Plant genome editing using engineered nucleases and success of CRISPR/Cas9 system

Moon Sajid1*, Zohaib Hassan1, Ghulam Hussain Sehrai1, Muhammad Adeel Rana1, Holger Puchta2, Abdul Qayyum Rao2

Abstract

Development of new plant breeding techniques have facilitated easy manipulation of plants at genetic level. Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR associated protein9 (Cas9) system is a valuable addition in programmable nucleases. The CRISPR/Cas9 system uses an RNA component to recognize a target DNA sequences and it has shown promising results with respect to simultaneous editing of multigenic plant traits. In this review, components of CRISPR/Cas9, their construction and its methods of delivery to plant cells are analyzed. Variation in nucleotide sequence of the protospacer adjacent motif, codon optimization and progress in web-based bioinformatic tools, will make CRISPR/Cas9 systems more efficient for plants. Development and optimization of protocols to efficiently target all plant species is still under development. Along with this, methods to inspect induced mutation and efficiency of the system have also been reviewed. Auxiliary improvements and understanding are still required to expand the CRISPR/Cas9 systems to target complex genome architectures and epigenetic elements.

Keywords:
New Plant Breeding Techniques, Clustered Regularly Interspaced Palindromic Repeats (CRISPR)/CRISPR-Associated Protein (Cas9), Multiplexing, Gene drive
Introduction
Along with basic research, development of new plant varieties that are pest resistant, herbicide tolerant and give high yield require efficient plant genome editing techniques [1-8]. Targeted genome editing has become one of the most important tool in plant genetic studies and research. Genome editing is based on the fact that sequence specific double strand DNA breaks (DSBs) are introduced into the plant genome. In principle, DSBs can be repaired by two different methods: by homologous recombination (HR) using a template with an identical sequence as matrix for the repair or simple by rejoining the broken ends by non-homologous end-joining (NHEJ) [9]. In case of NHEJ, the reaction is often error prone and results in disturbance of open reading frames [10]. While, in case of HR, external DNA fragment can be integrated in a controlled way at a specific site in the genome [11]. The current review is concentrating on changes induced by NHEJ. Developments of HR based technologies have been reviewed elsewhere recently [12,13]. Plant genome editing techniques have made it easy to identify and characterize new genes as well as to make changes in epigenetic elements [14]. For gene function studies and genetic improvement of plants, genetic mutants are developed and used. Natural mutants are proved very helpful in the characterization of various fundamental biological processes in the past. Nowadays, research has been shifted from natural mutants to artificially designed mutants [15].

Off target mutations produced by random mutagenesis and incomplete suppression caused by non-coding RNAs (ncRNAs), are among the hurdles in the development of genetic mutants, and in crop improvements [19]. The swift development in the field of sequence specific nucleases (SSNs) during the last few years, have not only made plant genome editing possible but also quite efficient [15]. Among programmable nucleases, meganucleases [20], zinc finger nucleases (ZFNs) [17] and transcription activator like effector nucleases (TALENs) [21] are the most commonly used (Table 1).

Table 1: General characteristics of programmable nucleases.

<table>
<thead>
<tr>
<th>Abbreviations</th>
<th>Meganuclease</th>
<th>ZFNs</th>
<th>TALENs</th>
<th>CRISPR/Cas9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identified from</td>
<td>Microbial mobile genetic elements</td>
<td>Eukaryotic gene expression regulators</td>
<td>Plant pathogenic bacterial species Xanthomonas</td>
<td>Adaptive immune system of bacteria and archaea</td>
</tr>
<tr>
<td>Target specificity/cleavage module</td>
<td>Target recognition domain / Nuclease domain</td>
<td>Zinc-finger domain / Non-specific FokI nuclease domain</td>
<td>TALE domain / Non-specific FokI nuclease domain</td>
<td>crRNA / Cas9 protein</td>
</tr>
<tr>
<td>No. of nucleases for an experiment</td>
<td>One meganuclease for one target site</td>
<td>A pair of ZFN for one target site</td>
<td>A pair of TALEN for one target site</td>
<td>One or more sgRNA and only one Cas9 nuclease</td>
</tr>
<tr>
<td>Success rate</td>
<td>Low</td>
<td>Low (24%)</td>
<td>High (&gt;99%)</td>
<td>High (&gt;90%)</td>
</tr>
<tr>
<td>Specific length of target site</td>
<td>14-40 bp</td>
<td>18-36 bp</td>
<td>30-40 bp</td>
<td>20-22 bp</td>
</tr>
<tr>
<td>Size</td>
<td>----</td>
<td>1kb x 2</td>
<td>3kb x 2</td>
<td>4.2 kb of Cas9 + 0.1 kb of sgRNA</td>
</tr>
<tr>
<td>Limitations</td>
<td>Difficult to design and screen Recognition site are missing in plant</td>
<td>Difficult to design and screen</td>
<td>Not suitable for multiplex gene editing</td>
<td>Difficult to design and screen</td>
</tr>
</tbody>
</table>

References [16] [17] [18] [15]
naturally as an adaptive immune system in bacterial and archaeal cells [23]. Although, several reviews are available on CRISPR/Cas9 system [24-27], the purpose of our review is to summarize recent advancements in the methods of CRISPR/Cas9 delivery and expression in plant cells along with analysis of induced mutations which can provide a huge breakthrough in the area of plant genome editing.

Methods

Literature survey and selection criteria

Data for this paper has been retrieved from Google Web Browser and Google Scholar by providing the key terms “Genome Editing Techniques”, “CRISPR Components”, “CRISPR for plants” and “CRISPR induced mutation analysis”. Research papers published in last 5 years were mainly consulted and data related to plant genome editing was selected for this review.

Discussion

What the breeders need?

The wish of plant varieties with the best traits is as old as agricultural practices. It started with selection of the best plants for the next generation, which is still carried out in many parts of the world. With the knowledge of genes that are responsible for specific traits and genetics, controlled hybridization of the best plants (plant breeding) was started [28]. Hybridization helps to unite different traits of two parent plants into one. Transfer of undesired traits (genes) and extended time required for development of one hybrid variety, are the main limitations of hybridization techniques. So, there was a need for target specific and time saving techniques for the development of desired plant varieties and plant transformation brought a revolution into classical breeding systems. Desired traits of plants can easily be modified with the least time comparable with old breeding system [29]. After that, improvements in transgenic technologies for modification of specific traits within the plants was required and researchers have now uncovered nature’s tool for genome editing including ZNFs and TALENs, along with the latest genome editing system, the CRISPR/Cas9 system.

Components of CRISPR/Cas9 system

The CRISPR/Cas9 system is a prokaryotic immune system that is found naturally in bacterial (48%) and archaeal (90%) genomes [30]. It plays a defensive role against invading genetic elements (DNA and RNA) that may be viral or plasmid in origin [31]. Three different types of CRISPR/Cas9 system (Type I, II and III) have been identified and this classification is based on a specific signature protein, better known as the Cas proteins. Type I and III systems share some common features including maturation and assembly of components which make them different from type II [23]. Type II system is more commonly exploited for genome editing because it is easily programmable and less complicated as compared with others [32].

The CRISPR/Cas9 system has three main components: Cas9 endonucleases, CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). Cas9 endonuclease form a complex with both RNA components (Fig. 2-a). crRNA plays its role in target recognition whereas TracrRNA plays its role in crRNA maturation and stabilization of Cas9 protein on the target site [23]. Both RNA components are non-coding and to some extent, complementary to each other. With
better understanding and knowledge, these RNA components are combined to form one chimeric RNA better known as single guided RNA (sgRNA) [23,24] (Fig. 2-b). With this combination, CRISPR/Cas9 system is now defined as a two-component system i.e. Cas9 endonuclease and sgRNA. sgRNA recognizes the target DNA sequences and direct the endonuclease to cleave it. Cleaved DNA ends will then re-unite through non homologous recombination (NHR) [33]. One thing that makes a DNA target specific for selection, is presence of a nucleotide sequence, known as protospacer adjacent motif (PAM). A PAM motif must be present in the targeted sequences because it helps the CRISPR/Cas9 system to differentiate between local and foreign genomes. The PAM sequence is also reported to be involved in an earlier association of CRISPR/Cas9 complex with the target region and it is the place where separation of targeted DNA strands is initiated [34]. Positions and sequences of the PAM motifs are CRISPR/Cas9 type specific. PAM usually starts with 5'-NGG-3' and is present at the downstream region of the target sequence. The Cas9 endonuclease have two independent regions like the RuvC and HNH regions. Both these domains are activated when CRISPR/Cas9 binds with the target sequence [35]. The HNH nuclease bind and cleaves the complementary strand while RuvC cleaves the non-complementary strand resulting in the blunt end formation which after re-union may give rise to frame shift mutations or indels [23].

**CRISPR/Cas9 system for plant genome editing**

CRISPR/Cas9 system was firstly used for microbial and mammalian genome editing and its first use for plant genome editing was not reported until August 2013 [36]. In the beginning, target specificity and efficiency of the CRISPR/Cas9 system was evaluated. Many reports have subsequently shown successful use of CRISPR/Cas9 systems for genetic modification of plants, including gene knockdown and knockin, genomic deletions or disruptions in *Arabidopsis*, *Nicotiana tabacum*, poplar, potato, tomato and for rice, wheat, *Zea mays* and petunia and for other plants as reviewed by Demirci et al. [25].

Following are the steps required for obtaining a genetically engineered plant by CRISPR/Cas9 system:

1) Selection of an appropriate target site (DNA segment) and sgRNA designing,

2) Construction of expression cassettes that carry Cas9 nuclease, sgRNA and other necessary genetic elements,

3) Delivery of “editing tools” in targeted cells,

4) Analysis of induced mutations in the genomic DNA of targeted cells,

5) Disposing external DNA elements [32,37]

**sgRNA expression system**

A regular and functional sgRNA is 98 nucleotides in length. This length also includes 20-nt of target sequence. A U3 or U6 promoter of small nuclear RNA (snRNA) is used for its expression in plant cells and RNA polymerase III is responsible for its transcription [15]. A number of methods like overlapping PCR, target-adaptor ligation and ribozyme methods have been successfully used to generate sgRNA expression cassettes along with its promoter and additional component. In ribozyme systems, functional sgRNA is formed by processing pre-sgRNA. Under this system, sgRNA is expressed by tissue specific promoters and RNA polymerase II transcribes it [38]. A PCR based intermediate cloning free technique has also been designed and it is used to quickly produce sgRNA expression systems. This technique uses either Golden Gate cloning or Gibson assembly to clone the complete sgRNA expression system in a CRISPR/Cas9 binary vector system [27].

**Cas9 expression system**

The original coding Cas9 gene is 4107 nucleotide long, but smaller Cas9 orthologues have also been reported in literature [15,39]. Nuclear localization signals (NLS) are fused with it to make it functional in eukaryotic cells. The genetic codes of Cas9 genes have also been optimized many times for its optimal expression in different organism. It is also reported that portions of the Cas9 genes can be expressed from separate genes. These portions then assemble in-vivo, if sgRNA is available [40]. Along with the genes of Cas9 and sgRNA, expression cassettes should have selective marker genes. An antibiotic resistant gene like hygromycin (*HPT*) and bialaphos (*Bar*) under 35S promoter can also be used for screening [41].

In a row of studies, host specific codon optimized genes were used because these genes are more efficient in translation as reported by [42,43]. Expression of codon optimized versions of Cas9 genes in rice or other
species of Gramineae family requires high GC content at 5’ terminal. Presence of this GC content mimics genes of this family as reviewed in previous literature [27].

Experimental design for delivery
In principle, two types of systems are used to deliver expression constructs in the plant cell: transient expression system and stable transformation in the host genome. Former method is helpful for short time expression of genes of interest in maximum number of cells whereas the later methods means permanent incorporation of transgene in the host genome that can be inherited to the next generations. Experimental design and planning decides the choice of expression system.

Transient expression systems enabled researchers to optimize and perform genome editing experiments easily. These expression methods include: protoplast transformation and Agroinfiltration methods. It has been reported that protoplast transient expression systems are highly efficient for DNA co-transfection [44]. Protoplast transfusion has advantages over other methods due to high levels of expression for both genes of CRISPR/Cas9 system, no matter if separate plasmids are used for single transfection. However, production of protoplast from plant tissues is a time-consuming process. This procedure involves removal of the cell wall using enzymatic digestion that is a critical step. Also, protoplast cultures are very delicate and can easily become contaminated [36]. A perfect alternative to the protoplast transformation is Agroinfiltration assays. It is easy to handle, less expensive and it can be carried out on intact plant, too. In this method, one or two different plasmids can be constructed that have genes of the CRISPR/Cas9 system. Newly designed constructs are introduced in Agrobacterium tumefaciens which is then injected into the plant [45]. It is also possible to introduce and express the Cas9 gene in the form of mRNA and sgRNA through microinjection in the plant cell [46].

A number of methods for stable nuclear transformation have also been successfully used but biolistic gene delivery by gene gun and Agrobacterium mediated transformation are most common. DNA damage and gratuitous build-ins in the genome are setbacks in wide spread use of biolistic methods [47]. Agrobacterium transformation is also a promising system and it has less side effects. It is a method of choice if genome editing experiment is planned to use NHEJ system for DSB repair and its successful use is reported for barley, maize and rice as reviewed by [27].

Multiplexing
Multiple genes interact with each other to control polygenic traits and to understand such traits, it is necessary to investigate the role of each gene involved in the group. The ease in designing a small sgRNA makes the CRISPR/Cas9 system a promising technique for editing multiple genes at the same time. Multiplex genome editing is helpful for improvement of complex traits of plants.

For multiplexing by CRISPR/Cas9 system, it is necessary that all required sgRNA genes are expressed in the cell at same time. A single vector having more than one sgRNA at different targets can be synthesized by multiple restriction enzymes or successive rounds of cloning but this process is time consuming [15]. The Golden Gate cloning is another promising method that has certain advantages over traditional methods. Many sgRNA genes can be added in a given arrangement by this method and it is not time consuming as reviewed by [27]. Ma et al. have reported successful editing of 7 FT-like gene simultaneously in a rice plant with CRISPR/Cas9 binary constructs developed by Golden Gate cloning. Gibson assembly method has also been explored for construction of CRISPR/Cas9 binary vectors. In this method, only one reaction is required to fit together multiple sgRNA genes [41]. A toolbox to design CRISPR/Cas9 multiplexing by using Golden Gate cloning and Golden Braid standard has been developed. Computer based applications like CRISPR Assembler applications and Golden Braid CRISPR domesticator help in in-silico designing of genome editing tools for plants [48].

Mechanism to synthesize multiple sgRNA from one polycistronic RNA has also been proposed. In their experiment, Xie et al. have used endogenous tRNA processing system to generate multiple sgRNA from one transcript. The authors have shown that a chimeric transcript having alternative tRNA and sgRNA units is effectively cleaved into functional tRNA and sgRNA units. Expression of sgRNA is also higher as compared with other methods. Successful generation of up to 8 sgRNA from one transcript and efficient editing of
respective target to produce DSBs have been achieved [49]. Multiplexing has been used to delete chromosomal regions and reports have shown successful deletion of up to 170-245 kb chromosomal portion from rice genome [50].

Detection of induced mutations
Methods to analyze induced mutations caused by programmable nucleases have an equal importance as the execution of experiment. A valid analysis assay assures the success and helps to calculate the efficiency of experiment. It also provides a base to conduct further experiments. Analysis methods get further importance when they are used to analyze new genome editing techniques like CRISPR/Cas9. The following methods are reported in literature for assessment of CRISPR/Cas9 induced mutations.

Analysis of restriction enzyme site
If a restriction enzyme site is present in the targeted DNA sequence, then the loss of the restriction site due to mutations caused by CRISPR/Cas9, can be used to check the success of an experiment. Blunt ends produced by Cas9 endonuclease in the target DNA is mainly located three base pairs before the PAM site. Restriction of the target DNA by Cas9 nuclease results in loss of restriction enzyme site due to DSB repair by the NHEJ repair system. During analysis, the target region is PCR amplified and treated with the restriction enzyme. When evaluated by agarose gel electrophoresis, the position of the DNA bands confirms the loss of restriction site. Several reports have shown a successful use of this method [49,51]. One inherent limitation of this method is that it limits the choice of target selection.

Surveyor assay is another way to analyze the success of experiment. It is based upon the cleavage of unpaired nucleotides present in DNA duplex by T7 Endonuclease I. PCR amplicons of targeted region are annealed and treated with endonuclease. Digested product is separated on agarose gel and stained with dye. Intensity of the bands determine the success of experiment as reported by [52]. Although it is less sensitive, surveyor assay can be used for any target sequence site and this is the reason of its extensive use. One thing that makes surveyor assay a method of choice for any experiment is that it does not need any specific sequence to be present in targeted DNA region as in the case of endonuclease assay and this characteristic makes it superior.

PCR-amplicon melting temperature method
Melting temperature of PCR amplicon depends upon the number of mismatch pairs in the DNA strand and it can also be used to differentiate between unchanged and mutated DNA strands. Melting temperature method is a type of qualitative assay because it only tells about the presence or absence of mutations and does not give information about the position and type of mutated nucleotides. This assay has also been employed to analyze the CRISPR/Cas9 results [53,54].

DNA sequencing methods
Sequencing of targeted regions or whole genome sequencing can also be used to analyze induced mutations by CRISPR/Cas9. Whole genome sequencing is most commonly used to detect off target mutations or rare mutations [54,55]. Sanger sequencing is particularly important for analysis of targeted region PCR amplicons [43]. Although, DNA sequencing is very expensive as compared to other available techniques, it is the most useful and reliable.

Efficiency of CRISPR/Cas9 system
Genome editing by programmable nucleases results in permanent change at targeted site and because of this, target specificity is very important for CRISPR/Cas9 system. Efficiency of CRISPR/Cas9 has been extensively analyzed [56-59]. Many studies have already demonstrated higher efficiency of CRISPR/Cas9 system for genome editing in plants as compared with the humans or mice [41]. There are many factors that define specificity and efficiency of CRISPR/Cas9 system. Some of these elements are: appropriate expression system, methods of transformation, selection of target site and targeted plants [15].

The ratio of CRISPR/Cas9 expression and functional nucleases delivered in the cell has also been reported to influence the efficiency. Limited amount of CRISPR/Cas9 are reported to cause less off-target [46,56]. The 1:1 DNA ratio of Cas9 enzyme to sgRNA is reported to be the best for nuclease activity of CRISPR Cas system [22]. Effects of concentration of Cas9 protein and sgRNA still need to be investigated along with the effects of duration of expression that may also play some role in target efficiency.

Although the tracrRNA is the same for every Cas9 system, the crRNA (or crRNA part of sgRNA) is variable because it decides the target specificity. It binds with the
target DNA sequence by simple Watson and Crick base pairing. The last 8-12 bases of crRNA also known as “seed sequence” along with its complementary targeted DNA region is very important for perfect target recognition and cleavage. Miss matches in PAM distal region can be tolerated to some extent as compare to the seed sequence [60].

The first nucleotide of sgRNA, is also likely to play a role in target selection. Some studies restrict the first nucleotide to be a purine, whereas other studies have shown that pyrimidines can also be used. Interaction of targeted regions with sgRNA scaffolds destabilize the CRISPR/Cas9 system and lowers its efficiency but for this, a minimum of six nucleotides must pair up with sgRNA scaffold [15]. It is also reported for rice that GC contents in target region also decide efficiency of CRISPR/Cas9 system and this high GC content results in higher off-target effects [41,56].

**Approaches to minimize off-target effects**

Careful selection of target sites by genome searching using BLAST or web-based tools proves to be helpful to minimize off-target effects. Optimization in CRISPR/Cas9 systems including modifications or change of nucleotides in Cas9 gene and sgRNA sequence can also minimize the off-target effects [61,62]. Cas9 specificity can be increased if a pair of Cas9 nickase variants are used to target opposite DNA strands. Both Cas9 nickase then produce a nick in their respective strand to cause a DSB. Target sequences are usually up to 100 base pair apart and properly repaired and it is improbable for off-target sites to be close enough to result in DSB [63,64]. The paired nickase approach has been successfully applied in Arabidopsis and rice by now [65-67].

One other potential approach to minimize off target effects is the development of catalytically dead Cas9 (dCas9). This variant of Cas9 is fused with FokI nuclease and known as dCas9-FokI.

**Outlook and future considerations**

Programmable nucleases are highly acceptable systems to improve model and crop plants effectively and efficiently. ZNFs and TALENs were used in routines for plant genome editing. Addition of CRISPR/Cas9 systems in NPBTs proves to be very useful. User friendly and easiness in sgRNA design makes CRISPR/Cas9 system superior over others. CRISPR/Cas9 systems use RNA for target recognition which help this system to recognize DNA sites that cannot be recognized by ZFNs and TALENs. These systems are highly sophisticated and reliable genome editing tools for both applied and basic plant research and breeding [68].

Engineered nucleases can help us to modify genetics of any plant species by gene insertions or deletions or through regulation of gene expression. It is now possible to regulate metabolic pathways to get desired products with ultimate enhanced plant yield. A better understanding of mechanisms involved in response to abiotic and biotic stress along with processes involved in nutrient and water absorption will also be investigated in near future [31]. “Gene drive” is one of the potential application of CRISPR/Cas9 system and it is successfully achieved in Drosophila [69]. Gene drive is a process by which a modified gene directs the modification of its sister allele on homologues chromosomes. If a gene in pollen grains is modified, it will also change its sister allele in the egg. This process spread a modified targeted gene in local and global environment in short time [70]. CRISPR/Cas9 system will be used to silence herbicide resistant genes in weeds to make them susceptible to herbicides and to introduce these genes in crops to make them resistant under gene drive methods. But field trials of CRISPR/Cas9 system for gene drive in plants still need extensive investigations [31,71]. Researchers are also working to use CRISPR/Cas9 as immune systems against DNA viruses as bacteria use it against RNA viruses [72].

**Conclusion**

Although plant genome editing does not has ethical concerns, still transgenic Cas9 genes and other marker genes can be removed in subsequent generations by segregation and result in non-transgenic genetically modified plant lines with negligible traces of transgenes [50,73]. However, CRISPR/Cas9 system still has some limitations that include off target effects [36]. Rapid development in Bioinfromatic tools for in-silico analysis will help us to minimize off target effects. Several reports have shown some optimizations to increase CRISPR/Cas9 target selection including changes in PAM recognizing domains to add other PAM motifs and incorporation of purines or pyrimidines as a first nucleotide of sgRNA [15,74-76]. With the successful
development and use of Cas9 orthologs, the CRISPR/Cas9 tool box will also be expanded to target some other genome elements including epigenetic elements [14]. First application in plants are promising [66,77,78]. CRISPR/Cas9 will also help to investigate how DNA binding proteins interact with specific region of genome and regulate gene expression.

It is impossible to foresee the future for programmable nucleases due to rapid development in the field of molecular biology. But discovery of CRISPR/Cas9 genome editing system will become a golden revolution for targeted genome engineering in the coming years.

Acknowledgment
We acknowledge Higher Education Commission (HEC) of Pakistan for providing access to world’s top journals.

References
54. Fauser F, Schiml S, Puchta H. Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in Arabidopsis thaliana. The Plant Journal, (2014); 79(2): 343-349.
Plant genome editing using engineered nucleases and success of CRISPR/Cas9 system


