Function and application of TAL effectors

by

Ting Li

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Genetics

Program of Study Committee:
Bing Yang, Major Professor
Diane Bassham
Jeffrey J. Essner
Martin H. Spalding
Yanhai Yin

Iowa State University

Ames, Iowa

2014

To my family

TABLE OF CONTENTS

ABSTRACT

The transcription activator-like effectors (TALE) are the largest family of the type III effector proteins from *Xanthomonas* spp. Each TALE contains an N-terminal signal for its secretion and translocation into host cells, a central modular DNA-binding repeat domain, and three nuclear localization signals (NLS) and an acidic activation domain (AD) at the C-terminus. Once internalized, TALEs facilitate the bacterial parasitism by transcriptionally activating the host target genes. TALE repeats recognize and bind to the target DNA sequence in a simple and unique "one to one" fashion, i.e., one repeat corresponds to one nucleotide and one type of repeat preferentially binds one of four necleotides, a code that makes TALE an ideal material to achieve the recognition of any preselected DNA sequence.

In my thesis I describe the development and functional characterization of TALENs, fusion proteins of native or artificial TALEs and the cleavage domain of the restriction enzyme FokI, *in vitro* and in yeast, and demonstrate the feasibility of customizing the novel TALEN to target and inactivate the yeast endogenous genes. I also investigated the efficiency of the gene editing through the repair of TALEN-caused double-stranded DNA breaks by non-homologous end-joining (NHEJ) and homologous recombination (HR). Furthermore, I utilized the designer TALENs to edit the TAL effector binding elements (EBEs) in the promoters of the disease susceptibility (S) genes targeted by the rice blight pathogen *Xanthomonas oryzae* pv. *oryzae* (Xoo). The gene editing disrupted the induction of the S genes by the naturally occurring TALEs (PthXo1 and AvrXa7) and thus generated the disease resistant rice lines. Finally, using Xoo-delivered designer TALEs to induce the expression of rice *SWEET* genes that encode sugar transporters in rice, I identified two disease susceptibility genes potentially targeted by the as yet unidentified TALEs in *Xanthomonas oryzae* pv. *oryzae*.

CHAPTER I

LITERATURE REVIEW

TALEN-MEDIATED GENOME EDITING: PROSPECTS AND PERSPECTIVES

David Wright*[1], Ting Li*[1], Bing Yang*[2] and Martin H. Spalding*[2]

* Department of Genetics, Development and Cell Biology, Iowa State University, Ames, Iowa 50011, USA

1 These authors contribute equally to this work.

2 To whom correspondence should be addressed (email byang@iastate.edu, mspaldin@iastate.edu)

Abstract

Genome Editing is the practice of making predetermined and precise changes to a genome by controlling the location of DNA double strand breaks (DSBs) and manipulating the cell's repair mechanisms. This technology results from harnessing natural processes that have taken decades and multiple lines of inquiry to understand. Through many false starts and iterative technology advances, the goal of Genome Editing is just now falling under the control of human hands as a routine and broadly applicable method. This review attempts to define the technique and capture the discovery process while following its evolution from meganucleases and zinc finger nucleases to the current state of the art - TALEN technology. We also discuss factors that influence success, technical challenges and future prospects of this quickly evolving area of study and application.

Key words: TAL effector nuclease (TALEN), zinc finger nuclease, meganuclease, homologous recombination (HR), genetic engineering, gene editing, genome editing, non-homologous end-joining (NHEJ)

Introduction

Recombination is a mechanism by which cells repair damaged DNA and exchange information between identical or similar DNA molecules. This process is involved in many stages of the cell cycle and reproduction, and it aides in replication of some viruses, mobile elements and plasmids and plays a part in duplications, deletions and rearrangements of genetic material. In short, recombination appears to be essential for all known forms of life by acting both to ensure integrity of the genome and to contribute the genomic diversity needed for the action of evolutionary forces.

The ability to manage and exploit this powerful process would bring direct benefit to scientific, environmental, medical, agricultural and industrial arenas. Understanding the basic mechanism of recombination has been the goal of decades of work, which has resulted in a stepwise advancement of techniques that allow it to be routinely managed in selected species, including some yeasts and bacteria and mouse and chicken cells. However, even with the insights gained so far, directing recombination as a general method for manipulation of genomes is limited, laborious and time consuming.

The advent of high throughput DNA sequencing and the enormous amount of data that are available from many "omics" programs underscores the need for a precise method to target and modify specific DNA bases within the genome of a broad range of organisms – we refer to this practice as Genome Editing. This review will discuss the history, challenges, solutions and future of nuclease-mediated Genome Editing with an emphasis on TALEN (Transcription-Activator-Like Effector Nuclease) technology.

Genetic recombination is a broad topic with many important discoveries, mechanisms and contributors that have expanded our current understanding of the subject. A key element in

recombination is the DNA double strand break (DSB), which can halt the cell cycle and lead to death if not immediately and efficiently repaired. Here, we will examine nuclease-induced DSBs and the two aspects of recombination that seem to play a major role in Genome Editing: non-homologous end-joining (NHEJ) and homologous recombination (HR).

NHEJ is a DSB repair pathway that utilizes microhomology at the exposed ends of a DSB to guide repair without a template (Figure 1.1A) (Lieber 2010). This type of repair may include resection to generate additional single stranded DNA when damage beyond a simple DSB is detected. Repair by this pathway generally leads to restoration of the original gene sequence (a precise repair), or disruption by a small insertion or deletion (an error-prone repair) (Shrivastav, De Haro et al. 2008, Mladenov and Iliakis 2011). The ratio of restoration to modification is not known, however it is the mutation aspect that we are interested in exploiting.

In principal, NHEJ has been a long standing laboratory practice for genetic manipulation. Techniques have been developed to randomly break DNA and induce arbitrary mutations typically using chemicals, radiation or exogenous DNA. Pools of mutagenized organisms can then be screened for interesting phenotypes or specific gene mutations. These techniques have led to many breakthroughs and have greatly enhanced scientific understanding of biological systems.

HR is another DSB repair pathway, but it utilizes extended regions of homology between the exposed ends of a DSB and a donor DNA molecule, which is used as a template for repair (Figure 1.1B) (San Filippo, Sung et al. 2008). Donor DNA is typically endogenous to the cell, however, it has been demonstrated by a multitude of HR experiments that exogenous DNA can fulfill this role. Unlike NHEJ, HR can replace missing information at the DSB locus; however, if the donor DNA template contains alternative information then a gene conversion event can take

place. Exogenous template guided HR confers precise control over the DSB repair process, which is more valuable in a Genome Editing sense; however, NHEJ appears to be the preferred repair pathway in higher eukaryotic organisms (Mladenov and Iliakis 2011).

Arguably, the most fruitful early eukaryotic recombination experiments were performed in the yeast *Saccharomyces cerevisiae*, which led to many fundamental discoveries and the development of new methods (Hinnen, Hicks et al. 1978, Orr-Weaver, Szostak et al. 1981, Orr-Weaver, Szostak et al. 1983, Rothstein 1983, Szostak, Orr-Weaver et al. 1983). Among the findings, was that this organism naturally has a very high rate of HR relative to that of NHEJ. Initial recombination experiments in plant and animal cells found that NHEJ usually dominated and that HR events were relatively rare. Efforts have been made to improve HR results, typically by selecting for rare events through a marker scheme, but the rate of HR still proved to be too low for practical use in these species (Smithies, Gregg et al. 1985, Paszkowski, Baur et al. 1988, Sedivy and Sharp 1989, Lee, Lund et al. 1990, Offringa, de Groot et al. 1990, Halfter, Morris et al. 1992, Miao and Lam 1995, Risseeuw, Offringa et al. 1995). From these experiments, HR has been estimated to occur at a frequency of $1x10^{-3}$ to $1x10^{-6}$. The notable exceptions, outside of yeast, include murine embryonic stem cells and chicken DT40 cells. Murine embryonic stem cells have a 10 to 100 fold higher rate of HR than other mouse cells, which has led to their successful use in the development of gene knockout strategies in mouse (Doetschman, Gregg et al. 1987, Thomas and Capecchi 1987, Doetschman, Maeda et al. 1988, Capecchi 1989, Koller and Smithies 1992). DT40 cells displayed 50 to 90% HR, relative to the NHEJ rate. However, later studies suggested that the HR rate is not unusually high, but that the NHEJ rate is suppressed in these cells (Buerstedde, Reynaud et al. 1990, Buerstedde and Takeda 1991, Iiizumi, Kurosawa et al. 2008). It should be noted that DT40 cells are immortalized pre-B-cells,

which limits their usefulness outside of studies involving cultured cell lines. While these experiments provided insight into recombination and other cellular processes and, in the case of mouse, lead to knockout mice, the techniques do not readily lend themselves to use in a broad range of organisms.

Many approaches to controlling recombination have been undertaken, such as positive and negative selectable markers, recombinases, transposons, integrases, virus vectors, peptide nucleic acids (PNA) triplex-forming oligonucleotides (TFO) and derivative technologies, chimeric DNA/RNA molecules (chimeraplasty), small fragment homologous replacement (SFHR), and over-expression of recombination mechanism components to name a few. These methods have met with some success, but none have proven as broadly applicable as nuclease-mediated DSB technology. DSB studies using standard restriction endonucleases (4 to 8 bp recognition sites) culminating in cell death and genome rearrangements as well as localized lesions resulting from the fact that these nucleases cleaved in millions of locations in a large genome, never the less, these studies generated a tremendous amount of information regarding the effects of DSBs on the cell (Thacker 1994, Stoilov, Mirkova et al. 1996). On the other hand, due to their relatively large recognition site requirements, mobile intron derived homing nucleases proved to be much more amenable to DSB studies and allowed progress in targeted recombination by limiting DSBs.

Meganuclease Technology

Mobile introns, typically found in phage, archaea, bacteria and lower eukaryotes, engage in a replicative process known as intron homing. This occurs when an intron from one allele is copied to an intron-less allele of a gene. The mechanism is initiated when an intron encoded

homing nuclease binds to an extended (typically 14 to 40 bp) recognition site in an intron-less allele, and then breaks the chromosome. The resulting DSB initiates repair, through HR, using the intron containing allele as a template. When the process is completed, both genes have the same homing-nuclease-gene-containing intron at the same position (Belfort and Perlman 1995, Belfort and Roberts 1997, Chevalier and Stoddard 2001).

A well characterized example is the I-SceI system in *Saccharomyces cerevisiae*. The mitochondrial 21S ribosomal RNA gene of ω+ strains contains a mobile intron, encoding the homing nuclease I-SceI, that is transmitted to ω- strains during mating with greater than 95% efficiency (Dujon, Slonimski et al. 1974, Netter, Petrochilo et al. 1974, Bos, Heyting et al. 1978, Faye, Dennebouy et al. 1979, Dujon 1980, Jacquier and Dujon 1985, Zinn and Butow 1985, Dujon 1989). I-SceI recognizes a target of 18 consecutive bp, which would be predicted to cut once in 68.7 billion bases, assuming a random genome (Colleaux, d'Auriol et al. 1986, Colleaux, D'Auriol et al. 1988, Monteilhet, Perrin et al. 1990). For reference, the size of the yeast genome is about 12.1 million bp and that of the human genome is approximately 3.3 billion bp.

The discovery, purification and characterization of I-SceI sparked an interest in studying DSBs in organisms other than yeast and enabled several key recombination studies in mammal, amphibian and plant cells during the 1990s (Puchta, Dujon et al. 1993, Rouet, Smih et al. 1994, Segal and Carroll 1995). Importantly, this and other work demonstrated an unprecedented increase in HR, with results as high as 10^{-2} to 10^{-3} for transfected mouse and transformed tobacco cells, making it clear that nuclease-mediated DSBs could induce high frequency recombination that might be applied universally (Choulika, Perrin et al. 1995, Jasin, Moynahan et al. 1996, Puchta, Dujon et al. 1996, Sargent, Brenneman et al. 1997). The remaining hurdle was how to introduce specific DSBs in targets for which an appropriate homing nuclease did not exist.

This work garnered substantial interest in engineering homing nucleases, also called meganucleases, for the task of Genome Editing. While academic researchers have contributed in this area, Cellectis and Precision Biosciences, Inc. have been the leaders in a large commercial effort. Engineering methods include dividing related nucleases into parts with random reassembly, mutation of areas that make DNA contact, changing amino acids in the DNA binding domain based on known interactions and computer modeling using algorithms (Gruen, Chang et al. 2002, Seligman, Chisholm et al. 2002, Chevalier, Turmel et al. 2003, Sussman, Chadsey et al. 2004, Chames, Epinat et al. 2005, Arnould, Chames et al. 2006, Ashworth, Havranek et al. 2006, Doyon, Pattanayak et al. 2006, Rosen, Morrison et al. 2006, Smith, Grizot et al. 2006, Chen, Wen et al. 2009, Ashworth, Taylor et al. 2010). Once assembled, libraries of new enzymes were screened for activity against a lengthy target of interest to find promising enzymes with altered specificity.

These screens represent key technology developments that have been used as tools for the advancement of many types of engineered nucleases; not just meganucleases. Among them are bacterial two hybrid methods, several bacterial marker gene elimination assays, and, arguably the most direct screening strategy, the yeast-based nuclease mediated recombination assay (Joung, Ramm et al. 2000, Gruen, Chang et al. 2002, Seligman, Chisholm et al. 2002, Gimble, Moure et al. 2003, Sussman, Chadsey et al. 2004, Chames, Epinat et al. 2005, Doyon, Pattanayak et al. 2006, Rosen, Morrison et al. 2006). Although there are variations to this method, this yeast-based assay basically works as follows (Figure 1.2): a disrupted and non-functional yeast marker gene is constructed such that an engineered nuclease target site and a selectable gene are imbedded within a small internal duplication of the marker gene coding sequence. This plasmid and a plasmid encoding an engineered nuclease are transformed separately into yeast cells of

opposite mating type. After mating, the cells are assayed for marker gene repair, which is dependent on target site cleavage and is used as a reporter to determine the initial quality of the engineered nuclease (Chames, Epinat et al. 2005). This method is suitable for manual or automated production and has identified many thousands of engineered nucleases in both academic and commercial laboratories (Arnould, Chames et al. 2006, Smith, Grizot et al. 2006, Doyon, McCammon et al. 2008, Townsend, Wright et al. 2009).

The success of meganuclease technology is just one facet of a continuously expanding and advancing field of academic study and commercial activity. While Genome Editing began with meganucleases, it certainly does not end there. Other highly successful tools for manipulating genomes combine two separate functional units: a DNA binding domain and an active element from a protein such as a transcription factor or a restriction endonuclease.

Zinc Finger Nuclease Technology

In the 1980s, transcription factor IIIA, an easily obtained and abundant egg protein from *Xenopus laevis*, was studied in great detail (Brown 1984). Analysis revealed that the protein contained zinc and nine tandem repeat units, each containing about 30 amino acids (Ginsberg, King et al. 1984, Miller, McLachlan et al. 1985). Initial structural models were built using a consensus from the repeats, then fitting this to the known structure of metalloproteins (Berg 1988). Later, NMR was used to gain insight into the architecture and, finally, the crystal structure was deduced (Lee, Gippert et al. 1989, Pavletich and Pabo 1991, Nakaseko, Neuhaus et al. 1992, Neuhaus, Nakaseko et al. 1992).

Through these studies, the zinc finger (ZF) DNA binding domain was defined. Today there are many recognized forms of this domain; however, the subject of this review is the ββα

structure that coordinates a zinc atom using a Cys_2His_2 motif to position the α helix into the major groove of a DNA strand for specific recognition of three consecutive bases (Pavletich and Pabo 1991). It should be noted that later analysis showed the α helix can have cross strand interactions with the base that precedes the ZF target, making it clear that each ZF can have a four base interaction with target site overlap between ZF modules (Fairall, Schwabe et al. 1993, Elrod-Erickson, Benson et al. 1998).

Attempts to engineer this structure began almost immediately and a small collection of mutant ZFs with altered DNA binding profiles was amassed. ZF arrays were assembled, generally using the consensus linker sequence TGEKP, and tested for altered target affinity and specificity. One such ZF array was used to both down-regulate an oncogene through binding alone and to up-regulate a CAT gene when fused to the activation domain VP16 (Choo, Sanchez-Garcia et al. 1994). Later, individual repressor domains fused to ZF arrays were used to demonstrate repressor activity (Papworth, Moore et al. 2003, Reynolds, Ullman et al. 2003).

The success of ZF-based artificial transcription factors led to the formation of Sangamo BioSciences, Inc. and Gendaq Ltd to exploit the new science. ZF arrays with three, four and six finger architectures were created along with modular methods and more powerful library based selection schemes to make custom ZF arrays (Choo and Klug 1994, Choo, Sanchez-Garcia et al. 1994, Rebar and Pabo 1994, Greisman and Pabo 1997, Isalan, Klug et al. 1998, Segal, Dreier et al. 1999, Choo and Isalan 2000, Joung, Ramm et al. 2000, Moore, Choo et al. 2001, Moore, Klug et al. 2001). Basically, individual ZFs with substitutions in the helix for residue positions -1, 2, 3 and 6 were generated then assembled into ZF array libraries followed by screening using a variety of *in vitro* and *in vivo* methods. Eventually, engineered ZF transcription factors were used to demonstrate repression of human immunodeficiency virus (HIV) and herpes simplex virus and

activation of vascular endothelial growth factor (VEGF) (Liu, Rebar et al. 2001, Papworth, Moore et al. 2003, Reynolds, Ullman et al. 2003). In the meantime, ZF technology slowly matured as the efforts of academic and commercial researchers continued, and ZFs soon caught the attention of a research group that specialized in nucleases.

Chandrasegaran's group began a detailed mutagenesis analysis of FokI, a type IIs restriction endonuclease, which led to an engineered nuclease revolution (Li, Wu et al. 1992, Li and Chandrasegaran 1993, Li, Wu et al. 1993, Kim and Chandrasegaran 1994). This nuclease consists of a DNA binding domain similar to a helix-turn-helix motif and a functionally separate and nonspecific nuclease domain (Figure 1.3) (Wah, Hirsch et al. 1997). Restriction endonucleases commonly function as dimers, each half binding one side of a palindromic target prior to cleavage. The FokI target sequence (5'-GGATG-3') is not palindromic and is typically singular in a given section of DNA. This arrangement should allow for a single FokI monomer to bind its target, impeding cleavage, since the nuclease domain contains a single catalytic center (Wah, Hirsch et al. 1997, Bitinaite, Wah et al. 1998, Wah, Bitinaite et al. 1998). Nonetheless, a DSB is made at the 3' end of the recognition site, after the 9th base on the top strand and the 13th base on the bottom strand; leaving a four base 5' overhang (Sugisaki and Kanazawa 1981). It is thought that either two separate DNA molecules, each with a FokI monomer bound, or a single DNA molecule with a FokI monomer bound and a free monomer must make momentary contact for a dimer interface to form and cleavage to occur (Bitinaite, Wah et al. 1998, Wah, Bitinaite et al. 1998, Vanamee, Santagata et al. 2001, Pernstich and Halford 2012). This is a *trans* cleavage model as opposed to the *cis* DNA binding site arrangement used by many classic palindromic restriction endonucleases to cleave within their recognition site. For example, a pair of EcoRI monomers must bind their target, one on either side of the palindromic recognition site

GAATTC, to form a dimmer interface then make a DSB (Pingoud, Fuxreiter et al. 2005). This seems a fine distinction, but the difference between these two types of interactions has implications for later technology development.

After realizing that the FokI DNA recognition domain was dispensable, the catalytic portion was combined with several different types of DNA binding motifs (Kim and Chandrasegaran 1994, Huang, Schaeffer et al. 1996, Kim, Cha et al. 1996, Kim, Kim et al. 1997, Kim, Smith et al. 1998). Among those tested were three distinct Cys_2His_2 ZFs; these were the first zinc finger nucleases (ZFNs). Initial analysis of ZFNs was performed *in vitro* using purified proteins and either monomeric or multiple tandem target sites, which relied on *trans*-cleavage for digestion to occur (Huang, Schaeffer et al. 1996, Kim, Cha et al. 1996, Smith, Berg et al. 1999). *Trans*-cleavage is inefficient at low protein concentrations, but becomes more effective as the concentration of nuclease is increased (Bitinaite, Wah et al. 1998). It wasn't until the Chandrasegaran and Carroll collaboration that the now-classic and more efficient pairing of two opposing ZFNs bound to DNA binding sites in a *cis* arrangement was utilized (Smith, Bibikova et al. 2000). This is now the most recognized visual model of engineered, FokI based nuclease binding for the creation of DSBs and control of the Genome Editing process (Figure 1.4A).

The combination of engineered ZF arrays and the FokI nuclease domain could theoretically generate an enzyme capable of producing a single DSB anywhere in a complex genome. To illustrate the point, in a genome made of random DNA, a ZFN pair that employs three, four, or six ZFs per half site has a combined 18, 24 or 36 base recognition site, giving it a unique target in 6.87×10^{10} bp, 2.81×10^{14} bp or 4.72×10^{21} bp respectively. For reference, the largest genome found to date, *Paris japonica*, carries an estimated 1.5×10^{11} bp of DNA (Pellicer, Fay et al. 2010).

Once efficacy of the design was established, the potential benefits of ZFN technology were obvious and work soon began to demonstrate and define the limits of the technique. Initial engineered ZFN experiments demonstrated HR using an extrachromosomal target in Xenopus oocytes, NHEJ and later HR at an endogenous locus in Drosophila and marker gene restoration in human cells (Bibikova, Carroll et al. 2001, Bibikova, Golic et al. 2002, Bibikova, Beumer et al. 2003, Porteus and Baltimore 2003). Afterward, NHEJ at a small artificial target in Arabidopsis and HR at an integrated marker gene in tobacco were demonstrated using pre-existing ZFNs (Lloyd, Plaisier et al. 2005, Wright, Townsend et al. 2005). This was followed by demonstrations of HR at three positions in an endogenous locus in tobacco and NHEJ and targeted gene insertion in maize using task specific engineered ZFNs, which confirmed the utility of ZFN technology in plants (Shukla, Doyon et al. 2009, Townsend, Wright et al. 2009). Following the initial success, many model and economically important organisms have been subjected to ZFN mediated Genome Editing with a notable record of success. Along with an ever broadening field of application came improved ZFN engineering methods that promised to put ZFN technology into the hands of more researchers (Carroll, Morton et al. 2006, Meng, Thibodeau-Beganny et al. 2007, Doyon, McCammon et al. 2008, Maeder, Thibodeau-Beganny et al. 2008, Kim, Lee et al. 2011, Sander, Dahlborg et al. 2011).

Even though ZFN development and adoption have outpaced meganuclease technology, numerous challenges became apparent to those applying the technique. The principal complication appears to be toxicity. This could be general or species specific and seems to be associated with ZF binding in some cases and ZFN generated off target DSBs for many others (Bibikova, Golic et al. 2002, Porteus and Baltimore 2003, Alwin, Gere et al. 2005). In the later instances, off target DSBs appear to be caused by poorly designed ZF arrays. For example, it has

been shown that some ZFs can tolerate up to eight mismatches in a 24 bp target; drastically increasing the potential for off target binding (Mussolino and Cathomen 2011). Another matter is the *trans* cleavage nature that FokI brings to the system; allowing ZFN monomers to cleave sites that are unrelated to the intended target. Some of the observed toxicity was reduced by the advent of improved ZF engineering protocols that took into account cross strand interaction between the a helix and the opposing base that precedes the ZF module as discussed above. Additional reductions in toxicity were realized when obligate heterodimeric forms of the FokI catalytic domain were produced. As discussed below, the heterodimer design ensures that a left ZFN interacts with a right ZFN and that left/left or right/right combinations are minimized, which presumably reduces the off target DSB occurrence in a cell. Since the initial development of ZF technology, great strides have been made in generating optimized ZFNs that are less toxic and more specific; however, not every sequence can be targeted because highly specific ZFs do not exist for all triplet targets in all contexts. Until recently, it was estimated that only one target in 500 bp could be utilized by ZFN technology using context assembly methods, but with the development of extended modular assembly, this estimate has dropped to one in 52 bp (Maeder, Thibodeau-Beganny et al. 2008, Sander, Dahlborg et al. 2011, Bhakta, Henry et al. 2013).

Today, Sangamo BioSciences is the dominant economic stake holder for ZF based technology with plant rights licensed to Dow AgroSciences LLC and ZFN manufacturing rights licensed to Sigma-Aldrich Corporation. ZFN technology will continue to play a significant role in the future of Genome Editing even as new technologies are developed. The preceding information was provided to explain the foundation for the creation of TALEN technology, which is the focus of the remainder of this review.

TALE Histroy, Function And Structure

TALE Technology

The transcription activator-like effectors (TALEs) represent the largest family of type III effector proteins from *Xanthomonas* spp., a group of Gram negative bacterial plant pathogens. TALE homologs can also be found in the bacterium *Ralstonia solanacearum* (Heuer, Yin et al. 2007, Li, Atef et al. 2013). TALEs are secreted into plant cells and function as transcription activators during infection. They contain an N-terminal bacterial secretion and translocation signal, a single unique DNA binding domain, termed repeat 0, followed by a central repeating modular DNA-binding domain, three C-terminal eukaryotic nuclear localization signals (NLS), and a highly conserved acidic activation domain (AD) (Figure 1.5A.) (Boch and Bonas 2010). The central repeats are nearly-identical except for two amino acids at positions 12 and 13, called Repeat Variable Di-residues (RVDs) (Moscou and Bogdanove 2009). Different TALEs vary in repeat number and RVD composition (Boch and Bonas 2010). It is their role as plant pathogenesis factors that initially captured the attention of researchers, although, the deciphering of the DNA binding code and eventually of the crystal structure has led to their widespread use as artificial transcription factors and nucleases. Major license holders of TALE related technology include: Two Blades Foundation, Cellectis and Life Technologies (now part of Thermo Fisher Scientific).

Natural TALE Function

TALEs are pathogenic bacterial proteins that have evolved as eukaryotic transcription factors. They are secreted into plant host cells through the type III secretion system and translocated into the nucleus, where their specific binding to the promoter of host susceptibility

(S) genes facilitates colonization, proliferation, and dissemination of the bacteria. An example is the interaction between TALE AvrBs3 and the gene *UPA20*, a plant transcription factor of the bHLH family, which functions to enlarge mesophyll cells. TALE-mediated activation of this gene leads to hypertrophy, which facilitates infection in pepper plant (Kay, Hahn et al. 2007). Examples from rice include *OsSWEET11* (also named *Os8N3*) and *OsSWEET14* (also named *Os11N3*). These two genes reportedly serve as sucrose efflux transporters, facilitating sucrose movement through intracellular regions of the plant. Pathogenic bacteria use TALEs to specifically over-activate these genes, presumably to supply nutrients for continued bacterial growth (Chen, Hou et al. 2010, Chen, Qu et al. 2012).

Occasionally, plants win the host-pathogen battle through evolution and natural selection. The TALE binding element (EBE) in the S gene can mutate so a TALE no longer binds, or an EBE can be created in the promoter of a resistance (R) gene to trap the otherwise pathogenic TALE, leading to plant resistance in either case. For example, in pepper, TALE AvrBs3 normally induces *UPA20*, but an R gene called *Bs3*, has evolved an AvrBs3 EBE to provide resistance to the pathogenic strain *Xanthomonas campestris* pv. *vesicatoria* (*Xcv*) through the onset of a series of defense responses, including a hypersensitive response, a rapid, localized cell death that restricts bacterial spread (Kay, Hahn et al. 2007, Romer, Hahn et al. 2007).

TALE DNA Recognition Code

Decryption of the TALE DNA binding code occurred 20 years after characterization of the first TALE. Yang et al. provided the first clue in 2000, when they reported that TALE AvrXa7 is a DNA binding protein with a preference for dA/dT oligos (Yang, Zhu et al. 2000). The second clue came from an analysis of three related TALEs and their level of activity toward

a common gene. The amino acid (aa) sequence of AvrBs3 and AvrBs4 are 97% identical and both have 17.5 repeats, but only AvrBs3 activates the *Bs3* gene in pepper. AvrBs3 and AvrHah1 are only 87% identical and the latter has 13.5 repeats, yet both activate *Bs3*. Closer analysis showed that the RVDs of AvrBs3 and AvrHah1 are closely related, while those of AvrBs3 and AvrBs4 are not. This correlation between RVD similarity and *Bs3* activation, but not overall aa identity and activation, suggested that RVD composition predominantly determines EBE binding specificity (Schornack, Minsavage et al. 2008). The third clue was the discovery that AvrBs3 directly binds to its EBE, and that the number of repeats nearly equals the DNA target nucleotide length. Essentially, the authors noticed that AvrBs3 has 17.5 repeats and the Bs3 EBE length is 18 bp (Romer, Hahn et al. 2007).

Finally, in 2009, two independent groups using different methods reported the TALE DNA binding code in the same issue of Science. From published data, both groups speculated that TALEs recognize their target using one RVD per nucleotide. The Bogdonove group used a bioinformatics approach to search for minimal entropy regions in gene promoters then confirmed the technique using known TALE-EBE combinations. Next they used RVDs from numerous TALEs to predicted unknown EBEs and correlated this to public microarray infection data. The authors used this information to statistically surmise that each RVD displayed a preference for specific nucleotides (Moscou and Bogdanove 2009). The Bonas group manually aligned AvrBs3 and its known EBE, then used these data to predict the EBEs of four TALEs with unknown DNA binding sites, which were then found in the promoters that each TALE is known to regulate. Following this, they used molecular biology methods to experimentally validate predictions (Boch, Scholze et al. 2009). The code, as it turns out, is relatively simple: the first base of the EBE is recognized by a region in the TALE N-terminus (repeat 0) and the remaining bases are

sequentially bound by one RVD per nucleotide with one type of RVD preferentially recognizing one specific nucleotide.

Many RVDs have been described in the literature, the four major RVDs are HD, NI, NG, and NN, which predominantly bind to cytosine (C), adenine (A), thiamine (T), and guanine (G) or A, respectively. However, recognition is not absolute. Each RVD has the capacity to recognize other bases; though the interactions indicated above are the most common. TALE binding occurs in a stepwise manner beginning at the 5' end of the EBE, which is normally occupied by a T, where repeat 0 initiates the process, then successive RVDs from the central repeat region bind to each consecutive base through the end of the EBE. It should be noted that TALEs are more sensitive to EBE mismatches at the 5' end relative to the 3' end. This is due in part to the polarity TALEs exhibit regarding target sites and the fact that repeat 0 is generally specific for T. Mutation of this initial T usually abolishes TALE binding activity, and other 5' mutations make a TALE less likely to bind, while 3' mutations are likely to have less of an effect since the more c-terminal repeat regions show less affinity for DNA (Meckler, Bhakta et al. 2013).

TALE Structure

Shortly after discovery of the TALE DNA binding code, the structure of this domain was solved through protein crystallization by two groups (Deng, Yan et al. 2012, Mak, Bradley et al. 2012). Crystal structure analysis sought to determine how TALE repeats are ordered and how they interact with specific DNA bases. Examination of TALE dHax3 and TALE PthXo1 indicated that the central repeat region is a right handed super helix wrapping around the sense strand of the target DNA, and each repeat is a left handed helix-loop-helix structure. The 12^{th} and

13th aa (RVD) of the repeats are located in the loop region where the 12th aa stabilizes the loop through hydrogen bonds with the protein backbone while the 13th aa determines base specificity when interacting with the major groove. Analysis of repeat 0, which precedes the central RVD repeats, revealed a degenerate helix-turn-helix motif. A tryptophan (W232) in the turn region specifically interacts with the conserved T in the EBE (Mak, Bradley et al. 2012). These crystal structures provided great insight into the TALE DNA binding domain architecture and will help guide future engineering efforts. Crystallization of the remaining TALE structures may give additional insights into TALE biology and engineering and may provide other future benefits that only the complete structure of this valuable protein can provide.

TALEN Architecture, Assembly And Applications

TALENs

Efforts to engineer TALEs began immediately after the DNA binding code was discovered. Obvious parallels between established ZFN technology and the promising TALE DNA binding domain meant that experimentation with TALE nucleases (TALENs) was inevitable. Like ZFNs, TALENs are generated by fusing the FokI nuclease element to a TALE DNA binding domain. This configuration necessitates that two TALENs dimerize to generate efficient nuclease activity. To assist this, TALEN target half sites are chosen such that TALEN pairs are arranged in an opposing orientation on opposite sides of doubled-strand DNA with an optimal spacer sequence between them. If successful, left and right TALENs bind their half sites, allowing the FokI domains to dimerize and cleave the chromosome to generate a DSB (Figure 1.5B).

Unlike ZFNs and meganucleases, TALEN technology has been broadly and rapidly adopted, which is due in large part to their binding affinity, specificity, ease of manufacture, high probability of generating functional nucleases and success rate in Genome Editing experiments. Additionally, the results of side-by-side comparisons of ZFNs and TALENs suggest that TALENs exhibit reduced nuclease associated cytotoxicity (Mussolino, Morbitzer et al. 2011). This may be due in part to longer DNA target sites and the T requirement at the 5' end of a target site, together resulting in greater specificity, but much more comparative research is needed before a final conclusion can be drawn. It is for these features and more that TALEN mediated Genome Editing is listed as one of the ten breakthroughs of 2012 by the journal Science.

FokI Homodimer and Heterodimer Forms

The common usage of the FokI domain in artificial nuclease designs is largely due to its thorough characterization, acceptance of alternative DNA binding domains, indiscriminate nuclease activity, history of success in ZFN technology and convenience. Despite its nearly universal adoption, the FokI domain has some drawbacks; most notably its engagement in *trans* cleavage events and its normal function as a homodimer. The *trans* cleavage characteristic, as discussed above, is not completely understood and requires a concentrated effort to resolve, but it does appear to be associated with nuclease concentration, making promoter choice a consideration during experimental design of artificial nucleases (Bitinaite, Wah et al. 1998, Beumer, Bhattacharyya et al. 2006, Pruett-Miller, Connelly et al. 2008). On the other hand, the homodimer problem has been the subject of a number of independent studies.

The difficulty of the homodimer becomes apparent when it is understood that artificial nuclease pairs designed with a homodimer nuclease can have three interactions: two left halves

or two right halves can make a functional nuclease just as easily as the intended interaction between the left and right sides. In essence, using the homodimer is like simultaneously introducing three different nucleases into a cell, then coupling this to the FokI *trans* cleavage complication and the probability of overwhelming the cell with DSBs becomes high. Both of these issues can decrease the overall effectiveness of Genome Editing by precipitating a high level of cell death and collateral damage in the genome of surviving cells.

To reduce off-target toxicity, several obligate heterodimer versions of FokI have been developed. The engineered versions are based on structure-guided design, DNA shuffling and mutagenesis (Miller, Holmes et al. 2007, Szczepek, Brondani et al. 2007, Guo, Gaj et al. 2010, Doyon, Vo et al. 2011, Sizova, Greiner et al. 2013). These designs have greatly reduced the toxic effect of off-target DSBs observed when using the homodimer form of FokI. Which of the available heterodimer designs is actually most beneficial probably depends on the cell type and experimental plan; however, a thorough testing of multiple heterodimers pairs under the same conditions would be a welcomed and insightful study. Example heterodimer pairs are summarized in Table 1.1.

TALEN Architectures

The TALE family of engineered nucleases is diverse; the protein itself is relatively large, it has a host codon bias, and the FokI attachment sites are numerous. These variables lead to a wide range of engineered TALE architectures in the literature. For example, TALENs with an N-terminal fusion are functional, but most reported architectures use a C-terminal fusion (Li, Huang et al. 2011). Additionally, FokI can be directly fused to the existing TALE structure, or the TALE can have a combination of N-terminal and/or C-terminal deletions. C-terminal

deletions affect the spacer length requirement (typically 12 to 33 bp) between half sites and may or may not remove the NLS and activation domains. N-terminal deletions generally remove the bacterial secretion and translocation signals and may include the addition of an NLS. Which specific TALE to use? What truncation, if any, and FokI fusion to use? What form of FokI to use? What promoter strength is right? To codon optimize or not? These and other questions are all relevant and are largely unanswered yet by comparative studies. Table 1.2 lists some of the more popular architectures.

TALEN assembly methods

Hitherto, many modular assembly methods have been developed to rapidly or robustly assemble TALEN fusion proteins: from PCR-based modules to plasmid-born assembly; from one step or two step "Golden Gate" cloning strategies to FLASH (Fast ligation-based automatable solid-phase high-throughput system); from LIC (ligation-independent cloning) to iterative capped assembly (Cermak, Doyle et al. 2011, Li, Huang et al. 2011, Briggs, Rios et al. 2012, Reyon, Tsai et al. 2012, Kim, Kweon et al. 2013, Schmid-Burgk, Schmidt et al. 2013); from customized self-assembly to commercially available TALEN services (Table 1.3). A popular method for constructing TALENs is based on a Golden Gate cloning strategy that conveniently avoids mutations generated by PCR methods, and eliminates labor-intensive gel isolation of DNA fragments. One-step Golden-Gate cloning, compared with two-step Golden-Gate cloning, not only eliminates the need to characterize intermediate constructs, it is also time saving, robotics-enabling and adaptable for automated high-throughput production. The one-step clone method is composed of ready-to-use trimer or tetramer plasmid libraries, which employs about 400 plasmids. Another strategy uses 16 dimers, 64 trimers or 256 tetramers was reported

as an alternative one step Golden-Gate strategy for TALEN construction (Kim, Kweon et al. 2013, Liang, Chao et al. 2013, Yang, Yuan et al. 2013, Zhang, Li et al. 2013). In addition, the FLASH method is reported to be robust enough to permit genome-scale construction of TALENs (Reyon, Tsai et al. 2012).

Resources

There is a growing selection of web based TALEN resources, including target selection utilities, TALE technology tutorials, troubleshooting guides, mutation detection methods, specific species applications and current literature. In addition, there are TALE engineering kits available from specific laboratories or organizations such as Addgene. These kits greatly simplify acquisition and application of TALE technology and are a generous resource for scientists who are interested in TALEN research. At Addgene, a nonprofit plasmid repository (http://www.addgene.org/TALEN/), there are currently five TALEN assembly kits available to the public from the Voytas/Bogdanove collaboration, Joung lab, the Zhang laboratory, the Hornung laboratory and the Musunuru/Cowan collaboration, which harness different strategies to rapidly assemble TALENs. Several open-sources, web-server programs are listed in Table 1.4 for selection of potential TALEN target sites.

TALEN Application

Quick and easy assembly, availability of powerful resources, cross species flexibility, plus a high rate of success make TALEN technology an easy choice for Genome Editing applications. In fact, more reports of TALEN usage have been published in the last few years than for all previous artificial-nuclease-based technology applications from the last few decades. TALEN technology has been used in a wide range of organisms, including yeast, plants, algae,

protozoann, nematodes, fish, insects, mammals, and human cells. The short time since the initial TALEN reports to the plethora of species applications demonstrates the simplicity and general applicability of TALEN technology.

Animal Applications

The application of artificial nuclease based technology in animal cells far outstrips all other research areas combined. Reports began with the modification of two endogenous genes in human cells (Miller, Tan et al. 2011). This research demonstrated that TALENs could mediate NHEJ events in up to 9% of cells at the NTF3 locus and generated NHEJ in 21% of cells and HR in 16% of cells at the CCR5 locus in human K562 cells. A report quickly followed that demonstrated successful TALEN targeting of the endogenous gene, ben-1 in *Caenorhabditis elegans* to produce NHEJ events in 3.5% of the progeny (Wood, Lo et al. 2011). Soon after, TALEN function was demonstrated in zebrafish (Huang, Xiao et al. 2011), where NHEJ reached 12.25% for the tnikb locus and was also detected at the dip2a gene. Both mutations showed heritable Genome Editing events could occur in a vertebrate.

More recent achievements demonstrated TALEN-mediated disruption of the mouse Y chromosome. Traditionally, it has been difficult to perform gene knockouts on the Y chromosome because Y is composed of highly repetitive DNA that proved difficult to target (Kato, Miyata et al. 2013, Wang, Hu et al. 2013). In a demonstration of the dual integrase cassette exchange (DICE) system, using human cells, it was shown that TALENs could mediate precise integration of a “landing pad” containing markers and phiC31 and Bxb1 integrase sites. The corresponding integrases were later used as a high fidelity system to replace the landing pad with genes of interest (Zhu, Gamboa et al. 2013). In a broader TALEN study, using *Bos taurus* (cow) and *Sus scrofa* (pig), it was demonstrated that 23 of 36 TALENs had high activity in

primary cells at 15 loci resulting in mono and biallelic knockouts. In addition, up to 75% of TALEN injected embryos resulted in knockout events (Carlson, Tan et al. 2012).

Disease Models

A critical application of TALEN technology is the production of disease models, both cell-based and whole-animal-based. These models can be used to follow protein expression, interactions and distribution in cells and to track the progression of disease through the life of an animal in response to specific therapeutics. Examples of reported TALEN generated models include human cells and the mouse. TALENs have reportedly been used to generate mutations in 15 genes using cultured somatic cells or human pluripotent stem cells, with the resulting phenotypes associated with dyslipidemia, insulin resistance, hypoglycemia, lipodystrophy, motor-neuron death, and hepatitis C infection (Ding, Lee et al. 2013). In mouse, C9orf72, a gene that is known to be involved in amyotrophic lateral sclerosis (ALS) and fus, a gene involved in frontotemporal lobar degeneration (FTLD) were targeted using TALENs. C9orf72 and fus were disrupted by NHEJ in 31% and 7.5% of pups respectively and fus engaged in HR in 6.8% of pups (Panda, Wefers et al. 2013).

Curing Disease

Just as TALENs have the power to create disease models, they also have the potential to cure genetic disorders or diseases. For example, it has been speculated that some mitochondrial based diseases could be cured by reducing the expression of a disease causing protein. To this end, TALEN pairs were constructed to eliminate mutant mitochondrial DNA in patient-derived cells, resulting in a reduction of disease bearing genomes in this organelle (Bacman, Williams et al. 2013). TALENs have also been used to correct mutations in human sickle cell disease and Duchenne Muscular Dystrophy bearing cell lines (Sun, Liang et al. 2012, Ousterout, Perez-

Pinera et al. 2013). Additionally, TALENs can have antiviral affects; for example, in a recent therapeutic strategy a TALEN pair was generated to target a conserved sequence in the DNA of several hepatitis B viral strains. When tested, it was confirmed that the TALEN pair reduced viral loads in human Huh7 cells and in a mouse model (Chen, Zhang et al. 2014). Genome Editing that leads to DNA elimination or gene correction may one day make locus specific gene therapy a safe and practical clinical application.

Plants

TALEN technology has been used in dicot species such as tobacco and Arabidopsis and in monocot species such as rice, *Brachypodium*, barley and maize (Christian, Qi et al. 2013, Liang, Zhang et al. 2013, Shan, Wang et al. 2013, Wendt, Holm et al. 2013, Zhang, Zhang et al. 2013). The broad species range of these successes suggests that other plant species likely will be modified by TALENs in the near future and that TALEN-based Genome Editing will be generally applicable to most plants. TALEN mediated Genome Editing in plants was first reported in rice (Li, Liu et al. 2012). In this work the major virulence TALE AvrXa7 EBE site in the *OsSWEET14* (*Os11N3)* promoter region was disrupted through NHEJ using an engineered TALEN pair in which the spacer region overlapped the AvrXa7 EBE. In plants that sustained an alteration, the wild type TALE AvrXa7 from *Xanthomonas oryzae* could no longer bind its target and activate this sucrose efflux transporter, which would have facilitated disease by increasing the availability of nutrients to the extracellular bacterial cells.

In more recent examples TALENs were used to modify genes in barely and maize to combat the problem of phytic acid, a major storage form of phosphate in many grains that is not digestible by monogastric animals. Phytic acid is an anti-nutritional, and undigested phytic acid passes into the environment, where phosphate released from it by microbial phytases contributes

to water quality problems, such as phytoplankton blooms. TALEN-mediated NHEJ at four loci involved in the Phytic acid biosynthesis pathway produced independent maize transformants at all four loci with target-gene disruptions in about 39.1% of the transformants (Liang, Zhang et al. 2013). In another approach, regulatory elements in a barley phytase gene promoter were targeted using TALENs. This approach produced NHEJ events in the targeted promoter of up to 31% of the transformants (Wendt, Holm et al. 2013). These studies indicate that TALEN-mediated Genome Editing technology can precisely modify predetermined loci in crop plants and that continued development of this critical tool could lead to improved yields, crops that are more environmentally friendly and that are disease or stress resistant.

Understanding Gene Function

The simplicity and speed of TALEN production coupled to its major use as a mediator of gene knockouts easily lends itself to the application of identifying the function of the thousands of genes whose functions remain unknown. For example, in cultured human cells, TALENs were harnessed to knockout the nuclear gene NDUFA9, a subunit of mitochondrial respiratory chain complex I, and enabled identification of its function as an assembly factor that stabilizes the junction between membrane and matrix arm (Stroud, Formosa et al. 2013). TALENS also can be used to expand knowledge by creating new model systems. For example, by using information from known *C. elegans* mutants, the species *Caenorhabditis briggsae*, was developed into a genetic model in a period of months instead of years. Hundreds of mutations were generated in dozens of genes, presenting the opportunity to make cross species comparisons and to identify new gene functions (Wei, Shen et al. 2014). Another area of interest for TALEN applications is determining the function of specific microRNAs, since their small target footprint makes them difficult to knockout by other means. In 2013, three independent groups successfully targeted

and disrupted microRNA with TALENs in a mouse model and in human cells, leading to the construction of a human cell microRNA deletion library containing 274 loci (Hu, Wallace et al. 2013, Kim, Wee et al. 2013, Takada, Sato et al. 2013).

The role of many genes varies in different developmental stages or in different tissues. The ability to generate inducible or tissue specific knockouts is essential to understanding gene function at a deeper level. This can be accomplished by placing TALEN pairs under the control of a heat shock promoter or tissue-specific promoter. This strategy has been applied in *C. elegans* and *Ciona* to generate mutations in certain developmental stages or specific somatic tissues (Cheng, Yi et al. 2013, Treen, Yoshida et al. 2014). On a finer scale, TALENs have been used to perturb chromosomal contacts through disruption of gene loops to gain a greater understanding of co-regulated genes (Fanucchi, Shibayama et al. 2013).

Large-scale Editing

When two pairs of TALENs are used to simultaneously target different loci in a cell, this action can result in a large deletion, inversion or translocation. This strategy has many benefits when studying gene clusters, microRNAs, long non-coding RNAs that do not respond to frame-shift mutations or the effects of gross genome rearrangement (Carlson, Tan et al. 2012, Liu, Luo et al. 2013). Studies have demonstrated that deletion fragment size can range from 795 bp to 5.5 Mb when a duel TALEN strategy is applied (Ma, Zhang et al. 2012, Gupta, Hall et al. 2013). In addition to deletions, inversion events also have been reported (Carlson, Tan et al. 2012, Gupta, Hall et al. 2013, Shan, Wang et al. 2013). For example, experiments in which two pairs of TALENs were simultaneously injected into pig fibroblast cells, resulted in large deletions and inversions in 10% and 4%, respectively, of injected cells. Fibroblast cells with desirable Genome Editing events were successfully used as nuclear donors for swine cloning, which clearly

demonstrated the efficacy of TALEN technology for use in these livestock (Carlson, Tan et al. 2012). Another intense area of interest that may benefit from studies of chromosomal rearrangement is the study of cancer cell biology. For example, modeling gene rearrangement through TALEN technology contributed to the discovery of a drug resistance mechanism in prostate cancer (Nyquist, Li et al. 2013). Furthermore, translocations induced by TALENs in human cells have been used to study specific cancers. In one such report, cancer cells were generated *de novo*, and a preexisting translocation in a cancer cell line was reversed using TALENs to restore the native chromosome arrangement (Piganeau, Ghezraoui et al. 2013).

Factors That Affect TALEN Efficiency

TALENs have been successfully used in everything from invertebrates to vertebrates, unicellular to multicellular organisms, and both dicotyledonous and monocotyledonous plants, encompassing more than 25 species and resulting in hundreds of publications. This broad application complicates any general evaluation of TALEN efficiency, because of a variety of factors, such as the species involved, donor DNA for HR work, type and amount of molecule delivered, delivery methods, TALEN architecture, target choice, and detection methods. A good assessment of TALEN effectiveness will depend on a thorough evaluation of the literature that is most relevant to the species of interest. With that said, some general information gleaned from the literature can help with experimental design.

Target Species

Two of the most important factors affecting TALEN efficiency and efficacy are probably the species and the cell type. These two parameters will largely dictate the delivery method and

the type of molecule delivered (e.g. DNA, RNA or protein), as well as the outcome. Most species engage in NHEJ to a greater extent than HR in response to DSBs, but the efficiency rates for detectable edits will vary between species and may even differ between stages of the cell cycle and cell types from the same species. While it is clear that NHEJ events can be obtained in most species, not enough information is available to make general statements about the probability of success with HR.

Donor DNA for HR

Relatively few thorough HR experiments have been reported, but it is clear that some species engage HR more readily than others. Important considerations for HR experiments are the length of the donor DNA molecule, the position of the desired change relative to the DSB site, the level of heterology, inclusion of the nuclease target site in the donor, and the size of inserted DNA, if any, relative to the target. Reported Donor DNAs range from about four kb down to oligonucleotides and can be single or double stranded. The window of incorporation of genetic changes can be narrow, as demonstrated with I-Sce I in mouse cells with 80% of events having track lengths of 58 bp or less with HR occurring on either side of the DSB (Elliott et al 1998). Although less stringently tested a similarly narrow HR window was observed in two independent HR experiments using human cells (Lee et al. 2012, Porteus 2006). In plants the story may be different, although the HR frequency was reduced from 4% at ~160 bp to 0.2% at ~1500 bp distance from the DSB, the HR window in tobacco is relatively wide (Townsend et al. 2009). Donor DNA divergence is another area that requires more study, however, it has been shown that as little as 1.2% heterology can reduce HR by about six fold in mammal cells (Elliott et al 1998). Additionally, removing the nuclease recognition site from the Donor DNA would

seem like an obvious improvement, however, reports suggest otherwise. For example, removal or inclusion of the nuclease recognition site in donor DNAs used in human or tobacco cells had little effect on HR efficiency (Urnov et al 2005, Porteus 2006, Townsend et al 2009). Furthermore several studies have shown relatively high incorporation rates for large DNA segments into DSB sites using plant and animal cells, which suggests that relatively large insertions may be treated differently than donor DNA containing many small changes (Porteus 2006, Shukula 2009, Wright et al. 2005). Until more information is available, longer donor DNAs with a short relative distance between the DSB and the desired DNA change with low heterology may be good choices for donor DNA design.

Delivery Molecule

If DNA is the delivery vehicle for TALENs, there are a number of structural parameters to consider, along with the fact that DNA will leave behind a trace in the genome. Important considerations should include promoter strength and tissue specificity, independent expression or coupled expression using, e.g., a FMDV 2A type linker, TALEN architecture and codon bias, to name a few. If TALEN traces in the genome are undesirable, then the sexual nature of the cell is also a consideration. For species that can breed, traces of the transgenic TALEN DNA can be removed genetically, but, there is a cost both of time and money associated with breeding programs. For species that cannot breed, the use of RNA or protein may be a better choice to avoid retention of the transgenic TALEN DNA in the final product.

RNA is a viable nuclease delivery form for some species, and it has the benefit of not leaving exogenous DNA in the cell (Carlson, Tan et al. 2012, Davies, Davies et al. 2013). However, use of RNA adds to the complexity of Genome Editing, because it is much more

susceptible to degradation than DNA. Additionally, the generation of RNA is more cumbersome than DNA. The RNA needs to be of high quality, produced in large quantity and have a 5' cap and a poly A tail to be effective. RNA may also limit delivery options or may not be an option for some species. With that said, a growing body of evidence suggests that there may be less toxicity associated with nucleases when exposure time is limited; use of mRNA is one way of limiting that time (Porteus and Baltimore 2003, Pruett-Miller, Reading et al. 2009).

A protein form of a nuclease can be taken up directly by some cell types, if a cell penetrating peptide is present or, as demonstrated ZFNs, the protein naturally passes through membranes (Gaj, Guo et al. 2012). Examples of cell penetrating peptides include the R9 (poly arginine) and TAT (from HIV) motifs. It was recently reported that TALENs bearing either cell penetrating peptide can be taken up by human cells and efficiently mediate NHEJ. A TALEN conjugated to the R9 motif reportedly led to gene disruption in 15% of HeLa cells while a TALEN fused to TAT led to a knockout in 5% of human induced pluripotent stem cells (Ru, Yao et al. 2013, Liu, Gaj et al. 2014). Coupling TALENs and cell penetrating peptides is a relatively new area of exploration and more research is needed before general conclusions can be drawn.

Delivery Method

TALEN delivery is largely dictated by the species and tissue type, but some general observations can be made. For animals there is a wider range of delivery methods available, and transfection efficiency is generally high, while plants have fewer delivery options, and the relative transformation efficiency is generally lower.

Nuclease delivery for animal cells includes such methods as nucleofection with RNA or DNA, lipofection, microinjection of either RNA or DNA, particle based methods, and virus

mediated transfection. Differences in concentration and form of the delivery molecule, either mRNA or DNA, can contribute to Genome Editing efficiency. For example, it has been demonstrated that modification frequencies in single cell rat embryos microinjected with DNA is lower than with mRNA and that observed frequencies may be dose dependent (Tesson, Usal et al. 2011). For viral delivery, it has been demonstrated that adenovirus can deliver TALEN genes into various human cell types, but rearrangements of TALENs was a problem when performing lentiviral vector mediated transfection (Holkers, Maggio et al. 2013).

For plants, the higher efficiency delivery methods reported fall into three basic areas: electroporation, physical methods such a biolistics, and *Agrobacterium tumefaciens* infection. Unlike many animal based studies, the goal of plant transformation generally is to regenerate whole plants that are capable of passing a modification along to their progeny. Electroporation most commonly involves DNA and is usually limited to the few species that can be regenerated from protoplasts. Biolistics and *A. tumefaciens* infection are more widely applicable and normally involve DNA, either as plasmids or T-DNA, respectively. However, compared to animal cells, both methods are relatively inefficient (Barampuram and Zhang 2011). A drawback of *A. tumefaciens* delivery is that transformation of a single cell by two different *A. tumefaciens*, each harboring one component of a TALEN pair, is difficult to achieve. Due to this constraint, both TALENs in a pair are normally encoded by one plasmid, either expressing the two TALENs separately or by using a single promoter and linking both TALENs by FMDV 2A translational skipping peptide (Li, Liu et al. 2012, Zhang, Zhang et al. 2013). A very recent paper demonstrated the delivery of nucleases to tobacco cells using a modified Geminivirus system that increased Genome Editing by one to two orders of magnitude over *A. tumefaciens* based delivery methods (Baltes, Gil-Humanes et al. 2014).

TALEN Architecture

As discussed above, there are a growing number of TALE architectures, which, when combined with the numerous FokI dimer options, can make choosing the optimal TALEN design difficult. Various combinations may greatly affect TALEN performance, stability and toxicity, which directly influence outcomes. Also, architecture will determine half-site target length options and spacer length requirements. Flexibility, assembly method, success in the literature and target site characteristics are all parameters that should be carefully considered when choosing a TALEN architecture.

Target Choice

To a certain degree, DNA composition, half-site recognition lengths and spacer length can affect TALEN performance. Half-sites commonly range from 12 to 24 RVDs, and spacer length is determined by length and structure of the C-terminus and FokI fusion point, with a shorter C-terminus generally decreasing the optimal spacer length. Base composition of the target should also be carefully considered. Some rules have been suggested in the literature, but these have not been rigorously tested. Beginning the target-half site with a T, use of longer half sites with a good mix of bases and considering spacer length requirements are good starting points for target selection. The type of intended recombination is also a factor. If gene disruption by NHEJ is desired, the target should be closer to the 5' end of the gene's coding sequence and placed in an exon or on an intron-exon border. Intron disruption is usually not a good choice for gene knockout. Target placement in one of the first few exons ensures a higher probability of early protein truncation, which is more likely to lead to a nonfunctional gene and less likely to leave a lingering product that can interfere with later results. If HR is the preferred outcome, the

target should be as close to the point of change as possible. This placement increases the probability that changes provided by the donor DNA will be incorporated at the DSB site.

Validation, Detection and Estimating Efficiency

Even though TALEN-based Genome Editing is generally efficient and most TALEN pairs function effectively, it is advisable to validate the function of newly designed TALENs before using them in transformation/transfection experiments to avoid wasting time with ineffective pairs. The validation process for TALENs and other nucleases, methods for detection of editing events, and calculations for estimating efficiency of gene modification can vary greatly and often depend on the preferences of the researcher. For example, validation assays can be performed *in vitro* or *in vivo*, mutation detection can utilize a variety of effective methods and calculations of efficiency can use differing parameters that are inherent to each system or are dependent on the preferences of the researcher. All of these variables can result in a variety of estimates for nuclease activity.

Many *in vitro* validation assays for nuclease function simply rely on plasmid cleavage, using purified or partially purified nuclease extracts, to confirm specific nuclease activity. On the other hand, *in vivo* assays often depend on transient expression of a disrupted reporter gene that is repaired through homologous recombination. Validating TALEN activity using a reporter disruption or recovery assay is relatively rapid compared to a full genome modification evaluation, which can be laborious and time consuming, especially when the gene of interest does not present a selectable or easily screenable phenotype. Different reporter genes can be used for this validation screen, such as *LacZ*, Luciferase, GFP or GUS. A good general assay is the yeast based recombination assay described above, but there are times when species specific

information is desirable. In these cases the reporter typically is disrupted by a small internal duplication that contains the TALEN target sequence, and cells from amenable species such as human, mouse, Arabidopsis, tobacco or rice protoplasts, or *N. Benthamiana* leaves are transiently co-transformed with the reporter construct and TALENs as singles or pairs. The cells are given time to activate the process by repairing the reporter gene, which is then assayed for restoration of function (Kim, Lee et al. 2011, Johnson, Gurevich et al. 2013, Zhang, Zhang et al. 2013).

For mutation detection, one of the most commonly used methods is a DNA mismatch cleavage assay. For this assay, wild type and putative mutant PCR products are annealed, then a mismatch sensitive endonuclease such as Cel1 (Surveyor) or T7 endonuclease I digestion is performed. Estimates of nuclease activity are made using gel densitometry to detect cleavage at the site of mismatches (Oleykowski, Bronson Mullins et al. 1998, Miller, Holmes et al. 2007, Kim, Lee et al. 2009). In other reported methods, cloned PCR products have been randomly sequenced or subjected to deep sequencing to evaluate mutation frequencies. Another reported detection method relies on restriction fragment length polymorphism (RFLP) analysis, but requires an appropriate restriction site in the spacer region to accurately estimate event frequency (Huang, Xiao et al. 2011). This method can be enhanced through pre-digestion of genomic DNA with the respective restriction enzyme before PCR amplification.

More sophisticated mutation detection methods have also been reported that use PCR product mismatches as a basis for an assay. For example, high resolution melt analysis (HRMA) takes advantage of thermostability difference between perfectly matched and mismatched PCR fragments (Dahlem, Hoshijima et al. 2012, Hu, Wallace et al. 2013). Additionally, the heteroduplex mobility assay (HMA) uses poly-acrylamide gel electrophoresis (PAGE) or

microchip electrophoresis for easy high throughput mutation screening. This assay relies on the observation that heteroduplexes have reduced gel mobility proportional to their degree of divergence (Ansai, Inohaya et al. 2014). A similar strategy is described for screening transmission efficiency by observing PAGE migration of relatively small PCR fragments and searching for altered fragment sizes (Wei, Liu et al. 2013).

A variety of methods also have been used for estimating efficiency using cells; some studies calculate a ratio based on the total number of cells, while others calculate the efficiency based on the total number of transfected or transformed cells. It is understandable why either total cell number or only the number of transfected or transformed cells might be used, but the latter calculation does result in a higher reported efficiency. Regardless of how estimates are made, it is important for the researchers to report exactly how the analysis was performed, so readers are not misled by reported results.

NHEJ Endonuclease Enhancement

As effective as TALENs are at inducing mutations through NHEJ, methods have been reported to improve their efficiency. The majority of DSBs do not produce mutations, because they are often efficiently and precisely repaired. Therefore, enhancing the Genome Editing potential of DSBs by decreasing the probability of precise repair is one way to improve overall TALEN efficiency. A recent example of this is the coupling of TALENs with an exonuclease. This strategy involves the use of an exonuclease to remove nucleotides from the DNA ends at the DSB before a precise repair can be made. Coupling TALENs with Exonuclease 1 (Exo1) increased mutagenesis efficiency up to 30% in rat cells (Mashimo, Kaneko et al. 2013), while coupling to Exonuclease 2 (Trex2) increased efficiency approximately 144% with no obvious

toxicity in human cells (Certo, Gwiazda et al. 2012). A paper describing a meganuclease TALE fusion also reports increased efficiency when Trex2 was included with this nuclease (Boissel, Jarjour et al. 2013).

Engineered TALE Transcription Factors

As discussed above, the native function of a TALE is to act as a eukaryotic transcription activator after being produced by a bacterial pathogen and secreted into a host cell. TALEs contain a nuclear localization signal and an acid activation domain, but their basic transcriptional mechanism has not been clearly demonstrated. In plants, a native TALE shifts the transcription start site of a target gene from its normal location to a position 44-61 bp downstream of the EBE, or in the case of some TALE activated plant defense genes, the transcription start site can be basically unchanged (Boch and Bonas 2010, Hummel, Doyle et al. 2012). Even with limited information, it is easy to imagine that engineered TALEs can be used as specific transcription factors for activation and repression of target genes.

When designing engineered TALE transcription factors, the position of the target sequence appears to be relatively flexible. Positions as far as 500 bp upstream of the native target gene transcription start site have reportedly been successful, as have positions upstream, downstream or overlapping the TATA box (Maeder, Linder et al. 2013). Additionally, multiple TALEs targeted to different regions of the same promoter appear to act synergistically (Maeder, Linder et al. 2013, Perez-Pinera, Ousterout et al. 2013). This may be due to the creation of additional open chromatin as a result of TALE binding, but contradictory results cloud this issue, and additional evidence is needed to determine whether TALEs bind better to open or compact

chromatin and whether they can open silent chromatin (Maeder, Linder et al. 2013, Perez-Pinera, Ousterout et al. 2013, Scott, Kupinski et al. 2014).

Additional modifications of engineered TALEs also have reportedly improved their functions as activators. For example, it has been demonstrated that the native acidic activation domain can be replaced with herpes virus VP16 activation domain (Zhu, Yang et al. 1999). In mammalian cells, TALEs usually display higher activity when using VP16 or VP64 fused with the most commonly used TALE truncation version, Δ152/+95 or when used with the FLASH system (Miller, Tan et al. 2011, Maeder, Linder et al. 2013). It has also been demonstrated that the TALE acidic activation domain can be replaced by a repressor, such as the mSin interaction domain (SID), and that this fusion can reduce expression of a target gene in plant and animal cells and that just physical binding can block transcription in bacteria (Table 1.5) (Cong, Zhou et al. 2012, Mahfouz, Li et al. 2012, Crocker and Stern 2013, Politz, Copeland et al. 2013).

TALE proteins could be fused to a variety of functional units other than nuclease, activator or repressor modules. Such fusions might include, methylases, demethylases, recombinases, transposases, integrases, flouresent proteins, and protein interaction domains to name a few. Many of these fusions could be used to generate new tools, control gene expression at an epigenetic level or study the effects of new protein interactions in a chromosomal context. Examples in the literature include A TALE TET1 hydroxylase fusion that was successfully used to demethylate CpG sites in human gene promoters to stimulate gene expression and a TALE Gin recombinase fusion that lead to recombination in bacterial and mammal cells (Mercer, Gaj et al. 2012, Maeder, Angstman et al. 2013).

Opportunity For Improvement

Progress in TALE-based technology development has occurred at an astounding pace, which speaks volumes about the positive change that a diverse group of research scientists can bring to a new technology when they have a common goal; in this case, creating a universally powerful tool for Genome Editing. On the other hand, there are still many areas that could be improved, including toxicity tests, alternatives for the T_0 recognition domain, evaluation of alternative RVDs, and an improved nuclease domain.

Toxicity Test

Toxicity was identified as an issue early in the development of nuclease based Genome Editing, and it still is an apparent issue with TALENs although it may be to a lesser degree. While progress has been made in testing the relative functionality of nuclease designs using methods such as the yeast-based nuclease assay, a simple, general test for nuclease toxicity has yet to be developed. This is complicated by the fact that toxicity is not a universal issue, but may be more of a species-dependent or cell-specific problem. Often it is easier to simply test several engineered nucleases, then use the one that works best. One way of screening for toxicity is to evaluate relative survival based on co-transformation with a marker gene. While this approach works, it is not always applicable or convenient, especially in plant cells that are not amenable to protoplast isolation or in animal cells in which nucleases are delivered to individual cells one at a time. Improvements in understanding toxicity and off target DNA binding will help, but some out-of-the box thinking will be required to solve this problem.

T_0 Domain

Another area for improvement is the repeat 0 region of the TALE DNA binding domain. Repeat 0 binds to the first base in an EBE, which is generally a T. Expanding the spectrum of this specificity would further increase TALEN flexibility by expanding the number of potential target sequences. Structural data show that repeat 0 has a helix-loop-helix motif and that T specificity is based on its interactions with W232 in the KQWSG sequence of the loop (Mak, Bradley et al. 2012). When mutations are generated in this region, the specificity for a T is relaxed; however, these gains appear to be context specific. That is to say, moving the mutation to a related TALE or using the mutation in a TALEN, TALE activator or a TALE recombinase changes the recognition pattern (Doyle, Hummel et al. 2013, Lamb, Mercer et al. 2013, Tsuji, Futaki et al. 2013). Additionally, it has been demonstrated that repeat 0 from *R. solanacearum* preferentially activates promoters containing EBEs with a 5' G. However, when tested in a PthXo1 construct, although the *R. solanacearum* repeat 0 changed specificity to G, it did so at the expense of activity (de Lange, Schreiber et al. 2013, Doyle, Hummel et al. 2013). Interestingly, an optimized TALE scaffold using the N terminal 207 a.a. plus the C terminal 63 a.a. from AvrXa10 reportedly exhibited similar activity whether a target DNA sequence contained a 5' C, A, T or G (Sun, Liang et al. 2012). These reports suggest that the specificity of repeat 0 for a T is not immutable, but they also suggest that a larger scale effort will be needed to effectively accommodate A, C and G at position 0 in many contexts.

Alternative RVDs

The most commonly used RVDs are HD, NI, NG, and NN for the recognition of C, A, T and G respectively. While these RVDs work well, they each have strengths and weaknesses. For

example, NN recognizes A about as well as G, HD interacts strongly with C, but is sensitive to 5-methylcytosine (5mC), NI has a relatively weak interaction with A, and NG interacts strongly with T, but also recognizes 5mC (Deng, Yin et al. 2012, Streubel, Blucher et al. 2012, Meckler, Bhakta et al. 2013, Lin, Fine et al. 2014).

Many other RVDs exist, but most have not been stringently tested. Of the few examined, it has been found that NK has a higher specificity for G over NN, but also lower affinity, which results in less TALEN activity in yeast and human cell based recombination assays (Huang, Xiao et al. 2011, Mahfouz, Li et al. 2011, Cong, Zhou et al. 2012). In contrast, RVD NH has been shown to have higher specificity for G over NN without sacrificing affinity (Christian, Demorest et al. 2012, Cong, Zhou et al. 2012, Streubel, Blucher et al. 2012, Meckler, Bhakta et al. 2013). Also, N* (missing the 13^{th} aa) was shown to bind 5mC as well as C, A, T and G; making it a generic RVD (Valton, Dupuy et al. 2012).

In addition to differences in binding affinity and specificity among the various RVDs, there exits polarity in the TALE DNA binding domain, with binding being strongest in the N-terminal region and weakest in the C-terminal end. This polarity makes a TALE much more sensitive to 5' target site mismatches than 3' mismatches, and at least one study has shown that RVD HD has a stronger affinity for C when located nearer to the N-terminus (Streubel, Blucher et al. 2012, Meckler, Bhakta et al. 2013, Lin, Fine et al. 2014).

The ability to specifically recognize modified and unmodified DNA has important implications in epigenetic and cancer research, but also has consequences for general targeting efficiency. Until the alternative RVDs are systematically tested for binding affinity, specificity and position effect, it will remain an open question as to which RVDs are best for any specific application.

Improved Nuclease Design

TALENs and ZFNs could benefit from an improved nuclease design. Both commonly use FokI, but FokI functions as a homodimer and engages in *trans* cleavage. Each of these characteristics contributes to cell toxicity, which has been partially addressed by improved heterodimeric forms of FokI. However, there is still much room for improvement. There are several reported examples of alternative nucleases being used, such a nickases, alternative dimeric nucleases and attempts to create monomeric TALENs that have the potential to improve application of TALENs under some conditions.

Nickases

A FokI nickase was produced by mutating the catalytic site (D450A) of one partner in a FokI heterodimer to induce a single strand break rather than a DSB (Kim, Kim et al. 2012, Ramirez, Certo et al. 2012, Wang, Friedman et al. 2012). The idea was to suppress the NHEJ rate in favor of HR, which worked; however, the HR rate was also greatly reduced (van Nierop, de Vries et al. 2009, Davis and Maizels 2011, Metzger, McConnell-Smith et al. 2011). A different approach to engineering a nickase reportedly fused the monomeric nickase MutH to a TALE DNA binding domain. MutH is a DNA mismatch repair endonuclease that recognizes the sequence GATC. This fusion protein was tested in a plasmid cleavage assay *in vitro* and shown to function as a nickase (Gabsalilow, Schierling et al. 2013). Further exploration of these and other TALE-based nickases may eventually lead to higher HR efficiencies.

Alternative Nucleases

Because FokI *trans* activity appears to cause some toxicity, the development of alternative nuclease partners may improve the general application of TALENs in Genome

Editing. The Type IIP restriction endonuclease PvuII was fused with a TALE, a catalytically inactive I-SceI and a zinc finger array. These new homodimeric nucleases were shown to have high specificity, but retained the PvuII recognition sequence as part of the overall target site (Fonfara, Curth et al. 2012, Schierling, Dannemann et al. 2012, Yanik, Alzubi et al. 2013). This work demonstrates that other nucleases can be paired with various DNA binding domains to create greater specificity. One wonders if a restriction enzyme with a four base specificity may be more versatile or if base specificity could be eliminated by mutagenesis in an alternative nuclease.

Monomeric nuclease

Because FokI nuclease requires dimerization and thus the use of TALEN pairs, the development of monomeric forms may improve the general application of TALENs in Genome Editing. A real breakthrough may come from a fusion of a nonspecific monomeric nuclease and a TALEN. Such a nuclease would streamline TALEN production by halving the development work load and simultaneously produce a more compact enzyme. Some attempts have been made to design a FokI monomer and to make TALE fusions to known monomeric nucleases.

A monomeric engineered nuclease has been reportedly isolated using a high throughput screen. The nuclease included a TALE DNA binding domain fused to tandem *FokI* cleavage domains separated by a 95 aa linker. Yeast and human assays indicated that, compared with a standard TALEN heterodimer, this nuclease only had 9% to 36% activity and exhibited significantly higher cytotoxicity. These characteristics may result from the shorter target leading to more off-target DSBs within the human genome, or they may be more related to the monomeric nature of this FokI design (Sun and Zhao 2014).

A “mega-TAL” fusion protein was described recently as another strategy for monomeric nuclease construction. This nuclease is composed of a TALE DNA binding domain and the site specific LAGLIDADG monomeric meganuclease I-Anil. In this fusion, I-Anil retains its native target site specificity, which is further enhanced by the TALE DNA binding domain. This architecture, when combined with the Trex2 exonuclease, demonstrated high specificity and increased rates of gene disruption, which exceeded 70% in human cells (Boissel, Jarjour et al. 2013).

In another monomeric example, a TALE DNA binding domain was fused to the site specific GIY-YIG monomeric meganuclease I-TevI catalytic domain. This domain was previously shown to be functionally separate from its native DNA binding domain, but it still retains partial site specific recognition, as the sequence CNNNG is required for nuclease action (Edgell, Stanger et al. 2004). Additionally, it was demonstrated that this I-TevI nuclease domain can be fused to other DNA binding domains to produce hybrid nucleases (Kleinstiver, Wolfs et al. 2012). The I-TevI::TALE fusion proved to be a functional monomeric nuclease, and, as expected, retained the CNNNG sequence requirement as part of its overall target sequence. This fusion was shown to have an activity of 8.9% compare to 6.9% for a standard TALE::FokI fusion in plants and 4.3% versus 9.2% respectively in hamster CHO-K1 cells (Beurdeley, Bietz et al. 2013).

Ultimately, development of a monomeric nuclease may need to be coupled to a trigger mechanism, such as a conformational shift that activates the nuclease only after tight TALE domain binding to the target site. Such a mechanism is suggested by the way the DNA binding domain of FokI may sequester its nuclease domain in the native structure (Wah, Hirsch et al. 1997).

Conclusions

The development of TALENs and related TALE technology is moving at a fast pace. The user friendly structure and availability of public resources have already changed the perception of DNA binding domain engineering from one of difficulty to one of ease, which has appealed to a broader audience of researches. In response, volumes of data have been published demonstrating TALEN mediated Genome Editing and the utility of other TALE based technologies in a wide variety of organisms. The technology still presents some challenges and there are many unknowns. Over time, solutions will be presented and many more discoveries will be made. TALE technology appears to be positioned to move targeted manipulation of the genome well beyond previous platforms and will be useful into the foreseeable future even as newer genome modifying technologies such as CAS9/CRISPR are developed.

Figures And Tables

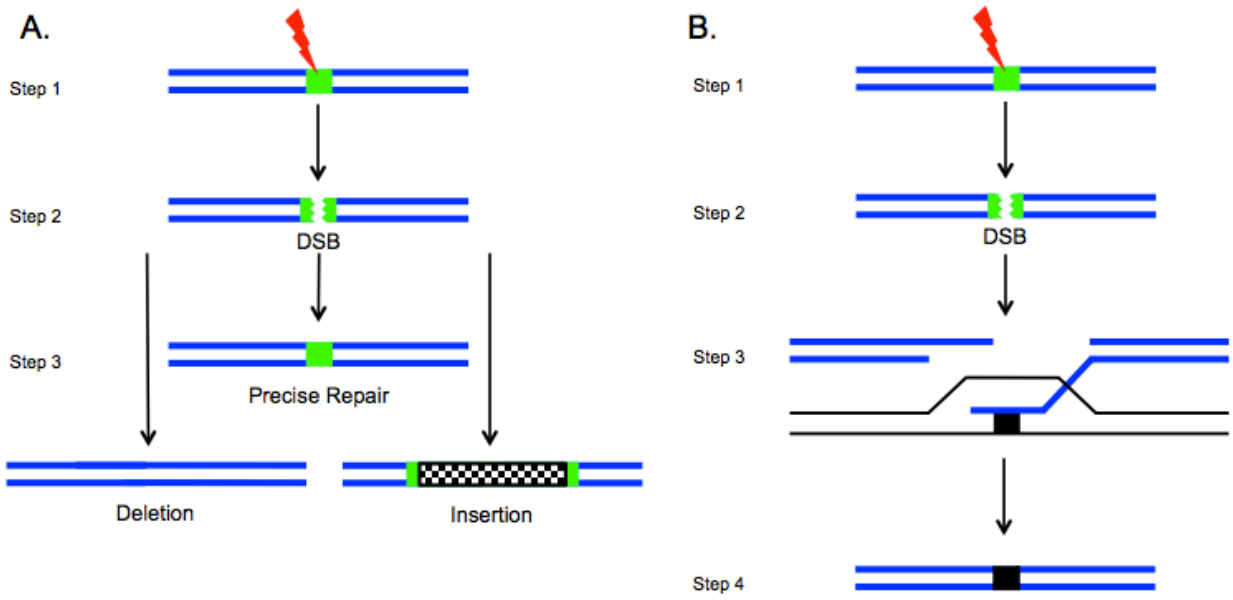


Figure 1.1. Simplified models of the NHEJ and HR processes are depicted in 1A and 1B. Double stranded chromosomal DNA is shown as two solid blue bars, the the green box is the DSB point, the red lighting symbol is an external DNA damaging agent, the checkered box is inserted DNA and the lighter black double bars are donor DNA carrying an orange box to represent change for HR. In A step 1, DNA is damaged resulting in a DSB as in step 2. In step 3, DNA is either repaired precisely or DNA is removed by repair enzymes and re-ligated resulting in a deletion as shown by the loss of the green box or DNA is inserted as depicted by addition of the checkered box. In B step 1, DNA is damaged resulting in a DSB as in step 2. In step 3, DNA is removed by repair enzymes then a free 3' end of the damaged DNA invades the donor DNA at a point of homology and repair enzymes copy information from the donor using the free 3' end of the damaged DNA as a primer. After extension, homology is found between the free 3' ends of the damaged DNA and the break is repaired by further polymerization and ligation as in step 4 resulting in gene conversion in the chromosome.

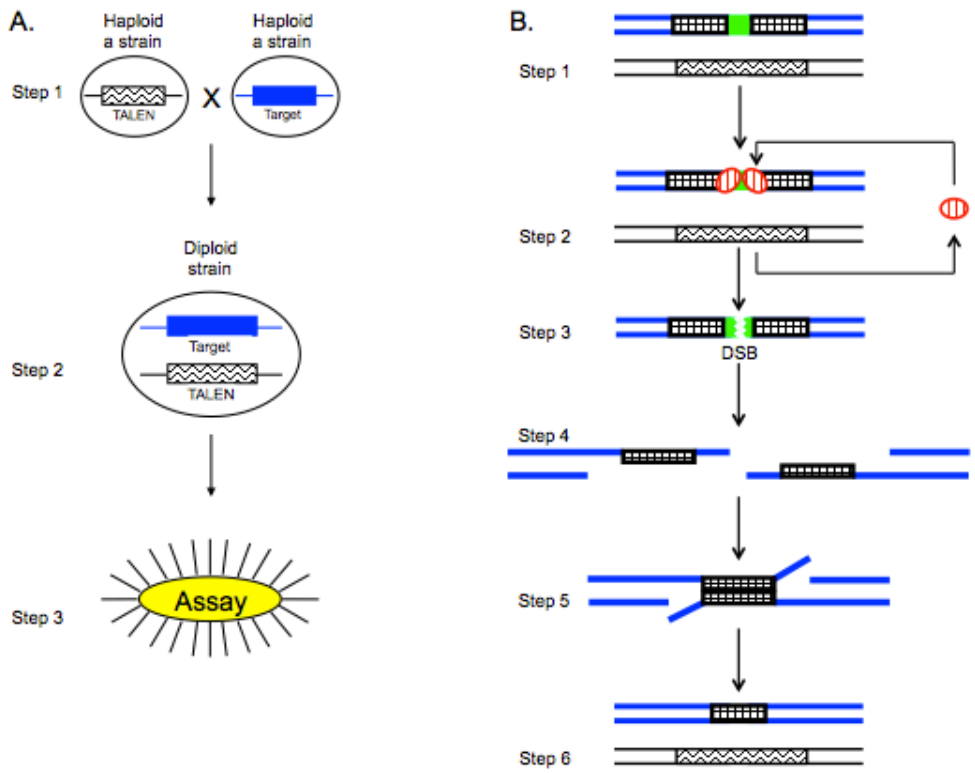


Figure 1.2. A simplified Macro and molecular depiction of the yeast recombination assay. For A, yeast are represented by large ovals with black outlines. The nuclease expression and nuclease target constructs are shown as labeled boxes within each yeast cell. Step 1 of A, haploid a and a yeast strains are transformed with either a nuclease expression vector or an inactive target assay vector containing an internal duplication and a nuclease target site. Transformed yeast cells are selected then crossed to form diploids as in step 2. Diploid yeast are then assayed by appropriate means to gauge the relative activity of the nuclease as indicated by step 3. For B, the target plasmid is shown as blue double bars representing the target with small internal duplications (black boxes with grid pattern) separated by the nuclease target (green box). The nuclease expression plasmid is represented by black bars with a nuclease coding region (black box with wavy lines) while nuclease protein is depicted as red striped ovals. Step 1 of B depicts the nuclease and inactive target vector in diploid yeast. In step 2, the nuclease is expressed then binds the target sequence in pairs where they cleave the target sequence as in step 3. In step 4, repair enzymes remove 5' DNA at the DSB site to generate free 3'ends that find homology with each other in the duplicated region as in step 5. Repair enzymes eliminate unpaired ends, fill in missing DNA and ligate the plasmid together to form a functional marker gene to be assayed as in step 6.

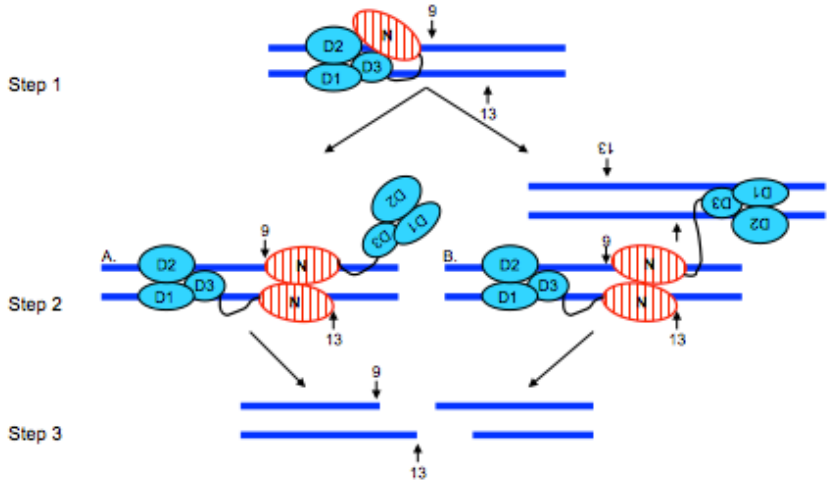


Figure 1.3. Simplified models of the FokI restriction endonuclease reaction process. Double blue bars represent double stranded DNA and FokI is represented by ovals with DNA binding subdomains labeled D1, D2, and D3 and a tethered nuclease domain labeled N. FokI cleavage points are indicated by arrows and the numbers 9 and 13. Crystal structure data indicates that FokI binds DNA as a monomer and further suggests that the nuclease domain is sequestered. Two cleavage models are shown in step 2, marked A. and B. In 2A, FokI bound to DNA is joined by a free monomer while the nuclease domains are extended and form a dimer interface to cleave the DNA at positions 9 and 13 as shown in step 3. In step 2B, FokI bound to DNA is joined another FokI monomer bound to DNA, the nuclease domains are extended and form a dimer interface to cleave at positions 9 and 13 as shown in step 3.

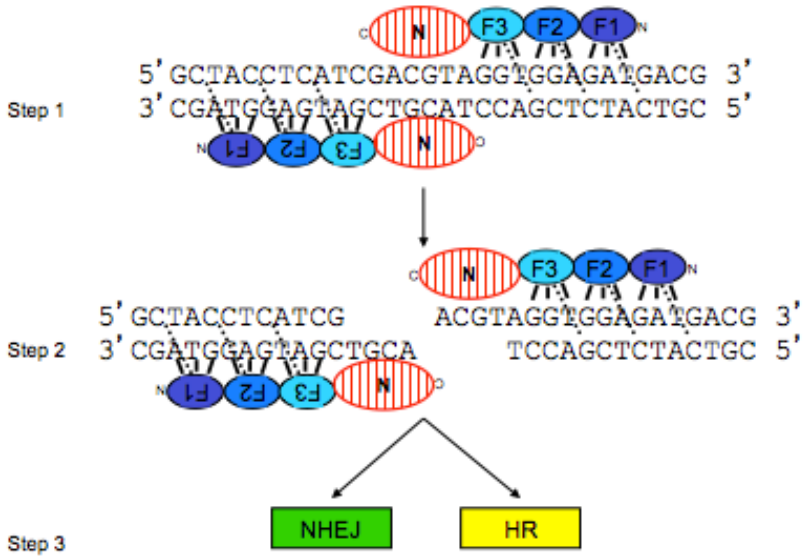


Figure 1.4. Depiction of a pair of zinc finger nucleases (ZFNs). Individual ZFs are shown as colored ovals marked F1, F2, F3 and the FokI nuclease domain is indicated by the red striped oval marked N. Small N and C indicate termini. Black bars indicate binding of individual ZFs to DNA triplets and dotted lines indicate cross strand interactions. In step 1, individual ZFNs bind their target DNA to form a pair with nucleases positioned over a six base spacer. In step 2, the nuclease domains form a dimer interface and cleave the target DNA, leaving a four base 5' overhang. This is followed by repair using a NHEJ or HR pathway as in step 3. Note that ZFNs bind DNA backward with respect to its N-terminus and the 5' DNA end.

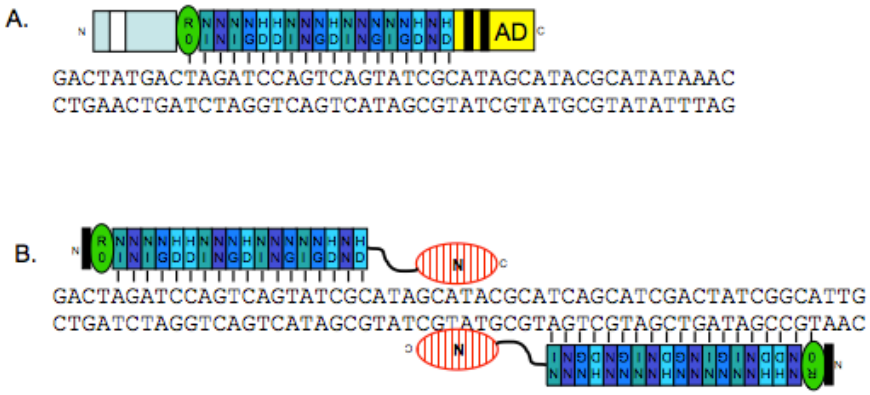


Figure 1.5. A TALE activator and a pair of Engineered TALENs are depicted. The N terminus of a TALE effector,7A, contains bacterial secretion and translocation signals as shown by the white box, repeat 0 is shown as a green oval marked with R0, individual RVDs are depicted as small colored rectangles with their nucleotide binding code indicated, the large yellow box marked AD, represents a transcription activation domain and nuclear localization signals (NLS) are shown as black boxes. In 7B the large red striped ovals in 7B, marked N, represent tethered FokI nuclease domains. Small N and C indicate termini. 7A represents a native 18 RVD TALE transcription effector that recognizes a target that is 19 bp long and activates transcription near the target site. 7B is an engineered 18 RVD TALEN pair that position FokI nuclease domains over a 12 base spacer where they will cleave the DNA to mediate NHEJ or HR. Note that the TALE DNA recognition domain binds DNA in a forward orientation with respect to its N-terminus and the 5' DNA end.

Table 1.1 Summary of examples of FokI heterodimer pairs used for site specific nucleases in gene editing studies.

Name	**Mutations**	**Reference**	**Notes**
RV-DA	Left D483R, I538V Right R487D, I499A	Szczepek et al. 2007	reduced activity *in vitro* and *in vivo*
KK-EL	Left Q486E:I499L Right E490K:I538K	Miller et al. 2007	
ELD-KKR	Left Q486E,I499L,N496D Right E490K,I538K,H537R	Doyon et al. 2011	higher nuclease activity and increased temperature tolerance
Sharkey	S418P and K441E	Guo et al. 2010	Improved nuclease activity and compatible with heterodimer forms
RE	S418R and K441E	Sizova et al. 2012	Increased DNA binding and decreased nuclease activity

Table 1.2 Relevant TALEN architectures information.

Name	Truncation N-terminal/C-terminal	Spacer Requirement (bp)	FokI type	Codon Bias	Reference
	-288/+295	16-31	homo	no	Li et al. 2011
	Δ152/+63	12-21	homo	no	Miller et al. 2011
	-287/+231	15-27	homo	no	Christian et al. 2010
	-153/+17, -153/+47	12, 12-15	KV/EA	no	Mussolino et al. 2011
	-111/+42	N/A	N/A	N/A	Meckler et al. 2013
	Δ152/+14	12-14	DAS/RR sharkey	Human codon	Kim et al. 2013
	Δ152/+18	13-16	homo	no	Christian et al. 2012
	-240/+63	14-16	NEL/CKK	Silkworm codon	Ma et al. 2012
Goldy	Modified Δ152/+63	15-16	homo	no	Bedell et al. 2012

Table 1.3 Commercial available TALEN production and services.

Company Name	**Location**
Cellectis Bioresearch	Paris, France
GeneCopoeia	Maryland, USA
PolyGene Transgenetics	Zurich, Switzerland
Stratech Scientific Ltd	Suffolk, UK
Excellgen	Seoul, Korea

Table 1.4 Open-source, web-server programs for selection of potential TALEN target sites selection utilities, TALE technology tutorials, troubleshooting guides, mutation detection methods, specific species applications and current literature.

Web Site	**Resources**	**Notes**	**Reference**
http://eendb.zfgenetics.org/index.php		Including TALENs, ZFNs, and CRISPR/Cas systems. Collects targeted genes in different organisms, offers knowledge about TALEN construction, genome modification methods, mutation detection methods	Xiao et al. 2013
http://egenome.org/			
http://www.genome-engineering.org/taleffectors/			
https://tale-nt.cac.cornell.edu/	TALEN-NT 2.0	TALEN design and off-target finder with support of restriction length fragment polymorphism (RLRP) mutation detection	Doyle et al. 2012
http://idtale.kaust.edu.sa	idTALE	TALEN design and search for potential TALENs binding sites within five different organisms	
http://www.talendesign.org/mojohand_main.php	Mojo hand	TALEN design with support of automated download and extract exons for most genes and restriction length fragment polymorphism (RLFP) mutation detection	
http://galaxy.informatik.uni-halle.de	TALgetter	predict TAL effector target site	Grau et al. 2013
https://www.addgene.org/	Plasmids kits and information		
http://www.hornunglab.de/TALEN.html			
http://www.talenlibrary.net/			
http://www.e-talen.org/E-TALEN/			
http://talengineering.org/platforms-flash.htm			
http://zifit.partners.org/ZiFiT/			
http://baolab.bme.gatech.edu/Research/BioinformaticTools/TAL_targeter.html	SAPTA	TALEN design by scoring algorithm based on TALEN activities measurement	Lin et al. 2014

Table 1.5 Reported repressor domain utilized in TALE repressor fusion protein in different organisms.

Organism	**Repression domain**	**References**
Arabidopsis	EAR-repression domain (SRDX)	Mahfouz et al. 2012
Mammalian cells	mSin interaction domain (SID)	Cong et al. 2012
E. coli	No repression domain fused, simply by physically block	Politz et al. 2013
Drosophila	Krüppel and Hairy repression domain, Hairy is stronger	Crocker and Stern 2013

CHAPTER 2

TAL NUCLEASES (TALNS): HYBRID PROTEINS COMPOSED OF TAL EFFECTORS AND FOKI DNA-CLEAVAGE DOMAIN

A paper published in *Nucleic Acids Research*‡

Ting Li [1,*], Sheng Huang [1,*], Wen Zhi Jiang [2], David Wright [1], Martin H. Spalding [1], Donald P. Weeks [2], Bing Yang [1, †]

[1] Department of Genetics, Development and Cell Biology, Iowa State University, Ames, IA 50011 and [2] Department of Biochemistry, University of Nebraska, Lincoln, NE 68588, USA

Abstract

DNA double-strand breaks enhance homologous recombination in cells and have been exploited for targeted genome editing through use of engineered endonucleases. Here we report the creation and initial characterization of a group of rare-cutting, site-specific DNA nucleases produced by fusion of the restriction enzyme FokI endonuclease domain (FN) with the high-specificity DNA binding domains of AvrXa7 and PthXo1. AvrXa7 and PthXo1 are members of the transcription activator-like (TAL) effector family whose central repeat units dictate target DNA recognition and can be modularly constructed to create novel DNA specificity. The hybrid FN-AvrXa7, AvrXa7-FN and PthXo1-FN proteins retain both recognition specificity for their target DNA (a 26 bp sequence for AvrXa7 and 24 bp for PthXo1) and the double-stranded DNA cleaving activity of FokI and, thus, are called TAL nucleases (TALNs). With all three TALNs, DNA is cleaved adjacent to the TAL-binding site under optimal conditions *in vitro*. When

‡ Reprinted with permission of *Nucleic Acids Research*, (2011) 39 (1): 359-372.

* Both authors contributed equally to this work.

† To whom correspondence should be addressed. Tel: 515 294 2968; Fax: 515 294 5256; Email: byang@iastate.edu

expressed in yeast, the TALNs promote DNA homologous recombination of a *LacZ* gene containing paired AvrXa7 or asymmetric AvrXa7/PthXo1 target sequences. Our results demonstrate the feasibility of creating a tool box of novel TALNs with potential for targeted genome modification in organisms lacking facile mechanisms for targeted gene knockout and homologous recombination.

Introduction

Perhaps the most significant application of endonucleases in the post-genome era is their coupling with custom-engineered proteins that recognize long stretches of DNA sequences and their application for targeting specific genes for knockout or gene replacement (1). The key component of these engineered nucleases is the DNA recognition domain that is capable of precisely directing the nuclease to the target site for the purpose of introducing a DNA double strand break (DSB). These breaks are principally repaired by one of two pathways, non-homologous end-Joining (NHEJ) or homologous recombination (HR). Repair by NHEJ often results in mutagenic deletions/insertions in the targeted gene. Moreover, DSBs can stimulate HR between the endogenous target gene locus and an exogenously introduced homologous DNA fragment with desired genetic information, a process called gene replacement or genome editing (2, 3, 4). To date the most promising method involving genome editing is the use of custom-designed zinc finger nucleases (ZFNs) (5). ZFN technology involves the use of hybrid proteins derived from the DNA binding domains of zinc finger (ZF) proteins fused with the nonspecific DNA cleavage domain of the endonuclease FokI (6, 7).

FokI endonuclease, a type IIS nuclease, was first isolated from the bacterium *Flavobacterium okeanokoites* (8). The nuclease consists of two separate domains, an N-terminal

DNA binding domain and a C-terminal DNA cleavage domain. The DNA binding domain recognizes the non-palindromic sequence 5'-GGATG-3' while the catalytic domain cleaves double-stranded DNA non-specifically at a fixed distance of 9 and 13 nucleotides downstream of the recognition site (9, 10). FokI exists as an inactive monomer in solution and becomes an active dimer upon binding to its target DNA and in the presence of specific divalent metals (11). It has been proposed that to form a functional complex, two molecules of FokI each bind to their recognition sequence in a double stranded DNA molecule and then dimerize in such a way that the two nuclease domains unite to form a functional endonuclease that cleaves both DNA strands at sites downstream of the FokI recognition sequence (12, 13, 14).

ZFNs can be assembled as modules that are custom-designed to recognize selected DNA sequences, typically 18 to 24 bp with a 4-7 bp spacer between two half sites (15, 16, 17). Following binding at the preselected site, a DSB is produced by the action of the ZFN's FokI cleavage domain (4). ZFN technology has been successfully applied to genetic modification in a variety of organisms, including yeast, plants, mammals, and even human cell lines [reviewed in (18, 19)]. Despite the promise of ZFN technology in basic and applied research, widespread adoption of this technology has been hampered by a bottleneck in custom-engineering zinc fingers that possess the requisite high specificity and affinity for preselected DNA target site. Such engineering is labor intensive, time consuming and associated with high rates of failure (20, 21). Since the effectiveness of these endonucleases depends almost solely on their DNA binding specificity, the required specificity theoretically can be supplanted by any high-fidelity DNA binding domain when fused with a functional endonuclease domain. The TAL effector proteins from a number of bacterial plant pathogens represent a potential source of modularly constructed DNA binding domains that could supply the needed DNA binding specificity.

TAL effectors belong to a large group of highly conserved bacterial proteins that exist in various strains of *Xanthomonas* spp. and are translocated into host cells by a type III secretion system and, thus, are called type III effectors [reviewed in (22)]. Once in host cells, some TAL effectors have been found to transcriptionally activate their corresponding host target genes either for strain virulence (ability to cause disease) or avirulence (capacity to trigger host resistance responses) dependent on the host genetic context (23, 24, 25, 26). Each of these bacterial effectors contains a functional nuclear localization motif and a potent transcription activation domain that are characteristic of eukaryotic transcription activators. Each TAL effector also contains a central repetitive region consisting of varying numbers of repeat units of 34 amino acids. It is this repeat region that recognizes a specific DNA sequence and determines the biological specificity of each effector [Figure 2.1 A, and reviewed in (27)]. Each repeat is nearly identical except for two variable amino acids at positions 12 and 13, the so called repeat variable di-residues (RVD) (28). Recent studies have revealed that the recognition of DNA sequences within the promoter regions of specific host target genes is defined by the repeat regions of TAL effectors. The DNA sequence recognition is based on a fairly simple code where one nucleotide of the DNA target site is recognized by the RVD of one repeat (i.e., one repeat/one nucleotide recognition). The sequential tandem array of repeats thus specifies the DNA sequence that will be bound (28, 29). The majority of naturally occurring TAL effectors contains repeat units in a range of 13 to 29 that presumably recognize DNA elements consisting of the same number of nucleotides. Thus, in theory, the so called TAL recognition code can be used as a guide for custom-design of novel TAL effectors with the required specificity to target a single locus in a genome.

AvrXa7 and PthXo1 are TAL type III effectors from *Xanthomonas oryzae* pv. *oryzae*

(Xoo), the causal pathogen of bacterial blight of rice. They each contain a unique combination of RVDs in 26 and 24 repeats, respectively (Figure 2.1B and 2.1C). For some Xoo strains, AvrXa7 is a key virulence factor in susceptible rice. On the other hand, it is also an avirulence determinant in the otherwise resistant plants containing the cognate resistance gene *Xa7* (30, 31). As the essential virulence factor, AvrXa7 activates the rice gene *Os11N3* to induce a state of disease susceptibility (manuscript submitted). The gene induction by AvrXa7 is mediated through its recognition of a specific DNA element within the promoter region of *Os11N3*, an element we refer to here as the effector binding element (EBE) (sequence shown in Figure 2.1B). Similarly, PthXo1 is also an essential virulence factor for some strains of *X. oryzae* pv. *oryzae* and activates the rice gene *Os8N3* to promote bacterial multiplication and disease development (24). PthXo1 recognizes an EBE of 24 nucleotides within the promoter of *Os8N3* (32) (sequence shown in Figure 2.1C).

As a proof-of-principle, we have tested the feasibility of generating a new type of DNA sequence-specific endonuclease by utilizing the sequence specificities of AvrXa7 and PthXo1 and the catalytic activity of the endonuclease FokI. Here we report the creation of TAL effector nucleases (TALNs) by fusing the full-length AvrXa7 TAL effector or the full-length PthXo1 TAL effector to the FokI nuclease domain and the characterization of their nuclease activities both *in vitro* and *in vivo* using a yeast cell assay.

Results

Construction of the chimeric genes for FN-AvrXa7, PthXo1-FN and PthXo1-FN

AvrXa7 and PthXo1 are two naturally occurring TAL effectors containing a central region of 26 and 24 repeat units, respectively, with, like their relatives, the last repeat containing

only the first 20 amino acid residues similar to other repeats. The specific arrays of 26 RVDs in AvrXa7 and 24 RVDs in PthXo1 make the target sequences of AvrXa7 and PthoXo1 unique in comparison with all other TAL effectors [Figure 2.1B and 2.1C; (26)]. AvrXa7 binds to a specific 26 bp promoter element, the effector binding element or EBE, in the rice *Os11N3* gene through its DNA binding repeats [Figure 2.1 B; (32); and submitted manuscript], while PthXo1 binds to a specific 24 bp sequence in the promoter region of *Os8N3* (24). The nucleotide "T" that precedes all known TAL effector target sites and is essential for the target gene activation (28, 29) is also present immediate upstream of the 1st nucleotide of EBEs of *Os8N3* and *Os11N3* and, therefore, is treated as part of EBE throughout this paper. We reasoned that a chimeric protein composed of AvrXa7 or PthXo1 and a DNA cleavage domain of an endonuclease might function in recognizing the *Os11N3* and *Os8N3* target sequences, respectively, and cleaving DNA adjacent to the recognition site. The DNA cleavage domain of the endonuclease FokI was chosen due to its well-documented nonspecific catalytic activity when linked with other DNA binding domains, such as zinc finger proteins. Reflecting the configuration of the chimeric protein we designed, FN-AvrXa7 was constructed by fusing the DNA sequence encoding the full-length AvrXa7 TAL effector downstream of the DNA sequence encoding the cleavage domain of FokI. AvrXa7 and PthXo1 constructs were also made that each express hybrid proteins with the FN domain at the C-terminus of each TAL effector. The resulting chimeric genes were predicted to encode proteins of 1645, 1648 and 1574 amino acid residues, respectively (Figure 2.1D, 2.1E and 2.1F). The 196 amino acid FokI domain is linked by 4 amino acid residues to the 1459 amino acids of AvrXa7 in FN-AvrXa7 and by 2 amino acids in both AvrXa7-FN and PthXo1-FN.

Transcription enhancement by FN-AvrXa7 in vivo

In addition to a DNA binding domain, the TAL effectors contain a potent C-terminal transcription activation domain (27). Thus, we reasoned that by placing the FokI nuclease domain at the N-terminus of TAL effectors, the activation domain would remain functional. We could then take advantage of transcription enhancement as an indirect measure of the DNA binding ability of the hybrid protein, FN-AvrXa7 in this case, or of any newly synthesized TAL fusion protein, in general.

We adapted a modified *Agrobacterium tumefaciens* mediated transient expression assay that has been successfully used for studying the interaction of TAL effectors with their target host genes (26, 29). In our case, the "reporter" construct contained the gene encoding a green fluorescence protein (GFP) under the control of the *Os11N3* promoter that contains the AvrXa7 EBE. Two "effector" constructs were made to express either AvrXa7 or FN-AvrXa7 under control of the strong and constitutive CaMV 35S promoter. Both reporter and effector genes were delivered by injection of *Nicotinana benthamiana* leaves with *Agrobacterium tumefaciens* to allow co-expression of genes from the effector and reporter constructs. Both AvrXa7- and FN-AvrXa7-containing constructs induced the expression of GFP while the construct lacking either AvrXa7 or FN-AvrXa7 did not (Figure 2.1G). These results indicated that the hybrid FN-AvrXa7 retained the DNA binding ability of AvrXa7, and that the transient expression assay could provide a means to test the DNA binding ability or specificity of TAL-derived hybrid proteins *in vivo*.

Overexpression and purification of FN-AvrXa7, AvrXa7-FN and PthXo1-FN

The FN-AvrXa7 and AvrXa7-FN were each cloned into an overexpression vector in frame with an N-terminal 6Xhistidine tag to allow for affinity chromatography purification on a

Ni-containing column after overproduction of the TAL hybrid protein in *E. coli*. The proteins were successfully isolated in a relatively high purity. The identity of FN-AvrXa7 and AvrXa7-FN were further confirmed with protein blot analysis using an antibody against the FLAG epitope that was integrated after the repeat regions of AvrXa7 and PthXo1. The FLAG-tagged proteins detected were approximately 175 KD, the calculated sizes of FN-AvrXa7 and AvrXa7-FN, respectively. IPTG-induced *E. coli* cells overexpressing FN-AvrXa7 or AvrXa7-FN did not exhibit any obvious growth defects (data not shown). Overexpression and purification of PthXo1-FN were achieved in a similar fashion to that for FN-AvrXa7 and AvrXa7-FN.

DNA binding activities of AvrXa7-FN and FN-AvrXa7 in vitro

AvrXa7-FN and FN-AvrXa7 purified from *E. coli* were tested for their DNA binding specificities *in vitro*. The abilities of purified AvrXa7-FN and FN-AvrXa7 to bind DNA targets *in vitro* were investigated using double-stranded oligonucleotides containing the AvrXa7 EBE of *Os11N3* or its mutated version, Os11N3M, containing five nucleotide substitutions near the 5' end of its EBE sequence (Figure 2.2A). The binding reactions were carried out in the solution lacking divalent metal cations such as Mg^{2+} to prevent cleavage of the oligonucleotides (see Materials and Methods). The electromobility shift (EMS) assay demonstrated that AvrXa7-FN and FN-AvrXa7 preferentially bound to the labeled double stranded DNA containing the authentic AvrXa7 EBE target sequence but did not effectively bind to the probe containing the mutated target sequence containing five nucleotide changes near the 5' end of the EBE (Figure 2.2B, left panels). Furthermore, AvrXa7-FN and FN-AvrXa7 binding to the ^{32}P-labeled AvrXa7 EBE could be competed away using increasing amounts of unlabeled DNA nucleotides of the same sequence (Figure 2.2B, middle panels). However, the binding to the labeled Os11N3 was not competed away with an excess of the variant oligonucleotide, Os11N3M (Figure 2.2B, right

panels).

DNA cleavage activities of FN-AvrXa7, AvrXa7-FN and PthXo1-FN in vitro

We also tested the abilities of FN-AvrXa7, AvrXa7-FN and PthXo1-FN to cleave substrate DNA individually or in combination *in vitro*. For experiments with individual FN-AvrXa7 and AvrXa7-FN TALNs, we chose a plasmid containing a cloned DNA fragment of the *Os11N3* promoter from rice. The plasmid pTOP/Os11N3 was first linearized at a unique restriction site (EcoNI) or digested with MluI (Figure 2.3A; 2.3B, lanes 2 and 3). The DNA was then incubated with FN-AvrXa7 or AvrXa7-FN at 37 ºC for 1 hr using the EMS assay buffer with the addition of 2.5 mM $MgCl_2$. The FN-AvrXa7 and AvrXa7-FN each cleaved the EcoNI-linearized pTOP/Os11N3 plasmid into two fragments of the expected size (0.8 kb and 2.1 kb) (Figure 2.3B, lanes 5 and 8). Incubation of the supercoiled pTOP/11N3 (Figure 2.3B, lane 1) with an appropriate amount of FN-AvrXa7 or AvrXa7-FN resulted in partial linearization of the plasmid (Figure 2.3B, lanes 4 and 7, respectively). Individual FN-AvrXa7 and AvrXa7-FN also cleaved the MluI predigested pTOP/11N3 into an expected pattern of DNA fragments (Figure 2.3B, lanes 6 and 9, respectively). Taken together, the results indicate that both FN-AvrXa7 and AvrXa7-FN mediated-cleavages occur at the AvrXa7 EBE site.

To investigate the specificity of cleaving action by FN-AvrXa7 in more detail, a plasmid containing a mutated binding site of AvrXa7 was generated from pTOP/Os11N3 using a linker scanning mutagenesis method to contain a six base pair insertion immediately downstream of the initial "T" residue of the *Os11N3* EBE. This EBE mutation abolished the ability of AvrXa7 to elicit GFP gene expression when placed in the *Os11N3* promoter in our *Os11N3* promoter assay (data not shown), which is consistent with the finding that the "T" immediately upstream of the 1st EBE nucleotide is essential for the TAL effector mediated activation of target genes (29).

Incubation of the same amount (1 ug) of EcoNI pretreated DNA of pTOP/Os11N3, the mutant pTOP/Os11N3m2 and pTOP/GFP (a plasmid containing a GFP gene sequence unrelated to the AvrXa7 EBE) with an appropriate amount of FN-AvrXa7 resulted in digestion of pTOP/Os11N3, but not the mutant- and GFP-containing plasmids. The specific and expected cleavage pattern observed with digestion of pTOP/Os11N3 DNA was not observed with the mutant pTOP/Os11N3m2 even with increasing amounts of FN-AvrXa7 protein. These experiments demonstrate that the FN-AvrXa7 is a highly selective endonuclease, having the ability to cleave double stranded DNA while discerning the preferred target site from a slight variant or unrelated DNA sequence.

Additional evidence for specific cleavage by TALNs in close proximity to EBE sites has come from cleavage patterns of DNA sequences that are targets for PthXo1-FN as well as AvrXa7-FN and PthXo1-FN in combination. Cleavages of the EBE site (Figure 2.3C) of plasmid pEBE-TD (Figure 2.3D) with either of the two TALNs (or EcoRI) cause linearization (or partial linearization) of the plasmid to produce a band of 3052 bp (Figure 2.3E, lanes 2, 4 and 6). To define the site of TALN-mediated DNA cleavage, post treatment of TALN-digested plasmid DNA with BglI was employed. BglI alone produces DNA fragments of 1786 bp and 1266 bp (Figure 2.3E, lane 3). Incubation with AvrXa7-NF alone results in incomplete digestion of the supercoiled plasmid and produces a band of DNA migrating at approximately 3 kb along with another band migrating with an apparent molecular size of >10 kb (lane 4). When BglI is added to the digestion mix 30 minutes after cutting with AvrXa7-FN has begun, four DNA fragments are observed that correspond to sizes interpreted to be 1786 bp, 1479 bp, 1266 bp and 307 bp (Figure 2.3E, lane 5). This pattern is fully compatible with the interpretation that the initial cutting by AvrXa7-FN was at the EBE target site, but was only partially complete. Post

incubation with BglI resulted in two BglI-BglI DNA fragments of 1786 bp and 1266 bp and two AvrXa7-FN-BglI bands of 1479 bp and 307 bp. When pEBE-TD is incubated with PthXo1-FN three DNA fragments are observed (Figure 2.3E, lane 6), a weak band co-migrating with uncut supercoiled plasmid, a band co-migrating with linearized DNA and a band >10 kb. When BglI is subsequently added to the reaction, again four bands of DNA are observed (Figure 2.3E, lane 7). Double digestion of pEBE-TD with AvrXa7-FN and PthXo1-FN using only one-fifth the amount of each TALN used in the reactions whose products are displayed in lanes 4 through 7, results once more in the appearance of four bands, but with somewhat greater representation of bands of molecular sizes of 1479 bp, 1266 bp and 307 bp – and significantly less non-specific DNA cleavage (compare lanes 5 and 7 with lane 8). The possible nature of the DNA band migrating with an apparent molecular size >10 kb is considered in the Discussion section below.

To further identify the major cleavage sites of the sense and antisense strand, the cleaved DNA fragments (expected sizes of ~840 bp and ~2120 bp) derived from pTOP/Os11N3 treated independently with AvrXa7-FN and FN-AvrXa7 were purified and subjected to sequencing. Each band was expected to contain part of the Os11N3 promoter fragment, but it was unclear which contained the actual target site. Primers that flank the original 0.4 kb *Os11N3* promoter fragment were chosen to sequence each of the digested plasmid bands. These primers presumably were able to sequence through the entire EBE site if it was present. For AvrXa7-FN cleavage, the right side primer (M13R on pTOPO) was used to sequence the sense strand of the 0.8 kb fragment, while M13F was used to sequence the antisense strand of the 2.1 kb fragment (Figure 2.4A). The reverse complementary DNA sequencing trace (Figure 2.4A, the upper chromatograph) from the sense strand of the 0.8 kb fragment ends two base pairs downstream of the last EBE nucleotide. On the other hand, the sequence from the antisense strand of 2.1 kb

fragment ends 17 base pairs downstream of last EBE nucleotide (Figure 2.4A, lower chromatograph), resulting in a 15 base pair cutting "zone" of each DNA fragment. These results indicate that the major cleavages of double stranded DNA by the action of AvrXa7-FN occur downstream of the EBE site, a finding that is consistent with the expected locations of the FokI nuclease domain and of the AvrXa7 binding domain given the configuration of the two components. For FN-AvrXa7 cleavage (Figure 2.4B), the sequence trace from the antisense strand of the 1.2 kb fragment ends 14 base pairs upstream of the first nucleotide of the EBE (Figure 2.4B, lower chromatograph), while the sequencing trace (reverse complementary to the original trace) generated from the sense strand of the 0.8 kb fragment terminated completely after the 6th nucleotide upstream of the last nucleotide of the EBE sequence (Figure 2.4B, upper chromatograph), but within the EBE site. These latter cutting sites were unexpected given the position of the FN domain relative to the FN-AvrXa7 DNA binding site and the presumable protection of the site by the AvrXa7 binding. Repeated sequencing of DNA fragments from additional experiments yielded similar results and, thus, provide no explanation for the observed, but unexpected, DNA cleavage pattern with FN-AvrXa7.

Stimulated homologous recombination in yeast by FN-AvrXa7, AvrXa7-FN, and AvrXa7-FN/PthXo1-FN

We sought to test the ability of the TALNs to bind and cleave target sequences *in vivo* by using a previously established yeast single strand annealing (SSA) assay (34, 36). In this assay, a "reporter" construct is coexpressed with an "effector" construct in yeast cells. The reporter construct contains a divided *LacZ* gene in which a duplicated 125 bp portion of the *LacZ* coding region has been created. The direct repeat within the *LacZ* gene is separated by a 1.2 kb sequence containing the *URA3* gene (Figure 2.5A) or a shorter 0.2 kb sequence (Figure 2.5B) and a

multiple cloning site (MCS). It is expected that the direct DNA repeats will undergo homologous recombination at high efficiency when a DSB is created between the repeats, resulting in a reconstituted and functional *LacZ* gene. Measurement of β-galactosidase (*LacZ* product*)* enzymatic activity was used to quantify the recombination frequency that, in turn, reflects the activity of TALNs in the presence of various target sequences (37, 38, 39).

For TALN assays, the *Saccharomyces cerevisiae* YHP500 strain carrying the reporter plasmid with only a single EBE of AvrXa7 EBE (x7-S) was mated with YHP499 harboring pCP3M-FN-AvrXa7/pCP4M or pCP3M-AvrXa7-FN/pCP4M (Figure 2.5). Cells carrying the reporter and effector plasmids in either combination did not show increased β-galactosidase activity compared with control cells transformed with two effector plasmids lacking any TALN (data not shown). The results suggested that a single EBE site was insufficient for TALNs FN-AvrXa7 or AvrXa7-FN to effectively cleave double stranded DNA at the target site or at any nearby site within the reporter plasmid *in vivo*. The results also suggested a possible requirement for binding of two TALNs in an orientation allowing efficient dimerization of their FokI nuclease domains – in a fashion similar to that needed for successful DNA cleavage by ZFNs. We, therefore, reasoned that two TAL EBE sites in a proper orientation and with an appropriate spacing could bring the FokI nuclease domains in sufficiently close vicinity to dimerize and consequentially execute a double strand cleavage. To test the hypothesis, we designed two constructs, one with two identical AvrXa7 EBE sites located in a head-to-head orientation (HD) for FN-AvrXa7 and the other in a tail-to-tail orientation (TD) for AvrXa7-FN. Another construct was made containing two EBE sites, one for AvrXa7 and one for PthXo1, appropriately situated to allow tail-to-tail (TD; FN to FN) interaction between EBE-bound AvrXa7-FN and PthXo1-FN. We also sought to determine the optimal range of spacer lengths between the two EBE sites

in each configuration to allow for efficient TALN cleavage (Figures 5A and 5B). For FN-AvrXa7, the head-to-head oriented dual EBE sites (HD) were separated by a series of spacers of 6, 14, 19, 24, 30, 35 and 40 bp in length (Figure 2.5A, DNA). Each reporter plasmid was coexpressed with pCP3M-FN-AvrXa7. Yeast cells containing constructs with spacers of 30 nucleotides or more exhibited -galactosidase activities that increased with the length of the spacer element employed (Figure 2.5C).

To compare the cleavage efficiency of the TALNs containing a C-terminal FN (using a tail-to-tail orientation of dual target sites) with efficiency obtained with a known ZFN under the similar context, we performed an assay with a ZFN consisting of the "original" BCR-ABL three-finger array and the FokI nuclease domain (35). For this experiment, YHP499 cells expressing BCR-ABL-FN were mated with YHP500 cells containing the reporter plasmid pzf-TD6 that contained a dual tail-to-tail target site for the BCR-ABL three-finger domain. As expected, the diploid cells exhibited high β-galactosidase enzymatic activity compared to cells lacking the effector plasmid (Figure 2.5D). For experiments with AvrXa7-FN, the tail-to-tail dual EBE sites were separated by spacers of 2, 5, 8 and 19 bp (Figure 2.5A). Among this collection of reporter plasmids, only the plasmid containing the 19 bp spacer produced increased -galactosidase in yeast cells when coexpressed with AvrXa7-FN. The magnitude of response was directly comparable to that obtained with the zinc finger nuclease (Figure 2.5D). Only background levels of -galactosidase expression were obtained when FN-AvrXa7 was substituted for AvrXa7-FN (data not shown). Furthermore, no stimulation of -galactosidase activity was observed in cells expressing AvrXa7-FN if either one or both AvrXa7 EBEs was mutated (Figure 2.5D). We further tested the ability of two different species of TALNs, AvrXa7-FN and PthXo1-FN, to act in concert to recognize and cleave an asymmetric target sequence separated by a serial spacer

(x7/o1-TDn3, n3=6, 11, 16, 21, 26, 31, 36 bp) (Figure 2.5B). The yeast cells containing reporter plasmids each with spacer in length of 16, 21, 26 or 31 exhibited significant increased -galactosidase activity in the presence of AvrXa7-FN and PthXo1-FN together (Figure 2.5D). Taken together, these results suggest requirements for dual EBE target sites, optimized spacer lengths between the EBEs as well as dimerization of TALN FN domains for efficient cleavage of double stranded DNA target sites by TALNs in living yeast cells.

Discussion

Many years of effort spent in elucidating the interaction between TAL effectors and their modulated host genes has led to a recent breakthrough in deciphering the DNA recognition code of TAL effectors (28, 29). The predictability and potential manipulability of the TAL central repeat domain for DNA binding specificities make TAL an excellent system for exploiting potential biotechnological applications. In the present study, we created chimeric TALNs, FN-AvrXa7, AvrXa7-FN and PthXo1-FN containing the entire AvrXa7 or PthXo1 TAL effectors and the nuclease domain of the FokI restriction enzyme either at the C-terminal end or the N-terminal end of each TAL effector. All three constructs were tested for the ability to bind to the respective EBE recognition site and to cleave adjacent DNA. Binding of FN-AvrXa7 to the AvrXa7 EBE *in vivo* was demonstrated by its ability to activate transcription of a GFP coding sequence driven by the rice *Os11N3* promoter that contains the AvrXa7 EBE binding site (Figure 2.1G). All three TALNs were successfully overproduced and purified from *E. coli* cells (AvrXa7 TALNs). FN-AvrXa7 and AvrXa7-FN each were shown to bind specifically to double-stranded oligonucleotides containing the AvrXa7 EBE target site, but not to a slightly modified version of the binding site in an EMS assay (Figure 2.2). Moreover, the purified AvrXa7-FN

and FN-AvrXa7 TALNs exhibited cleavage activity near the expected EBE binding site under optimized reaction conditions in an *in vitro* assay, the results of which were confirmed by DNA sequencing (Figures 2.3 and 2.4). Likewise, PthXo1-FN (alone and together with AvrXa7-FN) was shown to specifically cleave at its specific EBE DNA target site and produce the predicted sized DNA fragments (Figure 2.3). Finally, expression of the chimeric FN-AvrXa7, AvrXa7-FN and PthXo1-FN TALNs in yeast stimulated HR between internal repeats of a disrupted and non-functional reporter gene (*LacZ*) that contained appropriately paired AvrXa7 or asymmetric AvrXa7/PthXo1 target sites (Figure 2.5). These observations demonstrate the successful creation of functional TALNs and lead the way to future experimentation directed toward development of a technology for high-specificity gene knockout and homologous recombination in organisms that currently lack the ability to support either process in a practical manner.

FokI and its fusion proteins with zinc finger DNA binding domains have been extensively studied. The endonuclease domain (FokI nuclease, FN) by itself has no specificity for cleavage, but cuts DNA at a set distance from the binding site specified by the FokI DNA binding domain when the two domains are linked together (13, 14, 15). In this sense, several types of FN based fusion proteins have been successfully created that combine new DNA sequence binding specificities with the FN cleavage activities, with zinc finger nucleases (ZFNs) being the most familiar (6, 7, 41). Study has shown that fusion of FN to ZF motif does not change the DNA binding specificity of the ZF protein although it may cause slight decrease in binding affinity (41). We chose the FokI cleavage domain to fuse with members of the TAL effector family and, as a proof of principle, demonstrated the feasibility and generality of creating a new class of rare-cutting, site-specific DNA nucleases with sequence specificities attributable to the TAL effectors. The DNA binding features of TAL effectors make this group

of proteins or their repetitive domains desirable as the key component of such chimeric endonucleases for a number of applications, including various sorts of genome editing. For example, the majority of naturally occurring TAL effector proteins contains a large number of repeat units and, correspondingly, recognizes lengthy DNA target sites (32, 43). These TAL EBE sites are comparable in length to, or longer than, target sites of rare-cutting meganucleases or homing nucleases (i.e., 14 to 40 bp) as well as binding sites for artificial zinc finger proteins assembled from multiple single fingers (i.e., 18 or 24 bp) (5, 42). All TAL effector proteins investigated thus far exhibit high sequence specificity to the EBEs of their target genes (32, 43). The known code of TAL effectors predicts an alignment of a single type of repeat unit to a single nucleotide species (A, G, C or T) based on the specific di-residues at positions 12 and 13 in the repeat unit. This modular nature of the TAL repeat domain for DNA binding specificity suggests that techniques can be developed to produce an array of repeat units that can precisely recognize a unique, lengthy sequence of nucleotides in any given gene. If so, investigators will be able to create truly gene-specific TALNs for use in organisms with large genomes and lacking robust systems for homologous recombination.

Thus far, several TAL effectors have been found to function as transcription activators. Like many other transcription factors, TAL effectors may function as dimers to bind target DNA. However, to date, AvrBs3 is the only TAL effector shown to dimerize *in vitro* and in the cytoplasm *in vivo* before entry into nuclei of host cells (44). The sequence specificity of known TAL effectors that bind DNA can be aligned to only one strand of the target site which is usually asymmetric (28, 29). Thus, it is not yet clear if most TAL effector proteins form dimers or multimers in the presence of target DNA (or in the absence of DNA). The results from the present yeast SSA assay imply that TALNs do not form homo- or hetero-dimer at a single TAL

EBE site, or at least the dimerization of TAL subdomain does not facilitate the dimerization of FokI nuclease domains for effective double stranded DNA cleavage in yeast cells. More detailed structural studies of TAL effectors or TALNs likely will be needed to resolve this uncertainty, which may or may not negatively influence the ability to easily and successfully design sequence specific TAL effector nucleases in the future.

It has been established that for efficient double strand cleavage of target DNA dimerization of FokI monomer nuclease domains is required (11). Therefore, it is conceivable that TALNs need to dimerize for the efficient cleavage of DNA in solution where sufficient concentrations of purified proteins and substrates are present or *in vivo* where TALN and target substrate are otherwise limited. This could be achieved through various mechanisms, three of which are presented below. In one model, one EBE-bound TALN might form a dimer with another bound or unbound TALN through an as yet uncharacterized dimerization motif of TAL effector. In such a case, the TAL subdomain-mediated dimerization could bring the two FokI nuclease domains in close proximity near the binding site and allow DNA cleavage. Alternatively, one TALN monomer could bind to one EBE target site and, similar to the model proposed for the native FokI or hybrid ZFNs *in vitro* (13, 14), dimerization of two DNA bound-TALNs through the well characterized dimerization motif in the FokI nuclease subdomain could occur in close proximity and, thereby, support DNA cleavage *in trans* if sufficient concentrations of nucleases were present. Successful *in vitro* cleavage of DNA carrying a single AvrXa7 EBE by the FN-AvrXa7, AvrXa7-FN and PthXo1-FN TALNs (Figure 2.3) is consistent with this model. In a third model in which two tandem, head-to-head or tail-to-tail EBE sites are present, TALNs could bind to each of the EBE sites. This would bring the two FN domains of the two TALNs into sufficiently close proximity to allow dimerization and DNA cleavage. Our yeast

SSA data (Figure 2.5) is consistent with this latter model.

The function of native FokI is allosterically regulated through DNA and divalent metal binding. Without DNA binding and in the absence of divalent metal, FN is sequestered through tight interaction with the DNA recognition motifs of FokI and, thus, the FokI monomer maintains an idle state. Following binding of two FokI holoenzymes to the FokI recognition site and in the presence of metals, the two FokI nuclease domains are freed and can dimerize. This dimerization then allows double stranded DNA cleavage (14, 16, 17). It is possible that the interaction between the FN subddomain and the TAL DNA binding subdomain in hybrid TALN lacks such tight regulatory mechanism and, hence nuclease domains form dimers with less difficulty. That may explain why the apparent stringency of cleavage by the presently studied chimeric TALNs (and also ZFNs) is lower and leads to the non-specific cleavage observed in the presence of excess of TALNs. The structure (length and composition) of linker segment between the DNA binding and cleaving domains of FokI and its derived nucleases (i.e., ZFNs and, likely, TALNs) also dictate the enzymes' cleavage pattern, for example, the distance of cleavage sites from the DNA binding site (45, 46, 47). The linker segment of native FokI is 15 amino acids (residues 373-387) long and allows the FN to extend and cleave the sense strand 9 bp and antisense strand 13 bp downstream of the binding site (12). An 18 amino acid flexible linker of a ZFN accommodates effective cleavage of target spacer in a range of 6 to 18 bp with 8 bp as an optimum in *Xenopus* oocytes as determined in a single-strand annealing reporter assay (45). The latter study also reported that the dependence of efficient cleavage on spacer length *in vitro* differed from that under *in vivo* condition. Because the minimum-sized TAL effector fragment required for efficient DNA binding is unknown, the full-length AvrXa7 and PthXo1 were used to construct the TALNs in our study. Therefore, the N-terminal 288 amino acids in FN-AvrXa7 and

the C-terminal 295 amino acids in AvrXa7-FN and PthXo1-FN function as long inter-domain linkers between FN and the repeat DNA binding domain in TALNs. Such an extended inter-domain linker may allow significant "reach" for the nuclease domain to cut at a moderate distance away from the ends of the EBE or allow greater flexibility for the nuclease domain to cut within a moderately wide zone as we observed *in vitro* with our TALN enzyme assays and *in vivo* with our yeast assays. Future investigations will be required to determine how various combinations of inter-domain sequence and length affects the cleavage efficiency of TALNs [e.g., tests of TALNs consisting of the FN connected directly to the TAL effector repeat domain or through linkers of various lengths and composition].

Other important questions are presently under active investigation. One such question is the nature of the DNA fragment with an apparent molecular size >10 kb when supercoiled plasmid DNA is mixed with a TALN (lanes 4 and 7 in Figure 2.3B, lanes 4 and 6 in 3E). Such a band of DNA is not observed if the supercoiled plasmid is cleaved before mixing with the TALN (Figure 2.3B and 2.3E). Our working hypothesis is that the uppermost, slow migrating DNA band may be a complex between plasmid DNA and the TALN protein that exists prior to DNA cleavage (by the TALN or a restriction enzyme), but not after cleavage. Future analyses of the component(s) of this DNA band (i.e., ethidium bromide stained band) should resolve this present enigma. Another issue is TALN stability. We have observed that all of the TALNs created for the present study have quite short half-lives (i.e., ~30 to 45 minutes). This currently imposes serious practical constraints on the degree of purity of recombinant TALNs that can be obtained and on the duration and extensiveness of biochemical analyses, including accurate measurements of TALN binding affinities to DNA. Further attempts to discover conditions that stabilize TALNs overproduced in *E. coli* will be important undertakings along with endeavors to

determine if such instability does or does not exist *in vivo* in various cell types.

The work presented here unambiguously demonstrates several points: the ability to fuse TAL effectors with other proteins to create functional chimeric proteins; the specificity of DNA binding by TAL effectors when fused with a nonspecific nuclease domain; and the specific cleavage of DNA target sites by engineered TALNs both *in vitro* and *in vivo*. The newly emerging TALN-based approach could be an attractive alternative to the still improving ZFN-based or meganuclease-based (48) genomic tools for a wide variety of possible applications including targeted genome editing. However, a few basic questions remain unanswered regarding the feasibility of using TALNs for genome modification. First, can novel TALN DNA binding domains with the requisite specificity and affinity be synthesized based on the actual DNA target sequences? Although arbitrarily assembled, TAL effectors were able to activate promoters containing sequence elements synthesized based on the "code" (29), this capability is not yet demonstrated for a TALN fusion protein. Second, can two different TALNs work coordinately at preselected adjacent target sites? The reaction involving cleavage of a dual asymmetric PthXo1 EBE/AvrXa7 EBE site with a mixture of PthXo1-FN and AvrXa7-FN (2.3E, lane 7) hints this may be possible. That is, the two TALNs together produced somewhat better, albeit still incomplete, plasmid cleavage than when each TALN was used independently at 5 times the concentration employed in the dual cutting reaction – and with less nonspecific plasmid DNA cleavage (compare Figure 2.3E, lanes 5 and 7 with lane 8). Third, will the DNA recognition and cleavage by the TALNs occur in a chromosomal context in living cells? Data from the yeast experiment described in this paper (Figure 2.5) provide strong initial evidence that a TALN or sets of TALNs can successfully find, bind and cleave an EBE target site within yeast chromatin. Addressing these issues will assist in achieving the goal of generating a tool box of

TALNs for targeted genome editing based on the DNA binding specificity of custom-designed, synthetic TAL repeat domains and the DNA cleavage function of FokI or other nucleases.

Materials And Methods

Chimeric gene construction

Chimeric genes encoding fusions of TAL effector AvrXa7 or PthXo1 with the FokI nuclease domain (FN) at either the N-terminus or the C-terminus (i.e., FN-AvrXa7, AvrXa7-FN or PthXo1-FN) were constructed using standard *E. coli* strains and DNA techniques (33). The full-length AvrXa7 was first modified with PCR primers Tal-F and Tal-R to integrate the restriction sites KpnI and BglII upstream of the start codon at the 5' end and HindIII, XbaI and a stop codon containing SpeI at the 3' end based on the plasmid pZWavrXa7 (30). AvrXa7 without its repetitive central region was PCR amplified using primers Tal-F and Tal-R and cloned into pBluescript KS using KpnI and SpeI. Then the central repeat coding region was cloned back into the plasmid lacking the central repeat element using SphI, resulting in the plasmid pSK/AvrXa7. The SphI fragment for central repeat domain of AvrXa7 in pSK/AvrXa7 was replaced with that of PthXo1 from pZWpthXo1 (24), resulting in the plasmid pSK/PthXo1. The DNA fragment encoding the DNA cleavage domain (amino acids 388-583) of FokI (NCBI accession number J04623) was PCR amplified using the primers Fokn-F1 and Fokn-R1 and a plasmid containing the FokI gene as template. Fokn-F1 contained the restriction sites KpnI and BglII, while Fokn-R1 contained a BamHI restriction site. The product was cloned into the A/T cloning vector pGEM-T (Promega, Madison, Wisconsin). The KpnI and BamHI digested DNA fragment encoding FN was cloned into KpnI and BglII treated pSK/avrXa7 resulting in pSK/FN-AvrXa7 which contained the chimeric gene with the FN coding region at its 5' and AvrXa7 at its 3' end.

Similarly, FN coding sequence for C-terminal fusion of AvrXa7 and PthXo1 was PCR amplified with primers Fokn-F2 and Fokn-R2, which contain restriction sites for HindIII and SpeI, respectively. The HindIII and SpeI digested FN fragment was cloned into pSK/AvrXa7 and pSK/PthXo1 individually at their 3' ends. The accuracy of all PCR products was confirmed by sequencing. Primer sequences are provided in the Supplementary Data Table S1.

Transient expression assay for DNA binding activity of FN-AvrXa7

A construct containing green fluorescence protein (GFP) reporter gene cloned downstream of *Os11N3* promoter, which contains the AvrXa7 EBE sequence was made as follows. The coding region for the GFP gene in plasmid pEGFP (Clontech Laboratories, Mountain View, CA) was PCR amplified using primers GFP-F and GFP-R and cloned into pGEM-T for sequence confirmation. The GFP coding region with added restriction sites was cloned downstream of the promoter region containing the AvrXa7 EBE and upstream of the *Os11N3* terminator, resulting in pEBE-GFP. The GFP expression cassette was then cloned into pCAMBIA1300 (CAMBIA) at KpnI and HindIII restriction sites. The resulting construct was transformed into *Agrobacterium tumefaciens* strain EHA105 to create a "reporter" *Agrobacterium* strain. DNA encoding FN-AvrXa7 was cloned downstream of the cauliflower mosaic virus (CaMV) 35S promoter in a modified pCAMBIA1300 vector and mobilized into EHA105 to create an "effector" *Agrobacterium* strain. The effector strain containing AvrXa7 (lacking the FN domain) was similarly constructed to serve as a positive control. The reporter and the effector strains were co-infiltrated into *Nicotiana benthamiana* leaves. The inoculated leaves were checked for expression of GFP using a Leica M205 FA fluorescent stereomicroscope.

Expression and purification of FN-AvrXa7, AvrXa7-FN and PthXo1-FN

The chimeric gene FN-AvrXa7 was cloned into pPROEX HTb (Invitrogen, Carlsbad, CA) by ligating the BglII and SpeI digested FN-AvrXa7 fragment from pSK/FN-AvrXa7 into the BamHI and SpeI digested expression vector. Similarly, the BglII and SpeI digested AvrXa7-FN DNA fragment was cloned into pPROEX HTb for AvrXa7-FN overexpression. For PthXo1-FN, the expression vector pET28a (Novagen, Madison, WI) was used. The expression constructs were transformed into *E. coli* strain BL21 (DE3) for overexpression of the recombinant proteins with induction by isopropyl-1-thio-β-D-galactopyranoside (IPTG) following the manufacturer's instructions (Invitrogen, Carlsbad, CA). The 6Xhistidine tagged FN-AvrXa7, AvrXa7-FN and PthXo1-FN were purified with Ni-NTA agarose (Qiagen, Valencia, CA) and the protein concentrations were determined using the BioRad Bradford protein quantification kit (BioRad, Hercules, CA).

Electromobility shift (EMS) assay

The complementary oligonucleotides of Os11N3-F & Os11N3-R containing AvrXa7 EBE and Os11N3M-F & Os11N3M-R containing a mutated AvrXa7 EBE were annealed, respectively, and 5'-end labeled with [γ-^{32}P]ATP catalyzed by T4 polynucleotide kinase. The labeled oligonucleotide duplex DNA was mixed individually with AvrXa7-FN and FN-AvrXa7 in a 10 ul reaction volume containing Tris-HCl (15 mM, pH 7.5), KCl (40 mM), DTT (1 mM), glycerol (2.0%), poly(dI.dC) (50 ng/ul), EDTA (0.2 mM), ^{32}P-labeled DNA (final concentration, 5 fmol), and hybrid protein (final concentration, about 15 fmol). Unlabeled oligonucleotides were used as competitor probes, and were added in increasing amounts (final concentration, 0-250 fmol) to successive reaction. The binding reactions were kept at room temperature for 30 minutes before loading onto a 6% TBE polyacrylamide gel. After electrophoresis, the gel was

exposed to X-ray film for radioactive image capture.

In vitro DNA cleavage

A 406 bp genomic region of the rice *Os11N3* gene encompassing the AvrXa7 EBE was PCR amplified with the forward primer Os11N3P-F, and the reverse primer Os11N3P-R, then cloned into the cloning vector pTOPO (Invitrogen, Carlsbad, CA), resulting in pTOP/11N3. The plasmid was also used to generate pTOP/Os11N3m2 using linker scanning mutagenesis method with an insertion of 6 nucleotides (5'-cccggg-3') within the AvrXa7 EBE site. A DNA fragment containing the dual EBEs for PthXo1 and AvrXa7 in a tail-to-tail orientation was cloned into pPCR Script Amp (Stratagene, Santa Clara, CA), resulting in the plasmid pEBE-TD. For pTOP/Os11N3 and pTOP/Os11N3m2 digestions, the sequenced clone was digested to completion with EcoNI or MluI and purified, One microgram of uncut plasmid or the digested DNA was incubated individually with FN-AvrXa7 (225 ng or otherwise indicated in the text) and AvrXa7-FN (200 ng) in a volume of 15 uL. The buffer condition was the same as used for the EMS assay described above, but in the presence of 2.5 mM $MgCl_2$.

In vitro digestions of pEBE-TD with TALNs AvrXa7-FN and PthXo1-FN were performed differently. Assays for TALN activities with or without post digestion with restriction enzyme BglI were conducted *in vitro* using 1 g of pEBE-TD supercoiled plasmid DNA in a 30 L reaction volume containing Tris-HCl (20 mM), pH 8.5, NaCl (150 mM), $MgCl_2$ (2 mM), glycerol (5%), BSA (0.5 mg/ml), DTT (1 mM) and 5 g of Ni^+-affinity column-purified PthXo1-FN and/or AvrXa7-FN incubated for 30 min at 37. For reactions in which BglI was used after TALN digestion, 10 units of BglI were added and digestion continued for an additional 20 min. For reactions in which both PthXo1-FN and AvrXa7-FN were used simultaneously, the reaction conditions were identical except that only 1 g of each TALN was employed.

Yeast homologous recombination assay

The yeast strains YPH499 (*MAT*a *ura3-52 lys2-801_amber ade2-101_ochre trp1-*Δ*63 his3-*Δ*200 leu2-*Δ*1)* and YPH500 (*MAT*α *ura3-52 lys2-801_amber ade2-101_ochre trp1-*Δ*63 his3-*Δ*200 leu2-*Δ*1*) as well as expression vectors (pCP3, pCP4 and pCP5) were described and kindly provided by Dr. Dan Voytas (34). The pCP5 derived reporter construct containing a single AvrXa7 EBE was made by inserting the annealed oligonucleotides (EBES-F and EBES-R) into the BglII- and SpeI-digested pCP5. A set of duplexes of oligonucleotides (x7-HDn1-F and x7-HDn1-R, n1=6, 10, 14, 19, 24, 30, 35, 40 bp) containing the dual AvrXa7 EBEs in a head-to-head (corresponding to FN domain to FN domain of FN-AvrXa7) orientation separated by a serial of spacers (n1) were cloned into the BglII and SpeI digested pCP5 plasmid. Similar constructs with two AvrXa7 EBEs in a tail-to-tail orientation and separated by various lengths (2, 5, 8, 19 bp) of spacers were also made using duplexes of oligonucleotides (x7-TDn2-F and x7-TDn2-R, n2=2, 5, 8, 19 bp). For the 19 bp spacer configuration, an additional one reporter construct with a mutation in one AvrXa7 EBE and another with a mutation in each of the AvrXa7 EBEs were made. A third set of dual asymmetric EBEs with one corresponding to AvrXa7 and the other to PthXo1 in a tail-to-tail orientation were also cloned into pCP5 using PstI and SpeI. The duplexes of oligonucleotides are x7/o1-TDn3, n3=6, 11, 16, 21, 26, 31, and 36 bp. The expression vectors pCP3 and pCP4 were first modified with a linker sequence containing multiple cloning sites downstream of the translation elongation factor 1α (TEF1) promoter, resulting in pCP3M and pCP4M, respectively. The linker was made by annealing two oligonucleotides (Linker-F and Linker-R) and cloned into the XbaI and XhoI digested pCP3 and pCP4 individually. The chimeric genes FN-AvrXa7 and AvrXa7-FN were digested with BglII and SpeI and cloned into the BamHI and SpeI digested pCP3M vector, while the chimeric gene

PthXo1-FN was cloned into pCP4M by BglII and SpeI. The reporter plasmids each were transformed into the yeast mating strain YPH500 (MATα) and effector plasmids in a combination of pCP3M-FN-AvrXa7/pCP4M, pCP3M-AvrXa7-FN/pCP4M, and pCP3M-AvrXa7-FN/pCP4M-PthXo1-FN into YPH499 (MATa). Yeast cells from a single yeast colony carrying the reporter plasmid with little β-galactosidase background was mixed with cells from single colony with the effector plasmid (in a combination indicated in text) in triplicate on yeast nutrient medium (YPD) overnight, then cultured in synthetic complete medium lacking leucine, histidine and tryptophan to select for mated cells. The cells were then harvested for quantitative measurement of β-galactosidase activity by using the yeast β-galactosidase assay kit from Thermo Fisher Scientific (Rockford, IL) following the manufacture's manual. Enzyme activity is calculated based on the equation: β-galactosidase activity = $(1000 \times A_{420})/(t \times V \times OD_{660})$, t=time of incubation in minutes, V=volume of cells used in the assay in ml.

As a positive control to our yeast assay for TALN activity, we also constructed an effector plasmid containing a zinc finger nuclease and a reporter plasmid containing the corresponding ZF binding site. A DNA fragment encoding the "original" BCR-ABL three-finger array was obtained by the digestion of pGP-FB-orig BA (35) with XbaI and BamHI and cloned into pCP3, resulting in a translational fusion of BCR-ABL with a C-terminal FokI nuclease domain. The oligonucleotide duplex containing the dual ZFN target sites in a tail-to-tail orientation separated by a 6 bp spacer was cloned into pCP5 by BglII and SpeI, resulting in the reporter plasmid. The effector plasmid pCP4-ZFN and pCP3M were transformed into YPH499. The transformants were mated with YPH500 containing the reporter plasmid pzf-TD6 and measured for β-galactosidase activity as described for TALNs.

Acknowledgements

The authors thank Dr. Dan Voytas for providing the components of the yeast SSA system and Dr. Keith Joung for providing the plasmid pGP-FB-orig BA [obtained through Addgene (Addgene Inc., Cambridge, MA), also designated as Addgene plasmid 13420].

References

1. Le Provost, F., Lillico, S., Passet, B., Young, R., Whitelaw, B. and Vilotte, J.L. (2010) Zinc finger nuclease technology heralds a new era in mammalian transgenesis. *Trends Biotechnol,* **28,** 134-141.

2. Jasin, M. (1996) Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet.,* **12,** 224-228.

3. Vasquez, K.M., Marburger, K., Intody, Z. and Wilson, J.H. (2001) Manipulating the mammalian genome by homologous recombination. *Proc. Natl. Acad. Sci. U. S. A.,* **98,** 8403-8410.

4. Bibikova, M., Beumer, K., Trautman, J.K. and Carroll, D. (2003) Enhancing gene targeting with designed zinc finger nucleases. *Science,* **300,** 764.

5. Porteus, M.H. and Carroll, D. (2005) Gene targeting using zinc finger nucleases. *Nat. Biotechnol.,* **23,** 967-973.

6. Kim, Y.G. and Chandrasegaran, S. (1994) Chimeric restriction endonuclease. *Proc. Natl. Acad. Sci. U. S. A.,* **91,** 883-887.

7. Kim, Y.G., Cha, J. and Chandrasegaran, S. (1996) Hybrid restriction enzymes: Zinc finger fusions to fok I cleavage domain. *Proc. Natl. Acad. Sci. U. S. A.,* **93,** 1156-1160.

8. Sugisaki, H. and Kanazawa, S. (1981) New restriction endonucleases from *flavobacterium okeanokoites* (FokI) and micrococcus luteus (MluI). *Gene,* **16,** 73-78.

9. Szybalski, W., Kim, S.C., Hasan, N. and Podhajska, A.J. (1991) Class-IIS restriction enzymes--a review. *Gene,* **100,** 13-26.

10. Li, L., Wu, L.P. and Chandrasegaran, S. (1992) Functional domains in FokI restriction endonuclease. *Proc. Natl. Acad. Sci. U. S. A.,* **89,** 4275-4279.

11. Bitinaite, J., Wah, D.A., Aggarwal, A.K. and Schildkraut, I. (1998) FokI dimerization is required for DNA cleavage. *Proc. Natl. Acad. Sci. U. S. A.,* **95,** 10570-10575.

12. Wah, D.A., Hirsch, J.A., Dorner, L.F., Schildkraut, I. and Aggarwal, A.K. (1997) Structure of the multimodular endonuclease FokI bound to DNA. *Nature,* **388,** 97-100.

13. Wah, D.A., Bitinaite, J., Schildkraut, I. and Aggarwal, A.K. (1998) Structure of FokI has implications for DNA cleavage. *Proc. Natl. Acad. Sci. U. S. A.,* **95,** 10564-10569.

14. Vanamee, E.S., Santagata, S. and Aggarwal, A.K. (2001) FokI requires two specific DNA sites for cleavage. *J. Mol. Biol.,* **309,** 69-78.

15. Isalan, M.; Choo, Y. (2001) Engineering nucleic acid-binding proteins by phage display. Methods Mol. Biol. **148,** 417-429.

16. Pabo, C.O., Peisach, E. and Grant, R.A. (2001) Design and selection of novel Cys2His2 zinc finger proteins. *Annu. Rev. Biochem.,* **70,** 313-340.

17. Beerli, R.R. and Barbas III, C.F. (2002) Engineering polydactyl zinc-finger transcription factors. *Nat. Biotechnol.,* **20,** 135-141.

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

19. Carroll, D. (2008) Progress and prospects: zinc-finger nucleases as gene therapy agents. Gene Ther., **15**, 1463-1468.

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

21. Kim, J.S., Lee, H.J. and Carroll, D. (2010) Genome editing with modularly assembled zinc-finger nucleases. *Nat. Methods,* **7,** 91; author reply 91-2.

22. White, F.F., Potnis, N., Jones, J.B. and Koebnik, R. (2009) The type III effectors of *Xanthomonas*. *Mol. Plant. Pathol.,* **10,** 749-766.

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

24. Yang, B., Sugio, A. and White, F.F. (2006) *Os8N3* is a host disease-susceptibility gene for bacterial blight of rice. *Proc. Natl. Acad. Sci. U. S. A.,* **103,** 10503-10508.

25. 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 OsTFIIAgamma1 and OsTFX1 during bacterial blight of rice. *Proc. Natl. Acad. Sci. U. S. A.,* **104,** 10720-10725.

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

27. Gurlebeck, 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. *J. Plant Physiol.,* **163,** 233-255.

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

29. 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.

30. Yang, B., Zhu, W., Johnson, L.B. and White, F.F. (2000) The virulence factor AvrXa7 of *Xanthomonas oryzae* pv. *oryzae* is a type III secretion pathway-dependent nuclear-localized double-stranded DNA-binding protein. *Proc. Natl. Acad. Sci. U. S. A.,* **97,** 9807-9812.

31. Hopkins, C.M., White, F.F., Choi, S.H., Guo, A. and Leach, J.E. (1992) Identification of a family of avirulence genes from *Xanthomonas oryzae* pv. *oryzae*. *Mol. Plant Microbe Interact.,* **5,** 451-459.

32. Romer, P., Recht, S., Strauss, T., Elsaesser, J., Schornack, S., Boch, J., Wang, S. and Lahaye, T. (2010) Promoter elements of rice susceptibility genes are bound and activated by specific TAL effectors from the bacterial blight pathogen, *Xanthomonas oryzae* pv. *oryzae*. *New Phytol.,* 10.1111/j.1469-8137.2010.03217.x .

33. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1987) Molecular Cloning: A Laboratory Manual. Cold Pring Harbor Laboratory Press, Cold Spring Harbor, NY, U.S.A.

34. Townsend, J.A., Wright, D.A., Winfrey, R.J., Fu, F., Maeder, M.L., Joung, J.K. and Voytas, D.F. (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature,* **459,** 442-445.

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

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

37. Rudin, N. and Haber, J.E. (1988) Efficient repair of HO-induced chromosomal breaks in *Saccharomyces cerevisiae* by recombination between flanking homologous sequences. *Mol. Cell. Biol.,* **8,** 3918-3928.

38. Sugawara, N. and Haber, J.E. (1992) Characterization of double-strand break-induced recombination: Homology requirements and single-stranded DNA formation. *Mol. Cell. Biol.,* **12,** 563-575.

39. Fishman-Lobell, J., Rudin, N. and Haber, J.E. (1992) Two alternative pathways of double-strand break repair that are kinetically separable and independently modulated. *Mol. Cell. Biol.,* **12,** 1292-1303.

40. Kim, Y.G., Smith, J., Durgesha, M. and Chandrasegaran, S. (1998) Chimeric restriction enzyme: Gal4 fusion to FokI cleavage domain. *Biol. Chem.,* **379,** 489-495.

41. Smith, J., Berg, J.M. and Chandrasegaran, S. (1999) A detailed study of the substrate specificity of a chimeric restriction enzyme. *Nucleic Acids Res.,* **27,** 674-681.

42. Belfort, M. and Roberts, R.J. (1997) Homing endonucleases: Keeping the house in order. *Nucleic Acids Res.,* **25,** 3379-3388.

43. Romer, P., Recht, S. and Lahaye, T. (2009) A single plant resistance gene promoter engineered to recognize multiple TAL effectors from disparate pathogens. *Proc. Natl. Acad. Sci. U. S. A.,* **106,** 20526-20531.

44. Gurlebeck, D., Szurek, B. and Bonas, U. (2005) Dimerization of the bacterial effector protein AvrBs3 in the plant cell cytoplasm prior to nuclear import. *Plant J.,* **42,** 175-187.

45. Bibikova, M., Carroll, D., Segal, D.J., Trautman, J.K., Smith, J., Kim, Y.G. and Chandrasegaran, S. (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol. Cell. Biol.*, **21**, 289-297.

46. Handel, E., Alwin, S. and Cathomen, T. (2009) Expanding or restricting the target site repertoire of zinc-finger nucleases: the inter-domain linker as a major determinant of target site selectivity. *Mol. Ther.*, **17**, 104-111.

47. Shimizu, Y., Bhakta, M. S. and Segal, D. J. (2009) Restricted spacer tolerance of a zinc finger nuclease with a six amino acid linker. *Bioorg. Med. Chem. Lett.*, **19**, 3970-3972.

48. Grizot, S., Epinat, J.C., Thomas, S., Duclert A., Rolland, S., Pâques, F. and Duchateau, P. (2010) Generation of redesigned homing endonucleases comprising DNA-binding domains derived from two different scaffolds. *Nucleic Acids Res.*, **38**, 2006-2018.

Figures And Tables

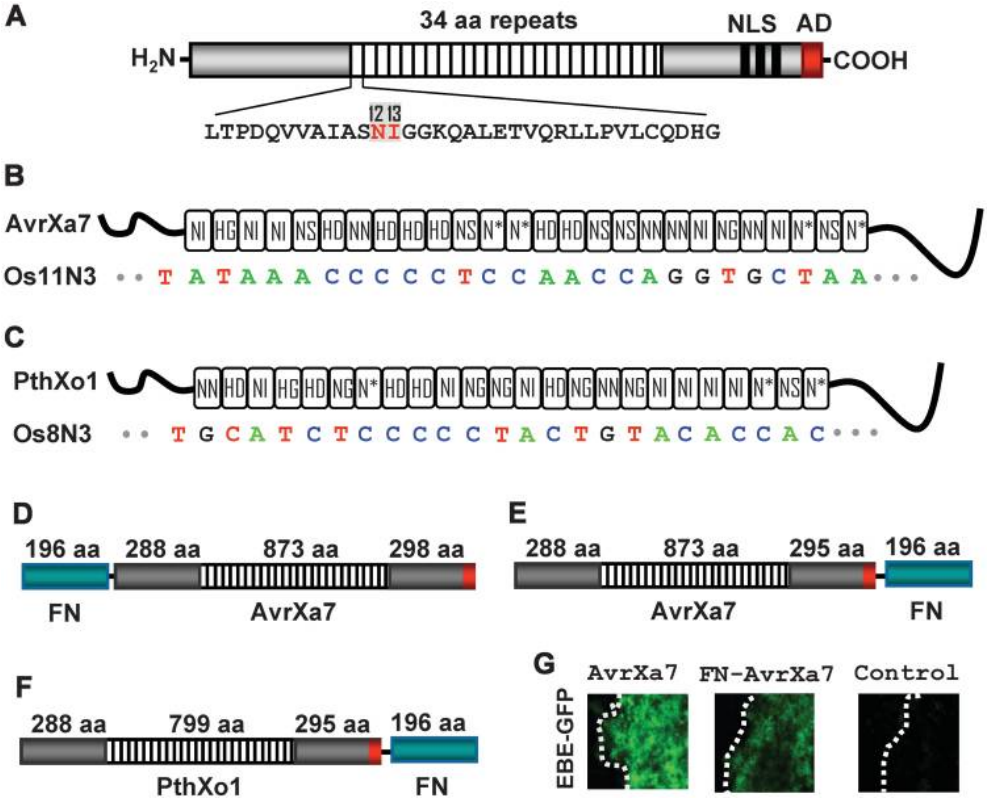


Figure 2.1. Schematics of TAL effectors AvrXa7 and PthXo1 as well as their fusions with FN. **(A)** A typical TAL effector contains a central region of 34 amino acid (aa) direct repeats (open boxes) and three nuclear localization motifs (NLS, black bars) as well as a transcription activation domain (AD, red solid box) at the C-terminus. The representative 34 aa repeat is shown below with the variable amino acid residues at the positions 12 and 13 in red. **(B)** and **(C)** AvrXa7 and PthXo1 contain central 26 and 24 tandem repeats (shown as boxes) of 34 amino acid residues, respectively. The repeats are highly conserved, except for residues at positions 12 and 13 (*, missing aa residue at position 13). The sequence recognition of AvrXa7 to *Os11N3* (and PthXo1 to *Os8N3*) is dictated by the order of the 34 aa repeat units that each contain at positions 12 and 13 a pair of amino acids that exactly recognize one nucleotide at one position in the DNA sequence of the effector binding element (EBE) of the *Os11N3* gene. **(D)**, **(E)** and **(F)** Schematics of TAL effector fusion proteins with FokI nuclease domain (FN) at the N-terminus of AvrXa7 (D) or at the C-termini of AvrXa7 (E) and PthXo1 (F). **(G)** Transient activation by AvrXa7 and FN-AvrXa7 of the *Os11N3* promoter (containing the AvrXa7 EBE) driving a GFP reporter gene (GFP-EBE). The plasmid containing the GFP reporter gene was co-injected along with a plasmid containing a gene encoding the AvrXa7 or AvrXa7-FN driven by the CaMV 35S promoter [or with a plasmid containing the 35S promoter but no AvrXa7 coding sequence (Control)]. The dotted line indicates the edge of the inoculation site.

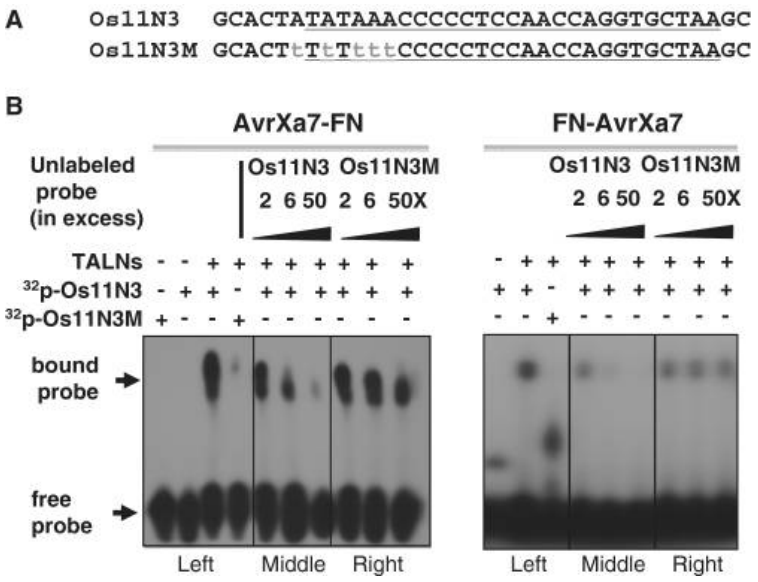
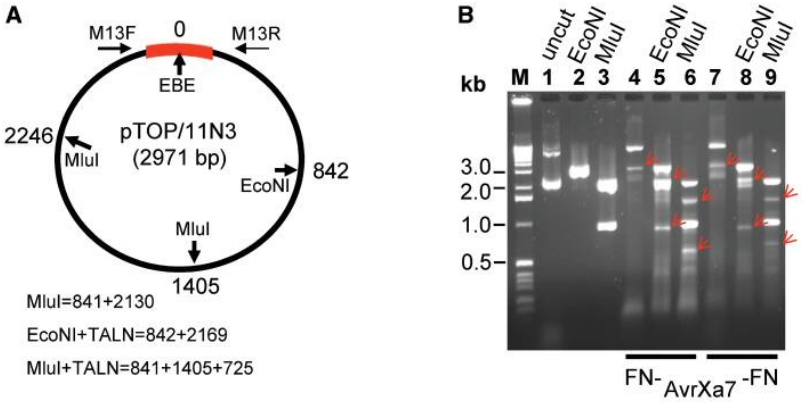


Figure 2.2. Binding specificity of AvrXa7-FN and FN-AvrXa7 fusion proteins to their target DNA. **(A)** Sense-strand sequences of the wild type (Os11N3) and mutant (Os11N3M) oligonucleotide duplexes used in the electrophoresis mobility shift (EMS) assays. **(B)** EMS assays demonstrating specificities of AvrXa7-FN and FN-AvrXa7 binding to the AvrXa7 EBE sequence. The set of gel images to the left depict results with AvrXa7-FN, while set of gel images to the right depict results with FN-AvrXa7. The left panels of each set show binding of AvrXa7-FN and FN-AvrXa7 to the authentic, ^{32}P-labeled, Os11N3 DNA target element but not to the mutated target. Competition assays (middle and right panels of each set) showing that the binding of ^{32}P-labeled Os11N3 EBE target by AvrXa7-FN and FN-AvrXa7 is effectively competed by excess amounts of nonradioactive Os11N3 oligonucleotides (middle panels of each set), but not by the nonradioactive mutated version of Os11N3 DNA (right panels). Positions of the bound and free probes are indicated at the left of the autoradiograph.



C

5'-TGCATCTCCCCCTACTGTACACCACcaaaagtgaattcgcatgcTTAGCACCTGGTTGGAGGGGGTTTATATA

EBEs ACGTAGAGGGGGATGACATGTGGTGgttttcacttaagcgtacgAATCGTGGACCAACCTCCCCCAAATATAT

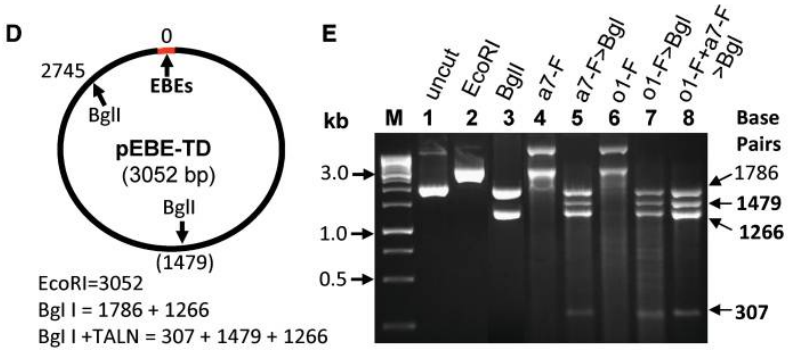


Figure 2.3. Target DNA digestion by the FN-AvrXa7, AvrXa7-FN and PthXo1-FN TALNs. (**A**) Circular plasmid containing a 400 bp *Os11N3* promoter region (thick bar in red) with one AvrXa7 EBE site (position 0). EcoNI cuts once at position 842 and MluI cuts twice at positions 1405 and 2246 relative to EBE of the plasmid. The expected sizes of fragments after cleavages are shown below the plasmid map. **(B)** Gel image of plasmid DNA treated with different enzymes individually or in combinations. M, 1 kb markers with size labeled on the left side. Different treatments are indicated above lanes 1 to 9 and in combination with TALNs labeled below lanes 4 to 9. Arrows in red indicate DNA fragment sizes expected if the NF-AvrXa7 or AvrXa7-FN nuclease cleavage takes place correctly at the AvrXa7 EBE target site in pTOP/Os11N3. (**C**) Structure of adjacent PthXo1 and AvrXa7 EBE sites (underlined) in a tail-to-tail orientation (to allow FN to FN proximity) and separated by 19 spacer nucleotides (lower case letters) which contains a unique EcoRI site (underlined). (**D**) Map of plasmid pEBE-DT with predicted cleavage sites for EcoRI (position 0), PthXo1-FN and/or AvrXa7-FN EBEs and BglI (positions 1479 and 2745). (**E**) Gel image of intact pEBE-DT (lane 1), EcoRI–linearized plasmid (2), BglI-cleaved plasmid (3), AvrXa7-FN digested plasmid DNA (4), AvrXa7-FN digested plasmid followed by BglI digestion (5), PthXo1-FN digested plasmid (6), PthXo1-FN digested plasmid followed by BglI digestion (7), and double digestion with PthXo1-FN and AvrXa7-FN (at one-fifth the concentrations of each protein used individually in reactions for lanes 3 to 6) followed by BglI digestion (8). The label “M” indicates the DNA marker lane with marker sizes provided in kilo-base (kb). Arrows to the right indicate sizes of DNA fragments expected from complete DNA digestion with BglI and partial digestion with PthXo1-FN (lane 5), AvrXa7-FN (lane 7), or PthXo1-FN combined with AvrXa7-FN (lane 8).

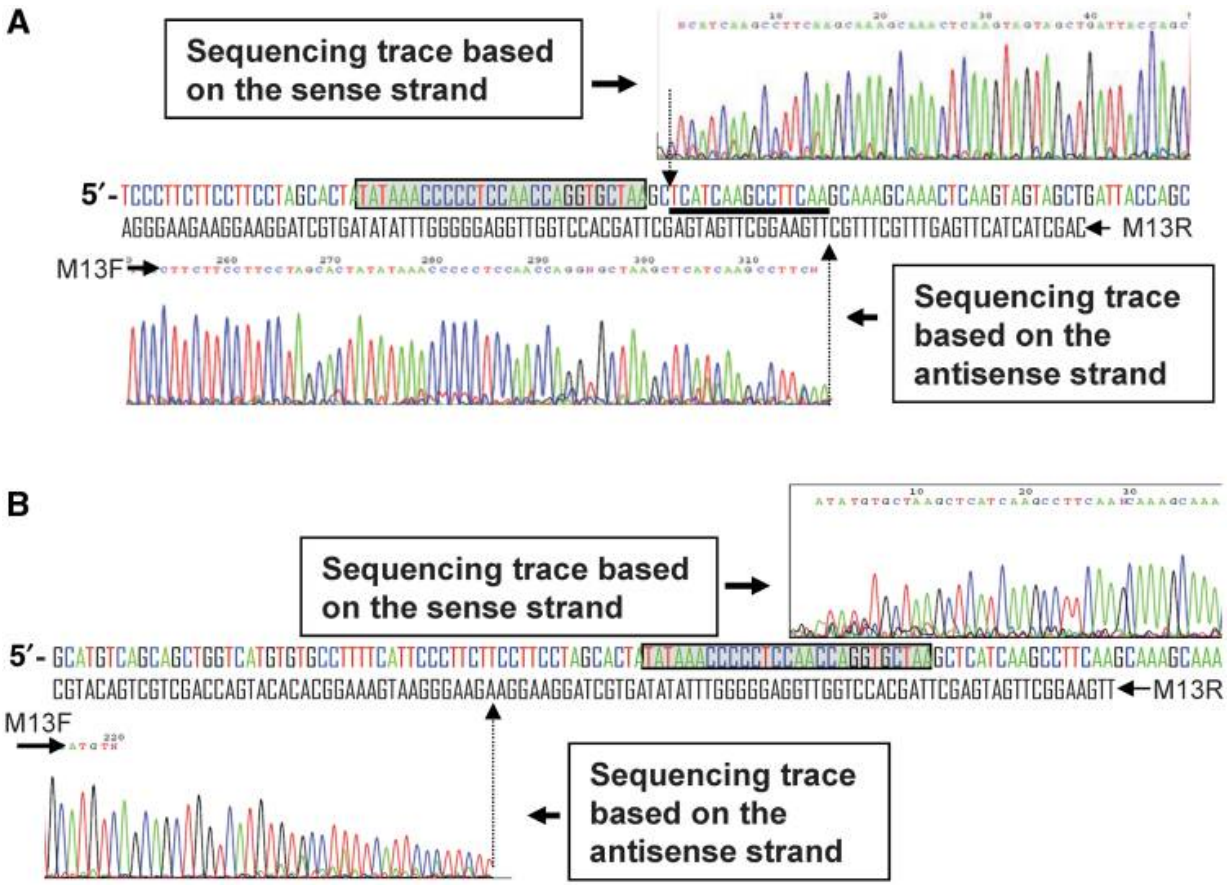


Figure 2.4. DNA sequencing revealing the major cleavage sites created by the AvrXa7-FN (A) and FN-AvrXa7 (B). **(A)** DNA sequencing chromatogram above the original *Os11N3* dsDNA sequence (sense strand colored for ease of viewing) is based on the sense strand of the 0.8 kb fragment purified from lane 8 in Figure 3B. The M13R primer was used for sequencing the sense strand. The chromatogram, which represents the sense strand sequence around the cleavage site, is in reverse-complement orientation for ease of viewing. The chromatogram below the dsDNA sequence is derived from the antisense strand of the 2.1 kb DNA fragment (also from lane 8 in Figure 3B) to the left of the predicted AvrXa7-FN cleavage site. The AvrXa7-FN binding site is boxed in shaded gray. The vertical arrows denote the obvious cleavage sites in both strands. The thick underline indicates the 15 bp region downstream of the EBE site in which the vast majority of AvrXa7-FN cleavage occurred. **(B)** DNA sequencing chromatograms of two DNA fragments derived from pTOP/11N3 treated with FN-AvrXa7 and purified from lane 5 in Figure 3B. Labeling is similar to that used for (A).

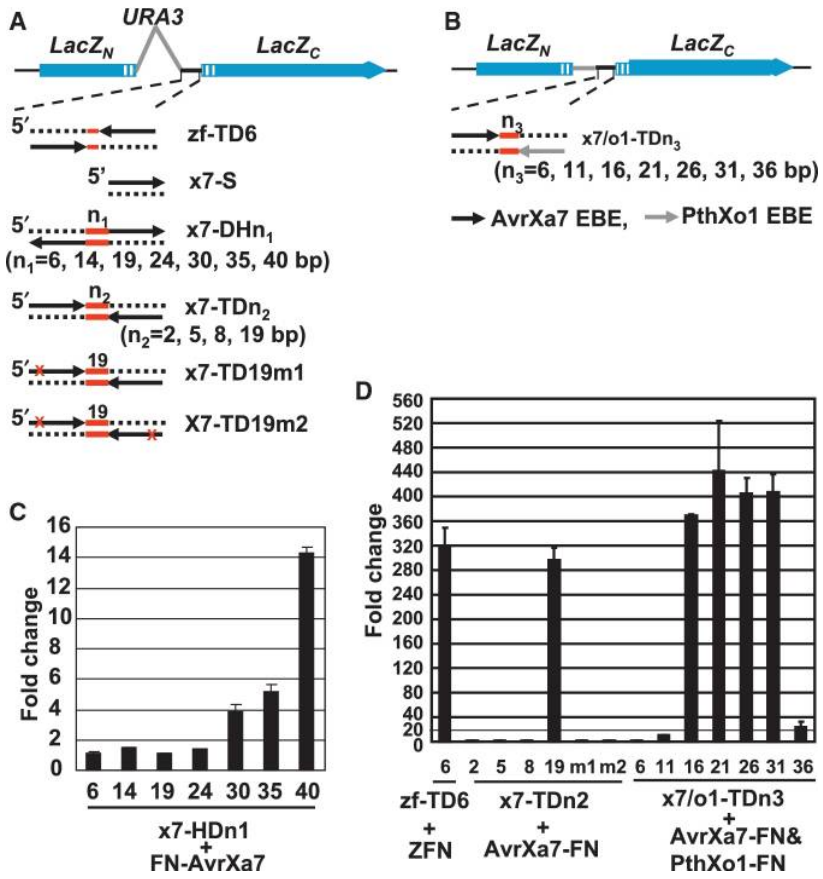

A
URA3
LacZN
LacZC
5′
zf-TD6
5′
x7-S
5′
n1
x7-DHn1
(n1=6, 14, 19, 24, 30, 35, 40 bp)
5′
n2
x7-TDn2
(n2=2, 5, 8, 19 bp)
5′
19
x7-TD19m1
5′
19
X7-TD19m2
B
LacZN
LacZC
n3
x7/o1-TDn3
(n3=6, 11, 16, 21, 26, 31, 36 bp)
AvrXa7 EBE,
PthXo1 EBE
C
Fold change
16
14
12
10
8
6
4
2
0
6 14 19 24 30 35 40
x7-HDn1
+
FN-AvrXa7
D
Fold change
560
520
480
440
400
360
320
280
240
200
160
120
80
40
20
0
6 2 5 8 19 m1 m2 6 11 16 21 26 31 36
zf-TD6
+
ZFN
x7-TDn2
+
AvrXa7-FN
x7/o1-TDn3
+
AvrXa7-FN&
PthXo1-FN

Figure 2.5. Yeast SSA assay for detection of FN-AvrXa7, AvrXa7-FN and PthXo1-FN TALN-induced homologous recombination. **(A)** Schematics of the reporter constructs (not drawn to scale) with dual BCR-ABL target site, AvrXa7 EBE site in different configurations drawn below. Two nonfunctional *LacZ* gene fragments (*LacZn* and *LacZc*, blue solid bars) were separated by the yeast *URA3* gene (gray line) and a multiple cloning site (MCS) (black line). The 125 bp duplicated *LacZ* coding sequences are represented as hatched blue boxes. The target sites in the reporter constructs are designated as zf-TD6 (ZFN dual sites), x7-S (single EBE site), x7-DHn1 (dual AvrXa7 EBEs in a head-to-head orientation separated by the red-lined spacers), x7-TDn2 (dual tail-to-tail AvrXa7 EBEs) followed by the number of spacer nucleotides. The constructs x7-TD19m1 and x7-TD19m2 contain a 19 bp spacer and carry a mutation in their EBE site caused by a 4 bp substitution (depicted as a red cross). **(B)** Schematics similar to (A) with a 125 bp fragment (gray line) instead of the *URA3* gene between the MCS and *LacZn*. **(C)** Activity of FN-AvrXa7 on reporters with different lengths of spacer between the AvrXa7 EBE sites (respective spacer lengths are given below each column). Fold change represents the β-galactosidase activity of yeast cells co-expressing the respective reporter and FN-AvrXa7 compared to cells containing the same reporter but lacking the FN-AvrXa7 expression vector. **(D)** Stimulation by the ZFN, AvrXa7-FN and AvrXa7-FN/pthXo1-FN constructs of their respective reporter genes containing various spacers (lengths in base pairs given below each column). Fold change represents the β-galactosidase activity of yeast cells co-expressing each reporter and its respective effector nuclease(s) (indicated below the columns) compared to those containing the same reporter but lacking the presence of genes encoding the respective nuclease(s). Error bars represent the standard deviation of three samples.

CHAPTER 3

MODULARLY-ASSEMBLED DESIGNER TAL EFFECTOR NUCLEASES FOR TARGETED GENE KNOCKOUT AND GENE REPLACEMENT IN EUKARYOTES

A paper published in *Nucleic Acids Research*‡

Ting Li [1, *], Sheng Huang [1, *], Xuefeng Zhao [2], David A. Wright [1], Susan Carpenter [3], Martin H. Spalding [1], Donald P. Weeks [4], Bing Yang [1, †]

[1] Department of Genetics, Development and Cell Biology; [2] Laurence H. Baker Center for Bioinformatics and Biological Statistics; [3] Department of Animal Science, Iowa State University, Ames, IA 50011 and [4] Department of Biochemistry, University of Nebraska, Lincoln, NE 68588, USA

Abstract

Recent studies indicate that the DNA recognition domain of TAL (transcription activator-like) effectors can be combined with the nuclease domain of FokI restriction enzyme to produce TAL effector nucleases (TALENs) that, in pairs, bind adjacent DNA target sites and produce double strand breaks between the target sequences, stimulating non-homologous end-joining and homologous recombination. Here, we exploit the four prevalent TAL repeats and their DNA recognition cipher to develop a "modular assembly" method for rapid production of designer TAL effector nucleases (dTALENs) that recognize unique DNA sequence up to 23 bases in any gene. We have used this approach to engineer ten dTALENs

---

‡ Reprinted with permission of *Nucleic Acids Research,* 2011, 39 (14): 6315-6325.

* These authors contributed equally to this work.

† To whom correspondence should be addressed. Tel: 515 294 2968; Fax: 515 294 5256; Email: byang@iastate.edu

to target specific loci in native yeast chromosomal genes. All dTALENs produced high rates of site-specific gene disruptions and created phenotypic mutant strains. Moreover, dTALENs stimulated high rates (up to 34%) of gene replacement by homologous recombination. Finally, dTALENs caused no detectable cytotoxicity and minimal levels of undesired genetic mutations in the treated yeast strains. These studies expand the realm of verified TALEN activity from cultured human cells to an intact eukaryotic organism and suggest that low-cost, highly dependable dTALENs can assume a significant role for gene modifications of value in human and animal health, agriculture and industry.

Introduction

Technologies for precise and efficient gene editing in living cells hold great promise in both basic and applied research, including therapeutic interventions for genetic diseases. These technologies exploit the ability of endonucleases to cause chromosomal double stranded DNA breaks (DSBs) and stimulate the subsequent breakage repair mechanisms in living cells (1, 2, 3). The two widely conserved, major repair pathways in eukaryotes are non-homologous end-joining (NHEJ) and homologous recombination (HR). NHEJ repair often results in mutagenic deletions/insertions and substitutions in the targeted gene; DSBs stimulate HR between the endogenous target gene locus and an exogenously introduced homologous donor DNA carrying the desired genetic alterations (4, 5, 6). At the forefront of these methods are custom-designed zinc-finger nucleases (ZFNs), which are hybrid proteins derived from the DNA binding domains of zinc finger (ZF) proteins and the nonspecific cleavage domain of the endonuclease FokI. However, despite their promise, widespread adoption of ZFNs is hampered by a bottleneck in custom-engineering zinc fingers with high

specificity and affinity for the DNA target sites (7, 8). Further, ZFN utility is somewhat limited by the number and location of potential target sites within a genome. Alternative strategies that overcome limitations in current technologies for targeted genome editing could greatly accelerate adoption of artificial nucleases for efficient gene disruption and gene replacement in a variety of heretofore recalcitrant eukaryotic organisms.

ZFN efficacy depends almost solely on the DNA binding specificity of their ZF domains (9), which theoretically can be supplanted by any high-fidelity DNA binding domain (10, 11, 12). TAL effector (TALE) proteins, a large group of bacterial plant pathogen proteins, have emerged as alternatives to ZF proteins. TALE proteins contain a varying number of centrally-located tandem 34-amino-acid repeats that mediate binding to a specific DNA target sequence, referred to as the effector binding elements (EBE). Each repeat is nearly identical except for two variable amino acids at positions 12 and 13, known as repeat variable di-residues (RVD). Polymorphism in the number of repeats (a range of 13 to 33) and in the RVD composition collectively determines the DNA binding specificity of individual TALE proteins. Remarkably, recognition of a specific DNA sequence is based on a fairly simple code wherein one base of the DNA target site is recognized by the RVD of one repeat (i.e., one repeat/one nucleotide). The sequential repeat arrangement in a single TALE protein thus specifies the contiguous DNA sequence that will be bound by that TALE protein (13, 14). TALE recognition studies also reveal a preference for certain RVDs over others in recognizing a particular nucleotide. In most cases, the RVDs asparagine and isoleucine (NI), histidine and aspartic acid (HD), asparagine and glycine (NG), and two asparagines (NN) each pair to recognize the nucleotides “A”, “C”, “T” and G”, respectively (13, 14). The prevalence of these four RVDs in the native TALEs made it possible to use them exclusively

to de novo synthesize or assemble TALE repeat arrays of up to 13 repeat units to target DNA sequences in plant and human cells (15, 16, 17). The addition of FokI nuclease domains to the C-termini of two paired synthetic TAL effectors (as in ZFNs) should allow for highly gene-specific gene targeting. Because of the modular nature of TALEs and the potential to use long DNA recognition sites, custom-made designer TALE nucleases (dTALENs) may overcome limitations of current ZFN technologies and may significantly advance the use of targeted genome editing for challenging long-term opportunities such as therapeutic repair of genetic diseases. Toward this end, we and others have recently shown that TALEs can be linked with FokI nuclease domain to direct targeted cleavage of DNA containing a specific EBE (17, 18, 19). More recently, Miller et al (20) demonstrated TALEN-mediated editing of endogenous genes in cultured human cells.

Here we report development of a modular assembly technology for custom-engineering dTALENs and characterization of ten such dTALENs for targeted gene modification. The method uses four basic repeats with RVDs of NI, NG, NN, and HD to generate 48 ready-to-use modules of single TAL repeat units that can be assembled up to 23 repeat units in any contiguous order. The modular assembly method is simple, fast and inexpensive and can be performed in most academic or industrial molecular biology laboratories. Remarkably, all ten dTALENs demonstrated efficient gene knockout and/or gene replacement in tests with three different chromosomal genes in yeast (*Saccharomyces cerevisiae*). The high success rate and facile synthesis of potent dTALENs against a variety of chromosomal targets further establishes dTALENs as an emerging and viable technology for precise gene modification in living cells. As previously practiced, all materials described in the present manuscript will be provided to other laboratories upon request.

Results

Tractability of yeast for testing gene modification by TALENs

We chose yeast, a classic eukaryotic model for studies of homologous recombination (30), as a platform to develop and test TALEN-based technologies for targeted gene modification based on the recent breakthroughs in the area of TALE research (13, 14). First, we determined if the yeast chromosome, which has not been subjected to genome modification using any artificial nucleases, is tractable for TALE nuclease-based modification. For this experiment, the EBEs for AvrXa7 and PthXo1 were precisely integrated in-frame between the first and second codons of the yeast *URA3* gene on chromosome 5. In parallel, DNA target sequences for the known ZFNs Zif268 and BCR-ABL were inserted into the identical *URA3* gene site for comparison of activities conferred by these two types of nucleases (Figure 3.1). The resulting yeast strains were prototrophic in uracil-free medium and sensitive to 5-fluoroorotic acid (5-FOA), indicating that the chimeric *URA3* genes were functional and intact. Yeast strains bearing the chimeric *URA3* genes were transformed with plasmids expressing the paired TALENs or ZFNs, grown for 5 days, plated on medium containing 5-FOA to select cells with an inactivated *URA3* gene. Approximately 0.9% of cells (out of ~10^6 cells) expressing paired TALENs produced 5-FOA resistant colonies while approximately 0.3% of cells expressing paired ZFNs yielded 5-FOA resistant colonies. In contrast, no 5-FOA resistant colonies were observed among ~10^6 cells (a ratio of <0.0001%) containing plasmids lacking a functional nuclease gene. Sequenced PCR products from the relevant alleles revealed all of the selected 5-FOA resistant clones harbor mutations (insertions and/or deletions that caused frame shift in the *URA3* genes) at the nuclease target sites (Figure 3.1). These results established similar gene disruptions and

comparable mutation rates elicited by ZFNs and TALENs targeted against genes in a native yeast chromosomal environment.

Modular dTALEN assembly

To fully realize the potential of TALENs, they must be custom-engineered to target any chromosomal DNA sequence of interest. By exploiting the repeat homology and the unique recognition sequence of the type IIS restriction enzyme BsmBI within each repeat of AvrXa7 or any TALE, we developed a method to assemble repeat domains in an exact predetermined order to recognize a specific DNA sequence in any gene of choice from any organism. Briefly, four AvrXa7 "core" repeats whose coding RVDs each recognize one of four nucleotides (i.e., NI, NG, NN and HD, respectively, for A, T, G, and C) were used to construct independent modules (single repeats) whose 5'- and 3'- ends were designed to form a unique 4-bp overhang with single-base polymorphism after digestion with BsmBI. The BsmBI site is immediately downstream of codons 18 and 19 of each repeat and BsmBI cleaves these two triplets into a 5'overhang of 4-bp at each end. The 18^{th} and 19^{th} codons are GCG CTG and can be modified into 8 variant triplets GC(A, T, G or C) (T, or C)TG (Figure 3.2A). Therefore, combinations of eight or fewer such overhangs, one at each end of a single repeat, were created without altering the encoded amino acids. This allowed the ordered ligation of 8 or fewer repeat modules in any predetermined sequence for construction of sub-arrays of repeats (Figure 3.2B). Each sub-array was cloned into the cloning vector pGEM-T and sequenced to confirm the correct number and order of the repeats. Similarly, multiple sub-arrays (2-3 in this study) were further assembled to match the order of nucleotides at the preselected genomic site. The AvrXa7-FN nuclease (18) lacking its repeat domain was used

as the scaffold for the assembled repeat domains, resulting in a finished dTALENs (Figure 3.2C).

To test the feasibility of our approach, we selected five distinct dual target sites (two in *URA3*, two in *LYS2*, and one in *ADE2*) based on the criteria: 1) "T" preceding each target sequence, 2) avoiding G rich blocks, and 3) 17-20 bp spacer between the inverted two EBE target sites. Accordingly, ten dTALENs were synthesized for gene targeting based on the preselected DNA coding sequences of the three yeast genes (Figure 3.3A). All ten dTALENs were expressed in yeast to levels comparable to those of hybrid nucleases made from PthXo1 or AvrXa7, except U3b-L which had somewhat lower expression.

To test the function of newly synthesized dTALENs and to reveal their relative DNA cleavage activity, we modified a transient and plasmid-borne single strand annealing (SSA) assay (31, 32) as a facile analytical tool. This method uses a yeast plasmid carrying a *lacZ* gene divided into upstream and downstream portions by insertion of two opposing EBEs, one recognized by a proven AvrXa7 TALEN (18) and the other EBE recognized by a candidate dTALEN. The separated *lacZ* fragments share at the 3' end of the upstream fragment and the 5' end of the downstream fragment a 125 bp segment of identical *lacZ* sequence. If a functional dTALEN is co-expressed in yeast cells with the AvrXa7 TALEN along with the target *lacZ* gene, it will bind to its target EBE sequence adjacent to the AvrXa7 TALEN and, thereby, create a double strand DNA break (DSB) (Figure 3.3B). The duplicated *lacZ* sequences are thus available for homologous recombination and restoration of an intact *lacZ* gene. The amount of β-galactosidase activity produced by the transformed yeast cells thus provides a measure of the amount of DNA cleavage supported by the candidate dTALEN. We initially tested this assay by first pairing the proven AvrXa7 TALEN with another proven

TALEN PthXo1 (18). The activity of this TALEN pair provides a standard against which the activity of any candidate dTALENs can be judged (Figure 3.3C). The ten newly produced dTALENs, each designed to recognize a specific 17 to 23 base DNA sequence in different yeast genes, were all found in yeast SSA tests to function nearly as well or better than the standard AvrXa7/PthXo1 TALEN pair (Figure 3.3C).

Efficient gene modification by dTALEN-induced NHEJ and HR

As a final evaluation of the function of the five pairs of dTALENs, we tested their ability to elicit site-specific DNA alterations at the preselected target sites in the *URA3, LYS2* and *ADE2* genes, which all have easily scored knockout phenotypes. Yeast cells were transformed with individual pairs of dTALEN-expressing plasmids and grown for five days on synthetic complete medium to allow accumulation and activity of the heteromeric dTALEN pair. Two yeast cultures were transformed separately with one or the other pair of dTALENs targeting the *URA3* gene and plated for 5-FOA selection of cells lacking a functional *URA3* gene. Likewise, two yeast cultures were transformed separately with one or the other pair of dTALENs targeting the *LYS2* gene and plated for α-aminoadipate (α-AA) selection of cells with *LYS2* gene mutations. *URA3* and *LYS2* mutants were obtained at a rate of approximately 10^{-4} to 10^{-2} (Figure 3.4A). Yeast cells transformed with the dTALENs pair targeting the *ADE2* gene were plated on medium containing limiting adenine concentrations that result in the formation of pink colonies if a functional *ADE2* gene is not present. Pink colonies appeared with a frequency of 0.15 % (Figure 3.4A). In contrast, ~10^6 yeast cells carrying plasmids lacking a functional dTALEN gene pair yielded no colonies resistant to 5-FOA or α-AA or with a pink color. Sequence analysis of PCR-amplified genomic DNA from the relevant target sites in several putative *URA3*, *LYS2*, and *ADE2* gene knockout mutants

revealed that all alleles harbored mutations at the dTALEN target sites as expected. A high proportion of the mutated loci contained deletions in a range from 1 to 75 bp (Figure 3.4B).

One experimentally and practically important virtue of DSBs caused by agents such as ZFNs (3) is that they increase rates of recombination between the DNA sequences within a broken gene and homologous endogenous or exogenously-supplied DNA sequences, which enables powerful gene replacement research opportunities. To determine if dTALEN-mediated DSBs stimulate homologous recombination (HR), we targeted the *URA3* gene for breakage with artificial dTALEN pairs in the presence of two different exogenously supplied DNA fragments, one containing a *URA3* gene interrupted by a neomycin phosphotransferase II (*NPTII*) expression cassette and the other a DNA fragment with the *URA3* ORF deleted. Both fragments contained at their 5' ends a 0.5 kb segment homologous to the 5' end of the *URA3* gene and, at their 3' ends, a 0.2 kb segment from the 3' end of the *URA3* gene. Yeast cells of YPH500c, a strain containing the functional *URA3* with the integrated EBEs for AvrXa7 and PthXo1, were transformed with one set of three different paired TALENs (i.e., U3a-L & U3a-R, U3b-L & U3b-R, and AvrXa7-FN & PthXo1-FN) and one of the two modified *URA3* gene constructs. The transformants were plated after five days of incubation on either selective medium containing 5-FOA (deleted *URA3* construct) or medium containing the neomycin-like antibiotic, G418 (*NPTII* interrupted *URA3*). Cells transformed with TALENs and the *URA3* ORF deletion construct (*DURA3*) showed frequencies of 5-FOA resistant colonies in the range of 4.5% to 27%. The negative control [transformed with donor DNA and plasmids lacking a nuclease gene (Empty Plasmid)] yielded 5-FOA resistant colonies at a rate of 0.01% (Figure 3.4C). The frequency of gene replacement for cells transformed with the *NPTII*-interrupted *URA3* gene construct (*NPTII*) was in the range of 9%

to 34% with the negative control (Empty Plasmid) displaying only ~0.1% gene replacement activity (Figure 3.4C). Overall, the enhancement of TALEN-induced gene replacement was between100 to 2700 fold higher than the control. The scale and consistency in the stimulation of HR by dTALENs suggests they have the potential to promote HR when used in eukaryotic cells that lack other sufficiently robust mechanisms to facilitate HR.

Genome-wide undesired mutations caused by TALENs

Some ZFNs have been found to be associated with toxicity in living cells (33, 34, 35, 36, 37, 38, 39). Whether such effect also exists for TALENs is unknown. To test the possibility, yeast cells were transformed with plasmids encoding six pairs of TALENs (one pair of native TALE-derived nucleases and five pairs of synthetic dTALENs targeting sequences in a size range between 17 bp to 27 bp) and one pair of known ZFNs [Zif268 and BCR-ABL with target sequences of 9 bp (33, 40)]. The transformed cells were grown in SC medium lacking leucine and histidine. TALEN-expressing yeast cells displayed no distinct phenotype in terms of cell viability and proliferation compared to the control, as did yeast strains with ectopic expression of paired ZFNs during five days of growth (Figure 3.5). Yeast cells expressing individually each of the eight introduced nuclease genes showed similar results. The results indicate the lack of any apparently deleterious effects on the viability of yeast cells expressing the tested TALENs and ZFNs under our experimental conditions.

Undesired genetic mutations (genotoxicity) due to promiscuous cleavage also have been reported for ZFNs (33, 40, 41), but it is unknown whether TALENs also induce such genotoxicity. It is possible that such mutations occurred in our cell survival experiment but did not visibly affect the viability of the yeast cells. The haploid nature and relatively small size (~12 mega bases) of the yeast genome in combination with the next-generation

sequencing technology enabled us to investigate any potential genome-wide undesired effects of TALENs. Five strains, including the parental strain YPH500 and four mutants, were chosen to investigate the occurrence of unintended mutations in addition to the site-specific mutations at the *URA3* locus mediated by the respective nucleases. The four mutant strains included one that contained a deletion in the chimeric *URA3* gene at the integrated EBE site targeted by the paired nucleases of natural TAL effectors AvrXa7 and PthXo1, one that contained an insertion in the chimeric *URA3* gene with integrated target sequences for the ZFNs Zif268 and BCR-ABL, and the other two that each contained a deletion mutation in the wild type *URA3* genes induced by the paired dTALENs U3a-L&-R and U3b-L&-R, respectively. The five strains were sequenced using Illumina/Solexa Genome Analyzer II, and their genomes were assembled with the coverage depth [(number of reads x Average read length)/(size of genome)] in a range of 135 to 170X (sequences available upon request). The assembled genomic sequence of each mutant strain was first screened for possible mutations at the sites that matched or loosely matched (defined as at least two-thirds of match as the cut-off) either intended sub-dual target sequence for the respective nucleases. No mutations were found at those locations other than the specific-site mutations mediated by the respective pairs of TALENs and ZFNs in each genome (Data not shown). However, alignment of the genomes of the individual nuclease treated strains with that of parental strain revealed a number of mutations in each strain (in a range of 3 to 5 mutations). These mutations were almost all nucleotide substitutions instead of predominant deletions/insertions at the nuclease target sites and they, therefore, were highly likely to be simultaneous mutations which might or might not be associated with the nuclease treatment.

Discussion

We have developed a simple, cost effective method to assemble functional TALE DNA binding domains, combine them with a nuclease module and confirm their activity against specific gene targets *in vivo*. Using this method, dTALENs were readily produced and rapidly validated using a facile yeast SSA assay. Importantly, these dTALENs were then demonstrated to function as pairs to mediate efficient gene disruption by NHEJ and gene replacement by HR at specific yeast chromosomal loci.

The DNA binding domain encoded by a TALE gene is modular by nature, having 13 to 33 RDV units that bear a simple code for target site recognition (13, 14). However, the high repeat homology imposes technical difficulty when using PCR-based de novo gene synthesis methods (42) for construction of the lengthy repeat arrays required for high-level DNA specificity. To simplify gene synthesis and reduce dTALEN production costs, a modular assembly technique was developed. TALE repeat domains have been assembled from the single RVD coding repeats to bind DNA targets predicted by the cipher; but they were assembled in a random way, not in a predetermined order and not based on the target DNA sequence (13). Our method involves creation of 48 ready-to-use modules and their assembly into repeat arrays in a controlled order based on any user chosen target DNA sequences (Figure 3.2). The ready-to-use modules can also be adapted for high-throughput dTALEN synthesis. By manipulating the combination of 5’ and 3’ unique overhangs of the modular sets, dTALENs with varying number of repeat units (up to 23 bp in this study) can be designed to better meet the criteria in choosing the dual target sites. Our facile modular TALEN assembly method contrasts with another recently reported method that involves

several rounds of PCR amplifications and ligations to assemble individual repeats into 12 repeat TAL effectors for gene activation in human cells was reported (16).

A modified plasmid-based yeast SSA *lacZ* assay was employed to determine activity of the candidate dTALENs before use for final gene targeting. This test system is based on assays developed to initially validate engineered meganucleases and ZFNs (31, 32) and was used previously in a preliminary form for TALEN testing (17, 18). The original yeast SSA assay relied on inverted repeat DNA targets and engineered homodimeric nucleases (32). For our modified assay, one candidate dTALEN was paired with the proven TALEN AvrXa7-FN (18) to target two adjacent EBEs, one for recognition by the candidate dTALEN and the other for binding of AvrXa7-FN. The paired TALENs derived from the natural TALEs AvrXa7 and PthXo1 were used as a standard. Another advantage of the modified SSA assay is that one EBE and the spacer remain constant so only the second EBE has to be replaced to test a different dTALEN. When testing many dTALENs, this represents a significant (~5X) cost saving due to using much shorter oligonucleotides to create a single (vs. double) TALEN site while allowing rapid evaluation of the activity of each individual candidate dTALEN. Moreover, construction of 3' to 3' EBE sites containing nonidentical DNA sequences avoids the technical challenges associated with cloning and stably maintaining inverted EBE sites containing identical TALEN recognition sequences. From a broader perspective, this plasmid-based transient assay should be readily adapted as a facile screening tool in animal and plant cell culture systems.

The ten dTALENs engineered for this study were designed in pairs to target the native sequences of the yeast genes, *URA3*, *LYS2* and *ADE2*. Each gene has a selectable or easily scored phenotype, so any dTALEN activity should be readily apparent. Expression of

dTALEN pairs under appropriate selections indicated that all five dTALEN pairs were functional in creating phenotypic mutant strains whose site-specific DNA alterations were all confirmed by genotyping (Figure 3.4). The activity of the five dTALEN pairs was comparable to the standard TALEN pair, AvrXa7 and PthXo1. Two pairs of dTALENs also were tested to compare their ability to mediate HR by targeting *URA3* for gene replacements against HR stimulated by the standard TALEN pair. The results indicated that dTALENs and standard TALENs increase HR at a comparable rate (Figure 3.4). The low rates of gene disruption via NHEJ in present study are probably attributable to the cryptic NHEJ repair pathway in yeast. The NHEJ repair of DSBs in yeast is mostly accurate. For example, repair to DSBs with 4-base overhangs led to an error frequency of only about 1% (43, 44) Only cells with mutagenic alterations in the intentionally targeted genes might be recovered (for *URA3* and *LYS2*) or detected (for *ADE2*) by using our procedures. Thus, while the number of dTALENs examined was somewhat limited, these experiments establish that dTALENs are active *in vivo* in promoting NHEJ- and HR-mediated gene modifications at endogenous loci in an intact, free-living, eukaryotic organism and, thus, verify the power of TALENs for targeted genome editing in eukaryotes beyond the TALEN-mediated gene disruption previously demonstrated in cultured human cells (20).

Taken together, the results of this study indicate that the modular assembly technique is valid, that the yeast SSA assay is a reliable and facile indicator of TALEN activity and that functional dTALENs can target a diverse range of genomic loci. Our study also establishes that yeast, as an intact eukaryotic organism, is a reliable platform to develop and potentially refine engineered nuclease based technology for targeted genome editing. The designer TALEN technology described here may overcome a number of critical hurdles (e.g. high

cost, limited plasticity and high rates of failure) faced in designing and producing other types of nucleases with DNA targeting capabilities (7, 8). The opportunity to build dTALENs with quite long DNA recognition domains bodes well for developing TALENs with exceptionally high accuracy in targeting any gene in any organism, including eukaryotes with highly complex genomes. Such technology has significant potential in experimental biology and medicine and in the development of products with value for human and animal health, agriculture and in a wide range of life sciences industries.

Materials And Methods

Yeast strains and growth conditions.

Yeast strains YPH499, YPH500 and RFY231 as previously described (18, 21) were grown in nutrient medium YPD or synthetic complete medium (SC) lacking the appropriate nutrients. 5-FOA was used at 0.1% in SC medium as described (22). α-AA was used at 0.2% in SC medium lacking the normal nitrogen source, but containing a small amount of lysine (30 mg/L) (23). The adenine limited medium is SC medium containing a limited amount of adenine (10 mg/L) and lacking leucine and histidine (24).

dTALEN constructs.

Four repeats, each encoding the RVD of NI, NG, HD or NN from AvrXa7, were used as the "core" repeats. [More recently, NK has been substituted for NN in the recognition of G nucleotides based on the observations of Miller et al. (20)]. Using a combination of 12 forward and 11 reverse primers (**Supporting Information Table S1**), twelve repeat sets were constructed from the 4 "core" repeats. For construction of the 1st 8-mer repeat array, combinations of the PCR primers were: TAL-Sph-F&TALcgct-R, TALcgct-F&TALcctt-R,

TALcctt-F&TALctct-R, TALctct-F&TALcgtt-R, TALcgtt-F&TALcttt-R, TALcttt-F&TALcatt-R, TALcatt-F&TALccct-R, and TALccct-F&TAL/Pst-R for repeat Set 1 to 8, respectively. For ligation of the 2nd and the 3rd 8-mer repeat arrays, two additional repeat sets each corresponding to Set 1 were constructed by using primers TAL/Pst-F&TALcctt-R and TAL/Bsr-F&TALcctt-R, respectively; another two repeat sets corresponding to Set 8 were also constructed by using primers TALccct-R&TAL/Bsr-R and Tailcact-F&TAL-Sph-R, respectively. In total 12 repeat sets were generated and then individually digested with BsmBI. Based on the base entity and order of the preselected target DNA sequence (e.g. any sequence combination of 22 bp long), one repeat from each of 8 repeat sets was sequentially selected for one ligation reaction to construct the 8-mer repeat array. Each set of ligated DNA was directly cloned into pGEM-T for sequencing. Once all confirmed, the 1st 8-mer array was digested with SphI and PstI, the 2nd 8-mers with PstI and BsrGI, and the 3rd array with BsrGI and SalI. The three purified DNA fragments were ligated into pSK/AvrXa7 (18) that was digested with SphI and SalI, resulting in pSK/dTALE plasmids. The repeat regions of individual dTALEs were swapped with the TALE repeat domain of AvrXa7 in pCP3M-AvrXa7-FN or pCP4M-AvrXa7-FN (18), resulting in the chimeric genes encoding the fusion of the full-length dTALE and the C-terminal FokI homodimeric cleavage domain. The expression level of these nuclease genes should be moderate due to the low copy number (about 1 copy per cell) of the centromeric plasmids pCP3M and pCP4M and the strong translation elongation factor 1α promoter (25).

Yeast SSA assay for dTALEN activity.

The individual target sites were constructed into pCP5 (18) in a tail-to-tail (3' to 3') orientation with AvrXa7 EBE. The oligonucleotide sequences for all TALEN EBEs were

provided in Table S2. AvrXa7 EBE (x7-EBE-F&R) was cloned into the PstI and SpeI sites of pCP5. The EcoRI site immediately upstream of SpeI site enabled all other TALEN EBEs to be cloned individually between EcoRI and SpeI sites, resulting in the reporter plasmids for individual TALENs in pair with AvrXa7-FN. The assay for individual TALENs each in pair with AvrXa7-FN was performed in a manner similar to that described (18). The assay was performed in triplicate.

Targeted gene disruption of URA3, LYS2 and ADE2 in yeast.

The Ty1 inserted *URA3* gene (*ura3-52*) in YPH500 was restored to a functional *URA3*, resulting in strain YPH500a. Similarly, the target sequences for ZFNs (Zif268 and BCR-ABL) and for TALENs of AvrXa7 and PthXo1 were individually integrated into the *URA3* gene immediately downstream of the start codon and used to restore the *ura3-52* mutant, resulting in strain YPH500b and YPH500c, respectively. YPH500b was transformed with plasmids pCP3M/Zif268-FN and pCP4M/BCR-ABL-FN. DNA fragment for the zinc finger protein Zif268 was PCR amplified from pMal-Zif268 (kindly provided by David J. Segal) using primers Zif268-F and Zif268-R and cloned in frame with FokI cleavage domain in pCP3M. Construct pCP4M/BCR-ABL-FN was described previously (18). YPH500c was transformed with pCP3M/AvrXa7-FN and pCP4M/PthXo1-FN, two plasmids previously described (18). YPH500a was transformed with plasmids expressing the paired dTALENs U3a_L & U3a_R and U3b_L & U3b_R. The respective yeast strains were transformed with plasmids pCP3M and pCP4M as negative control for each paired nucleases. The transformants were grown on the SC medium lacking histidine and leucine for 5 days before plating on the SC medium containing 0.1% 5-FOA for selection of resistant colonies and in parallel on SC medium without 5-FOA to test for plating efficiency. Genomic DNA extracted

from a number of 5-FOA resistant colonies for each pair of nucleases was used for PCR amplification of the relevant regions. The PCR products were sequenced using the respective primers. Similarly, gene disruption of *LYS2* and *ADE2* was performed in yeast strain RFY231. See **Supporting Information** for detailed information about the creation of these strains and gene disruption.

HR-based URA3 gene replacement stimulated by dTALENs.

Donor DNA constructs were made each with the ORF of *URA3* deleted (pΔura3) and replaced by the *NPTII* expression cassette (pΔura3::Kan). YPH500c was transformed with donor construct and individual pair of plasmids expressing AvrXa7-FN & PthXo1-FN, dTALENs U3a-L & -R, U3b-L & -R, and the plasmids lacking a nuclease gene as control. The transformed cells were grown in SC media lacking histidine and leucine for five days, then plated on SC medium supplemented with either 5-FOA (for pΔura3) or YPD medium supplemented with 200 mg/L of G418 (for pΔura3::Kan), the duplicated cells were in parallel plated on the SC medium or YPD medium for plating efficiency. See **Supporting Information** for additional details.

Cell growth assay.

YPH500 cells were transformed with individual plasmids as indicated in the text. Three single colonies for each plasmid were picked and grown to a concentration of OD_{600}=1.0. The cells were serially diluted and applied to appropriate solid medium for growth, which was observed daily for five days.

Solexa sequencing and data analysis.

The Genomic DNA from each of five yeast strains derived from YPH500, which is congenic with S288C (26), was extracted as described by Philippsen et al. (27). The DNA

processing and sequencing were performed according to manufacturer's instruction for the Illumina/Solexa Genome Analyzer II at the Iowa State University DNA facility. The Illumina short reads for each strain were aligned to the yeast reference genome S288C to assemble each genome using the BWA software (28). The consensus sequences and polymorphisms among the five sequenced strains and S288C were delineated using SAMtools (29).

Acknowledgements

The authors thank Dr. Dan Voytas for providing the components of the yeast SSA system and Drs. Keith Joung and David Segal for providing materials to make constructs of ZFNs BCR-ABL and Zif268, respectively.

Reference

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

2. Carroll,D. (2008) Progress and prospects: Zinc-finger nucleases as gene therapy agents. *Gene Ther.,* **15,** 1463-1468.
3. Urnov,F.D., Rebar,E.J., Holmes,M.C., Zhang,H.S. and Gregory,P.D. (2010) Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.,* **11,** 636-646.

4. Jasin,M. (1996) Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet.,* **12,** 224-228.

5. Vasquez,K.M., Marburger,K., Intody,Z. and Wilson,J.H. (2001) Manipulating the mammalian genome by homologous recombination. *Proc. Natl. Acad. Sci. U. S. A.,* **98,** 8403-8410.

6. Bibikova,M., Beumer,K., Trautman,J.K. and Carroll,D. (2003) Enhancing gene targeting with designed zinc finger nucleases. *Science,* **300,** 764.

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

8. Kim,J.S., Lee,H.J. and Carroll,D. (2010) Genome editing with modularly assembled zinc-finger nucleases. *Nat. Methods,* **7,** 91; author reply 91-2.

9. Kim,Y.G., Cha,J. and Chandrasegaran,S. (1996) Hybrid restriction enzymes: Zinc finger fusions to fok I cleavage domain. *Proc. Natl. Acad. Sci. U. S. A.,* **93,** 1156-1160.

10. Boch,J. and Bonas,U. (2010) Xanthomonas AvrBs3 family-type III effectors: Discovery and function. *Annu. Rev. Phytopathol.,* **48,** 419-436.

11. White,F.F., Potnis,N., Jones,J.B. and Koebnik,R. (2009) The type III effectors of xanthomonas. *Mol. Plant. Pathol.,* **10,** 749-766.

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

13. 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.

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

15. Morbitzer,R., Romer,P., Boch,J. and Lahaye,T. (2010) Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *Proc. Natl. Acad. Sci. U. S. A.,* **107,** 21617-21622.

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

17. 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.

18. Li,T., Huang,S., Jiang,W.Z., Wright,D., Spalding,M.H., Weeks,D.P. and Yang,B. (2010) TAL nucleases (TALNs): Hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.,* **39,** 359-372.

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

20. Miller,J.C., Tan,S., Qiao,G., Barlow,K.A., Wang,J., Xia,D.F., Meng,X., Paschon,D.E., Leung,E., Hinkley,S.J., et al. (2011) A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.,* **29,** 143-148.

21. Kolonin,M.G. and Finley,R.L.,Jr. (1998) Targeting cyclin-dependent kinases in drosophila with peptide aptamers. *Proc. Natl. Acad. Sci. U. S. A.,* **95,** 14266-14271.

22. Boeke,J.D., Trueheart,J., Natsoulis,G. and Fink,G.R. (1987) 5-fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods Enzymol.,* **154,** 164-175.

23. Chattoo,B.B., Sherman,F., Azubalis,D.A., Fjellstedt,T.A., Mehnert,D. and Ogur,M. (1979) Selection of lys2 mutants of the yeast SACCHAROMYCES CEREVISIAE by the utilization of alpha-AMINOADIPATE. *Genetics,* **93,** 51-65.

24. Ugolini,S. and Bruschi,C.V. (1996) The red/white colony color assay in the yeast saccharomyces cerevisiae: Epistatic growth advantage of white ade8-18, ade2 cells over red ade2 cells. *Curr. Genet.,* **30,** 485-492.

25. Mumberg,D., Muller,R. and Funk,M. (1995) Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene,* **156,** 119-122.

26. Johnston,M. and Davis,R.W. (1984) Sequences that regulate the divergent GAL1-GAL10 promoter in saccharomyces cerevisiae. *Mol. Cell. Biol.,* **4,** 1440-1448.

27. Philippsen,P., Stotz,A. and Scherf,C. (1991) DNA of saccharomyces cerevisiae. *Methods Enzymol.,* **194,** 169-182.

28. Li,H. and Durbin,R. (2009) Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics,* **25,** 1754-1760.

29. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and 1000 Genome Project Data Processing Subgroup. (2009) The sequence Alignment/Map format and SAMtools. *Bioinformatics,* **25,** 2078-2079.

30. Aylon,Y. and Kupiec,M. (2004) DSB repair: The yeast paradigm. *DNA Repair (Amst),* **3,** 797-815.

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

32. Townsend,J.A., Wright,D.A., Winfrey,R.J., Fu,F., Maeder,M.L., Joung,J.K. and Voytas,D.F. (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature,* **459,** 442-445.

33. Porteus,M.H. and Baltimore,D. (2003) Chimeric nucleases stimulate gene targeting in human cells. *Science,* **300,** 763.

34. Alwin,S., Gere,M.B., Guhl,E., Effertz,K., Barbas,C.F.,3rd, Segal,D.J., Weitzman,M.D. and Cathomen,T. (2005) Custom zinc-finger nucleases for use in human cells. *Mol. Ther.,* **12,** 610-617.

35. Porteus,M.H. (2006) Mammalian gene targeting with designed zinc finger nucleases. *Mol. Ther.,* **13,** 438-446.

36. Miller,J.C., Holmes,M.C., Wang,J., Guschin,D.Y., Lee,Y.L., Rupniewski,I., Beausejour,C.M., Waite,A.J., Wang,N.S., Kim,K.A., et al. (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.,* **25,** 778-785.

37. Szczepek,M., Brondani,V., Buchel,J., Serrano,L., Segal,D.J. and Cathomen,T. (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat. Biotechnol.,* **25,** 786-793.

38. Cornu,T.I., Thibodeau-Beganny,S., Guhl,E., Alwin,S., Eichtinger,M., Joung,J.K. and Cathomen,T. (2008) DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Mol. Ther.,* **16,** 352-358.

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

40. Wright,D.A., Thibodeau-Beganny,S., Sander,J.D., Winfrey,R.J., Hirsh,A.S., Eichtinger,M., Fu,F., Porteus,M.H., Dobbs,D., Voytas,D.F., et al. (2006) Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat. Protoc.,* **1,** 1637-1652.

41. Beumer,K., Bhattacharyya,G., Bibikova,M., Trautman,J.K. and Carroll,D. (2006) Efficient gene targeting in drosophila with zinc-finger nucleases. *Genetics,* **172,** 2391-2403.

42. Hoover,D.M. and Lubkowski,J. (2002) DNAWorks: An automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.,* **30,** e43.

43. Wilson,T.E. and Lieber,M.R. (1999) Efficient processing of DNA ends during yeast nonhomologous end joining. evidence for a DNA polymerase beta (Pol4)-dependent pathway. *J. Biol. Chem.,* **274,** 23599-23609.

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

Figures And Tables

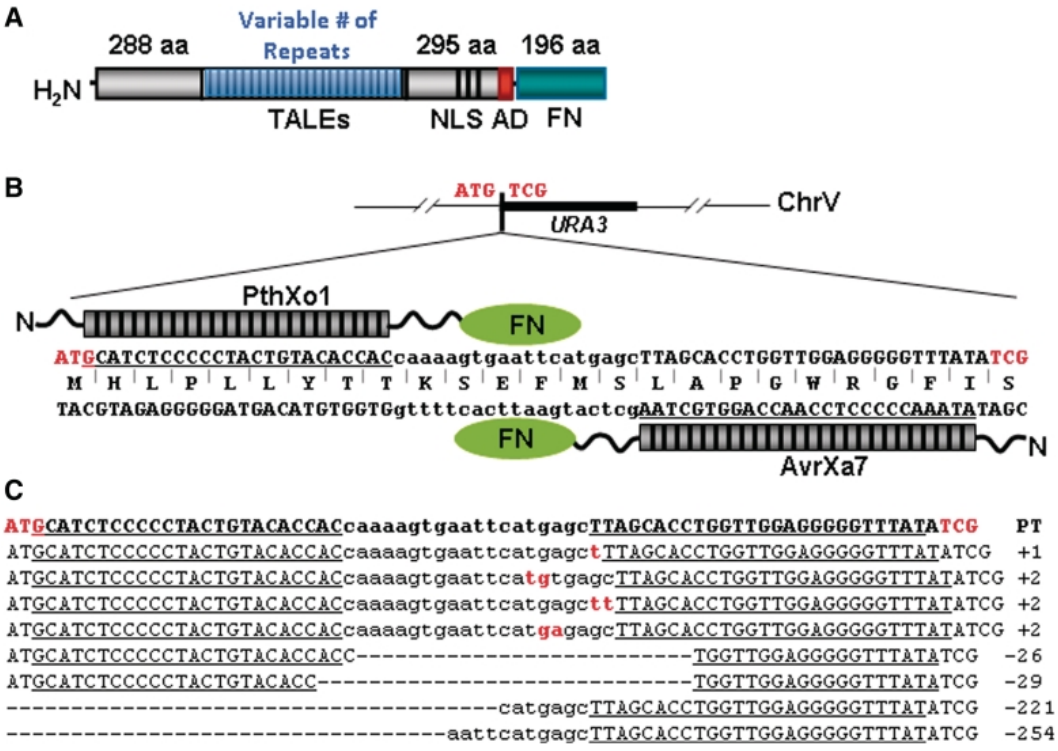
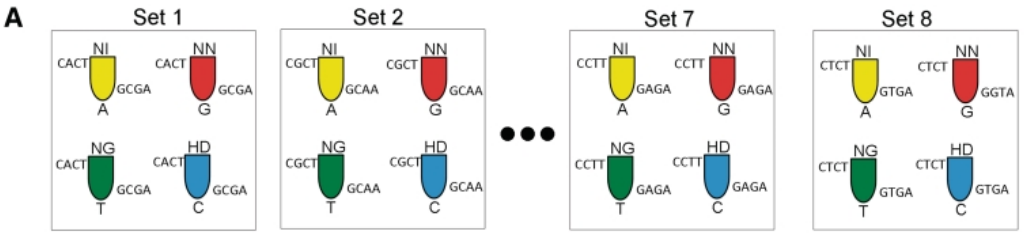


Figure 3.1. DNA alterations by natural TAL effector-derived TALENs at chromosomally-integrated paired EBE target sites in the *URA3* gene. **(A)** Schematic of TALENs used in the study. Full-length TALEs were fused with the homodimeric cleavage domain of FokI (FN) with number of amino acids (aa) shown above each region (NLS, nuclear localization motifs; AD, transcription activation domain). **(B)** Target sequences between the first and second codons of the yeast *URA3* gene for cleavage by hybrid TALENs derived from AvrXa7 and PthXo1. **(C)** Alignment of genomic sequences of mutants and their parental strain at the TALEN target site. The number of nucleotides inserted (lowercase letters in red) or deleted (dashes) compared to parental sequences (PT) from each mutant colony is indicated on the right of each sequence.



↓ Synthesize 8 gene sets each with 4 core RVD-coding modules and bearing 5′ and 3′ termini as dictated by the table to the right

↓ Digest with BsmB1 to generate unique 5' and 3' overhangs with single nucleotide polymorphisms (as dictated by table to the right)

↓ Select a 16 (or 24) bp DNA recognition site (e.g., AGGTACTCGAATCCTG)

↓ For 1st 8-mer repeat, pick an NI from set 1, an NN from set 2, an NN from set 3, etc., etc. to create a pool of 8 single-repeat genes

↓ Anneal and ligate the 8 single-repeat genes into a predetermined order to produced the desired 8-mer

Set	5' Overhang	3' Overhang
1	CACT	GCGA
2	CGCT	GCAA
3	CGTT	GAAA
4	CTTT	GTAA
5	CATT	GGGA
6	CCCT	GGAA
7	CCTT	GAGA
8	CTCT	GTGA

↓ Pooled ligation of 8-mer arrays

↓ Cloning into the pAvrXa7-FN scaffold plasmid

↓ Generation of dTALEN recognizing 16 (or 24) bp target site

Figure 3.2. Design and modular construction of dTALENs. (A) TALEN repeat gene sets for modular construction of multi-repeat TALEN genes. Each set contains four single-repeat genes each encoding one of the four "core" TALEN repeat modules containing NI, NG, NN, and HD RVDs with binding specificity for A, T, G and C nucleotides, respectively. The core repeats in each set (boxed) contains a 5' and 3' termini unique to that particular set – as listed in the inset table. For construction of a TALEN repeat recognizing a specific 8 nucleotide DNA recognition sequence (e.g., AGGTACTC), an NI repeat gene is selected from set 1, an NN repeat gene from set 2, an NN repeat gene from set 3, an NG repeat gene from set 4, etc. The 5' and 3' terminal regions are designed such that BsmBI digestion results in generation of 5' and 3' 4-base overhangs. Because of the unique complementarity of 3' overhangs from set 1 genes with the 5' overhang of genes from set 2, the complementarity of the 3' overhangs of set 2 with 5' overhangs from set 3, etc., the annealing and ligation of overhangs results in one, and only one, ordered alignment of the 8 repeat genes that, when translated, will specifically recognize and bind (in the present example) the AGGTACTC DNA recognition sequence. (B) Once two (or three) such blocks of repeats are constructed, they are combined in a similar ordered fashion to create a TALE repeat region that is capable of binding a specific 16 (or 24) base pair target site. (C) The assembled TALE repeat domains are cloned into the TALEN repeat-deficient pAvrXa7-FN scaffold to create a candidate dTALEN.

Repeat	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
U3a-L	NI	NG	NI	NI	NN	NN	NI	NI	HD	NN	NG	NN	HD	NG	NN	HD	N*						
U3a-R	NI	NG	NG	NI	NI	NI	NG	NI	NN	HD	NG	NG	NN	NN	HD	NI	NN	HD	N*				
U3b-L	NI	NG	NN	NI	HD	NI	HD	HD	HD	NN	NN	NG	NN	NG	NN	NN	N*						
U3b-R	NI	HD	NG	NN	NG	NG	NN	NI	NN	HD	HD	NI	NI	NG	NN	HD	N*						
L2a-L	NN	HD	NI	NG	NI	NG	HD	HD	HD	HD	HD	NI	NN	HD	HD	NI	NN	NI	HD	NI	NI	NI	N*
L2a-R	NI	NI	NG	NI	NI	HD	NI	NI	NG	HD	NI	NI	HD	HD	HD	NI	HD	NN	NG	NN	NN	NG	N*
L2b-L	NG	HD	HD	NI	NI	HD	HD	HD	NI	NI	HD	HD	HD	NG	NI	NG	HD	NG	NG	NG	HD	NI	
L2b-R	NN	NG	HD	NG	NI	HD	HD	NI	NI	NN	NI	NI	HD	NI	HD	HD	HD	NG	NG	NI	NN	NN	N*
A2-L	HD	HD	NN	HD	NI	NI	NN	HD	NI	NI	NN	HD	NN	NG	NN	NN	NI	NI	NG	NG	NI	NI	N*
A2-R	NI	HD	HD	NG	NN	NN	HD	NI	NI	NN	NG	NN	NI	NN	HD	NI	NN	HD	HD	HD	HD	NI	N*

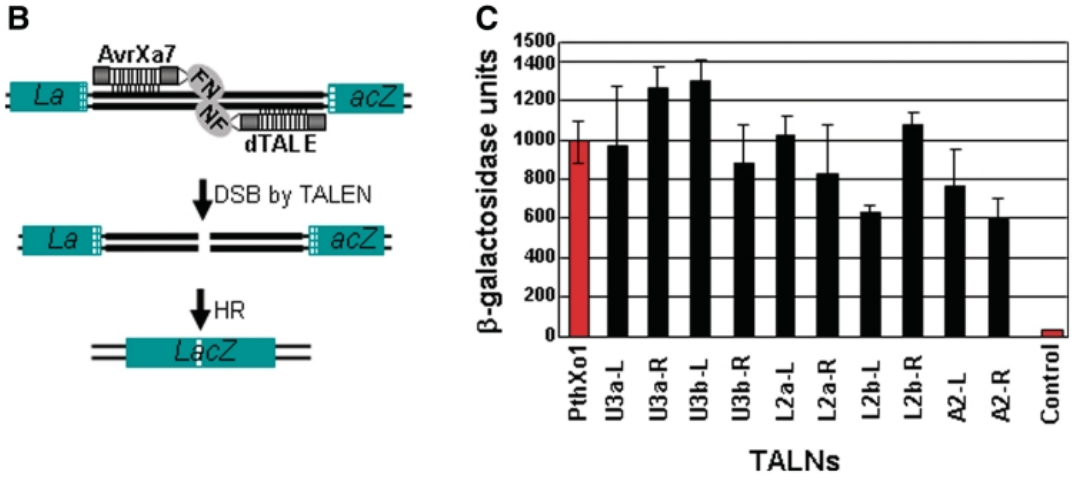


Figure 3.3. Yeast SSA assay of modularly assembled dTALENs targeting three endogenous yeast genes. **(A)** RVD sequences within repeat modules of ten custom-synthesized dTALENs. N* represents dTALEN repeat modules with the 13th amino acid missing. **(B)** Schematic of the yeast SSA assay for measuring dTALEN activity based on plasmid-borne HR. Individual candidate dTALENs are assayed in combination with AvrXa7-FN for their ability to stimulate the recombination between the duplicated regions of *LacZ* gene (hatched boxes), leading to formation of a functional *lacZ* gene. **(C)** Activities of individual dTALENs in creating DSBs as detected in a β-galactosidase assay. Control denotes the β-galactosidase activity (<5 units) of yeast cells lacking a functional TALEN gene. Error bars denote s.d.; n=3.

A

Percent gene modification by NHEJ

Target Gene	Knockout Phenotype	AvrXa7 PthXo1	Zif268 BCR-ABL	U3a-L U3a-R	U3b-L U3b-R	L2a-L L2a-R	L2b-L L2b-R	A2-L A2-R	Empty Plasmid
URA3	5-FOA Resistant	0.9%	0.3%	5.4%	1.7%	—	—	—	$<10^{-6}$
LYS2	α-AA Resistant	—	—	—	—	0.4%	1.8%	—	$<10^{-6}$
ADE2	Pink Colonies	—	—	—	—	—	—	.15%	$<10^{-6}$

B **TALEN-generated insertions/Deletions**

```
                TATAAGGAACGTGCTGCtactcatcctagtcctgttGCTGCCAAGCTATTTAATA  URA3a
                TATAAGGAACGTGCTGCtactcatccTtagtcctgttGCTGCCAAGCTATTTAATA   +1
U3a-L           TATAAGGAACGTGCTGCtactcatGcctagtcctgttGCTGCCAAGCTATTTAATA   +1
U3a-R           TATAAGGAACGTGCTGCtactcatcctaTAgtcctgttGCTGCCAAGCTATTTAATA  +2
                TATAAGGAACGTGCTGCtactcatc-tagtcctgttGCTGCCAAGCTATTTAATA    -1
                TATAAGGAACGTGCT-----------agtcctgttGCTGCCAAGCTATTTAATA    -12
                TATAAGGAACG----------------------TGCTGCCAAGCTATTTAATA    -24

                TGCATATCCCCCAGCCAGACAAAccatttacttaggtgttgcTAAACCACGTGGGTTGATTGTTA  LYS2a
                TGCATATCCCCCAGCCAGACAAAccatttacttaggTtgttgcTAAACCACGTGGGTTGATTGTTA   +1
                TGCATATCCCCCAGCCAGACAAAccatttacttagGgtgttgcTAAACCACGTGGGTTGATTGTTA   +1
L2a-L           TGCATATCCCCCAGCCAGACAAAccatttacttaggtGTgttgcTAAACCACGTGGGTTGATTGTTA  +2
L2a-R           TGCATATCCCCCAGCCAGACAAAccatttacttag-tgttgcTAAACCACGTGGGTTGATTGTTA    -1
                TGCATATCCCCCAGCCAGACAAAccatttacttagg-----cTAAACCACGTGGGTTGATTGTTA    -7
                TGCATATCCCCCAGCCAGACAAAccattta----------CTAAACCACGTGGGTTGATTGTTA    -11

                TCCGCAAGCAAGCGTGGAATTAAaacaattatcgctggagctggTGGGGCTGCTCACTTGCCAGGTA  ADE2
                TCCGCAAGCAAGCGTGGAATTAAaacaattatcgctGggagctggTGGGGCTGCTCACTTGCCAGGTA  +1
                TCCGCAAGCAAGCGTGGAATTAAaacaattatcGCgctggagctggTGGGGCTGCTCACTTGCCAGGTA +2
                TCCGCAAGCAAGCGTGGAATTAAaacaattatcgct-gagctggTGGGGCTGCTCACTTGCCAGGTA   -1
                TCCGCAAGCAAGCGTGGAATTAAaacaattat--ctggagctggTGGGGCTGCTCACTTGCCAGGTA   -2
A2-L            TCCGCAAGCAAG--------------------------------TGGGGCTGCTCACTTGCCAGGTA  -32
A2-R            TCCGCAAGCAAGCGTG----------------------------CTGCTCACTTGCCAGGTA     -33
                TCCGCAAGCAAGCGT-------------------------------------------------     -57
                TCCGCAAGCAAGCGTG------------------------------------------------     -60
                TCCGCAAGCAAGCGT-------------------------------------------------     -75
```

C **Percent gene modification by HR**

Donor Template	HR Phenotype	U3a-L U3a-R	U3b-L U3b-R	AvrXa7 PthXo1	Empty Plasmid
DURA3	5-FOA Resistant	**27%**	**4.5%**	8.5%	0.01%
NPTII	G418 Resistant	**34%**	**9%**	16%	0.1%

Figure 3.4. dTALEN induced gene modifications by NHEJ and HR. **(A)** The frequency of gene disruption induced by five sets of paired dTALENs at five specific gene target site as measured by the numbers of colonies with the indicated mutant phenotypes. "—" denotes not applicable. **(B)** TALEN-induced insertion/deletion mutations at the three of five gene loci tested. Genomic sequences from each mutant clone at the relevant loci are aligned with the respective wild type sequences. dTALEN target sites are underlined. The number of nucleotides inserted (bold uppercase letters) or deleted (dashes) is indicated to the right of each sequence. **(C)** TALEN-induced HR as measured by the percentage of yeast colonies displaying the indicated phenotypes.

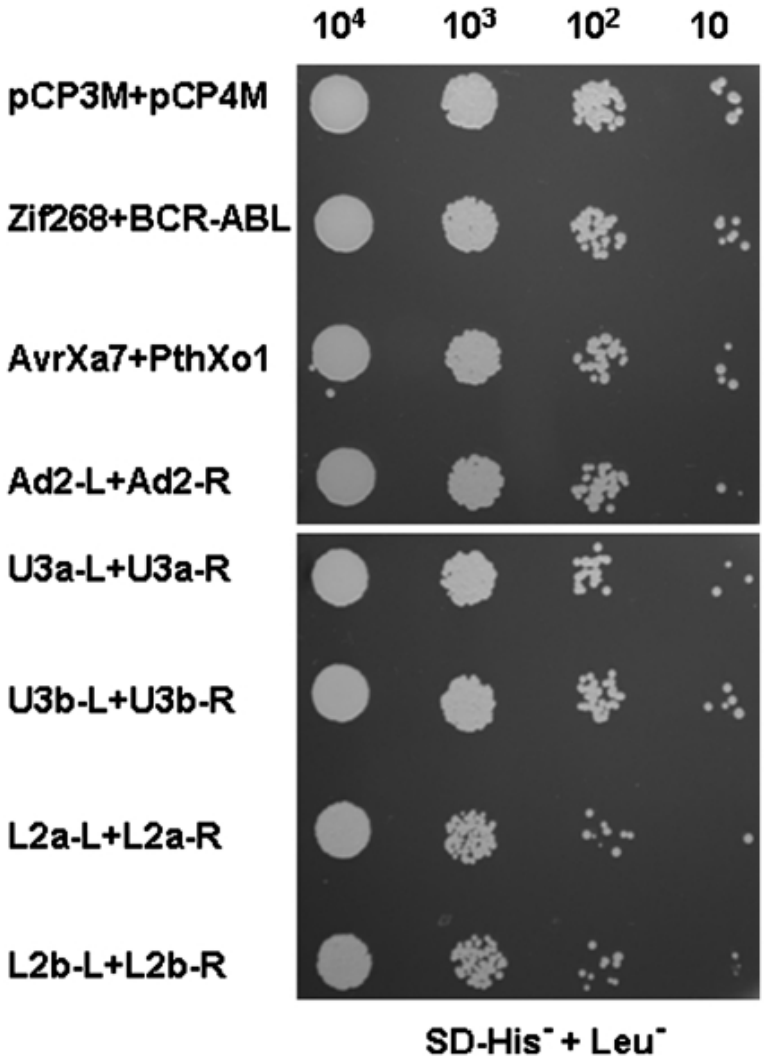


Figure 3.5. Growth of yeast cells expressing the paired TALEN and paired ZFN genes. Strains containing plasmids encoding the indicated nuclease genes (on the left side of each row) or lack thereof (pCP3M & pCP4M) were serially diluted (from 10^4 to 10 cells), applied as spots to SC medium plates lacking histidine (for pCP3M and its derived plasmids) and leucine (for pCP4M and its derived plasmids) and allowed to grow for four days.

CHAPTER 4

HIGH EFFICIENCY TALEN-BASED GENE EDITING FOR PRODUCING DISEASE RESISTANCE IN CROP PLANTS

A paper published in *Nature Biotechnology*‡‡

Ting Li[1], Bo Liu[1], Martin H. Spalding[1], Donald P. Weeks[2], Bing Yang[1 *]

[1]Department of Genetics, Development & Cell Biology, Iowa State University, IA 50011, USA; [2]Department of Biochemistry, University of Nebraska-Lincoln, NE 68588 , USA.

Abstract

We report the use of TALEN technology to thwart the ability of *Xanthomonas oryzae* pv. *oryzae* to cause bacterial blight, a major disease with significant impact on rice production worldwide. Strong disease resistance was engineered using two independent designer TALEN pairs to destroy the DNA target sites for binding of the bacterial TAL effectors, AvrXa7 and PthXo3, within the promoter region of the rice *Os11N3* gene. Genetic crosses were used to remove the TALEN genes and produce rice plants lacking transgenes or other foreign DNA elements. These studies provide a demonstration of the potential power of TALEN technology for improvement of agricultural crops.

---

‡ Reprinted with permission of Nature Biotechnology, 2012, **30**, 390–392.
* To whom correspondence should be addressed. E-mail: byang@iastate.edu

Introduction

Transcription activator-like effectors (TALEs) of *Xanthomonas oryzae* pv. *oryzae* (Xoo) contribute to pathogen virulence by transcriptionally activating specific rice disease susceptibility (S) genes [1,2]. TALE nucleases (TALENs), fusion proteins derived from the DNA recognition repeats of native or customized TALEs and the DNA cleavage domains of FokI [3-5], have been used to create site-specific gene modifications in plant cells [6,7], yeast [8], animals [9-12] and even human pluripotent cells [13]. Here we exploit TALEN technology to edit a specific S gene in rice to thwart the virulence strategy of *Xanthomonas oryzae* and, thereby, engineer heritable genome modifications for resistance to bacterial blight, a devastating disease in a crop that feeds half of the world's population.

Results

We targeted the rice bacterial blight susceptibility gene *Os11N3* (also called *OsSWEET14*) for TALEN-based disruption. This rice gene encodes a member of the SWEET sucrose efflux transporter family and is hijacked by *Xanthomonas oryzae* pv. *oryzae* using its TAL effectors AvrXa7 or PthXo3 to activate the gene and thus divert sugars from the plant cell to satisfy its nutritional needs and enhance its persistence [2,14]. The *Os11N3* promoter contains an effector binding element (EBE) for AvrXa7, overlapping with another EBE for PthXo3 and the TATA box (**Fig. 4.1a and Supplementary Fig. 4.1**). Two pairs of designer TALENs (Pair 1 and Pair 2) were independently deployed to induce mutations in these overlapping EBEs of the *Os11N3* promoter and thus to interfere with the virulence function of AvrXa7 and PthXo3 but not the developmental function of *Os11N3* (**Supplementary Fig. 4.1**). The TALE repetitive regions used for nuclease fusions included

the native AvrXa7 and three designer TALE repetitive regions custom-synthesized using a modular assembly method [8]. Each designer TALEN contained 24 repeat units for recognition of a specific set of 24 contiguous nucleotides at the target sites (**Supplementary Fig. 4.1**).

For each pair of TALEN genes, one TALEN gene (half of the pair) was under the control of the 35S promoter of cauliflower mosaic virus and the other gene was driven by the maize ubiquitin 1 promoter, comprising a specific TALEN pair in a single plasmid (**Supplementary Fig. 4.2**). Each plasmid also contained a marker gene for hygromycin resistance. These constructs were introduced into rice embryonic cells using *Agrobacterium tumefaciens*, and individual transformant cells were selected, propagated and regenerated into whole plants (T0). The *Os11N3* promoter regions from a number of independent hygromycin resistant callus lines and the segregating progeny (T1) of self-pollinated T0 plants were PCR-amplified and sequenced to detect potential sequence alterations. For TALEN Pair 1 genes, two of five examined callus lines contained biallelic mutations (**Supplementary Fig. 4.3**). Of twenty three randomly selected T1 progeny produced from self-pollination of seven independent T0 plants transformed with TALEN Pair 1 genes, about half (48%) carried mono- or bi-allelic mutations (including the four mutations detected in the two previously examined callus lines) (**Fig. 4.1b**). Approximately two-thirds (63%) of the randomly selected T1 plants (n=30) generated from self-pollination of sixty-six independent T0 plants from the two independent transformation experiments carried mutations that were induced by the TALEN Pair 2 genes (**Fig. 4.1c** and **Fig. 4.1d**). In total, sixteen distinct mutations, including 6 homozygous lines, were detected in fifty-three T1 plants from TALEN Pair 1 and Pair 2. The majority of these mutations were small deletions that left the TATA box intact, with the exception of two deletions in heterozygous lines containing a wild type allele (**Fig. 4.1e** and

Supplementary Fig. 4.4). Bacterial infection assays using leaf tip clipping method on other T1 plants (n=627) generated from TALEN Pair 2 (Experiment 1) and not previously genotyped demonstrated that approximately 48% of the treated plants exhibited resistance to infection by pathogenic Xoo in terms of length of leaf lesions (1~4 cm for resistance vs. 10 ~14 cm for susceptibility) (**Supplementary Fig. 4.5**). DNA sequence analyses of 27 such Xoo-resistant T1 plants confirmed the presence of homozygous monoallelic or heterozygous biallelic EBE mutations and revealed 17 additional, distinct mutant haplotypes (**Supplementary Fig. 4.6**). All mutant plants were morphologically normal compared to wild type plants, indicating that the developmental function of *Os11N3* was not disrupted.

Forty plants from the second generation (T2) of three self-pollinated T1 plants were also genotyped by sequencing to determine the heritability of 3 TALEN generated mutations, all of which, whether homozygous or heterozygous, were passed on to T2 plants (**Supplementary Fig. 4.7**).

To determine the effects of TALEN-directed mutations, we investigated whether the pathogenic strain of Xoo that is dependent on AvrXa7 or PthXo3 for virulence is able to either induce the modified *Os11N3* gene in T2 homozygous plants or cause disease. The modified *Os11N3* gene was no longer inducible by AvrXa7 or PthXo3 delivered by the pathogenic strain of the bacterium [ME2(*avrXa7*) or ME2(*pthXo3*)] in T2 plants homozygous for either the 9, 6, 15, or 4 bp deletion (**Fig. 4.1f** for AvrXa7**, Supplementary Fig. 4.8a** for PthXo3). The loss of induction was specific to *Os11N3*, as the induction of Os04g19960, a transposon coding gene collaterally targeted by AvrXa7, was not prevented (**Fig. 4.1g**). Similarly, the induction of another S gene (*Os8N3*/*OsSWEET11*) by PthXo1 in the T2 mutant plants remained unaffected (**Supplementary Fig. 4.8b**). These TALEN modified T2 plants

also showed strong resistance to infection of AvrXa7- or PthXo3-dependent Xoo strain but not PthXo1-dependent pathogenic Xoo strain as determined by symptoms (**Fig. 4.1h** for AvrXa7) and by quantitative measurement of the lengths of leaf lesions in a standard pathogenesis assay described in Supplemental Methods and Materials (**Supplementary Fig. 4.9**).

We also investigated the possibility of using genetic segregation to obtain genetically modified rice lacking any selection marker and TALEN gene. The polymerase chain reaction (PCR) assay using primers for amplification of the hygromycin resistance gene and for amplification of the TALEN genes failed to detect the presence of either gene in 5 out of 37 T1 plants that contained the desired genetic modifications in the *Os11N3* promoter and that were disease resistant (**Supplementary Fig. 4.10**). Although these data clearly demonstrate the absence of intact TALEN and hygromycin resistance genes, further sequencing the genomes of several mutants and the Kitake parental line will be needed to conclusively demonstrate that all of the transgene fragments have been removed.

Discussion

The rice *Os11N3* gene is induced by 32 of 40 Xoo strains collected worldwide (manuscript in preparation). However, polymorphisms in the *Os11N3* gene that prevent induction by AvrXa7- and/or PthXo3-dependent Xoo strains and also confer disease resistance have not been identified in rice germplasm. The approaches described here for precisely and efficiently editing the disease susceptibility elements in *Os11N3* and for the subsequent removal of transfer DNA (T-DNA) sequences by classical genetics likely can be applied directly to elite rice varieties to simultaneously or sequentially edit multiple

susceptibility genes (e.g. *Os11N3* and *Os8N3*), leading to resistance to the major forms of bacterial blight. Present methods using TALEN-based technology in rice should be easily modified for application to other plant species and, thus, holds significant promise in facilitating gene modification-based research and crop improvement.

Materials and Methods

TALEN design and construction.

Two pairs of designer TALENs (Pair 1: dTALEN L1 and dTALEN R1 and Pair 2: TALEN AvrXa7 and dTALEN R2) were used in the present studies. TALEN AvrXa7 was based on a native AvrXa7 TAL effector and contained a truncated transcription activation domain fused at its C-terminus with a wild type FokI DNA cleavage domain [3,8]. The other three TALENs employed in these studies were made using previously described "modular assembly" methods[8]. Such assembly used four different genes encoding TAL effector DNA binding domains that, due to specific codons in position 12 and 13 of the coding sequence, are capable, respectively, of recognizing either an A, G, C or T residue in the DNA sequence of a particular target effector binding element (EBE). Each designer TALEN contained a wild type FokI DNA cleavage domain [3,8]. All TALENs used a complete TAL effector N-terminus. Diagrams showing the structure of each TALEN and the EBE DNA sequence to which it matches are shown in **Supplementary Figure 4.1**.

Construction of TALEN expression plasmids and rice transformation.

The two promoters used to express the paired TALEN genes were the maize ubiquitin 1 promoter (ubi1) and the 35S promoter of cauliflower mosaic virus (35S) [15,16]. One of the

paired TALEN genes was cloned downstream of the 35S promoter at BamHI and SpeI sites in a binary vector and the other under the control of the *ubi1* promoter was cloned into the BamHI and SacI sites of an intermediate vector. The ubi1-TALEN gene expression cassette was excised with HindIII and moved into the HindIII site of the binary plasmid containing the 35S-TALEN gene expression cassette. The resultant plasmids were mobilized into *Agrobacterium tumefaciens* strain EHA105 by electroporation. *Agrobacterium*-mediated transformation of the rice cultivar Kitake was conducted according to a previously described protocol [17].

DNA sequencing analysis of regions in the Os11N3 gene targeted by TALENs.

Genomic DNA from individual plants was extracted using the CTAB method as described [18]. Forward primer, 5'- TCCCTTAACTAGGACAACTTGGA-3', and reverse primer, 5'- CCGGATCCAGCCATTGCAGCAAGATCTTG-3', were used to amplify a region of ~ 550 bp with the preselected target sites located in the middle. The PCR products from individual plants were sequenced using an internal primer, 5'-CATGGCTGTGATTGATCAGG-3'. Each sequencing chromatogram was manually analyzed for polymorphisms within a trace.

Quantitative RT-PCR analysis of rice gene inducibility.

Bacterial inoculums with optical density of 1.0 at 600 nm (OD_{600}) were infiltrated into rice leaves by using needle-less syringe as described [18]. The bacterial strains used were Xoo strain PXO99ME2 (hereafter designated as ME2, a PXO99 derivative strain lacking TAL effector PthXo1 with concomitant loss of strain virulence), ME2(*avrXa7*),

ME2(*pthXo3*) and ME2(*pthXo1*), three ME2 transformants containing the respective TAL effector genes *avrXa7*, *pthXo3* and *pthXo1*. Both ME2(*avrXa7*) and ME2(*pthXo3*) induce *Os11N3* and retain virulence, while ME2(*pthXo1*) retains virulence by inducing another S gene (*Os8N3*) [1]. Total RNA of the inoculated portion of leaves was extracted using TRI reagent from Ambion 24 hours after bacterial inoculation. RNA concentration and quality were determined using an ND-1000 Nanodrop spectrophotometer (Nanodrop Technologies). One microgram of RNA from each sample was treated with DNase 1 (Invitrogen) followed by cDNA synthesis using the iScript cDNA synthesis kit (Bio-Rad). cDNA derived from 25 ng of total RNA was used for detection of gene induction by AvrXa7, PthXo3 and PthXo1 using real-time quantitative PCR analyses. PCR was performed on Stratagene's Mx4000 multiplex quantitative PCR system using the iQ SYBR Green Supermix kit (Bio-Rad). In addition to inducing *Os11N3*, AvrXa7 also "collaterally" induces another rice gene Os04g19960, which encodes a putative retrotransposon protein, but is not associated with disease susceptibility in rice. Gene-specific primers for *Os11N3* are 5'-GAGAAGAAGGTAGCTGCATGAGTG-3' and 5'-TCATGGAAGGAACCCTTACAGGTTG-3', primers for Os04g19960 are 5'-AGAAGGCGTAGGCATTCACAT-3' and 5'- ACATTAACACAGCACACGTCAAC-3', and primers for *Os8N3* are 5'-GACTCCATGTCCCCGATCTCC-3' and 5'-CACCACCTCGACCTTGTGCA-3'. The rice general transcription factor *TFIIAγ5* expression was used as an internal control with primers 5'-CTACTCAGCCAATAAATTGATAACTGC-3'and 5'-CAATTTCTACTACTCATCGTTTAG-3'. The average threshold cycle (Ct) was used to

determine the fold change of gene expression. The $2^{\Delta\Delta Ct}$ method was used for relative quantification [19].

PCR determination of the presence or absence of T-DNA in genetically modified rice plants.

Genomic DNA was extracted as described [18] and used for PCR amplification of fragments from the *Os11N3* gene promoter, hygromycin phosphotransferase (*hpt*) gene and TALEN genes. Primers for the *Os11N3* gene were: forward, 5'-CATGGCTGTGATTGATCAGG-3' and reverse, 5'-CCGGATCCAGCCATTGCAGCAAGATCTTG-3'; primers for the *hpt* gene were: Hyg-F, 5'- CCGCTCGTCTGGCTAAGATC-3' and 35S-R, 5'- CGCTGAAATCACCAGTCTCTC-3'; and primers for the TALEN genes were: FokI-F, 5'-CAGCTAGTGAAATCTGAATTGG-3' and Nos-R, 5'-CATCGCAAGACCGGCAACAGG-3'.

Surveyor nuclease cleavage assay for detection of nucleotide insertions and deletions.

Genomic DNA was extracted using the CTAB method as described[18] from individual callus lines and subjected to PCR amplification of the *Os11N3* promoter region (~ 550 bp) using the gene-specific primers (5'- TCCCTTAACTAGGACAACTTGGA-3' and 5'-CCGGATCCAGCCATTGCAGCAAGATCTTG-3'). The Surveyor nuclease (Surveyor mutation detection kit, Transgenomic) was used to treat the PCR products following the manufacturer's instruction. The treated DNA was subjected to electrophoresis in a 1.5% agarose gel and visualized by staining with ethidium bromide.

Disease resistance assay.

Fully expanded leaves of rice plants were inoculated using leaf tip clipping method. In this previously described procedure [20], scissors blades are immersed in bacterial suspension (OD_{600}=0.5) of *avrXa7*-containing strain PXO86, *pthXo3*-containing strain ME2(*pthXo3*) and *pthXo1*-containing strain PXO99 immediately prior to clipping each target leaf. Symptoms were scored 12-14 days after inoculation by measuring lesion length. Plants were categorized as resistant (R) if lesion lengths were shorter than 4 cm or susceptible if lesions were longer than 8 cm.

Acknowledgements

We thank David Wright and Changfu Yao for helpful suggestions, to Lingui Xue and Ji Luo for technical assistance, and to Frank F. White and Stephen Howell for critical reading of the manuscript. This work was supported by grants from the National Science Foundation (0820831 to B.Y., MCB-0952323 to M.H.S., MCB-0952533 and EPSCoR grant 1004094 to D.P.W.).

References

1. Yang, B, A Sugio, FF White. 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.

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

3. Li,T., Huang,S., Jiang,W.Z., Wright,D., Spalding,M.H., Weeks,D.P. and Yang,B. 2010. TAL nucleases (TALNs): Hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucleic Acids Res., 39: 359-372.

4. 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.

5. Miller,J.C., Tan,S., Qiao,G., Barlow,K.A., Wang,J., Xia,D.F., Meng,X., Paschon,D.E., Leung,E., Hinkley,S.J., et al. 2011. A TALE nuclease architecture for efficient genome editing. Nat. Biotechnol., 29: 143-148.

6. Mahfouz,M.M., Li,L., Shamimuzzaman,M., Wibowo,A., Fang,X. and Zhu,J.K. 2011. De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. Proc. Natl. Acad. Sci. U. S. A., 108: 2623-2628.

7. Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF. 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res. 39(12):e82.

8. Li T, Huang S, Zhao X, Wright DA, Carpenter S, Spalding MH, Weeks DP, Yang B. Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. Nucleic Acids Res 2011; 39:6315-6325.

9. Sander JD, Cade L, Khayter C, Reyon D, Peterson RT, Joung JK, Yeh JR. 2011. Targeted gene disruption in somatic zebrafish cells using engineered TALENs. Nat Biotechnol. 5;29(8):697-8.

10. Huang P, Xiao A, Zhou M, Zhu Z, Lin S, Zhang B. 2011. Heritable gene targeting in zebrafish using customized TALENs. Nat Biotechnol. 29(8):699-700.

11. Tesson L, Usal C, Ménoret S, Leung E, Niles BJ, Remy S, Santiago Y, Vincent AI, Meng X, Zhang L, Gregory PD, Anegon I, Cost GJ. 2011. Knockout rats generated by embryo microinjection of TALENs. Nat Biotechnol. 29(8):695-6.

12. Wood AJ, Lo TW, Zeitler B, Pickle CS, Ralston EJ, Lee AH, Amora R, Miller JC, Leung E, Meng X, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Meyer BJ. 2011. Targeted genome editing across species using ZFNs and TALENs Science. 333(6040):307.

13. Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassady JP, Cost GJ, Zhang L, Santiago Y, Miller JC, Zeitler B, Cherone JM, Meng X, Hinkley SJ, Rebar EJ, Gregory PD, Urnov FD, Jaenisch R. 2011. Genetic engineering of human pluripotent cells using TALE nucleases. Nat Biotechnol. 29(8):731-4.

14. Chen LQ, Qu XQ, Hou BH, Sosso D, Osorio S, Fernie AR, Frommer WB. 2012. Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. Science 335:207-211.

15. Christensen AH, Quail PH. 1996. Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. Transgenic Res. 5(3):213-8.

16. Odell JT, Nagy F, Chua NH. 1985. Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. Nature. 6;313(6005):810-2.

17. Hiei Y, Ohta S, Komari T, Kumashiro T. 1994. Efficient transformation of rice (Oryza sativa L.) mediated by Agrobacterium and sequence analysis of the boundaries of the T-DNA. Plant J. 6(2):271-82.

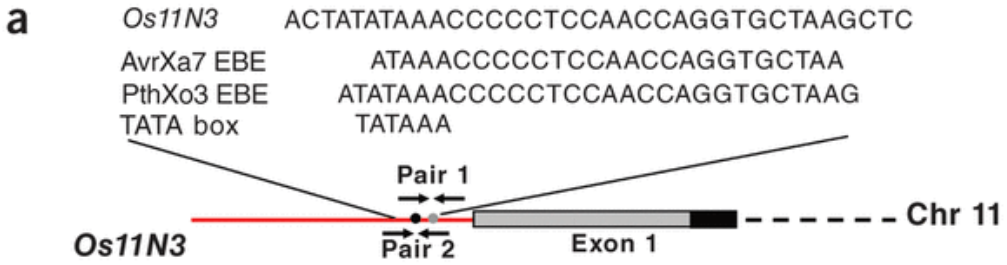
18. Edwards K, Johnstone C, Thompson C. 1991. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res. 19(6):1349.

19. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 25(4):402-8.

20. Kauffman HE, Reddy APK, Hsieh SPY, Merca SD. 1973. An improved technique for evaluating resistance of rice varieties to *Xanthomonas oryzae*. Plant Disease Rep. 57, 537-541.

Figures and Tables

a



b

Pair 1 Exp. 1 Number of T_1 plants	Total	Genotype: wt	−49/−49	−49/−8	−5a/−6a	−4a/+2	Unique mutations
	23	12	5	4	1	1	6

c

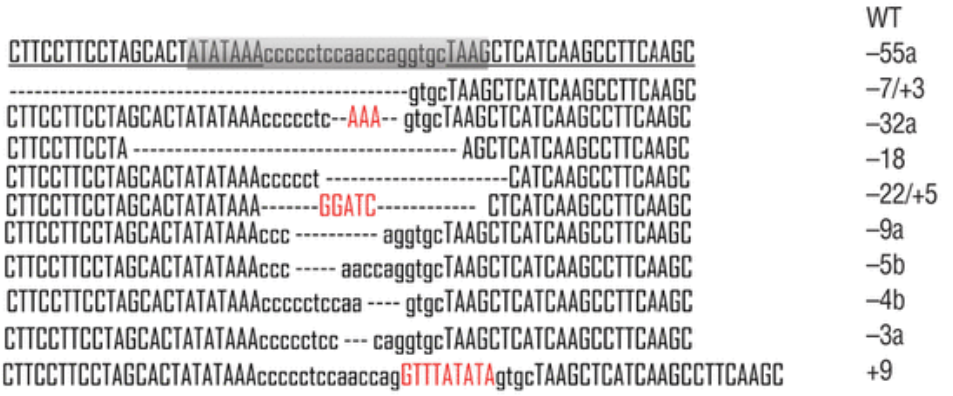
Pair 2 Exp. 1 Number of T_1 plants	Total	Genotype: wt	+9/+9	−9a/−9a	−5b/−5b	−4b/−4b	−55a/(−7/+3)	−9a/wt	UD*	Unique mutations
	15	5	1	1	1	1	1	1	4	6

d

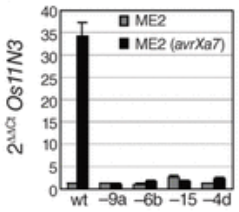
Pair 2 Exp. 2 Number of T_1 plants	Total	Genotype: wt	−18/−18	(−22/+5)/−5b	(−22/+5)/wt	−32a/wt	−3a/wt	UD*	Unique mutations
	15	6	1	1	3	1	1	2	5

UD*, undermined (impossible to unambiguously determine DNA sequence).

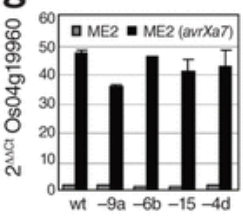
e



f



g



h

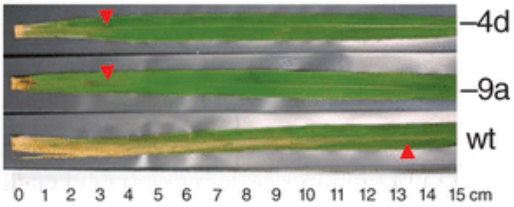
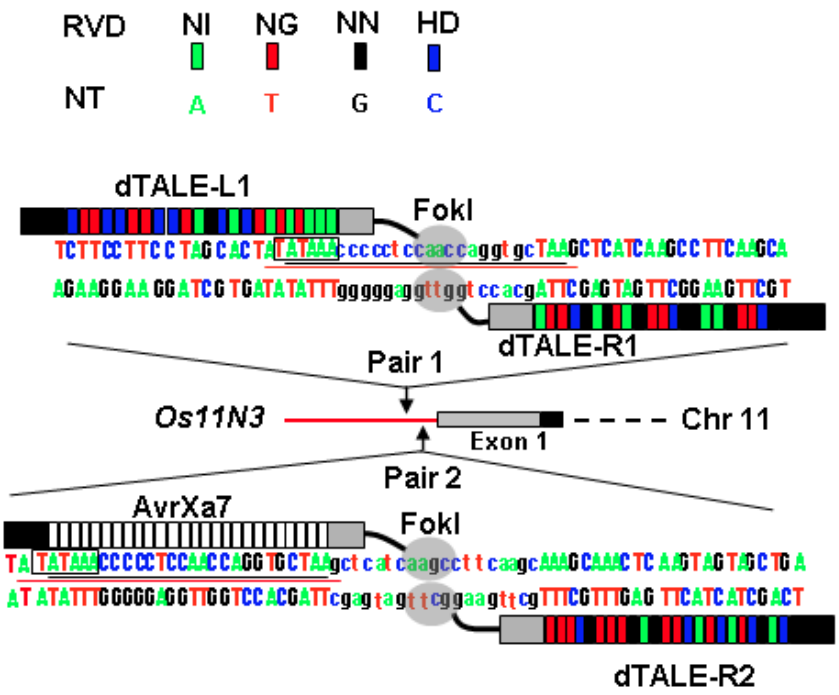
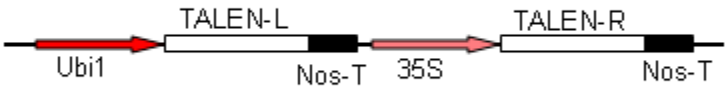


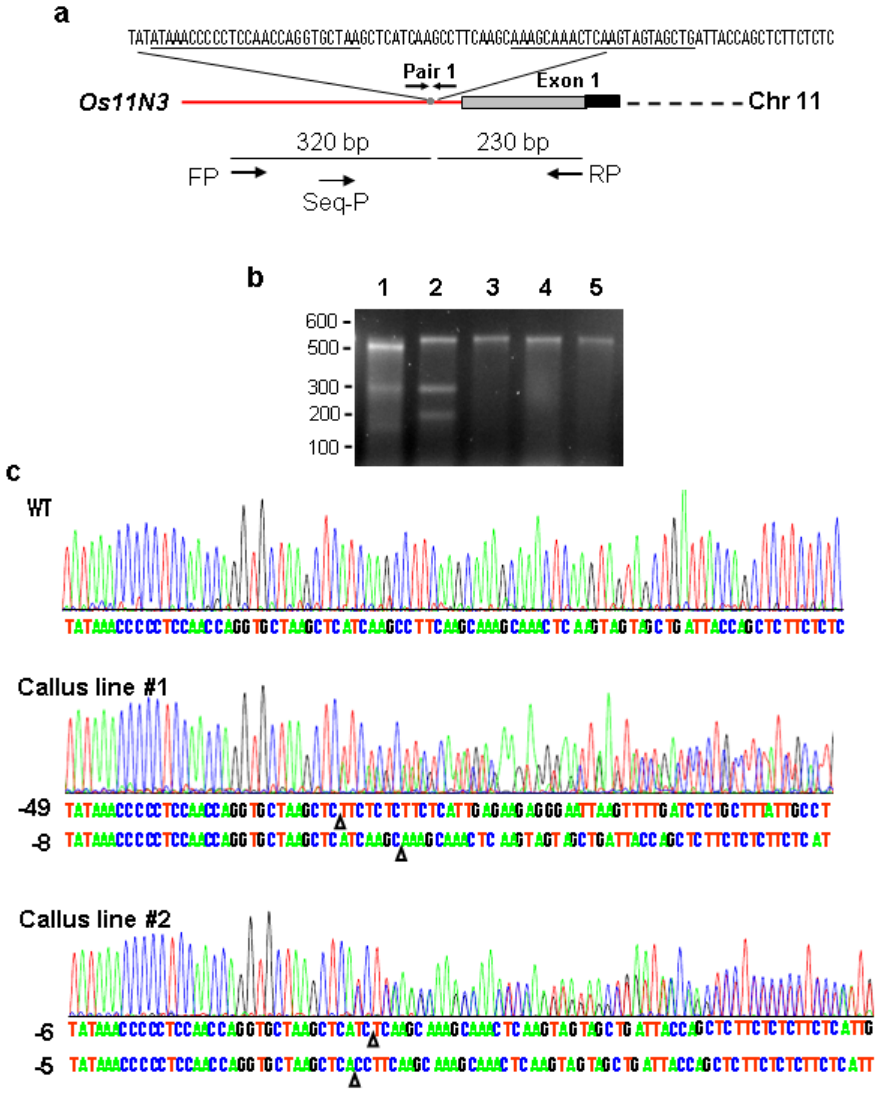
Fig. 4.1. High efficiency targeted gene editing using TALENs. (**a**) Overlapping elements targeted by two pairs (Pair 1 and 2) of designer TALENs in the *Os11N3* promoter. (**b**) - (**d**) Genotypes of progeny (T1) of primary transgenic plants (T0) derived from TALEN-expressing embryonic cells from three independent transformation experiments (Exp.). Each of the two alleles of an individual plant are designated as being wild type (wt), having a nucleotide insertion (+) or a deletion (-) and are separated top and bottom by a dividing line. The designation [-55/(-7/+3)] indicates that one allele contains a deletion of 55 bp and that the other allele has both a deletion of 7 bp and an insertion of 3 bp. (**e**) Sequences of *Os11N3* mutations induced by the Pair 2 TALENs with deletions (dashes) and insertions (red letters). TALEN-binding sequences are underlined in WT and the overlapping EBEs shaded in grey. (**f**), (**g**) Expression of *Os11N3* and Os04g19960 induced by AvrXa7 in plants of different genotypes. Quantitative RT-PCR was performed with RNA derived from treatments of nonpathogenic Xoo strain ME2 and pathogenic ME2(*avrXa7*). $2^{\Delta\Delta Ct}$ is a measure of transcript abundance for a selected gene [*Os11N3* (f) or *Os04g19960* (g)] relative to transcript abundance produced from a constitutively expressed gene (*OsTFIIAγ5*) as determined by relative PCR cycle thresholds (C_t). (**h**) Resistance phenotype displayed by two T2 mutant plants compared with the disease susceptibility phenotype of a nontransgenic wild type rice plant.



Supplementary Figure 4.1. TALENs and their DNA targets in the promoter of chromosomal *Os11N3* gene. Four basic modular repeats whose repeat variable di-residue (RVD) (colored blocks) recognizing one nucleotide (NT) in the target site are used to assemble the DNA binding domain of each designer TALEN. The two pairs of nucleases (Pair 1 and 2) are fusions between the DNA cleavage domain of FokI (FokI) and the native (AvrXa7) or customized TAL effector (dTALE). The last 40 amino acids at C-terminus of dTALE-L1 and AvrXa7 are truncated to avoid the inappropriate induction of *Os11N3* by the activation domain. The other two TAL effectors (dTALE-R1 and dTALE-R2) contain the complete C-terminus. All four TAL effectors contain the complete N-terminus. *Os11N3* promoter contains an effector binding element (EBE) for AvrXa7 (underlined in black), an EBE for PthXo3 (underlined in red) and the TATA box (boxed). Lower letters represent regions wherein two FokI domains dimerize and cause a double stranded DNA break.

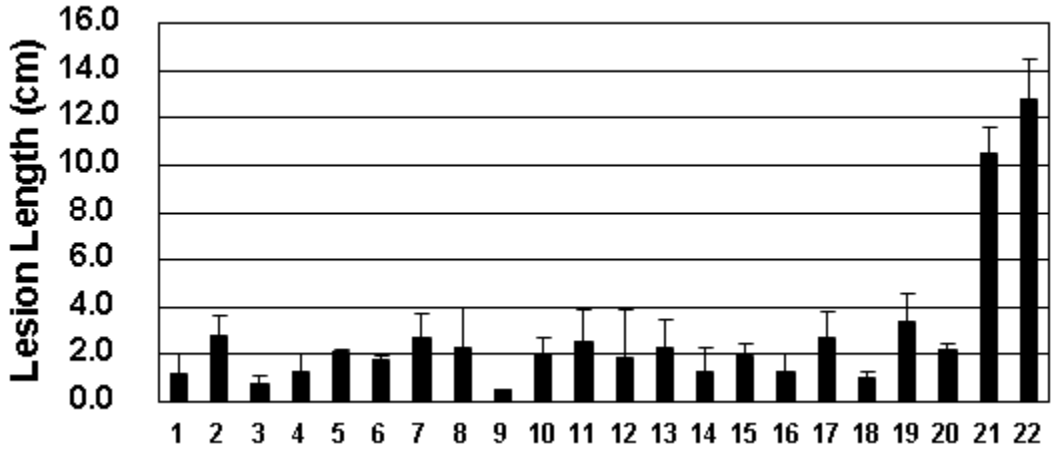


Supplementary Figure 4.2. Schematic diagram of a two-gene expression cassette in a single binary vector designed for *Agrobacterium*-mediated rice transformation. The expression cassette includes a promoter [maize ubiquitin 1 promoter (Ubi1) (red arrow) to drive expression of the TALEN-L gene (open box), the cauliflower mosaic virus 35S gene promoter (35S) (red arrow) to allow transcription of the TALEN-R gene (open box)] and a gene terminator (Nos-T) (black box).


a
TATATAAACCCCCTCCAACCAGGTGCTAAGCTCATCAAGCCTTCAAGCAAAGCAAACTCAAGTAGTAGCTGATTACCAGCTCTTCTCTC
Pair 1
Exon 1
Os11N3
Chr 11
320 bp
230 bp
FP
RP
Seq-P
b
1 2 3 4 5
600
500
300
200
100
c
WT
TATAAACCCCCTCCAACCAGGTGCTAAGCTCATCAAGCCTTCAAGCAAAGCAAACTCAAGTAGTAGCTGATTACCAGCTCTTCTCTC
Callus line #1
-49
TATAAACCCCCTCCAACCAGGTGCTAAGCTCTTCTCTCTTCTCATTGAGAAGAGGGAATTAAGTTTTGATCTCTGCTTTATTGCCT
-8
TATAAACCCCCTCCAACCAGGTGCTAAGCTCATCAAGCAAAGCAAACTCAAGTAGTAGCTGATTACCAGCTCTTCTCTCTTCTCAT
Callus line #2
-6
TATAAACCCCCTCCAACCAGGTGCTAAGCTCATCTCAAGCAAAGCAAACTCAAGTAGTAGCTGATTACCAGCTCTTCTCTCTTCTCATTG
-5
TATAAACCCCCTCCAACCAGGTGCTAAGCTCACCTTCAAGCAAAGCAAACTCAAGTAGTAGCTGATTACCAGCTCTTCTCTCTTCTCATT

Supplementary Figure 4.3. Analysis of site-specific mutations within the *Os11N3* gene promoter in T0 callus lines expressing the Pair 1 TALENs. (a) Schematics of the *Os11N3* gene promoter target site with primers (FP & RP) for PCR amplification of promoter sequence (550 bp) and primer (Seq-P) for sequencing the PCR products. (b) Analysis of PCR products derived from five individual callus lines using the Surveyor nuclease cleavage assay (see Methods and Materials) and showing two out of five calli (#1 and #2) with biallelic mutations. This assay is designed to detect and cleave "loop-out" regions between hybrids of wild type promoter DNA sequences and DNA sequences in promoters containing TALEN-generated nucleotide deletions or insertions. If loop-out sequences are present, cleavage by the single-strand DNA-specific Surveyor nuclease should generate two DNA fragments, one in the range of 230 bp and the other in the range of 330 bp. Molecular sizes are indicated in base pairs at the left side of the ethidium bromide stained gel image. (c) DNA sequencing chromatograms of three DNA fragments derived from wild type (WT) tissue and two callus lines (#1, #2) each containing biallelic mutations (i.e., a deletion of 49 bp in one *OS11N3* allele and a 8 bp deletion in the other alleles of the *OS*11N23 gene in callus #1; a deletion of 6 bp in one allele in the promoter of *OS11N3* in callus #2 and a 5 bp deletion in the other allele).

Supplementary Figure 4.4. Sequence of *Os11N3* gene mutations in T1 plants induced by the Pair 1 TALENs. Deletions and insertions are indicated by dashes and red letters, respectively. TALEN-binding sequences are underlined in the wild type (WT) gene sequence. Numbers and letters designating each individual mutant (with numbers reflecting the length of nucleotide deletions or insertions) are indicated to the right side of the DNA sequence.



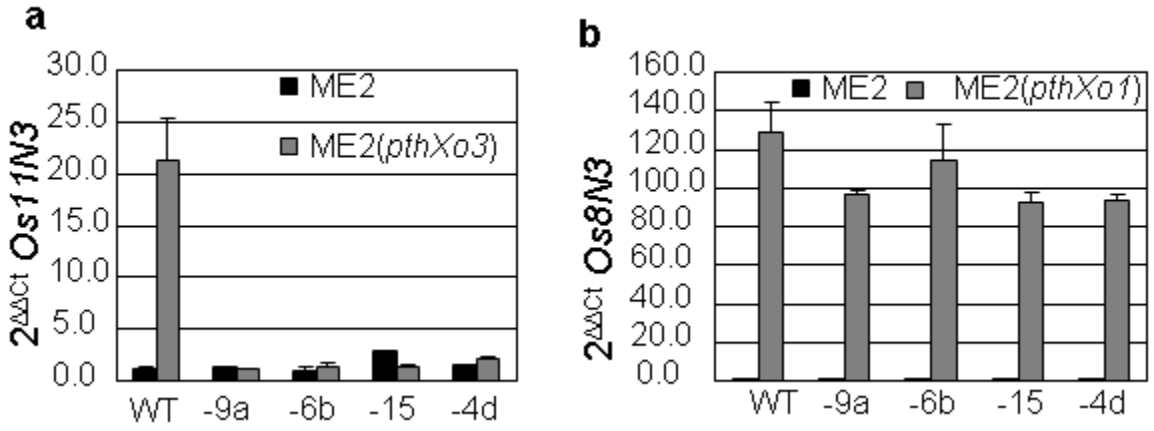
Supplementary Figure 4.5. Disease resistance in transgenic rice T1 plants. Lesion lengths caused by infection with a pathogenic AvrXa7-dependent Xoo strain were measured 14 days after inoculation of 20 T1 mutant plants (1 – 20) generated from TALEN pair 2 and 2 wild type Kitake plants (21 – 22). The T1 plants contained either homozygous monoallelic or heterozygous biallelic EBE mutations as confirmed by genotyping through sequencing. Leaf lesion lengths of 1~4 cm indicate disease resistance and lesion lengths of 10 ~14 cm indicate disease susceptibility. Error bars indicate 1 SD.

Supplementary Figure 4.6. Additional haplotypes detected in T2 plants carrying *Os11N3* gene mutations produced with Pair 2 TALENs. Deletions and insertions are indicated by dashes and red letters, respectively. Numbers and letters designating each individual mutant (with numbers reflecting the length of nucleotide deletions or insertions) are indicated to the right side of the DNA sequence. TALEN-binding sequences are underlined in wild type (wt).

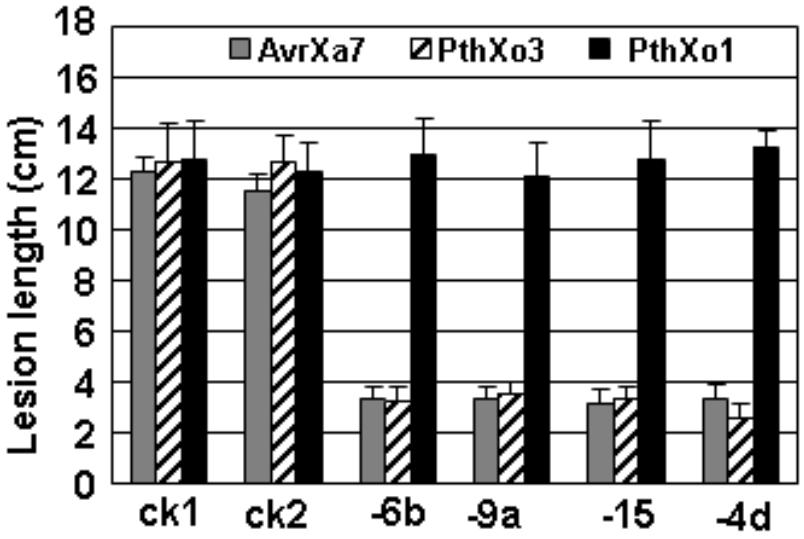
	Genotype						
T1	-9a / -9a	-4d / -57			-9a / wt		
T2	-9a / -9a	-4d / -4d	-4d / -57	-57 / -57	-9a / -9a	-9a / wt	wt / wt
Ratio*	10/10	5/16	8/16	3/16	6/14	6/14	2/14

* Ratio denotes the number of T2 plants with the genotypes specified in T2 row to total number of T2 plants genotyped by sequencing the *Os11N3* promoter region.

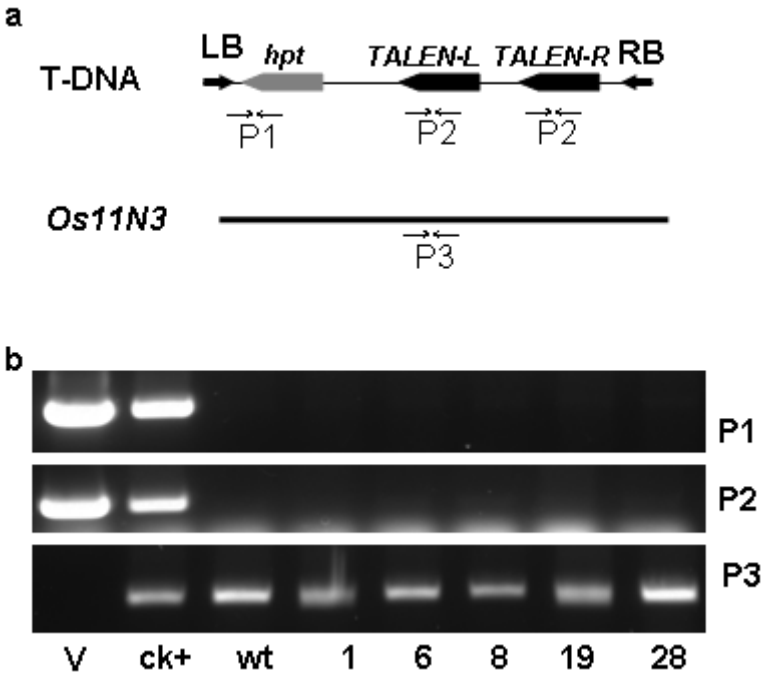
Supplementary Figure 4.7. Genetic Segregation of forty T2 plants derived from self-pollination of three T1 plants associated with the Pair 2 TALENs. Each of the two alleles of an individual plant are designated as being wild type (wt) or having a nucleotide deletion (-) and are separated top and bottom by a dividing line.



Supplementary Figure 4.8. Expression of *Os11N3* (**a**) and *Os8N3* (**b**), respectively, induced by PthXo3- and PthXo1-dependent Xoo strains in T2 plants of different genotypes (indicated below each column). Quantitative RT-PCR was performed with RNA derived from treatments of nonpathogenic Xoo strain ME2 and pathogenic Xoo strains ME2(*pthXo3*) and ME2(*pthXo1*). Transcript levels of the rice gene *OsTFIIAγ5* were used as a reference for measurements of *Os11N3* and *Os8N3* transcript levels.



Supplementary Figure 4.9. Severity of disease damage to wild type and transgenic rice plants caused by AvrXa7-, PthXo3-, and PthXo1-dependent Xoo strains. Lengths of lesions in wild type plants (CK1), segregating T2 transgenic plants with intact *Os11N3* (CK2) and T2 transgenic plants homozygous for *Os11N3* promoter mutations of 6 bp (-6b), 9 bp (-9a), 15 bp (-15) and 4 bp (-4d) deletions, respectively, were measured 14 days post inoculation with different TAL effector Xoo strains as indicated.



Supplementary Figure 4.10. Removal of T-DNA sequences containing TALEN genes from TALEN-modified rice plants using genetic crossing. (a) Schematic diagrams of the transfer DNA (T-DNA) region containing a paired set of TALENs and the *Os11N3* gene aligned with the paired primer sets P1, P2 and P3 used for specific gene segment detection by PCR amplification. LB and RB represent the left and right border sequences for *Agrobacterium*-mediated gene transfer; *hpt* represents the hygromycin resistance gene. (b) Gel images of PCR products obtained with the primer sets of P1, P2 and P3 for hygromycin phosphotransferase gene (*hpt*), TALEN genes and *Os11N3* promoter, respectively. Labels below gel images represent: v, binary vector DNA; ck+, a positive control of DNA from a transgenic plant containing the T-DNA region depicted in (a) and the *Os11N3* gene; wt, DNA from a nontransgenic, wild type rice plant; number, individual T1 plants selected from genetic crosses to lack the T-DNA region, but retain a functional *Os11N3* promoter region containing inactivated or deleted AvrXa7 and PthXo3 EBE sites.

CHAPTER 5

DESIGNER TAL EFFECTORS INDUCE DISEASE SUSCEPTIBILITY AND RESISTANCE TO *XANTHOMONAS ORYZAE* PV. ORYZAE IN RICE

A paper published by *Molecular plant*[‡]

Ting Li[a], Sheng Huang[a,b], Junhui Zhou[a], Bing Yang[a,1]

[a] Department of Genetics, Development and Cell Biology, Iowa State University, Ames, IA 50011, USA

[b] Present address: State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, College of Life Science and Technology, Guangxi University, Nanning, Guangxi 530005, China

Abstract

TAL (transcription activator-like) effectors from *Xanthomonas* bacteria activate the cognate host genes, leading to disease susceptibility or resistance dependent on the genetic context of host target genes. The modular nature and DNA recognition code of TAL effectors enable custom-engineering of designer TAL effectors (dTALE) for gene activation.

However, the feasibility of dTALEs as transcription activators for gene functional analysis has not been demonstrated. Here we report the use of dTALEs, as expressed and delivered by the pathogenic *Xanthomonas oryzae* pv. oryzae (Xoo), in revealing the new function of two previously identified disease-related genes and the potential of one

[‡]Reprinted with permission of *Molecular Plant*, 2013, 6 (3): 781-789.

[1] To whom correspondence should be addressed. Email byang@iastate.edu, tel. +1-515-294-2968, fax +1-515-294-5256

developmental gene for disease susceptibility in rice/Xoo interactions. The dTALE gene *dTALE-xa27*, designed to target the susceptible allele of the resistance gene *Xa27*, elicited a resistant reaction in the otherwise susceptible rice cultivar IR24. Four dTALE genes were made to induce the four annotated *Xa27* homologous genes in rice cultivar Nipponbare, but none of the four induced *Xa27*-like genes conferred resistance to the dTALE-containing Xoo strains. A dTALE gene was also generated to activate the recessive resistance gene *xa13*, an allele of the disease susceptibility gene *Os8N3* (also named *Xa13* or *OsSWEET11*, a member of sucrose efflux transporter SWEET gene family). The induction of *xa13* by the dTALE rendered the resistant rice IRBB13 (*xa13*/*xa13*) susceptible to Xoo. Finally, *OsSWEET12*, an as yet uncharacterized SWEET gene with no corresponding naturally occurring TAL effector identified, conferred susceptibility to the Xoo strains expressing the corresponding dTALE genes. Our results demonstrate that dTALEs can be delivered through the bacterial secretion system to activate genes of interest for functional analysis in plant.

Key Words: TAL effector; rice; *Xanthomonas*; *Xa27*; disease susceptibility; disease resistance; designer TAL effector.

Introduction

Plant pathogens of *Xanthomonas* bacteria essentially depend on a type III secretion system for pathogenesis in their host plants (Bonas et al. 1991; Zhu et al. 2000; Cho et al. 2008). The type III secretion system in *Xanthomonas* secretes a suite of effector proteins, including TALEs (transcription activator-like effectors), into host cells (White et al. 2009). Once internalized, TALEs function as transcription activators to mediate host gene

expression by binding to the promoters of host resistance (R) genes or susceptibility (S) genes, triggering resistance responses or inducing disease susceptibility, respectively (Gu et al. 2005; Yang et al. 2006; Kay et al. 2007; Antony et al. 2010). In addition to the nuclear localization motif and trans-activating domain, each TALE contains a central region of multiple 34- to 35-amino acid direct repeats that are nearly identical except the 12^{th} and 13^{th} amino acid residues (so called repeat variable di-amino-acids, or RVD) (Boch and Bonas 2010). The combination of repeat number and composition of RVDs of individual TALEs determine the specificity of the targeted genes (Bogdanove et al. 2010; Boch and Bonas 2010). Each repeat of TALE recognizes contiguously one nucleotide of target DNA, hereinafter referred to as EBE for effector binding element, in a simple cipher (Boch et al. 2009; Moscou and Bogdanove 2009). Among many native TALE repeats, four predominant types of repeats each recognize preferentially one of four nucleotides of target DNA. Therefore, the TALE recognition "code" can be used to guide custom-engineering of DNA binding domains with novel specificity to the user-chosen DNA sequences (Morbitzer et al. 2010; Zhang et al. 2011; Li et al. 2011). A variety of methods have been developed and used to synthesize TALE repeats that can be subsequently used for different fusions, in most cases for TAL effector nucleases (TALENs) [see review by Bogdanove and Voytas (Bogdanove and Voytas 2011)].

Bacterial blight of rice, caused by *Xanthomonas oryzae* pv. oryzae (Xoo), is a major rice disease in Asia and western Africa, often resulting in yield losses of up to 50% and sometimes loss of the entire crop (Mew 1987). The TALEs are important pathogenicity determinants in Xoo/rice interaction. Several TALEs have been found to be essential virulence factors of Xoo in susceptible rice, inducing host S genes and subsequently

promoting pathogen infection and disease development (Yang and White 2004; Yang et al. 2006; Yu et al. 2011; Antony et al. 2010). For example, some Xoo strains use TALE PthXo1 to induce the S gene *Os8N3* (*Xa13* or *OsSWEET11*) for disease in susceptible rice. Intriguingly, some genetic variations in the promoter of *Os8N3* are non-responsive to PthXo1 and confer disease resistance to PthXo1-dependent Xoo strains, and those S gene alleles are collective named as the recessive resistance gene *xa13* (Yang et al. 2006; Chu et al. 2006; Chen et al. 2010). The "resistance" is not due to the active defense but a lack of S gene induction and loss of disease susceptibility (Yang et al. 2006). To encounter TALEs, plants have evolved R genes, such as *Xa27*, that recognize TALEs through the promoter sequence and activate defense process once induced. *Xa27*, recognizing the cognate TALE AvrXa27, is the representative of an unusual class of dominant R genes in plant (Gu et al. 2004; Gu et al. 2005). Both resistant (*Xa27*) and susceptible (*xa27*) alleles contain identical coding sequences but only resistant cultivars express *Xa27* upon infection by Xoo strains expressing AvrXa27 (Gu et al. 2005). Expression of *Xa27* depends on the direct interaction between its promoter and AvrXa27 (Romer et al. 2009). When fused to the nonspecific pathogen inducible *OsPR1* promoter, the induced Xa27 conferred resistance to Xoo strains regardless of the presence of AvrXa27 (Gu et al. 2005). Two additional, as yet uncharacterized, rice R genes (*Xa10* for AvrXa10 and *Xa7* for AvrXa7) have similar requirements for the transcription activation domain and nuclear localization motifs of their cognate TALEs as required for AvrXa27-mediated induction of *Xa27* (Zhu et al. 1998; Yang et al. 2000). However, it remains unclear how many, if any, other TALE-recognizing R genes exist in rice genome.

In this work, we describe the feasibility of using custom-engineered TALEs to investigate the functionality of host target genes involved in Xoo/rice interaction. Specifically, by using artificial TALEs, we show that *xa13* can be induced by *Xanthomonas oryzae* pv. oryzae and confer disease susceptibility, lending further evidence for *Os8N3* (*OsSWEET11* or *Xa13*) as an S gene. This approach allowed us to identify another SWEET gene (*OsSWEET12*) that can act as an S gene, provided that the Xoo pathogen contains a corresponding TALE. We also show that the *xa27* allele can be activated and triggers resistance to the bacterium expressing a corresponding dTALE. However, four other *Xa27* family members do not appear to have the potential to act as R genes since their activation by dTALEs did not trigger resistance.

Results

Design of TAL effectors for targeted gene activation

We used an improved modular assembly method to synthesize TAL effector DNA binding domains. Four types of modular repeats with RVDs of NI, NG, NN and HD were used to recognize the respective target nucleotides of A, T, G and C. Instead of ligating eight pre-digested PCR-derived single-repeat units in one reaction as developed in our previous study (Li et al. 2011), the improvement involved digesting and ligating eight or fewer plasmid borne single-repeat units into one receptor plasmid in one tube by adapting the "Golden-Gate" cloning strategy described by Eagler et al. (Engler et al. 2008). Three arrays of 8-mers were ligated into scaffold of repeat-less *avrXa10*, resulting in a full-length dTALE presumably retaining specificity for the user-chosen DNA sequence (Figure 5.1). The target DNA sequences in the promoter of genes of interest were selected and included based on the

criteria: 1) Potential or obvious TATA box upstream of the transcription initiation site or predicted translation start site (if no cDNA sequence existed); 2) "T" preceding each EBE sequence (or at zero position of target site); 3) EBE about 23 bp long.

Designer TAL effector targeting susceptible allele of dominant R gene Xa27 elicited resistance responses

To test whether a dTALE could activate a host gene and produce a phenotype once translocated into host cells by pathogenic *Xanthomonas* bacterium, we carried out a proof-of-concept experiment with a susceptible allele (*xa27*) of *Xa27*. The alleles of *Xa27* between IR24 (referred to as *xa27*/*xa27*) and its near isogenic line IRBB27 (*Xa27*/*Xa27*) contain identical coding sequences but differ in the EBEs for AvrXa27 (Gu et al. 2005; Romer et al. 2009). The IR24 allele was irresponsive to AvrXa27 in gene induction due to incompatibility between the repeat domain of AvrXa27 and the promoter sequence and thus did not confer resistance to the pathogenic Xoo strains containing AvrXa27 (Gu et al. 2005) (Figure 5.2A). No naturally occurring TAL effector has been identified that is able to induce *xa27* and elicit resistance response in IR24. dTALE-xa27 was designed to recognize a sequence in the promoter of *xa27* (Figure 5.2A) and introduced into PXO99, an *avrXa27*-containing Xoo strain that caused disease in IR24 but resistance response in IRBB27 (Gu et al. 2005). PXO99 caused susceptible reaction in IR24 (yellow coloration at the inoculation sites), consistent with the previous studies (Hopkins et al. 1992; Gu et al. 2005), but elicited a typical hypersensitive response (brown coloration at the inoculation site) in the presence of dTALE-xa27 in IR24 (Figure 5.2B). The presence of dTALE-xa27 in PXO99 also restricted the bacterial growth in IR24 leaves by about 40 folds of population per inoculation site compared to PXO99 alone (Figure 5.2C). Furthermore, *xa27* is only induced by bacteria

containing dTALE-xa27 but not by PXO99 alone and mock inoculation (water) in a quantitative RT-PCR assay (Figure 5.2D). The results indicate that dTALE-xa27 is active and an avirulence factor corresponding to *xa27* that is, in turn, a disease resistance gene and is functionally equivalent to *Xa27* corresponding AvrXa27 in a gene-for-gene manner. The results also demonstrate, as proof-of-principle, the validity of expression, secretion and functionality of designer TAL effectors in rice/Xoo interaction.

Four annotated Xa27-like genes in Nipponbare did not confer resistance responses even induced by the corresponding dTALEs

The annotated Nipponbare genome encodes a family of four homologs with varying identities to XA27, referred to as *Xa27*-like genes (XALs), the most closely related one (*XAL1*, Os06g39810) shares 91% identity with XA27 of IRBB27 (Gu et al. 2005). Sequences for other three XAL genes could not be identified from the sequenced indica rice variety 93-11, and their entities in the un-sequenced IR24 genome are unknown (data not shown). We asked whether other members of *Xa27* family could also function as R genes upon induction in a gene-for-gene manner. Four dTALEs were designed to target the promoters of four *Xa27*-like genes in Nipponbare (Figure 5.3A). Each dTALE was introduced into PXO99, also a virulent strain in Nipponbare. PXO99 transformants with individual dTALEs caused strong water-soaking, a diagnostic feature of susceptible reaction (Figure 5.3B). The gene inductions of individual *Xa27*-like genes by individual dTALEs were validated in a quantitative RT-PCR (Figure 5.3C-F).

dTALE acted as a virulence factor and overcame xa13 recessive resistance

TAL effector PthXo1 induced the S gene *Os8N3* and subsequently a state of disease susceptibility, while the allelic *xa13* conferred resistance to PthXo1-dependent Xoo due to its

non-inducibility by PthXo1 and could be overcome by Xoo strains that induced an alternative S gene (e.g. *Os11N3* by AvrXa7 and/or PthXo3) (Chu et al. 2006; Yang et al. 2006; Antony et al. 2010). We hypothesized that *xa13*-mediated resistance could also be overcome by Xoo strain that directly induced *xa13* gene. To test this idea, we designed a TAL effector (dTALE11) based on the conserved promoter sequences of the allelic *xa13* and *Os8N3* (Figure 5.4A and B). *dTALE11* gene was introduced to ME2, a non-pathogenic mutant of PXO99 with *pthXo1* gene knocked out and lack of any other major TAL effector (Yang and White 2004). As expected, *dTALE11*-containing ME2 induced both *Os8N3* and *xa13* in IR24 and IRBB13, respectively (Figure 5.4C and D), and the inductions conferred disease susceptibility to ME2 in both varieties (Figure 5.4E), indicating a function of *xa13* as disease susceptibility gene upon induction. In agreement with prior studies (Yang and White 2004; Yang et al. 2006; Antony et al. 2010), the ability of ME2 to cause disease was severely attenuated due to the lack of any major virulent TAL effector to induce the corresponding S gene in both IR24 and IRBB13; reintroduction of *pthXo1* restored ME2 the ability to cause disease in IR24 but not in IRBB13 due to the difference in gene induction of *Os8N3* and the allelic *xa13*, respectively (Figure 5.4C and D).

dTALE targeted gene activation revealed the potential of OsSWEET12 as a disease susceptibility gene

Finally, we sought to further validate the feasibility of dTALEs for targeted gene activation and to expand their application to gene function analysis of uncharacterized genes. Rice genome contains at least twenty-one *SWEET* (or *N3*) genes with a phylogenetic clade that harbors two known S genes (*OsSWEET11* and *OsSWEET14*) and three additional

uncharacterized SWEETs (Chen et al. 2010). We designed two dTALEs to target one of them, Os03g22590 (also named *OsSWEET12*) in Nipponbare, by binding to its two putative TATA boxes (Figure 5.5A and B). The dTALE genes were transformed into ME2. Again as expected, ME2 transformant of each dTALE gene induced *OsSWEET12* expression in Nipponbare 24 hours post inoculation (Figure 5.5C), and also induced a state of disease susceptibility in term of disease development (Figure 5.5D), indicating that both dTALEs were virulence factors and *OsSWEET12* was a potential disease susceptibility gene for bacterial blight upon transcriptional activation. Similarly, test on other two SWEETs is in progress and the preliminary data indicates they also appear to contribute disease susceptibility upon induction (data not shown).

Discussion

The bacterial blight of rice represents an excellent plant/pathogen system for studying the biology of TAL effectors due to the large reservoir of TALEs in each strain of *Xanthomonas oryzae* pv. oryzae and the diverse roles of TALEs in pathogenesis of the disease (Nino-Liu et al. 2006; White and Yang 2009). Indeed, many years of effort spent in understanding the interaction between TALEs and their host target genes has led to the discovery of the DNA recognition cipher of TALEs (Boch et al. 2009; Moscou and Bogdanove 2009). The programmability of the TALE central repeat domains enables custom-engineering DNA-binding proteins with predictable specificities feasible to manipulate gene function in eukaryotes (Bogdanove and Voytas 2011). In the present study, we expanded the utilization of TALE-based technologies from targeted gene editing to gene activation and function analyses involved in Xoo/rice interaction (Li et al. 2012). We

generated seven artificial or designer TAL effectors based on the promoter sequences of six loci in rice (Figures 5.2A, 5.3A, 5.4A and B, and 5.5A and B) and transformed them into *Xanthomonas oryzae* pv. oryzae. Gene expression analyses indicated all dTALEs were active in trans-activating target genes in rice (Figures 5.2D, 5.3C-5.3F, 5.4C and D, and 5.5C), achieving a 100% success rate in engineering active dTALEs. Gene activation of three alleles (*xa27*, *xa13* and *OsSWEET12*) led to phenotypic changes in disease resistance or susceptibility in response to Xoo infection (Figures 5.2B and C, 5.4E, and 5.5D), suggesting the feasibility of this approach to the gene functional analysis.

Prior studies demonstrated the gene activation by artificial TALEs, but the associated phenotypic changes were lacking. For example, dTALEs designed for activation of tomato *Bs4*, *Arabidopsis EGL3* and *KNAT1* could activate the respective genes upon *Agrobacterium*-mediated ectopic expression of dTALEs, but the expected phenotypic alterations were not observed (Morbitzer et al. 2010). Also in mammalian and human cells, dTALEs with an architecture of a viral nuclear localization motif combined with the VP16 or its derived transcription activation domain were only demonstrated in up-regulating the target genes upon ectopic expression (Zhang et al. 2011; Geissler et al. 2011; Bultmann et al. 2012). Therefore, realization of the potential of dTALEs for targeted gene activation in basic and applied researches needs more studies.

Bacterial blight of rice involves molecular interactions of genes or gene products from both pathogen and rice, and gene regulation in host through pathogen effectors appears to play an important role in the pathogenesis of disease (White and Yang 2009). dTALEs provide an useful tool box to transiently activate host genes of interest and to quickly assess the associated host phenotypic and physiological changes in response to pathogen infection.

As a case in point, we quickly assessed the four homologous genes of the R gene *Xa27* for their functionality in conferring resistance responses and one sucrose efflux transporter gene for its vulnerability to bacterial disease. Likewise, dTALEs may be used for genome-enabled functional analysis of genes and networks involved not only in rice/Xoo interaction but also probably in other plant disease systems.

Materials And Methods

Plant Material and Growth Conditions

Rice (*Oryza sativa*) varieties IR24, IRBB13, Nipponbare were used in the study. Seeds of Nipponbare (accession number PI 514663) were obtained from the USDA-Agricultural Research Service National Small Grains Collection. IR24 and IRBB13 seeds were kindly provided by the International Rice Research Institute (courtesy of Casiana Vera Cruz). All rice plants were grown in growth chamber with temperature of 28°C, relative humidity of 75%, and photoperiod of 12 hours.

Plasmids, Bacterial Strains and Growth Conditions

Plasmids used in this study included those in the modular assembly kit we developed to synthesize designer TAL effectors (detailed information available upon request), pZWavrXa10 [*avrXa10* in pBlueScript II KS, unique HindIII site after *avrXa10* coding region (Zhu et al. 1999)], pHM1 [broad host range plasmid (Hopkins et al. 1992)], and the plasmids containing *dTALEs* as specified in the text. *E. coli* strains included XL1-Blue MRF′ [*F′proAB lacI*^{q}*ZΔM15 Tn10* (Tet^{r}), Stratagene, La Jolla, CA, U.S.A.], DB3.1 [F– *gyr*A462 *end*A1 Δ(sr1-*rec*A) *mcr*B *mrr hsd*S20(rB–, mB–) *sup*E44 *ara*-14 *gal*K2 *lac*Y1 *pro*A2 *rps*L20(SmR) *xyl*-5 λ– *leu mtl*1, Life Technologies], JM109 [*end*A1, *rec*A1, *gyr*A96,

thi, *hsd*R17 (rk–, mk+), *rel*A1, *sup*E44, Δ(*lac-pro*AB), F′, *tra*D36, *pro*AB, *laq*IqZΔM15, Promega], C2110 [Nal^r Rif^r (*polA1, rha, his*) (Leong et al. 1982)]. *Xanthomonas oryzae* pv. oryzae strains included PXO99 (Hopkins et al. 1992) , its derivative ME2 (Yang and White 2004), and their transformants with plasmids as specified in the text. *E. coli* cells were grown in Luria Broth with a standard culture technique (Ausubel et al. 1993). *X. oryzae* pv. oryzae was grown in either nutrient broth (Difco Laboratories) or typtone sucrose medium (tryptone, 10 g/liter; sucrose, 10 g/liter; glutamic acid, 1 g/liter) at 28°C. Antibiotics used in this study were carbenicillin (100 ug/ml), cephalexin (10 ug/ml), kanamycin (50 ug/ml), tetracycline (12.5 ug/ml) and spectinomycin (100 ug/ml).

dTALE Construction

Four "core" repeats from *avrXa7* encoding RVDs of NI, NG, NN and HD were PCR-amplified with specially designed primers and the PCR products were individually cloned into the A/T cloning vector pGEM-T (Promega) to generate fifty-one plasmids, each containing a single modular repeat. Sequence information for the primers and modular repeats is available upon request. A receptor vector (pTL-n) was made to clone 8 or fewer single repeats together in a predetermined order by using the "Golden Gate" cloning strategy (Engler et al. 2008). Then three repeat arrays were cloned into the repeat-deleted pZWavrXa10, resulting in individual TAL effector genes consisting of 5'- and 3'-coding sequences of *avrXa10* and the newly synthesized repeat domains. The kit with the fifty-three plasmids is available from the Yang laboratory under a Material Transfer Agreement (MTA) and the detailed protocol is also available upon request.

Disease Assay

Fully expanded leaves of rice plants at the stages indicated in the text were inoculated using leaf tip clipping method for lesion measurement (Kauffman et al. 1973) and syringe infiltration for observation of water-soaking/hypersensitive reaction and for bacterial population counting as described (Yang and White 2004). Bacterial suspensions of optical density of 1.0 at 600 nm (OD_{600}=1.0) were used for syringe infiltration and OD_{600}=0.5 for clipping inoculation. One-way analysis of variance (ANOVA) statistical analyses were performed on all measurements. The Tukey honest significant difference test was used for post-ANOVA pair-wise tests for significance, set at 5% ($P < 0.05$).

Gene Expression Analyses

The rice leaves were inoculated with indicated bacterial strains and used for total RNA extraction 24 hours post inoculation. RNA was extracted using the TRI reagent from Ambion, and RNA concentration and quality were measured using an ND-1000 Nanodrop spectrophotometer (Nanodrop Technologies). One microgram of RNA from each inoculation with bacteria as indicated in the text was treated individually with amplification grade DNase1 (Invitrogen) followed by cDNA synthesis using the iScript Select cDNA synthesis kit (Bio-Rad). cDNA derived from 25 ng of total RNA was used for each real-time PCR with gene-specific primers. *TFIIAγ5* expression was used as an internal control. The RT-PCR was performed on Strategene's Mx4000 multiplex quantitative PCR system using the iQ SYBR Green Supermix kit (Bio0Rad). The average threshold cycle (Ct) was used to determine the fold change of gene expression. The $2^{\Delta\Delta Ct}$ method was used for relative quantification (Livak and Schmidttgen, 2001). The primer sequences are provided in Supplemental Table 1 online.

Acknowledgements

The research was supported by the National Science Foundation (Award 0820831) and the Iowa State University faculty startup fund.

References

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

Ausubel, F., Brent, R., Kingston, R., Moore, D., Seidman, J., Struhl, K. *Current Protocols in Molecular Biology.* 1st ed. New York: Canada John Wiley and Sons; 1993.

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U. Breaking the code of DNA binding specificity of TAL-type III effectors. Science 2009; 326:1509-1512.

Boch J, Bonas U. Xanthomonas AvrBs3 family-type III effectors: Discovery and function. Annu Rev Phytopathol 2010; 48:419-436.

Bogdanove AJ, Schornack S, Lahaye T. TAL effectors: Finding plant genes for disease and defense. Curr Opin Plant Biol 2010; 13:394-401.

Bogdanove AJ, Voytas DF. TAL effectors: Customizable proteins for DNA targeting. Science 2011; 333:1843-1846.

Bonas U, Schulte R, Fenselau S, Minsavage GV, Staskawicz BJ, Stall RE. Isolation of a gene cluster from*Xanthomonas campestris* pv.*vesicatoria* that determines pathogenicity and the hypersensitive response on pepper and tomato. Mol Plant Microbe Interact 1991; 4:81-88.

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

Chen LQ, Hou BH, Lalonde S, Takanaga H, Hartung ML, Qu XQ, Guo WJ, Kim JG, Underwood W, Chaudhuri B, Chermak D, Antony G, White FF, Somerville SC, Mudgett MB, Frommer WB. Sugar transporters for intercellular exchange and nutrition of pathogens. Nature 2010; 468:527-532.

Chen LQ, Qu XQ, Hou BH, Sosso D, Osorio S, Fernie AR, Frommer WB. Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. Science 2012; 335:207-211.

Cho HJ, Park YJ, Noh TH, Kim YT, Kim JG, Song ES, Lee DH, Lee BM. Molecular analysis of the hrp gene cluster in xanthomonas oryzae pathovar oryzae KACC10859. Microb Pathog 2008; 44:473-483.

Chu Z, Yuan M, Yao J, Ge X, Yuan B, Xu C, Li X, Fu B, Li Z, Bennetzen JL, Zhang Q, Wang S. Promoter mutations of an essential gene for pollen development result in disease resistance in rice. Genes Dev 2006; 20:1250-1255.

Engler C, Kandzia R, Marillonnet S. A one pot, one step, precision cloning method with high throughput capability. PLoS One 2008; 3:e3647.

Geissler R, Scholze H, Hahn S, Streubel J, Bonas U, Behrens SE, Boch J. Transcriptional activators of human genes with programmable DNA-specificity. PLoS One 2011; 6:e19509.

Gu K, Tian D, Yang F, Wu L, Sreekala C, Wang D, Wang GL, Yin Z. High-resolution genetic mapping of Xa27(t), a new bacterial blight resistance gene in rice, oryza sativa L. Theor Appl Genet 2004; 108:800-807.

Gu K, Yang B, Tian D, Wu L, Wang D, Sreekala C, Yang F, Chu Z, Wang GL, White FF, Yin Z. R gene expression induced by a type-III effector triggers disease resistance in rice. Nature 2005; 435:1122-1125.

Hopkins CM, White FF, Choi SH, Guo A, Leach JE. Identification of a family of avirulence genes from xanthomonas oryzae pv. oryzae. Mol Plant Microbe Interact 1992; 5:451-459.

Kauffman HE, Reddy APK, Hsiek SP, and Marca SD. An improved technique for evaluating resistance of race varieties to*Xanthomonas oryza*. Plant Dis Rep 1973; 57:537-541.

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

Leong SA, Ditta GS, Helinski DR. Heme biosynthesis in rhizobium. identification of a cloned gene coding for delta-aminolevulinic acid synthetase from rhizobium meliloti. J Biol Chem 1982; 257:8724-8730.

Lievak, KJ., Schmittgen, TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-ΔΔC(t)) method. Methods 25; 402-408

Li T, Huang S, Zhao X, Wright DA, Carpenter S, Spalding MH, Weeks DP, Yang B. Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. Nucleic Acids Res 2011; 39:6315-6325.

Li T, Liu B, Spalding MH, Weeks DP, Yang B. High-efficiency TALEN-based gene editing produces disease-resistant rice. Nat Biotechnol 2012; 30:390-392.

Mew TW. Current status and future prospects of research on bacterial blight of rice. Annu Rev Phytopathol 1987; 25:359-382.

Morbitzer R, Romer P, Boch J, Lahaye T. Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. Proc Natl Acad Sci U S A 2010; 107:21617-21622.

Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL effectors. Science 2009; 326:1501.

Nino-Liu DO, Ronald PC, and Bogdanove A. *Xanthomonas oryzae*pathovars: Model pathogens of a model crop. Mol Plant Pathol 2006; 7:303-324.

Romer P, Recht S, Lahaye T. A single plant resistance gene promoter engineered to recognize multiple TAL effectors from disparate pathogens. Proc Natl Acad Sci U S A 2009; 106:20526-20531.

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

White FF, Potnis N, Jones JB, Koebnik R. The type III effectors of xanthomonas. Mol Plant Pathol 2009; 10:749-766.

White FF, Yang B. Host and pathogen factors controlling the rice-xanthomonas oryzae interaction. Plant Physiol 2009; 150:1677-1686.

Yang B, Zhu W, Johnson LB, White FF. The virulence factor AvrXa7 of xanthomonas oryzae pv. oryzae is a type III secretion pathway-dependent nuclear-localized double-stranded DNA-binding protein. Proc Natl Acad Sci U S A 2000; 97:9807-9812.

Yang B, White FF. Diverse members of the AvrBs3/PthA family of type III effectors are major virulence determinants in bacterial blight disease of rice. Mol Plant Microbe Interact 2004; 17:1192-1200.

Yang B, Sugio A, White FF. Os8N3 is a host disease-susceptibility gene for bacterial blight of rice. Proc Natl Acad Sci U S A 2006; 103:10503-10508.

Yu Y, Streubel J, Balzergue S, Champion A, Boch J, Koebnik R, Feng J, Verdier V, Szurek B. Colonization of rice leaf blades by an african strain of xanthomonas oryzae pv. oryzae depends on a new TAL effector that induces the rice nodulin-3 Os11N3 gene. Mol Plant Microbe Interact 2011; 24:1102-1113.

Zhang F, Cong L, Lodato S, Kosuri S, Church GM, Arlotta P. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol 2011; 29:149-153.

Zhu W, Yang B, Chittoor JM, Johnson LB, White FF. AvrXa10 contains an acidic transcriptional activation domain in the functionally conserved C terminus. Mol Plant Microbe Interact 1998; 11:824-832.

Zhu W, Yang B, Wills N, Johnson LB, White FF. The C terminus of AvrXa10 can be replaced by the transcriptional activation domain of VP16 from the herpes simplex virus. Plant Cell 1999; 11:1665-1674.

Zhu W, MaGbanua MM, White FF. Identification of two novel hrp-associated genes in the hrp gene cluster of xanthomonas oryzae pv. oryzae. J Bacteriol 2000; 182:1844-1853.

Figures and Tables

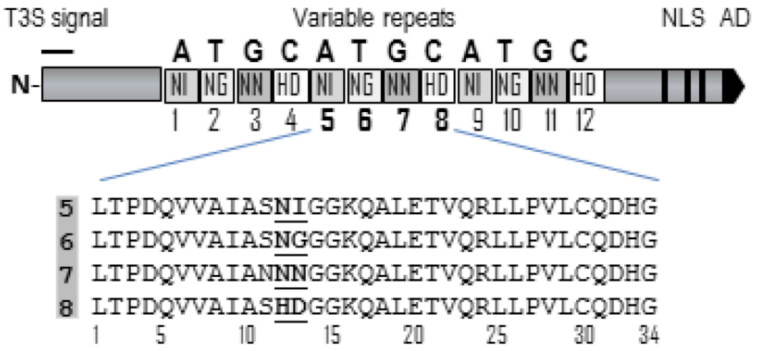


Figure 5.1. Structure of a typical design TAL effector (dTALE). The dTALEs consist of a scaffold of the repeat-less AvrXa10 and the custom-engineered TAL effector repeat domains. Each dTALE contains an N-terminal type III secretion signal (T3S signal), three functional nuclear localization motifs (NLS) and an acidic trans-activating domain (AD), and a varying number of repeats as each represented by the 12th and 13th amino acids (Variable repeats). Four types of 34-amino-acid repeats, each recognizing one of four nucleotides, are used to assemble the repeat domain of each dTALE.

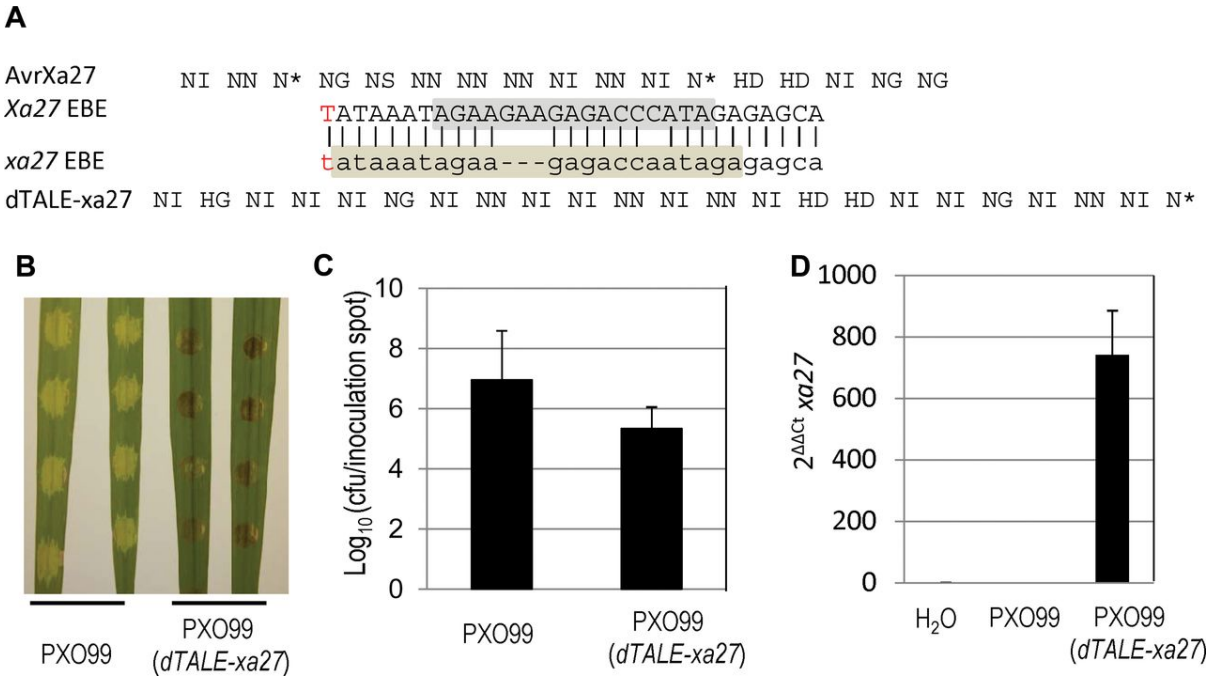


Figure 5.2. dTALE-xa27 induced the susceptible allele of *Xa27* and triggered resistance responses in rice IR24. **(A).** Promoter sequence for dTALE-xa27 (*xa27* EBE, shaded in tan) in IR24 and the TAL effector AvrXa27 binding element of *Xa27* (*Xa27* EBE, shaded in grey) in IRBB27. **(B).** Disease reactions of IR24 leaves to PXO99 strains containing or lacking dTALE-xa27. Hypersensitive reaction (HR) is indicated by tissue browning within 48 hours compared to susceptible reaction (yellow coloration) at the inoculation sites. **(C).** Populations of PXO99 and derivative in leaves of IR24. Populations are expressed as colony forming unit (cfu) per inoculation spot 4 days post inoculation. Lowercase letters above columns indicate significant differences at the <0.01 level using Tukey statistic and ANOVA analysis. Bars indicate 1 SD. **(D).** Gene induction of *xa27* in IR24. RNA from leaves treated with water, PXO99 and dTALE-xa27-containing PXO99 was used for quantitative RT-PCR with gene-specific primers and rice TFIIA gamma subunit gene in chromosome 5 used as an internal control.

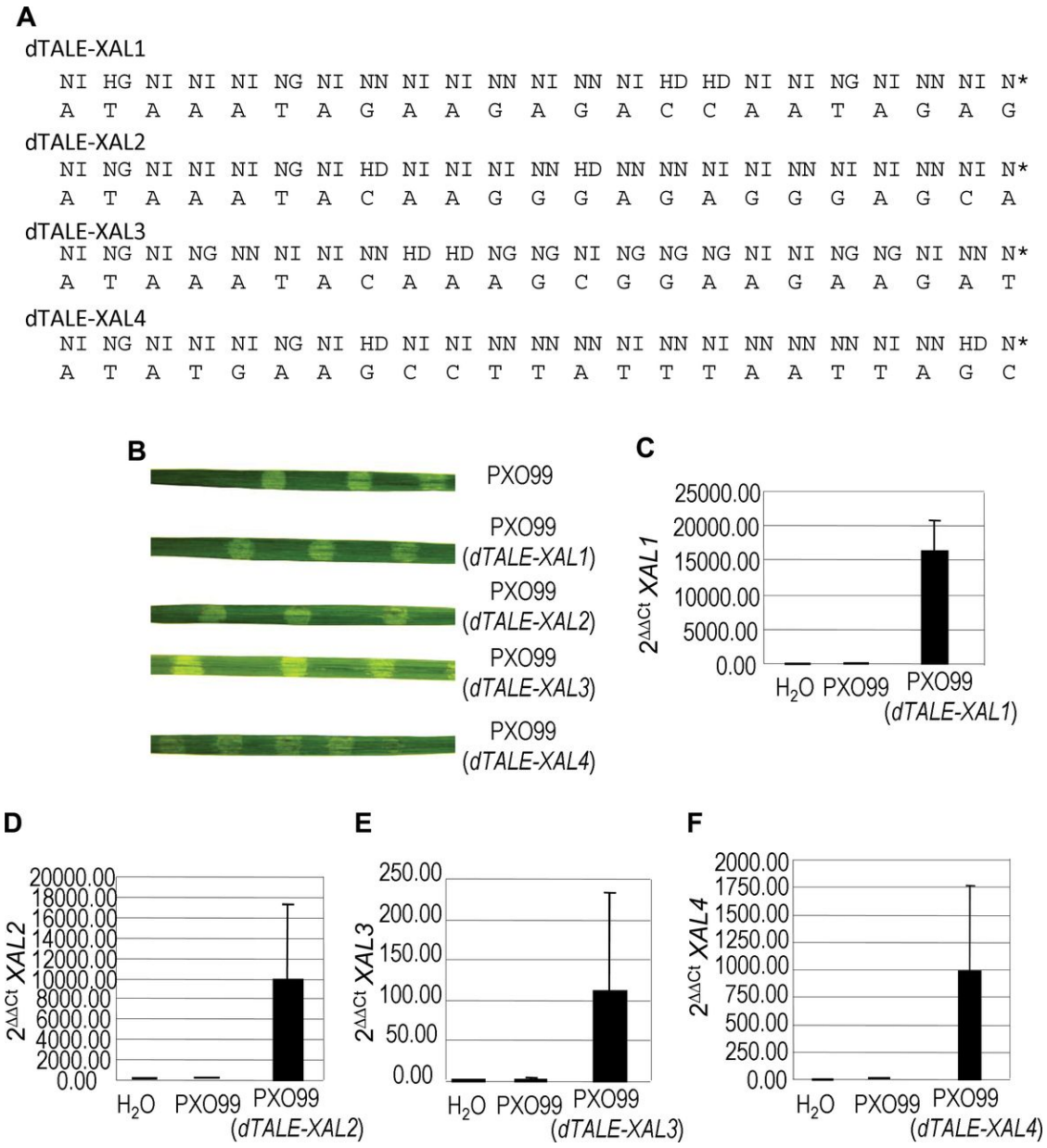


Figure 5.3. Four dTALEs targeted four *Xa2*-like genes in Nipponbare for gene activation and disease reaction. **(A).** Four dTALEs and their corresponding EBEs in the respective promoters of *Xa27* homologs. **(B).** Disease reactions of Nipponbare leaves in response to PXO99 strains containing different dTALEs. **(C), (D), (E), (F).** Gene induction of individual *Xa27* homologs in response to different Xoo strains as detected by RT-PCR with gene-specific primers. The rice *TFIIA* gamma subunit gene was used as an internal control.

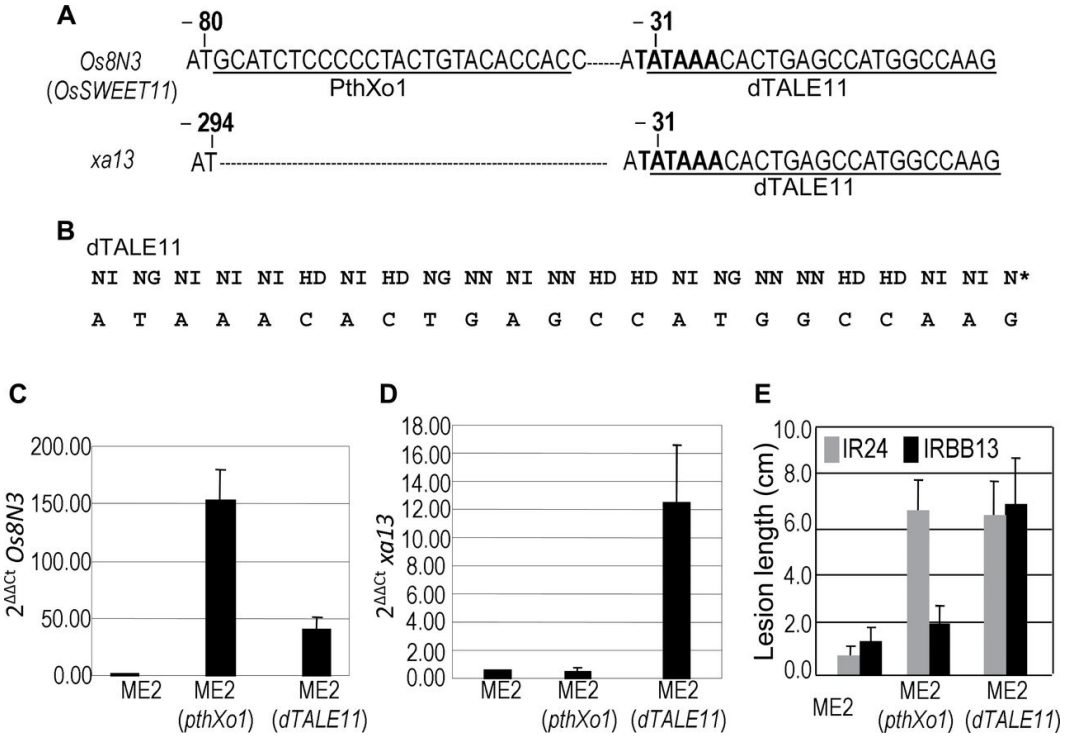


Figure 5.4. dTALE11 induced and overcame resistant gene *xa13*. **(A).** Promoter sequences of the allelic *Os8N3* and *xa13* with TAL effector binding elements (EBEs). PthXo1 recognizes *Os8N3* (also named *OsSWEET11,* or *Xa13*) in susceptible rice IR24 but not *xa13* in the IR24-derived near isogenic resistant line IRBB13. dTALE11 was designed to activate both alleles by recognizing the identical promoter sequences (underlined and overlapping with the TATA box) in IR24 and IRBB13. **(B).** dTALE11 with RVDs and their corresponding target nucleotides (EBE). **(C), (D).** Gene induction detected with RT-PCR. Leaf RNA of IR24 and IRBB13 from different treatments was subjected to quantitative RT-PCR with gene specific primers. PCR for *TFIIAγ5* was used as an internal control. ME2 is a *pthXo1* knockout PXO99. **(E).** Disease severities (lesion lengths) caused by different Xoo strains in IR24 (grey bars) and IRBB13 (dark bars). Lesions were measured 10 days post inoculation. Error bars indicate ±SD.

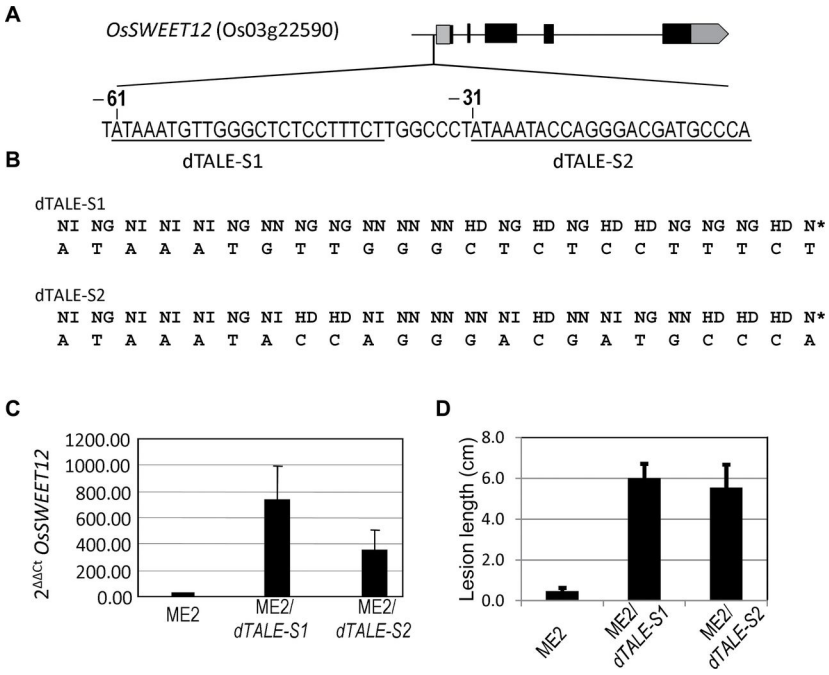


Figure 5.5. *OsSWEET12* induction conferred disease susceptibility to Xoo. (A). Two dTALE binding elements (underlined) in the promoter of *OsSWEET12*. (B). Repeat units (RVDs) of two dTALEs and their EBEs. (C). *OsSWEET12* induction as detected by RT-PCR. Leaf RNA of rice Nipponbare treated with different Xoo strains was subjected to RT-PCR with gene-specific primers. RT-PCR with *TFIIAγ5* was used as an internal control. (D). Lesion lengths caused by Xoo in Nipponbare leaves. Lesions in 4-week old plants were measured 10 days post inoculation. Error bars indicate ±SD.

CHAPER 6
CONCLUSION AND FUTURE WORK

General Conclusions

The “one to one” TALE DNA recognition code enables the TALEs to become powerful technologies in synthetic biology and also in areas that TALEs can be used as site specific DNA binding proteins, such as nucleases, transcription factors and epigenetic modifiers. TALE-based technologies are applicable in a plethora of organisms. My work, in collaboration with others, on TALEs has largely focused on technology development and utilization, especially in the area of DNA-targeting using the TALE derived nucleases and transcription activators.

Basically, I have first demonstrated the TALE could function as DNA binding domain similar to zinc finger proteins. After fused with the FokI cleavage domain at the C-termius, TALE nuclease proteins could generate DNA double-strand break at the expected sites. And the activities were comparabe to the activities of the well-characterized ZFNs (zinc finger nucleases).

After determining the activity of TALENs derived from the naturally TALEs, I conceived a method to synthesize novel TALE DNA binding domains for any preselected DNA sequence by assembling four types of modular TALE repeats (RVD NI, NG, NN and HD recognizing A, T, G, and C, respectively). Initially, I developed a restriction/ligation method with the PCR-amplified modular repeats. This method could assemble 8 repeats in an order of any combination. Later, I have improved the modular assembly method by combining the Golden Gate cloning strategy with a plasmid-borne repeat library. The improved method is simpler and more efficient to make active designer TALENs.

With the proof of concept demonstration of generating functional TALENs *in vitro* and *in vivo*, I went on to modify the genes of interest with custom-engineered TALENs in rice coupled with rice tissue culture and transformation. The first target is the EBE of *Os11N3* (or *OsSWEET14*) for the native TALEs AvrXa7 and PthXo3. The modified rice lines with EBE changes showed bacterial disease resistance because of the prevention of AvrXa7 and PthXo3 from binding and further transcriptional activation of the rice susceptibility gene *OsSWEET14*. The TALEN-induced genetic changes were also capable of transmitting into the progeny plants. Importanly, the TALEN transgenes could be genetically segregated out in some of progeny plants, resulting in lines with desired modification but lacking any "foreign" DNA left over in the genome. The study is the first case of using TALEN technology to create stable and heritable trait of agromonically significance.

Additionally, my work has demonstrated the feasibility in using designer TALEs to probe the gene function in rice. TALEs function as transcription activator by binding to the EBEs in the promoters of target genes after the bacterial pathogen injects them into host cells through a type III secretion system. By using the Xoo bacterium to produce and deliver designer TALEs, I was able to show one member of SWEET gene family (*OsSWEET12*) was a potential disease susceptibility gene for bacterial blight of rice provided it would be induced by an as yet unknown TALE.

Future Prospects

Other TALE fusion proteins

TALEs bind to double-stranded DNA of specific sequences, while it has been shown that TALE is also capable of binding to DNA-RNA hybrid strand with the specificity

determined by the interaction between the TALE repeats and the sequence of DNA strand. The physical binding is able to block the accessibility of other macromolecules to the RNA strand, pointing to a prospect of potential use of designer TALEs to interfere with the DNA replication and retroviral infections (Yin et al. 2012).

TALEs fused with different domains can function differently for various avenues of DNA-targeting. For example, TALE-derived recombinases can be used to recombine DNA sequences in bacterial and mammalian cells (Mercer et al. 2012); TALEs fused with fluorescent proteins to specifically target telomeres or satellite repeats located at centromeres are used for chromosome study and *in vitro* detection in fixed cells (Ma et al. 2013, Thanisch et al. 2013); TALEs fused with transposase are capable of selectively integrating single copy of transposon into a safe harbor (Owens et al. 2013).

In addition to genome editing, TALEs have also been applied in epigenome modification at specific genomic loci. For instance, TALEs fused with the catalytic domain of chromotin-modifying enzyme ten-eleven translocation 1 (TET1) are utilized for specific CpG demethylation and analysis of gene regulation (Maeder et al. 2013). In a similar case, Mendenhall et al. fused TALEs with lysine-specific demethylase 1 (LSD1) to selectively inactivate enhancer-associated chromatin state through demethylation of H3K4 mono- and di-methylation in a sequence specific manner (Mendenhall et al. 2013). Most recently, two components (SUVH2 and SUVH9) of RNA-directed DNA methylation complex in *Arabidopsis thaliana* were shown to bind to methylated DNA and recruit Pol V to reinforce RNA directed DNA methylation. When fused with a zinc finger protein, SUVH2 is capable of recruit Pol V to the ZF target site and establish DNA methylation and gene silencing (Johnson et al. 2014). Likewise, SUVH2, when fused with the rogrammable TALEs, is

capable of inducing site directed DNA methylation for gene silencing and functional analysis. Therefore, TALE-directed site specific DNA methylation, demethylation and histone modification may signicantly facilitate the epigenetic and chromatin state study.

Opportunity for Improvement

Progress in TALE-based technology development has occurred at an astounding pace, which speaks volumes about the positive change that a diverse group of research scientists can bring to a new technology when they have a common goal; in this case, creating a universally powerful tool for Genome Editing. On the other hand, there are still many areas that could be improved, including toxicity tests, alternatives for the T_0 recognition domain, evaluation of alternative RVDs, and an improved nuclease domain.

Toxicity Test

Toxicity was identified as an issue early in the development of nuclease based Genome Editing, and it still is an apparent issue with TALENs. While progress has been made in testing the relative functionality of nuclease designs using methods such as the yeast-based nuclease assay, a simple, general test for nuclease toxicity has yet to be developed. This is complicated by the fact that toxicity is not a universal issue, but may be more of a species-dependent or cell-specific problem. Various solutions for toxicity have been suggested, but most do not address the central issue, and often it is simpler to simply test several engineered nucleases, then use the one that works best. This screening for toxicity can be performed by evaluating relative survival based on co-transformation with a marker gene. While this approach works, it is not always applicable or convenient, especially in plant cells that are not amenable to protoplast isolation or in animal cells in

which nucleases are delivered to individual cells one at a time. Some out-of-the box thinking will be required to solve this problem.

T_0 Domain

Another area for improvement is the repeat 0 region of the TALE DNA binding domain. Repeat 0 binds to the first base in an EBE, which is generally a T. Expanding the spectrum of this specificity would further increase the flexibility for identification of effective TALEN target sequences. Structural data show that repeat 0 has a helix-loop-helix motif and that T specificity is based on its interactions with W232 in the KQWSG sequence of the loop. When mutations are generated in this region, the specificity for a T is relaxed; however, these gains appear to be context specific. That is to say, moving the mutation to a related TALE or using the mutation in a TALEN, TALE activator or a TALE recombinase changes the recognition pattern (Lamb et al. 2013, Tsuji et al. 2013, Doyle et al. 2013). Additionally, it has been demonstrated that repeat 0 from *R. solanacearum* preferentially activates promoters containing EBEs with a 5' G. However, when tested in a PthXo1 construct, although the *R. solanacearum* repeat 0 changed specificity to G, it did so at the expense of activity (de Lange et al. 2013, Doyle et al. 2013). Interestingly, an optimized TALE scaffold using the N terminal 207 a.a. plus the C terminal 63 a.a. from AvrXa10 reportedly exhibited similar activity whether a target DNA sequence contained a 5' C, A, T or G (Sun et al. 2012). These reports suggest that the specificity of repeat 0 for a T is not immutable, but they also suggest that a larger scale effort will be needed to effectively accommodate A, C and G at position 0 in many contexts.

Alternative RVDs

The most commonly used RVDs are HD, NI, NG, and NN for the recognition of C, A, T and G respectively. While these RVDs work well, they each have strengths and weaknesses. For example, NN recognizes A about as well as G, HD interacts strongly with C, but is sensitive to 5-methylcytosine (5mC), NI has a relatively weak interaction with A, and NG interacts strongly with T, but also recognizes 5mC (Streubel et al. 2012, Meckler et al. 2013, Lin et al. 2014, Deng et al. 2012b).

Many other RVDs exist, but most have not been stringently tested. Of the few examined, it has been found that NK has a higher specificity for G over NN, but also lower affinity, which results in less TALEN activity in yeast and human cell based recombination assays (Huang et al. 2011, Mahfouz et al. 2011, Cong et al. 2012). In contrast, RVD NH has been shown to have higher specificity for G over NN without sacrificing affinity (Christian et al. 2012, Streubel et al. 2012, Mechler et al. 2013, Cong et al. 2012). Also, N* (missing the 13^{th} aa) was shown to bind 5mC as well as C, A, T and G; making it a generic RVD (Valton et al. 2012).

In addition to differences in binding affinity and specificity among the various RVDs, there exits polarity in the TALE DNA binding domain, with binding being strongest in the N-terminal region and weakest in the C-terminal end. This polarity makes a TALE much more sensitive to 5' target site mismatches than 3' mismatches, and at least one study has shown that RVD HD has a stronger affinity for C when located nearer to the N-terminus (Streubel et al. 2012, Meckler et al. 2013, Lin et al. 2014).

The ability to specifically recognize modified and unmodified DNA has important implications in epigenetic and cancer research, but also has consequences for general

targeting efficiency. Until the alternative RVDs are systematically tested for binding affinity, specificity and position effect, it will remain an open question as to which RVDs are best for any specific application.

Improved TALE Nuclease Design

TALENs and ZFNs could benefit from an improved nuclease design. Both commonly use FokI, but FokI functions as a homodimer and engages in *trans* cleavage. Each of these characteristics contributes to cell toxicity, and improved heterodimeric forms of FokI have reduced much of the toxicity. However, there is still much room for improvement. There are several reported examples of alternative structures, such a nickases, alternative dimeric nucleases and attempts to create monomeric TALENs that have the potential to improve application of TALENs under some conditions.

Nickases

A FokI nickase was produced by mutating the catalytic site (D450A) of one partner in a FokI heterodimer to induce a single strand break rather than a DSB (Ramirez et al. 2012, Wang et al. 2012, Kim et al. 2012). The idea was to suppress the NHEJ rate in favor of HR, which worked. However, the HR rate was also greatly reduced (Metzger et al. 2011, Davis and Maizels. 2011, van Nierop et al. 2009). A different approach to engineering a nickase reportedly fused the monomeric nickase MutH to a TALE DNA binding domain. MutH is a DNA mismatch repair endonuclease that recognizes the sequence GATC. This fusion protein was tested in a plasmid cleavage assay *in vitro* and shown to function as a nickase

(Gabsalilow et al. 2013). Further exploration of these and other TALE-based nickases may eventually lead to higher HR efficiencies.

Alternative Nucleases

Because FokI *trans* activity causes some toxicity, the development of alternative nuclease partners may improve the general application of TALENs in Genome Editing. The Type IIP restriction endonuclease PvuII was fused with a TALE, a catalytically inactive I-SceI and a zinc finger array. These new homodimeric nucleases were shown to have high specificity, but retained the PvuII recognition sequence as part of their overall target site (Fonfara et al. 2012, Schierling et al. 2012, Yanik et al. 2013). This work demonstrates that other nucleases can be paired with various DNA binding domains to create greater specificity. One wonders if a restriction enzyme with a four base specificity may be more versatile or if base specificity could be eliminated in an alternative nuclease.

Monomeric TAL effector nucleases

Because FokI nuclease requires dimerization and thus the use of paired TALENs, the development of monomeric forms may improve the general application of TALENs in Genome Editing. A real breakthrough may come from the fusion of a nonspecific monomeric nuclease or cleavage domain and a TALE. Such a nuclease would streamline TALEN production by halving the development work load and simultaneously produce a more compact enzyme. Some attempts have been made to design a FokI monomer and to make TALE fusions to known monomeric nucleases.

A monomeric engineered nuclease has been reportedly isolated using a high throughput screen. The nuclease included a TALE DNA binding domain fused to tandem *FokI* cleavage domains separated by a 95 aa linker. Yeast and human assays indicated that, compared with a standard TALEN heterodimer, this nuclease only had 9% to 36% activity and exhibited significantly higher cytotoxicity. These characteristics may result from the shorter target leading to more off-target DSBs within the human genome, or they may be more related to the monomeric nature of this FokI design (Sun and Zhao, 2014).

A "mega-TAL" fusion protein was described recently as another strategy for monomeric nuclease construction. This nuclease is composed of a TALE DNA binding domain and the site specific LAGLIDADG monomeric meganuclease I-Anil. In this fusion, I-Anil retains its native target site specificity, which is further enhanced by the TALE DNA binding domain. This architecture, when combined with the Trex2 exonuclease, demonstrated high specificity and increased nuclease activity, which resulted in the modification of 70% of human cells compared to wild type I-Anil (Boissel et al. 2013).

In another monomeric example, a TALE DNA binding domain was fused to the site specific GIY-YIG monomeric meganuclease I-TevI catalytic domain. This domain was previously shown to be functionally separate from its native DNA binding domain, but it still retains partial site specific recognition, as the sequence CNNNG is required for nuclease action (Edgell et al. 2004). Additionally, it was demonstrated that this I-TevI nuclease domain can be fused to other DNA binding domains to produce hybrid nucleases (Kleinstiver et al. 2012). The I-TevI::TALE fusion proved to be a functional monomeric nuclease, and, as expected, retained the CNNNG sequence requirement as part of its overall target sequence. This fusion was shown to have an activity of 8.9% compare to 6.9% for a standard

TALE::FokI fusion in plants and 4.3% versus 9.2% respectively in hamster CHO-K1 cells (Beurdeley et al. 2013).

REFERENCES CITED

Alwin, S., et al. (2005). "Custom zinc-finger nucleases for use in human cells." Mol Ther 12(4): 610-617.

Ansai, S., et al. (2014). "Design, evaluation, and screening methods for efficient targeted mutagenesis with transcription activator-like effector nucleases in medaka." Dev Growth Differ 56(1): 98-107.

Arnould, S., et al. (2006). "Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets." J Mol Biol 355(3): 443-458.

Ashworth, J., et al. (2006). "Computational redesign of endonuclease DNA binding and cleavage specificity." Nature 441(7093): 656-659.

Ashworth, J., et al. (2010). "Computational reprogramming of homing endonuclease specificity at multiple adjacent base pairs." Nucleic Acids Res 38(16): 5601-5608.

Bacman, S. R., et al. (2013). "Specific elimination of mutant mitochondrial genomes in patient-derived cells by mitoTALENs." Nat Med 19(9): 1111-1113.

Baltes, N. J., et al. (2014). "DNA Replicons for Plant Genome Engineering." Plant Cell.

Barampuram, S. and Z. J. Zhang (2011). "Recent advances in plant transformation." Methods Mol Biol 701: 1-35.

Belfort, M. and P. S. Perlman (1995). "Mechanisms of intron mobility." J Biol Chem 270(51): 30237-3024

Belfort, M. and R. J. Roberts (1997). "Homing endonucleases: keeping the house in order." Nucleic Acids Res 25(17): 3379-3388.

Berg, J. M. (1988). "Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins." Proc Natl Acad Sci U S A 85(1): 99-102.

Beumer, K., et al. (2006). "Efficient gene targeting in Drosophila with zinc-finger nucleases." Genetics 172(4): 2391-2403.

Beurdeley, M., et al. (2013). "Compact designer TALENs for efficient genome engineering." Nat Commun 4: 1762.

Bhakta, M. S., et al. (2013). "Highly active zinc-finger nucleases by extended modular assembly." Genome Res 23(3): 530-538.

Bibikova, M., et al. (2003). "Enhancing gene targeting with designed zinc finger nucleases." Science 300(5620): 764.

Bibikova, M., et al. (2001). "Stimulation of homologous recombination through targeted cleavage by chimeric nucleases." Mol Cell Biol 21(1): 289-297.

Bibikova, M., et al. (2002). "Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases." Genetics 161(3): 1169-1175.

Bitinaite, J., et al. (1998). "FokI dimerization is required for DNA cleavage." Proc Natl Acad Sci U S A 95(18): 10570-10575.

Boch, J. and U. Bonas (2010). "Xanthomonas AvrBs3 family-type III effectors: discovery and function." Annu Rev Phytopathol 48: 419-436.

Boch, J., et al. (2009). "Breaking the code of DNA binding specificity of TAL-type III effectors." Science 326(5959): 1509-1512.

Boissel, S., et al. (2013). "megaTALs: a rare-cleaving nuclease architecture for therapeutic genome engineering." Nucleic Acids Res.

Bos, J. L., et al. (1978). "An insert in the single gene for the large ribosomal RNA in yeast mitochondrial DNA." Nature 275(5678): 336-338.

Briggs, A. W., et al. (2012). "Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers." Nucleic Acids Res 40(15): e117.

Brown, D. D. (1984). "The role of stable complexes that repress and activate eukaryotic genes." Philos Trans R Soc Lond B Biol Sci 307(1132): 297-299.

Buerstedde, J. M., et al. (1990). "Light chain gene conversion continues at high rate in an ALV-induced cell line." Embo j 9(3): 921-927.

Buerstedde, J. M. and S. Takeda (1991). "Increased ratio of targeted to random integration after transfection of chicken B cell lines." Cell 67(1): 179-188.

Capecchi, M. R. (1989). "Altering the genome by homologous recombination." Science 244(4910): 1288-1292.

Carlson, D. F., et al. (2012). "Efficient TALEN-mediated gene knockout in livestock." Proc Natl Acad Sci U S A 109(43): 17382-17387.

Carroll, D., et al. (2006). "Design, construction and in vitro testing of zinc finger nucleases." Nat Protoc 1(3): 1329-1341.

Cermak, T., et al. (2011). "Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting." Nucleic Acids Res 39(12): e82.

Certo, M. T., et al. (2012). "Coupling endonucleases with DNA end-processing enzymes to drive gene disruption." Nat Methods 9(10): 973-975.

Chames, P., et al. (2005). "In vivo selection of engineered homing endonucleases using double-strand break induced homologous recombination." Nucleic Acids Res 33(20): e178.

Chen, J., et al. (2014). "An efficient antiviral strategy for targeting hepatitis B virus genome using transcription activator-like effector nucleases." Mol Ther 22(2): 303-311.

Chen, L. Q., et al. (2010). "Sugar transporters for intercellular exchange and nutrition of pathogens." Nature 468(7323): 527-532.

Chen, L. Q., et al. (2012). "Sucrose efflux mediated by SWEET proteins as a key step for phloem transport." Science 335(6065): 207-211.

Chen, Z., et al. (2009). "Directed evolution of homing endonuclease I-SceI with altered sequence specificity." Protein Eng Des Sel 22(4): 249-256.

Cheng, Z., et al. (2013). "Conditional targeted genome editing using somatically expressed TALENs in C. elegans." Nat Biotechnol 31(10): 934-937.

Chevalier, B., et al. (2003). "Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-CreI and I-MsoI." J Mol Biol 329(2): 253-269.

Chevalier, B. S. and B. L. Stoddard (2001). "Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility." Nucleic Acids Res 29(18): 3757-3774.

Choo, Y. and M. Isalan (2000). "Advances in zinc finger engineering." Curr Opin Struct Biol 10(4): 411-416.

Choo, Y. and A. Klug (1994). "Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage." Proc Natl Acad Sci U S A 91(23): 11163-11167.

Choo, Y., et al. (1994). "In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence." Nature 372(6507): 642-645.

Choulika, A., et al. (1995). "Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae." Mol Cell Biol 15(4): 1968-1973.

Christian, M., et al. (2013). "Targeted mutagenesis of Arabidopsis thaliana using engineered TAL effector nucleases." G3 (Bethesda) 3(10): 1697-1705.

Christian, M. L., et al. (2012). "Targeting G with TAL effectors: a comparison of activities of TALENs constructed with NN and NK repeat variable di-residues." PLoS One 7(9): e45383.

Colleaux, L., et al. (1986). "Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into E. coli as a specific double strand endonuclease." Cell 44(4): 521-533.

Colleaux, L., et al. (1988). "Recognition and cleavage site of the intron-encoded omega transposase." Proc Natl Acad Sci U S A 85(16): 6022-6026.

Cong, L., et al. (2012). "Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains." Nat Commun 3: 968.

Crocker, J. and D. L. Stern (2013). "TALE-mediated modulation of transcriptional enhancers in vivo." Nat Methods 10(8): 762-767.

Dahlem, T. J., et al. (2012). "Simple methods for generating and detecting locus-specific mutations induced with TALENs in the zebrafish genome." PLoS Genet 8(8): e1002861.

Davies, B., et al. (2013). "Site specific mutation of the Zic2 locus by microinjection of TALEN mRNA in mouse CD1, C3H and C57BL/6J oocytes." PLoS One 8(3): e60216.

Davis, L. and N. Maizels (2011). "DNA nicks promote efficient and safe targeted gene correction." PLoS One 6(9): e23981.

de Lange, O., et al. (2013). "Breaking the DNA-binding code of Ralstonia solanacearum TAL effectors provides new possibilities to generate plant resistance genes against bacterial wilt disease." New Phytol 199(3): 773-786.

Deng, D., et al. (2012). "Structural basis for sequence-specific recognition of DNA by TAL effectors." Science 335(6069): 720-723.

Deng, D., et al. (2012). "Recognition of methylated DNA by TAL effectors." Cell Res 22(10): 1502-1504.

Ding, Q., et al. (2013). "A TALEN genome-editing system for generating human stem cell-based disease models." Cell Stem Cell 12(2): 238-251.

Doetschman, T., et al. (1987). "Targetted correction of a mutant HPRT gene in mouse embryonic stem cells." Nature 330(6148): 576-578.

Doetschman, T., et al. (1988). "Targeted mutation of the Hprt gene in mouse embryonic stem cells." Proc Natl Acad Sci U S A 85(22): 8583-8587.

Doyle, E. L., et al. (2013). "TAL Effector Specificity for base 0 of the DNA Target Is Altered in a Complex, Effector- and Assay-Dependent Manner by Substitutions for the Tryptophan in Cryptic Repeat -1." PLoS One 8(12): e82120.

Doyon, J. B., et al. (2006). "Directed evolution and substrate specificity profile of homing endonuclease I-SceI." J Am Chem Soc 128(7): 2477-2484.

Doyon, Y., et al. (2008). "Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases." Nat Biotechnol 26(6): 702-708.

Doyon, Y., et al. (2011). "Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures." Nat Methods 8(1): 74-79.

Dujon, B. (1980). "Sequence of the intron and flanking exons of the mitochondrial 21S rRNA gene of yeast strains having different alleles at the omega and rib-1 loci." Cell 20(1): 185-197.

Dujon, B. (1989). "Group I introns as mobile genetic elements: facts and mechanistic speculations--a review." Gene 82(1): 91-114.

Dujon, B., et al. (1974). "Mitochondrial genetics IX: A model for recombination and segregation of mitochondrial genomes in saccharomyces cerevisiae." Genetics 78(1): 415-437.

Edgell, D. R., et al. (2004). "Coincidence of cleavage sites of intron endonuclease I-TevI and critical sequences of the host thymidylate synthase gene." J Mol Biol 343(5): 1231-1241.

Elrod-Erickson, M., et al. (1998). "High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition." Structure 6(4): 451-464.

Fairall, L., et al. (1993). "The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition." Nature 366(6454): 483-487.

Fanucchi, S., et al. (2013). "Chromosomal contact permits transcription between coregulated genes." Cell 155(3): 606-620.

Faye, G., et al. (1979). "Inserted sequence in the mitochondrial 23S ribosomal RNA gene of the yeast Saccharomyces cerevisiae." Mol Gen Genet 168(1): 101-109.

Fonfara, I., et al. (2012). "Creating highly specific nucleases by fusion of active restriction endonucleases and catalytically inactive homing endonucleases." Nucleic Acids Res 40(2): 847-860.

Gabsalilow, L., et al. (2013). "Site- and strand-specific nicking of DNA by fusion proteins derived from MutH and I-SceI or TALE repeats." Nucleic Acids Res 41(7): e83.

Gaj, T., et al. (2012). "Targeted gene knockout by direct delivery of zinc-finger nuclease proteins." Nat Methods 9(8): 805-807.

Gimble, F. S., et al. (2003). "Assessing the plasticity of DNA target site recognition of the PI-SceI homing endonuclease using a bacterial two-hybrid selection system." J Mol Biol 334(5): 993-1008.

Ginsberg, A. M., et al. (1984). "Xenopus 5S gene transcription factor, TFIIIA: characterization of a cDNA clone and measurement of RNA levels throughout development." Cell 39(3 Pt 2): 479-489.

Greisman, H. A. and C. O. Pabo (1997). "A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites." Science 275(5300): 657-661.

Gruen, M., et al. (2002). "An in vivo selection system for homing endonuclease activity." Nucleic Acids Res 30(7): e29.

Guo, J., et al. (2010). "Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases." J Mol Biol 400(1): 96-107.

Gupta, A., et al. (2013). "Targeted chromosomal deletions and inversions in zebrafish." Genome Res 23(6): 1008-1017.

Halfter, U., et al. (1992). "Gene targeting in Arabidopsis thaliana." Mol Gen Genet 231(2): 186-193.

Heuer, H., et al. (2007). "Repeat domain diversity of avrBs3-like genes in Ralstonia solanacearum strains and association with host preferences in the field." Appl Environ Microbiol 73(13): 4379-4384.

Hinnen, A., et al. (1978). "Transformation of yeast." Proc Natl Acad Sci U S A 75(4): 1929-1933.

Holkers, M., et al. (2013). "Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells." Nucleic Acids Res 41(5): e63.

Hu, R., et al. (2013). "Targeting human microRNA genes using engineered Tal-effector nucleases (TALENs)." PLoS One 8(5): e63074.

Huang, B., et al. (1996). "Splase: a new class IIS zinc-finger restriction endonuclease with specificity for Sp1 binding sites." J Protein Chem 15(5): 481-489.

Huang, P., et al. (2011). "Heritable gene targeting in zebrafish using customized TALENs." Nat Biotechnol 29(8): 699-700.

Hummel, A. W., et al. (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 Phytol 195(4): 883-893.

Iiizumi, S., et al. (2008). "Impact of non-homologous end-joining deficiency on random and targeted DNA integration: implications for gene targeting." Nucleic Acids Res 36(19): 6333-6342.

Isalan, M., et al. (1998). "Comprehensive DNA recognition through concerted interactions from adjacent zinc fingers." Biochemistry 37(35): 12026-12033.

Jacquier, A. and B. Dujon (1985). "An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene." Cell 41(2): 383-394.

Jasin, M., et al. (1996). "Targeted transgenesis." Proc Natl Acad Sci U S A 93(17): 8804-8808.

Johnson, R. A., et al. (2013). "A rapid assay to quantify the cleavage efficiency of custom-designed nucleases in planta." Plant Mol Biol 82(3): 207-221.

Joung, J. K., et al. (2000). "A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions." Proc Natl Acad Sci U S A 97(13): 7382-7387.

Kato, T., et al. (2013). "Production of Sry knockout mouse using TALEN via oocyte injection." Sci Rep 3: 3136.

Kay, S., et al. (2007). "A bacterial effector acts as a plant transcription factor and induces a cell size regulator." Science 318(5850): 648-651.

Kim, E., et al. (2012). "Precision genome engineering with programmable DNA-nicking enzymes." Genome Res 22(7): 1327-1333.

Kim, H. J., et al. (2009). "Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly." Genome Res 19(7): 1279-1288.

Kim, S., et al. (2011). "Preassembled zinc-finger arrays for rapid construction of ZFNs." Nat Methods 8(1): 7.

Kim, Y., et al. (2013). "A library of TAL effector nucleases spanning the human genome." Nat Biotechnol 31(3): 251-258.

Kim, Y. G., et al. (1996). "Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain." Proc Natl Acad Sci U S A 93(3): 1156-1160.

Kim, Y. G. and S. Chandrasegaran (1994). "Chimeric restriction endonuclease." Proc Natl Acad Sci U S A 91(3): 883-887.

Kim, Y. G., et al. (1997). "Construction of a Z-DNA-specific restriction endonuclease." Proc Natl Acad Sci U S A 94(24): 12875-12879.

Kim, Y. G., et al. (1998). "Chimeric restriction enzyme: Gal4 fusion to FokI cleavage domain." Biol Chem 379(4-5): 489-495.

Kim, Y. K., et al. (2013). "TALEN-based knockout library for human microRNAs." Nat Struct Mol Biol 20(12): 1458-1464.

Kleinstiver, B. P., et al. (2012). "Monomeric site-specific nucleases for genome editing." Proc Natl Acad Sci U S A 109(21): 8061-8066.

Koller, B. H. and O. Smithies (1992). "Altering genes in animals by gene targeting." Annu Rev Immunol 10: 705-730.

Lamb, B. M., et al. (2013). "Directed evolution of the TALE N-terminal domain for recognition of all 5' bases." Nucleic Acids Res 41(21): 9779-9785.

Lee, K. Y., et al. (1990). "Homologous recombination in plant cells after Agrobacterium-mediated transformation." Plant Cell 2(5): 415-425.

Lee, M. S., et al. (1989). "Three-dimensional solution structure of a single zinc finger DNA-binding domain." Science 245(4918): 635-637.

Li, L., et al. (2013). "Characterization and DNA-binding specificities of Ralstonia TAL-like effectors." Mol Plant 6(4): 1318-1330.

Li, L. and S. Chandrasegaran (1993). "Alteration of the cleavage distance of Fok I restriction endonuclease by insertion mutagenesis." Proc Natl Acad Sci U S A 90(7): 2764-2768.

Li, L., et al. (1992). "Functional domains in Fok I restriction endonuclease." Proc Natl Acad Sci U S A 89(10): 4275-4279.

Li, L., et al. (1993). "C-terminal deletion mutants of the FokI restriction endonuclease." Gene 133(1): 79-84.

Li, T., et al. (2011). "TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain." Nucleic Acids Res 39(1): 359-372.

Li, T., et al. (2011). "Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes." Nucleic Acids Res 39(14): 6315-6325.

Li, T., et al. (2012). "High-efficiency TALEN-based gene editing produces disease-resistant rice." Nat Biotechnol 30(5): 390-392.

Liang, J., et al. (2013). "FairyTALE: A High-Throughput TAL Effector Synthesis Platform." ACS Synth Biol.

Liang, Z., et al. (2013). "Targeted Mutagenesis in Zea mays Using TALENs and the CRISPR/Cas System." Journal of Genetics and Genomics.

Lieber, M. R. (2010). "The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway." Annu Rev Biochem 79: 181-211.

Lin, Y., et al. (2014). "SAPTA: a new design tool for improving TALE nuclease activity." Nucleic Acids Res.

Liu, J., et al. (2014). "Cell-Penetrating Peptide-Mediated Delivery of TALEN Proteins via Bioconjugation for Genome Engineering." PLoS One 9(1): e85755.

Liu, P. Q., et al. (2001). "Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A." J Biol Chem 276(14): 11323-11334.

Liu, Y., et al. (2013). "Inheritable and precise large genomic deletions of non-coding RNA genes in zebrafish using TALENs." PLoS One 8(10): e76387.

Lloyd, A., et al. (2005). "Targeted mutagenesis using zinc-finger nucleases in Arabidopsis." Proc Natl Acad Sci U S A 102(6): 2232-2237.

Ma, S., et al. (2012). "Highly efficient and specific genome editing in silkworm using custom TALENs." PLoS One 7(9): e45035.

Maeder, M. L., et al. (2013). "Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins." Nat Biotechnol 31(12): 1137-1142.

Maeder, M. L., et al. (2013). "Robust, synergistic regulation of human gene expression using TALE activators." Nat Methods 10(3): 243-245.

Maeder, M. L., et al. (2008). "Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification." Mol Cell 31(2): 294-301.

Mahfouz, M. M., et al. (2012). "Targeted transcriptional repression using a chimeric TALE-SRDX repressor protein." Plant Mol Biol 78(3): 311-321.

Mahfouz, M. M., et al. (2011). "De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks." Proc Natl Acad Sci U S A 108(6): 2623-2628.

Mak, A. N., et al. (2012). "The crystal structure of TAL effector PthXo1 bound to its DNA target." Science 335(6069): 716-719.

Mashimo, T., et al. (2013). "Efficient gene targeting by TAL effector nucleases coinjected with exonucleases in zygotes." Sci Rep 3: 1253.

Meckler, J. F., et al. (2013). "Quantitative analysis of TALE-DNA interactions suggests polarity effects." Nucleic Acids Res 41(7): 4118-4128.

Meng, X., et al. (2007). "Profiling the DNA-binding specificities of engineered Cys2His2 zinc finger domains using a rapid cell-based method." Nucleic Acids Res 35(11): e81.

Mercer, A. C., et al. (2012). "Chimeric TALE recombinases with programmable DNA sequence specificity." Nucleic Acids Res 40(21): 11163-11172.

Metzger, M. J., et al. (2011). "Single-strand nicks induce homologous recombination with less toxicity than double-strand breaks using an AAV vector template." Nucleic Acids Res 39(3): 926-935.

Miao, Z. H. and E. Lam (1995). "Targeted disruption of the TGA3 locus in Arabidopsis thaliana." Plant J 7(2): 359-365.

Miller, J., et al. (1985). "Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes." Embo j 4(6): 1609-1614.

Miller, J. C., et al. (2007). "An improved zinc-finger nuclease architecture for highly specific genome editing." Nat Biotechnol 25(7): 778-785.

Miller, J. C., et al. (2011). "A TALE nuclease architecture for efficient genome editing." Nat Biotechnol 29(2): 143-148.

Mladenov, E. and G. Iliakis (2011). "Induction and repair of DNA double strand breaks: the increasing spectrum of non-homologous end joining pathways." Mutat Res 711(1-2): 61-72.

Monteilhet, C., et al. (1990). "Purification and characterization of the in vitro activity of I-Sce I, a novel and highly specific endonuclease encoded by a group I intron." Nucleic Acids Res 18(6): 1407-1413.

Moore, M., et al. (2001). "Design of polyzinc finger peptides with structured linkers." Proc Natl Acad Sci U S A 98(4): 1432-1436.

Moore, M., et al. (2001). "Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units." Proc Natl Acad Sci U S A 98(4): 1437-1441.

Moscou, M. J. and A. J. Bogdanove (2009). "A simple cipher governs DNA recognition by TAL effectors." Science 326(5959): 1501.

Mussolino, C. and T. Cathomen (2011). "On target? Tracing zinc-finger-nuclease specificity." Nat Methods 8(9): 725-726.

Mussolino, C., et al. (2011). "A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity." Nucleic Acids Res 39(21): 9283-9293.

Nakaseko, Y., et al. (1992). "Adjacent zinc-finger motifs in multiple zinc-finger peptides from SWI5 form structurally independent, flexibly linked domains." J Mol Biol 228(2): 619-636.

Netter, P., et al. (1974). "Mitochondrial genetics. VII. Allelism and mapping studies of ribosomal mutants resistant to chloramphenicol, erythromycin and spiramycin in S. cerevisiae." Genetics 78(4): 1063-1100.

Neuhaus, D., et al. (1992). "Solution structures of two zinc-finger domains from SWI5 obtained using two-dimensional 1H nuclear magnetic resonance spectroscopy. A zinc-finger structure with a third strand of beta-sheet." J Mol Biol 228(2): 637-651.

Nyquist, M. D., et al. (2013). "TALEN-engineered AR gene rearrangements reveal endocrine uncoupling of androgen receptor in prostate cancer." Proc Natl Acad Sci U S A 110(43): 17492-17497.

Offringa, R., et al. (1990). "Extrachromosomal homologous recombination and gene targeting in plant cells after Agrobacterium mediated transformation." Embo j 9(10): 3077-3084.

Oleykowski, C. A., et al. (1998). "Mutation detection using a novel plant endonuclease." Nucleic Acids Res 26(20): 4597-4602.

Orr-Weaver, T. L., et al. (1981). "Yeast transformation: a model system for the study of recombination." Proc Natl Acad Sci U S A 78(10): 6354-6358.

Orr-Weaver, T. L., et al. (1983). "Genetic applications of yeast transformation with linear and gapped plasmids." Methods Enzymol 101: 228-245.

Ousterout, D. G., et al. (2013). "Reading frame correction by targeted genome editing restores dystrophin expression in cells from Duchenne muscular dystrophy patients." Mol Ther 21(9): 1718-1726.

Panda, S. K., et al. (2013). "Highly efficient targeted mutagenesis in mice using TALENs." Genetics 195(3): 703-713.

Papworth, M., et al. (2003). "Inhibition of herpes simplex virus 1 gene expression by designer zinc-finger transcription factors." Proc Natl Acad Sci U S A 100(4): 1621-1626.

Paszkowski, J., et al. (1988). "Gene targeting in plants." Embo j 7(13): 4021-4026.

Pavletich, N. P. and C. O. Pabo (1991). "Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A." Science 252(5007): 809-817.

Pellicer, J., et al. (2010). "The largest eukaryotic genome of them all?" Botanical Journal of the Linnean Society 164(1): 10-15.

Perez-Pinera, P., et al. (2013). "Synergistic and tunable human gene activation by combinations of synthetic transcription factors." Nat Methods 10(3): 239-242.

Pernstich, C. and S. E. Halford (2012). "Illuminating the reaction pathway of the FokI restriction endonuclease by fluorescence resonance energy transfer." Nucleic Acids Res 40(3): 1203-1213.

Piganeau, M., et al. (2013). "Cancer translocations in human cells induced by zinc finger and TALE nucleases." Genome Res 23(7): 1182-1193.

Pingoud, A., et al. (2005). "Type II restriction endonucleases: structure and mechanism." Cell Mol Life Sci 62(6): 685-707.

Politz, M. C., et al. (2013). "Artificial repressors for controlling gene expression in bacteria." Chem Commun (Camb) 49(39): 4325-4327.

Porteus, M. H. and D. Baltimore (2003). "Chimeric nucleases stimulate gene targeting in human cells." Science 300(5620): 763.

Pruett-Miller, S. M., et al. (2008). "Comparison of zinc finger nucleases for use in gene targeting in mammalian cells." Mol Ther 16(4): 707-717.

Pruett-Miller, S. M., et al. (2009). "Attenuation of zinc finger nuclease toxicity by small-molecule regulation of protein levels." PLoS Genet 5(2): e1000376.

Puchta, H., et al. (1993). "Homologous recombination in plant cells is enhanced by in vivo induction of double strand breaks into DNA by a site-specific endonuclease." Nucleic Acids Res 21(22): 5034-5040.

Puchta, H., et al. (1996). "Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination." Proc Natl Acad Sci U S A 93(10): 5055-5060.

Ramirez, C. L., et al. (2012). "Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects." Nucleic Acids Res 40(12): 5560-5568.

Rebar, E. J. and C. O. Pabo (1994). "Zinc finger phage: affinity selection of fingers with new DNA-binding specificities." Science 263(5147): 671-673.

Reynolds, L., et al. (2003). "Repression of the HIV-1 5' LTR promoter and inhibition of HIV-1 replication by using engineered zinc-finger transcription factors." Proc Natl Acad Sci U S A 100(4): 1615-1620.

Reyon, D., et al. (2012). "FLASH assembly of TALENs for high-throughput genome editing." Nat Biotechnol 30(5): 460-465.

Risseeuw, E., et al. (1995). "Targeted recombination in plants using Agrobacterium coincides with additional rearrangements at the target locus." Plant J 7(1): 109-119.

Romer, P., et al. (2007). "Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene." Science 318(5850): 645-648.

Rosen, L. E., et al. (2006). "Homing endonuclease I-CreI derivatives with novel DNA target specificities." Nucleic Acids Res 34(17): 4791-4800.

Rothstein, R. J. (1983). "One-step gene disruption in yeast." Methods Enzymol 101: 202-211.

Rouet, P., et al. (1994). "Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells." Proc Natl Acad Sci U S A 91(13): 6064-6068.

Ru, R., et al. (2013). "Targeted genome engineering in human induced pluripotent stem cells by penetrating TALENs." Cell Regeneration 2(1): 1-8.

San Filippo, J., et al. (2008). "Mechanism of eukaryotic homologous recombination." Annu Rev Biochem 77: 229-257.

Sander, J. D., et al. (2011). "Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA)." Nat Methods 8(1): 67-69.

Sargent, R. G., et al. (1997). "Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination." Mol Cell Biol 17(1): 267-277.

Schierling, B., et al. (2012). "A novel zinc-finger nuclease platform with a sequence-specific cleavage module." Nucleic Acids Res 40(6): 2623-2638.

Schmid-Burgk, J. L., et al. (2013). "A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes." Nat Biotechnol 31(1): 76-81.

Schornack, S., et al. (2008). "Characterization of AvrHah1, a novel AvrBs3-like effector fromXanthomonas gardneriwith virulence and avirulence activity." New Phytologist 179(2): 546-556.

Scott, J. N., et al. (2014). "TALE proteins bind to both active and inactive chromatin." Biochem J 458(1): 153-158.

Sedivy, J. M. and P. A. Sharp (1989). "Positive genetic selection for gene disruption in mammalian cells by homologous recombination." Proc Natl Acad Sci U S A 86(1): 227-231.

Segal, D. J. and D. Carroll (1995). "Endonuclease-induced, targeted homologous extrachromosomal recombination in Xenopus oocytes." Proc Natl Acad Sci U S A 92(3): 806-810.

Segal, D. J., et al. (1999). "Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences." Proc Natl Acad Sci U S A 96(6): 2758-2763.

Seligman, L. M., et al. (2002). "Mutations altering the cleavage specificity of a homing endonuclease." Nucleic Acids Res 30(17): 3870-3879.

Shan, Q., et al. (2013). "Rapid and efficient gene modification in rice and Brachypodium using TALENs." Mol Plant 6(4): 1365-1368.

Shrivastav, M., et al. (2008). "Regulation of DNA double-strand break repair pathway choice." Cell Res 18(1): 134-147.

Shukla, V. K., et al. (2009). "Precise genome modification in the crop species Zea mays using zinc-finger nucleases." Nature 459(7245): 437-441.

Sizova, I., et al. (2013). "Nuclear gene targeting in Chlamydomonas using engineered zinc-finger nucleases." Plant J 73(5): 873-882.

Smith, J., et al. (1999). "A detailed study of the substrate specificity of a chimeric restriction enzyme." Nucleic Acids Res 27(2): 674-681.

Smith, J., et al. (2000). "Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains." Nucleic Acids Res 28(17): 3361-3369.

Smith, J., et al. (2006). "A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences." Nucleic Acids Res 34(22): e149.

Smithies, O., et al. (1985). "Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination." Nature 317(6034): 230-234.

Stoilov, L. M., et al. (1996). "Restriction endonucleases induce chromosomal aberrations in barley." Mutagenesis 11(1): 119-123.

Streubel, J., et al. (2012). "TAL effector RVD specificities and efficiencies." Nat Biotechnol 30(7): 593-595.

Stroud, D. A., et al. (2013). "Gene knockout using transcription activator-like effector nucleases (TALENs) reveals that human NDUFA9 protein is essential for stabilizing

the junction between membrane and matrix arms of complex I." J Biol Chem 288(3): 1685-1690.

Sugisaki, H. and S. Kanazawa (1981). "New restriction endonucleases from Flavobacterium okeanokoites (FokI) and Micrococcus luteus (MluI)." Gene 16(1-3): 73-78.

Sun, N., et al. (2012). "Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease." Mol Biosyst 8(4): 1255-1263.

Sun, N. and H. Zhao (2014). "A single-chain TALEN architecture for genome engineering." Mol Biosyst 10(3): 446-453.

Sussman, D., et al. (2004). "Isolation and characterization of new homing endonuclease specificities at individual target site positions." J Mol Biol 342(1): 31-41.

Szczepek, M., et al. (2007). "Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases." Nat Biotechnol 25(7): 786-793.

Szostak, J. W., et al. (1983). "The double-strand-break repair model for recombination." Cell 33(1): 25-35.

Takada, S., et al. (2013). "Targeted gene deletion of miRNAs in mice by TALEN system." PLoS One 8(10): e76004.

Tesson, L., et al. (2011). "Knockout rats generated by embryo microinjection of TALENs." Nat Biotechnol 29(8): 695-696.

Thacker, J. (1994). "The study of responses to 'model' DNA breaks induced by restriction endonucleases in cells and cell-free systems: achievements and difficulties." Int J Radiat Biol 66(5): 591-596.

Thomas, K. R. and M. R. Capecchi (1987). "Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells." Cell 51(3): 503-512.

Townsend, J. A., et al. (2009). "High-frequency modification of plant genes using engineered zinc-finger nucleases." Nature 459(7245): 442-445.

Treen, N., et al. (2014). "Tissue-specific and ubiquitous gene knockouts by TALEN electroporation provide new approaches to investigating gene function in Ciona." Development 141(2): 481-487.

Tsuji, S., et al. (2013). "Creating a TALE protein with unbiased 5'-T binding." Biochem Biophys Res Commun 441(1): 262-265.

Valton, J., et al. (2012). "Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation." J Biol Chem 287(46): 38427-38432.

van Nierop, G. P., et al. (2009). "Stimulation of homology-directed gene targeting at an endogenous human locus by a nicking endonuclease." Nucleic Acids Res 37(17): 5725-5736.

Vanamee, E. S., et al. (2001). "FokI requires two specific DNA sites for cleavage." J Mol Biol 309(1): 69-78.

Wah, D. A., et al. (1998). "Structure of FokI has implications for DNA cleavage." Proc Natl Acad Sci U S A 95(18): 10564-10569.

Wah, D. A., et al. (1997). "Structure of the multimodular endonuclease FokI bound to DNA." Nature 388(6637): 97-100.

Wang, H., et al. (2013). "TALEN-mediated editing of the mouse Y chromosome." Nat Biotechnol 31(6): 530-532.

Wang, J., et al. (2012). "Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme." Genome Res 22(7): 1316-1326.

Wei, C., et al. (2013). "TALEN or Cas9 - rapid, efficient and specific choices for genome modifications." J Genet Genomics 40(6): 281-289.

Wei, Q., et al. (2014). "Rapid Creation of Forward-Genetics Tools for C. briggsae Using TALENs: Lessons for Nonmodel Organisms." Mol Biol Evol 31(2): 468-473.

Wendt, T., et al. (2013). "TAL effector nucleases induce mutations at a pre-selected location in the genome of primary barley transformants." Plant Mol Biol 83(3): 279-285.

Wood, A. J., et al. (2011). "Targeted genome editing across species using ZFNs and TALENs." Science 333(6040): 307.

Wright, D. A., et al. (2005). "High-frequency homologous recombination in plants mediated by zinc-finger nucleases." Plant J 44(4): 693-705.

Yang, B., et al. (2000). "The virulence factor AvrXa7 of Xanthomonas oryzae pv. oryzae is a type III secretion pathway-dependent nuclear-localized double-stranded DNA-binding protein." Proc Natl Acad Sci U S A 97(17): 9807-9812.

Yang, J., et al. (2013). "ULtiMATE system for rapid assembly of customized TAL effectors." PLoS One 8(9): e75649.

Yanik, M., et al. (2013). "TALE-PvuII Fusion Proteins - Novel Tools for Gene Targeting." PLoS One 8(12): e82539.

Zhang, Y., et al. (2013). "Transcription activator-like effector nucleases enable efficient plant genome engineering." Plant Physiol 161(1): 20-27.

Zhang, Z., et al. (2013). "A simple and efficient method for assembling TALE protein based on plasmid library." PLoS One 8(6): e66459.

Zhu, F., et al. (2013). "DICE, an efficient system for iterative genomic editing in human pluripotent stem cells." Nucleic Acids Res.

Zhu, W., et al. (1999). "The C terminus of AvrXa10 can be replaced by the transcriptional activation domain of VP16 from the herpes simplex virus." Plant Cell 11(9): 1665-1674

Zinn, A. R. and R. A. Butow (1985). "Nonreciprocal exchange between alleles of the yeast mitochondrial 21S rRNA gene: kinetics and the involvement of a double-strand break." Cell 40(4): 887-895.

Zhang Z, Li D, Xu H, Xin Y, Zhang T, Ma L, Wang X, Chen Z, Zhang Z. A simple and efficient method for assembling TALE protein based on plasmid library. PLoS One. 8(6):e66459.

Zhu F, Gamboa M, Farruggio AP, Hippenmeyer S, Tasic B, Schüle B, Chen-Tsai Y, Calos MP. 2013. DICE, an efficient system for iterative genomic editing in human pluripotent stem cells. Nucleic Acids Res. 2013 Dec 4. [Epub ahead of print]

Zhu, W, B Yang, JM Chittoor, LB Johnson and FF White. 1998. AvrXa10 contains an acidic transcriptional activation domain in the functionally conserved C-terminus. Molecular Plant-Microbe Interactions 11:824 –832.

Zhu, W, B Yang, N. Wills, L B Johnson and FF White. 1999. The C terminus of AvrXa10 can be replaced by the transcriptional activation domain of VP16 from the Herpes Simplex Virus. Plant Cell 11:1665 –1674.

ACKNOWLEDGEMENTS

I would like to thank my major professor, Dr. Bing Yang, for his invaluable guidance, encouragement and support throughout the course of this research. His scientific spirits inspire me and help me to build experimental skills and develop critical thinking skills.

I also express my sincere appreciation to other members of my POS committee, Dr. Diane C. Bassham, Dr. Jeffrey J. Essner, Dr. Martin H. Spalding, and Dr. Yanhai Yin for their helpful advice.

In addition, I wish to thank my collaborators for the productive research as well as all other members of the Yang Lab, past and present, Bo Liu, Dr. Sheng Huang, Junhui Zhou, Dr. Congfeng Song, Dr. Huanbin Zhou, Honghao Bi, Sinian Char, Dr. Juying Long, Zhiyuan Ji, Dr. Fei Lin, Dr. Liang Wu, Chih-Ying Chen for their scientific discussion and friendship. I am fortunate to have such an excellent working environment.

Finally, my deep gratitude also goes to my family for encouragement, especially to my husband Changfu Yao and my baby Ada X. Yao for standing by me and supporting me throughout my dissertation work. Without them, I cannot achieve what I have accomplished in my research.