACG**GTGCAGCGGGTGCAGCGG**CTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGACACCGAAGTCTTC L T P E Q V V A I A S X X G G K Q A L E T **V Q R V Q R** L L P V L C Q A H G L T P
p3 (37aa v2)	GAAGACTTACCGGAGCAGGTGGTGGCCATCGCCAGCNNNNNNGGCGGCAAGCAGGCGCTGGAGACG**GTGCAGCGGGTGCAGCGG**CTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTCACCAAGTCTTC P E Q V V A I A S X X G G K Q A L E T **V Q R V Q R** L L P V L C Q A H G L T
p4 (37aa v2)	GAAGACTTCACCCCGGAGCAGGTGGTGGCCATCGCCAGCNNNNNNGGTGGCAAGCAGGCGCTGGAGACG**GTGCAGCGGGTGCAGCGG**CTGTTGCCGGTGCTGTGCCAGGCCCATGGCCTGACTCAAGTCTTC T P E Q V V A I A S X X G G K Q A L E T **V Q R V Q R** L L P V L C Q A H G L T

[1] Duplicated/deleted regions are underlined and shown in bold, RVDs are highlighted in yellow, *BsaI* restriction sites are labeled in light blue and overhangs in red.

Appendix Table 3.1.4: Aberrant repeat modules generated or used in this study.

aberrant repeat	repeat position						
	RVD	**1**	**2**	**3**	**4**	**5**	**6**
42aa	**HD**	+	+	+	+	+	+
	NN	+	+	+	+	+	+
	NI	+ / s	+[a] / s	+ / s	+ / s	+ / s	+
	NG	+	+	+	+	n/c	+
40aa	**HD**	+	+	+	+	+	(+)
	NN	+	+[a]	+[a]	+	+	(+)
	NI	+	+	+	+	+	(+)
	NG	+	+	+	+	+	(+)
	N*	n/c	+	+	n/c	n/c	(+)
37aa v1	**HD**	+	+	+	+	n/a	n/a
	NN	+	+	+	+	n/a	n/a
37aa v2	**HD**	+	+	+	+	n/c	n/c
	NN	+	+	n/c	+	n/c	n/c
36aa v1	**HD**	n/c	+	+[b]	n/c	n/c	n/c
	NN	n/c	+	+	n/c	n/c	n/c
	NI	+	+	+	n/c	n/c	n/c
	NS	+[b]	+	n/c	n/c	n/c	n/c
36aa v2	**HD**	n/c	+	n/c	n/c	n/c	n/a
	NN	n/c	+	n/c	n/c	n/c	n/a
	NI	n/c	+	n/c	n/c	n/c	n/a
30aa	**HD**	+	+	+	+	n/a	n/a
	NN	+	+	+	+	n/a	n/a
	NI	+	+[a]	+	+	n/a	n/a
	NG	+	+	+	+	n/a	n/a
28aa	**HD**	n/c	+	+	n/c	n/c	n/c
	NN	+	+	+	n/c	n/c	n/c
	NI	n/c	+	+	n/c	n/c	n/c
	NG	+	+	+	n/c	n/c	n/c

\+ Aberrant repeat module was constructed in this study.

+[a] Aberrant repeat module was available prior to this study.

+[b] Aberrant repeat module was constructed by Stefanie Mücke.

+ / s Aberrant repeat module and a stop repeat model were constructed in this study.

(+) Aberrant repeat version at this position requires a special assembly vector, three were constructed, namely L-A(40aa), A-B(40aa) and B-C(40aa)

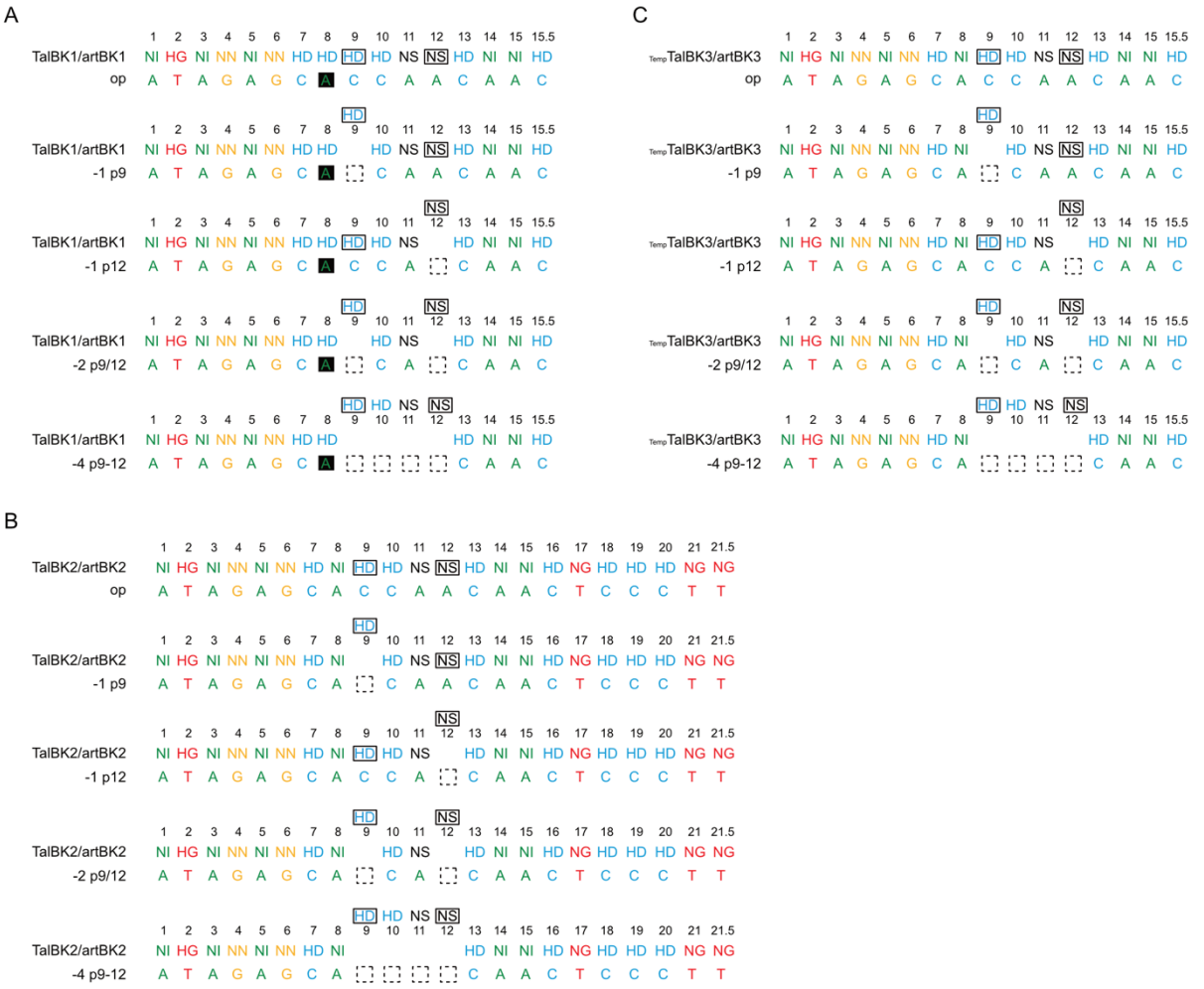
n/c Aberrant repeat module for this position was not constructed.

n/a Aberrant repeat cannot be placed at this position due to an incompatibility between the overhangs used in the Golden TALE assembly kit and the nature of the aberrant repeat.

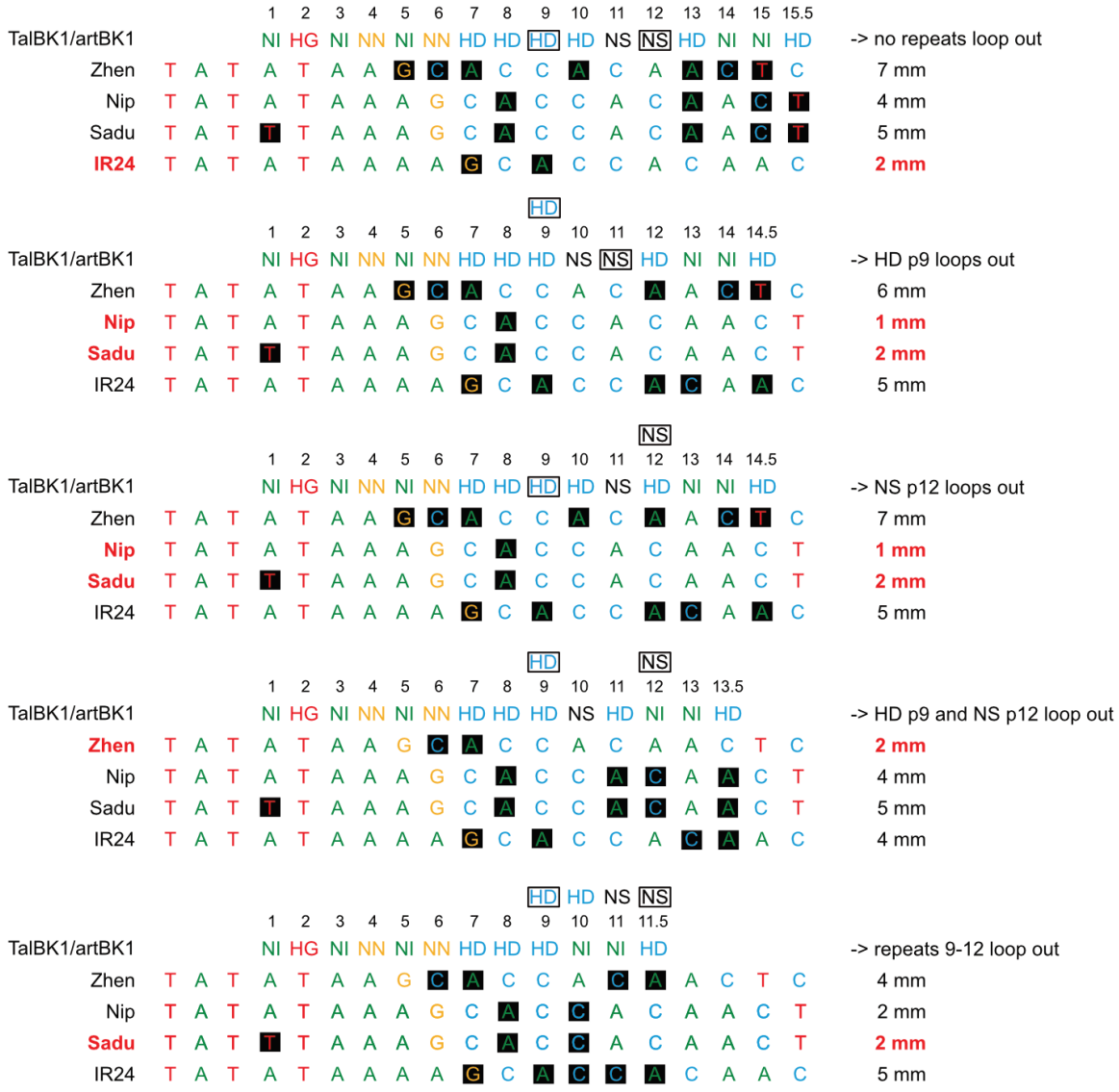
note: All generated modules are fully compatible with the “Golden TAL Technology” cloning kit (Geißler *et al.*, 2011).

Appendix Table 3.1.5: Sequenced *OsSWEET13* promoter alleles.

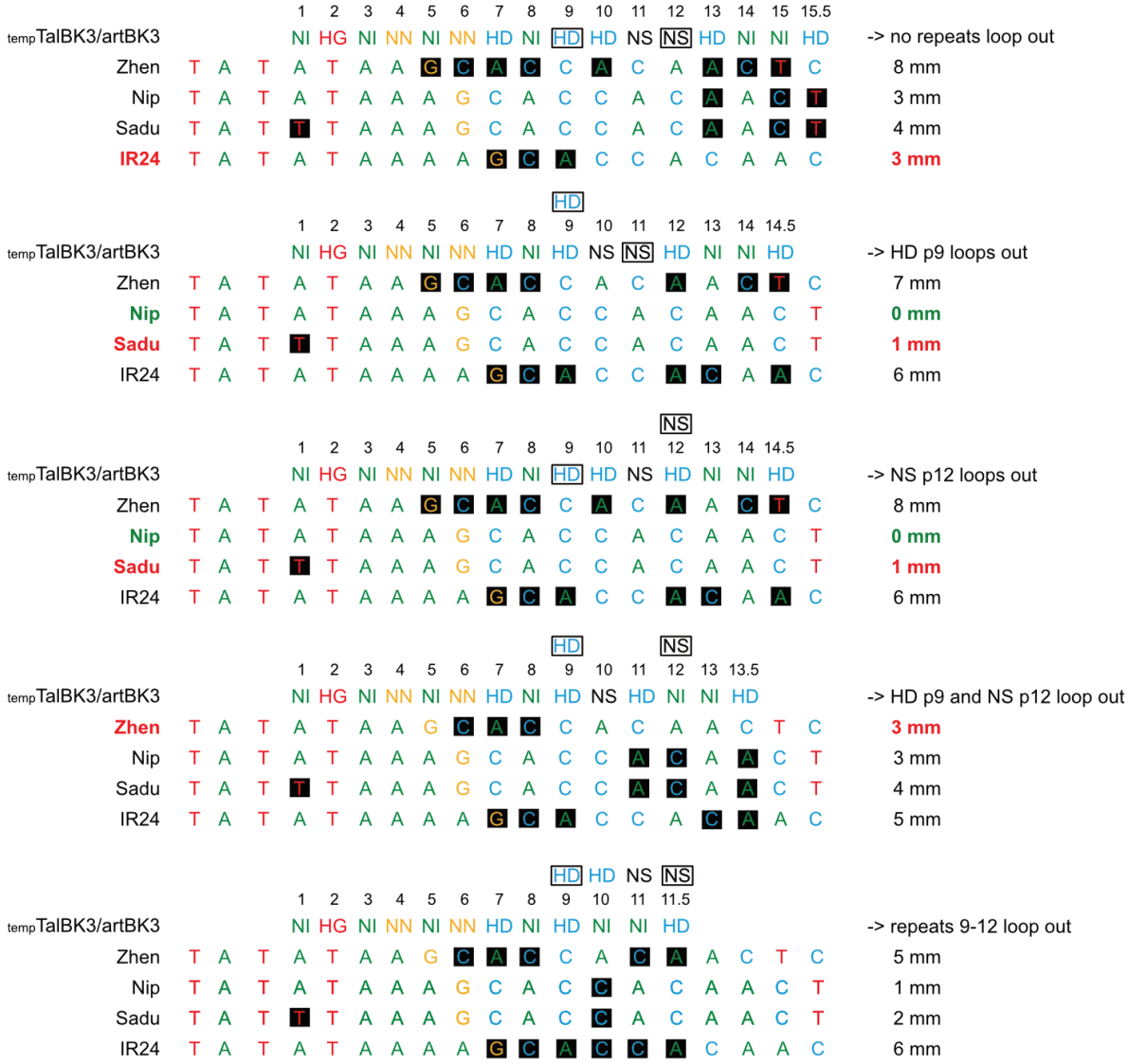
rice cultivar	sequence result for the region covering the putative TALE binding site
Aswina	TATATAA**A**-GCACCACAACTCCCTT
Cypress	TATATAA**A**-GCACCACAACTCCCTT
Nipponbare	TATATAA**A**-GCACCACAACTCCCTT
Pokkali	TATATAA--GCACCACAACTCCCTT
Ryada	TATATAA**AA**GCACCACAACTCCCTT
Sadu Cho	TAT**T**TAA**A**-GCACCACAACTCCCTT



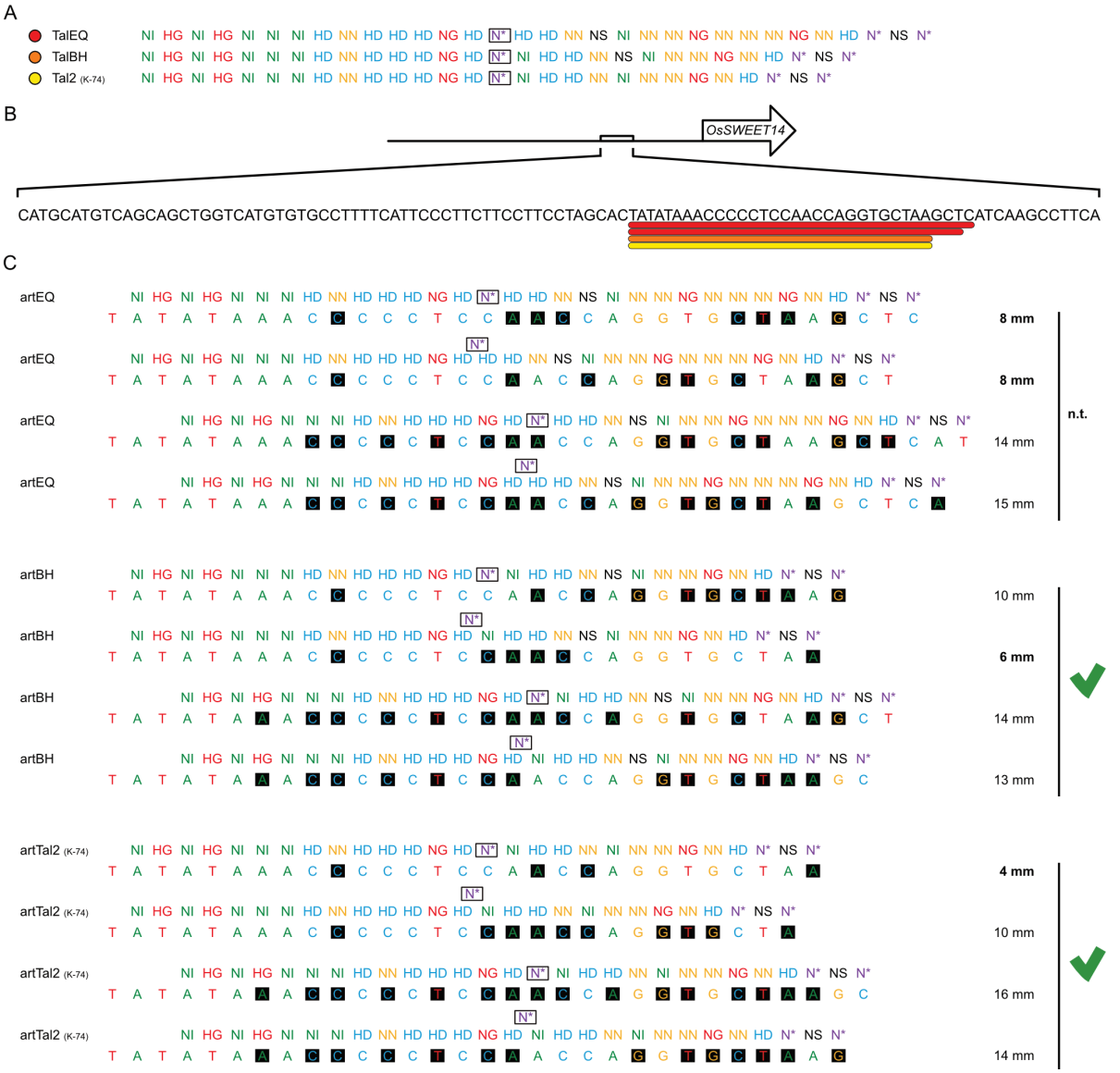
Appendix Figure 3.1.1: Binding possibilities of the three TalBK class members if combined with different target boxes. The target sequence is either an optimal box or one of several different frameshift derivatives containing a deletion of either one nucleotide at position 9 (-1 p9) or 12 (-1 p12), two nucleotides at position 9 and 12 (-2 p9/12) or four nucleotides (-4 p9-12). **(A)** TALE TalBK1/artBK1. **(B)** TalBK2/artBK2 and **(C)** tempTalBK3/artBK3. Aberrant repeats are boxed, deleted positions within the target boxes are indicated by dashed squares. Mismatched positions within the target boxes are shaded in black. All target sequences were designed to perfectly match TalBK2/artBK2.



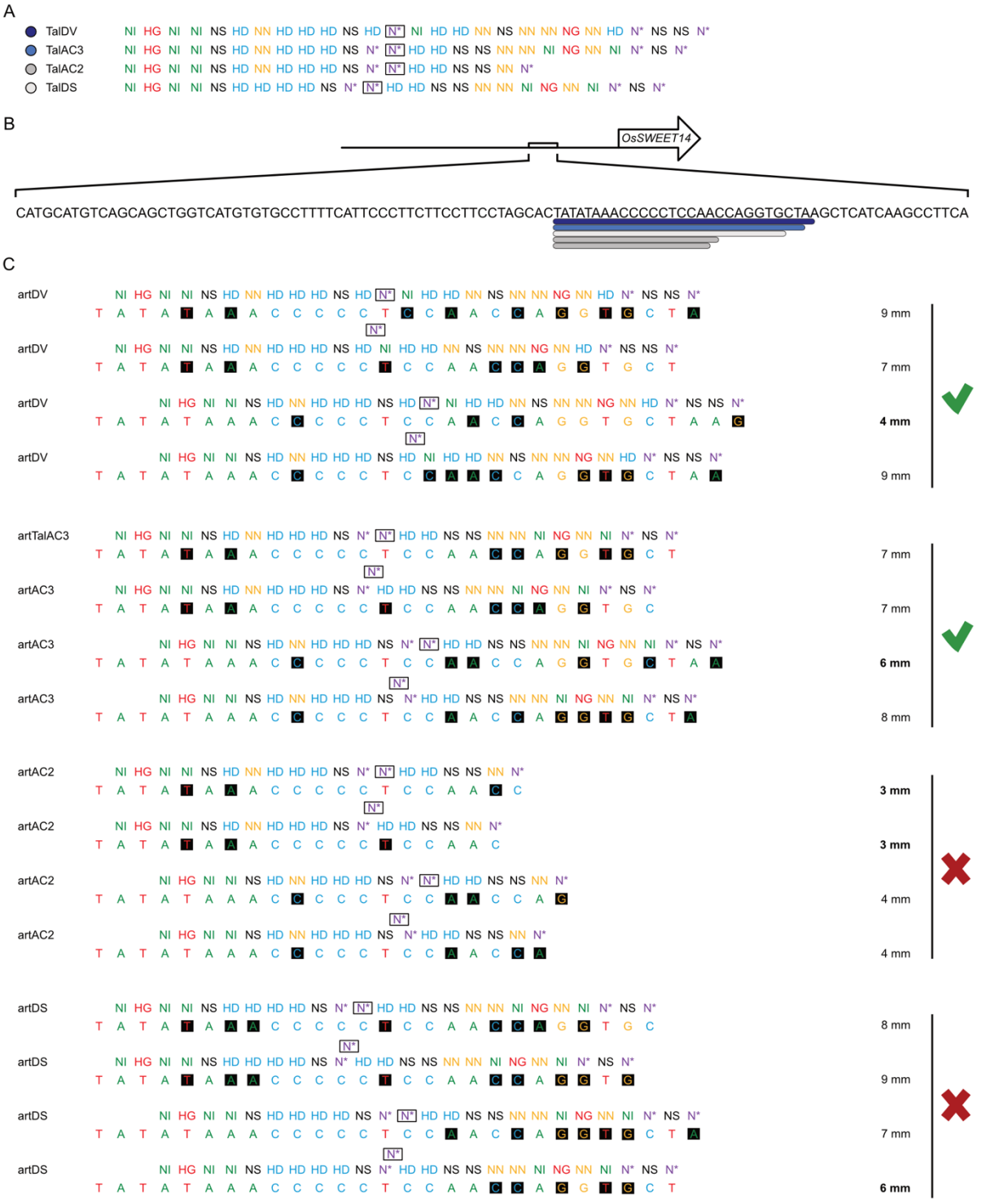
Appendix Figure 3.1.2: Best binding options of TalBK1/artBK1 in the four known *OsSWEET13* promoter variants. Aberrant repeats are boxed. A mismatch to the target box is indicated by a black square. The lowest number of mismatches (mm) for each rice cultivar is indicated in bold. If the promoter fragment from this cultivar was activated in the GUS reporter assay, it is coloured green, if not, it is labelled in red.



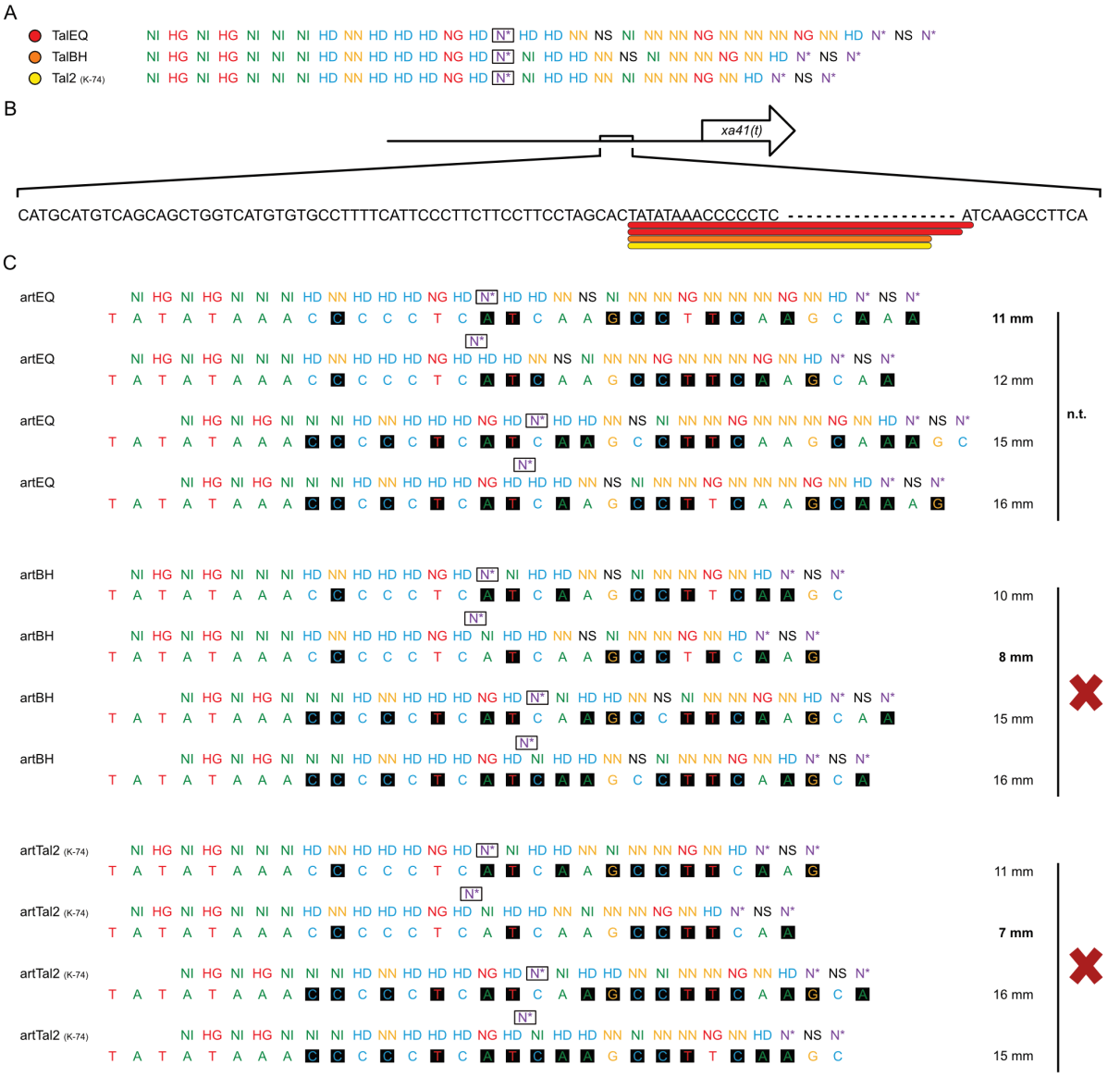
Appendix Figure 3.1.3: Best binding options of tempTalBK3/artBK3 in the four known *OsSWEET13* promoter variants. Aberrant repeats are boxed. A mismatch to the target box is indicated by a black square. The lowest number of mismatches (mm) for each rice cultivar is indicated in bold, if the promoter fragment from this cultivar was activated in the GUS reporter assay, it is coloured in green, if not, it is labelled in red.



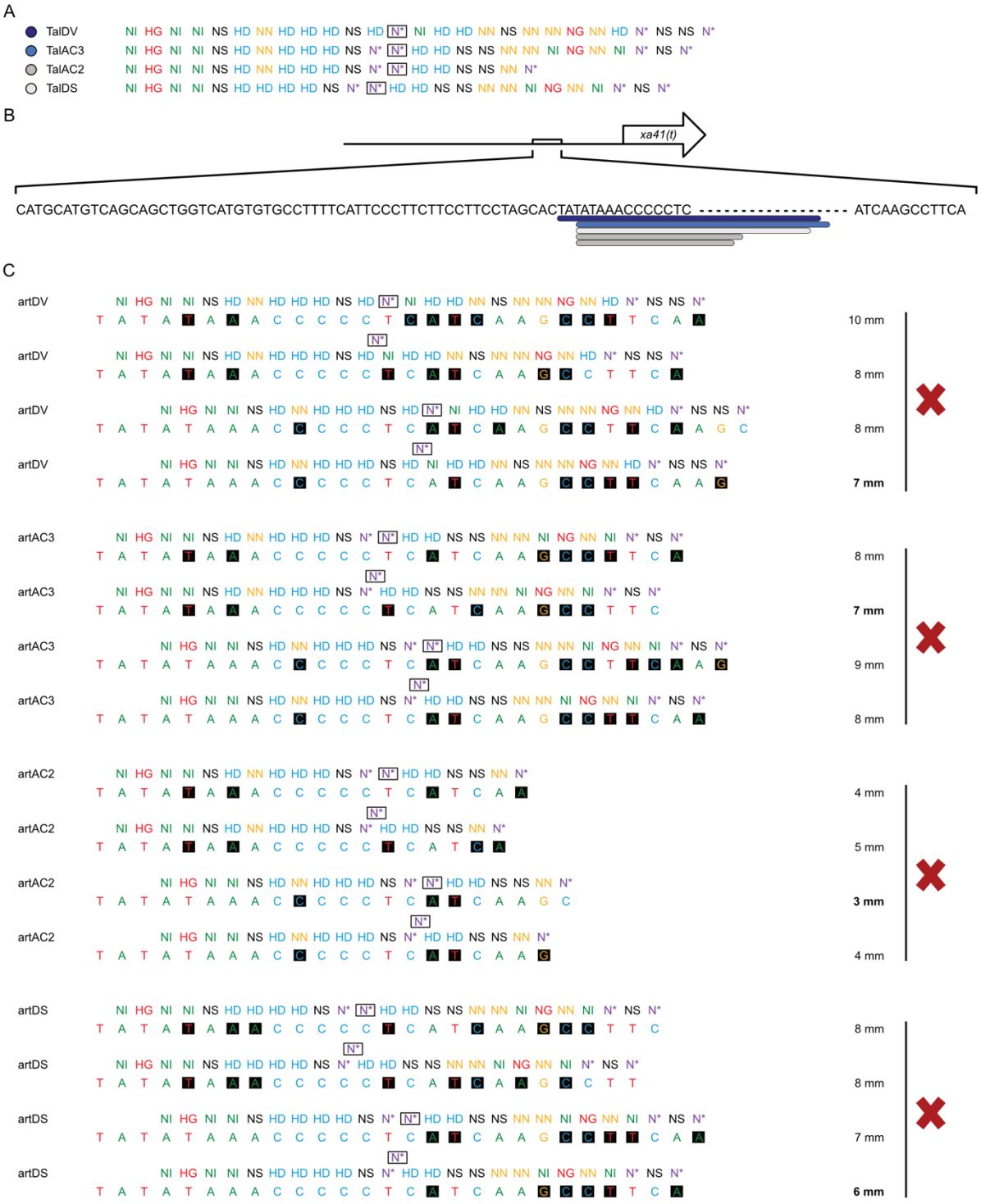
Appendix Figure 3.1.4: TalBH and TalBH-like TALEs/TALE classes and their binding possibilities in the *OsSWEET14* promoter. (A) RVD sequences and **(B)** location of the putative binding sites in the *OsSWEET14* promoter for the two TALE classes TalEQ and TalBH as well as Tal2 from K-74. The aberrant repeat is boxed. The colours of the putative target sites correspond to the different TALE/TALE classes. **(C)** All three TALEs/TALE classes can theoretically use one of two binding frames and doing so with either the aberrant repeat partaking in DNA binding or by looping it out. The first two rows show the TALEs using the first binding frame, the second two rows show the second binding frame starting 2 bp later. The binding option with the least number of mismatches (mm) is shown in bold. TALEs/TALE classes that were able to induce the *OsSWEET14* reporter construct are indicated by a green tick those unable to do so with a red cross. TalEQ was not tested (n.t.).



Appendix Figure 3.1.5: TalAC and TalAC-like TALEs/TALE classes and their binding possibilities in the *OsSWEET14* promoter. (A) RVD sequences and **(B)** location of the putative binding sites in the *OsSWEET14* promoter for the three TALE classes TalDV, TalAC and TalDS. The aberrant repeat is boxed. The colours of the putative target sites correspond to the different TALEs/TALE classes. **(C)** All TALEs/TALE classes can theoretically use one of two binding frames and doing so with either the aberrant repeat partaking in DNA binding or by looping it out. The first two rows show the TALE using the first binding frame, the second two rows show the second binding frame starting 2 bp later. The binding option with the least number of mismatches (mm) is shown in bold. TALEs/TALE classes that were able to induce the *OsSWEET14* reporter construct are indicated by a green tick, those unable to do so with a red cross.



Appendix Figure 3.1.6: TalBH and TalBH-like TALEs/TALE classes and their binding possibilities in the *OsSWEET14* promoter allele *xa41(t)*. **(A)** RVD sequences and **(B)** location of the putative binding sites in the *OsSWEET14* promoter allele *xa41(t)* for the two TALE classes TalEQ and TalBH as well as Tal2 from K-74. The aberrant repeat is boxed. The colours of the putative target sites correspond to the different TALEs/TALE classes. **(C)** All three TALEs/TALE classes can theoretically use one of two binding frames and doing so with either the aberrant repeat partaking in DNA binding or by looping it out. The first two rows show the TALE using the first binding frame, the second two rows show the second binding frame starting 2 bp later. The binding option with the least number of mismatches (mm) is shown in bold. TALEs/TALE classes that were able to induce the *OsSWEET14* reporter construct are indicated by a green tick, those unable to do so with a red cross. TalEQ was not tested (n.t.).



Appendix Figure 3.1.7: TalAC and TalAC-like TALEs/TALE classes and their binding possibilities in the *OsSWEET14* promoter allele *xa41(t)*. **(A)** RVD sequences and **(B)** location of the putative binding sites in the *OsSWEET14* promoter allele *xa41(t)* for the three TALE classes TalDV, TalAC and TalDS. The aberrant repeat is boxed. The colours of the putative target sites correspond to the different TALEs/TALE classes. **(C)** All TALEs/TALE classes can theoretically use one of two binding frames and doing so with either the aberrant repeat partaking in DNA binding or by looping it out. The first two rows show the TALE using the first binding frame, the second two rows show the second binding frame starting 2 bp later. The binding option with the least number of mismatches (mm) is shown in bold. TALEs/TALE classes that were able to induce the *OsSWEET14* reporter construct are indicated by a green tick, those unable to do so with a red cross.

Appendix Table 3.2.1: Complete list of genes used to define the genomic TALE clusters in *Xoc*.

Cluster	*Xoc* strain	flanking genes used to define genomic TALE clusters (locus tag / protein ID) [1]	
I		**Esterase-lipase-thioesterase family protein CDS**	**tRNA uridine 5-carboxymethylaminomethyl modification enzyme MnmG CDS**
	BLS279	*ACU10_00725* / AKN95550.1	*ACU10_00790* / AKN95558.1
	B8-12	*ACU13_00730* / AKN91808.1	*ACU13_00790* / AKN91816.1
	RS105	*ACU12_RS00725* / WP_014505232.1	*ACU12_RS00785* / WP_047340288.1
	L8	*ACU14_RS00725* / WP_014505232.1	*ACU14_RS00785* / WP_047340288.1
	BLS256	*XOC_4538* / AEQ98595.1	*XOC_4523* / AEQ98581.1
	CFBP2286	*ACU11_20980* / AKO21540.1	*ACU11_20915* / AKO21532.1
	BXOR1	*ACU15_00730* / AKN99290.1	*ACU15_00795* / AKN99298.1
	CFBP7342	*BE73_00845* / AJQ85855.1	*BE73_00905* / AJQ85864.1
	CFBP7331	*ACU16_21300* / AKO06235.1	*ACU16_21235* / AKO06226.1
	CFBP7341	*ACU17_RS21305* / WP_014505232.1	*ACU17_RS21240* / WP_024711822.1
II		**NAD-binding domain 4 protein CDS**	**Prolyl oligopeptidase-like protein CDS**
	BLS279	*ACU10_01260* / AKN98824.1	*ACU10_01405* / AKN98829.1
	B8-12	*ACU13_01260* / AKN95095.1	*ACU13_01405* / AKN95100.1
	RS105	*ACU12_RS01255* / WP_014505119.1	*ACU12_RS01395* / WP_048483900.1
	L8	*ACU14_RS01255* / WP_014505119.1	*ACU14_RS01395* / WP_048483900.1
	BLS256	*XOC_4410* / AEQ98476.1	*XOC_4374* / AEQ98445.1
	CFBP2286	*ACU11_20430* / AKO22007.1	*ACU11_20295* / AKO22001.1
	BXOR1	*ACU15_01270* / AKO02470.1	*ACU15_01390* / AKO02472.1
	CFBP7342	*BE73_01415* / AJQ85953.1	*BE73_01565* / AJQ85971.1
	CFBP7331	*ACU16_20615* / AKO06713.1	genes between cluster II and III inverted [3]
	CFBP7341	*ACU17_RS20610* / WP_024712024.1	genes between cluster II and III inverted [3]
III		**beta-N-acetylglucosaminidase CDS**	**sperimidine synthase CDS**
	BLS279	*ACU10_01900* / AKN95737.1	*ACU10_01965* / AKN95746.1
	B8-12	*ACU13_01900* / AKN91997.1	*ACU13_01965* / AKN92006.1
	RS105	*ACU12_RS01890* / WP_048482950.1	*ACU12_RS01955* / WP_048482957.1
	L8	*ACU14_RS01885* / WP_048482950.1	*ACU14_RS01950* / WP_048482957.1
	BLS256	*XOC_4258* / AEQ98331.1	*XOC_4242* / AEQ98317.1
	CFBP2286	*ACU11_19795* / AKO21349.1	*ACU11_19725* / AKO21340.1
	BXOR1	*ACU15_01885* / AKN99475.1	*ACU15_01955* / AKN99483.1
	CFBP7342	*BE73_02210* / AJQ86068.1	*BE73_02290* / AJQ86077.1
	CFBP7331	genes between cluster II and III inverted [3]	*ACU16_19895* / AKO06023.1
	CFBP7341	genes between cluster II and III inverted [3]	*ACU17_RS19900* / WP_024712584.1
IV		**Chemotaxis protein methyltransferase CDS**	**gpmA CDS**
	BLS279	*ACU10_06970* / AKN96552.1	*ACU10_06875* / AKN96538.1
	B8-12	*ACU13_07015* / AKN92822.1	*ACU13_06910* / AKN92806.1
	RS105	*ACU12_RS06995* / WP_048483189.1	*ACU12_RS06910* / WP_011258259.1
	L8	*ACU14_RS06975* / WP_048483189.1	*ACU14_RS06870* / WP_011258259.1
	BLS256	*XOC_1589* / AEQ95757.1	*XOC_1554* / AEQ95734.1
	CFBP2286	*ACU11_07145* / AKO19270.1	*ACU11_07035* / AKO19253.1
	BXOR1	*ACU15_13700* / AKO01393.1	*ACU15_13780* / AKO01406.1
	CFBP7342	*BE73_15430* / AJQ88285.1	*BE73_15520* / AJQ88300.1
	CFBP7331	*ACU16_14925* / AKO05216.1	*ACU16_14995* / AKO05228.1
	CFBP7341	*ACU17_RS14715* / WP_082348299.1	*ACU17_RS14785* / WP_011258259.1
		Integral membrane protein CDS	**Glutathione reductase CDS**
V	**BLS279**	*ACU10_07545* / AKN96640.1	*ACU10_07490* / AKN96633.1
	B8-12	*ACU13_07590* / AKN92910.1	*ACU13_07535* / AKN92903.1
	RS105	*ACU12_RS07570* / WP_014502734.1	*ACU12_RS07515* / WP_024710670.1
	L8	*ACU14_RS07545* / WP_014502734.1	*ACU14_RS07490* / WP_024710670.1
	BLS256	*XOC_1719* / AEQ95883.1	*XOC_1702* / AEQ95870.1
	CFBP2286	*ACU11_07705* / AKO19359.1	*ACU11_07650* / AKO19352.1
	BXOR1	*ACU15_13175* / AKO01303.1	*ACU15_13200* / AKO01307.1
	CFBP7342	*BE73_14830* / AJQ88181.1	*BE73_14855* / AJQ88185.1
	CFBP7331	*ACU16_14400* / AKO05128.1	*ACU16_14425* / AKO05132.1
	CFBP7341	*ACU17_RS14190* / WP_024712694.1	*ACU17_RS14215* / WP_024710670.1

		LysR family protein CDS	Methyl-accepting chemotaxis protein CDS
VI	**BLS279**	*ACU10_08190* / AKN96754.1	*ACU10_08265* / AKN96765.1
	B8-12	*ACU13_08245* / AKN93025.1	*ACU13_08320* / AKN93036.1
	RS105	*ACU12_RS08225* / WP_041183337.1	*ACU12_RS08300* / WP_048483255.1
	L8	*ACU14_RS08195* / WP_041183337.1	*ACU14_RS08270* / WP_048483255.1
	BLS256	*XOC_2870* / AEQ96975.1	*XOC_2854* / AEQ96961.1
	CFBP2286	*ACU11_12965* / AKO20232.1	*ACU11_12890* / AKO20221.1
	BXOR1	*ACU15_08060* / AKO00473.1	*ACU15_08140* / AKO00482.1
	CFBP7342	*BE73_13900* / AJQ88023.1	*BE73_13820* / GI:764029780
	CFBP7331	*ACU16_08700* / AKO04220.1	*ACU16_08785* / AKO04230.1
	CFBP7341	*ACU17_RS08530* / WP_024712334.1	*ACU17_RS08630* / WP_024710578.1
VII		**Efflux pump membrane transporter CDS**	**Serine protease CDS**
	BLS279	*ACU10_09825* / AKN97022.1	*ACU10_10010* / AKN97049.1
	B8-12	*ACU13_09880* / AKN93291.1	*ACU13_10065* / AKN93319.1
	RS105	*ACU12_RS09865* / WP_014503420.1	*ACU12_RS10050* / WP_048483327.1
	L8	*ACU14_RS09825* / WP_014503420.1	*ACU14_RS10000* / WP_048483327.1
	BLS256	*XOC_2507* / AEQ96626.1	*XOC_2475* / AEQ96605.1
	CFBP2286	*ACU11_11245* / AKO19951.1	*ACU11_11115* / AKO19935.1
	BXOR1	*ACU15_09750* / AKO00739.1	*ACU15_09815* / AKO00747.1
	CFBP7342	*BE73_12120* / AJQ87726.1	*BE73_11855* / AJQ87692.1
	CFBP7331	*ACU16_10570* / AKO04514.1	*ACU16_10655* / AKO04525.1
	CFBP7341	cluster not present [2]	cluster not present [2]
VIII		**Adenosylmethionine-8-amino-7-oxononanoate aminotransferase CDS**	**Rhs element Vgr protein CDS**
	BLS279	*ACU10_12035* / AKN99006.1	*ACU10_12075* / AKN97395.1
	B8-12	*ACU13_12090* / AKN95278.1	*ACU13_12125* / AKN93663.1
	RS105	*ACU12_RS12070* / WP_041183213.1	*ACU12_RS12105* / WP_048483403.1
	L8	*ACU14_RS12025* / WP_041183213.1	*ACU14_RS12070* / WP_048483403.1
	BLS256	*XOC_2015* / AEQ96162.1	*XOC_2004* / AEQ96155.1
	CFBP2286	*ACU11_09025* / AKO21810.1	*ACU11_08990* / AKO19577.1
	BXOR1	*ACU15_11840* / AKO02671.1	*ACU15_11875* / AKO01087.1
	CFBP7342	*BE73_09510* / AJQ87301.1	*BE73_09475* / AJQ87296.1
	CFBP7331	*ACU16_12840* / AKO06573.1	*ACU16_12870* / AKO04883.1
	CFBP7341	*ACU17_RS12645* / WP_024711000.1	*ACU17_RS12675* / WP_048488594.1
IX		**UvrABC system protein B CDS**	**Putative acyltransferase CDS**
	BLS279	*ACU10_12895* / AKN97526.1	*ACU10_12865* / AKN97522.1
	B8-12	*ACU13_12945* / AKN93793.1	*ACU13_12915* / AKN93789.1
	RS105	*ACU12_RS12925* / WP_014503764.1	*ACU12_RS12895* / WP_048483435.1
	L8	*ACU14_RS12885* / WP_014503764.1	*ACU14_RS12855* / WP_048483435.1
	BLS256	*XOC_2896* / AEQ96999.1	*XOC_2890* / AEQ96994.1
	CFBP2286	*ACU11_13095* / AKO20253.1	*ACU11_13065* / AKO20249.1
	BXOR1	*ACU15_07930* / AKO00451.1	*ACU15_07960* / AKO00455.1
	CFBP7342	*BE73_14335* / AJQ88093.1	*BE73_14005* / AJQ88043.1
	CFBP7331	*ACU16_08555* / AKO04197.1	*ACU16_08595* / AKO04201.1
	CFBP7341	*ACU17_RS08390* / WP_024710852.1	*ACU17_RS08425* / WP_024712414.1
X		**type IV secretion protein Rhs CDS**	**type VI secretion protein CDS**
	BLS279	cluster not present [2]	cluster not present [2]
	B8-12	cluster not present [2]	cluster not present [2]
	RS105	cluster not present [2]	cluster not present [2]
	L8	cluster not present [2]	cluster not present [2]
	BLS256	cluster not present [2]	cluster not present [2]
	CFBP2286	cluster not present [2]	cluster not present [2]
	BXOR1	*ACU15_14525* / AKO01539.1	*ACU15_14540* / AKO01540.1
	CFBP7342	cluster not present [2]	cluster not present [2]
	CFBP7331	cluster not present [2]	cluster not present [2]
	CFBP7341	cluster not present [2]	cluster not present [2]
XI		**Putative NADH dehydrogenase/ NAD(P)H nitroreductase CDS**	**ABC transporter ATP-binding protein CDS**
	BLS279	*ACU10_18240* / AKN98404.1	*ACU10_18210* / AKN98399.1

	B8-12	*ACU13_18320* / AKN94681.1	*ACU13_18285* / AKN94676.1
	RS105	*ACU12_RS18315* / WP_014504809.1	*ACU12_RS18280* / WP_014504794.1
	L8	*ACU14_RS18235* / WP_014504809.1	*ACU14_RS18195* / WP_014504794.1
	BLS256	*XOC_4041* / AEQ98129.1	*XOC_4024* / AEQ98112.1
	CFBP2286	*ACU11_18760* / AKO21197.1	*ACU11_18685* / AKO21188.1
	BXOR1	*ACU15_02775* / AKN99603.1	*ACU15_02845* / AKN99612.1
	CFBP7342	*BE73_03555* / AJQ86288.1	*BE73_03625* / AJQ86299.1
	CFBP7331	*ACU16_03335* / AKO03344.1	*ACU16_03405* / AKO03353.1
	CFBP7341	*ACU17_RS03175* / WP_014504809.1	*ACU17_RS03245* / WP_014504794.1
XII		**short chain dehydrogenase CDS**	**Thioredoxin CDS**
	BLS279	*ACU10_18920* / AKN98512.1	*ACU10_18965* / AKN98518.1
	B8-12	*ACU13_19000* / AKN94786.1	*ACU13_19045* / AKN94792.1
	RS105	*ACU14_RS18920* / n.a.	*ACU12_RS19040* / WP_011257407.1
	L8	*ACU14_RS18920* / n.a.	*ACU14_RS18965* / WP_011257407.1
	BLS256	*XOC_0479* / AEQ94699.1	*XOC_0457* / AEQ94692.1
	CFBP2286	*ACU11_02045* / AKO18446.1	*ACU11_02000* / AKO18440.1
	BXOR1	*ACU15_18475* / AKO02172.1	*ACU15_18510* / AKO02176.1
	CFBP7342	CDS present but not annotated*	*BE73_22100* / AJQ89407.1
	CFBP7331	*ACU16_02215* / AKO03171.1	*ACU16_02160* / AKO03165.1
	CFBP7341	*ACU17_RS02235* / n.a.	*ACU17_RS02195* / WP_011257407.1

[1] Mobile genetic elements and hypothetical proteins were not used to define the border for a genomic TALE cluster.

[2] This cluster does not occur in the indicated strain.

[3] Due to a genetic recombination event in this strain the genes between TALE cluster II and III are inverted, and thus the position of the indicated border genes is swapped as well.

Appendix Table 3.2.2: Predicted optimal target sequence for all unique *Xoc* TALE class versions.

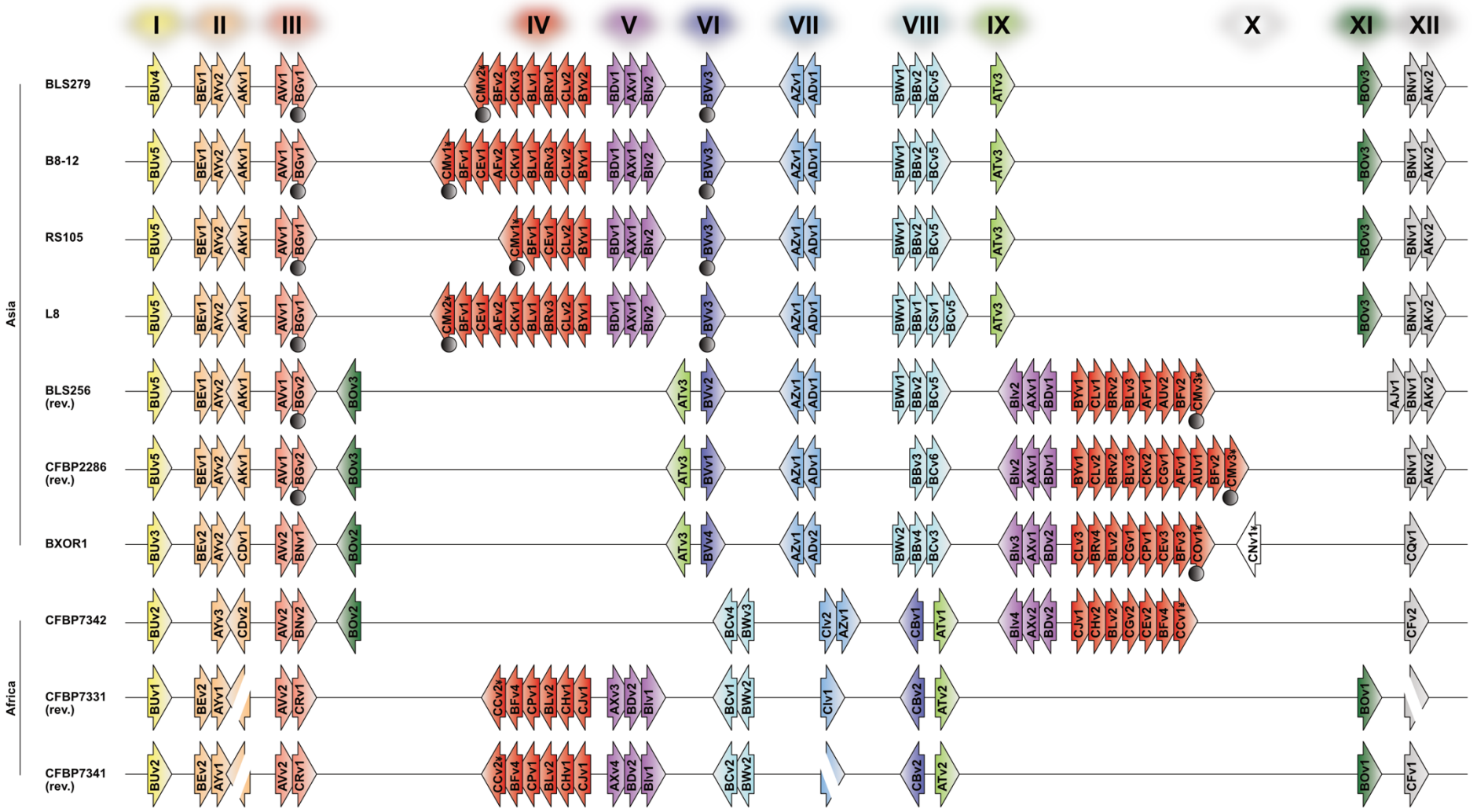
TALE class	Version	RVD sequence / Predicted optimal target sequence
AD	v1	NN HD NS NG HD NN HD NI HD NN HD NN HD NN HD NN NN NN NN NN NN NN HD NG T G C A T C G C A C A C G C G C G G G G G G G C T
AD	v2	NN HD NS NG HD NN N* NI HD NS HD NN HD NN HD NN NN NN NN NN NN NN HD NG T G C A T C G C A C A C G C G C G G G G G G G C T
AD	v3	NN HD NS NG HD NN N* NI HD NS HD NN HD NN NN NN NN NN NN NN NN NN HD NG T G C A T C G C A C A C G C G C G G G G G G G C T
AF	v1	NI NN NI NI NI NI N* NI NG NN NN NN NI NS NG NG T A G A A A A C A T G G G A A T T
AF	v2	NI NN NI NI NI NI N* NI NG NN NN NN NG NS NG HD T A G A A A A C A T G G G T A T C
AF	v3	NI NN NN NI NI NI HD NS HG NN NN NN NI NI NG HD T A G G A A A C A T G G G A A T C
AF	v4	NI NN NN NI NI NI HD NS HG NN NN NN NI NI HG HD T A G G A A A C A T G G G A A T C
AF	v5	NI NN NN NI NI NS HD NS HG NN NN NN NI NI NG HD T A G G A A A C A T G G G A A T C
AJ	v1	NN HD NI NK HD HD HD NG NI NN HD HD NG T G C A G C C C T A G C C T
AK	v1	HD HG HD HG N* NN NG HD NN HD NG NG NN HD N* NG NG T C T C T C G T C G C T T G C C T T
AK	v2	HD HD HD HD HD NG HD NN HD NG HG NN HD N* NG NG T C C C C C T C G C T T G C C T T
AT	v1	NN HD HD HH NN NG NG NG HD NI NG HD HH NG NG NG NG T G C C C G T T T C A T C C T T T T
AT	v2	NN HD HD HH NN NG NS NG HD NI NG HD HH NG NG NG NG T G C C C G T A T C A T C C T T T T
AT	v3	NN HD HD HH NN NG NS NG HD NI NG HD HH NG NG NG HA NG T G C C C G T A T C A T C C T T T C T
AU	v1	HD HD NG NI HG HD HG NI NI NN HD HG HG HG NG NI NG T C C T A T C T A A G C T T T T A T
AU	v2	HD HD NG NI HG HD HG NI HD NN HD HG HG HG NG NI NG T C C T A T C T A C G C T T T T A T
AV	v1	NN NS NG NI HD HD NG NN HD NI NG N* N* NG HD HD NS NI NN NG NG T G A T A C C T G C A T C C T C C A A G T T
AV	v2	NN NS NG NI HD HD NG NN HD NI NG N* N* NG HD HD NS NI NN NG N* T G A T A C C T G C A T C C T C C A A G T C
AX	v1	NI N* NI NI NN N* NI NK HD HD HD NG NI NN HD NN HD NG T A C A A G C A G C C C T A G C G C T
AX	v2	NI N* NI NI NN N* NI NK HD HD HD NG NI NN ND N* HD NG T A C A A G C A G C C C T A G C C C T
AX	v3	NI N* NI NI NN N* NI NK HD HD HD NG NI NN ND NN ND NG T A C A A G C A G C C C T A G C G C T
AY	v1	HD HD NI HD HD HG HD NN ND ND HG NG NN HG NG ND HD NG T C C A C C T C G C C T T G T T C C T
AY	v2	HD HD HD HD HD NG HD NN HD ND NG NG NN HG ND ND HD NG T C C C C C T C G C C T T G T C C C T
AY	v3	HD HD NI HD HD HG HD NN ND ND NG NG NN HG HG ND HD NG T C C A C C T C G C C T T G T T C C T
AZ	v1	HH NN HD HD NG N* HD NN HD ND N* NG NS NS NN HD HD NG T C G C C T C C G C C C T A A G C C T
BB	v1	NS HG NS NG NS NN NG NN NG HD NN HD HD NG T A T A T A G T G T C G C C T
BB	v2	NS HG NI NG NS NN NG NN NG HD NN HD HD NG T A T A T A G T G T C G C C T
BB	v3	NI NG NI NG NS NN NG NN NG HD NN HD HD NG T A T A T A G T G T C G C C T
BB	v4	NN HG NI NG NI NN NG NN NG HD NN HD HD N* T G T A T A G T G T C G C C C

BC	v1	NN NG NI NN NG NN ND N* NG N* HN NN HD NN NI NN HD HD NG NG NG HD NG HD NG T G T A G T G C C T C G G C G A G C C T T T C T C T
BC	v2	NN NG NI NN NG NN ND N* NG N* HN NN HD NN NI NN HD HD NG NG NG HD NG T G T A G T G C C T C G G C G A G C C T T T C T
BC	v3	NN HG NI NN NG NN ND N* NG N* HN NN HD NN NI NN HD HD NG NG NG HD N* T G T A G T G C C T C G G C G A G C C T T T C C
BC	v4	NI HG NI NN NG NN ND N* NG N* HN NN HD HN NN NN HD HD NG NG NG HD NG T A T A G T G C C T C G G C G G G C C T T T C T
BC	v5	NS HG NI NG NS NN ND N* NG N* HN NN HD NS NI NN HD HD NG NG HG HD NG T A T A T A G C C T C G G C A A G C C T T T C T
BC	v6	NS HG NI NG NS NN ND N* NG N* HN NN HD NS NI NN HD HD NG HD HG HD NG T A T A T A G C C T C G G C A A G C C T C T C T
BD	v1	NN HD NI NI NG NN NN HD NN NI NI NN NN NN NG NI HD NG T G C A A T G G C G A A G G G T A C T
BD	v2	NN HD NI NI NG NN NN HD HH NI NI NN NN NN NG NI HD NG T G C A A T G G C C A A G G G T A C T
BE	v1	HD NN HD N* NS NN HG NI NI NS NI NG HD NN HD HD NG HG NG T C G C C A G T A A A A T C G C C T T T
BE	v2	HD NN HD N* NS NN NG NI NI NG HY NG HD NN HD HD NG HG NG T C G C C A G T A A T A T C G C C T T T
BF	v1	NN NN ND N* NS NS YG NI SN ND HD NG HD NS NN HD NG T G G C C A A T A G C C T C A G C T
BF	v2	NN NN ND N* NS NS YG NI SN HD HD NG HD NS NN HD NG T G G C C A A T A G C C T C A G C T
BF	v3	NN NN HD N* NS NS YG NI SN HD HD NG HD NS NN HD NG T G G C C A A T A G C C T C A G C T
BF	v4	NN NN HD N* NS SN YG NI SN HD HD NG HD NS NN HD NG T G G C C A G T A G C C T C A G C T
BG	v1	NI NN NI HD NN NG HD NN HD HG HD HG HG HD HD NG T A G A C G T C G C T C T T C C T
BG	v2	NI NN HN NN NI NG HD NN HD HG HD HG HG HD HD NG T A G G G A T C G C T C T T C C T
BI	v1	NN HD NI HD NN NG NG HD NN HD N* NN HD HD HD NG HD NN HD NS NI NG T G C A C G T T C G C C G C C C T C G C A A T
BI	v2	NN HD NI HD NN NG NG HD NN HD N* NN HD HD HD NG T G C A C G T T C G C C G C C C T
BI	v3	NN HD NI HD NS NG NG HD NN HD N* NN HD HD HD NG T G C A C A T T C G C C G C C C T
BI	v4	NN HD NI HD NI NG NG HD NN HD N* NN HD HD NN HD HD NG T G C A C A T T C G C C G C C G C C T
BL	v1	NI HG N* NN N* HD HG HD NN HD NN HG NG NS N* HD T A T C G C C T C G C G T T A C C
BL	v2	NI NG N* NN N* HD HG HD NN HD NN HG HG NS NG HD T A T C G C C T C G C G T T A T C
BL	v3	NI HG N* NN N* HD HG HD NN HD NK HG NN NN NG HD T A T C G C C T C G C G T G G T C
BN	v1	NI NN NI NI NG HD NG NN NI NN NI NN HD T A G A A T C T G A G A G C
BN	v2	NI NN NI NI NG HD NG NN NI NN NI NN NG T A G A A T C T G A G A G T
BO	v1	NN HG HD NG HD QD NI NH HG NI NN NN HD NS NG NN HD NG NG NG NG T G T C T C C A G T A G G C A T G C T T T T
BO	v2	NN HG HD NG HD HD NI NH HG NI NN NN HD NS NG NN HD NG NG NG NG T G T C T C C A G T A G G C A T G C T T T T
BO	v3	NN NG HD NG HD HD NI NH HG NI NN NN HD NS NG NN HD NG NG N* T G T C T C C A G T A G G C A T G C T T C
BR	v1	HD HD NN NN NG N* HD NI NG HD NG NS ND HA ND N* ND NN HD NN NN HD HD N* NN NG HD T C C G G T C C A T C T A C C C C C G C G G C C C G T C
BR	v2	HD HD NN NN NG N* HD NI NG HD NG NS HD HA ND N* ND NN HD NN NN HD HD N* NN NG HD T C C G G T C C A T C T A C C C C C G C G G C C C G T C
BR	v3	HD HD NN NN NG N* HD NI NG HD NG NS HD HA HD N* HD NN HD NN NN HD HD NG NN NG HD T C C G G T C C A T C T A C C C C C G C G G C C T G T C
BR	v4	HD HD NN NN NG N* HD NI NG HD NG NS ND N* HD N* HD NG HD NN NN HD HD NG NN NG HD T C C G G T C C A T C T A C C C C C T C G G C C T G T C
BU	v1	NI N* NI HG NI NI N* NG HD NN NI NG HD NN NS NN NG NG T A C A T A A C T C G A T C G A G T T
BU	v2	NI HD NI N* NI HG NI NI N* NG HD NN NI NG HD NN NS NN NG NG T A C A C A T A A C T C G A T C G A G T T

BU	v3	NS HD NI N* NI HG NI NI N* NG HD NN NI NG HD NN NS NN NG NG T A C A C A T A A C T C G A T C G A G T T
BU	v4	NS HD NI N* NI HG NI NI NG NG HD NN NI NG HD NN NS NN NI NN T A C A C A T A A T T C G A T C G A G A G
BU	v5	NS HD NI N* NI HG NI NI N* NG HD NN NI NG HD NN NS NN NG NN T A C A C A T A A C T C G A T C G A G T G
BV	v1	NN HD NI NI NN HA NN NS NI NI HD HA HA HA HD HD HD HA HD N* T G C A A G C G A A A C C C C C C C C C C
BV	v2	NN HD NI NI NN HA NN NS NS NI HD HA HA HA HD HD HD HA HD N* T G C A A G C G A A A C C C C C C C C C C
BV	v3	NN HD NI NI NN HA NN NS NS NI HD HA HA HA HD HD HD HA HD N* T G C A A G C G A A A C C C C C C C C C C
BV	v4	NN HD NI NI NN HA NN NS NS NI HD HA HA HA HD HD HD NI HD N* T G C A A G C G A A A C C C C C C C C A C C
BW	v1	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 N* T A T A T G T G T C G G T C G A G C C T G G C C C C C
BW	v2	NI NG NI NG NN NG HH NG HD NN NN HG HD NN NS NN HD HD NG NA NN HD HD HD HD N* T A T A T G T C T C G G T C G A G C C T G G C C C C C
BW	v3	NI HG NS NG NN NG HH NG HD NN NN HG HD NN NS NN HD HD NG NA NN HD HD HD HD N* T A T A T G T C T C G G T C G A G C C T G G C C C C C
BY	v1	HD HD NC HD NC NG HD HH HD NI NG N* NS N* HD HD NS NI HG HD NG T C C C C C T C C C A T C A C C C A A T C T
BY	v2	HD HD HD HD HD NG HD HH HD NI NG N* NS N* HD HD NS NI HG HD NG T C C C C C T C C C A T C A C C C A A T C T
CB	v1	NN HD NS NS NN HA NN NS NS NI HD HD HD HA HD NN ND NN HD N* T G C A A G C G A A A C C C C C G C G C C
CB	v2	NS HD NI NS NN HA NN NS NS NI HD HD HD HA HD NN HD NN HD N* T A C A A G C G A A A C C C C C G C G C C
CC	V1	NI HD N* HG NG HG HD HD N* HD H* T A C C T T T C C C C T
CC	V2	NI HD HD N* HD H* T A C C C C T
CD	v1	HD HG N* NN NG HD NN HD NN NG NN HD N* NG NG T C T C G T C G C G T G C C T T
CD	v2	HD HG N* NN HG HD NN HD NN NG NN HD NN HD NG T C T C G T C G C G T G C G C T
CE	v1	NN HD NI N* NI NI NK HD HD HD HG NI NN HD HD T G C A C A A G C C C T A G C C
CE	v2	NN HD NI N* NI NI NK HD HD HD NG HD NN HD HD T G C A C A A G C C C T C G C C
CE	v3	NN HD NI N* NI NI NK HD HD HD NG NI NN HD HD T G C A C A A G C C C T A G C C
CF	v1	NN NG NI N* NS NG NN NN HG ND ND HD NG T G T A C A T G G T C C C T
CF	v2	NN NG NS HD NI NG NN NN HG ND ND HG NG T G T A C A T G G T C C T T
CG	v1	HD HG N* NN NG NN NI NN NI NN NN HD HG NG HD NG T C T C G T G A G A G G C T T C T
CG	v2	HD HD NN NN NN NI NN NI NN NN HD HG NG T C C G G G A G A G G C T T
CH	v1	HD HD NN NN NG N* HD NI NG HD NG NS ND HA HD N* HH NG NI NN NN HD HD NG NN NG HD T C C G G T C C A T C T A C C C C C T A G G C C T G T C
CH	v2	HD HD NN NN NG N* HD NI NG HD NG NS ND HA HD N* HH NG NI NN NN HD HD NG HD T C C G G T C C A T C T A C C C C C T A G G C C T C
CI	v1	HD HD NN NN NG N* HD NI NG HD NG NG NS ND HA HD N* N* NN NI NN NN HD NG T C C G G T C C A T C T T A C C C C C G A G G C T
CI	v2	HD HD NN NN NG N* HD NI NG HD NG NG NS ND HA HD N* HD NG NI NN NN HD HD NG HD T C C G G T C C A T C T T A C C C C C T A G G C C T C
CJ	v1	NI HG NI NG HN NG NN HD ND NG HD NI HG NG HD HH N* HH N* HD NN NG HD HD NN NN NG NN NI NV T A T A T G T G C C T C A T T C C C C C C G T C C G G T G A A
CK	v1	NI HG N* NI SN HD HD HD NN NI NK HD NN NN HD NG T A T C A G C C C G A G C G G C T
CK	v2	NI HG N* NI SN HD HD HD NN HD NK HD HD NN HD NG T A T C A G C C C G C G C C G C T
CK	v3	NI NG N* NI SN HD HD HD NN HD NK HD NN HD HD N* T A T C A G C C C G C G C G C C C
CL	v1	NI HG NI N* HN NG NN ND HD NG HD NI HG HG HD NN HD HH N* HD NN NN NG HD NN NN NG NN NI NG T A T A C G T G C C T C A T T C G C C C C G G T C G G T G A T

CL	**v2**	NI HG NI NG HN NG NN ND HD NG HD NI HG HG HD NN HD HH N* HD NN NN NG HD NN NN NG NN NI NG T A T A T G T G C C T C A T T C G C C C C G G T C G G T G A T
CL	**v3**	NI HG NI HG HN NG NN HD ND NG HD NI HG HG HD NN HD HH N* HD NN NN NG HD NN NN NG NN NI NV T A T A T G T G C C T C A T T C G C C C C G G T C G G T G A A
CM	**V1**	NI ND HG HG HG HG NG HG HG HD HD HD HD NN NN HD HG HH H* H* NN HD H* T A C T T T T T T T C C C C G G C T C T T G C T
CM	**V2**	NI ND HG HG HG NG HG HG HD HD HD NN NN HD HG HH H* H* NN HD H* T A C T T T T T T T C C C G G C T C T T G C T
CM	**V3**	NS ND HG HG HG NG HG HG HD HD HD NN NN HD HG HH H* H* NN HD H* T A C T T T T T T T C C C G G C T C T T G C T
CN	**v1**	NS HD N* NG HY NG NS NG HD N* T A C C T A T A T C C
CO	**v1**	NI HG HG HG HG NG HD HD HD HD HD NG N* NI NN HD HD H* T A T T T T T C C C C C T C A G C C T
CP	**v1**	HD HG HD NI NN ND HD HD NN NN NI HD T C T C A G C C C G G A C
CQ	**v1**	NN NI HN NG NN NN NG NI N* NI NG NN NN HG HD HD NG T G A G T G G T A C A T G G T C C T
CR	**v1**	NN NN HD NI NG HD HD HD NG NI HD HD NG T G G C A T C C C T A C C T
CS	**v1**	NS HG NI NG NS NN ND N* NG N* HN NN HD NS NS NN NG NN NG HD NN HD HD NG T A T A T A G C C T C G G C A A G T G T C G C C T

The optimal target sequences were predicted using AnnoTALE v1.2 (Grau *et al.*, 2016). Truncated or fragmentend TALEs were not taken into account.



Appendix Figure 3.2.1: TALE cluster affiliation in the 10 completely sequenced *Xoc* strains with regard to the newly assigned TALE class versions. The figure is equivalent to Figure 3.2.1 but all TALE classes are additionally subgrouped into different class versions. Each version respresents a unique RVD sequence that is shared by all other class members with the same version number.

Appendix Table 3.2.3: List of all unique *Xoc* TALE class versions.

TALE class	Version	Cluster affiliation	RVD sequence [1 2]	TALE class presence										# in *Xoc*	# in *Xoo*	pSKA2 construct	*A. tumefaciens*	pSKX1 construct	*X. oryzae*
				BLS79	B8-12	RS105	L8	BLS256	CFBP2286	BXOR1	CFBP7342	CFBP7331	CFBP7341						
AD [3]	v1	VII	NN HD NS NG HD NN HD NI HD NN HD NN HD NN HD NN NN NN NN NN NN NN HD NG	x	x	x	x	x	x	-	-	-	-	6	-	-	-	-	-
	v2	VII	NN HD NS NG HD NN N* NI HD NS HD NN HD NN HD NN NN NN NN NN NN NN HD NG	-	-	-	-	-	-	x	-	-	-	1	12	x	x	x	x
	v3	-	NN HD NS NG HD NN N* NI HD NS HD NN HD NN NN NN NN NN NN NN NN NN HD NG	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
AF [3]	v1	IV	NI NN NI NI NI NI N* NI NG NN NN NN NI NS NG NG	-	-	-	-	x	x	-	-	-	-	2	-	x	x	-	-
	v2	IV	NI NN NI NI NI NI N* NI NG NN NN NN NG NS NG HD	-	x	-	x	-	-	-	-	-	-	2	-	-	-	x	x
	v3	-	NI NN NN NI NI NI HD NS HG NN NN NN NI NI NG HD	-	-	-	-	-	-	-	-	-	-	-	7	x	x	x	x
	v4	-	NI NN NN NI NI NI HD NS HG NN NN NN NI NI HG HD	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
	v5	-	NI NN NN NI NI NS HD NS HG NN NN NN NI NI NG HD	-	-	-	-	-	-	-	-	-	-	-	3	-	-	-	-
AJ	v1	XII	NN HD NI NK HD HD HD NG NI NN HD HD NG	-	-	-	-	x	-	-	-	-	-	1	-	x	x	x	x
AK	v1	II	HD HG HD HG N* NN NG HD NN HD NG NG NN HD N* NG NG	x	x	x	x	x	x	x	-	-	-	6	-	x	x	x	x
	v2	XII	HD HD HD HD HD NG HD NN HD NG HG NN HD N* NG NG	x	x	x	x	x	x	x	-	-	-	6	-	x	x	x	-
AT	v1	IX	NN HD HD HH NN NG NG NG HD NI NG HD HH NG NG NG NG	-	-	-	-	-	-	-	x	-	-	1	-	x	x	x	x
	v2	IX	NN HD HD HH NN NG NS NG HD NI NG HD HH NG NG NG NG	-	-	-	-	-	-	-	-	x	x	2	-	x	x	-	-
	v3	IX	NN HD HD HH NN NG NS NG HD NI NG HD HH NG NG NG HA NG	x	x	x	x	x	x	x	-	-	-	7	-	x	x	-	-
AU	v1	IV	HD HD NG NI HG HD HG NI NI NN HD HG HG HG NG NI NG	-	-	-	-	-	x	-	-	-	-	1	-	x	x	x	x
	v2	IV	HD HD NG NI HG HD HG NI HD NN HD HG HG HG NG NI NG	-	-	-	-	x	-	-	-	-	-	1	-	x	x	x	x
AV	v1	III	NN NS NG NI HD HD NG NN HD NI NG N* N* NG HD HD NS NI NN NG NG	x	x	x	x	x	x	-	-	-	-	6	-	-	-	-	-
	v2	III	NN NS NG NI HD HD NG NN HD NI NG N* N* NG HD HD NS NI NN NG N*	-	-	-	-	-	-	x	x	x	x	4	-	x	x	x	-
AX	v1	V	NI N* NI NI NN N* NI NK HD HD HD NG NI NN HD NN HD NG	x	x	x	x	x	x	x	-	-	-	7	-	x	x	x	x
	v2	V	NI N* NI NI NN N* NI NK HD HD HD NG NI NN ND N* HD NG	-	-	-	-	-	-	-	x	-	-	1	-	x	x	x	x
	v3	V	NI N* NI NI NN N* NI NK HD HD HD NG NI NN ND NN ND NG	-	-	-	-	-	-	-	-	x	x	2	-	-	-	-	-
AY	v1	II	HD HD NI HD HD HG HD NN ND ND HG NG NN HG NG ND HD NG	-	-	-	-	-	-	-	-	x	x	2	-	x	x	-	-
	v2	II	HD HD HD HD HD NG HD NN HD ND NG NG NN HG ND ND HD NG	x	x	x	x	x	x	x	-	-	-	7	-	x	x	x	x
	v3	II	HD HD NI HD HD HG HD NN ND ND NG NG NN HG HG ND HD NG	-	-	-	-	-	-	-	x	-	-	1	-	-	-	-	-
AZ	v1	VII	HH NN HD HD NG N* HD NN HD ND N* NG NS NS NN HD HD NG	x	x	x	x	x	x	x	x	-	-	8	-	x	x	x	x
BB	v1	VIII	NS HG NS NG NS NN NG NN NG HD NN HD HD NG	-	-	-	x	-	-	-	-	-	-	1	-	x	x	-	-
	v2	VIII	NS HG NI NG NS NN NG NN NG HD NN HD HD NG	x	x	x	-	x	-	-	-	-	-	4	-	x	x	-	-
	v3	VIII	NI NG NI NG NS NN NG NN NG HD NN HD HD NG	-	-	-	-	-	x	-	-	-	-	1	-	x	x	x	x
	v4	VIII	NN HG NI NG NI NN NG NN NG HD NN HD HD N*	-	-	-	-	-	-	x	-	-	-	1	-	-	-	-	-
BC	v1	VIII	NN NG NI NN NG NN ND N* NG N* HN NN HD NN NI NN HD HD NG NG NG HD NG HD NG	-	-	-	-	-	-	-	-	x	-	1	-	x	x	-	-
	v2	VIII	NN NG NI NN NG NN ND N* NG N* HN NN HD NN NI NN HD HD NG NG NG HD NG	-	-	-	-	-	-	-	-	-	x	1	-	x	x	x	x
	v3	VIII	NN HG NI NN NG NN ND N* NG N* HN NN HD NN NI NN HD HD NG NG NG HD N*	-	-	-	-	-	-	x	-	-	-	1	-	x	x	-	-

TALE class	Version	Cluster affiliation	RVD sequence [1,2]	TALE class presence										# in *Xoc*	# in *Xoo*	pSKA2 construct	*A. tumefaciens*	pSKX1 construct	*X. oryzae*
				BLS79	B8-12	RS105	L8	BLS256	CFBP2286	BXOR1	CFBP7342	CFBP7331	CFBP7341						
BC	v4	VIII	NI HG NI NN NG NN ND N* NG N* HN NN HD HN NN NN HD HD NG NG NG HD NG	-	-	-	-	-	-	-	x	-	-	1	-	-	-	-	-
	v5	VIII	NS HG NI NG NS NN ND N* NG N* HN NN HD NS NI NN HD HD NG NG HG HD NG	x	x	x	x	x	-	-	-	-	-	5	-	x	x	-	-
	v6	VIII	NS HG NI NG NS NN ND N* NG N* HN NN HD NS NI NN HD HD NG HD HG HD NG	-	-	-	-	-	x	-	-	-	-	1	-	x	x	x	x
BD	v1	V	NN HD NI NI NG NN NN HD NN NI NI NN NN NN NG NI HD NG	x	x	x	x	x	x	x	-	-	-	7	-	x	x	x	x
	v2	V	NN HD NI NI NG NN NN HD HH NI NI NN NN NN NG NI HD NG	-	-	-	-	-	-	-	x	x	x	3	-	x	x	x	x
BE	v1	II	HD NN HD N* NS NN HG NI NI NS NI NG HD NN HD HD NG HG NG	x	x	x	x	x	x	-	-	-	-	6	-	x	x	x	x
	v2	II	HD NN HD N* NS NN NG NI NI NG HY NG HD NN HD HD NG HG NG	-	-	-	-	-	-	x	-	x	x	3	-	x	x	-	-
BF	v1	IV	NN NN ND N* NS NS YG NI SN ND HD NG HD NS NN HD NG	-	x	x	x	-	-	-	-	-	-	3	-	x	x	-	-
	v2	IV	NN NN ND N* NS NS YG NI SN HD HD NG HD NS NN HD NG	x	-	-	-	x	x	-	-	-	-	3	-	x	x	x	-
	v3	IV	NN NN HD N* NS NS YG NI SN HD HD NG HD NS NN HD NG	-	-	-	-	-	-	x	-	-	-	1	-	x	x	-	-
	v4	IV	NN NN HD N* NS SN YG NI SN HD HD NG HD NS NN HD NG	-	-	-	-	-	-	-	x	x	x	3	-	-	-	-	-
BG	v1	III	NI NN NI HD NN NG HD NN HD HG HD HG HG HD HD NG	x	x	x	x	-	-	-	-	-	-	4	-	x	x	x	x
	v2	III	NI NN HN NN NI NG HD NN HD HG HD HG HG HD HD NG	-	-	-	-	x	x	-	-	-	-	2	-	x	x	-	-
BI	v1	V	NN HD NI HD NN NG NG HD NN HD N* NN HD HD HD NG HD NN HD NS NI NG	-	-	-	-	-	-	-	-	x	x	2	-	x	x	-	-
	v2	V	NN HD NI HD NN NG NG HD NN HD N* NN HD HD HD NG	x	x	x	x	x	x	-	-	-	-	6	-	x	x	x	x
	v3	V	NN HD NI HD NS NG NG HD NN HD N* NN HD HD HD NG	-	-	-	-	-	-	x	-	-	-	1	-	x	x	-	-
	v4	V	NN HD NI HD NI NG NG HD NN HD N* NN HD HD NN HD HD NG	-	-	-	-	-	-	-	x	-	-	1	-	x	x	-	-
BL	v1	IV	NI HG N* NN N* HD HG HD NN HD NN HG NG NS N* HD	x	x	-	x	-	-	-	-	-	-	3	-	x	x	x	x
	v2	IV	NI NG N* NN N* HD HG HD NN HD NN HG HG NS NG HD	-	-	-	-	-	-	x	x	x	x	4	-	x	x	-	-
	v3	IV	NI HG N* NN N* HD HG HD NN HD NK HG NN NN NG HD	-	-	-	-	x	x	-	-	-	-	2	-	x	x	-	-
BN	v1	XI	NI NN NI NI NG HD NG NN NI NN NI NN HD	x	x	x	x	x	x	x	-	-	-	7	-	x	x	x	x
	v2	III	NI NN NI NI NG HD NG NN NI NN NI NN NG	-	-	-	-	-	-	-	x	-	-	1	-	x	x	x	x
BO	v1	X	NN HG HD NG HD QD NI NH HG NI NN NN HD NS NG NN HD NG NG NG NG	-	-	-	-	-	-	-	-	x	x	2	-	-	-	-	-
	v2	X	NN HG HD NG HD HD NI NH HG NI NN NN HD NS NG NN HD NG NG NG NG	-	-	-	-	-	-	x	x	-	-	2	-	x	x	x	x
	v3	X	NN NG HD NG HD HD NI NH HG NI NN NN HD NS NG NN HD NG NG N*	x	x	x	x	x	x	-	-	-	-	6	-	x	x	-	-
BR	v1	IV	HD HD NN NN NG N* HD NI NG HD NG NS ND HA ND N* ND NN HD NN NN HD HD N* NN NG HD	x	-	-	-	-	-	-	-	-	-	1	-	x	x	x	-
	v2	IV	HD HD NN NN NG N* HD NI NG HD NG NS HD HA ND N* ND NN HD NN NN HD HD N* NN NG HD	-	-	-	-	x	x	-	-	-	-	2	-	-	-	-	-
	v3	IV	HD HD NN NN NG N* HD NI NG HD NG NS HD HA HD N* HD NN HD NN NN HD HD NG NN NG HD	-	x	-	x	-	-	-	-	-	-	2	-	-	-	-	-
	v4	IV	HD HD NN NN NG N* HD NI NG HD NG NS ND N* HD N* HD NG HD NN NN HD HD NG NN NG HD	-	-	-	-	-	-	x	-	-	-	1	-	-	-	-	-
BU	v1	I	NI N* NI HG NI NI N* NG HD NN NI NG HD NN NS NN NG NG	-	-	-	-	-	-	-	-	x	-	1	-	x	x	x	x
	v2	I	NI HD NI N* NI HG NI NI N* NG HD NN NI NG HD NN NS NN NG NG	-	-	-	-	-	-	-	x	-	x	2	-	x	x	-	-
	v3	I	NS HD NI N* NI HG NI NI N* NG HD NN NI NG HD NN NS NN NG NG	-	-	-	-	-	-	x	-	-	-	1	-	x	x	-	-
	v4	I	NS HD NI N* NI HG NI NI NG NG HD NN NI NG HD NN NS NN NI NN	x	-	-	-	-	-	-	-	-	-	1	-	-	-	x	x
	v5	I	NS HD NI N* NI HG NI NI N* NG HD NN NI NG HD NN NS NN NG NN	-	x	x	x	x	x	-	-	-	-	5	-	-	-	-	-

TALE class	Version	Cluster affiliation	RVD sequence [1,2]	TALE class presence										# in *Xoc*	# in *Xoo*	pSKA2 construct	*A. tumefaciens*	pSKX1 construct	*X. oryzae*
				BLS79	B8-12	RS105	L8	BLS256	CFBP2286	BXOR1	CFBP7342	CFBP7331	CFBP7341						
BV	v1	VI	NN HD NI NI NN HA NN NS NI NI HD HA HA HA HD HD HD HA HD N*	-	-	-	-	-	x	-	-	-	-	1	-	x	x	x	x
	v2	VI	NN HD NI NI NN HA NN NS NS NI HD HA HA HA HD HD HD HA HD N*	-	-	-	-	x	-	-	-	-	-	1	-	x	x	-	-
	v3	VI	NN HD NI NI NN HA NN NS NS NI HD HA HA HA HD HD HD HA HD N*	x	x	x	x	-	-	-	-	-	-	4	-	x	x	-	-
	v4	VI	NN HD NI NI NN HA NN NS NS NI HD HA HA HA HD HD HD NI HD N*	-	-	-	-	-	-	x	-	-	-	1	-	x	x	-	-
BW	v1	VIII	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 N*	x	x	x	x	x	-	-	-	-	-	5	-	x	x	x	-
	v2	VIII	NI NG NI NG NN NG HH NG HD NN NN HG HD NN NS NN HD HD NG NA NN HD HD HD HD N*	-	-	-	-	-	-	x	-	x	x	3	-	x	x	x	x
	v3	VIII	NI HG NS NG NN NG HH NG HD NN NN HG HD NN NS NN HD HD NG NA NN HD HD HD HD N*	-	-	-	-	-	-	-	x	-	-	1	-	x	x	-	-
BY	v1	IV	HD HD NC HD NC NG HD HH HD NI NG N* NS N* HD HD NS NI HG HD NG	-	x	x	x	x	x	-	-	-	-	5	-	x	x	x	x
	v2	IV	HD HD HD HD HD NG HD HH HD NI NG N* NS N* HD HD NS NI HG HD NG	x	-	-	-	-	-	-	-	-	-	1	-	x	x	-	-
CB	v1	VI	NN HD NS NS NN HA NN NS NS NI HD HD HD HA HD NN ND NN HD N*	-	-	-	-	-	-	-	x	-	-	1	-	-	-	-	-
	v2	VI	NS HD NI NS NN HA NN NS NS NI HD HD HD HA HD NN HD NN HD N*	-	-	-	-	-	-	-	-	x	x	2	-	x	x	x	x
CC¥	v1	IV	NI HD N* HG NG HG HD HD N* HD H*	-	-	-	-	-	-	-	x			1	-	-	-	-	-
	v2	IV	NI HD HD N* HD H*	-	-	-	-	-	-	-	-	x	x	2	-	-	-	-	-
CD	v1	II	HD HG N* NN NG HD NN HD NN NG NN HD N* NG NG	-	-	-	-	-	-	x	-	-	-	1	-	x	x	-	-
	v2	II	HD HG N* NN HG HD NN HD NN NG NN HD NN HD NG	-	-	-	-	-	-	-	x	-	-	1	-	-	-	-	-
CE	v1	IV	NN HD NI N* NI NI NK HD HD HD HG NI NN HD HD	-	x	x	x	-	-	-	-	-	-	3	-	x	x	x	x
	v2	IV	NN HD NI N* NI NI NK HD HD HD NG HD NN HD HD	-	-	-	-	-	-	-	x	-	-	1	-	x	x	-	-
	v3	IV	NN HD NI N* NI NI NK HD HD HD NG NI NN HD HD	-	-	-	-	-	-	x	-	-	-	1	-	-	-	-	-
CF	v1	XI	NN NG NI N* NS NG NN NN HG ND ND HD NG	-	-	-	-	-	-	-	-	-	x	1	-	x	x	-	-
	v2	XI	NN NG NS HD NI NG NN NN HG ND ND HG NG	-	-	-	-	-	-	-	x	-	-	1	-	x	x	x	x
CG	v1	IV	HD HG N* NN NG NN NI NN NI NN NN HD HG NG HD NG	-	-	-	-	-	x	x	-	-	-	2	-	x	x	x	x
	v2	IV	HD HD NN NN NN NI NN NI NN NN HD HG NG	-	-	-	-	-	-	-	x	-	-	1	-	x	x	x	x
CH	v1	IV	HD HD NN NN NG N* HD NI NG HD NG NS ND HA HD N* HH NG NI NN NN HD HD NG NN NG HD	-	-	-	-	-	-	-	-	x	x	2	-	x	x	-	-
	v2	IV	HD HD NN NN NG N* HD NI NG HD NG NS ND HA HD N* HH NG NI NN NN HD HD NG HD	-	-	-	-	-	-	-	x	-	-	1	-	x	x	x	x
CI	v1	VII	HD HD NN NN NG N* HD NI NG HD NG NG NS ND HA HD N* N* NN NI NN NN HD NG	-	-	-	-	-	-	-	-	x	-	1	-	x	x	x	x
	v2	VII	HD HD NN NN NG N* HD NI NG HD NG NG NS ND HA HD N* HD NG NI NN NN HD HD NG HD	-	-	-	-	-	-	-	x	-	-	1	-	x	x	-	-
CJ	v1	IV	NI HG NI NG HN NG NN HD ND NG HD NI HG NG HD HH N* HH N* HD NN NG HD HD NN NN NG NN NI NV	-	-	-	-	-	-	-	x	x	x	3	-	-	-	x	x
CK	v1	IV	NI HG N* NI SN HD HD HD NN NI NK HD NN NN HD NG	-	x	-	x	-	-	-	-	-	-	2	-	x	x	x	x
	v2	IV	NI HG N* NI SN HD HD HD NN HD NK HD HD NN HD NG	-	-	-	-	-	x	-	-	-	-	1	-	x	x	x	x
	v3	IV	NI NG N* NI SN HD HD HD NN HD NK HD NN HD HD N*	x	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
CL	v1	IV	NI HG NI N* HN NG NN ND HD NG HD NI HG HG HD NN HD HH N* HD NN NN NG HD NN NN NG NN NI NG	-	-	-	-	x	-	-	-	-	-	1	-	x	x	x	x
	v2	IV	NI HG NI NG HN NG NN ND HD NG HD NI HG HG HD NN HD HH N* HD NN NN NG HD NN NN NG NN NI NG	x	x	x	x	-	x	-	-	-	-	5	-	x	x	-	-
	v3	IV	NI HG NI HG HN NG NN HD ND NG HD NI HG HG HD NN HD HH N* HD NN NN NG HD NN NN NG NN NI NV	-	-	-	-	-	-	x	-	-	-	1	-	x	x	x	x

TALE class	Version	Cluster affiliation	RVD sequence [1,2]	TALE class presence										# in *Xoc*	# in *Xoo*	pSKA2 construct	*A. tumefaciens*	pSKX1 construct	*X. oryzae*
				BLS79	B8-12	RS105	L8	BLS256	CFBP2286	BXOR1	CFBP7342	CFBP7331	CFBP7341						
CM¥	v1	IV	NI ND HG HG HG HG NG HG HG HD HD HD HD NN NN HD HG HH H* H* NN HD H*	-	x	x	-	-	-	-	-	-	-	2	-	-	-	-	-
	v2	IV	NI ND HG HG HG NG HG HG HD HD HD NN NN HD HG HH H* H* NN HD H*	x	-	-	x	-	-	-	-	-	-	2	-	-	-	-	-
	v3	IV	NS ND HG HG HG NG HG HG HD HD HD NN NN HD HG HH H* H* NN HD H*	-	-	-	-	x	x	-	-	-	-	2	-	-	-	-	-
CN¥	v1	XII	NS HD N* NG HY NG NS NG HD N*	-	-	-	-	-	-	x	-	-	-	1	-	-	-	-	-
CO¥	v1	IX	NI HG HG HG HG NG HD HD HD HD HD NG N* NI NN HD HD H*	-	-	-	-	-	-	x	-	-	-	1	-	-	-	-	-
CP	v1	IV	HD HG HD NI NN ND HD HD NN NN NI HD	-	-	-	-	-	-	x	-	x	x	3	-	x	x	x	x
CQ	v1	XI	NN NI HN NG NN NN NG NI N* NI NG NN NN HG HD HD NG	-	-	-	-	-	-	x	-	-	-	1	-	x	x	x	x
CR	v1	III	NN NN HD NI NG HD HD HD NG NI HD HD NG	-	-	-	-	-	-	-	-	x	x	2	-	x	x	x	x
CS	v1	VIII	NS HG NI NG NS NN ND N* NG N* HN NN HD NS NS NN NG NN NG HD NN HD HD NG	-	-	-	x	-	-	-	-	-	-	1	-	x	x	x	x

¥ TALEs with this symbol contain truncated N- and C-terminal regions in their natural form and were not cloned or analysed in further detail

[1] Repeats of aberrant length are underlined, TalBG members contain the variant with 30aa, TalBV members the variant 36aa v2

[2] RVD positions highlighted in yellow differ from the ones in the respective naturally occurring TALE but they contain a RVD with identical specificity (NG instead of HG, HD instead of ND or vice versa)

[3] Constructs for the *Xoo* TALEs TalAD v2 and TalAF v3 were obtained by S. Mücke (Mücke *et al*., 2019)

Appendix Table 3.2.4: Novel repeat modules generated in this study.

RVD	repeat position					
	1	2	3	4	5	6
HH	x	x	x	x	x	x
HA	x	x	x	x	x	x
ND	x	x	x	x	x	x
HM		x				
AN			x			
NA		x				
NC			x		x	
HN					x	
HY					x	
SN			x			
YG	x					

Appendix Table 3.2.5: Published *Xoc* TALE target genes and their expression values after infection with the 10 completely sequenced *Xoc* strains.

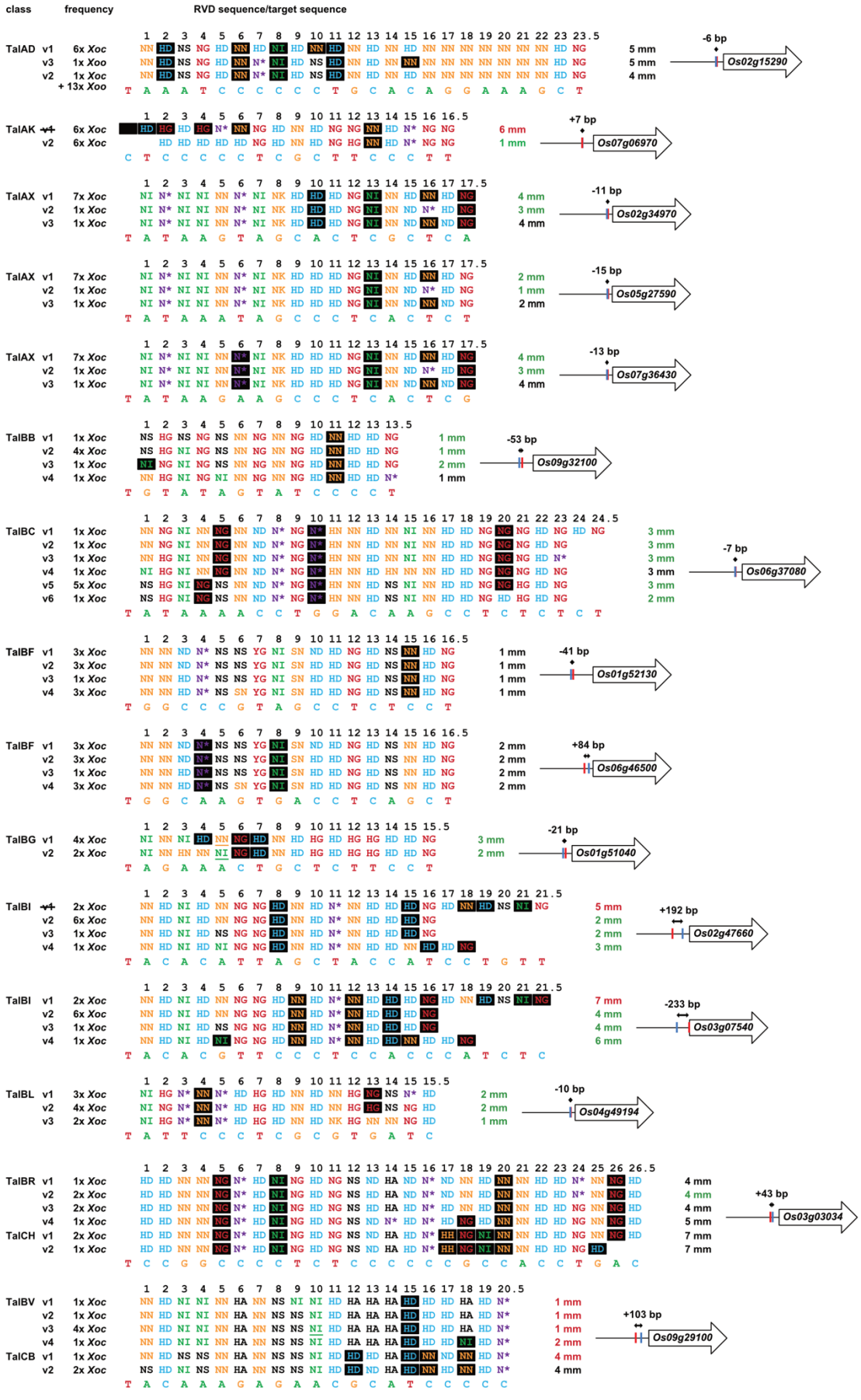
</= 1
</= 2-10
</= 11-40
</= 41-80
</= 81-160
> 161

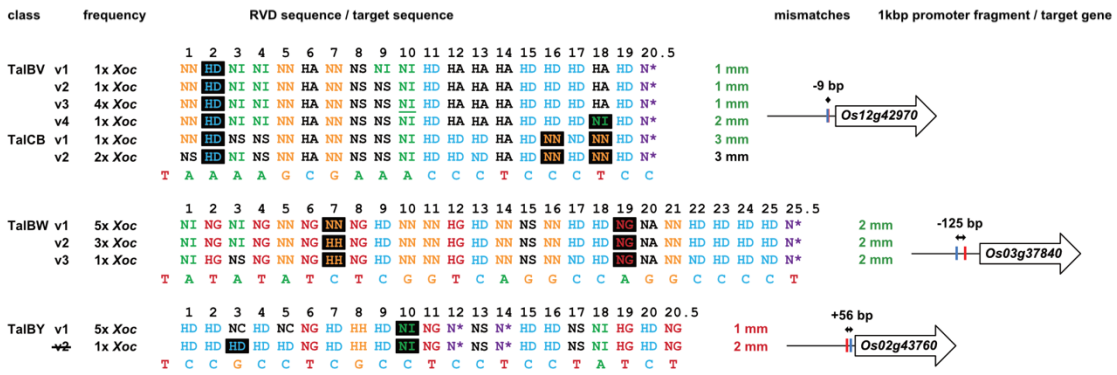
Locus ID	protein name or annotated function	original TALE name (BLS256)	TALE class member	*Xoc* strain / TALE presence [1] fold change [2]									
				BLS79	B8-12	RS105	L8	BLS256	CFBP2286	BXOR1	CFBP7342	CFBP7331	CFBP7341
Os02g15290	OsVQ7	Tal5a	AD	v1	v1	v1	v1	v1	v1	v2			
Os07g06970	OsHEN1	Tal1c	AK	v2	v2	v2	v2	v2	v2				
Os02g34970	no apical meristem protein	Tal3b	AX	v1	v1	v1	v1	v1	v1	v1	v2	v3	v4
Os05g27590	wound-induced protein	Tal3b	AX	v1	v1	v1	v1	v1	v1	v1	v2	v3	v4
Os07g36430	expressed protein	Tal3b	AX	v1	v1	v1	v1	v1	v1	v1	v2	v3	v4
Os09g32100	expressed protein	Tal4b	BB	v2	v2	v2	v1	v2	v3	v4			
Os06g37080	OsAAO1	Tal4c	BC	v5	v5	v5	v5	v5	v6	v3	v4	v1 *	v2
Os01g52130	OsSULTR3;6	Tal2g	BF	v2	v1	v1	v1	v2	v2	v3	v3	v3	v3
Os06g46500	monocopper oxidase	Tal2g	BF	v2	v1	v1	v1	v2	v2	v3	v3	v3	v3
Os01g51040	16k transmembrane protein	Tal9b	BG	v1	v1	v1	v1	v2	v2				
Os02g47660	OsbHLH079	Tal3c	BI	v2	v2	v2	v2	v2	v2	v3	v4 *	~~v1 *~~	~~v1 *~~
Os03g07540	OsbHLH153	Tal3c	BI	v2	v2	v2	v2	v2	v2	v3	v4 *	v1 *	v1 *
Os04g49194	OsDOX-1	Tal2d	BL	v1	v1		v1	v3	v3	v2	v2	v2	v2
Os03g03034	OsDOX-2	Tal2c	BR CH	v1	v3		v3	v2	v2	v4	v2 *	v1	v1
Os09g29100	Cyclin-4D-1	Tal6	BV	v3	v3	v3	v3	v2	v1	v4			
Os12g42970	GATA zinc finger	Tal6	BV CB	v3	v3	v3	v3	v2	v1	v4	v1	v2	v2
Os03g37840	OsHAK16	Tal4a	BW	v1	v1	v1	v1	v1		v2	v3	v2	v2
Os02g43760	ubiquitin hydrolase	Tal2a	BY	~~v2~~	v1	v1	v1	v1	v1				

The TALE targets listed in this table were initially discovered due to a correlation analysis that compared their upregulation with the presence of absence of TALEs from the *Xoc* strain BLS256 (Cernadas *et al*., 2014), consequently, the original names of these TALEs in BLS256 are indicated as well as their TALE class affiliation according to the AnnoTALE classification system (Grau *et al*., 2016). The shown expression values are derived from Wilkins *et al*., 2015.

[1] Indicated are all unique TALE class versions.

[2] Shown are the fold change values of each gene in combination with one of the 10 analysed *Xoc* strains. The different shades of white, grey and black correspond to the expression of the respective gene. Yellow cells indicate that no reliable data was available for this gene.


class
frequency
RVD sequence/target sequence
TalAD v1 6x Xoc
v3 1x Xoo
v2 1x Xoc
+ 13x Xoo
T A A A T C C C C C C T G C A C A G G A A A G C T
5 mm
5 mm
4 mm
-6 bp
Os02g15290
TalAK v4 6x Xoc
v2 6x Xoc
C T C C C C C T C G C T T C C C T T
6 mm
1 mm
+7 bp
Os07g06970
TalAX v1 7x Xoc
v2 1x Xoc
v3 1x Xoc
T A T A A G T A G C A C T C G C T C A
4 mm
3 mm
4 mm
-11 bp
Os02g34970
TalAX v1 7x Xoc
v2 1x Xoc
v3 1x Xoc
T A T A A A T A G C C C T C A C T C T
2 mm
1 mm
2 mm
-15 bp
Os05g27590
TalAX v1 7x Xoc
v2 1x Xoc
v3 1x Xoc
T A T A A G A A G C C C T C A C T C G
4 mm
3 mm
4 mm
-13 bp
Os07g36430
TalBB v1 1x Xoc
v2 4x Xoc
v3 1x Xoc
v4 1x Xoc
T G T A T A G T A T C C C C T
1 mm
1 mm
2 mm
1 mm
-53 bp
Os09g32100
TalBC v1 1x Xoc
v2 1x Xoc
v3 1x Xoc
v4 1x Xoc
v5 5x Xoc
v6 1x Xoc
T A T A A A A C C T G G A C A A G C C T C T C T C T
3 mm
3 mm
3 mm
3 mm
3 mm
2 mm
-7 bp
Os06g37080
TalBF v1 3x Xoc
v2 3x Xoc
v3 1x Xoc
v4 3x Xoc
T G G C C C G T A G C C T C T C C T
1 mm
1 mm
1 mm
1 mm
-41 bp
Os01g52130
TalBF v1 3x Xoc
v2 3x Xoc
v3 1x Xoc
v4 3x Xoc
T G G C A A G T G A C C T C A G C T
2 mm
2 mm
2 mm
2 mm
+84 bp
Os06g46500
TalBG v1 4x Xoc
v2 2x Xoc
T A G A A A C T G C T C T T C C T
3 mm
2 mm
-21 bp
Os01g51040
TalBI v4 2x Xoc
v2 6x Xoc
v3 1x Xoc
v4 1x Xoc
T A C A C A T T A G C T A C C A T C C T G T T
5 mm
2 mm
2 mm
3 mm
+192 bp
Os02g47660
TalBI v1 2x Xoc
v2 6x Xoc
v3 1x Xoc
v4 1x Xoc
T A C A C G T T C C C T C C A C C C A T C T C
7 mm
4 mm
4 mm
6 mm
-233 bp
Os03g07540
TalBL v1 3x Xoc
v2 4x Xoc
v3 2x Xoc
T A T T C C C T C G C G T G A T C
2 mm
2 mm
1 mm
-10 bp
Os04g49194
TalBR v1 1x Xoc
v2 2x Xoc
v3 2x Xoc
v4 1x Xoc
TalCH v1 2x Xoc
v2 1x Xoc
T C C G G C C C C T C T C C C C C C G C C A C C T G A C
4 mm
4 mm
4 mm
5 mm
7 mm
7 mm
+43 bp
Os03g03034
TalBV v1 1x Xoc
v2 1x Xoc
v3 4x Xoc
v4 1x Xoc
TalCB v1 1x Xoc
v2 2x Xoc
T A C A A A G A G A A C G C A T C C C C C
1 mm
1 mm
1 mm
2 mm
4 mm
4 mm
+103 bp
Os09g29100



Appendix Figure 3.2.2: Predicted TALE binding sites within the promoter regions of all published *Xoc* TALE target genes. TALEs are depicted by their RVDs and aligned to the predicted EBEs found in the promoter regions of their published target genes. RVDs are shown for all unique TALE class versions. Class versions whose presence did not yield an induction of the target gene in the original RNA-Seq experiment (Wilkins *et al.*, 2015) are crossed out. The abundace of each version is indicated for each TALE class version (TalAD occurs additionally in several *Xoo* strains). Mismatched RVDs are boxed in black. The number of predicted mismatches (mm) is shown next to the RVD sequence and its colour indicates whether or not this version showed activity in the GUS reporter assay. Green indicates solid activity, yellow a weak activity (of less than 5 pmol MU/min/µg protein), red no activity and black that this version was not tested in the GUS assay. The target genes are depicted as arrows and a schematic of the first 1000 bp of each promoter is shown (beginning upstream of the first ATG). The relative distance between the TALE binding site (blue) and the annotated transcriptional start site (red) is indicated in base pairs (bp).

Appendix Table 3.2.6: Most promising putative new *Xoc* TALE target genes selected from the target prediction published by Erkes *et al*., 2017.

Locus ID	protein name or annotated function	TALE class	*Xoc* strain: BLS79	B8-12	RS105	L8	BLS256	TALE class Presence [1]: CFBP2286	BXOR1	log2 folgd change [2]: CFBP7342	CFBP7331	CFBP7341
Os01g32439	**expressed protein**	**AJ**					v1					
Os06g17020	**UDP-glucosyltransferase**	**AJ**					v1					
Os02g51110	**OsNIP2;1 (aquaporin)**	**AV**	v1	v1	v1	v1	v1	v1	v2	v2	v2	v2
Os02g02190	**nitrate transporter**	**AX**	v1	v1	v1	v1	v1	v1	v1	v2	v3	v4
Os07g26520	**expressed protein**	**AX**	v1	v1	v1	v1	v1	v1	v1	v2	v3	v4
Os11g07600	ABC transporter [3]	AY	v2	v2	v2	v2	v2	v2	v2	v3	v1	v1
Os04g49560	zinc finger protein [3]	AY	v2	v2	v2	v2	v2	v2	v2	v3	v1	v1
Os05g39200	hypothetical protein [3]	BD	v1	v1	v1	v1	v1	v1	v1	v2	v2	v2
Os02g51900	cytokinin-glucosyltransferase	BI	v2	v2	v2	v2	v2	v2	v3	v4	v1	v1
Os09g21380	expressed protein	BN	v1	v1	v1	v1	v1	v1	v1	v2		
Os04g33810	RNA recognition motif containing protein	CD	--	--	--	--	--	--	v1	v2	--	--
		AK	v1	v1	v1	v1	v1	v1	--	--	--	--
Os12g42200	**cation-hydrogen exchanger**	**CG**						v1	v1	v2		
Os06g36270	**putative receptor-like kinase 5**	**CK**	v3	v1	--	v1	--	--	--	--	--	--
		CP	--	--	--	--	--	v2	v1	--	v1	v1
Os01g40290	**expressed protein**	**CL**	v2	v2	v2	v2	v1	v2	v3	--	--	--
		CJ	--	--	--	--	--	--	--	v1	v1	v1
Os05g42150	**OsGH3.4 (indole-3-acetic acid-amido synthetase)**	**CL**	v2	v2	v2	v2	v1	v2	v3	--	--	--
		CJ	--	--	--	--	--	--	--	v1	v1	v1
Os10g39840	**OsXTH20 (glycosyl hydrolase family 16)**	**CL**	v2	v2	v2	v2	v1	v2	v3	--	--	--
		CJ	--	--	--	--	--	--	--	v1	v1	v1

</= 0
</= 0.5
</= 0.9
</= 1.9
</= 2.9
</= 3.9
> 4.0

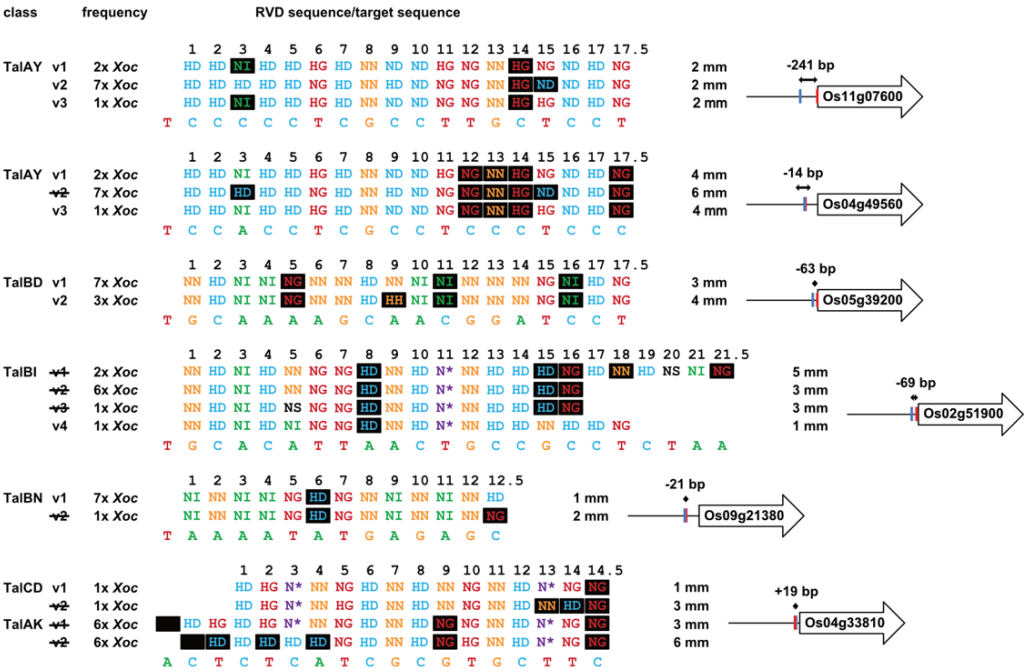
The genes shown in this table were manually selected from a larger set of putative *Xoc* TALE target genes predicted by Erkes *et al*., 2017. The prediction is based on the re-analysis of an RNA-Seq experiment originally published by Wilkins *et al*., 2015.

[1] Indicated are the specific versions of each TALE class, the version numbers were assigned according to chapter 3.2.1.

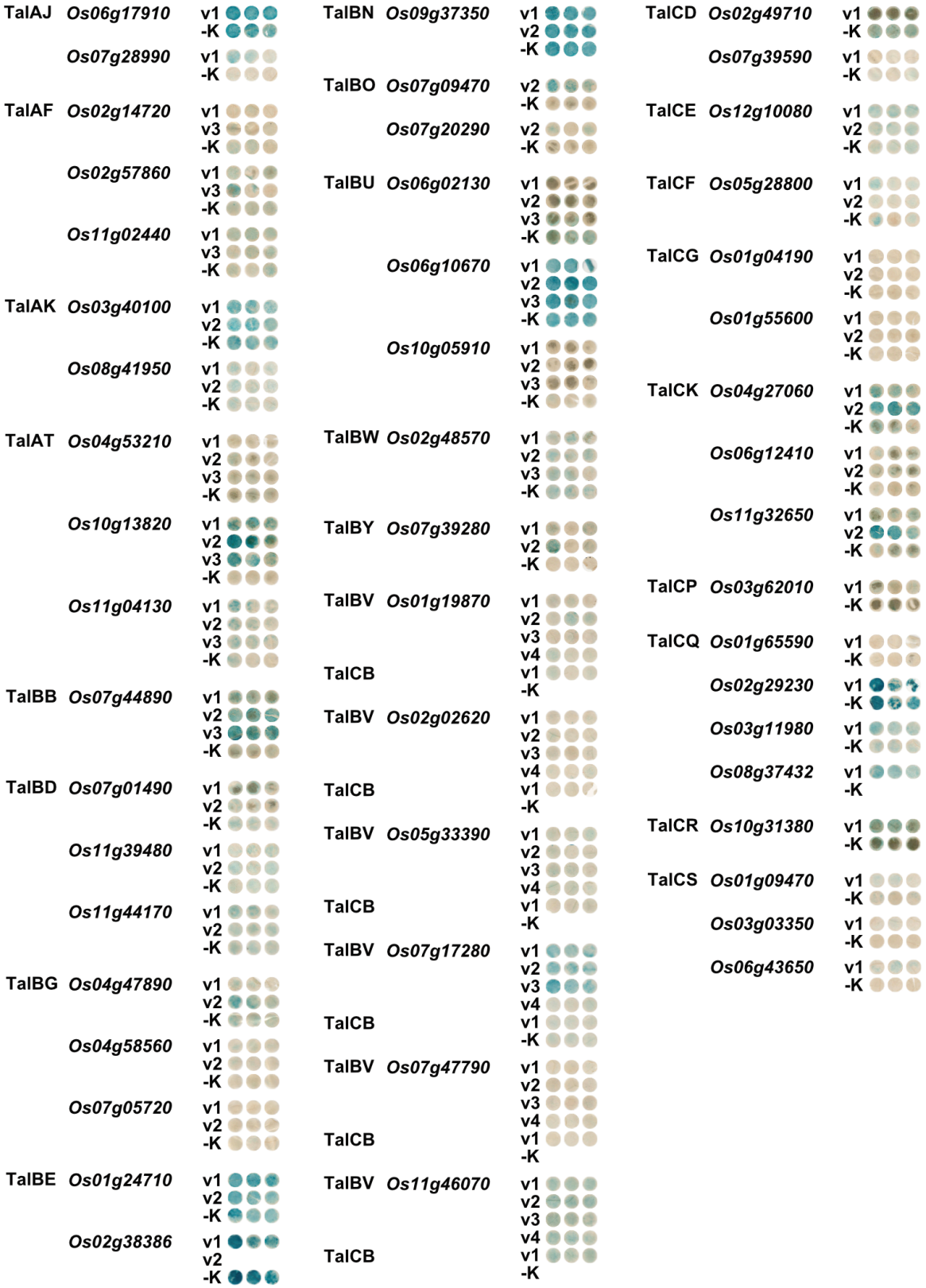
[2] Shown are the log 2-fold values for each gene in response to the infection with the 10 completely sequenced *Xoc* strains.

[3] The different shades of white, grey and black correspond to the expression value of the respective gene.

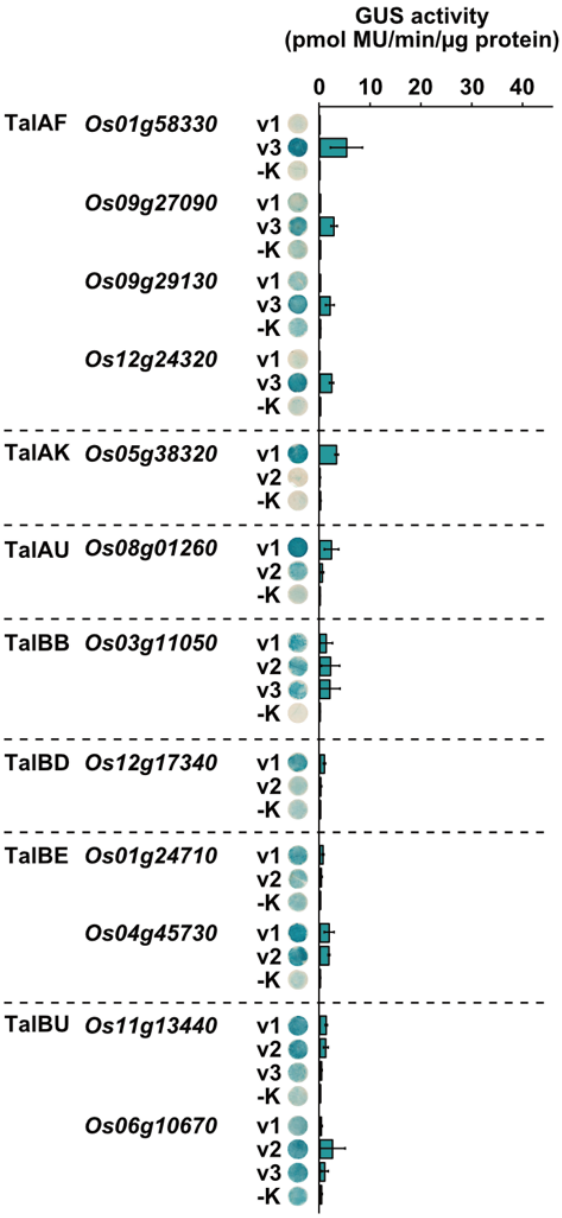
[4] this target was selected for experimental validation but the PCR amplification its their reporter fragments was not successful.



Appendix Figure 3.2.3: Putative TALE binding sites within the promoters of all rice genes from Appendix Table 3.2.6. that were not yet analysed experimentally. TALEs are depicted by their RVDs and they are aligned to the predicted EBEs found in the promoter regions of their putative target genes. RVDs are shown for all unique TALE class versions. Class versions whose presence did not yield an induction of the target gene in the original RNA-Seq experiment are crossed out (RNA-Seq originally by Wilkins *et al*., 2015; re-analysis and target predictions by Erkes *et al*., 2017). The frequency of occurrence within the 10 completely sequenced *Xoc* strains is indicated for each TALE class version. Mismatched RVDs are boxed in black. The number of predicted mismatches (mm) is shown next to the RVD. The putative target genes are depicted as arrows and a schematic of the first 1000 bp of each promoter is shown (beginning upstream of the first ATG). The relative distance between the TALE binding site (blue) and the annotated transcriptional start site (red) is indicated in base pairs (bp).



Appendix Figure 3.2.4: Results of the qualitative GUS assay for all analysed rice genes whose promoters showed no or nearly no TALE-dependent induction. For each gene, a promoter fragment was amplified from the rice cultivar Nipponbare and cloned in front of a GUS reporter gene. The fragment encompassed always the first 1000 bp upstream of the first ATG of the CDS. The resulting reporter constructs were either inculated in combination with a TALE expression construct or without a TALE as a negative control (-K). Each combination was inoculated into three different plants and a leaf disc was harvested from each plant.



Appendix Figure 3.2.5: Results of the quantitative GUS assay for all putative *Xoc* TALE targets that were induced in response to a TALE but yielded only a weak GUS activity. For each gene, a promoter fragment was amplified from the rice cultivar Nipponbare and cloned in front of a GUS reporter gene. The fragment encompassed always the first 1000 bp upstream of the first ATG of the CDS. The resulting reporter constructs were either inculated in combination with a TALE expression construct or without a TALE as a negative control (-K). A representative leaf disc from the qualitative GUS staining is shown. Error bars represent standard deviaton (n = 3).

full links for figure 1.1

06.09.2018 16:30

Xoo http://www.knowledgebank.irri.org/decision-tools/rice-doctor/rice-doctor-fact-sheets/item/bacterial-blight

Xoc http://www.knowledgebank.irri.org/training/fact-sheets/pest-management/diseases/item/bacterial-leaf-streak

06.09.2018 16:30

Xoo https://www.cabi.org/isc/datasheet/56956

Xoc https://www.cabi.org/isc/datasheet/56977

List of publications

Mücke S., Reschke M., Erkes A., Schwietzer C.A., **Becker S.**, Streubel J., Morgan R.D., Wilson G.G., Grau J., and Boch J. (2019). Transcriptional reprogramming of rice cells by *Xanthomonas oryzae* TALEs. Frontiers in Plant Science *10*, 162.

Becker S. and Boch J. (2016). Single-molecule biophysics: TALEs spin along, but not around. Nature Chemical Biology *12*, 766-768.

Raschke A., Ibañez C., Ullrich K.K., Anwer M.U., **Becker S.**, Glöckner A., Trenner J., Denk K., Saal B., Sun X., Ni M., Davis S.J., Delker C., Quint M. (2015). Natural Variants of *ELF3* Affect Thermomorphogenesis by Transcriptionally Modulating *PIF4*-Dependent Auxin Response Genes. BMC Plant Biology *15*, 197.

Danksagung

Mein besonderer Dank gilt Prof. Dr. Jens Boch für das in mich gesetzte Vertrauen, den fachlichen Rat, die angeregten Diskussionen und, nicht zuletzt, für die Möglichkeit diese Arbeit in seiner Arbeitsgruppe anfertigen zu können.

Ich möchte mich außerdem bei Prof. Dr. Edgar Maiß für die Erstellung des zweiten Gutachtens bedanken.

Weiterhin bedanke ich mich bei Prof. Dr. Ulla Bonas in deren Arbeitsgruppe ich meine Promotion beginnen durfte sowie bei bei Dr. Jan Grau für die Hilfe in bioinformatischen Fragen.

Bedanken möchte ich mich auch bei allen Mitgliedern der AG Bonas in Halle und der AG Pflanzenbiotechnologie in Hannover für die stete Bereitschaft zu helfen sowie für die produktive und angenehme Arbeitsatmosphäre.

Ich bedanke mich bei allen TAs, im Besonderen allerdings bei Tjorven für die effiziente, angenehme und vor allem unterhaltsame Zusammenarbeit in den letzten 2 Jahren.

Auch bedanke ich mich bei den Teilnehmern der Brettspieleabende in Hannover, besonders bei Gerrich der die meisten davon initiiert hat.

Bedanken möchte ich mich auch noch einmal explizit bei allen Doktoranden mit denen ich zusammenarbeiten durfte, im Besonderen aber bei Caro, Steffi, Maik, Jana und Annekatrin – für den großartigen Zusammenhalt, die Diskussionen und Ratschläge, die unterhaltsamen Freizeitunternehmungen, die Toleranz eurerseits sowie die aufbauenden Worte in schlechten Zeiten, also grundsätzlich einfach für alles.

Weiterhin bedanke ich mich bei meinen Freunden die mich auch in Zeiten geringen Kontakts nie vergessen haben, vor allem Christian und Benni.

Und schließlich bei meiner Familie, meinen Eltern und natürlich noch einmal bei Annekatrin – für eure Unterstützung und eure Geduld!

Curriculum vitae

Name:	Sebastian Becker
Address:	Altenauer Weg 5
	30419 Hannover
Birth:	11.11.1987
	in Bernburg (Saale) (Germany)

Education

since 10/2018	**Senior Research Scientist, Leibniz-Universität Hannover** Institute of Plant Genetics, Department of Plant Biotechnology, Prof. Dr. Jens Boch
10/2015-09/2018	**Doctorate (continued), Leibniz-Universität Hannover** Institute of Plant Genetics, Department of Plant Biotechnology, Prof. Dr. Jens Boch
03/2014-09/2015	**Doctorate, Martin-Luther-Universität Halle-Wittenberg** Institute of Biology, Department of Genetics, Prof. Dr. Ulla Bonas, supervisor Dr. Jens Boch doctoral thesis: "Analysis of TALE repeats of aberrant length and identification of novel candidate target genes for *Xanthomonas oryzae* pv. *oryzicola* TALEs"
09/2011-02/2014	**Master of Science Biology, Martin-Luther-Universität Halle-Wittenberg** master thesis: „Molekulare und funktionelle Charakterisierung von TAL-Effektor-ähnlichen Proteinen aus *Burkholderia rhizoxinica*“
08/2008-09/2011	**Bachelor of Science Biology, Martin-Luther-Universität Halle-Wittenberg** bachelor thesis: „*Quantitative Trait Loci* Analyse von auxin- und temperaturinduzierten Wachstumsveränderungen in *Arabidopsis thaliana*"
7/2007	**Higher education entrance qualification, Gymnasium Carolinum in Bernburg (Saale)**