CHAPTER 2

Guilherme Souza Prado
Thaísa Tessutti Pinheiro
Josias Correa de Faria
Rosana Pereira Vianello

Genome editing via non-homologous end-joining (NHEJ) and ribonucleoproteins (RNP)

Introduction

Genome editing is a tool that allows the manipulation of genetic material to induce mutations in regions of interest so that the organism presents a desirable phenotype, and, in some cases, may not be considered transgenic. Because of this, its application in plants of agronomic interest has aroused great interest, mainly for simplifying regulatory steps.

Four main classes of nucleases are used in new gene-editing technologies, namely: meganucleases, zinc-finger nucleases (ZFNs); transcription activator-like effector nucleases (TALENs); and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) -associated (Cas) (Gaj et al., 2016). Meganucleases are endonucleases that cleave DNA in specific regions, recognizing sequences larger than 12 bp (base pairs). The LAGLIDADG family of meganucleases comprises I-CreI and I-SceI, which are the first enzymes that were used for gene editing. As only a few amino acid residues make contact with nucleotides, these enzymes have been manipulated to cleave genes at specific locations (Paques; Duchateau, 2007). Also, ZFNs are artificial enzymes that were also one of the first to be used to induce targeted mutations in plants. These enzymes were generated from the fusion between Zinc-Finger domains and FokI restriction enzyme domains. Like the other nucleases used in gene editing, the ZFN inserts breaks in the double-strand (DSBs- double-strand break) at specific locations in the DNA that are going to be repaired and, through failures during the repair mechanism, mutations can arise (Carroll, 2011). The main problem with the use of this system is the high toxicity of this enzyme and also the fact that it generates many off-target effects (Cornu et al., 2008; Ramirez et al., 2008), which can impair functions of genes that should not have its functioning changed (Zhang et al., 2015). With the consolidation of gene editing as a biotechnological tool and the search for more viable alternatives, the TALEN technique was developed from the fusion of TAL effectors (transcription

activator-like), proteins derived from pathogens that infect plants and can bind to DNA, to the FokI endonuclease domain (Moscow; Bogdanove, 2009). TALENs can be generated from a plasmid library that allows the assembly of the protein of interest in a combinatorial way. Although this technique has advantages over ZFN, because it uses proteins that are less toxic and easier to manipulate (Mussolino et al., 2014), it is still considered more laborious when compared to CRISPR (Nemudryi et al., 2014).

Among the techniques that have made genome editing more accessible and revolutionary, the CRISPR/Cas system stands out. This is due to the simplicity, low cost, and high efficiency of the technique when compared to other editing tools, such as ZFNs and TALENs. Due to this, the CRISPR/Cas system has become widely used in molecular biology laboratories on a routine basis in recent years and is currently the main genome editing tool used. The CRISPR system is based on the induction of mutations in DNA through cuts in the double-strand, in an extremely specific and targeted manner. These cuts are promoted by endonucleases from the Cas family and directed by the so-called guide RNAs (gRNA). They will be repaired by one of the two main repair pathways that the cell presents when the double-strand is damaged: homologous recombination (HR) or non-homologous end-joining (NHEJ) (Cardona; Morales, 2014). More details can be found in Chapter 1.

We will focus here on the use of the CRISPR/Cas system to induce mutations via NHEJ. The NHEJ mechanism acts in most of the cell cycle and is not based on homology to repair DNA, causing mutations such as insertions or deletions (indels) or even substitutions. In this way, the expression of the genes in the place where the repair occurred by NHEJ can, in general, be interrupted, and the encoded proteins may not be functional (Cui et al., 2019). To clarify the process of genome editing by CRISPR, this chapter aims to address the details of the NHEJ mechanism, as well as the use of ribonucleoprotein systems (RNPs) to promote editing, in addition to providing a practical procedure for the design of experiments.

Mechanisms of genetic repair

The CRISPR system acts promoting specific cuts in the double-stranded DNA in genomic regions in a targeted manner using a guide RNA (gRNA) which is then repaired, may leading to the modulation of genes or promoter regions. Thus, clarifying the main repair systems used by the cell to reconstruct the double-strand becomes essential for understanding the technique. The importance of the genetic information being transmitted through the generations in a precise way is evident, both by the efficient replication system and by the complexity and investment of

the cell in a repair system specialized in the correction of the diverse types of DNA disorders.

With this objective, to minimize the modifications in DNA, the cells have a mechanism of response to damage. One of them is the DDR (DNA Damage Response) system, which is efficient in DNA repair and acts by several pathways and enzymes, depending on the type of injury that has occurred. In this mechanism, when cells suffer damage, the replication process is suspended until the repair is carried out (Silva; Ideker, 2019). For this, a signaling cascade occurs generating changes in the chromatin, leaving the damaged end accessible to DNA repair proteins (Riches et al., 2008). With the damaged DNA end exposed, the Mre11-Rad50-NBS1 (MRN) complex binds to the DSB, recruiting proteins members of the MAT (Mutated Ataxia Telangiectasia) family, which regulate the response to the damage through phosphorylation (Lavin; Kozlov, 2007). Thus, cell cycle checkpoints and chromatin remodeling are activated; everything so that the DNA is repaired before the cell cycle continues (Silva; Ideker, 2019). However, in multicellular animals, if the damage to the genetic material is irremediable, apoptosis is activated so that the lesion is not perpetuated (Norbury; Zhivotovsky, 2004).

For the cell to repair the lesion in the genetic material, five main repair pathways are activated at different moments in the cell cycle. When the damage to the genetic material occurs in only one of the DNA strands (SSB, single-strand break), the repair mechanisms are of the mismatch repair (MMR) type, base excision repair (BER), and nucleotide excision repair (NER). However, when DNA is damaged in both strands (DSB), repair pathways can occur by homologous recombination (HR) or by non-homologous end-joining (NHEJ) (Cardona; Morales, 2014), which are the molecular basis of CRISPR technology.

Non-homologous end-joining (NHEJ)

The NHEJ pathway does not require template DNA and is responsible for 75% of repairs to the genetic material of animal cells (Mao et al., 2008). This repair path takes about 30 minutes to complete, while HR is completed within 7 hours or more (Mao et al., 2008). In the NHEJ pathway, when the DSB is repaired, indels are generally formed which, by modifying the original DNA sequence, can cause gene inactivation (Her; Bunting, 2018). NHEJ repair employs, for this, a series of essential factors that work to detect the DSBs, as well as align and repair the ends so that they can be connected again (Her; Bunting, 2018).

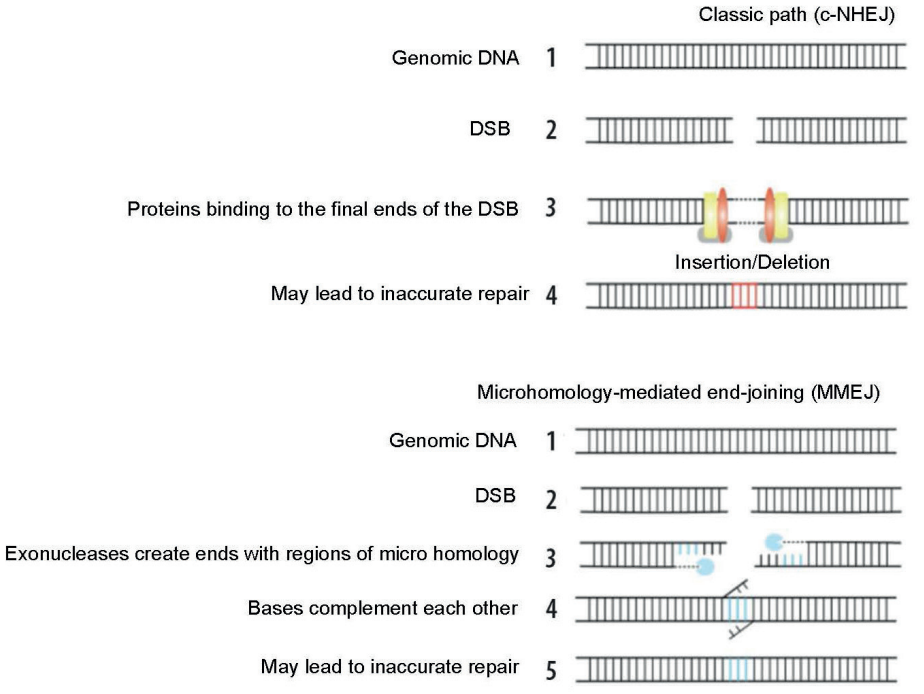


Figure 1. Repair pathways by non-homologous end-joining (NHEJ). (A) Classic non-homologous end-joining (c-NHEJ), which can lead to insertions or deletions (indels). (B) Final union mediated by microhomology (MMEJ), which always results in deletions.

Source: adapted from Deriano and Roth (2013).

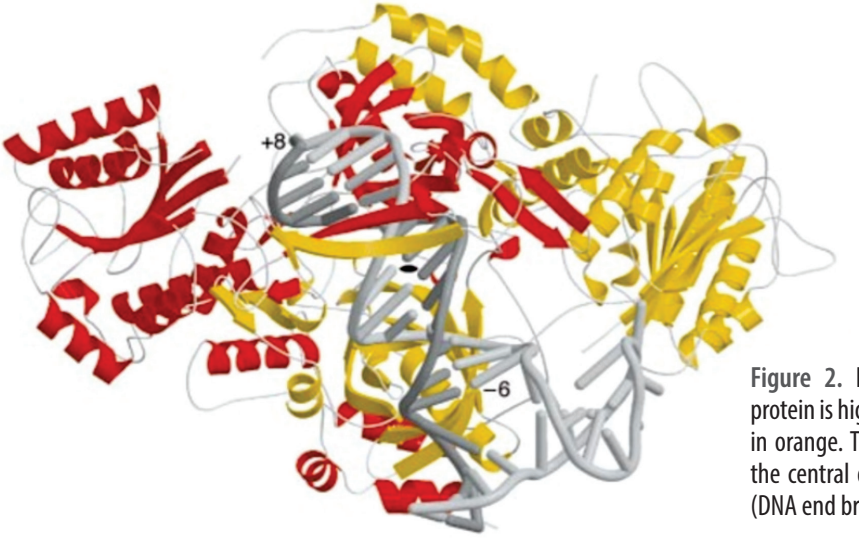


Figure 2. Ku complex: DNA. Ku70 protein is highlighted in red, and Ku80 in orange. The terminal base pairs of the central duplex are numbered +8 (DNA end broken) and -6.

Source: adapted from Walker et al. (2001).

Biochemical mechanism

In the presence of the DSB, the DDR damage response mechanism is activated. Thereby, a cascade of phosphorylation is activated, generating changes in chromatin, which leave the damaged end accessible to DNA repair proteins (Riches et al., 2008). In plants (Steinert et al., 2016) and mammals (Hartlerode; Scully, 2009), both HR and NHEJ are triggered.

NHEJ, in turn, can be subdivided into two routes: classic (c-NHEJ), which is the main chromosomal repair and rearrangement route (Lieber, 2010), and backup (b-NHEJ), also called alternative (a-NHEJ) or microhomology-mediated end-joining (MMEJ) (Deriano; Roth, 2013). This second pathway is activated if functional problems occur with c-NHEJ. However, it does not have such a precise repair mechanism, which generally leads to chromosomal translocations and excessive deletions (Deriano; Roth, 2013) (Figure 1).

The c-NHEJ pathway can be divided into three main stages: DSB recognition, DNA processing, and ligation (Yang et al., 2016). DSB recognition depends on Ku70-Ku80 (Ku) proteins, a heterodimer with two subunits of 70 kDa and 80 kDa, which intertwine in a ring that surrounds the end of the DSB, forming the Ku:DNA complex (Figure 2). This complex keeps the ends of the DSB relatively close so that there are no additional translocations or mergers (Downs; Jackson, 2004).

As Ku:DNA complex is formed at each end of the DSB, protein factors are recruited, such as the catalytic subunit of the DNA-dependent protein kinase (DNA-PKcs), complementary transversal X-ray repair protein 4 (XRCC4), factor type XLF, PAXX (paralogous to XRCC4 and XLF) and DNA-Ligase IV (Lig4), which keep the ends of the DNA close, forming a bridge between them for processing and ligation to occur (Davis et al., 2014).

If the DSB generates simple ends, they can be directly connected after the DSB is recognized. However, in most DSBs there are incompatible ends due to chemical modifications or incompatibility of the ends of the DNA strands, which end up preventing the strands from being connected directly. Thus, the DNA must be processed, and, for that, exonucleases degrade the 5'or 3' ends, generating regions of microhomology. This process, known as a recession, allows the two ends to become compatible.

During the recession of the final ends of the DSB, the DNA-PKcs act in a complex with the endonuclease Artemis which, after being phosphorylated by the DNA-PKcs, cleaves bonds from the single-strand and between the double-strands. Other factors involved in this process are polynucleotide kinase phosphatase (PNKP),

aprataxin (APTX), APTX-polynucleotide kinase-phosphatase-like factor 1 (APLF or PALF), werner (WRN), DNA polymerases, and deoxynucleotidyl terminal transferase (TdT) (Lieber, 2010).

After DSB processing, the ends are ready to be connected. All DNA processing and binding are coordinated by DNA-PKcs, which phosphorylate several proteins (Jiang et al., 2015). These proteins can also be autophosphorylated or trans-phosphorylated by other factors, such as ATM and ataxia related to telangiectasia (ART) (Davis et al., 2014). For double-strand binding to occur, specific ligases and the XRCC4-XLF-Lig4 complex are recruited, which is the main factor involved. For the formation of this complex, XRCC4 is anchored to the Ku:DNA-PKcs complex and, together with the XLF factor, forms a bridge with Lig4, promoting the connection of the DSB (Andres et al., 2012).

Genome editing via NHEJ – practical application

The application of the CRISPR/Cas system for genome editing in plants has been used successfully to obtain genetically improved cultivars. The characteristics manipulated with this tool range from nutritional improvement, efficient use of nutrients (Liang et al., 2014), disease resistance (Wang et al., 2014, Peng et al., 2017), tolerance to abiotic stresses (Mushtaq et al., 2018) until the increase in productivity (Ma et al., 2016). Thus, one of the aspects of improvement using CRISPR is the knockout, which eliminates gene function through indels derived from NHEJ repair.

Aiming at the resistance of wheat plants (*Triticum aestivum*) to the powdery mildew fungus (*Blumeria graminis* f. Sp. *Tritici*), gene editing was carried out via TALEN to introduce targeted mutations in three homologous genes encoding the Molo-Resistance Locus (MLO) proteins (*TaMLO-A1*, *TaMLO-B1,* and *TaMLO-D1*), and also editing via CRISPR to knockout the *TaMLO-A1* gene. MLO proteins act as negative regulators, repressing the plant's defense pathways against powdery mildew in uninfected tissues (Piffanelli et al., 2002). Thus, when the homologous MLO genes were edited via TALEN and CRISPR, there was an increase in the resistance of wheat plants to powdery mildew when compared to unedited plants. For the CRISPR construct, an expression vector containing the Cas9 CDS, the Ubiquitin 1 corn promoter, and the sgRNA was used. The vector was introduced into the wheat genome via biolistics, resulting in a mutation efficiency in regenerated plants equal to 5.6% (Wang et al., 2014).

In another study, researchers worked with maize (*Zea mays*), which has a high concentration of phytic acid, an antinutrient that is not digested by monogastric

animals and impairs the absorption of nutrients. With the use of CRISPR, a knockout was performed on the gene *Zea mays* Inositol Phosphate Kinase (*ZmIPK*), generating edited plants with a low concentration of phytic acid. In this study, the genetic transformation via protoplast transfection using two guides constructed to direct Cas9 to the *ZmIPK* gene showed a mutation efficiency of 16.4% and 19.1% for each guide used (Liang et al., 2014).

For soybean (*Glycine max* (L.) Merrill), the first successful work conducted as a proof of concept using the CRISPR system to perform directed mutagenesis in roots (hairy roots) demonstrated the effectiveness of the technique by turning off endogenous genes and the *bar* gene (Cai et al., 2015). In this study, seven vectors were constructed, one with gRNAs for the transgene *bar* and other six with gRNAs targeting two endogenous soybean genes (*GmFEI2* and *GmSHR*), with three guides for each gene (*GmFEI2-SP1*, *GmFEI2-SP2,* and *GmFEI2- SP3* for the *GmFEI2* gene and *GmSHR-SP1*, *GmSHR-SP2* and *GmSHRSP3* for *GmSHR*). Cas9 expression was driven by the maize Ubiquitin promoter, the gRNAs were driven by the *Arabidopsis* U6 promoter, and the GFP gene expression, for rapid visualization of transgenic hairy roots, was driven by the CAMV 35S promoter. The genetic transformation method used was based on *Agrobacterium rhizogenes* and, as plant material, transgenic homozygous plants were used for the *bar* gene (for editing the *bar* gene) and wild plants, not genetically transformed, for the editing of the *GmFEI2* and *GmSHR* genes. For the transgene *bar*, 11 mutations were identified in the 30 independent transgenic hairy roots analyzed, with a frequency of 1.3% to 21.0%. As for endogenous genes, in the *GmFEI2-SP1* guide, mutations were identified in 18 of the 30 independent transgenic hairy roots, with an indel frequency between 0.6% and 18.8%. Mutations at the site of the *GmFEI2-SP2* guide were identified in 28 of the 30 hairy roots, with the frequency of indels ranging from 1.0% to 95.0%. For the *GmFEI2-SP3* guide, 3 mutations were identified in the 30 independent transgenic hairy roots analyzed, with indels frequency between 5.4% and 27.0%. As for the *GmSHR-SP2* gene, 10 out of 22 hairy roots were identified, with an indel frequency of 8.7% to 30.0%. For *GmSHR-SP1*, mutations were found in 15 of the 30 roots analyzed, with the frequency of indels ranging from 2.3% to 21.3%. At the *GmSHR-SP2* site, mutations were found in 10 of the 22 roots analyzed, with the frequency of indels ranging from 8.7% to 30.0% (Cai et al., 2015). In *GmSHR-SP3* mutations were identified in 10 of the 28, with frequency ranging from 2.8% to 28.7%.

Promoter modulation and optimization

The use of NHEJ to modulate gene expression can also be applied to regulatory elements, such as promoter regions. This strategy makes the transcriptional activity, based on the modulation of transcription factors access to the promoter, be altered and, as a consequence, there is a change in the expression of genes (Cong et al., 2013). When promoter sequences are modified using genome editing techniques, it is called in vivo promoter engineering (IPE) (Pandiarajan; Grover, 2018). Some studies have conducted this approach aiming at generating variants of promoters with altered transcriptional activity (Li et al., 2020).

To obtain rice plants more adapted to soils with high salinity content, CRISPR technology was used to generate indels in regions of functional promoter units, the cis-regulatory elements (CREs) of the *OsRAV2* gene. This system was established to confirm that deletions in CREs of the promoter sequence of the *OsRAV2* gene (in GT-1 elements) would be related to resistance to saline stress. The *RAV* (Related to ABI3/VP1) genes encode proteins involved in the response to abiotic stress, with *OsRAV2* being specific in the response to environments with excess salinity in rice. In this way, a vector containing Cas9 and a gRNA directed to the GT-1 elements of the promoter region of the *OsRAV2* gene was constructed, and the genetic transformation was carried out via *Agrobacterium tumefaciens*. In this study, 12 plants were edited, and all showed the desired targeted mutations, which promoted the resistance of the plants to soils with high salt concentration (Duan et al., 2016).

In a study involving citrus canker, caused by *Xanthomonas citri*, a pathogen responsible for significant losses in the cultivation of *Citrus* spp., CRISPR was used in *Citrus sinensis* Osbeck, aiming at the modulation of the Lateral Organ Boundaries 1 (*CsLOB1*) gene promoter, associated with susceptibility to the pathogen. The gene had its induction in response to the infection stopped, and this approach generated plants with greater resistance to citrus canker compared to the wild type. To obtain the plants, a vector was coded for Cas9, in addition to the AtU6-1 promoter directing the expression of the gRNA directed to the promoter region of the *CsLOB*1 gene. The genetic transformation was carried out by *A. tumefaciens*, and the mutation efficiency was between 32.5% and 90.7% (Peng et al., 2017).

Ribonucleoprotein (RNPs) nuclease-sgRNA

RNPs: DNA-free system

Genome editing or engineering is, in most cases, conducted using plasmid vector systems carrying genes that, when integrated into the host's genome, encode the expression of the products necessary for the editing mechanism, being a nuclease and an RNA-guide (gRNA), in addition to a transformant selection marker gene, which confers resistance to an antibiotic or herbicide, when the target organism is a plant, and, eventually, reporter genes (Xie; Yang, 2013).

With this, new characteristics are incorporated into the host both due to the mutagenesis triggered in the host through the editing and due to the exogenous genes, that are introduced, which configures the event obtained as a transgenic GMO. In this case, concerns and problems related to regulation in biosafety and bioethics can become obstacles for the developed biotechnological product to become a market reality. To circumvent and/or avoid barriers that delay the arrival of the edited organism to the market, researchers usually appeal to methods that employ the elimination of inserted sequences utilizing hybridization and Mendelian segregation techniques, so that the edited organism is characterized only as GM, but not as transgenic, since the exogenous genetic material is lost (Cyranoski, 2015).

However, these techniques may become unfeasible for obtaining non-transgenic edited plants that have asexual reproduction processes, such as grapes (*Vitis* spp.), bananas (*Musa* spp.), and potatoes (*Solanum tuberosum*), due to the absence of recombination that would promote the elimination of transgenes in the offspring. They may also be little viable or little advantageous in the case of perennial plants, such as passion fruit (*Passiflora* spp.), eucalyptus (*Eucalyptus* spp.), and cherry (*Prunus* spp.), due to the impracticality of the transgenic elimination process, which becomes too long. Thus, plants with a short life cycle are more likely to be obtained through this method due to their greater practicality (Woo et al., 2015).

Given this scenario, the list of target plants becomes narrow, existing the need to develop or find alternatives to circumvent this limitation. Therefore, one of the strategies is the use of DNA-free gene-editing systems, from which ribonucleoproteins (RNPs) are the most well-known and its use in recent years has continued to expand to several species (Kanchiswamy, 2016).

In genome editing, RNPs are ribonucleoprotein complexes composed of a nuclease and one or more gRNAs. The gRNA is obtained by in vitro or chemical

synthesis, and it can also eventually be optimized so that its non-specificity is reduced, by truncating 17 nucleotide residues, unlike the 20 residues normally used for Cas9, for example. The gRNA is also stabilized when protected against degradation by RNases due to complexation with the nuclease (Kanchiswamy, 2016). The nuclease, in turn, can be obtained commercially from suppliers of biotechnology companies or expressed in recombinant form in a heterologous system, commonly bacterial systems based on *Escherichia coli*. After expression, the recombinant protein is purified, preferably by bioaffinity in a nickel chromatographic column, based on a 6-His-tag fused to one end of the nuclease.

According to reports in the scientific literature, RNPs are generally used to edit the genome of animal cells and, in comparison to plasmid vectors, they present very low cytotoxicity to the host. This is because cytotoxicity is often associated with the plasmid transfection process itself and with some reagents used to conduct the transfection.

Some advantages of RPNs compared to plasmids are highlighted below:

a) The plasmid mechanism is more complex, since it depends on the molecular machinery of transcription (for the nuclease and the gRNA) and translation (for the nuclease), since there is a need to integrate the genes of interest in the target genome, in the nucleus, so that the transcripts are produced and, in the case of genetic information for the nuclease, exported to the cytoplasm for gene translation, with the nuclease returning to the nucleus to perform the editing. With this, a continuous and complex transit is observed between different subcellular compartments, while in the use of RNPs the mechanism occurs directly since the preformed complex immediately goes to the nucleus and performs its function (Figure 3) (Amirkhanov; Stepanov, 2019; Wilbie et al., 2019).

b) The use of plasmids leads to gene integration, which culminates in the persistence of the expression of the editing agents, who remain functionally active for a long time, maintaining the editing activity, which increases the likelihood of off-target effects. On the other hand, RNPs do not depend on gene integration to carry out the editing, which occurs immediately after the delivery of the complex in the nucleus, with a peak of mutagenesis after about 24 hours, being rapidly degraded and therefore reducing the occurrence of off-target effects. (Amirkhanov; Stepanov, 2019).

c) In addition to the integration of the genes of interest, the plasmid backbone can also be integrated into the host genome and persist, even after the use of hybridization techniques, leaving molecular "footprints" (Mao et al., 2018),

which can become another problem in the regulation of the edited plant, as it characterizes a transgenic organism.

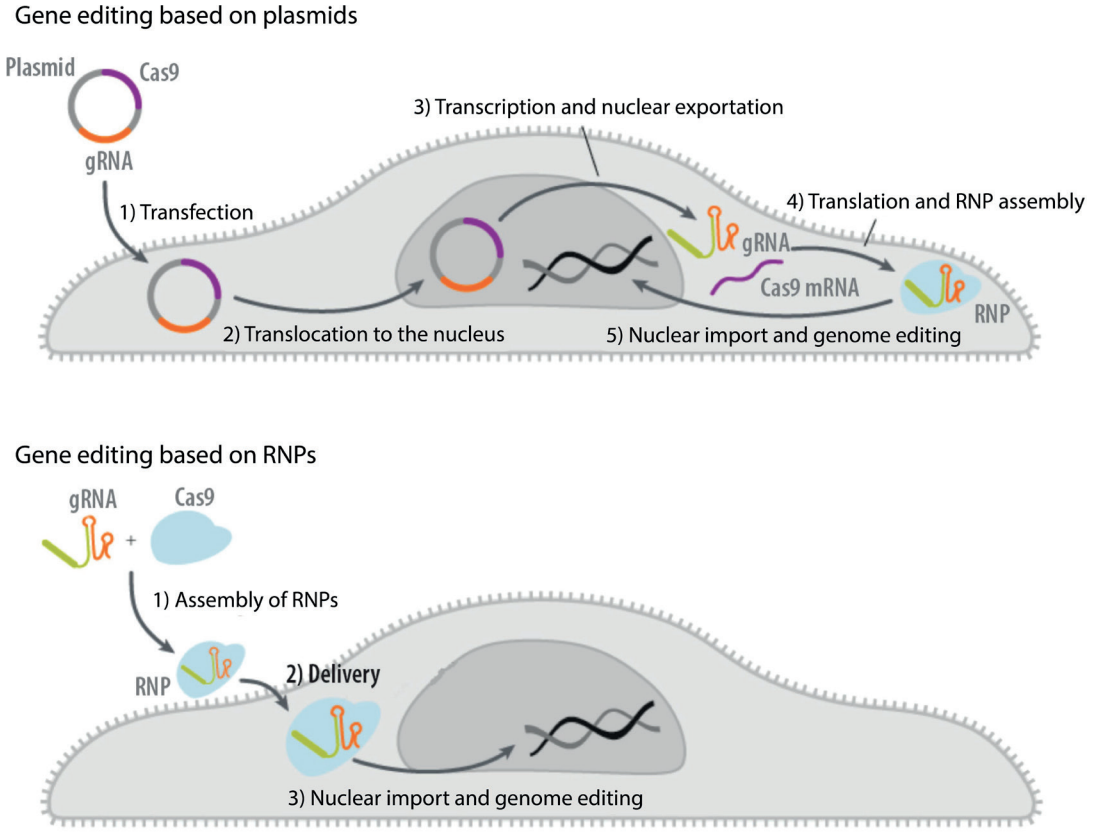


Figure 3. Plasmids versus RNPs in genome editing. (A) Plasmid vector carrying genes for the encoding of a Cas9 protein and a gRNA, showing stages of a more complex transit between different cellular compartments until the genome editing. (B) RNP is composed of a Cas9 complexed with a gRNA, showing a simple step of direct transit to the nucleus to carry out the editing of the genome.

Source: adapted from Spencer (2020).

RNPs are not the only DNA-free strategy for genome editing. Other systems have also been developed and their applications have been demonstrated not only for animal cells but also for plants of economic interest. Among them, IVTs (in vitro transcripts), and vectors based on viral replicons are alternatives that also dispense genomic integration, each of which has advantages and disadvantages

(Gil-Humanes et al., 2017; Liang et al., 2018). For the various advantages offered by vectors based on viral replicons, see Chapter 3.

Through the IVT system, mRNA molecules encoding the nuclease are delivered and, although IVT-derived mRNAs have the advantage of not integrating into the genome, mRNA molecules are more subject to degradation by RNases in the host's intracellular environment, resulting in less system stability. Also, long RNA molecules, such as those encoding the nuclease, are easily degraded during laboratory manipulation processes, making preparation difficult. Viral vectors, which are often used in replicon-based systems, require additional steps for cloning and genetic engineering that are unnecessary in the RNPs system (Gil-Humanes et al., 2017; Liang et al., 2018). While some viral vectors for animal cells lead to the integration of the viral genome into the host's DNA, in plants this does not occur, and the replicon remains active in the episomal form in the cytoplasm, dispensing the transgenic elimination steps of the edited plant.

As an example of reports of the application of RNPs in plants, Woo et al. (2015) established genome editing in plant systems via RNP for the first time, having as host organisms: *Arabidopsis thaliana*, rice (*Oryza sativa*), lettuce (*Lactuca sativa*) and tobacco-coyote (*Nicotiana attenuata*). The authors performed the transfection of protoplasts via PEG with pre-assembled RNPs, obtaining editing efficiency of up to 46% (8.4% to 44% for *A. thaliana*, rice, and tobacco, and 46% for lettuce). No off-target effects were detected and editing remained stable for the next generation. The mutations generated proved to be indistinguishable from the genetic variations that occurred naturally.

Svitashev et al. (2016) and Young et al. (2019), in turn, demonstrated the applicability of RNPs in maize (*Zea mays*), using particle bombardment for the transfection of the complex in immature embryos, aiming at the knockout of the *LIG1*, *Ms26,* and *Ms45* genes, located on different chromosomes. More specifically, Young et al. (2019) studied the off-target effects of this strategy compared to the use of plasmids, demonstrating the absence of these effects in the corn genome. Liang et al. (2017) compared RNPs with plasmid vectors in wheat (*Triticum aestivum*), verifying similarity of on-target activity, but five times less off-target effect when using RNPs, corroborating what was mentioned about the correlation between prolonged editing activity in the cell and the increase in off-target cases. Finally, Liang et al. (2018) compared RNPs with IVTs for *T. aestivum*, demonstrating that both were equivalent in terms of efficacy while eliminating genomic integration and giving rise to DNA-free (non-transgenic) edited plants. Other plant species have also been transfected with RNPs, such as grape (*Vitis vinifera*), apple (*Malus*

domestica) (Malnoy et al., 2016), petunia (*Petunia x hybrida*) (Subburaj et al., 2016), and potato (*S. tuberosum*) (Andersson et al., 2018)

In this case, the authors of these works used protoplasts as explants for transfection. Of these, only Andersson et al. (2018) regenerated tissues from the transfected protoplasts, obtaining sprouts derived from induced calli. For this work, the analysis of the editing efficiency was conducted using the leaf tissue itself as a source of genomic DNA sample, in which they concluded that 9% of the events obtained contained the expected mutation in at least one of the alleles of the target gene. On the other hand, both Malnoy et al. (2016) and Subburaj et al. (2016) analyzed genomic DNA samples obtained directly from protoplasts, without undergoing seedling regeneration. With that, Malnoy et al. (2016) reached an editing efficiency of 0.1% for grapes and 0.5% to 6.9% for apples, while Subburaj et al. (2016) obtained an efficiency of 2.4% to 21% in petunia.

As it is possible to observe, RNPs have wide-ranging applicability, with the potential to encompass other useful approaches for the development of edited and economically important plants that are DNA-free. Thus, like plasmids, IVTs, and vectors based on viral replicons, RNPs are suitable for generating knockout via indels from the NHEJ mechanism, as well as they can be adapted to knock-in from the HDR mechanism, which will be detailed in the next chapter. Besides, transcriptional regulation for editing and modulation of promoters via indels, aiming at overexpression or knockdown of genes, as well as for base editing (base editing) and the latest genome-accurate editing technology, prime editing, it is possible to be executed in plants using RNPs, which opens opportunities for research that demonstrate for the first time this viability and its success.

However, the limitation of the use of RNPs arises with approaches that require the persistence of the nuclease-sgRNA complex acting in the nucleus. This is because the nucleases of RNPs are degraded by intracellular proteinases, losing their biological activity as soon as their task of genome editing is completed. Therefore, studies involving modulation and epigenetic engineering, using activators and inhibitors via dCas9 in CRISPRa/CRISPRi system or studies aimed at imaging based on genomic marking with fluorescent probes, for example, become unfeasible through RNPs. Additionally, obtaining RNPs requires the in vitro production of sgRNA transcripts, which, once produced, have low durability due to their high sensitivity (Liang et al., 2018). Also, due to the factor mentioned concerning degradation, there is a short window available to RNPs to perform the editing, which usually leads to low mutagenesis efficiency by this mechanism. Finally, when working with RNPs, there is no use of selection marker genes that would assist in the discrimination of transformed explants, which makes RNPs a more laborious technique, as it

requires an additional step of molecular characterization to identify edited plants (Kanchiswamy, 2016).

In short, deepening and extending the use of RNPs in species already tested and in others not yet evaluated, respectively, is of great importance, within the technical limits of the approach, and can significantly impact world agriculture and, consequently, the quality of human life.

Delivery strategy and internalization of RNPs

The delivery of the molecular agents of genome editing is a fundamental step towards obtaining edited plants, and the parameters associated with it are directly linked to the success and practicality in the development of these events. Based on this, a wide variety of delivery methods for RNPs is already used or can be complemented with strategies not yet established for plants, although they are already widely applied to animal cells (Figure 4). These methods can be classified into:

1) **Simple transfection:** covers techniques for direct introduction of RNPs - although it is also used for plasmid systems and other mechanisms - in the host cell by physical or chemical methods, without the participation of complex and sophisticated reagents or systems to facilitate transfection. Examples:
   a) Electroporation: in the case of plants, it consists of the application of electric pulses in protoplasts, to promote the rapid opening of pores in the cell membrane, allowing the introduction of RNPs in the cell.
   b) PEG: polyethylene glycol is a polymer that promotes the internalization of RNPs and is often used for transfection of protoplasts as explants.
   c) Particle bombardment (biolistics): method mostly used for plants, consisting of the acceleration of microparticles composed, generally, of gold or tungsten, at speeds above 1.500 km/h, using the gene gun equipment. In this procedure, the RNP complex is prepared and precipitated on the microparticles, following well-established protocols. It is a technique classically used in the genetic transformation of plants. More information about this technique can be found in Chapter 5 of the book "Manual of genetic transformation of plants" (Vianna et al., 2015).
2) **Nanostructured systems**: includes polymers or other substances, produced in the form of particles on a nanometer scale, which act as nanocarriers for RNPs (Zuris et al., 2015). Li et al. (2015) reviewed potential non-viral vectors

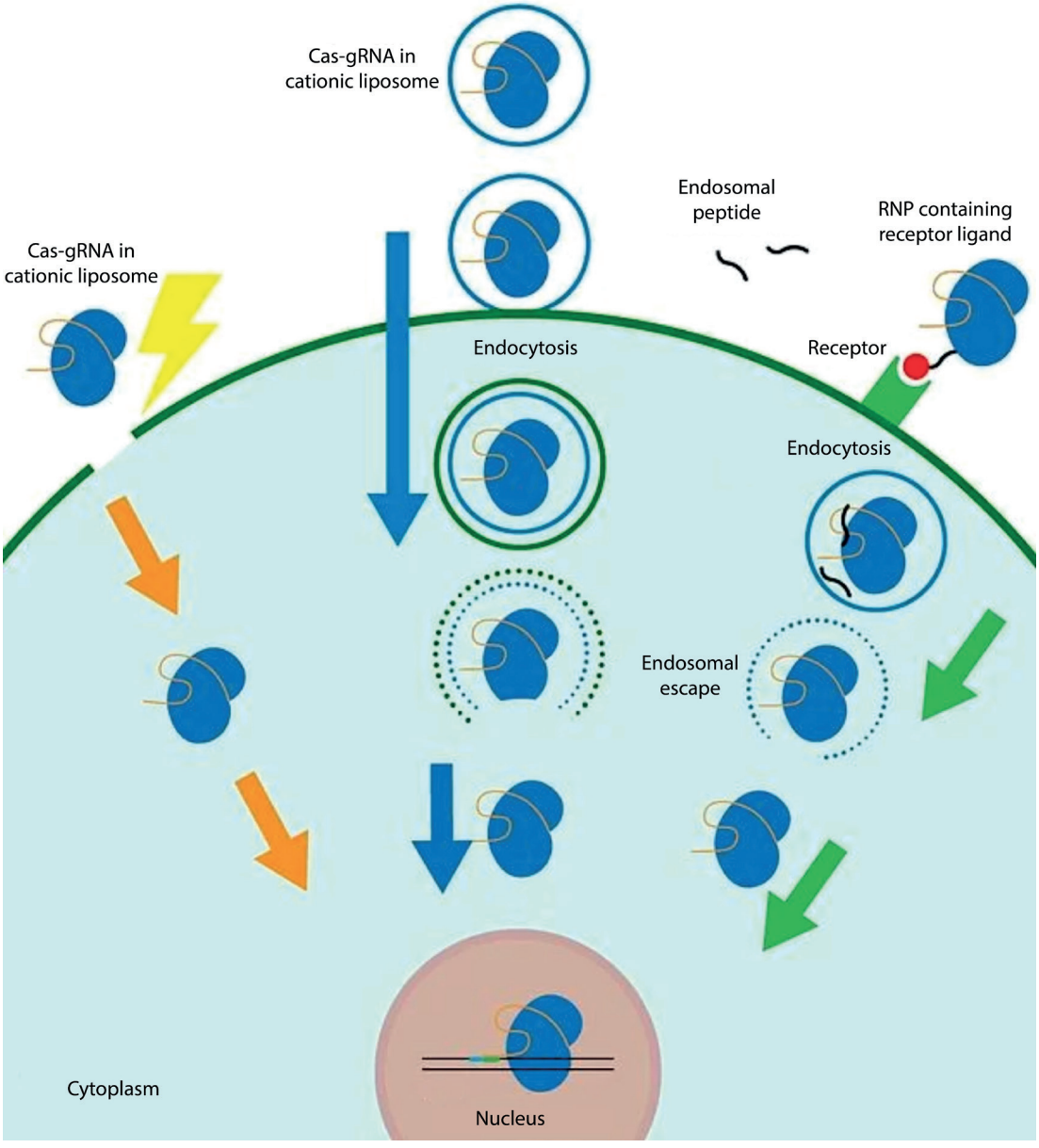


Figure 4. Delivery methods and internalization of RNPs. (A) Simple transfection - exemplified by electroporation. (B) Nanostructured systems - exemplified by the encapsulation of RNP in cationic liposomes. (C) Ligand-receptor mechanism - exemplified by RNP fused to a peptide and endosomal ligand. All methods culminate in editing the nuclear genome, with or without endocytosis as an intermediate step.

Source: adapted from Hampstead (2020).

in the form of nanostructured systems or nanomaterials for the delivery of CRISPR/Cas, which can perfectly be used shortly for editing in plants via RNPs, such as:

a) Cationic liposomes: spherical lipid bilayers with positive charges, enabling the binding to negatively charged nucleases, which, in turn, are complexed with sgRNA. They can be composed of DOTAP - N-[1-(2,3-dioleoyloxy)propyl]-N,N,N-trimethylammonium methylsulfate, DOTMA - N-[1-(2,3-dioleyloxy)propyl]-N,N,N-trimethylammonium chloride, DOSPA - 2,3-dioleyloxy-N-[2-(sperminecarboxamido)ethyl]-N,N-dimethyl -1-propanaminium trifluoroacetate and DMRIE - 1,2-dimyristyloxypropyl-3-dimethylhydroxyethylammonium bromide.

b) Solid lipid nanoparticles (SLNs): nanoparticles consisting of lipids with a high melting point, especially fatty acids (e.g. stearic acid and palmitic acid), with a solid core and covered by surfactant molecules.

c) Polymeric nanoparticles: polymers in the form of nanometric particles, which may be composed of PEI (polyethyleneimine), PLL (poly-L-lysine), and PDMAEMA - poly [2- (dimethylamino) ethyl methacrylate, for example.

d) Chitosan nanoparticles: nanoparticles composed of a polysaccharide derived from chitin deacetylation and with cationic properties.

e) Dendrimers: highly positively charged polymers, functioning as "proton sponges", with a high affinity for negatively charged molecules. Poly(amidoamine) (PAMAM) is one of the main examples.

f) Nanoparticles of silica (Torney et al., 2007).

g) Magnetic nanoparticles (Zhao et al., 2017).

h) Carbon nanotubes (Demirer et al., 2019).

Although the use of RNPs for genome editing in plants is sporadic, their reports in the literature mostly involve the use of simple transfection. Only recently, Liu et al. (2020) conducted the transfection of tobacco protoplasts (*Nicotiana tabacum* cv. Bright Yellow-2) based on nanostructured systems for lipofection, using two lipid reagents - Lipofectamine 3000 and RNAiMAX- for the assembly of cationic liposomes, obtaining delivery efficiency of 66% and 48% for these reagents, respectively. For Lipofectamine 3000, the editing efficiency was 6%. Although lipofection is well established for genome editing in animal cells, only in this study a transition from nanomaterials to CRISPR/Cas was established in plant systems.

3) **Ligand-receptor mechanism**: cover the logic that the nuclease responsible for the editing process can be fused to peptide or protein ligands of specific cell receptors (Rouet et al., 2018). Although there are reports of the use of this approach for editing in animal cells, its applicability to CRISPR in plants has not yet been demonstrated, and neither do authors in the scientific literature have perspectives in this regard.

Excluding particle bombardment, for all methods and techniques used for delivery, it is important to emphasize the need to establish protocols that facilitate the use of these systems, mainly for in vitro regeneration of plants from protoplasts, since they are explants most used to conduct these procedures. However, knowing that some species are recalcitrant to regeneration from protoplasts, the development of such protocols would be unfeasible, and other strategies based on alternative types of explants should be chosen. Anyhow, for those species for which this procedure is possible, this benefit arises when working with RNPs.

Experimental procedure

Part 1: in silico

The first step to perform genome editing via CRISPR/Cas is to choose genomic targets of the desired species to be edited. This choice will serve as a basis to subsequently identify, within the respective genes, sites with sequences complementary to the gRNA used to guide the nuclease, since they meet the requirements of the respective PAM sequence required by the enzyme for cleavage. Also, it is essential to perform the prediction of potential off-target sites, to propose safer candidate gRNAs regarding genomic specificity. Finally, choosing a suitable delivery strategy, in particular, depending on the type of explant required for transfection, as well as an appropriate expression vector with appropriate configuration of its genetic elements, is the final and fundamental step - although they vary according to the objectives and the type of each study - a set of procedures that must be conducted in a complementary way to the *in-silico* analysis, as will be detailed below.

Identification and selection of genomic targets

Aiming at knockout in coding sequences (CDS) or transcriptional regulation of promoters (generating variants of promoters), both via NHEJ, the target gene must

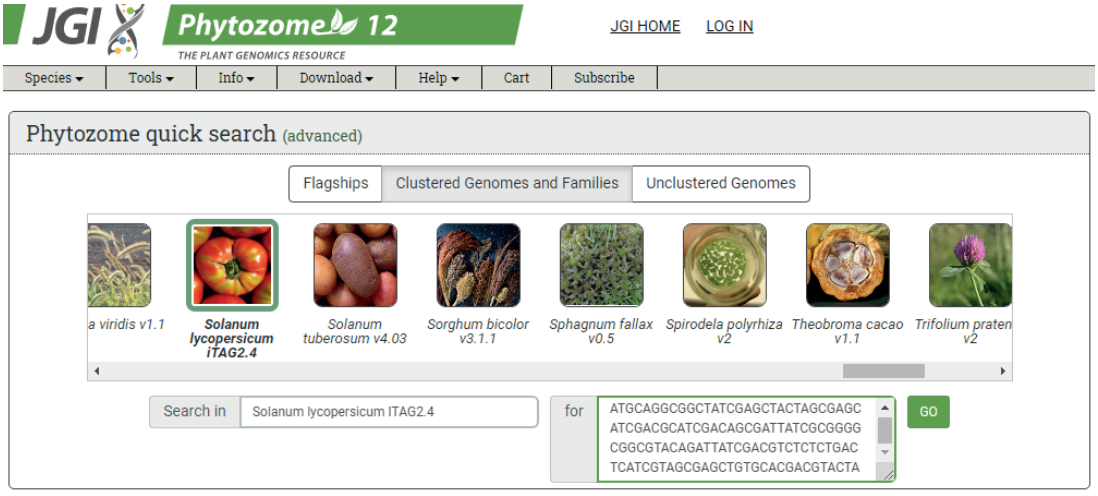


Figure 5. Phytozome v12.1. Search for sequences across the genome of different possible plant species. A search bar for inserting the nucleotide sequence of the selected species, on the right, is shown, enabling BLAST-mediated search, using the "GO" function.

Figure 6. Search for mature transcript sequences in Phytozome v12.1. Differentiation stage of the genetic elements that compose the nucleotide sequence of the gene after transcription, which will serve as a model for similarity in the analysis of candidate gRNAs. In green, the sequence related to the 5'-UTR region (useful for the selection of gRNAs for the transcriptional modulation of promoters). In blue, sequences related to neighboring exons (useful for the selection of gRNAs aiming at the knockout of the gene), separated by light and dark tones.

be characterized in terms of the nucleotide sequences that compose it, including its regulatory elements. In this case, it is possible to obtain the nucleotide sequences in the annotated database of the National Center for Biotechnology Information (NCBI)[1], using terms that identify the species and/or the sequence, as well as accession numbers, when possible. Another possibility is to obtain the complete gene sequence from smaller sequences (query), through similarity with sequences deposited (subject) in the database. For that, this information can be obtained using the different modalities of the Basic Local Alignment Search Tool (BLAST)[2] for different types of sequences.

However, the BLAST resource can be accessed indirectly and in a personalized way for many of the main plant species of economic importance, as, for example, in the Phytozome v12.1[3] database (Figure 5).

With the identification of the target sequence (Figure 6), the nucleotide composition can be detailed in the "Sequence" tab. The subsection "Transcript Sequence" provides the components of the mature transcript differentiated by colors between 5'-UTR, CDS, and 3'-UTR, which can be useful for the design of gRNAs for both approaches (CDS and regulatory elements). As in the NCBI, the sequence can be obtained in FASTA format and stored for gRNA analysis.

At the same time, it is recommended that in the case of CDS as a target, the sequence of the polypeptide chain should also be accessed and stored in any text editing tool, for another type of analysis, as will be shown below. In this case, in Phytozome it is possible to access the respective sequence in the subsection "Peptide Sequence", within the same tab "Sequence", while in NCBI the respective sequences are shown on the page of the characteristics table of each access, in the section "CDS".

gRNA design

Having the target site delimited in silico and containing its genetic elements, one should, depending on the strategy, design one or more gRNAs complementary to a regulatory element or the CDS, taking into account the target sequence at the genomic level. To promote the knockout of a gene, the functionality of the expression product must be eliminated by truncating the encoded polypeptide chain. Therefore, one must start from two fundamental premises:

[1] Available at: https://www.ncbi.nlm.nih.gov/nuccore

[2] Available at: https://phytozome.jgi.doe.gov/pz/portal.html

[3] Available at: https://phytozome.jgi.doe.gov/pz/portal.html

a) The knockout should occur based on mutagenesis in more distal 5' regions, that is: the more upstream the CDS cleavage occurs, the less likely that native and functional protein domains will remain expressed based on the remaining sequence in the frame.

b) The pairing of the gRNA with the genome must occur in regions restricted to specific and unique exons, that is: the PAM sequence and the cleavage site must be located in the same exon. Therefore, locating these sequences in introns or exon-intron junctions can make the candidate sequence irrelevant to effective genome editing with concrete results.

Considering these factors to start the second stage, a very practical, intuitive, and broad tool for the search and prediction of candidate gRNAs is CRISPOR[4] (Concordet; Haeussler, 2018), which will be detailed below, instructing the reader to select the best guides for the aforementioned purposes:

1) When accessing CRISPOR (Figure 7), the target sequence must be inserted in the nucleotide sequence bar in "Step 1". If you prefer, you can assign a name to the sequence in that same section.

   Note: the sequence to be searched can be genomic or cDNA. In both cases, one must pay attention to the second premise exposed above, discarding candidate gRNAs that match introns or that cover more than 1 exon, respectively.

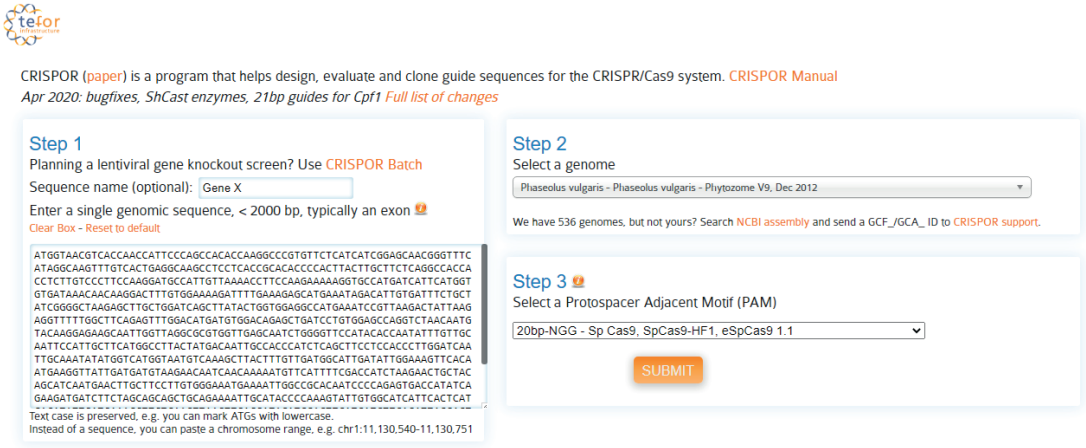


Figure 7. Search for candidate gRNAs at CRISPOR. Input window of the target sequence for the design of the candidate gRNAs, detailing the bars of insertion of the sequence (Step 1), selection of the target genomes options (Step 2), and selection of the PAM sequence (Step 3).

[4] Available at: http://crispor.tefor.net/

2) Then, in the "Step 2" section, from the 536 available genomes, select the target host genome to be edited. If the genome of the species is not available, it is possible to disable the selection of a specific genome by selecting the option "No genome". In this case, the output will have less data richness, without specific details.

3) Finally, in the "Step 3" section, select the PAM corresponding to the PAM sequence of the nuclease with which you will work on the bench and whose abbreviation is shown next to each generic sequence. The PAM selection already includes the standard length of the gRNA sequences to be returned as an output, which varies according to the nuclease used in the editing. After that, click on "Submit" to start the search.

4) Right after, new information will be shown on a new page (Figure 8). First, the target sequence will be made available in a gray box, with residues identified every 10 nucleotides. Under the residues, the initial gRNA sequences overlaid on the target will appear. Depending on the specificity of each gRNA, they will appear in three different colors: green (sequences with high specificity), yellow (medium specificity), or red (low specificity). It is strongly recommended that only green colored sequences be selected. The sequences can be aligned to the sequence Forward (starting with nucleotides and ending with dashes), or Reverse (starting with dashes and ending with nucleotides) to the target sequence of the input.

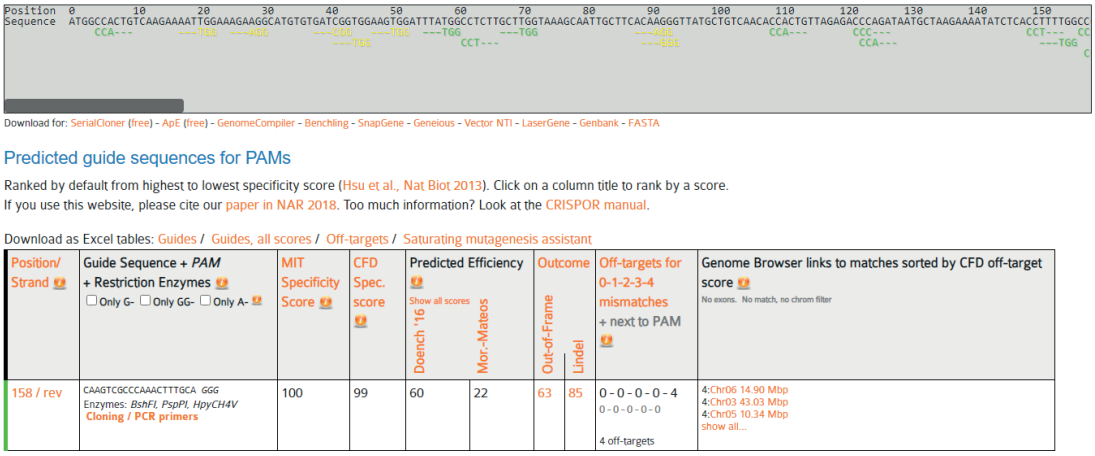


Figure 8. Result of the prediction of candidate gRNAs in CRISPOR. Output window of candidate gRNA sequences, showing the overlap of each in the target sequence provided, its nucleotide composition, as well as the location on the respective strand and on-target and off-target parameters, detailing each of the cases of potential offs -targets.

Note 1: when working with nucleases that cleave the target sequence in both strands, it is irrelevant to select Forward or Reverse candidates.

Note 2: it is possible to click on each sequence of the gray box to be directed to the respective information in the table that contains all the gRNAs.

5) The information table, presented in the Predicted guide sequences for PAMs section, contains the main data on all candidate gRNAs found according to the established parameters. This table can be exported and saved as a file with .xls extension, by clicking on the different options just above it (Download as Excel tables). As for the table parameters, the listed gRNAs are arranged in decreasing order of specificity (on-target, MIT Specificity Score, which ranges from 0 to 100), which is inversely proportional to the probability of occurrence of off-target effects using the respective guide. In the first column of the table (Position/Strand), the position and the strand of the target sequence that contains complementarity to the gRNA are informed. In the second column (Guide sequence + PAM + restriction enzymes), the sequences of each guide, containing the PAM, are shown, as well as the restriction sites present and capable of cleaving each guide. This column also indicates, when appropriate, an observation that the guide may not be adequate because it contains termination signs in the form of sequential thymine residues (T_n). It is also possible to restrict guides initiated with specific nucleotide residues (this restriction is generally used when working with U6 and U3 promoters to direct the expression of sgRNA, although even when selecting such promoters to direct expression, these residues can be added artificially to the 5'end of any sgRNA to make them compatible with the respective promoter). The CFD and Specificity Score columns are related to the specificity and cleavage efficiency scores, but less useful for choosing gRNAs. The Outcome column represents parameters related to the result of the cleavage, such as out-of-frame and frameshifts, caused by deletions and indels, respectively. In this case, the values also vary from 0 to 100, with the number representing the theoretical percentage of events/clones having these editing results. The column Off-targets for 0-1-2-3-4 mismatches indicates the number of cases of non-target sequences in the host genome that contain 0 to 4 mismatches, in that order, serving as a crucial parameter to define the best guides that represent less risk of generating off-target effects. The smaller the number of cases, especially if there are few or no cases with the least number of mismatches, the more specific the guide will be. The selected guides should have no cases with less than 3 mismatches, especially in the seed region of the gRNA. The last column details the genomic position of each of the off-target cases, detailing the pairing when hovering the mouse over each case and marking the mismatches

between the residues, which makes it easier to verify if the occurrence of the mismatches is within or outside the seed.

In addition to CRISPOR, for some plant species - especially those of great economic importance, such as soy, rice, and maize, it is possible to design gRNAs with other computational tools, such as CRISPR-P v2.0[5] (Lei et al., 2014). In this case, just access the main page of the program and click on the Submit tab. In the new window (Figure 9), you must select the target species, the PAM sequence, the U6 or U3 promoter and the length of the desired gRNA sequence to be returned (it can vary from 15 to 22 residues according to the program default), also inserting the nucleotide sequence to be edited. The sgRNA scaffold sequence is already pre-established. After that, click on Submit to search for gRNAs.

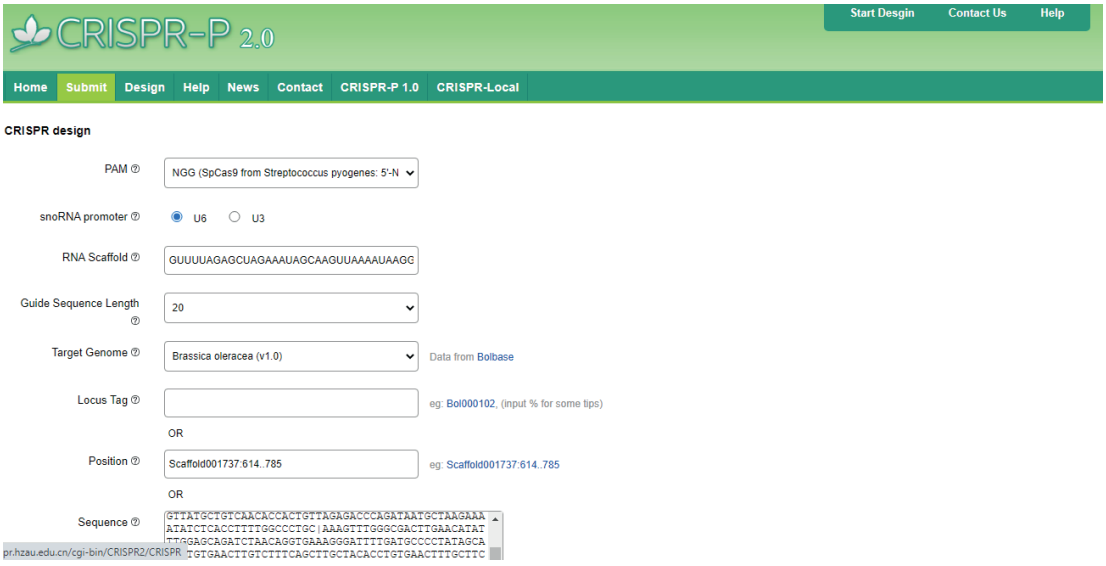


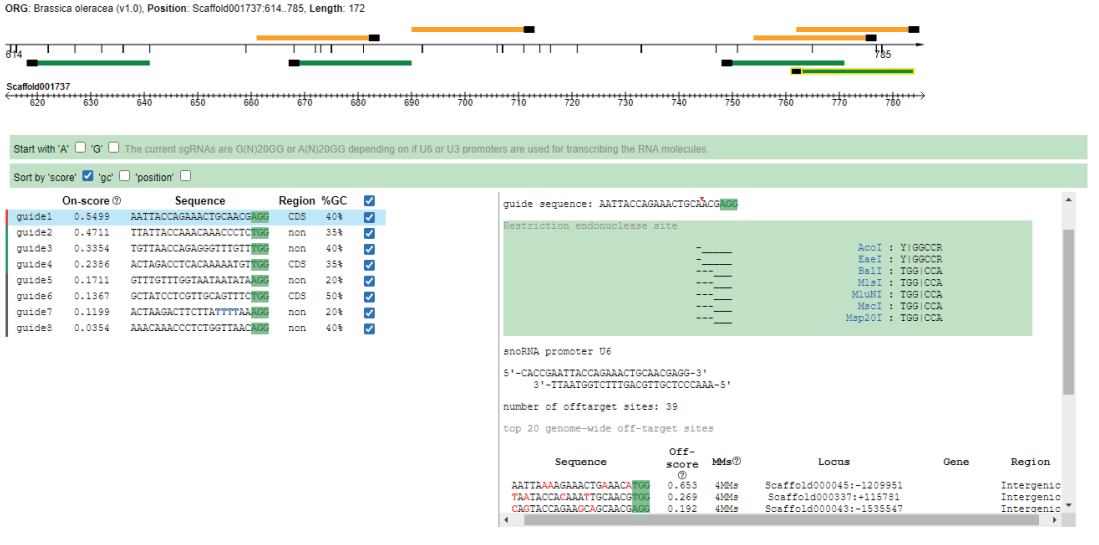
Figure 9. Search for candidate gRNAs in CRISPR-P v2.0. Input window of the target sequence for the design of candidate gRNAs, in the Submit tab of the page, detailing the selection bars of the PAM sequence, the length of the guide, the target genome (species), and the insertion of the nucleotide target sequence.

After submission, a new page (Figure 10) will open with details of the candidate guides for selection. On this page, the gRNAs will be represented graphically in line with the sequence in the upper panel, with the Forward and Reverse sequences in different colors. In the lower panel, the candidate sequences are listed, with several

[5] Available at: http://crispr.hzau.edu.cn/cgi-bin/CRISPR2/SCORE

parameters informed on the left (% GC, on-target score, nucleotide composition, the presence or not in the coding region). On the right, the number of off-target cases is reported, with details of each of them regarding the nucleotide composition (with mismatches highlighted in red), the number of mismatches, chromosomal location, and the presence or not in a coding region (each case is classified as intergenic or CDS and, in the latter case, the corresponding gene ID is indicated in the column on the side). The restriction sites present in the sequence of each predicted gRNA are also informed, which can be useful in the stage of subcloning the guides in the expression vector, for example.

Additionally, it is possible to design more than one gRNA for the same target gene, which ensures a greater probability of occurrence of the knockout of that gene, in addition to being able to excise fragments of great length due to the production of at least two DSB sites. Going further, multiple gRNAs can be designed and used, usually cloned into the same expression vector, and not necessarily directed to the same gene, but to different target genes (Cong et al., 2013), in a strategy called "multiplex genome engineering". Multiplex engineering, therefore, allows modulating the expression of several genes associated with highly divergent or related characters converging to a common character.



	On-score	Sequence	Region	%GC
guide1	0.5499	AATTACCAGAAACTGCAACGAGG	CDS	40%
guide2	0.4711	TTATTACCAAACAAACCCTCTGG	non	35%
guide3	0.3354	TGTTAACCAGAGGGTTTGTTTGG	non	40%
guide4	0.2386	ACTAGACCTCACAAAAATGTTGG	CDS	35%
guide5	0.1711	GTTTGTTTGGTAATAATATAAGG	non	20%
guide6	0.1367	GCTATCCTCGTTGCAGTTTCTGG	CDS	50%
guide7	0.1199	ACTAAGACTTCTTATTTTAAAGG	non	20%
guide8	0.0354	AAACAAACCCTCTGGTTAACAGG	non	40%

Sequence	Off-score	MMs	Locus	Gene	Region
AATTAAAGAAACTGAAACATGG	0.653	4MMs	Scaffold000045:-1209951		Intergenic
TAATACCACAAATTGCAACGTGG	0.269	4MMs	Scaffold000337:+115781		Intergenic
CAGTACCAGAAGCAGCAACGAGG	0.192	4MMs	Scaffold000043:-1535547		Intergenic

Figure 10. Result of the prediction of candidate gRNAs in CRISPR-P v2.0. Output window of candidate gRNA sequences, showing the overlap of each one in the target sequence provided, its nucleotide composition, as well as present restriction sites and on-target and off-target parameters, detailing each of the cases of potential off-targets.

In order to conduct the multiplex approach, it is common for the designed gRNAs to be inserted into the expression vector so that they are arranged on polycistronic cassettes. For this purpose, in addition to the classic configuration of different gRNAs arranged in different expression cassettes (that is, in a monocistronic form), with expression directed by the respective promoters, the gRNA units can be interspersed by sequences of transfer RNA (tRNA), in a strategy based on the endogenous processing of tRNA molecules by RNases for excision of transcripts *in vivo* conditions, promoting the release of fused gRNAs and thus allowing them, once free in the nucleus, to be able to associate with the nuclease to guide them to different target sites in genomic DNA, also increasing the efficiency of genome editing (Xie et al., 2015; Wolabu et al., 2020). However, the tRNA-gRNA-based processing system is based on the use of promoters for RNA Polymerase III to direct the expression of these synthetic genes (He et al., 2017). The use of these promoters, however, like the U6 and U3 variants, has some limitations (Zhang et al., 2017): they are constitutively active and, therefore, lacking cellular or tissue specificity; they are poorly characterized for some organisms and are not suitable for use in in vitro transcription due to the lack of commercial versions of RNA Polymerase III.

To circumvent these obstacles, an alternative way to conduct multiplex engineering is to use promoters for RNA Polymerase II. These promoters, in principle, also have some limitations in their use for genome engineering (Zhang et al., 2017), such as the transcripts they direct in the expression undergo post-transcriptional processing (capping, splicing, and polyadenylation), modifications in the transcript that can functionally inactivate the gRNA; the mature RNA is transported to the cytosol, being confined in a distinct cell compartment from where it must act together with the Cas nuclease, which constitutes a physical barrier that blocks the editing activity. However, when these promoters are associated with a second processing strategy, they can become sufficient to promote editing. This strategy, called "RGR" (Ribozyme-gRNA-Ribozyme), has as a principle the use of gRNAs flanked by ribozymes at both 5′ and 3′ ends of the guides (Lee et al., 2016), being generally hammerhead ribozymes (HH) and Hepatitis D Virus (HDV).

Based on this, the principle of this mechanism is similar to the principle of the tRNA-gRNA strategy, although it is based on the autocatalytic activity of ribozymes from the moment they are transcribed, which also promotes the release of gRNAs in the cell nucleus, bypassing the processes that a mature RNA would receive, as mentioned above. Thus, it becomes possible to produce gRNAs from any promoters in any organism, making the choices of promoters to conduct genome editing virtually unlimited (He et al., 2017). Although this strategy can also be used for a single gRNA with a promoter-driven expression for RNA Polymerase II, using monocistronic cassettes similar to those mentioned on promoters for RNA

Polymerase III, it is quite convenient for the multiplex approach, as mentioned earlier, in a polycistronic configuration.

Identification of possible off-targets

Although both tools presented in the previous topic are capable of predicting potential off-target sites for each designed gRNA, there are specialized tools in this task and with greater accuracy. Among them, the Cas-OFFinder[6] tool (Bae et al., 2014) is quite complete, intuitive, with a simple and more widely used layout, and is therefore recommended by the authors for this procedure, which is detailed below:

1) Accessing Cas-OFFinder (Figure 11), the gRNA sequences selected in the previous step must be input simultaneously. In this window, the tool is subdivided into 3 sections: (a) PAM Type: first, the PAM sequence corresponding to the nuclease to be worked with must be selected; (b) Target Genome: here the group of organisms within which the species containing the target genome is located must be selected, being the groups: Vertebrates, Insects, Plants, Others and Non-reference Vertebrates, and then the species with its reference genome; (c) Query Sequences: in this section, the residue sequences of the gRNAs are inserted one below the other (in separate lines), without the PAM sequence, and a maximum limit of mismatches can be established, giving greater objectivity to the analysis. In the end, click Submit to start the search.

2) Then, new information will be shown on a new page (Figure 12). It presents two sections: (a) Summary: presents, for each sequence of gRNA, a line with the number of on-target cases and another line with the number of off-target cases, according to the number of mismatches; similar to what is shown in CRISPOR; (b) Details: for both on-target and off-target cases, the pairing between the gRNA and the genomic target is shown, as well as the indication on which chromosome the target is located, in addition to its starting position showing the nucleotide residue and sequence orientation on the strand (sense +, or anti-sense -). With this information, it is possible to return to Phytozome and locate the sequence of the off-target site, characterizing it in terms of expression. In this case, it is recommended to check first if the off-target sites are in intergenic or transcribed sequences. Being in transcribed sequences, it is important to determine the region of the sequence, evaluating its importance in the performance of the final product, so that, once the knockout is performed, the result does not generate harmful effects to the organism

[6] Available at: http://www.rgenome.net/cas-offinder/

or interfere with the analysis. For example cleavages in 5'-UTR regions and exons are often critical; therefore, a gRNA designed with a potential off-target in these cleavages should not be prioritized, unlike others located in intronic regions (except in consensus regions that are preponderant for splicing).

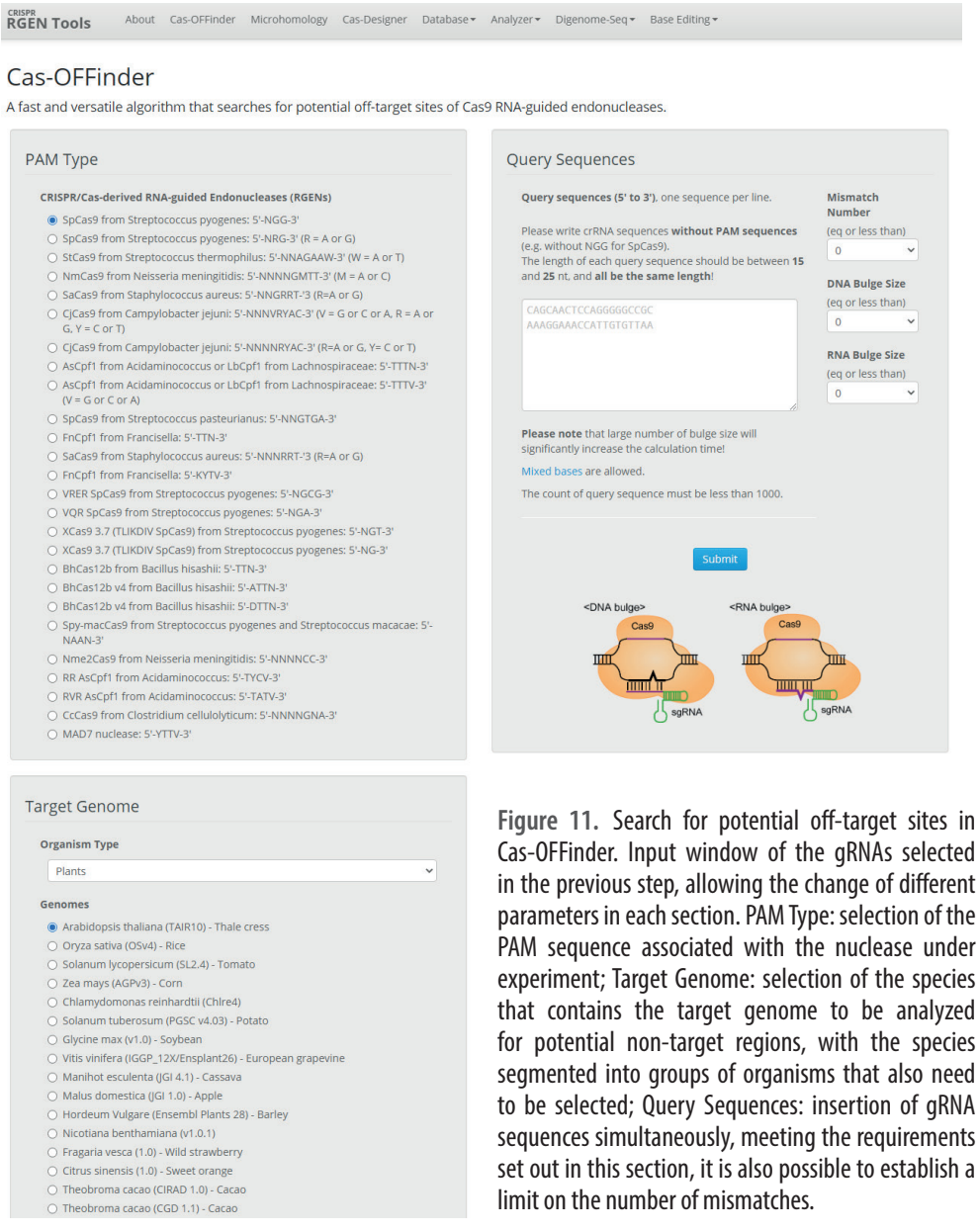


Figure 11. Search for potential off-target sites in Cas-OFFinder. Input window of the gRNAs selected in the previous step, allowing the change of different parameters in each section. PAM Type: selection of the PAM sequence associated with the nuclease under experiment; Target Genome: selection of the species that contains the target genome to be analyzed for potential non-target regions, with the species segmented into groups of organisms that also need to be selected; Query Sequences: insertion of gRNA sequences simultaneously, meeting the requirements set out in this section, it is also possible to establish a limit on the number of mismatches.

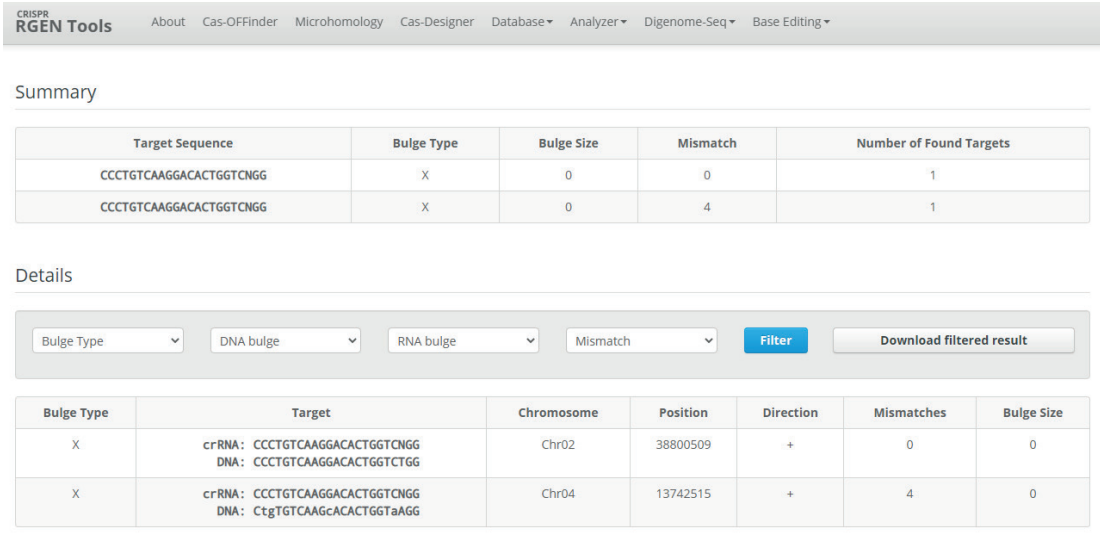


Figure 12. Result of the prediction of potential off-target sites in Cas-OFFinder. Output window of the gRNAs selected in the previous step, with the analysis of the cases of potential off-target sites found in the reference genome. The Summary section summarizes, for the same gRNA, on-target cases on one line (shown by 0 mismatches, being the target sequence itself in the genome) and off-target on another line (shown by 1 case with 4 mismatches, in this example). In the Details section, both cases are detailed regarding the pairing between the gRNA and the similar genomic sequence, but informing the chromosome location, in the Chromosome and Position columns.

Knockout simulation and homology prediction

Once verified and filtered only gRNAs with good specificity and without critical off-target activity, when there are potential off-target sites, it is important to simulate the knockout resulting from the editing, so that homology can be predicted with functional protein domains from the remaining polypeptide chain since it is expected to be truncated and its biological activity to be abolished. It is important to note that this type of analysis should be performed only when the objective is the knockout by editing, via indels, the coding sequences of target proteins. For this, the appropriate procedure must be performed as follows:

1) An in-silico translation should be conducted from the nucleotide residue where theoretically cleavage by the nuclease will occur so that a truncated polypeptide chain will be generated for later verification of homology. It is possible to conduct it in the online tool ExPASy Translate Tool[7]. From the cleavage point, select the upstream sequence and insert it into the nucleotide

[7] Available at: https://web.expasy.org/translate/

sequence bar (Figure 13). It is recommended, as an output format, to keep the Compact option selected, as well as the results shown on both tapes, enabling both options, without the need to select the specific genetic code. Finally, click on Translate!. It is important to remember that not every indel will cause an earlier stop codon to form along the target sequence or, more broadly, a frameshift that alters the polypeptide chain 3' from the cleavage point, modifying its secondary and/or tertiary three-dimensional structure and, thus, making native biological activity unfeasible. This is because insertions or deletions of nucleotide residues may occur in multiple numbers of 3, which keeps the rest of the sequence in the frame, not necessarily causing this deleterious effect due to the presence or absence of a few amino acid residues. However, due to the unpredictability of the type of indel to occur in each event, one must work on assuming the occurrence of non-multiple of 3 indels, enabling simulation.

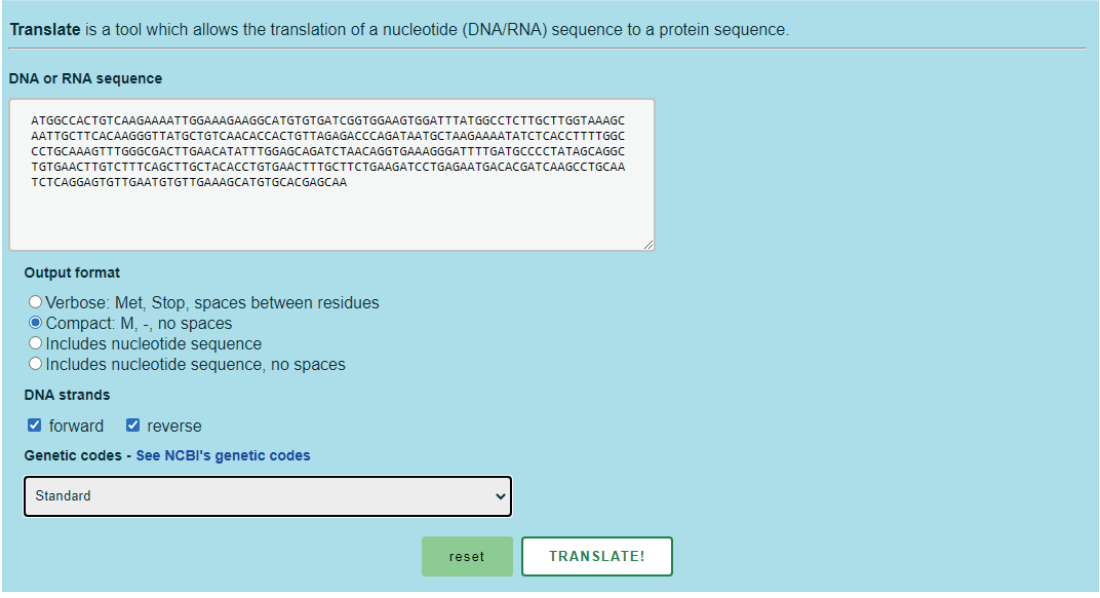


Figure 13. In silico translation of truncated sequences in the ExPASy Translate Tool. Input window of the simulated nucleotide sequence regarding truncation, showing different possible output formats and the possibility of representing 3 frames (Forward or Reverse only) or 6 frames (Forward and Reverse).

2) With the result of the analysis, a new page is opened (Figure 14), with the results of the *in-silico* translation, in all 6 possible ORFs (Open Reading Frames). Since the input sequence is usually inserted from the first residue of the start codon (ATG), all results in this way will be found in Frame 1 (5'-3'), with the truncated sequence marked continuously in red.

Figure 14. Result of in silico translation in ExPASy Translate Tool. Output window of the translated sequence, in 6 different ORFs (Open Reading Frames), with the sequence of the polypeptide chain, in this case, present in Frame 1 5'-3'.

3) Having the polypeptide sequence truncated, it is possible to copy and insert it in the online tool for predicting homology with functional protein domains, InterPro[8]. On this page (Figure 15), the sequence is entered in FASTA format, conducting the search and keeping the pre-established parameters. After that, a page with the analysis output (Figure 16) will be opened, informing about the presence or absence of protein families prediction referring to the remaining polypeptide sequence. If there is no prediction, it is likely that the resulting sequence does not form functional domains that perform a residual biological activity, which implies deleting the protein function associated with the gene subjected to the knockout.

[8] Available at: http://www.ebi.ac.uk/interpro/

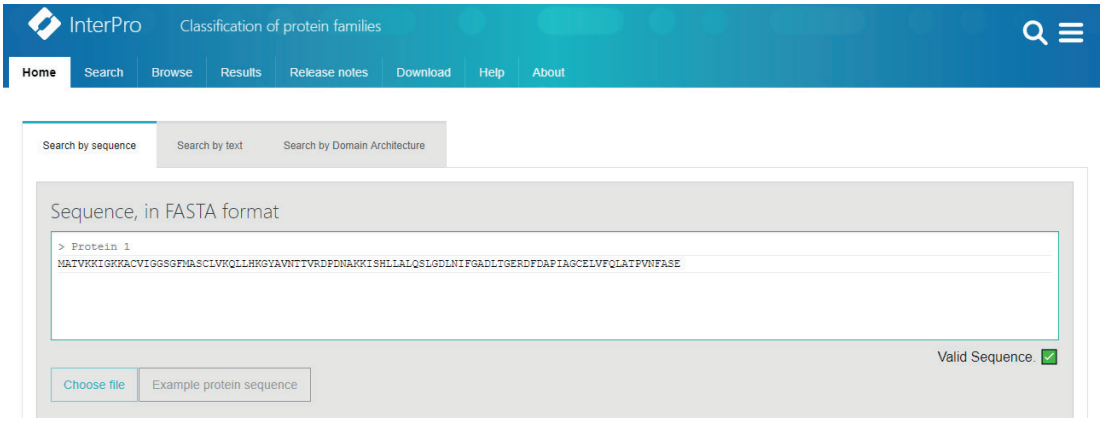


Figure 15. Search for protein homology in InterPro. Input window of the truncated polypeptide sequence to predict homology with functional protein domains by association with protein families.

Figure 16. Result of homology prediction with functional domains in InterPro. Output window of the truncated polypeptide sequence, showing the absence of homology with protein families, indicating that the remaining sequence has no known biological activity.

Once the absence of homology is confirmed, the gRNA meets all the requirements and properties of a useful candidate and is more likely to perform its function without the occurrence of undesirable effects and can be validated in vitro and

prepared for subcloning in expression vectors (approach: use of plasmids) or used alone (approach: IVT) or complexed to a nuclease (approach: RNP).

Part 2: in vitro

The second step to perform genome editing via CRISPR/Cas consists of in vitro validation of the isolated and mediating molecules in the editing process, which can be carried out using commercial kits. Because the genetic elements that make up an expression vector (in approaches to the use of plasmids) and that transfection methods vary greatly according to the target species and the explant used, the outline of these steps will not be detailed here, since this must be established in a particular way in each research project. However, the in vitro validation of the designed guides and associated nucleases is a more universal process for the different organisms to be edited.

Cleavage efficiency and validation of the nuclease-sgRNA complex

sgRNA must be able to guide the nuclease to the target site. Furthermore, the process will only be viable if, in addition to this capacity, the nuclease is efficient in promoting the cleavage of the sequence at the expected cleavage site. Thus, it is possible to verify the activity of both employing an in vitro cleavage efficiency test, in which the target sequence is amplified by Polymerase Chain Reaction (PCR) and the PCR product is cleaved by the complex. The formation of digestion fragments from the PCR product indicates that both the sgRNA and the nuclease perform their function normally, being able to proceed with genome editing in vivo. Otherwise, the absence of cleavage products indicates the lack of efficiency in the function of at least one of these components of the complex. For this, it is necessary to obtain the RNP corresponding to the complex that will be formed in vivo. In this case, obtaining the nuclease and sgRNA can occur as described in the topic RNPs: DNA-free systems, of this chapter. For works that use the RNP system for transfection, after validation, the complex can now be used directly to proceed with the editing. In the case of plasmid systems or transcripts, the corresponding nucleotide sequences of the validated molecules must be used.

A simple and succinct procedure (GenScript, 2018) for this validation, based on an in vitro test kit, can be followed as described below:

1) Perform the amplification of the target region by PCR (Note: it is possible to design pairs of primers for both the target region and for potential off-targets in CRISPOR, from the search output information table). The PCR product will

be used as a substrate for digestion with RNP nuclease. It is recommended that the product contains at least 200 bp flanking the sequence to which the gRNA is paired, both 5'and 3', being preferred that the PCR product has about 1 kb for better visualization of digestion.

2) Prepare 16 µL of RNP in a nuclease-free tube, according to the following reaction system, adapted from GenScript (2018):

Component	Volume/Concentration
sgRNA	3,6 pmol
Nuclease Cas (1 µg/µL)	1,5 pmol
Reaction buffer (Cas) (10X)	2 mL
H_2O nuclease-free	13,6 µL
Final volume	16 µL

3) Conduct the RNP assembly by incubating the reaction system tube at 37 °C for 10 min.

4) Add 450 ng of the PCR product to the RNP mix assembled in step 2, making up the volume with nuclease-free water to 20 µL and swirling slightly to promote mixing.

5) Incubate the product-RNP mixture for at least 30 min, at 37° C.

6) Check the occurrence or not of digestion through visualization based on agarose gel electrophoresis.

Part 3: in vivo

Finally, after the transfection and regeneration of candidate events for edited plants has been carried out, it is necessary to characterize the editings that occurred in vivo using different techniques, each presenting its advantages and disadvantages. Here, a step-by-step of each one will not be detailed, but the reader will be guided with literature about them, as well as with pertinent general information, to search on how to conduct them according to their preference, suitability to the design of their study, and available materials, reagents, and equipment in your laboratory.

Identification of edited plants

For the candidate events to be characterized in terms of altering the genomic sequence, it is of great interest to filter, when possible, plants that present phenotype, reducing the volume of evaluations, aiming at making the characterization more practical, simple, and less laborious.

When working with plasmid systems and genomic integration, it is possible, based on the expression vector used, and its genetic elements, to select explants that regenerate in a selective medium with resistance guaranteed by the marker with which they work, or that are visually observable as to the characteristic that the inserted reporter gene confers. However, when working with IVTs or RNPs, for example, selection based on these characteristics becomes impracticable, since it is generally not applicable. Thus, if it is not possible to conduct a screening based on a selective agent, or on the phenotype itself resulting from genome editing, the characterization must proceed for the selection of mutants. For this selection, the following methods are some of those that can be used in genotyping:

- Sequencing: very useful to detail the type and composition of the indel generated after cleavage and repair, informing on each of the strands the exact sequence of the nucleotide residues, both using the Sanger method, and NGS (Next-Generation Sequencing) approaches. In general, sequencing is usually a central or complementary approach to other less informative methods at the level of the genomic sequence, being used, when applied as a complementary method, to characterize only previously genotyped events.
- PCR-RE (PCR-restriction enzyme): based on an amplification step of the regions to be evaluated, which are usually the target sequences, and the regions of possible off-target. Then, a restriction enzyme digestion with a recognition site is carried out encompassing the cleavage site by the editing's nuclease. Thus, the digestion of the PCR product reveals that the restriction site remains intact and, therefore, there was no editing; the absence of digestion reveals that the restriction site was eliminated due to the indel resulting from genome editing. Although frequently used, the limitation of this technique lies in the need for the existence of such restriction sites, which is not always the case, given the reduced diversity of candidate gRNAs that are appropriate. For more information, read the work of Xie and Yang (2013).
- T7E1 / Surveyor ™ assay: based on the ability to recognize and cleave heteroduplex DNA molecules (derived from mismatches between strands) by T7 endonuclease and Surveyor enzymes. These enzymes recognize DNA molecules whose strands are mismatched. For this reason, they are useful in

detecting heteroduplexes, which correspond to heterozygous individuals for editing. These are simple and inexpensive techniques when compared to sequencing, but they do not reveal the nucleotide sequence of the regions evaluated. They also lack robust reproducibility and are susceptible to generating false-positive results when polymorphisms or random mutations occur in the evaluated sequences, although their use is commonly reported in the literature. Also, only heterozygous individuals for the editings can be detected. For more information, read the work by Vouillot et al. (2015).

- qPCR (PCR quantitative): there are countless variations of qPCR for genotyping event indels, both at the genomic level and at the transcriptomic level (via RT-qPCR), whose primers designed overlap with the target site to be edited, so that the basis of this technique lies in the fact that the occurrence or not of mutagenesis is indicated by the failure or success, respectively, of amplification. This is because editing tends to cancel the primer ringing in the edited region, interfering with the generation of the amplicon. However, smaller indels tend to interfere to a lesser extent in amplification. For this, the HRM (High-Resolution Melting) technique can be an alternative to identify smaller indels, as it consists of changing the fluorescence pattern in the analysis of the melting curve, which is generated during the formation of the DNA duplex. Thus, based on the profile of the curve, it is possible to clearly distinguish a wild individual (wild-type, WT) from a mutant. In general, qPCR/RT-qPCR is an accurate technique, but with a slightly higher relative cost than some other genotyping techniques. Thus, we recommend that the reader understand the details of the main methodological aspects from the work of Yu et al. (2014), Falabella et al. (2017), and Li et al. (2018).
- Western blot: a technique that consists of labeling a target protein using an antibody conjugated to a molecule that produces a visible signal after an enzymatic reaction, the antibody being directed to recognize a specific region of the target protein. The protein is detected on a membrane after transfer from an SDS-PAGE (polyacrylamide gel electrophoresis with sodium dodecyl sulfate). If the protein is present in its complete form, the region recognized by the antibody remains present, resulting in a colorimetric signal due to the catalyzed reaction; in the absence of the complete form, no signal is theoretically detected. It is a less used method for the characterization since, in addition to being more laborious and requiring more work time, it is possible to perform the characterization at the genomic or transcriptomic level, without the need for evaluation at the proteomic level. Besides, there is the possibility of generating false positives, since the antibody can cross-react with proteins structurally similar to the target protein, even though it

is already truncated due to editing. Normally, the detection of Cas protein is conducted, checking for events that express the nuclease and, therefore, are at first undergoing an editing process (Endo et al., 2019).

- Polyacrylamide gel (PAGE): the polyacrylamide gel forms a mesh with properties that generate a high-resolution capacity of DNA molecules, which allows detecting differences of up to about 2 bp between the strands and between different samples. Thus, it is a very accurate alternative for detecting all length ranges of the indels, despite being a little laborious, and not very fast. For more information, see the work of Zhu et al. (2014).
- Agarose gel (AGE) 4%-6%: although agarose gels, in their usual working concentrations (1% to 3%) are not useful for distinguishing indels by CRISPR/Cas, and therefore for genotyping events, Bhattacharya and Van Meir (2019) recently developed a simple method for genotyping using more concentrated gels, with 4% to 6% agarose. In this approach, indels of at least 3 bp can be detected visually based on the mobility property of heteroduplex molecules along with the gel mesh. Initially, heterozygotes can be differentiated from homozygotes after being subjected to 34 cycles of PCR. Subsequently, mutant homozygotes can be differentiated from wild homozygotes after an additional denaturation/renaturation cycle through hybridization with a wild-type control, which leads to the formation of homoduplexes in the case of wild homozygotes and heteroduplexes in the case of mutant homozygotes. The technique is quite advantageous in terms of practicality and simplicity compared to others such as qPCR, PAGE, Western blot, and T7E1/Surveyor tests, but so far it has been used only for animals, being promising to test its application in plants.

References

ALBERTS, B.; JOHNSON, A.; LEWIS, J.; RAFF, M.; ROBERTS, K.; WALTER, P. **Biologia molecular da célula**. 5. ed. Porto Alegre: Artmed, 2010. 1396 p.

AMIRKHANOV, R. N.; STEPANOV, G. A. Systems of delivery of CRISPR/Cas9 ribonucleoprotein complexes for genome editing. **Russian Journal of Bioorganic Chemistry,** v. 45, p. 431–437, Nov. 2019.

ANDERSSON, M.; TURESSON, H.; OLSSON, N.; FALT, A-S.; OHLSSON, P.; GONZALEZ, M. N.; SAMUELSSON, M.; HOFVANDER, P. Genome editing in potato via CRISPR-Cas9 ribonucleoprotein delivery. **Physiologia Plantarum,** v. 164, n. 4, p. 378–384, Dec. 2018.

ANDRES, S. N.; VERGNES, A.; RISTIC, D.; WYMAN, C.; MODESTI, M.; JUNOP, M. A human XRCC4-XLF complex bridges DNA. **Nucleic Acids Research,** v. 40, n. 4, p. 1868–1878, Feb. 2012.

BAE, S.; PARK, J.; KIM, J. S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. **Bioinformatics,** v. 30, n. 10, p. 1473-1475, May 2014.

BHATTACHARYA, D.; VAN MEIR, E. G. A simple genotyping method to detect small CRISPR-Cas9 induced indels by agarose gel electrophoresis. **Scientific Reports,** v. 9, article 4437, 2019.

CAI, Y. ; CHEN, L.; LIU, X.; SUN, S.; WU, C.; JIANG, B.; HAN, T. ; HOU, W. CRISPR/Cas9- mediated genome editing in soybean hairy roots. **PLoS One**, v. 10, n. 8, e0136064, Aug. 2015.

CARDONA, Y. T.; MORALES, M. A. M. Main repair mechanisms for damages in the DNA molecule. **Biosalud,** v. 13, n. 2, p. 95-110, jul./dic. 2014.

CARROLL, D. Genome engineering with zinc-finger nucleases. **Genetics,** v. 188, n. 4, p. 773-782, Aug. 2011.

CICCIA, A.; ELLEDGE, S. J. The DNA damage response: making it safe to play with knives. **Molecular Cell,** v. 40, n. 2, p. 179-204, Oct. 2010.

CONCORDET, J.; HAEUSSLER, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. **Nucleic Acids Research,** v. 46, n. 1, p. 242-245, July 2018.

CONG, L.; RAN, F. A.; COX, D.; LIN, S.; BARRETTO, R.; HABIB, N.; HSU, P. D.; WU, X.; JIANG, W.; MARRAFFINI, L. A.; ZHANG, F. Multiplex genome engineering using CRISPR/Cas systems. **Science,** v. 339, n. 6121, p. 819-823, Feb. 2013.

COOK, E. H.; SCHERER, S. W. Copy-number variations associated with neuropsychiatric conditions. **Nature,** v. 455, n. 7215, p. 919-923, Oct. 2008.

CORNU, T. I.; THIBODEAU-BEGANNY, S.; GUHL, E.; ALWIN, S.; EICHTINGER, M.; JOUNG, J. K.; CATHOMEN, T. DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. **Molecular Therapy,** v. 16, n. 2, p. 352-358, Feb. 2008.

CUI, Y.; DONG, H.; MA, Y.; ZHANG, D. Strategies for applying nonhomologous end-joining-mediated genome editing in Prokaryotes. **ACS Synthetic Biology,** v. 8, n. 10, p. 2194–2202, 2019.

CYRANOSKI, D. **CRISPR tweak may help gene-edited crops bypass biosafety regulation**. 2015. Available at: https://www.nature.com/news/crispr-tweak-may-help-gene-edited-crops-bypass-biosafety-regulation-1.18590. Accessed on: 02 jun. 2020.

DAVIS, A. J.; CHEN, B. P.; CHEN, D. J. DNA-PK: a dynamic enzyme in a versatile DSB repair pathway. **DNA Repair (Amst),** v. 17, p. 21–29, May 2014.

DEMIRER, G. S.; ZHANG, H.; GOH, N. S.; GONZÁLEZ-GRANDÍO, E.; LANDRY, M. P. Carbon nanotube-mediated DNA delivery without transgene integration in intact plants. **Nature Protocols,** v. 14, p. 2954-2971, Oct. 2019.

DERIANO, L.; ROTH, D. B. Modernizing the non-homologous end-joining repertoire: alternative and classical NHEJ share the stage. **Annual Review of Genetics**, v. 47, p. 433-455, 2013.

DOWNS, J. A.; JACKSON, S. P. A means to a DNA end: the many roles of Ku. **Nature Reviews Molecular and Cell Biology,** v. 5, p. 367-378, May 2004.

DUAN, Y. B.; LI, J.; QIN, R. Y.; XU, R. F.; LI, H.; YANG, Y. C.; MA, H.; LI, L.; WEI, P. C.; YANG, J. B. Identification of a regulatory element responsible for salt induction of rice OsRAV2 through *ex situ* and *in situ* promoter analysis. **Plant Molecular Biology,** v. 90, n. 1/2, p. 49-62, Jan. 2016.

ENDO, M.; MIKAMI, M.; ENDO, A.; KAYA, H.; ITOH, T.; NISHIMASU, H.; NUREKI, O.; TOKI, S. Genome editing in plants by engineered CRISPR–Cas9 recognizing NG PAM. **Nature Plants,** v. 5, p. 14–17, Jan. 2019.

FALABELLA, M.; SUN, L.; BARR, J.; PENA, A. Z.; KERSHAW, E. E.; GINGRAS, S.; GONCHAROVA, E. A.; KAUFMAN, B. A. Single-step qPCR and dPCR detection of diverse CRISPR-Cas9 gene editing events *in vivo*. **G3: Genes, Genomes, Genetics**, v. 7, n. 10, p. 3533-3542, Oct. 2017.

GAJ, T.; SIRK, S. J.; SHUI, S. L.; LIU, J. Genome-editing technologies: principles and applications. **Cold Spring Harb Perspectives in Biology**, v. 8, n. 12, p. a023754, Dec. 2016.

GENSCRIPT. **CRISPR**: Ribonucleoprotein (RNP) user manual. New Jersey, 2018. p. 7.

GIL-HUMANES, J.; WANG, Y.; LIANG, Z.; SHAN, Q.; OZUNA, C. V.; SÁNCHEZ-LEÓN, S.; BALTES, N. J.; STARKER, C.; BARRO, F.; GAO, C; VOYTAS, D. F. High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. **The Plant Journal,** v. 89, n. 6, p. 1251-1262, Mar. 2017.

GLASS, Z.; LEE, M.; LI, Y.; XU, Q. Engineering the delivery system for CRISPR-based genome editing. **Trends in Biotechnology,** v. 36, n. 2, p. 173-185, Feb. 2018.

GRIFFITHS, A. J. F.; MILLER, J. H.; SUZUKI, D. T.; LEWONTIN, R. C.; GELBART, W. M. **An introduction to genetic analysis**. 7th ed. New York: W. H. Freeman, 2000. 860 p.

HAMPSTEAD, A. **CRISPR 101**: Ribonucleoprotein (RNP) delivery. Addgene. Available at: https://blog.addgene.org/crispr-101-ribonucleoprotein-rnp-delivery. Accessed on: 02 jun. 2020.

HARTLERODE, A. J.; SCULLY, R. Mechanisms of double-strand break repair in somatic mammalian cells. **Biochemistry Journal,** v. 423, n. 2, p. 157–168, Sept. 2009.

HE, Y.; ZHANG, T.; YANG, N.; XU, M.; YAN, L.; WANG, L.; WANG, R.; ZHAO, Y. Self-cleaving ribozymes enable the production of guide RNAs from unlimited choices of promoters for CRISPR/Cas9 mediated genome editing. **Journal of Genetics and Genomics,** v. 44, n. 9, p. 469-472, Sept. 2017.

HER, J.; BUNTING, S. F. How cells ensure correct repair of DNA double-strand breaks. **Journal of Biological Chemistry,** v. 293, n. 27, p. 10502–10511, July 2018.

HOTTA, A.; YAMANAKA, S. From genomics to gene therapy: induced pluripotent stem cells meet genome editing. **Annual Review of Genetics,** v. 49, p. 47-70, 2015.

JIANG, W.; CROWE, J. L.; LIU, X.; NAKAJIMA, S.; WANG, Y.; LI, C.; LEE, B. J.; DUBOIS, R. L.; LIU, C.; YU, X.; LAN, L.; ZHA, S. Differential phosphorylation of DNA-PKcs regulates the interplay between end-processing and end-ligation during non-homologous end-joining. **Molecular Cell,** v. 58, n. 1, p. 172–185, Apr. 2015.

KANCHISWAMY, C. N. DNA-free genome editing methods for targeted crop improvement. **Plant Cell Reports,** v. 35, n. 7, p. 1469-1474, July 2016.

KREBS, J. E; GOLDSTEIN, E. S.; KILPATRICK, S. T; LEWIN, B. **Lewin's Genes X**. 10th ed. London: Jones & Bartlett, 2011. 930 p.

LAVIN, M. F.; KOZLOV, S. ATM activation and DNA damage response. **Cell Cycle,** v. 6, n. 8, p. 931–942, Apr. 2007.

LEE, R. T. H.; NG, A. S. M.; INGHAM, P. W. Ribozyme mediated gRNA generation for *In Vitro* and *In Vivo* CRISPR/Cas9 mutagenesis. **PLoS One,** v. 11, n. 11, e0166020, Nov. 2016.

LEI, Y.; LU, L.; LIU, H. Y.; LI, S.; XING, F.; CHEN, L. L. CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants. **Molecular Plant,** v. 7, n. 9, p. 1494-1496, Sept. 2014.

LI, L.; HE, Z.; WEI, X.; GAO, G.; WEI, Y. Challenges in CRISPR/Cas9 delivery: potential roles of nonviral vectors. **Human Gene Therapy,** v. 26, n. 7, p. 452-462, July 2015.

LI, Q.; SAPKOTA, M.; VAN DER KNAAP, E. Perspectives of CRISPR/Cas-mediated *cis*-engineering in horticulture: unlocking the neglected potential for crop improvement. **Horticulture Research,** v. 7, n. 36, Mar. 2020.

LI, S.; LIU, S.; LIU, Y.; LU, H.; TAN, Y.; HUANG, J.; WEI, P.; SHU, Q. Y. HRM-facilitated rapid identification and genotyping of mutations induced by CRISPR/Cas9 mutagenesis in rice. **Crop Breeding and Applied Biotechnology**, v. 18, n. 2, p. 184-191, Apr./June 2018.

LIANG, Z.; CHEN, K.; LI, T.; ZHANG, Y.; WANG, Y.; ZHAO, Q.; LIU, J.; ZHANG, H.; LIU, C.; RAN, Y.; GAO, C. Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. **Nature Communications,** v. 8, article 14261, Jan. 2017.

LIANG, Z.; CHEN, K.; ZHANG, Y.; LIU, J.; YIN, K.; QIU, J. L.; GAO, C. Genome editing of bread wheat using biolistic delivery of CRISPR/Cas9 *in vitro* transcripts or ribonucleoproteins. **Nature Protocols**. v. 13, p. 413-430, Mar. 2018.

LIANG, Z.; ZHANG, K.; CHEN, K.; GAO, C. Targeted mutagenesis in *Zea mays* using TALENs and the CRISPR/Cas system. **Journal of Genetics and Genomics,** v. 41, n. 2, p. 63–68, Feb. 2014.

LIEBER, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. **Annual Review of Biochemistry,** v. 79, p. 181–211, 2010.

LIU, W.; RUDIS, M. R.; CHEPLICK, M. H.; MILLWOOD, R. J.; YANG, J. P.; ONDZIGHI-ASSOUME, C. A.; MONTGOMERY, G. A.; BURRIS, K. P.; MAZAREI, M.; CHESNUT, J. D.; STEWART, C. N. Lipofection-mediated genome editing using DNA-free delivery of the Cas9/gRNA ribonucleoprotein into plant cells. **Plant Cell Reports,** v. 39, p. 245-257, 2020.

MA, X.; ZHU Q.; CHEN Y.; LIU Y-G. CRISPR/Cas9 platforms for genome editing in plants: developments and applications. **Molecular Plant,** v. 9, n. 7, p. 961-974, July 2016.

MALNOY, M.; VIOLA, R.; JUNG, M. H.; KOO, O. J.; KIM, S.; KIM, J. S.; VELASCO, R.; KANCHISWAMY, C. N. DNA-free genetically edited grapevine and apple protoplast using CRISPR/Cas9 ribonucleoproteins. **Frontiers in Plant Science,** v. 7, n. 1904, 2016.

MALZAHN, A.; LOWDER, L.; QI, Y. Plant genome editing with TALEN and CRISPR. **Cell Bioscience,** v. 7, n. 21, Apr. 2017.

MAO, C. Z.; ZHENG, L.; ZHOU, Y. M.; WU, H. Y.; XIA, J. B.; LIANG, C. Q.; GUO, X. F.; PENG, W. T.; ZHAO, H.; CAI, W. B.; KIM, S. K.; PARK, K. S.; CAI, D. Q.; QI, X. F. CRISPR/Cas9-mediated efficient and precise targeted integration of donor DNA harboring double cleavage sites in Xenopus tropicalis. **The FASEB Journal,** v. Online, Jun. 2018.

MAO, Z.; BOZZELLA, M.; SELUANOV, A.; GORBUNOVA, V. Comparison of nonhomologous end-joining and homologous recombination in human cells. **DNA Repair (Amst),** v. 7, n. 10, p. 1765–1771, Oct. 2008.

MONTGOMERY, S. B.; GOODE, D. L.; KVIKSTAD, E. et al. The origin, evolution, and functional impact of short insertion-deletion variants identified in 179 human genomes. **Genome Research**, v. 23, n. 5, p. 749–761, May 2013.

MOSCOU, M. J.; BOGDANOVE, A. J. A simple cipher governs DNA recognition by TAL effectors. **Science,** v. 326, n. 5959, p. 1501, Dec. 2009.

MUSHTAQ, M.; BHAT J. A.; MIR, Z. A.; SAKINA, A.; ALI, S.; SINGH, A. K.; TYAGI, A.; SALGOTRA, R. K.; DAR, A. A.; BHAT, R. CRISPR/Cas approach: a new way of looking at plant-abiotic interactions. **Journal of Plant Physiology**, v. 224–225, p. 156-162, May/June 2018.

MUSSOLINO, C.; ALZUBI, J.; FINE, E. J.; MORBITZER, R.; CRADICK, T. J.; LAHAYE, T.; BAO, G.; CATHOMEN, T. TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity. **Nucleic Acids Research,** v. 42, n. 10, p. 6762–6773, June 2014.

NEMUDRYI, A. A.; VALETDINOVA, K. R.; MEDVEDEV, S. P.; ZAKIAN, S. M. TALEN and CRISPR/Cas genome editing systems: tools of discovery. **Acta Naturae,** v. 6, n. 3, p. 19-40, July 2014.

NUSSBAUM, R. L.; MCINNES, R. R.; WILLARD, H. F. **Genética médica**. 7. ed. Rio de Janeiro: Guanabara Koogan, 2008. 640 p.

OTTO, S. P. The evolutionary consequences of polyploidy. **Cell**, v. 131, n. 3, p. 452-462, Nov. 2007.

PANDIARAJAN, R.; GROVER, A. *In vivo* promoter engineering in plants: are we ready? **Plant Scienc**e, v. 277, p. 132-138, Dec. 2018.

PÂQUES, F.; DUCHATEAU, P. Meganucleases and DNA double-strand break induced recombination: perspectives for gene therapy. **Current Gene Therapy,** v. 7, n. 1, p. 49–66, Feb. 2007.

PENG, A.; CHEN, S.; LEI, T.; XU, L.; HE, Y.; WU L.; YAO, L.; ZOU, X. Engineering canker resistant plants through CRISPR/Cas9targeted editing of the susceptibility gene *CsLOB1* promoter in citrus. **Plant Biotechnology Journal,** v. 15, n. 12, p. 1509–1519, Dec. 2017.

RAMIREZ, C. L.; FOLEY, J. E.; WRIGHT, D. A.; MULLER-LERCH, F.; RAHMAN, S. H.; CORNU, T. I.; WINFREY, R. J.; SANDER, J. D.; FU, F.; TOWNSEND, J. A.; CATHOMEN, T.; VOYTAS, D. F.; JOUNG, J. K. Unexpected failure rates for modular assembly of engineered zinc fingers. **Nature Methods,** v. 5, n. 5, p. 374–375, May 2008.

RICHES, L. C.; LYNCH, A. M.; GOODERHAM, N. J. Early events in the mammalian response to DNA double-strand breaks. **Mutagenesis,** v. 23, n. 5, p. 331–339, Sept. 2008.

ROUET, R.; THUMA, B. A.; ROY, M. D. et al. Receptor-mediated delivery of CRISPR-Cas9 endonuclease for cell-type-specific gene editing. **Journal of the American Chemical Society,** v. 140, n. 21, p. 6596-6603, 2018.

SEHN, J. K. Insertions and deletions (Indels). In: KULKARNI, S; PFEIFER, J. (ed.). **Clinical genomics**. London: Academic Press, 2015. p. 129-150.

SILVA, E.; IDEKER, T. Transcriptional responses to DNA damage. **DNA Repair (Amst)**. v. 79, p. 40-49, July 2019.

SPENCER, N. Y. **Are you still using CRISPR plasmids? Consider ribonucleoprotein (RNP) instead!** IDT. Available at: https://www.idtdna.com/pages/education/decoded/article/are-you-still-using-crispr-plasmids-consider-ribonucleoprotein-(rnp)-instead!. Accessed on: 02 jun. 2020.

STEINERT, J.; SCHIML, S.; PUCHTA, H. Homology-based double-strand break-induced genome engineering in plants. **Plant Cell Reports,** v. 35, n. 7, p. 1429–1438, July 2016.

STRAUSS, B. S. Mechanisms of mutation**.** In: WEISS, R. E.; REFETOFF, S. (ed.). **Genetic diagnosis of endocrine disorders.** 2nd ed. Amsterdam: Academic Press, 2015. p. 3-18.

SUBBURAJ, S.; CHUNG, S. J.; LEE, C.; RYU, S-M.; KIM, D. H.; KIM, J-S.; BAE, S.; LEE, G-J. Site-directed mutagenesis in *Petunia* × *hybrida* protoplast system using direct delivery of purified recombinant Cas9 ribonucleoproteins. **Plant Cell Reports,** v. 35, n. 7, p. 1535–1544, July 2016.

SVITASHEV, S.; SCHWARTZ, C.; LENDERTS, B.; YOUNG, J. K.; CIGAN, A. M. Genome editing in maize directed by CRISPR-Cas9 ribonucleoprotein complexes. **Nature Communications,** v. 7, article 13274, Nov. 2016.

TORNEY, F.; TREWYN, B. G.; LIN, V. S.; WANG, K. Mesoporous silica nanoparticles deliver DNA and chemicals into plants. **Nature Nanotechnology**, v. 2, p. 295-300, Apr. 2007.

TORRES, E. M.; WILLIAMS, B. R.; AMON, A. Aneuploidy: cells losing their balance. **Genetics,** v. 179, n. 2, p. 737-746, June 2008.

VIANNA, G. R.; ARAGÃO, F. J. L.; RECH, E. L. O sistema biobalístico. In: BRASILEIRO, A. C. M.; CARNEIRO, V. T. C. (ed.). **Manual de transformação genética de plantas**. 2. ed. Brasília. DF: Embrapa, 2015. p. 89-104.

VOUILLOT, L.; THÉLIE, A.; POLLET, N. Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. **G3: Genes, Genomes, Genetics,** v. 5, n. 3, p. 407-415, Mar. 2015.

WALKER, J. R.; CORPINA, R. A.; GOLDBERG, J. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. **Nature,** v. 412, n. 6847, p. 607–614, Aug. 2001.

WANG, Y.; CHENG, X.; SHAN, Q.; ZHANG, Y.; LIU, J.; GAO, C.; QIU, J. L. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. **Nature Biotechnology,** v. 32, n. 9, p. 947–951, Sept. 2014.

WILBIE, D.; WALTHER, J.; MASTROBATTISTA, E. Delivery aspects of CRISPR/Cas for in vivo genome editing. **Accounts of Chemical Research,** v. 52, n. 6, p. 1555-1564, Jun. 2019.

WOLABU, T. W.; CONG, L.; PARK, J-J.; BAO, Q.; CHEN, M.; SUN, J.; XU, B.; GE, Y.; CHAI, M.; LIU, Z.; WANG, Z-Y. Development of a highly efficient multiplex genome editing system in outcrossing tetraploid alfalfa (*Medicago sativa*). **Frontiers in Plant Science,** v. 11, 1063, July 2020.

WOLT, J. D.; WANG, K.; YANG, B. The regulatory status of genome-edited crops. **Plant Biotechnology Journal,** v. 14, n. 2, p. 510–518, Feb. 2016.

WOO, J. W.; KIM, J.; KWON, S. I.; CORVALÁN, C.; CHO, S. W.; KIM, H.; KIM, S-G.; KIM, S-T.; CHOE, S.; KIM, J-S. DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. **Nature Biotechnology,** v. 33, p. 1162–1164, Oct. 2015.

XIE, K.; MINKENBERG, B.; YANG, Y. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. **Proceedings of the National Academy of Sciences of the United States of America,** v. 112, n. 11, p. 3570-3575, Mar. 2015.

XIE, K.; YANG, Y. RNA-guided genome editing in plants using a CRISPR/Cas system. **Molecular Plant,** v. 6, n. 6, p. 1975–1983, Nov. 2013.

YANG, K.; GUO, R.; XU, D. Non-homologous end-joining: advances and frontiers. **Acta Biochimica et Biophysica Sinica,** v. 48, n. 7, p. 632-640, July 2016.

YOUNG, J.; ZASTROW-HAYES, G.; DESCHAMPS, S. et al. CRISPR-Cas9 editing in maize: systematic evaluation of off-target activity and its relevance in crop improvement. **Scientific Reports,** v. 9, article 6729, Apr. 2019.

YU, C.; ZHANG, Y.; YAO, S.; WEI, Y. A PCR-based protocol for detecting indel mutations induced by TALENs and CRISPR/Cas9 in zebrafish. **PLoS ONE,** v. 9, n. 6, e98282, June 2014.

ZHANG, T.; GAO, Y.; WANG, R.; ZHAO, Y. Production of guide RNAs *in vitro* and *in vivo* for CRISPR using ribozymes and RNA polymerase II promoters. **Bio-Protocol,** v. 7, n. 4, e2148, Feb. 2017.

ZHANG, X-H.; TEE, L. Y.; WANG, X-G.; HUANG, Q-S.; YANG, S-H. Off-target effects in CRISPR/Cas9-mediated genome engineering. **Molecular Therapy Nucleic Acids,** v. 4, e264, 2015.

ZHAO, X.; MENG, Z.; WANG, Y.; CHEN, W.; SUN, C.; CUI, B.; CUI, J.; YU, M.; ZENG, Z.; GUO, S.; LUO, D.; CHENG, J. Q.; ZHANG, R.; CUI, H. Pollen magnetofection for genetic modification with magnetic nanoparticles as gene carriers. **Nature Plants,** v. 3, p. 956-964, Nov. 2017.

ZHU, X.; XU, Y.; YU, S.; LU, L.; DING, M.; CHENG, J.; SONG, G.; GAO, X.; YAO, L.; FAN, D.; MENG, S.; ZHAND, X.; HU, S, TIAN, Y. An efficient genotyping method for genome-modified animals and human cells generated with CRISPR/Cas9 system. **Scientific Reports,** v. 4, n. 6420, Sept. 2014.

ZURIS, J. A.; THOMPSON, D. B.; SHU, Y.; GUILINGER, J. P.; BESSEN, J. L.; HU, J. H.; MAEDER, M. L.; JOUNG, J. K.; CHEN, Z. Y.; LIU, D. R. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. **Nature Biotechnology**, v. 33, n. 1, p. 73-80, Jan. 2015.