NEW BREAKTHROUGH CRISPR/CAS9 BIOTECHNOLOGY OF GENOME EDITING FOR CREATION OF ELITE CROPS IN KAZAKHSTAN

Abstract. The potential role of the CRISPR/Cas9 technology is reviewed. The technology, a genome-editing tool called CRISPR/Cas9, revolutionized the life sciences when it appeared on the market in 2013. CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats, associated with protein 9 - Cas9), is RNA-regulated protection mechanism in bacteria and archaea, in which based new “crazy” popular technology of gene editing in human, animal and higher plants. CRISPR/Cas9 system is a simple, inexpensive and versatile tool for genome editing, resulting in a groundswell of research based on the technique which popularity in the last 4.5 years has become known as the ‘CRISPR craze’.

In bacteria organism, the Cas9 nuclease associates with two RNAs, the CRISPR RNA (crRNA) and the trans-activating crRNA (tracrRNA), to direct sequence-specific cleavage of foreign DNA. This bacterial acquired immune system has been shown to be effective for gene editing in mammalian cells, and is now routinely used as an effective genome engineering tool in multiple high organisms including crops.

Elite crops cultivars with improved resistance to fungal or viral pathogens will benefit farmers and the local economy in Kazakhstan by increasing harvest yields, and grain quality with high economic benefit through establishment of disease resistant crops genotypes using CRISPR/Cas9 plant genome editing. CRISPR/Cas9 system for plant genome editing is a breakthrough technology in breeding - prospects for application in near future, could be exempt from GM classification.

Keywords: genome engineering, genome editing, CRISPR/Cas9, gene knock-out, crops, breeding perspective, non-GMO classification.

Abbreviations
CRISPR - clustered regularly interspaced short palindromic repeats
Cas - CRISPR-associated (protein)
crRNA - CRISPR RNA
DSB - double-strand DNA break
HDR - homology-directed repair
NHEJ - non-homologous end joining
PAM - protospacer adjacent motif
sgRNA - single guide
shRNA - short hairpin RNA
tracrRNA - trans-activating crRNA
Introduction
The biomedical award – ‘Breakthrough Prize’ - was received by Jennifer Doudna from the United States and Emmanuelle Charpentier from Germany in 2013 - for the discovery of the CRISPR/Cas9 mechanism, which protects bacteria from virus attacks. This system determines the viruses by DNA and releases the Cas9 enzyme, which acts as scissors cutting the genetic material of the aggressor. However, scientists found that using CRISPR/Cas9 it is possible to rewrite defective human DNA, replacing damaged genes with healthy ones. In this case, the place of the spacer in such RNA occupies the sequence chosen by the researcher. The Cas9 protein is able to "learn" and connect to such a synthetic CRISPR RNA (it is called a "guide") and becomes programmed to recognize and cut the corresponding place in the DNA. The groups of J. Doudna and E. Charpentier [1] demonstrated the possibility of such an approach in vitro.

Almost at the same time, groups of George Church and his former graduate student Feng Zhang from the Broad Institute in MIT have shown that the bacterial protein Cas9 and the RNA guide are able to "work", recognize and aim to cut DNA in cells of higher organisms, particularly in higher plants. MIT managed to apply for a patent a day earlier [2, 3].

It is now proving useful in the plant science community as a powerful tool for the improvement of agricultural crops. It is estimated that up to 40% of harvest is lost worldwide to pests/diseases threatening our food supply. The major crops – wheat, rice, barley, are susceptible to many viral and fungal diseases that can result in drastically reduced yield and poor quality grain [4-11]. Crops cultivars with improved resistance to fungal or viral pathogens will benefit farmers and the local economy by increasing harvest yields and grain quality.

Results and Discussion
CRISPR/Cas9: What is it and how does it work?
In August 2013, five reports were published discussing the first application of CRISPR/Cas9-based genome editing in plants [4-8]. This first group of studies already demonstrated the immense versatility of the technology in the field of plant biology by embracing the model species Arabidopsis thaliana and Nicotiana benthamiana as well as crops such as rice, by using a range of transformation platforms (protoplast transfection, agroinfiltration and the generation of stable transgenic plants), by targeting both endogenous genes and transgenes and by exploiting both NHEJ and HR to generate small deletions, targeted insertions and multiplex genome modifications [9-11]. Subsequent work focused on additional crop species such as sorghum and rice [12], wheat [13, 14] and maize. Most recently, the CRISPR/Cas9 system was shown to work in tomato hairy roots following transformation with Agrobacterium rhizogenes [15] and was the first genome editing platform used in the fruit crop sweet orange [16].

Interestingly, four independent groups have shown that the CRISPR/Cas9 system can introduce biallelic or homozygous mutations directly in the first generation of rice and tomato transformants, highlighting the exceptionally high efficiency of the system in these species [9, 17-19]. It was also shown in Arabidopsis, rice and tomato that the genetic changes induced by Cas9/gRNA were present in the germ line and segregated normally in subsequent generations without further modifications [6, 17-21].

CRISPR stands for Clustered Regularly Interspaced Short Palindromic Repeats. The discovery of the type II prokaryotic CRISPR “immune system” has allowed for the development for an RNA-guided genome editing tool that is simple, easy and quick to implement. The CRISPR/Cas9 system consists of a single monomeric protein and a chimeric RNA. A 20-nt sequence in the gRNA confers sequence specificity and cleavage is mediated by the Cas9 protein. Watson–Crick base pairing with the target DNA sequence is the basis for gRNA-based cleavage, making sophisticated protein engineering for each target unnecessary. Only a 20 nt in the gRNA is needs to be modified to facilitate the recognize a different target (Figure 1).

CRISPR/Cas9 immune defense system in bacteria and archaea
CRISPR/Cas9 is a new technology for editing genomes of higher organisms, based on the immune system of bacteria. This system is based on specific areas of bacterial DNA, short palindromic cluster repeats, or CRISPR. Between identical replicas there are different DNA fragments - spacers, many of
Advantages of CRISPR/Cas9 genome editing

Main advantages of CRISPR/Cas9 are in terms of simplicity, accessibility, cost and versatility. Advantages of the CRISPR/Cas9 system in comparison with ZFNs or TALENs have no objections. Multiplex editing with the CRISPR/Cas9 system simply requires the monomeric Cas9 protein and any number of different sequence-specific gRNAs. In contrast, multiplex editing with ZFNs or TALENs requires separate dimeric proteins specific for each target site. Unlike its predecessors, the CRISPR/Cas9 system does not require any protein engineering steps, making it much more straightforward to test multiple gRNAs for each target gene. Furthermore, only 20 nt in the gRNA sequence need to be changed to confer a different target specificity, which means that cloning is also unnecessary. Any number of gRNAs can be produced by in vitro transcription using two complementary annealed oligonucleotides. This allows the inexpensive assembly of large gRNA libraries so that the CRISPR/Cas9 system can be used for high-throughput functional genomics applications, bringing genome editing within the budget of any molecular biology laboratory [27, 28].

CRISPR components and mechanism

The natural CRISPR/Cas9 consists of a Cas9 protein, a CRISPR RNA (crRNA), and a trans-activating crRNA (tracrRNA). In gene editing applications, crRNA and tracrRNA are often fused into a single guide RNA (sgRNA). The crRNA is the part that is complementary to the target DNA sequence, thus recognizing the sequence to be cleaved. The tracrRNA is a small helper RNA that enables maturation of the crRNA and functions as a scaffold for the crRNA-Cas9 interaction. In nature, the crRNA and tracrRNA segments exist as a duplex. Synthetically, they can be engineered as one seamless fusion sequence known as single-guide RNA (sgRNA). The ribonuclease protein invades the target with crRNA guide sequence by forming a 20-bp RNA/DNA hybrid and displacing the opposite DNA strand after it encounters a protospacer adjacent motif (PAM), such as NGG.

The design of gRNA and choice of the nuclease depends on the desired application. The right choice of components is crucial for a successful experiment. As the gRNA will guide the nuclease to the cutting site, it is important to ensure that the design yields minimum off-target (unintended) cuts while also providing maximum on-target efficiency. Nevertheless, while designing gRNA for KO experiments, it is essential to ensure that the sequence conforms to the following criteria. Sequences that target exons are preferred to minimize the risk of splicing of the targeted gene from the mRNA. In addition, sequences coding for N-terminus of protein are selected, as they are more likely to damage the region critical for protein-function. To edit the genome using the CRISPR/Cas9 system, a single protein (unlike TALEN, Zn-fingers) is used, and the RNA guide can be created in a short time in the laboratory or purchased [20 - 28].

Several platforms are available to help design guide RNAs by predicting their on-target and off-target activity. We have listed some of these below: Synthego’s free design tool is one of the fastest and most efficient design tools available for researchers. The tool offers easy design of synthetic sgRNAs with up to 97% editing efficiency and the lowest off-target effects. It includes a library of more than 100,000 genomes and 9,000 species, and offers a convenient way to order your guides within the tool. You can also use the tool to validate gRNAs designed using other platforms. Free websites for gRNAs creation in silico: CRISPRdirect http://crispr.dbcls.jp/, CRISPR-PLANT www.genome.arizona.edu/ crispr, CRISPR-P http://ebi.hzau.edu.cn/crispr/, E-CRISP http://www.e-crisp.org/ E-CRISPR, CRISPR RGEN Toolhttp://www. genome.net/cas-designer/, Synthego, Addgene, Origene, [29] etc.

Selecting a nuclease. The most commonly used system in cell applications is derived from the Streptococcus pyogenes Type II CRISPR/Cas9 system and consists of the Cas9 protein as the nuclease
that cleaves double-strand DNA when guided by the crRNA and tracrRNA (dual RNA). Cas-protein is, though very large [1]. If in standard systems, several proteins are assembled into a complex that binds CRISPR RNA, and then this complex recognizes the viral DNA target and attracts another protein that "bites" the viral DNA, then in a system that was lucky to study Charpentier, one protein called Cas9, performs all these functions: and binds CRISPR RNA, and recognizes the target, and "bites" it. The Cas9 nuclease recognizes the motif 5’-NGG-3’.

However, other nucleases that exist in nature or have been developed in the lab can be used depending on the specific need of the experiment. Cas9 nickase is a modified version of the Cas9 protein, which nicks a single DNA strand, rather than generating a DSB.

Cpf1 is another common nuclease, which stands for ‘CRISPR from Prevotella and Francisella’, recognizes and binds a different PAM 5’TNN-3’, is preferred for experiments relying on the HDR repair outcome and does not require a tracrRNA.

S. aureus Cas9 (SaCas9) from the species Staphylococcus aureus, recognizes the same PAM as SpCas9, much smaller by length of about 1kb.

CRISPR loci in a bacterium contain "spacers" (viral DNA inserted into a CRISPR locus) that in type II adaptive immune systems were created from invading viral or plasmid DNA (called "protospacers"). On subsequent invasion, Cas9 nuclease attaches to tracrRNA: crRNA which guides Cas9 to the invading protospacer sequence. But Cas9 will not cleave the protospacer sequence unless there is an adjacent PAM sequence. The spacer in the bacterial CRISPR loci will not contain a PAM sequence, and will thus not be cut by the nuclease. But the protospacer in the invading virus or plasmid will contain the PAM sequence, and will thus be cleaved by the Cas9 nuclease [4]. For editing genes, guideRNAs (gRNAs) are synthesized to perform the function of the tracrRNA: crRNA complex in recognizing gene sequences having a PAM sequence at the 3’-end [7, 8].

The canonical PAM is the sequence 5’-NGG-3’ where "N" is any nucleobase followed by two guanine ("G") nucleobases [9]. Guide RNAs (gRNAs) can transport Cas9 to anywhere in the genome for gene editing, but no editing can occur at any site other than one at which Cas9 recognizes PAM. The canonical PAM is associated with the Cas9 nuclease of Streptococcus pyogenes (designated SpCas9), whereas different PAMs are associated with the Cas9 proteins of the bacteria Neisseria meningitidis, Treponema denticola, and Streptococcus thermophilus [10]. 5’-NGA-3’ can be a highly efficient non-canonical PAM for human cells, but efficiency varies with genome location [11]. Attempts have been made to engineer Cas9s to recognize different PAMs to improve ability of CRISPR-Cas9 to do gene editing at any desired genome location [12]. Using either Agrobacterium tumefaciens or by trans-fecting plasmids that encode them, programmable nucleases can be delivered into plant cells, where these nucleases cleave chromosomal target sites in a sequence-dependent manner. The result is site-specific DNA double-strand breaks (DSBs) whose repair by endogenous systems results in targeted genome modifications.

So, the mechanism of genomic editing using CRISPR/Cas9 includes the next points. To correct the "wrong" gene, we need a very precise molecular "scalpel" that will find a mutant sequence of nucleotides and can "cut" it from DNA. This "scalpel" is Cas9. With the help of an RNA guide, the sequence of which coincides with the desired place, he can make a break in the right place of the genome. The recognition of the target takes place on a length of 20-30 nucleotides. On average, sequences of this length are found once in the human or plant genome, which allows for accuracy.

For genome editing purposes, generation of a targeted double-strand DNA break (DSB) is the key event that opens up multiple repair options both for the cell and the genome engineer [19, 20]. Such breaks are generally repaired by one of two pathways, homology-directed repair (HDR) or non-homologous end joining (NHEJ). Cells use NHEJ more frequently than HDR because the latter requires a template homologous to the regions flanking the break and to insertions or deletions (INDELs) at its position which can result in functional knock out of a gene [1, 2, 14, 22, 23, 29] (Figures 1, 4).

In all phases of the cell cycle other than S phase, a homologous region of the chromosome is rarely in close enough proximity to act as this template, and thus NHEJ acts as a stop gap to quickly repair the break and maintain chromosomal integrity. NHEJ is an error-prone process that uses ligases, nucleases and polymerases to reseal a break, and generally results in nucleotides being inserted or deleted (indels) in an unpredictable process. If the break occurs in a protein-coding region, these indels will often result in a frame shift mutation and subsequent premature stop codon, abrogating the protein’s function.
targeted DSB engineered in a coding sequence of a protein of interest may thus result in a loss-of-function allele. As the efficacy of this process varies based on the target site and experimental conditions, a mixture of homozygous null, heterozygous and unmodified cells will be present, and this heterogeneity must be taken into account when interpreting observed phenotypes [23, 30, 31].

The magic of CRISPR is in its ability to force a DSB event. Cells must repair DSBs, or risk dying. Thus, all of the editing that comes from CRISPR is due to the cell’s innate ability to repair itself. The cell will not die from inserting a rupture into DNA, as it will be corrected by a healthy copy of the pair chromosome due to the natural process of DNA repair. If the pair chromosome is not present, as in the case of hemophilia, it is possible to insert the "right" gene site into the cell simultaneously with Cas9 and RNA guide and use it as a template for healing the inserted rupture [29-31].

In other worlds, NHEJ typically leads to a frameshift mutation and a knockout of the targeted genetic element’s function. NHEJ-mediated gene knockouts are the simplest form of targeted modification, and these could be used e.g., to eliminate genes that negatively affect food quality, to confer susceptibility to pathogens or to divert metabolic flux away from valuable end-products [14]. The insertion of large sequences by NHEJ or HR would allow the introduction of transgenes at defined loci that promote high-level transcription and do not interfere with the activity of endogenous genes.

Alternatively, if the objective of the experiment is to replace the targeted genetic element with a different sequence (e.g., gene insertion, single-base editing, etc.), the cell can be directed towards an alternative repair pathway, homology directed repair (HDR). To accomplish this, a homologous DNA template bearing the desired sequence must be introduced in the cell, along with the CRISPR components. A certain number of cells will use this template to repair the broken sequence via homologous recombination, thereby incorporating the desired edits into the genome.

Using CRISPR/Cas9, you can do multiplex editing of several wrong genes at once. To do this, just enter the protein Cas9 and several different RNA guides. Each of them will send Cas9 to its own target, and together they will eliminate the genetic problem. In general, the described mechanism functions due to the principle of complementarity, which was first proposed by Jim Watson and Francis Crick (1952) in their famous model of double-stranded DNA.

In the other opinion, there are three types of genome modifications. First type of modification is the insertion of small insertions and deletions (INDEL - INsertion or DELetion) to knock out any gene, while INDEL can lead to a shift of the reading frame or the destruction of the splice site. The second type of modification is the introduction of target DNA sequence into a specific region of genome. The third type of modification is creation of large deletions (from hundreds to hundreds of thousands of bases) [30-31].

**Specialty of CRISPR/Cas9 editing for monocots**

A codon optimized Cas9 protein and a gRNA are expressed from a single vector and provided as ready-to-use, transfection-grade DNA. The native Cas9 coding sequence was codon optimized for expression in dicots [32-38] and monocots [39-42], respectively. The monocot Cas9 constructs contain a monocot U6 promoter for sgRNA expression. The plant selection markers include hygromycin B resistance gene, neomycin phosphotransferase gene, and the bar gene (phosphinothricin acetyl transferase) [29, 43].

In the field of plants genome engineering, cereals such as wheat, maize and barley with big complicated genomes are difficult material for transformation [39-42]. At the same time, cereals are the main food products of more than half of the world’s population and serve as important sources of vegetable protein, carbohydrates, vitamins, mineral salts, bioethanol and cellulose. The first reports of successful genome editing using the CRISPR/Cas9 dated to 2013, so there are still no proven methods for genomic editing in cereal plants with complex genomes [29, 44, 45].

The process of monocots genomic editing includes the following stages: 1) selection of target sequences and design of gRNA; 2) construction of genetic vectors carrying the nuclease Cas9, sgRNA; 3) delivery of "editing tools" to plant cells; 4) detection of changes in genomic DNA; 5) clean the expression cassette with foreign DNA [29].

Cassettes of expression of specific gRNA includes promoter, gRNA, and terminator. Size of one gRNA expression cassette is 400-500 nucleotides. The promoters U3 and U6 from rice, wheat and maize
have been successfully used for gRNA monocots [6-11, 46]. Start for transcription initiation from the U3 promoter is A nucleotide, and from U6 promoter – G nucleotide.

Cassettes for Cas9 expression, as a rule, consist of a promoter adapted by the codon composition of Cas9 coding sequence, the nuclear localization signal and terminator. Optimized codon composition in accordance with the frequency of codon using in human genes, Cas9 has been successfully used to edit the genomes of a number of plants, including barley [27, 28]. At the same time, it was shown that Cas9, optimized for the codon composition of plant genes (both monocots and dicots), is more effective for monocot plants than the codon-optimized human genes [43, 46].

Many genes of Gramineae family plants are characterized by increased GC-composition of 5′ region of coding part. Cas9, according to GC composition corresponding to the structure of the cereal genes (62.5%), with a total GC composition of 54.2%, demonstrated high efficiency of rice genome editing [29]. In monocot plants promoters of ubiquitin genes of maize or rice have been successfully used for Cas9 expression. Optimally composed expression cassettes let to achieve 90% of the editing efficiency in rice T₀ [36]. Genes of resistance to hygromycin (HPT) and bialaphos (Bar), under the control of the 35S promoter from Cauliflower mosaic virus, are often used as the selective markers for editing genomes of monocotyledonous plants.

Procedure of CRISPR/Cas9 genome editing tools includes: 1. Design sgRNA. Requires the identification of target sites with specific sequence criteria, while also avoiding the potential for off-target effects. 2. Transcribe and screen sgRNA in vitro. Quickly transcribe any sgRNA in vitro at high yields, without ligation. Don’t waste time delivering ineffective sgRNAs to cells - test the cleavage efficiencies of individual sgRNAs in vitro before performing gene editing in your target cells. 3. Deliver sgRNAs and Cas9 into cell. There are several options for delivering sgRNAs and Cas9 to your target cells: for plasmid delivery, Cas9/sgRNA co-expression vectors allow seamless insertion of sgRNAs, and express bright fluorescent markers. Use gEasicles, which are cell-derived nanovesicles, for efficient delivery of active Cas9 ribonucleoprotein (RNP) complexes to a broad range of cell types with reduced off-target effects and very low cytotoxicity. To prevent genomic integration of Cas9, using this single reagent to transfect target cells with Cas9 mRNA and sgRNAs without cytotoxic effects. This AAV2-based system for delivery of sgRNAs and Cas9 enables efficient gene editing in difficult-to-transfect cells without genomic integration of Cas9. 4. For sgRNAs and Cas9 delivering into barley, we are using our elaborated A. tumefaciens - mediated pollen germ – line transformation technique. 5. Detection Cas9 protein and confirmation gene editing. Confirm that Cas9 protein is being expressed in target cells. Ensure that your cell population contains mutations at your target locus by using a mismatch detection assay that outperforms a CEL-1 based assay. 6. Genotype determination. If there are indels on one or both copies of your target gene, Cas9/sgRNA-mediated in vitro cleavage reaction can accurately determine cell’s genotype after gene editing. 7. Identity indels. Characterize CRISPR/Cas9-induced indels with a simple four-step protocol using the Guide-it Indel Identification Kit [43].

For example, we study knock out of epigenetic factor of viruses’ translation initiation eIF4E: sc-9976, caused many virus translation initiation in barley. RNAi silencing of eIF4E has conferred resistance to multiple viruses in melon, cucumber and broad spectrum resistance to potyviruses in tomato [47]. More recently, Arabidopsis complete resistance to Turnip Mosaic Virus has been successfully engineered by editing eIF4E using the CRISPR/Cas9 tool [27]. CRISPR/Cas9 constructs and plasmids for use in plants have been shown in figures 5, 6. Electrophoreograms of Cas9 Streptococcus pyogenes recombinant protein in combination with DNA, sgRNA are represented in figures 7. Epigenetic factor of viruses’ translation initiation eIF4E is represented in figure 8, [46].

Delivery of CRISPR components into the cell

The CRISPR machinery is delivered in cells using different methods depending on the cell type and format of the CRISPR components [6-9, 23, 29, 30, 31].

In a lipid-based delivery system, cationic lipid reagents facilitate delivery of biomolecules into cells. This method is high-throughput, shows low cell-toxicity, and is applicable in various cell types. Traditionally used for delivering nucleic acids in cells, lipid-delivery method has recently been optimized for delivering ribonucleoprotein (RNP) format of CRISPR components in cells.
Electroporation enables delivery of the CRISPR machinery in cell types that are difficult to transform using lipid-based delivery systems. Application of a controlled, short electric pulse to the cells forms pores in the cell membrane, allowing entry of foreign material.

Nucleofection is a variant of electroporation, in which the electric pulse is optimized such that the nuclear membrane of the cells also forms pores. The CRISPR components are thus directly delivered inside the nucleus.

Microinjection is commonly used to inject the Cas9 and gRNA ribonucleoprotein complex in embryos, although it can also be used in cells. Zebrafish, mouse, and most recently human embryos have been manipulated using this technique.

Plant CRISPR/Cas9 products are intended for Agrobacterium-mediated plant transformation or biolistic microparticle bombardment or protoplast transformation.

Agrobacterium and bioballistic are the main methods of monocots transformation [9]. There are two different approaches to the expression of targeted genetic constructs in plant cells: transient (temporal) expression and stable transformation of the genome, which allows the production of transgenic plants [29]. Transient expression involves the introduction of a genetic construct simultaneously into the maximum number of cells without selecting the transformed variants, which results in a local peak of transgene expression in the cell population within a short time after transformation. A stable transformation consists of the insertion of genetic construct in the genomic DNA of a single plant cell, from which a transgenic plant can later be obtained by regeneration and growing on special selective media. The production of monocot transgenic plants takes a long time, for many cultures it is laborious and difficult technology [44, 45].

Most of the early work was carried out by the method of transient expression of genetic constructs in plant cells. The effect of the CRISPR/Cas9 system was demonstrated on wheat protoplasts [9, 13], cells of immature sorghum embryos [12], protoplasts and rice calli [11, 12], apple and maize protoplasts [35, 40]. At the same time, there were first reports of transgenic rice plants with mutations produced at target loci [36, 38], a little later there were publications of similar studies on wheat [39, 41, 42] and barley [33].

Agrobacterial transformation is an effective method of "tools" delivery if genome changes use such a mechanism of reparation as non-homologous stitching of the ends. This method was successfully used for genome editing of maize [40], rice [11, 18, 19] and barley [29, 33]. Bioballistics was successfully used for wheat [13], maize [40] and rice [19]. Experimental evidence suggests that editing the genomes of rice and maize with using of homologous recombination mechanism was more effective in the case of bioballistics transformation [29].

We are developing a stable germ-line genetic transformation in wheat, soybean and barley by agrobacterium pipetting of targeted genes into stigma of flower before anthesis and transporting them into mature but not divided zygote, using natural pollen tubes.

Analysis of CRISPR editing

After transfecting the cells, the efficiency of DNA editing using CRISPR needs to be determined. A qualitative approach involves treatment of the CRISPR-edited DNA and the non-CRISPR edited DNA (control) with an enzyme that cuts DNA at mismatched sites. Gene deletions by the CRISPR system often result in mismatched bases during DNA repair. Thus, the CRISPR-edited DNA shows multiple small fragments after size-based separation in a gel, while the control shows a single band of uncut DNA. This method is a simple and crude way of estimating the CRISPR editing efficiency.

The quantification consists in determining the proportion of mutated cells in the case of cell culture or the ratio of plants carrying and not carrying mutations at the target locus in the production of transgenic plants. A qualitative assessment includes determining the type of mutations obtained. The most informative method is the amplification, cloning and sequencing of a fragment of genomic DNA containing a target sequence. Restriction endonucleases and PCR are used to enrich the preparation with the modified variants and to estimate the fraction of the mutated DNA [14]. The insertion of two double-stranded breaks into the knockout gene (due to multiplex editing) causes deletions, which are easily detected by PCR on the basis of the analysis of the length of the amplification product in the gel and sequence [11, 23, 29].
Alternatively, next generation sequencing (NGS) is a gene sequencing method that can be used for accurate and quantitative analysis of the editing efficiency of CRISPR. It also provides additional information regarding off-target edits in the DNA. However, due to the high cost of this method, alternative cost-effective and semi-quantitative methods are also used.

Optimization of on-target activity

Although the CRISPR/Cas9 system is an excellent tool for genome editing, the extent of off-target mutation needs to be investigated in more detail as well as the differences in cleavage efficiency among different but perfectly matched targets [23, 27, 28]. When attempting to design sgRNAs to target a gene of interest, CRISPR technology presents an embarrassment of riches, as the number of potential sgRNA sequences scales with the size of the gene. As either the coding or template strand of the DNA may serve as a target, the *S. pyogenes* PAM site (NGG) appears on average once every 8 nt. How, then, to choose from the dozens to hundreds of potential sgRNA sequences? Avoidance of off-target activity, discussed in more detail below, may be used to eliminate some sgRNA sequences, and target-specific features may be incorporated to choose sgRNAs that are more likely to be effective. For example, to generate loss-of-function alleles of protein coding genes, targeting closer to the N-terminus increases the chance that a frameshift allele will be deleterious, as more of the coding sequence will be disrupted. Likewise, for CRISPR technology, proximity to the transcriptional start site is critical for recruiting appropriate factors [18, 23, 30, 31]. The first genome-wide libraries were designed according to these general criteria, but did not take into account any sequence-specific information that may enhance on-target activity.

Interestingly, an extended PAM sequence beyond the canonical NGG motif was shown to affect activity, with CGGH being the most-optimal sequence (where H = A, C or T). This quantitative analysis led to a predictive model for designing optimal sgRNA sequences for any target of interest. Initial studies have shown that there are sequence features that affect the ability of Cas9 to bind sgRNAs, cleave DNA, and result in a loss-of-function allele. Further characterization of these rules, and incorporation of them into library design, will be critical for successful deployment of genome-wide libraries, and will result in libraries with progressively higher fractions of active sgRNAs.

Minimizing off-target activity

While optimizing on-target activity has clear ramifications for use of CRISPR technology, understanding off-target effects is equally important to avoid erroneous interpretation of experimental results. Initial experiments in mammalian cells showed that *S. pyogenes* Cas9 cleavage activity tolerates a number of mismatches between the sgRNA and the DNA target [10]. In general, mismatches closer to the 5’ end of the RNA are more tolerated than mismatches close to the PAM. Currently, there are not enough data to create fully predictive models of when an sgRNA will lead to appreciable levels of off-target DNA cleavage, as the exact base composition of the mismatch appears to affect activity [10, 23]. Given the tolerance of mismatches by Cas9 and the likelihood of finding certain sequence motifs in multiple locations in the genome, it is essential to consider possible off-target locations when designing sgRNA sequences. Several strategies have been developed to minimize the off-target effects of Cas9. One uses a shortened sgRNA of 17 nt, rather than the standard 20 nt sequence; under the conditions tested, the 17 nt [23].

Prospects for application of breakthrough technologies in Breeding: the CRISPR/Cas9 system for plant genome editing

“Although future studies are needed to examine the germ line transmission and heritability of the CRISPR/Cas-induced mutations and to evaluate any potential off-target effects of the CRISPR/Cas, our results here suggest that the CRISPR/Cas technology will make targeted gene editing a routine practice not only in model plants but also in crops”- Z. Feng [2].

The potential efficiency of CRISPR/Cas9 technology is much higher than traditional breeding approaches, and excludes the residual portion of the donor genome in editing crops (Figure 9).

About 10 years ago the first results on genome editing were achieved on plants, and during the last 4 years, thanks to the use of the relatively simple and convenient CRISPR/Cas9 system, there has been a sharp increase in the number of published works reporting successful editing genome of plants, including
the directed modification of economically valuable genes of cultivated plants (potatoes, cabbage, tomato, maize, rice, wheat, barley, soybeans, sorghum) [10-15, 17-20, 29-42, 44, 45]. Published works demonstrate the possibility of obtaining non-transgenic plants using CRISPR/Cas9 system with specific predetermined mutations stably inherited in generations. This possibility offers the challenge to obtain varieties with predetermined mono- and olygogenic traits.

The main practical advantage of CRISPR/Cas9 is the ease of multiplexing. The simultaneous introduction of DSBs at multiple sites can be used to edit several genes at the same time and can be particularly useful to knock out redundant genes or parallel pathways. The same strategy can also be used to engineer large genomic deletions or inversions by targeting two widely spaced cleavage sites on the same chromosome [7, 13, 19, 32]. It is possible to edit several genes simultaneously by introducing multiple or long DNA breaks in the genome to embed a whole complex of useful genes that will be transmitted in the offspring as a single locus.

The average term for the creation of a stable genotype by CRISPR/Cas9 technology is 2 years, which is 3-4 times faster than 10-12 years by backcross methods of conventional breeding, or 2-3 times faster than modern methods of marker-assisted selection (MAS) + marker double haploid (MDG) + backcross (BC) or obtaining double haploid hybrids [36 - 39].

Studies of 145 target genes in 15 crops obtained for 4.5 years from 2013 to 2018, demonstrated the possibility of obtaining modified non-transgenic plants. Editing of 37 genes was related to improvement of crops yield and stress resistance. In these studies, the ability to get transgene-free modified plants was widely demonstrated. In most of these cases, modifications resulted in knockout of the genes such as negative growth and development regulators or negative regulators of plant resistance. In most cases, the phenotype of modified plants was assessed, and the presence of desired changes was shown. Essential success has been achieved over a short period since the first publications on CRISPR/Cas application in plants [32 - 45, 47].

Generally, the CRISPR/Cas9 system for plant genome editing is a breakthrough technology in breeding, and the main prospect for creation of elite high yielding crops today and in the nearest future with social and economic benefit, in comparison with other breeding approaches: conventional breeding, non-targeted mutagenesis (like chemical mutagenesis) with polygene control of indexes, double haploids and marker assisted selections.

**Discover regulations for gene editing**

In the United States, a product-oriented concept has been adopted and it is established that CRISPR/Cas9 genome edited plants are not GMOs, as contain no recombinant foreign DNA.

In Europe, Russia and Kazakhstan, a process-oriented concept is still adopted and plants with edited genomes are GMOs.

Removal of undesirable plasmid DNA including the Cas9 and guide RNA achieved following segregation and screening of ‘clean’ plants in the next generation that carry only the edited event.

In the case of using CRISPR/Cas9, several methods are possible for creating of non-transgenic modified plants: 1) by using programmable nucleases, including on the basis of temporary expression of nuclease components using agroinfiltration or viral vectors, or direct delivery of components in the form of functional gRNA and Cas9 protein; 2) by integrating the transgenes gRNA and Cas9 protein into a chromosome different from that of the edited gene, so as to get rid of transgenic structures due to independent inheritance in the offspring; 3) transient expression of structures carrying elements of the CRISPR/Cas9 system, without integrating them into genomic DNA. It is assumed that the temporary presence of nucleases and gRNA in the cell may be sufficient to introduce the necessary changes in the genome. It is shown that this principle can be realized in monocot plants [29, 48].

Although the European regulatory framework for genetically modified crops focuses on the process and not the product (hence two identical plants produced by conventional mutagenesis and genetic engineering would be regulated differently under the current guidelines), there is hope and confidence that plants altered by the excision of a few nucleotides using genome editing tools such as CRISPR/Cas9 would not be classified as genetically modified organisms [49, 50]. The classification of genome edited plants is currently under review to decide whether new breeding technologies including CRISPR/Cas9 are exempt from GM classification.
CRISPR in the future

CRISPR has received a lot of attention primarily due to its ability to genetically edit living organisms. However, while this side of CRISPR occupies the spotlight, researchers have begun tinkering with the technology to unlock its vast potentials that go beyond the applications discussed so far.

Scientists are now using a modified version of CRISPR to explore epigenomics -the genome-wide set of chemical groups that adorn DNA and its associated histone packaging proteins. Previously, researchers were merely able to catalogue the correlation between epigenetic markers and gene expression in cells. Now, a CRISPR complex that is capable of acetylating histone proteins at precise locations dictated by the complex’s gRNA has been developed [22-26]. Such technologies can shed light on the causal relationship between epigenetic markers and gene expression in the future.

CRISPR is also enabling elucidation of large portions of the human genome, the function of the vast majority of which is unknown. Scientists have long been trying to identify the location and function of ‘non-gene’ genetic elements that do not code for proteins but are thought to have important regulatory roles in expression. CRISPR is allowing researchers to knock out these previously uncharted regions to study their role in the cell [31, 32].

Multiple gRNAs targeting the same promoter also demonstrate synergistic effects, indicating that tuning the level of transcriptional control is possible using this approach [25]. Furthermore, multiple gRNAs targeting different promoters allow the simultaneous inducible regulation of different genes. Two independent research groups have already extended this approach by layering CRISPR regulatory devices based on either transcriptional activators or repressors to create functional cascaded circuits. In this context, another peculiar feature of the CRISPR/Cas9 system is the ability to use orthogonal Cas9 proteins to separately and simultaneously carry out genome editing and gene regulation in the same cell [26]. Functional genomics attempts to understand the genome by perturbing the flow of information from DNA to RNA to protein, in order to learn how gene dysfunction leads to disease. CRISPR/Cas9 technology is the newest tool in the geneticist's toolbox, allowing researchers to edit DNA with unprecedented ease, speed and accuracy, and representing a novel means to perform genome-wide genetic screens to discover gene function [44, 45].

CRISPR is not only paving the way for us to solve the most difficult of problems in the life sciences, but it is also enabling the scientific community to explore dimensions of the genome that we’ve been unable to study up until this point. Due to its adaptability across a wide range of species and its simplicity of use, CRISPR/Cas9 has quickly revolutionized genome engineering. The CRISPR/Cas9 technology promises to deliver some truly stunning advances within the coming decades, particularly in relation to human therapeutics, agricultural biology, and basic scientific research.

Conclusion

CRISPR/Cas9 technology has revolutionized gene manipulation capabilities in many species including crops. The multitude of functions that can be performed with CRISPR/Cas9 and its many derivatives make it a molecular tool that will open new opportunities in the complicated world of plant–pathogen interactions and help design durable crop resistance to pathogens.

CRISPR tool will help integrate omics data in order to fully understand and increase crop improvement. Generally, the CRISPR/Cas9 system for plant genome editing is a breakthrough technology in breeding, and prospect for creation of elite high yielding crops today and in the nearest future with great social and economic benefit.

The agricultural applications described review represent only the very first, initial uses of this exciting technology, and we can expect many more valuable opportunities for agriculture in the near future. Although we may be in the heyday of CRISPR technology, we remain at the early stages of fully understanding the system and expanding its potential.

Funding

We acknowledge funding for this research from the Scientific Council of Ministry of Education and Sciences of Republic of Kazakhstan, Project IRN APO5132774, 2018-2020.
Cas9 is directed to its DNA target by base pairing between the gRNA and DNA. A PAM motif downstream of the gRNA-binding region is required for Cas9 recognition and cleavage. Cas9/gRNA cuts both strands of the target DNA, triggering endogenous DSB repair. For a knockout experiment, the DSB is repaired via the error-prone NHEJ pathway, which introduces an inDel at the DSB site that knocks out gene function. In a knock-in experiment, the DSB is repaired by HDR using the donor template present, resulting in the donor DNA sequence integrating into the DSB site [Origene. https://www.origene.com/products/gene../crispr-cas9, 2015].

Figure 2 - Overview of CRISPR/Cas bacterial immune system

(A) A typical structure of CRISPR locus; (B) Illustration of new spacer acquisition and invading DNA cleavage [Feng Z., Yan W. and Xiong G., 2014].
Mature crRNA guides Cas9 to the target site of invading phage DNA. The DNA single-strand matching crRNA and opposite strand are cut, respectively, by the HNH nuclease domain and RuvC-like nuclease domain of Cas9, generating a DSB at the target site [Feng Z., Yan W. and Xiong G., 2014].

The crRNA sequence (blue) and tracrRNA sequence (green) are fused together by a short loop (purple) to create an sgRNA. The 20 underlined nucleotides can be programmed to recognize any DNA sequence of interest. For S. pyogenes Cas9, the NGG PAM is required immediately downstream of the target site. The two strands of DNA are cut by the HNH and RuvC nuclease domains of Cas9. [Hartenian, Doench, 2015]
Figure 5 - Scheme of T-DNA unit for monocot CRISPR/Cas9 editing [https://www.sigmaaldrich.com/technical-documents/articles/biology/crispr-cas9-genome-editing.html]

Figure 6 - CRISPR/Cas9 knockout plasmid

Figure 7 - CAS9 *Streptococcus pyogenes* recombinant protein [Origene. https://www.origene.com/products/gene.../crispr-cas9, 2015].

Figure 8 - Epigenetic factor of viruses’ translation initiation eIF4E: sc-9976

Figure 9 - Advantages of CRISPR technology for Breeding [Khlestkina, Shumny, 2016]
REFERENCES

[41] CRISPR/Cas9 Nuclease RNA-guided Genome Editing. Sigma-Aldrich (2017) https://www.sigmaaldrich.com/.../crispr-cas9-genome-editing...
арналаң қарапайым, арзы және әмбедің құрамын бөліп табылады. Бактерия ағызғының Cas9 қулақша, бөтен ДНК қызметкер сіквystals-спецификалық бағыттайдың екі PKM, PKH CRISPR (crRNA) және трансбезендерінің crPHK (tracrRNA) қауымдалырады. Бұл бактериялардың құрылығы құлап, сутқоқрақтар мен жәрдімші өсімдіктер ағызғының гендері дәрежелі реанықталады, пайдаланылатын қатар, қаңғар көзде жәрдімші өсімдіктер геномынан реанықталады және селекцияда тәжірибі қолданылады. Сапарқұлақтың және вирусті патогендерге тәжірибі сұлулауын әкімдірлігін қолданысы құлғының сорты, Қазақстанда фермерлер мен жерғілікті экономикаға, өнімділігінің әр түрі мен әдет сапасының жақсыраққа байланысты пайда екеледі.

Түйін сөзлер: геномды инженерия, геномды реанықталу, CRISPR/Cas9, ән содауы, әуемдіраушылық дәрежелер, селекция перспективасы, ГМО емес жіктеу.

О.И. Кершаннык

Институт биология и биотехнологии растений КН МОН РК, Алматы, Казахстан

НОВАЯ РЕВОЛЮЦИОННАЯ БИОТЕХНОЛОГИЯ РЕДАКТИРОВАНИЯ ГЕНОМА CRISPR/CAS9
ДЛЯ СОЗДАНИЯ ЭЛИТНЫХ СОРТОВ СЕЛЬСКОХОЗЯЙСТВЕННЫХ КУЛЬТУР В КАЗАХСТАНЕ

Аннотация. В статье представлен обзор литературы по прорывной технологии, инструмент реадактирования генома под названием CRISPR/Cas9, которая произвела революцию в области науки о жизни с тех пор, как она появилась на рынке в 2013 году. CRISPR/Cas9 (Стр строгие регулярные чередующиеся короткие палindrome повторы, связанные с белком 9 - Cas9), является РНК регулируемым защитным механизмом в бактериальных и археальных, на котором основана новая популярная технология реадактирования генома у людей, животных и высших растений. Система CRISPR/Cas9 представляет собой простой, недорогой и универсальный инструмент для редактирования генома, лежащий в основе фундаментальных и прикладных исследований, популярность которого за последние 4,5 года стала известна как «сумасшествие CRISPR».

В организм бактерий нуклеаза Cas9 ассоциируется с двумя РНК, РНК CRISPR (crRNA) и трансактивирующей crPHK (tracrRNA), направляющейся сиквystals-специфическое расщепление чужеродной ДНК, в настоящее время широко используется как эффективный инструмент редактирования генома у высших организмов, включая улучшение сортов сельскохозяйственных культур.

Система CRISPR/Cas9 для редактирования генома приоритетных сельскохозяйственных культур является прорывной технологией в селекции – важнейшей перспективой для применения в ближайшем будущем, и может быть исключена из классификации ГМ.

Ключевые слова: геномная инженерия, редактирование генома, CRISPR/Cas9, нокаут гена, сельскохозяйственные культуры, перспективы селекции, не-ГМО классификация.

Information about authors:
Kershanskaya Olga Ivanovna, Institute Plant Biology and Biotechnology, SC MES RK, 45 Timiryazev str., Almaty, 050040, Kazakhstan, Chief Scientific Researcher, Doctor in Biological Sciences, Professor, Academician of Petri Prima Academy of Sciences, Russia. e-mail: gen_o_kersh@mail.ru; Tel. +7 727 3947557; mob. +7 7057140271