TALENs and CRISPR/Cas9 for Rice-Genome Editing

Bing Yang
Iowa State University
Ames, Iowa

byang@iastate.edu

Rice, *Oryza sativa* L., is an important staple crop that feeds more than half of the world's population. Cultivated for more than 5,000 years, a huge number of natural and cultivated genotypes exist, representing the richest genetic resource among crop species. Rice also serves as a model species for cereal crops due to its syntenic relationship with other grass species (Gale and Devos, 1998). It was the first crop to have its genome sequenced (Goff *et al.*, 2002; Yu *et al.*, 2002), and the genome sequence of cultivar Nipponbare has been well annotated (Sakai *et al.*, 2013). Rice has been a subject of genetic manipulation by a variety of approaches, including "traditional" mutagenesis (*i.e.* chemical- and radiation-induced mutation) and forward and reverse genetics.

Research in our laboratory has been focusing on host/microbe interactions by using bacterial blight of rice as a model. The pathogen, *Xanthomonas oryzae* pv. *oryzae*, causes an important disease in Asia and Western Africa, resulting in severe losses in rice-grain production (Nino-Liu *et al.*, 2006).

The objectives of our research include:

- identification of virulence factors that facilitate the pathogen's ability to colonize the host and develop symptoms,
- identification of host targets for the pathogen's virulence factors that cause disease susceptibility,
- elucidation of the molecular mechanism of disease susceptibility and resistance mediated by the interactions between the pathogen's virulence factors and the host's target genes or gene products, and
- engineering resistance based on the best information obtained from study of the disease, an approach that is widely applicable to other related diseases in crops.

We are particularly interested in a group of bacterial proteins from *Xanthomonas*, transcription-activator-like (TAL) effectors, on which some strains of *Xanthomonas* depend for pathogenesis. More than 100 distinct TAL-effector genes have been cloned or identified

from sequenced bacterial genomes; they are highly conserved, with about 90% identity at the nucleotide or amino acid level (Boch and Bonas, 2010). A typical TAL effector consists of the N-terminal domain, middle repetitive region and C-terminal domain. The N-terminus contains a secretion signal used by the bacterial type-III secretion system to translocate the TAL effectors from bacterial cells into host plant cells. The C-terminus contains three functional nuclear localization motifs for transporting TAL effectors into the nuclei of host cells. The C-terminus also contains a functional trans-activating domain, a characteristic feature of eukaryotic transcription activators. The most striking feature of TAL effectors is the central repeats that are mainly 34 amino acids in length. The repeats are nearly identical except for the two amino acids at positions 12 and 13, the so-called variable di-amino acids. The di-amino acids actually determine the specificity of DNA binding for each repeat. More than 20 types of repeats exist in native TAL effectors, but four types are predominant and each corresponds to one of four nucleotides with NI to A, NG to T, NN to G, and HD to C (using the single-letter code for amino acids[1]) (Boch *et al.*, 2009; Moscou and Bogdanove, 2009). The discovery of the TAL-effector recognition code has two immediate implications:

- the code could be used to predict and validate the DNA sequences of host-target genes involved in bacterial diseases of crops, and
- the code could be used to guide the custom-engineering of DNA-binding proteins or domains with novel specificity.

Fusion Proteins of TAL Effectors and the FokI Nuclease Domain are Active Nucleases

As one of the obvious applications with programmable TAL-effector DNA-binding domains, TAL-effector nucleases (TALENs) have become a promising genetic tool for basic and applied research. We started working on development of TALEN technology by using TAL effectors AvrXa7 and PthXo1 and their respective target sequences. First, the two native TAL effectors were fused with the nuclease domain of the restriction enzyme FokI, generating the fusion proteins (putative TALENs). Second, the known DNA sequences respectively targeted by PthXo1 and AvrXa7 in the promoters of rice genes *Os8N3* and *Os11N3* were fused into a non-functional reporter gene *LacZ*; the two halves of *LacZ* contained the duplicated 120-bp regions separated by the PthXo1- and AvrXa7-targeted sequences in *Os8N3* and *Os11N3*. Third, constructs expressing the fusion proteins of AvrXa7-FokI and PthXo1-FokI were co-expressed with the *LacZ*-containing reporter construct in yeast. If functional as site-specific nucleases, PthXo1-FokI and AvrXa7-FokI would cause DNA double-strand breaks (DSBs) at the spacer regions between the two halves of the *LacZ* gene, and the repair to DSBs would lead to reconstitution of a functional *LacZ* gene and β-galactosidase activity, the gene product of *LacZ*. Our proof-of-concept experiments, indeed, demonstrated the feasibility of producing active TALENs by fusing the TAL effectors as DNA-binding domains and the FokI nuclease domain. The work was published in *Nucleic Acids Research* (Li *et al.*, 2011a).

[1]N=asparagine; I=isoleucine; G=glycine; H=histidine; and D=aspartic acid.

We further demonstrated the ability to engineer novel TAL-effector DNA-binding domains by using four types of TAL-effector repeats that contain NI, NG, NN and HD at positions 12 and 13 of the repeats that preferentially recognize the nucleotides A, T, G and C, respectively. The four types of TAL repeats were assembled by a modular assembly method. Briefly, each of the repeats as independent modules contains a unique 4-bp overhang with single-base polymorphism after digestion with BsmBI at its 5'- and 3'-ends. For construction of an 8-repeat TALEN array recognizing a specific 8-nucleotide sequence, a corresponding repeat is selected from each set of the repeats for the specific nucleotide of that position. By putting 8 separate repeats together through DNA ligation, an array of 8 repeats can be assembled. Similarly, one or two additional 8-repeat arrays can be assembled; further ligation of two or three 8-repeat arrays results in 16 or 24 repeats of novel TAL-effector binding domains and TALENs, once fused with FokI nuclease domain, recognizing the user-chosen 16 or 24 bp of the target sequence.

To demonstrate the capacity of engineered designer TALENs (dTALENs) to modify endogenous gene loci in eukaryotic cells, we chose genes *URA3*, *LYS2* and *ADE2* as the targets in yeast. Five pairs of TALENs were designed and engineered by using our TALEN modular-assembly method with two pairs targeting two sites of *URA3*, two pairs targeting two sites of *LYS2* and one pair targeting one site of *ADE2*. When transferred with each pair of TALEN genes, the yeast cells were selected for mutations that were derived from the repair to the TALEN-induced DSBs on growth media containing chemicals such as 5-fluoroorotic acid (5-FOA, for *ura3*-mutated cells), or α-aminoadipate (α-AA for cells with mutated *lys2*), or containing limiting adenine concentrations that result in the formation of pink colonies formed by *ade2*-mutant cells. The results indicated that each pair of TALENs indeed induced site-specific mutations that were either insertions or deletions in the intended target genes with corresponding phenotypes of yeast cells. The results were published also in *Nucleic Acids Research* (Li *et al.*, 2011b).

TALEN Applications in Rice to Engineer Disease Resistance to Bacterial Blight

Our ultimate goal in developing TALEN technology is to apply it for our basic scientific research and for breeding disease-resistant rice varieties. As a case in point, we focus on bacterial blight of rice. Years of research effort by our group and by scientists in other laboratories around the world allow us to propose a working model for the outcome of rice-blight disease controlled by the interactions between the TAL effectors from the bacterial pathogen *Xanthomonas oryzae* pv. *oryzae* (Xoo) and the host target disease-susceptibility (S) genes.

First, the pathogen produces and translocates its virulence proteins, including TAL effectors, into host cells through a type-III secretion system; once internalized, TAL effectors are localized into the nuclei of the host cells and bind to the promoter elements [effector-binding elements (EBEs)] of S genes; TAL effectors transcriptionally activate the S-gene expression, which leads to more susceptibility of host plants to bacterial infection. It has been discovered that some Xoo isolates depend on the TAL effector PthXo1 to activate *SWEET11*, one of the sugar-transporter genes in rice. Some isolates

use PthXo3, AvrXa7, TalC or Tal5 to induce *SWEET14* for disease susceptibility, whereas some pathogenic isolates utilize *SWEET13* for disease (Yang *et al.*, 2006; Antony *et al.*, 2010; Streubel *et al.*, 2013).

It would be ideal to make a small change at the TAL-effector binding site in the S gene so that the TAL effector would not recognize the target site of the S gene and the modified plants would become resistant to the bacterial infection. Collectively mutating all three S-gene promoters would enable the plants to become durable and broadly resistant, *i.e.* to all of the Xoo strains that depend on the induction of either one or all of the S genes. The precise modification at the intended site of the S gene would be achieved by using TALENs designed to specifically cause DSBs in the S genes. The TALEN-mutagenized S gene would provide disease resistance similarly to a naturally occurring S-gene mutant such as *xa13,* a resistant rice gene that has a mutation at the PthXo1 binding site in the promoter of *SWEET11* (Chu *et al.*, 2006; Yang *et al.*, 2006).

The precise gene mutagenesis using TALENs in rice also involves multiple steps of transgenics (Li *et al.*, 2014). The steps include 1) immature rice embryos are induced to generate embryogenic callus cells in tissue-culture medium; 2) the callus cells (calli) are transformed with a TALEN-gene-containing construct that also has an antibiotic-resistance gene for selection of transgenic cells or plants containing the TALEN construct; 3) the transformed calli are selected on hygromycin, an antibiotic that kills the non-transformed wild-type embryogenic cells; 4) transgenic plantlets are generated from calli that may contain desired mutations at the intended site caused by TALENs; and 5) the transgenic plants are genotyped for site-specific mutations.

To first demonstrate the feasibility of employing TALENs to effect site-specific mutations at the TALEN-binding site of the S gene and thus produce rice plants resistant to bacterial blight, we engineered two pairs of TALENs that targeted the AvrXa7- and PthXo1-binding sites in the promoter of *SWEET14* (or *Os11N3*). We expected that the TALEN-induced mutations that interfered with the inducibility of *SWEET14* by AvrXa7 and PthXo1, would not affect the developmental function of *SWEET14* in rice. The T1 generation of the transgenic rice plants, in the progeny of selfed primary transgenic plants, were genotyped by PCR-amplifying the relevant region and sequencing the amplicons. Indeed, site-specific mutations of small insertions/deletions were found at the promoter of *SWEET14* at high efficiency. Also, the rice plants homozygous for the 4-bp or 9-bp deletion were resistant to bacterial blight when inoculated with Xoo isolates that depended on AvrXa7. Importantly, the construct that contains the transgenic TALEN genes and the hygromycin-resistant gene could be segregated out in some progeny through genetic crossing, resulting in plants that contained only the desired mutations and valuable agronomic traits, but not the transgenes. The results were published in *Nature Biotechnology* (Li *et al.*, 2012).

The transgene-free rice plants allow us to carry out the second round of gene modification by using constructs that contain new TALENs and the hygromycin-resistant gene, in this case by targeting another SWEET S gene, *SWEET11*. One pair of TALENs was engineered based on the PthXo1-recognizing DNA sequence in the promoter of *SWEET11*. Similarly to the production of *SWEET14* mutations, rice plants with muta-

tions occurring at the intended site of *SWEET11* were obtained. Wild-type rice was also used to generate mutations in *SWEET11* with the engineered TALENs. Those plants contained mutations in both *SWEET14* and *SWEET11* promoters as expected, and were subjected to bacterial inoculation with a collection of 95 Xoo isolates. The plants with only *SWEET11* mutations were resistant to 16 of 95 Xoo isolates; plants with only *SWEET14* mutations were resistant to 71 of the 95 isolates, whereas plants with both *SWEET11* and *SWEET14* mutations were resistant to 87 of the 95 Xoo isolates. The results clearly demonstrate the broad spectrum of resistance in rice plants with mutagenesis of two S genes (unpublished data).

We have been using TALEN technology also to generate genetic materials of rice to gain basic understanding of the roles of rice SWEET genes in plant growth, development and production in addition to disease susceptibility. There are 15 SWEET genes that are highly homologous and expressed in different types of tissues. Our goal is to generate knockout plants of individual genes or combination of multiple genes. So far, we have generated knockout plants of 8 SWEET genes. Characterization of the mutant plants and production of the rest of the SWEET genes are in progress.

CRISPR/Cas9 System for Targeted Gene-Editing in Rice

The type-II CRISPR/Cas RNA-guided nucleases are the most recent addition to the tool kit of sequence-specific nucleases. Intense interest has been focused on the CRISPR/Cas9 system from *Streptococcus pyogenes* following initial reports of its successful use for gene editing (Jinek *et al.*, 2012). In this system, Cas9 nuclease coupled with tracrRNA (trans-activating crRNA) can be guided by a ~20 nt guide (or seed) sequence in crRNA (CRISPR RNA) to hybridize with a specific complementary DNA sequence (*i.e.*, the target site) that is followed by a 5'-NGG or 5'-NAG PAM (protospacer adjacent motif) sequence to induce a precise cleavage of the target sequence 3–4 base pairs upstream of the PAM site (Jinek *et al.*, 2012). Alternatively, parts of crRNA and tracrRNA sequence can be fused in a synthetic gene to produce a single-guide RNA (sgRNA) that is equally as effective as the separate crRNA and tracrRNA complex in targeting a specific DNA sequence for Cas9-directed cleavage (Cong *et al.*, 2013; Mali *et al.*, 2013). Unlike ZFNs (zinc-finger nucleases) and TALENs, the CRISPR/Cas9 system, referred to here as Cas9/sgRNA, is DNA-methylation insensitive. It is also more affordable, remarkably easier to use, and well-suited for multiplex gene targeting and high-throughput genome-wide gene editing at a similar or even higher efficiency than ZFNs and TALENs.

We have developed a Gateway-based Cas9/sgRNA system for rice-gene editing. Specifically and briefly, a common Cas9-expressing destination vector and an intermediate vector for cloning up to two oligo-derived sgRNA genes were constructed. The major cloning work involves sequential insertion of two oligo-derived small dsDNAs in the intermediate vector. The resulting sgRNA construct can be easily combined with the master Cas9 binary vector using Gateway clonase into a single construct wherein, in rice cells, Cas9 is expressed under the maize ubiquitin 1 promoter while the sgRNA genes are expressed under the rice U6 promoters. Restriction enzymes BtgZI and BsaI will be used to create 4-bp overhangs downstream of the U6 (or U3) promoter in the

intermediate vector, and the complementary oligos with appropriate 4-bp overhangs, after being annealed to each other, will be phosphorylated, annealed and cloned into the respective pENTR-sgRNA.

As an example, we made a construct expressing Cas9 and the sgRNA-targeting the rice SWEET13-coding region. The construct was transferred into rice embryogenic callus cells and transgenic plants were obtained through tissue culture and transformation similarly to the TALEN work in rice. Nine independent transgenic lines were obtained; PCR-amplification of the relevant region and sequencing of the PCR amplicons revealed a high efficiency of mutagenesis at the targeted site of *SWEET13*. Each line contained the mutations that occurred independently on the two chromosomes (so-called di-allelic mutations), resulting in a mutagenesis frequency of 100%.

We also demonstrated proof-of-efficiency of Cas9/sgRNAs in producing large chromosomal deletions (115–245 kb) involving three clusters of genes in rice protoplasts and verification of deletions of two clusters in regenerated T0 generation plants. Part of our Cas9/sgRNA work has been published in *Nucleic Acids Research* (Zhou *et al.*, 2014).

Conclusions

- We have demonstrated that fusion proteins of the native TALE and the FokI nuclease domain enabled site-specific DSBs (Li *et al.*, 2011a);
- We developed a modular assembly method to engineer designer TAL effectors with novel DNA-binding domains, and the custom-made TALENs were capable of inducing gene editing in yeast (Li *et al.*, 2011b, 2014);
- TALENs were successfully applied to edit the promoters of two disease-susceptibility SWEET genes to render the otherwise susceptible rice resistant to a broad range of bacterial-blight pathogen field isolates (Li *et al.*, 2012);
- CRISPR/Cas9 has been established for genome editing in rice, leading to extremely high efficiency in small/local DNA changes and efficient large chromosomal-segment deletions (Jiang *et al.*, 2013; Zhou *et al.* 2014).

Acknowledgments

I thank colleagues in my laboratory and collaborators in the laboratories of Dr. Marty H. Spalding at Iowa State University, Dr. Frank F. White at Kansas State University and Dr. Donald Weeks at the University of Nebraska-Lincoln. The work presented here has been funded by grants from the US National Science Foundation, the US Department of Agriculture, and from the ISU Research Foundation.

References

Antony G *et al.* (2010) Rice xa13 recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene Os-11N3. The Plant Cell 22 3864–3876.

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

Boch J *et al.* (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326 1509–1512.
Chu Z (2006) Promoter mutations of an essential gene for pollen development result in disease resistance in rice. Genes and Development 20 1250–1255.
Cong L *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339 819–823.
Gale MD Devos KM (1998) Comparative genetics in the grasses. Proceedings of the National Academy of Sciences of the USA 95 1971–1974.
Goff SA *et al.* (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science 296 92–100.
Jiang W *et al.* (2013 Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucleic Acids Research 41(20) e188 doi:10.1093/nar/gkt780.
Jinek M *et al.* (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337 816–821.
Li T *et al* (2011a) TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucleic Acids Research 39 359–372.
Li T *et al.* (2011b) Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. Nucleic Acids Research 39 6315–6325.
Li T *et al.* (2012) High-efficiency TALEN-based gene editing produces disease-resistant rice. Nature Biotechnology 30 390–392.
Li T *et al.* (2014) TALEN utilization in rice genome modifications. Methods 69 9–19.
Mali P *et al.* (2013) RNA-guided human genome engineering via Cas9. Science 339 823–826.
Moscou MJ Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. Science 326 1501.
Nino-Liu DO *et al.* (2006) *Xanthomonas oryzae* pathovars: Model pathogens of a model crop. Molecular Plant Pathology 7 303–324.
Sakai H *et al.* (2013) Rice annotation project database (RAP-DB): An integrative and interactive database for rice genomics. Plant and Cell Physiology 54: e6.
Streubel J *et al.* (2013) Five phylogenetically close rice SWEET genes confer TAL effector-mediated susceptibility to Xanthomonas oryzae pv. oryzae. New Phytologist 200 808–819.
Yang B *et al.* (2006) Os8N3 is a host disease-susceptibility gene for bacterial blight of rice. Proceedings of the National Academy of Sciences of the USA 103 10503–10508.
Yu J *et al.* (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science 296 79–92.
Zhou H *et al.* (2014) Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. Nucleic Acids Research doi: 10.1093/nar/gku806.

BING YANG'S research interests focus on: the molecular mechanisms of plant/microbe interactions and crop disease-resistance engineering by using the bacterial blight of rice as a model; and development and application of TALEN and CRISPR technologies for targeted genome editing in plants. Over the past 15 years, he has identified and characterized several important naturally occurring TAL effectors in the rice pathogen *Xanthomonas oryzae* for their disease-promoting ability, and, most recently, he has helped harness the disease-causing TAL effectors for targeted gene editing. His group generated the first disease-resistant crop species by using the TALEN technology.

DR. YANG has a PhD in plant pathology from Kansas State University. He is an associate professor in the Department of Genetics, Development and Cell Biology at Iowa State University.