Review: Plant Genome Editing Targeted by RNA-Guided DNA Endonuclease CRISPR/Cas9

Norimoto Murai

Department of Plant Pathology and Crop Physiology, College of Agriculture, and AgCenter Biotechnology Laboratory, Louisiana Experiment Station, LSU Agricultural Center, Louisiana State University, Baton Rouge, LA, USA
Email: nmurai@lsu.edu

Abstract

This review chronicles the development of the research on CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeat/CRISPR associated protein 9) during the last 30 years from the discovery of CRISPR sequence, of biological significance and of the molecular mechanism for adaptive bacterial immunity. It describes recent works on structural and functional diversity of CRISPR/Cas systems, and on three-dimensional structure-based improvements of on-target specificity of CRISPR/Cas9 and Cpf1 endonucleases. The review ends with the application of CRISPR/Cas9 to targeted editing of plant genomes. Importantly, plant commodities modified by CRISPR-Cas9 have not been considered as genetically modified organisms (GMO) as long as foreign DNAs from plant pests were not introduced, according to the recent determination by the USDA.

Keywords

CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeat/CRISPR Associated Protein 9), on- and off-Target Sequences, PAM (Protospacer Adjacent Motif)

1. Introduction

Restriction endonucleases were discovered originally as a bacterial defensive mechanism (restriction-modification system) against the invading phages or plasmid DNA in 1970 [1]. A similar but RNA-based adapted heritable immune system of bacteria was found recently using CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeat/CRISPR associated protein 9) [2] [3]. The CRISPR sequence was first identified in *Escherichia coli* at the 3'-end flanking region of *iap* gene coding for an aminopeptidase responsible for alkaline phos-

phatase isozyme conversion [4]. Five direct repeats of highly homologous sequences of 29 nucleotides contain a dyad symmetry of 14 base pairs (underlined) CGG**TTTA**TCCCCGC**TG/AG/A**CGCGGGG**AA**CTC and with homologous nucleotides found in at least two DNA segments in boldface type. The CRISPR repeats were separated by unique 32 nucleotide spacers. After the recent development of repeat pattern recognition algorithm, these repeat-spacer-repeat patterns of CRISPR loci were found widely distributed occurring in 50% and 87% of bacterial and archaeal genomes so far completely sequenced, respectively [5].

The biological significance of the CRISPR locus was not evident immediately at the time of discovery and thereafter for nearly 20 years. An investigation in *Streptococcus thermophilus*, a key species in the yogurt and cheese production has elucidated the clues to important function of CRISPR [2]. The CRISPR locus of wild-type strain DGCC7710 consists of 33 non-contiguous direct CRISPR repeats separated by stretches of variable spacer sequences located adjacent to Cas7 gene. Comparative analyses of CRISPR and spacer sequences in various *S. thermophilus* strains indicated differences in the number and type of spacers between phage-resistant and phage-sensitive strains [2]. Nine phage-resistant mutants were generated independently by challenging wild-type strain DGCC7710 with two distinct but closely related virulent bacteriophages 858 and 2972. The CRISPR loci of phage-resistant strains were found modified at the 5'-leader end by the insertion of one to four novel spacers (protospacers) derived from phage genomic sequences [2]. Removal or addition of particular bacterial spacers resulted in the modification of phage-resistance phenotype of the cell. For example, the presence of spacers S1 and S2 was linked to the resistance to bacteriophage 858 but not to 2972. Thus, the resistance specificity for bacterial phage is determined by the sequence similarity between the bacterial genome spacer and phage sequence, while the resistance mechanism is provided by CRISPR-associated (Cas) enzyme machinery [2].

The molecular mechanism for adaptive bacterial immunity was elucidated in the type II CRISPR system with dual-RNA-guided DNA endonucleases making a site-specific double-stranded (ds) DNA cleavage [6]. The *in vitro* reconstitution experiments for double-stranded DNA cleavage required the Cas9 enzyme, spacer 2 sequence with functional protospacer adjacent motif (PAM), 42-nucleotide CRISPR-derived RNA (crRNA) containing a spacer 2 sequence, and 75-nucleotide *trans*-activating crRNA (trancrRNA) [6]. A Cas9 enzyme has one HNH nuclease domain (a zinc finger domain including two pairs of conserved histidine (H) and one asparagine (N), an inactivating point mutation at H840A) in the middle, and three RuvC-like nuclease domains (resolvase cleaving the Holliday junction in DNA repair), an amino terminal domain (an inactivating point mutation at D10A) and two domains adjacent to HNH domain. Cas9 enzymes with point mutation experiments showed that the Cas9 HNH nuclease domain cleaves the complementary strand whereas the amino-terminal Cas9 RuvC-like domain cleaves the non-complementary strand [6]. The posi-

tions of cleavage sites were determined predominantly at three nucleotides 5'-upstream of the PAM of within the target sequence of each strand by termination of primer extension in the sequencing reaction. The cleavage sites of complementary strand were at predominantly three and four nucleotides 3'-downstream of the PAM when the size of 5'-labelled products were determined by gel electrophoresis. The cleavage sites of non-complementary strand were at eight to three nucleotides 5'-upstream of PAM [6]. These observations are in consistent with the sequences of deletion products of CRISPR/Cas9 reaction. Electrophoresis mobility-shift assay indicated tracrRNA is required for target DNA recognition of crRNA [6]. 5'-Terminal 48 to 23 nucleotides of mature tracrRNA were the minimal region that crRNA is able to guide Cas9-mediated DNA cleavage [6]. The 5'-end of the mature crRNA was base-paired to the complementary strand via the 20 nucleotides target sequence including the seed sequence [7]. The target DNA/crRNA complex structure was stabilized by tracrRNA via 20 base-pairs between the 3'-downstream sequence of crRNA and the 5'-end 10 to 33 nucleotides of tracrRNA. Dual RNA structure directed the CRISPR-associated protein Cas9 to introduce double-stranded break in targeted DNA. When engineered as a single RNA chimera, the crRNA (5'-20 nucleotides complementary sequences and the 3'-downstream 12 nucleotides) joined by GAAA-loop with tracrRNA (5'-23 to 48 nucleotides) also directs sequence-specific Cas9 dsDNA cleavage [6]. The inclusion of the 3'-end tail sequence from 48 to 85 of tracrRNA provided the most active architecture of the sgRNA, probably due to the increased stability of sgRNA [8]. The mismatched target sequence experiments indicated that a contiguous stretch of at least 13 base pairs between the crRNA and target DNA site proximal to PAM is required for efficient target cleavage [6]. The PAM conforms to an NGG consensus sequence in type II system [6]. Cleavage assay with the PAM mutations in either or both complementary or non-complementary strands indicated that NGG consensus sequence is required for Cas9-mediated DNA cleavage in the non-complementary strand. Native gel mobility assay was used to determine the binding affinity of complex for target sequence. Both G nucleotides of the PAM are required for the affinity of Cas9-tracrRNA:crRNA to the target DNA and the subsequent DNA cleavage. The PAM sequence is specifically recognized by Cas9 as a prerequisite for target DNA binding and possibly strands separation to allow strand invasion and R-loop formation [6]. The results suggest a new experimental approach for a genome-editing using a single guide RNA programmable to a particular target sequence [9] [10].

2. Structural and Functional Diversity of CRISPR/Cas Systems

So far two distinct classes (class 1 and 2) of CRISPR/Cas systems have been reported, each class subdivided to three different types, class 1 subdivided to type I, III and IV, and class 2 divided to types II, V and VI [11] with 19 subtypes [5].

The class 1 systems are distributed in both bacterial and archeal species, while class 2 systems are found almost exclusively restricted to bacteria. The class 1 systems appear to be the evolutionary ancestral one, and the class 2 systems have evolved from the class 1 after insertion of transposable elements (casposons) encoding various nucleases, based on the module organization comparison and sequence contents analysis [11]. The inverted terminal repeats of casposon may have contributed to the prototype CRISPR repeat sequence.

RNA-based adaptive immune systems were developed independently as the CRISPR system in bacteria and archaea, and RNA interference (RNAi) in eukaryote [12]. There are parallels and distinctions between these two systems. In each case long RNA transcription precursors are processed into smaller RNAs, which act as sequence-specific guides with a high-affinity seed sequence to target complementary DNA sequences in the CRISPR system, or RNA sequences (mRNA, viral RNA) in the RNAi. In eukaryotic RNAi microRNA, small inhibitory RNA, or other small RNA species form RNA-induced silencing complex with Agonaute protein [12].

CRISPR-mediated adaptive immune systems follow three distinct steps; step 1 for insertion of new protospacer (Adaptation), step 2 for CRISPR-derived RNA (crRNA) processing (Expression), and step 3 for target interference (Interference), or effector module (crRNA and target binding) and target cleavage [11]. For step 1 Adaptation (spacer insertion) the CRISPR/Cas loci share the common Cas1 nuclease/integrase and Cas2 nuclease between class 1 and 2. After an invading viral DNA enters the host cells, Cas1 and 2 proteins select a protospacer sequence of viral DNA and integrate a new spacer at the 5'-leader end of the CRISPR array containing the start site of CRISPR transcription [11]. For step 2 Expression the CRISPR locus is transcribed into the pre-crRNA which is then processed into mature crRNA guide by Cas6 nuclease (class 1) or RNase III (class 2) [11]. For step 3 Interference the Cas-crRNA complex scans the invading DNA for a complementary DNA sequence, and the selected target DNA is degraded by a Cas nuclease [11].

The CRISPR repetitive loci serve as molecular vaccination cards by maintaining a genetic record of prior encounters with invading viral or plasmid DNA [11]. Protospacers are flanked by a short motif, protospacer adjacent motif (PAM) located on the 5'-(type I and V), or 3'-side of protospacers (type II) [11]. Long crRNAs precursors are transcribed from 5'-end leader/promoter (A + T)-rich sequences of CRISPR repeats, and processed at approximately 30- to 65-nucleotide intervals with repeat sequences recognized and cleaved into short crRNA [11]. In type I and III system of class1, a CRISPR-specific endoribonuclease Cas6e cleaves 8 nucleotides upstream of each spacer sequence [11]. In type II of class 2 systems including Cas9, crRNA and a *trans*-activating antisense RNA (tracrRNA) forms a RNA duplex base-paired through the complementary sequence to CRISPR repeats [11]. The *trans*-activating anti-sense RNA is transcribed from a part of a single-subunit crRNA-effector complex, downstream of

Cas9 endonuclease. The RNA duplexes are recognized and cleaved by cellular RNase III. Cleavage intermediates are trimmed further by exo- or endonuclease to approximately 40-nucleotide mature crRNA with the 5'-end repeat sequence removed and 3'-end 20-nucleotide repeat sequences remained as a 3'-handle [11]. The cleaved crRNA-tracrRNA hybrid complexes are bound and stabilized by Cas9, which triggers a conformational change toward a state compatible with target scanning, recognition and interference [11].

Class 1 and 2 of CRISPR systems demonstrate a distinct modular organization for the Interference function [11]. Class 1 has a multi-subunit crRNA-effector complex, and class 2 has a single-subunit multi-module crRNA-effector complex [11]. The effector modules of class 1 share four common proteins, Cas7, Cas5, Small and Large Subunits [11]. Target cleavage for type I is mediated by N-terminal nuclease HD phosphohydrolase domain (Cas3") and the C-terminal helicase domain (Cas3') [11]. Cas10/Large Subunit proteins are responsible for target cleavage for type III. Type III-A targets plasmid DNA *in vivo*, while type III-B cleaves only single-stranded RNA substrates *in vitro* [11]. Type I and III share similar effector modules and strikingly similar overall architecture despite a low sequence similarity of individual effector module components [11]. In class 2 a single large Cas9 protein for type II, Cpf1 protein for type V and C2c2 protein for type VI are responsible for interference function.

For target interference in type II of class 2 system, selection of a target sequence relies on recognition of non-self complementary protospacer sequence (20 - 22 nucleotides) including the seed sequence (7 - 8 nucleotides), and protospacer adjacent motif (PAM, 3 nucleotides) [11]. The 5'-side of the crRNA spacer sequence recognizes the single-stranded target DNA and forms a high-affinity seed sequence which becomes a target of double-stranded DNA break by Cas9 endonuclease [7]. Matching PAM and the close proximal seed sequence is a crucial quality control step that is required for the complete displacement of the non-complementary strand of the target DNA by the crRNA guide forming R-loop conformation [7]. A major step in Cas9 activation is the re-orientation of the structural lobes after the crRNA/tracrRNA binding, resulting in the formation of a central channel that accommodates the target DNA. The PAM 5'-TGG-3'nucleotides are identified by the mostly stranded region of Cas9 endonuclease that positions DNA duplex and promote local duplex melting [13]. This allows the Cas9-RNA duplex to probe the identity of the nucleotides of the seed sequence, driving further step-wise destabilization of DNA duplex.

3. Improved On-Target Specificity of CRISPR-Cas9 and Cpf1 Endonucleases

Molecular structures of a CRISPR-Cas9 R-loop complex were determined by cryo-electron microscopy on crystal structure of wild-type *Streptococcus pyogenes* Cas9 (SpCas9) bound to single guide RNA (sgRNA) at 4.5A° resolution,

and also on Cas9/sgRNA/30 bp target dsDNA containing a canonical 5'-TGG-3' PAM at 6.0A° resolution [13]. SpCas9 bends the DNA helix by 30° angle, providing the structural distortion needed for R-loop formation. The structural rearrangement forces displaced DNA strand position near the RuvC nuclease domain active site and further position the HNH nuclease domain adjacent to the target DNA strand cleavage site in a conformation essential for concerted double-stranded DNA break [13]. The R-loop formation triggers a conformational change of the two nuclease domains (HNH and RuvC) of SpCas9, which adapt an active state for blunt-end double-stranded break [13].

CRISPR-Cas9 endonucleases cause double-stranded DNA breaks at guide RNA-targeted genomic site, but also at off-target sites that are not fully complementary to the guide RNA sequence. Structure-based site-directed mutagenesis of SpCas9 endonuclease was conducted to reduce off-target sites cleavages, and the effects of mutations were evaluated in human embryonic kidney cells using indel formation at the on-target site and three known off-target sites [14]. Based on the molecular structure of CRISPR-Cas9 R-loop complex, 31 positively charged residues (Lys, His, Arg) were identified at positions between the HNH and RuvC nuclease domains and PAM-interacting domain in SpCas9 nuclease [13]. These 31 positively charged residues within the non-target strand grove were substituted singly or in combination by neutral alanine residue [14]. The rationale for the above approach is that neutralization could weaken non-target strand binding and encourage rehybridization between the target and non-target DNA strands, thus requiring more stringent Watson-Crick base pairing between the RNA guide and the target DNA strand [14]. Five out of 31 mutants (R780A, K810A, K848A, K855A, H982A) were shown to reduce at least ten-fold the off-target cleavage activity compared with wild-type SpCas9 and also to maintain the original on-target DNA breakage efficiency [14]. After combination of these mutation sites, eight of 34 combinatory mutants retained wild-type on-target efficiency with no detectable off-target indel levels. One single mutant K855A and two combinatory mutants, K810A/K1003A/R1060A and K848A/K1003A/R1060A demonstrated high efficiency with wild-type levels of on-target indel formation and specificity [14]. These mutants were named as the enhance specificity variants or eSpCas9 (1.1).

An alternative approach for reducing the off-target DNA cleavage was based on a hypothesis that off-target effects of SpCas9 might be minimized by decreasing non-specific interactions with its target DNA site [15]. Based on the structural studies they identified four residues (N497, R661, Q695, Q926) of SpCas9 endonuclease making direct hydrogen bonds with phosphodiester chain of target DNA [13]. Both N497 and R661 interact with the10th and 11th PAM proximal nucleotides, Q926 interacts with the 12th and 13th nucleotides, and Q695 interacts with the 19th and 20th nucleotide [13]. The four residues were substituted by alanine singly or in combination, and the effect of single, double, triple and quadruple combinatory mutations were evaluated for insertion or de-

letion mutations on-target and off-target cleavage sites [15]. They used the genome-wide unbiased identification of double-stranded breaks enabling by sequencing (GUIDE-seq) method, and targeted amplification sequencing for direct measurement of the frequencies of indel mutations [15]. SpCas9 endonuclease with four combinatory mutations N497A, R661A, Q695A and Q926A maintained on-target activities comparable to wild-type SpCas9 with 85% of single-guide RNAs (sgRNAs) tested in human cells. This high-fidelity SpCas9 enzyme (SpCas9-hf1) reduced all or nearly all off-target DNA cleavage undetectable assessed by the above two methods, and also can be highly effective to reduce off-target effects of sgRNA to simple TG repeat sequences [15]. The specificity of high-fidelity SpCas9 enzyme was enhanced further by additional mutation D1135E, and two mutations for hydrophobic or stacking interaction of residues L169A and Y450A that interacts with the 5th and 7th PAM proximal nucleotides, respectively [15]. Essentially the same dissociation constants were observed with wild type enzyme, and high fidelity variants eSpCas9(1.1) and SpCas9-hf1 [16].

A functional approach has been taken to elucidate the interactions of protein domains fluorescently labeled by Cy3/Cy5 dyes and that with target DNA substrate resulting in the enzymatic activation for double-stranded DNA break [16]. Single-molecule Forster resonance energy transfer (smFRET) analysis is a quantitative approach to distinguish between the active (FRET efficiency = 0.97) and inactive state (FRET efficiency = 0.45) of HNH nuclease domain of the SpCas9 enzymes [16]. The majority of WT enzyme was in an active state with both on-target and mismatched target sequences. Only approximately 32% of SpCas9-hf1 molecules had the HNH active state with an on-target substrate while the remaining 68% trapped in the inactive state. Single mismatch of target sequence at the PAM-distal end essentially eliminated the HNH active state and moved FRET efficiency to the HNH inactive state [16]. Similar effects were also observed with eSpCas9(1.1) variants. The evidence from further analysis was consistent with a proposal that the high-fidelity SpCas9 enzymes might reduce off-target cleavage by raising the threshold for HNH conformational activation when bound to target DNA sequences [16]. A non-catalytic domain within the Cas9 recognition lobe (REC 3) interacts with the RNA-DNA heteroduplex and undergoes conformational changes upon target binding [13]. Deletion of REC3 from the WT SpCas9 lowered FRET efficiency to the HNH inactive state and reduced the cleavage rate by 1000-fold [16]. Complementation *in vitro* of REC3 domain to the REC3-deletion mutant restored the DNA cleavage rate. Thus, the REC3 might act as an allosteric effector, and allow for the activation of HNH nuclease activity and for the regulation of overall activity of double-stranded DNA breakages [16]. Another domain REC 2 and HNH domain appear to be tightly coupled to form the conformational proofreading mechanism ensuring catalytic competence [13]. REC2 might sterically stops HNH in the conformational check point when SpCas9 is bound to off-target sequences. Five clusters of

residues within the REC3 are located within 5A° of the RNA-DNA interface [13]. Mutations of four conserved amino acids in cluster 1 (N692A, M694A, Q695A and H696A) suppressed off-target cleavage while retaining target binding affinities comparable to WT demonstrated hyper-accuracy of a new Cas9 variant (HpaCas9) against mismatches at positions 1 through 18 of target sequences [16]. The GUIDE-seq analysis demonstrated that the new variant HypaCas9 showed dramatically improved genome wide specificity compared with WT SpCas9 and showed equivalent or better genome-wide specificity in comparison to both SpCas9-hf1 and eSpCas9(1.1) for all sgRNAs examined [16]. It is concluded that REC3 binding to the RNA-DNA duplex is necessary for re-orienting REC2 enabling HNH docking to the active site for double-stranded DNA break.

To minimize off-target cleavages it is also best to target non-repetitive sequences that do not have closely matched sites elsewhere in the genome. On-target specificity of CRISPR-Cas9 could be improved to identify potential off-target genomic sequences with following criteria [8]: 1), potential off-target sequences should not followed with 5'-NGG or 5'-NAG PAM sequences; 2) guide sequences with fewer than three mismatches should be avoided; 3) off-target site with at least two mismatches lie within the PAM-proximal region should be avoided; 4) a maximal number of mismatches should be consecutive or spaced less than four bases apart. The amount of SpCas9 and sgRNA could be titrated quantitatively to optimize on to off-target cleavage ratio.

Cfp1 nucleases for type V-A, one from *Acidaminococcus sp. BV3L6* (AsCpf1) and another from *Lachnospiraceae bacterium ND2006* (LbCpf1) showed similar on-target efficiencies as with SpCas9 enzyme at the overlapping target sites in human cells [17] [18]. Cpf1 requires only a single 42-nucleotide crRNA, and Cfp1/crRNA complexes cleave target DNA molecules in the absence of a tracrRNA [19]. Cfp1 recognizes its PAM 5'-TTN-3' through a combination of base and shape readout. Two Cfp1 nucleases are highly sensitive to mismatched crRNA nucleotides at most positions between 1 and 18 proximal to the PAM assessed by T7 endonuclease I (T7EI) mismatch cleavage assays [17] [18]. Both Cfp1 nucleases are highly specific in its targeting of human DNA, as off-target cleavage were not detectable for more than 20 different crRNA using GUIDE-seq and targeted deep sequencing analyses [17] [18]. Cpf1 generates staggered double-stranded DNA breaks with 4- or 5-nucleotide 5'-overhangs [19]. In the Cpf1 structure, the unique nuclease domain is positioned so as to cleave the target strand outside of the heteroduplex. The analysis suggests that the specificities of AsCfp1 nucleases may approach that of the high-fidelity SpCas9-hf1 enzyme described above [17]. C2c2 from *Leptotrichia shahii* is class 2 type VI CRISPR-Cas effector with the RNA-guided ribonuclease function and is capable of interference against RNA phage [20].

To understand the mechanism by which SpCas9 molecule search the 20 nucleotides-target sequence within the genome, the movement of catalytically inac-

tive SpCas9 molecules (dSpCas9) were visualized by labeling with the enhanced Green Fluorescent Protein, and by tracking single-particle in living mouse cell nuclei [21]. The target sequence and protospacer of mouse genome are bound with histone complex within the chromatin structure. The analysis showed three-dimensional diffusion of Cas9 movement *in vivo*, and off-target binding events take 750 milliseconds [21]. Searching is dependent of the local chromatin environment with less sampling and slower movement within heterochromatin region of chromosomes [21].

In the bacterial genome of free-living *Escherichia coli*, the chromosomal DNA sequences are considered free from any bound proteins and exist essentially as a naked DNA. The search kinetics of dCas/single-stranded RNA was studied in living *E. coli* by combining single-molecule fluorescence microscopy and bulk restriction-protection assay [22]. Binding of Cas9 is more time-consuming than simple binding of transcription factor for correct base recognition, because Cas9 binds and also unwinds the DNA double-helix to test for correct base pairing to the guide RNA. The dCas molecules were visualized by fusing with fluorescent protein YPet [22]. DNA-bound dCas/YPet molecules are detectable as individual diffraction-limited spots after five second image acquisition time while non-bound molecules are seen as the diffuse fluorescent background [22]. Under these conditions dCas was demonstrated to take six hours to find the correct target sequence suggesting that each potential target is bound for less than 30 milliseconds that is 20 times faster than 750 milliseconds in eukaryotic mouse cells [21]. To achieve fast targeting, both dCas9 and its guide RNA molecules have to be present at extremely high saturating concentrations.

4. Plant Genome Editing Targeted by RNA-Guided CRISPR/Cas9 Endonuclease

Applications of CRISPR/Cas9 to plant genome-editing were first reported in 2013 for monocotyledonous crop plants of rice and wheat [23] [24] as well as for dicotyledonous model plants *A. thaliana* and *Nicotiana benthamiana* [25] [26]. S*treptococcus pyrones Cas*9 (*SpCas*9) was modified to optimize the bacterial gene codons for plant expression and used with the attachment of plant nuclear localization signals (NLS). SpCas9 DNA endonuclease was expressed under the control of either 2 × 35S promoter of Califlower Mosaic Virus [23] or maize ubiquitin promoter [24]. Single guide RNA (sgRNA) was expressed under control of RNA Polymerase III U3 or U6 prompoter. The lower expression efficiency of sgRNA appeared to be a limiting factor for optimal mutation efficiency. SpCas9 and sgRNA were introduced to protoplasts for transient expression assay, or particle-bombarded to callus or *Agrobacterium*-mediated leaf infection for transgenic plant assay [25] [26]. Rice phytoene desaturase gene *OsPDS*, three other genes (*OsBDH*2, *Os*02*g*23823, *OsMPK*2), and common wheat gene (*TaMLO*) were targeted for mutation [23]. The effects of mutation were screened for the resistance to restriction enzyme digestion or for indel formation

[23]. Single-nucleotide deletions were observed at the intended double-stranded DNA breakage sites due to non-homologous end joining (NHED), as well as a large deletion of up to 32 nucleotides at around intended breakage site [23]. Homology-directed repair (HDR) of gene modification was also demonstrated by targeted insertion of 12 bp DNA encoding KpnI and EcoRI sites at the double-stranded DNA cleavage site of *OsPDS* after co-transformation of Cas9, sgRNA and single stranded 72-bases long nucleotide [23].

A single-nucleotide substitution, deletion or insertions were detected in the Arabidopsis protoplasts at or around the intended double-stranded DNA breakage sites of *AtPDS* and *AtFLS*2 (*Flagellin Sensitive* 2) gene with relatively low mutation rates of 1.1% to 5.6% [25]. In *N. benthamiana* protoplasts, much higher mutagenesis frequencies for *NbPDS* (37.7%) and *NbFL*2 (38.5%) were resulted with considerable DNA deletions or insertions but rare single-nucleotide substitution at the intended site [25]. Only single-nucleotide substitutions or insertions were detected when targeted to an Arabidopsis gene for *Receptor for Activated C Kinase* 1 (*RACK*1) [25]. When two tandem sgRNAs were co-expressed with 24-bp spacer between two adjacent target sequences in *AtPDS*3, deletion of up to 48 bp genomic sequence was observed with a mutation frequency of 7.7%. Homology-directed repair of replacement of *NbPDS* gene was demonstrated at the frequency of 9.0% with the 5'- and 3'-homology arm sequences of 533 bp and 114 bp in the HDR template [25]. Preassembled CRISPR-Cas9/guide RNA complexes were introduced to protoplasts of *A. thaliana*, *N. attenuata*, lettuce and rice [27]. Targeted mutagenesis frequencies in regenerated plants were reported up to 46%.

A non-functional mutant gene for Green Fluorescent Protein (GFP) was used as a reporter of double-stranded DNA breakage by CRISPR/Cas9/sgRNA followed by Non-Homologous End Joining repair system to generate a functional GFP in Arabidopsis, tobacco, rice and sorghum [28]. The GFP reporter system had a distinct advantage of monitoring the products of DNA cleavage and repair reactions with highly visual sensitivity at specific excitation and detection wavelength of 448 nm and 500 - 550 nm of GFP, respectively. The coding sequence of mutant GFP gene was shifted out of frame just the downstream of ATG start codon with the insertion of the 20 bases of target sequence of sgRNA and AGG PAM sequence [28]. Repaired functional GFP genes were enriched from site-specific PCR amplification by loss of the restriction endonuclease site for *ApaL*1 in Arabidopsis and tobacco, and for *SexA*1 or *BsrG*1 in rice. DNA sequencing of repaired functional GFP showed small deletions of one to 17 nucleotides or insertion of one to three nucleotides in Arabidopsis and tobacco, and small deletion of one to eight nucleotides and one nucleotide insertion in rice [28].

A positive selectable phenotype of the herbicide tolerance to bispyribac sodium was introduced by point mutations of rice gene for acetolactate synthase at amino acid residues 548 (W548L, tryptophan T**G**G to leucine T**T**G) and residues

637 (S637I, serine AGT to isoleucine ATT) [29] [30]. Biallelic gene targeting was directed by two sequential introductions of Cas9 followed by sgRNA and DNA template for homology-directed repair events [30]. The successful HDR events introduced new restriction enzyme *Mfe1* site. Positive transformation frequency for herbicide-tolerance was relatively low at maximum 1% [30]. It was not clear whether the bialleleic mutations are introduced after double-stranded DNA breakage via a mechanism of homology-directed repair or non-homology end-joining repair system. Transposon *piggyBac* derived from the cabbage looper moth was used to introduce the two mutations for the herbicide tolerance after Agrobacterium-mediated transformation [29]. Cleaved amplified polymorphic sequence analysis showed that the transposon was excised precisely with high-frequency (92.5% to 100%) without leaving the footprint [29].

CRIPSR/Cas9 gene editing methods were applied to other crop plants, tomato [31], maize [32], soybean [33] and sweet orange [34]. Loss of function mutation of *ARGONOUTE*7 (*SALARGO*7) resulted in a distinct recognizable phenotype in 48% of T0 plants, first formed leaves with petiole-less leaflets and later formed radialized leaves, and compound flat leaves becoming needle-like or wiry in mostly infertile mature plants [31]. Large deletions of up to 140 bp, and small deletions of one to three bp were detected using two 20 bp target sequences of sgRNA separated by 68 bp. An expected size of 68 bp deletion was found only in one of 29 T0 plants, suggesting the significantly lower probability of simultaneous two double-stranded DNA breakages having taking places than asynchronous cuts. CRISPR/Cas9-generated mutations were stably transmitted in tomato through the germline [31]. The mutation rate of 3.2% to 3.9% was reported when the CRISPR/Cas9/sgRNA was targeted to the Phytene Desaturase (*CsPDS*) gene of sweet orange [34]. Deletions of one to 12 bp and nucleotide replacements were detected by DNA sequencing.

Lowder *et al*. [35] in 2015 proposed a comprehensive molecular toolbox for multifaceted applications of CRISPR/Cas9 methodology to plants using Golden Gate and Gateway cloning methods. The toolbox was used in tobacco, Arabidopsis and rice for multiplexed gene editing, transcriptional activation via VP64, or repression via SRDX of endogenous genes. CRISPR/Cas9 and up to three sgRNAs were introduced to plant nuclei after *A. tumefaciens*-mediated transformation. Two tobacco genes for *Flagellin-Sensitive*2 (*FLS*2) and *Brassinosteroid Kinase* 1 (*BAK*1) were targeted by two sgRNAs resulting in a deletion of up to 2 kbp [35]. Rice genes for *Rice Young Seedling Albino* (*OsYSA*) and *Rice Outermost Cell-specific Gene*5 (*OsROC*5) were targeted using rice protoplasts. In transgenic rice plants two sgRNAs targeted to two genomic sites of *OsYSA* 200 bp apart resulted in deletions of 204 to 210 bp with a high frequency. Non-homologous end-joining repairs of *OsYSA* and *OsROC* genes were demonstrated with 6 or 9 bp deletions or on to 51 bp insertions with more than 50 % frequencies [5]. Transcriptional activation of a gene for *Arabidopsis Production of Anthocyanin Pigment* 1 (*AtPAP*1) was shown up to seven-fold using PP64 ac-

tivator.

Gemini virus was used to attempt to show non-integrating introduction of HDR-directed mutation to tomato genome after Agrobacterium-mediated transformation [36]. Bean yellow dwarf virus vector carried the CRISPR/Cas9/ sgRNA system, targeting the promoter insertion of the Myb transcription factor for the Anthocyanin gene (intense purple color phenotype) via homology-directed repair. Approximately 10% of transformed kanamycin-resistant calli showed purple phenotype after gene targeting in T0 progeny [36]. The frequency of purple calli was ten-fold less without the use of Gemini virus vector, suggesting that the high-copy number of virus vector (hundreds to thousands of copies per cell) could have increased gene-targeting frequency by ten fold. DNA sequencing showed a perfect match to the expected sequence at the right and left homology junction in the 69% of purple calli. No purple spots were observed in the absence of kanamycin-selection for transformation, suggesting that T-DNA construct of Gemini virus containing CRISPR/Cas9/sgRNA and anthocyanin promoter could have been integrated into the tomato genome prior to the transcription of the Gemin virus vector [36]. Targeted HDR-directed mutation was stably transmitted to the T1 and T2 progeny.

The use of Gemini virus is limited due to the narrow host-range and small passenger gene size. The Gemini virus vector is restricted to a small genome size (ca. 3 kb single stranded DNA) with 1.5 kb passenger gene. The host-range could be narrow to a single crop (turnip, spinach) or could be broader to dicots, monocots, or higher plants. Conventional *A. Tumefaciens* enjoyed a broad host-range from the monocot/dicot Angiosperms, Gymnosperms, other higher and lower plants (fungi) to animal cells (human).

5. Concluding Remark

The gene-editing method using CRISPR-Cas9 has the number of advantages over transgenic plant production using *A. tumefaciens*. Importantly, plant commodities modified by CRISPR-Cas9 have not been considered as genetically modified organisms (GMO) as long as foreign DNAs from plant pests were not introduced, according to the recent determination by the USDA. One of six genes encoding polyphenol oxidase was knocked out by CRISPR-Cas9, resulting in the 30% reduction in the enzyme activity and in less browning of common white button mushroom (*Agaricus bisporus*) [37]. Gene-editing using CRISPR-Cas9 has been so far applicable to the model plants (*Arabidopsis thaliana*, *Nicotiana benthamiana*), monocotyledonous (rice, maize, wheat, sorghum) and dicotyledonous crop plants (soybean and sweet orange), and other 30 crops (apple, lettuce, potato, tomato and white button mushroom) [38].

Acknowledgements

Biotechnology Laboratory, Louisiana Experiment Station, LSU Agricultural Center, Louisiana State University.

References

[1] Smith , H.O. and Wilcox, K.W. (1970) A Restriction Enzyme from *Hemophilus influenzae*. I. Purification and General Properties. *Journal of Molecular Biology*, **51**, 379-391. https://doi.org/10.1016/0022-2836(70)90149-X

[2] Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P. (2007) CRISPR Provides Acquired Resistance against Viruses in Prokaryotes. *Science*, **315**, 1709-1712. https://doi.org/10.1126/science.1138140

[3] Samson, J.E., Magadan, A.H., Sabri, M. and Moineau, S. (2013) Revenge of the Phages: Defeating Bacterial Defenses. *Nature Review Microbiology*, **11**, 1044-1050.

[4] Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. and Nakata, A. (1987) Nucleotide Sequence of the *iap* Gene, Responsible for Alkaline Phosphatase Isozyme Conversion in *Escherichia coli* and Identification of the Gene Product. *Journal of Bacteriology*, **169**, 5429-5433. https://doi.org/10.1128/jb.169.12.5429-5433.1987

[5] Jackson, S.A., McKenzie, R.E., Fagerlund, R.D., Kieper, S.N., Fineran, P.C. and Brouns, S.J.J. (2017) CRISPR-Cas: Adapting to Change. *Science*, **356**, eaal5056. https://doi.org/10.1126/science.aal5056

[6] Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012) A Programmable Dual-RNA-guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, **337**, 816-821. https://doi.org/10.1126/science.1225829

[7] Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns S.J.J. and Severinov, K. (2011) Interference by Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) RNA Is Governed by a Seed Sequence. *Proceedings of National Academy of Science U.S.A.*, **108**, 10098-10103. https://doi.org/10.1073/pnas.1104144108

[8] Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., Cradick, T.J., Marraffini, L.A., Bao, G. and Zhang, F. (2013) DNA Targeting Specificity of RNA-guided Cas9 Nuclease. *Nature Biotechnology*, **31**, 827-832. https://doi.org/10.1038/nbt.2647

[9] Doudna, J.A. and Charpentier, E. (2014) The New Frontier of Genome Engineering with CRISPR-Cas9. *Science*, **346**, Article ID: 1258096. https://doi.org/10.1126/science.1258096

[10] Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. and Zhang, F. (2013) Multiplex Genome Engineering via CRISPR/Cas Systems. *Science*, **339**, 819-823. https://doi.org/10.1126/science.1231143

[11] Mohanraju, P., Makarova, K.S., Zetsche, B., Zhang, F., Koonin, E.V. and van der Oost, J. (2016) Diverse Evolutionary Roots and Mechanistic Variations of the CRISPR-Cas Systems. *Science*, **353**, aad5147-1-12. https://doi.org/10.1126/science.aad5147

[12] Wiedenheft, B., Sternberg, S.H. and Doudna, J.A. (2012) RNA-Guided Genetic Silencing Systems in Bacteria and Archaea. *Nature*, **482**, 331-338. https://doi.org/10.1038/nature10886

[13] Jiang, F., Taylor, D.W., Chen, J.S., Kornfeld, E., Zhou, K., Thompson, A.J., Nogales, E. and Doudna, J.A. (2016) Structures of a CRISPR-Cas9 R-Loop Complex Primed

for DNA Cleavage. *Science*, **351**, 867-871. https://doi.org/10.1126/science.aad8282

[14] Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D., Yan, W.X. and Zhang, F. (2016) Rationally Engineered Cas9 Nucleases with Improved Specificity. *Science*, **351**, 84-88. https://doi.org/10.1126/science.aad5227

[15] Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z. and Joung, J.K. (2016) High-Fidelity CRISPR-Cas9 Nucleases with No Genome-Wide Off-Target Effects. *Nature*, **529**, 490-495. https://doi.org/10.1038/nature16526

[16] Chen, J.S., Dagdas, Y.S., Kleinstiver, B.P., Welch, M.M., Sousa, A.A., Harrington L.B., Sternberg, S.H., Joung J.K., Yildiz, A. and Doudna J.A. (2017) Enhanced Proofreading Governs CRISPR-Cas9 Targeting Accuracy. *Nature*, **550**, 407-410. https://doi.org/10.1038/nature24268

[17] Kleinstiver, B.P., Tsai, S.Q., Prew, M.S., Ngueyen, N.T., Welch, M.M., Lopez, J.M., McCaw, Z.R., Aryee, M.J. and Joung, J.K. (2016) Genome-Wide Specificities of CRISPR-Cas Cpf1 Nucleases in Human Cells. *Nature Biotechnology*, **34**, 863-868. https://doi.org/10.1038/nbt.3620

[18] Kim, D., Kim, J., Hur, J.K., Been, K.W., Yoon, S.-H. and Kim, J.S. (2016) Genome-Wide Analysis Reveals Specificities of Cpf1 Endonucleases in Human Cells. *Nature Biotechnology*, **34**, 869-874. https://doi.org/10.1038/nbt.3609

[19] Zetsche, B., Gootenberg J, Abudayyeh, O.O., Slaymaker, J.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Reev, A., Koonin, E.V. and Zhang, F. (2015) Cpf1 Is a Single RNA-guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*, **163**, 759-771. https://doi.org/10.1016/j.cell.2015.09.038

[20] Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B.T., Shmakov, S., Makarova, K.S., Semenova, E., Minakhin, L., Severinov, K., Regev, A., Lander, E.S., Koonin, E.V. and Zhang, F. (2016) C2c2 Is a Single-Component Programmable RNA-Guided RNA-Targeting CRISPR Effector. *Science*, **353**, aaf5573-1 to 8. https://doi.org/10.1126/science.aaf5573

[21] Knight, S.C., Xie, L., Deng, W., Guglielmi, B., Witokowsky, L.B., Bosanac, L., Zhang, E.T., Beheiry, M.E., Masson, J.-B., Dahan, M., Liu, Z., Doudna, J.A. and Tjian, R. (2015) Dynamics of CRISPR-Cas9 Genome Interrogation in Living Cells. *Science*, **350**, 823-826. https://doi.org/10.1126/science.aac6572

[22] Jones, D.L., Leroy, P., Unoson, C., Fange, D., Curic, V., Lawson, M.J. and Elf, J. (2017) Kinetics of dCas9 Target Search in *Escherichia coli*. *Science*, **357**, 1420-1424. https://doi.org/10.1126/science.aah7084

[23] Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J.J., Qiu, J.-L. and Gao, C. (2013) Targeted Genome Modification of Crop Plants (Rice and Wheat) Using CRISPR-Cas System. *Nature Biotechnology*, **31**, 686-688. https://doi.org/10.1038/nbt.2650

[24] Miao, J., Guo, D., Zhang, J., Huang, Q., Qin, G., Zhang, X., Wan, J., Gu, H. and Qu, L.-J. (2013) Targeted Mutagenesis in Rice Using CRISPR-Cas System. *Cell Research*, **23**, 1233-1236. https://doi.org/10.1038/cr.2013.123

[25] Li, J.-F., Norville, J.E., Aach, J., McCormack, M., Zhang, D., Bush, J., Church, G.M. and Sheen, J. (2013) Multiplex and Homologous Recombination-Mediated Genome Editing in *Arabidospsis* and *Nicotiana benthamiana* Using Guide RNA and Cas9. *Nature Biotechnology*, **31**, 688-691. https://doi.org/10.1038/nbt.2654

[26] Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J.D. and Kamoun, S. (2013) Targeted Mutagenesis in the Model Plant *Nicotiana benthamiana* Using Cas9 RNA-Guided Endonuclease. *Nature Biotechnology*, **31**, 691-693.

https://doi.org/10.1038/nbt.2655

[27] Woo, J.W., Kim, J., Kwon, S., Corvalan, C., Cho, S.W., Kim, H., Kim, S.-G., Kim, S.-T., Choe, S. and Kim, J.-S. (2015) DNA-Free Genome Editing in Plants with Preassembled CRISPR-Cas9 Ribonucleoproteins. *Nature Biotechnology*, **10**, 1162-1164. https://doi.org/10.1038/nbt.3389

[28] Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B. and Weeks, D.P. (2013) Demonstration of CRISPR/Cas9/sgRNA-Mediated Targeted Gene Modification in Arabidopsis, Tobacco, Sorghum and Rice. *Nucleic Acid Research*, **41**, e188. https://doi.org/10.1093/nar/gkt780

[29] Nishizawa-Yokoi, A., Endo, M., Ohtsuki, N., Saika, H. and Toki, S. (2015) Precision Genome Editing in Plants via Gene Targeting and *piggyBac*-Mediated Marker Excision. *The Plant Journal*, **81**, 160-168. https://doi.org/10.1111/tpj.12693

[30] Endo, M., Mikami, M. and Toki, S. (2015) Biallelic Gene Targeting in Rice. *Plant Physiology*, **170**, 667-677. https://doi.org/10.1104/pp.15.01663

[31] Brooks, C., Nekrasov, V., Lippman, Z.B. and Van Eck, J. (2014) Efficient Gene Editing in Tomato in the First Generation Using the Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-Associated9 System. *Plant Physiology*, **166**, 1292-1297. https://doi.org/10.1104/pp.114.247577

[32] Svitashev, S., Young, J.K., Schwarz, C., Gao, H., Falco, S.C. and Cigan, A.M. (2015) Targeted Mutagenesis, Precise Gene Editing, and Site-Specific Gene Insertion in Maize Using Cas9 and Guided RNA. *Plant Physiology*, **169**, 931-945. https://doi.org/10.1104/pp.15.00796

[33] Li, Z., Liu, Z.-B., Xing, A., Moon, B.P., Koelihoffer, J.P., Huang, L., Ward, R.T., Clifton, E., Falco, S.C. and Cigan, A.M. (2015) Cas9-Guided RNA Directed Genome Editing in Soybean. *Plant Physiology*, **169**, 960-970. https://doi.org/10.1104/pp.15.00783

[34] Jia, H. and Wang, N. (2014) Targeted Genome Editing of Sweet Orange Using Cas9/sgRNA. *PLoS ONE*, **9**, e93806. https://doi.org/10.1031/journal.pone.0093806

[35] Lowder, L.G., Zhang, Y., Baltes, N.J., Paul, J.W., Tang, X., Zheng, X., Voytas, D.F., Hsieh, T.-F., Zhang, Y. and Qi, Y. (2015) ACRISPR/Cas9 Toolbox for Multiplexed Plant Genome Editing and Transcriptional Regulation. *Plant Physiology*, **169**, 971-985. https://doi.org/10.1104/pp.15.00636

[36] Cermak, T., Baltes, N.J., Cegan, R., Zhang, Y. and Voytas, D.F. (2015) High Frequency, Precise Modifications of the Tomato Genome. *Genome Biology*, **16**, 232. https://doi.org/10.1186/s13059-015-0796-9

[37] Waltz, E. (2016) Gene-Edited CRISPR Mushroom Escapes US Regulation. *Nature*, **532**, 293. https://doi.org/10.1038/nature.2016.19754

[38] Maxmen, A. (2017) Engineered Apple Tests US Consumer's Appetite. *Nature*, **551**, 149-150. https://doi.org/10.1038/551149a