Validation of sgRNA targeting for CRISPR/Cas9 editing in *Brachypodium distachyon.*

Jorge Martínez-Fortún

PhD Thesis

Supervisors

Dr. Dylan Phillips.

Lecturer in Genetics.

Aberystwyth University.

Prof. Huw Jones.

Chair in Translational Genomics for Plant Breeding.

Aberystwyth University.

Thesis Word Count: 57,246

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Candidate Name: Jorge Martínez-Fortún

Signature:

Date: 25/03/2021

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).*

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signature:

Date: 25/03/2021

*[*this refers to the extent to which the text has been corrected by others]*

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signature:

Date: 25/03/2021

Abstract

Genome editing techniques have been in the spotlight for several years due to the potential they have in research and crop development and the uncertainty about the possible regulatory frameworks necessary for their use as a crop development tool. CRISPR/Cas9 has been the most prominent of all these tools, given the simplicity of the system and the high accuracy and efficiency when inducing targeted modifications in a specific sequence. Along the years, many improvements have been achieved in the use of CRISPR/Cas9 as a lab tool, including optimisations in the system such as codon optimisation of the Cas9, sgRNA length and sequence optimisation and the use of different promoters for the expression of the editing machinery. At the same time, different crops have been edited using this tool. One of the most relevant clades commercially, the grasses, has been the focus of many studies reporting the use of CRISPR/Cas9. Within this group, *Brachypodium distachyon* is rising as a prominent model due to its many advantages for its use in lab conditions when compared to other plants in the family. Although CRISPR/Cas9 has been used in this species, many of the optimisations previously mentioned are still to be achieved.

With this context in mind, this project focused on the use of CRISPR/Cas9 in *Brachypodium distachyon*, finding new transformation methods for the introduction of the editing machinery in different tissues of the plant, including protoplasts and embryos, and investigating the possibility of using different promoters for the expression of the sgRNA. Additionally, the optimisation of the sgRNA sequence for its use in plants was assessed, proposing a new model for the prediction of editing efficiencies based on published sources. Finally, the regulatory context of edited crops was reviewed, focusing on the comparison of the possible outcomes obtained from different traditional breeding techniques and the possibility of the use of genome editing to replicate these outcomes. Finding possible equivalences between traditional breeding techniques and genome editing will help the case for an equivalent and proportional regulatory framework for the future.

Table of Contents.

List of figures.

List of tables

Acknowledgements.

During my time as a PhD student, many people have had an impact on me as a researcher and as a person, people who without their support completing this project would have not been possible. Firstly, I would like to thank my supervisors, Drs. Huw Jones and Dylan Phillips, for their guidance and support both inside and outside the lab, and for giving me the opportunity of going through this journey. I would also like to thank everyone I had the pleasure of sharing a laboratory with, including Dr. Lorenz Fuchs, Chris Henderson and Dr. Anne Maddison.

The student community in Aberystwyth also deserves gratitude, including the SLN group, the lunch guys and the Aberystwyth Spanish community. They have made me feel at home when home was so far away and without them, I would not have gotten to the place I am today.

I would also like to thank Noelia, the rock that supported me during the hardest times of the project and has stopped me from crumbling apart during these hard times.

I would like to thank all my friends and loved ones in Spain. I would like to thank them for supporting me even when I disappeared and was too busy to send anything. I would also like to thank my family, Julio, Maria Victoria and Ana for supporting me and being someone I could always rely on when things got rough.

Finally, I would like to make a special mention and express my gratitude to all the essential workers that have made the greatest sacrifices during these hard times.

Chapter 1. Introduction.

1.1. *Brachypodium distachyon* as a model plant.

Research in plant genetics often relies on a reference plant species. These model plant species can be used to extrapolate the results to plants within the same group, depending on the process studied. Plants such as bryophytes have been used for studying basic processes in plant biology, such as the protein mechanisms of the alternation of gametophytic and sporophytic generations (Mosquna et al., 2009; Okano et al., 2009), which relates to most land plants. More specific processes, such as floral identity, require model species that are more closely related, such as *Anthirrinum majus* (Snapdragon) or *Arabidopsis thaliana* (from now on Arabidopsis, Coen and Meyerowitz, 1991). The transfer of knowledge to other plants based on findings on a model will depend on the relationship between both plants and the processes studied.

Another consideration to make when selecting a model is the feasibility of its use in a research setting. This includes the logistics of its growth in controlled conditions and in large numbers, which requires a plant that has a small size, short life cycle and produces a large number of seeds per generation. In addition to these natural properties, additional characteristics such as crossing methods, *in vitro* growth and tissue culture techniques are important for its use in the laboratory. Once a plant become adopted by the research community it will lead to the development of other relevant resources, such as genome sequencing, transcriptome analysis or protein profiling. The development of techniques such as transformation, random and targeted mutagenesis will help a plant become established as a model. Another factor to consider would be the popularity of its use in research. This drives a positive feedback loop, since the easier it is to study, the more research will be done, which will mean new techniques and findings will be made, which will make it easier to study.

The usefulness of a model relies on how easy it is to work with and how the studies made in the model can relate to crops with commercial value. However, a model will only be useful when an advantage is present when compared to working directly into the crop of interest. The main advantage will be the speed findings can be obtained, and studies can be done faster in the model plant given the characteristics previously described. Then, these findings can be transferred to the commercial crop.

A wide variety of plant models have been used to represent different clades or biological processes. (Figure 1.1, Chang et al., 2016). Focusing on genetics research, one of the key features is the availability of an assembled genome sequence. Genome size varies a lot between plant genomes (Michael, 2014), from the 184 Gb from *Paris japonica* (Pellicer et al., 2010) to the 82 Mb in *Urticularia gibba* (Ibarra-Laclette et al., 2013). Having a sequenced and assembled genome naturally leads to genome annotation studies, which then can be used to determine gene function. Ultimately this knowledge can be transferred from the model plant to the plant of interest once equivalent resources have been developed.

The most prominent model plant is *Arabidopsis thaliana*. This eudicot is found in the *Brassicaceae* family, which contains plants of such economic importance as mustards, cabbages, turnips, cauliflowers, kale, rocket and radish. *Brassicaceae* is located within the Rosid clade, which includes other important plants like grapes (*Vitis*), legumes (*Fabaceae*) or fruit crops (*Rosaceae*). This highlights the importance of *Arabidopsis* given its location in the tree of plant life. *Arabidopsis* is compact, at 25 cm tall, and has a generation time of 6 weeks. This makes it perfect for its growth in controlled conditions in high numbers. The availability of inbred nearly isogenic lines has also facilitated its use in the laboratory and functional genetics research, such as Columbia-0. At the same time, Arabidopsis also allows outbreeding to obtain plants with combinations of traits from specific desired parentals.

Thorough the years, different techniques have been developed for the study of functional genetics (Salinas and Sanchez-Serrano, 2006). The development of these techniques has led to a deep

understanding of the biology and genetics of Arabidopsis. It has a manageable and described genome, consisting of 125 Mb distributed among 5 chromosomes (The Arabidopsis Genome Initiative, 2000), not including the 154,478 bp of chloroplast genome (Sato et al., 1999). The methylation state of this genome has been described as well (Yang et al., 2015). How this genome is expressed in different conditions has also been studied (Mockler et al., 2007; Winter et al., 2007).

All these molecular data have been obtained thanks to a wide variety of laboratory techniques described and optimized for this plant. One of the key protocols has been the introduction of foreign DNA using *Agrobacterium tumefaciens* mediated transformation (Clough and Bent, 1998; Krysan et al., 1999). The use of this technique, together with screening methods, provides a mechanism of introducing a transfer-DNA (T-DNA), which can be modified to include a desired DNA sequence, making possible the insertion of foreign DNA sequences. If the T-DNA lands in a gene it will cause an insertional mutation, allowing for large T-DNA libraries to be developed (Szabados et al., 2002).

One of the families with the highest commercial value is the *Poaceae*, which includes the plants known as grasses (Soreng *et al.* 2015). Plants from this group provide food that constitutes around 50% of the calory intake worldwide (Alexandratos and Bruinsma, 2012). This group is found in the monocot clade, which is outside the eudicot clade Arabidopsis belongs to. This implies that findings from Arabidopsis may not be directly translatable to the *Poaceae*. An example of both cases is the flowering and heading processes, where both Arabidopsis and the cereals share the mechanism for the change from somatic meristem to reproductive meristem, but differ in the way they respond to cold conditions during the vernalization process (Greenup et al., 2009). This implies that Arabidopsis may not be the optimal model for the *Poaceae* family. Over the years, different crops have been proposed as a model, such as the green foxtail (*Setaria viridis*, Brutnell et al., 2010; Huang et al., 2016), maize (*Zea mays*, Strable and Scanlon, 2009; Portwood et al., 2019), rice (*Oryza sativa*, Izawa and Shimamoto, 1996; Hong et al., 2019) and wheat (*Triticum aestivum*, Appels et al., 2018), together with *Brachypodium distachyon* (from now on *Brachypodium*, Draper et al., 2001; Scholthof et al., 2018). When comparing these crops to Arabidopsis as a model (Holland and Jez, 2018), each one of these crops have their advantages and disadvantages as a model (summarized in Table 1.1, reviewed in Borrill, 2019).

Among all the species reviewed in Table 1.1, *Brachypodium* possesses the natural traits necessary in a model plant, such as a reduced height (around 20 cm), a rapid life cycle (2-3 months) and high number of seed in each generation. These make its growth manageable under glasshouse conditions, with described optimal conditions for its growth both in soil and *in vitro* (Bablak et al., 1995; Betekhtin et al., 2020). Another advantage of *Brachypodium* as a model is the fact that it is self-compatible (Khan and Stace, 1999), allowing both inbreeding and outbreeding approaches. A wide range of techniques have been developed for *Brachypodium* (Sablok et al., 2017), including *Agrobacterium*-mediated transformation (Alves et al., 2009; Lee et al., 2011; Thole et al., 2012).

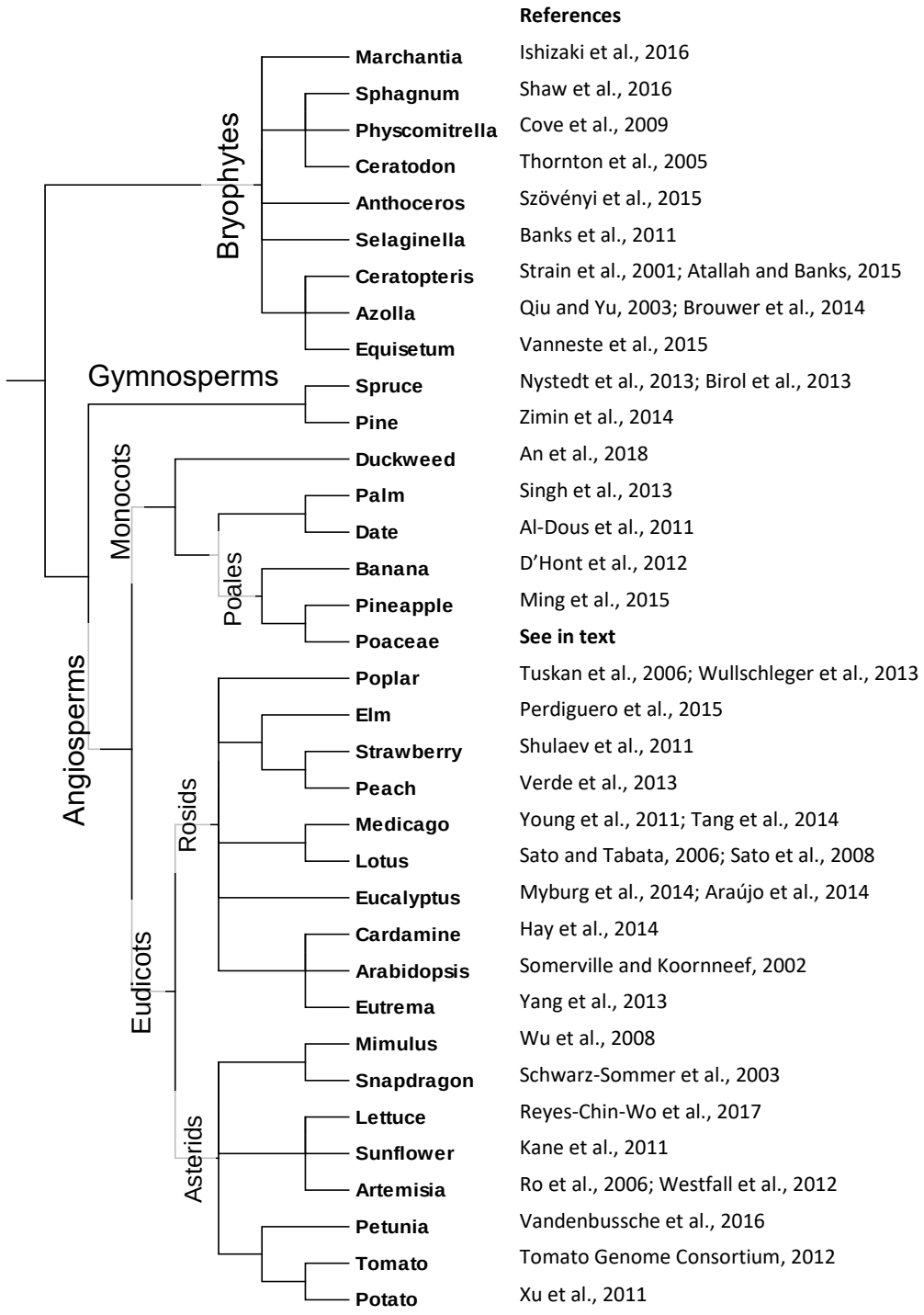


Figure 1.1. Taxonomic grouping of different vascular model plants shows different represented groups.

List of species related by taxonomic grouping that have been defined as models for different processes. The list of species was taken from Chang et al. (2016) and the taxonomic grouping was done by designing a Newick Tree displayed using MEGAX (Kumar et al., 2018).

Table 1.1. Summary of different traits related to the viability of some cereal crops as a model (Adapted from Borrill, 2019).

		Arabidopsis thaliana	*Brachypodium distachyon*	*Setaria viridis*	*Zea mays* ssp. *mays*	*Oryza sativa ssp. japonica*	*Triticum aestivum*
Classical traits	Height	20–25 cm (Krämer, 2015)	15–20 cm (Li et al., 2012a)	≈30 cm (Huang et al., 2016)	>2m (Gyenes-Hegyi et al., 2002)	≈1m (Weng et al., 2014)	≈1m (Wang et al., 2010)
	Fast lifecycle (months)	2 (Chang *et al.*, 2016)	2 (Brutnell, 2015)	2 (Brutnell, 2015)	6–12 (Chang *et al.,* 2016) 2 for rapid lifecycle varieties (McCaw et al., 2016)	4 (Chang *et al.*, 2016) 2 for rapid lifecycle varieties (Hu et al., 2018b)	4–6 (Watson et al., 2018) 2 for rapid lifecycle varieties or speed breeding (Bugbee and Koerner, 1997; Watson et al., 2018)
	Genome Size (ploidy) [I]	≈135 Mb (2x=10)	≈355 Mb (2x=10)	≈395.7 Mb (2x=18) (Huang et al., 2016)	≈2400 Mb (2x=20)	≈500 Mb (2x=12)	≈17000 Mb (commonly 6x = 42)
Derived resources	Genetic resources	TAIR	JGI Phytozome	JGI Phytozome	MaizeGDB	RGAP funRiceGenes	IWGSC
	Genome assembly [II]	TAIR10	v3.1	v1.0	AGP v3 (B73) v1.1 (PH207)	v7	v1
	Expression atlas [III]	Arabidopsis eFP[IV]	Brachypodium eFP[IV] gene2function.de	JGI Phytozome	Maize eFP[IV]	Rice eFP[IV] RiceXPro	Wheat eFP[IV] wheat-expression.com
	Mutant collections [III]	TAIR T-DNA (Sessions et al., 2002; Alonso et al., 2003) Transposon (Parinov et al., 1999; Kuromori et al., 2004)	JGI T-DNA (Bragg et al., 2012)	Not available	MaizeGDB Transposon (Settles et al., 2007; Williams-Carrier et al., 2010)	RiceGE T-DNA (Zhang et al., 2006) Retrotransposon (Miyao et al., 2003) Fast neutron (Li et al., 2017)	EMS mutations (Krasileva et al., 2017) Wheat TILLING

I. Data obtained from Ensembl Genomes unless otherwise specified.

II. Latest version of the genome assembly found in JGI Phytozome. If different varieties are available they are specified between brackets.

III. For further information see references in the text about each one of the species.

IV. Electronic Fluorescence Pictographs from The Bio-Analytic Resource for Plant Biology (Bar.toronto) (Winter et al., 2007)

To further facilitate genetic studies in *Brachypodium*, isogenic inbred lines have been developed, such as Bd21-3 (Vogel et al., 2006a), being the basis of the description of its genetic characteristics. Its 5 chromosome genome structure has been described cytologically (Idziak et al., 2015), with the full genomic (Vogel et al., 2010) and chloroplast sequence (Bortiri et al., 2008) available. A collection of Expressed Sequence Tags (ESTs) has been obtained to describe its transcriptome (Vogel et al., 2006b). Different microarray experiments have been converted into an Electronic Fluorescent Pictograph that can be used to compare gene expression in different conditions and plant tissues (Sibout et al., 2017). The methylation state found in the genome of *Brachypodium* has also been described (Eichten et al., 2016). Random mutagenesis using sodium azide, EMS (Dalmais et al., 2013), or T-DNA insertions (Bragg et al., 2012) has been used to obtain further information on gene function.

The description of these characteristics helped establish *Brachypodium* as a plant model and has been used to study different traits, such as response to day-night cycles (MacKinnon et al., 2020) and light (Tran et al., 2018), root (Watt et al., 2009; Nam et al., 2020) and shoot growth (Schillaci et al., 2021; Martin et al., 2020), seed development (Laudencia-Chingcuanco and Vensel, 2008; Larré et al., 2010; Guillon et al., 2011), cell wall composition (Rancour et al., 2012; Coomey et al., 2020), xylan biosynthesis (Petrik et al., 2020) and arbuscule density (Müller et al., 2020), in addition to abiotic (Verelst et al., 2013; Rivera-Contreras et al., 2016; Sade et al., 2020; Lu et al., 2020; Yoon & Seo, 2021) or biotic stress (Parker et al., 2008; Fitzgerald et al., 2015; Powell et al., 2017; Kouzai et al., 2020; Scholthof, 2020).

Knowledge gained from studies focused on *Brachypodium* have been transferred to key crop species, such as wheat (Wu et al., 2020; Wang et al., 2020a; Yoon et al., 2021; Wang et al., 2020b; Lu et al., 2020), ginseng (Smith et al., 2021), barley (Hong et al., 2020), rice (Sade et al., 2020), triticale (Zaidi et al., 2020) and sugarcane (Anur et al., 2020; Falter & Voight, 2016), in addition to Arabidopsis (Roh et al., 2021; Wang et al., 2020c; Kim et al., 2020a; Kuromori et al., 2021; Rowlands, 2021).

When comparing *Brachypodium* and Arabidopsis, the number of references related to *Brachypodium* is notably low when compared to the latter. However, a similar trend is observed (Figure 1.2). This implies the strength of *Brachypodium* as a model, together with the need for more work to put it on level with Arabidopsis.

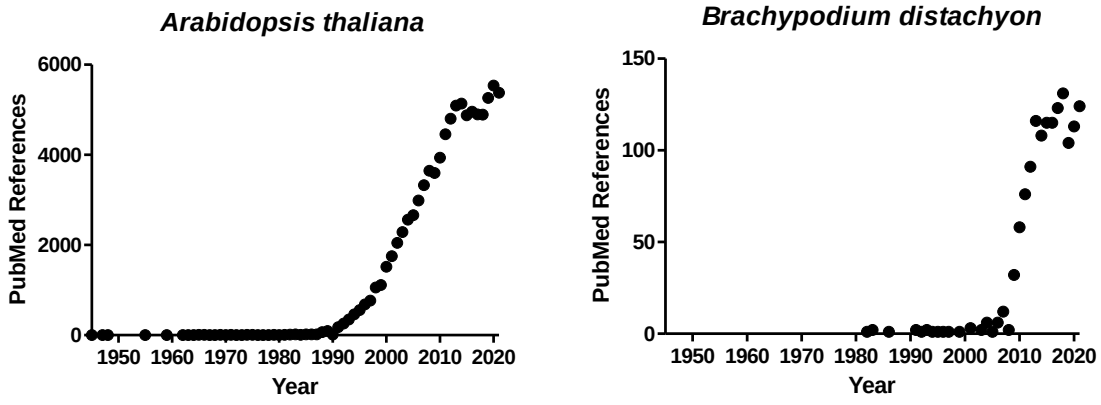


Figure 1.2. The evolution of references with time related to *Brachypodium* shows its potential as a model crop.
Number of references collected in the database PubMed. Data collected by January 2022.

Given the characteristics previously described, the use of *Brachypodium* in the laboratory for genetic studies has an advantage when compared to other crops in the family, highlighting height, generation time and genome size. The use of *Brachypodium* as a model will enable the quick finding of genes of interest and relating them to gene function to extrapolate the results to other crops.

1.2. Genome Editing.

Genome editing techniques are able to deliver targeted changes in a DNA sequence. Targeting relies on interactions between the editing machinery and the genomic DNA, which can be protein-DNA interactions (Meganucleases, TALEN and ZFN), ribonucleoprotein-DNA interactions (CRISPR/Cas) or chimeroplast-DNA interactions (ODM). Depending on the editing machinery causing the modifications, there are two broad types of genome editing: nuclease-based editing (MN, TALEN, ZFN and CRISPR/Cas) and base-directed editing (ODM). Given the plasticity of nuclease-based editing tools, changes similar to base-directed editing can be obtained.

1.2.1. DNA Repair Mechanisms used in Nuclease-based Genome Editing.

Most nuclease-based editing tools are based on the induction of double strand breaks (DSB) in the genomic sequence. These cuts can be repaired using four different mechanisms: Classical Non-Homologous End Joining (C-NHEJ), Alternative Non-homologous End Joining (alt-EJ) and Homologous Recombination (HR), including Single-Strand Annealing (SSA) and Synthesis-Dependant Strand Annealing (SDSA) (Figure 1.3).

When a DSB occurs, Ku heterodimers (Ku70/80) will cover each one of the ends of the break (Yaneva et al., 1997; Hammarsten et al., 2000). Ku70/80 binds to the DNA, forming the DNA-Protein Kinase Complex (DNA-PK, Walker et al., 2001). The DNA-PK complex will be the stepping stone for the DNA Ligase IV (Lig4) to ligate the ends of the DSB (Ellenberger and Tomkinson, 2008). Lig4 is able to ligate strands independently from one another (Ma et al., 2004) and to ligate incompatible ends (Gu et al., 2007), processes necessary and specific to repair via C-NHEJ. To bind to the DNA-PK complex, non-enzymatic proteic factors such as XRCC4 and XLF will have to be bound to the complex (Koch et al., 2004; Palmbos et al., 2008; Riballo et al., 2009). These proteins will comprise the core of the C-NHEJ repair mechanism. Lig4 binds to the complex within 10 minutes from the DSB (Wu et al., 2008b). The quickness of this process, however mutagenic, will protect the integrity of the genome (Difilippantonio et al., 2000).

Extensive damage may require the use of alt-EJ, implying a resection of the damaged DNA. After removal of the Ku proteins (Audebert et al., 2004; Wang et al., 2006), the MRN (for MRE11, RAD50 and NSB1) complex will position itself on the DSB. To keep the ends of the break together to help synapsis, PARP-1 has been suggested to interact with the DSB site when Ku does not bind, recruiting the factors required for alt-EJ (Audebert et al., 2004, 2006; Wang et al., 2006; Sallmyr et al., 2008; Della-Maria et al., 2011). Regarding the MRN complex, MRE11 has been related to microhomology pairing, the first step necessary for alt-EJ repair (Zhang and Paull, 2005; Williams et al., 2008). The complex will act as a manganese dependant nuclease that will cleave both strands of the DNA trimming the damaged nucleotides (Trujillo et al., 1998). RAD50 will act as a ATPase (Hopfner et al., 2000), which will change the activity of MRE11 between exonuclease and endonuclease (Majka et al., 2012). The suggested role of NSB1 in this complex is structural or protein recruitment. To fill the gaps created by the resection process, a DNA polymerase from the PolX family will be required (Daley and Wilson, 2005).

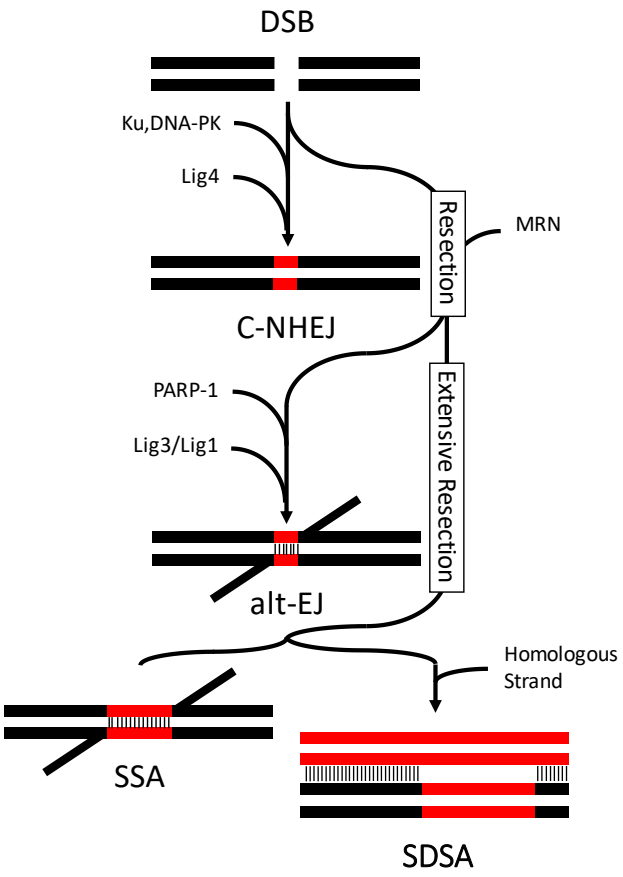


Figure 1.3. Simplified diagram of the different DSB repair mechanisms.
Schematic representation of different DSB repair mechanisms. In black the two DNA strands of the gDNA are represented, with the repaired DNA in red. The connecting lines represent the different possibilities depending on the existence and extension of the resection process and the proteins involved in the respective processes. In vertical lines between the DNA strands, homologies in the sequence are shown.

The alt-EJ process is also independent from Lig4. In mammalians, Lig3 (Audebert et al., 2004; Wang et al., 2005; Sallmyr et al., 2008; Simsek et al., 2011) and Lig1 (Liang et al., 2008; Simsek et al., 2011) have been suggested to fulfil this role. In contrast with C-NHEJ, alt-EJ will base the final ligation of the ends on the appearance and microhomologies around the DSB and end clipping. This can lead to genomic rearrangements or mutations causing bigger changes in the original sequence than C-NHEJ.

If the extensive resection occurs, other proteins such as exonucleases, helped by helicases, will trim even further, leaving long ssDNA chains that can be repaired using HR (Symington and Gautier, 2011; Sturzenegger et al., 2014). This process will allow mechanisms such as Single Strand Annealing (SSA) or Synthesis-Dependant Strand Annealing (SDSA), which rely on macro-homologies. In SSA, the extensive resection will continue until a homologous sequence is found in both sides of the break, annealing both ssDNA chains left by it and proceeding to ligate the breaks. This usually leads to loss of genetic material. The macro-homologies necessary for SSA repair of a DSB can also come from a different chromosome, leading to chromosomal translocations by exchanging a chromosome arm (Pacher et al., 2007). In plants, the proteins specific for this process have not been identified. In other eukaryotes RAD52 is suggested to play a role in annealing and recruiting the machinery necessary for completing the ligation using DNA Ligase I (Ceccaldi et al., 2016).

Another HR pathway, SDSA, uses an homologous dsDNA as a template for repairing the ssDNA that just formed during the resection. The template invasion will be solved in two different ways. DSBR (DSB Repair) will be based on Holliday Junctions, using the homologous strand in the dsDNA as a template and leading to a crossover event once the second strand of the ssDNA is synthesized. SDSA (Synthesis-Dependant Strand Annealing) will be based on the formation of a replication bubble with the invading ssDNA and the homologous dsDNA, which will be extended until a macro-homology is formed. This then will be sealed in a SSA-like manner (Gorbunova and

Levy, 1999). These processes will be highly conservative, trying to minimize the loss of genetic material and to keep the sequence as close to the original as possible.

Most nuclease-based editing will rely on DSB repair to achieve the targeted genome modification. Usually, that break will be repaired by the C-NHEJ pathways, causing insertion/deletions (indels) in the targeted sequence. *A priori*, HR pathways will be a conservative manner of fixing the breaks. However, this has been exploited by scientists to induce specific mutations by providing an homologous strand containing the desired modification.

1.2.2. Targeting Mechanisms.

1.2.2.1. MegaNucleases.

Meganucleases (MNs), were the first type of protein proposed as an editing tool (Reviewed in Silva et al., 2011 and Stoddard, 2011). The original function of these proteins, also known as homing endonucleases, is to cleave DNA in specific sites to remove type II introns or inteins (Dujon et al., 1989; Lambowitz and Belfort, 1993; Belfort and Perlman, 1995; Belfort and Roberts, 1997; Chevalier and Stoddard, 2001). MNs are based on the eukaryotic LAGLIDADG family of homing endonucleases (Orlowski et al., 2007; Zhao et al., 2007). This motif is essential for enzymatic activity, and it can appear as a single motif (e.g. I-CreI, Heath et al., 1997) or two motifs (e.g. PI-SceI, Duan et al., 1997), being SceI the prototypical MN designed for genome editing. The functional domain has been related to HNH-type nucleases (Dalgaard et al., 1997). The αββαββα structure that the proteins had the LAGLIDADG motif in the terminal part of the first helix, comprising the catalytic center of the protein, with the β strands creating the DNA binding region that attaches to the mayor grooves of the genomic DNA (Figure 1.4, Flick et al., 1998; Jurica et al., 1998). These proteins are able to bind specifically to sequences between 14 and 40 base pairs long. Other regions in the proteins have been found to be important for targeting (Prieto et al., 2007).

By interchanging or modifying the α/β domains, the specificity of the original protein targets has been modified (Chevalier et al., 2002; Epinat et al., 2003; Silva et al., 2006). The change in specificity have also been obtained by introducing specific mutations in the proteins based on a semi-rational design (Seligman et al., 2002; Sussman et al., 2004; Doyon et al., 2006; Rosen et al., 2006; Chen et al., 2009). Inducing modifications in only one of the strands instead of both of them via DSB induction has been obtained by the use of targeted Mega Nickases (I-SceI, Niu et al., 2008; I- AniI, (McConnell Smith et al., 2009). Targeted gene insertion has also been obtained using MNs (Choulika et al., 1994, 1995).

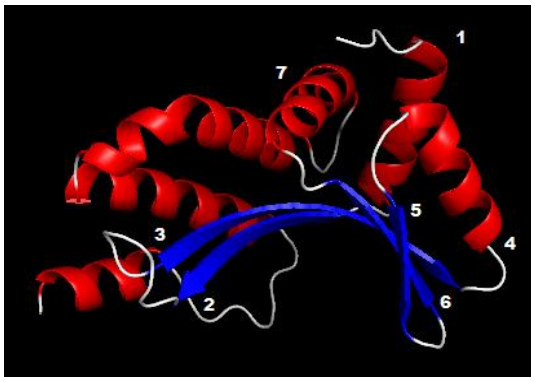


Figure 1.4. LAGLIDADG Domain in Homing Nucleases.
Representation of the LAGLIDADG domain in Homing Nucleases. Labeled from 1 to 7 are shown the αββαββα structures in the protein. Image obtained from the I-CreI sequence of *Chlamydomonas reindhartii* from UniprotKB. The structures was obtained using Phyre2 (Kelley et al., 2015) and the image was obtained using The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

Protein engineering of MNs has been proven complicated, given the difficulties in predicting specific structures within the protein when inducing modifications of the DNA targeting

domains. In spite of these challenges, an online tool for MNs design is available (Taylor et al., 2012). MNs will be useful as an editing tool when there is an existing site that has been previously described, but the plasticity of the tool for general genome editing is limited. Nevertheless, some examples of plants edited using this technique have been reported (Gao et al., 2010).

1.2.2.2. Zinc Finger Nucleases (ZFN).

Zinc Finger (ZF) domains are one of the most common DNA binding domains in nature, appearing in Zinc Finger proteins (ZFPs), a group of eukaryotic transcription factors. To convert a ZF into an editing tool a FokI endonuclease was attached in order to produce a DSB (Reviewed in Urnov et al., 2010).

ZFs are able to bind to the DNA thanks to a tandem array of Cys2-His2 fingers (Figure 1.5A, B), which will bind specifically to approximately 3 nucleotides each (Wolfe et al., 2000; Miller et al., 2001). Different ZFs can be linked sequentially to recognise longer sequences, but each one of the individual ZF will still recognise 3 base pairs. Each ZF will interact with the target DNA and with the surrounding ZF domains (Pavletich and Pabo, 1991). The DNA-interacting residues will also interact with the nucleotides surrounding the target triplet (Fairall et al., 1993; Pavletich and Pabo, 1993; Houbaviy et al., 1996; Nolte et al., 1998; Wolfe et al., 2001; Segal et al., 2006), which makes the design of a specific ZF to a specific triplet more complicated.

An array of ZF will help targeting a FokI nuclease to a specific sequence (Figure 1.5C, Nolte et al., 1998). FokI only produce a DSB when it dimerizes due to the weakness of the interaction with the DNA (Vanamee et al., 2001), and ensures the specificity of the cut. To target a specific position in the genome, the dimerization requires a specific spacing and orientation between the ZF. This will allow targeted modifications using Zing Finger Nucleases (ZFNs).

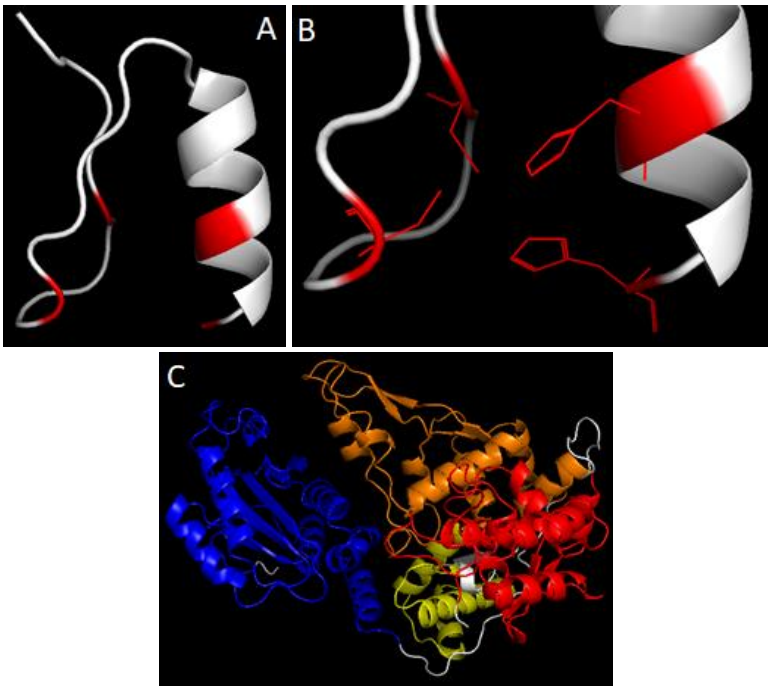


Figure 1.5. Parts of a Zinc Finger Nuclease.
(A, B) Single Zinc Finger protein structure obtained from the sequence described in Brown, Sander, & Argos (1985). Highlighted in red are the 2 Cysteines and 2 Histidines responsible of DNA interaction. (C) FokI nuclease structure obtained using the sequence described in Pernstich & Halford (2012). In blue is shown the nuclease domain and in orange, red and yellow the 3 recognition domains. The structures were obtained using Phyre2 and the images was obtained using The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

1.2.2.3. TALEN.

Transcription Activating-Like Effector Nucleases (from now on TALEN), as ZFNs, are based on the use of proteins to recognise specific DNA sequences in order to target a FokI nuclease. To achieve this sequence specific targeting, the *Xanthomonas* pathogen Transcription Activator-Like Effectors (TALE) domains were used. These proteins have a DNA binding domain based on several repeats of 34 amino-acids (Figure 1.6A) which will surround the DNA double strand. The protein sequence of these proteins will be highly conserved, except for the twelfth and thirteenth residues, the Repeat-Variable Di-Residues (RVD). The twelfth residue will stabilize the loop, while the thirteenth will bind directly to the DNA, recognizing a different base depending on this RVD (Figure 1.6B, Moscou and Bogdanove, 2009). Depending on the RVD, a TALE will recognize a specific base, being NI for Adenine, NG for Thymine, HD for Cytosine and NN for Guanine/Adenine (Osakabe and Osakabe, 2015). The combinations of residues that recognize each one of the possible nucleotides found in the DNA are already described (Boch et al., 2009), making the design of targeting proteins simple.

As for the ZFNs, the activation of the FokI nuclease attached to the TALE repeats will depend on the dimerization of the protein. This implies that the mechanism of design and action of the TALEN will be similar to a ZFN.

It has been shown that regions outside the binding and endonuclease domain affect the editing efficiency. Examples of these are the effects of truncation (Miller et al., 2011; Mussolino et al., 2011; Sun et al., 2012, 2013) or deletion (Voytas, 2013) of the amino and carboxy-terminal regions of the TALEN backbone, which showed an increase in editing efficiency when compared with complete (Cermak et al., 2011; Li et al., 2012b) TALEN backbones (Voytas, 2013).

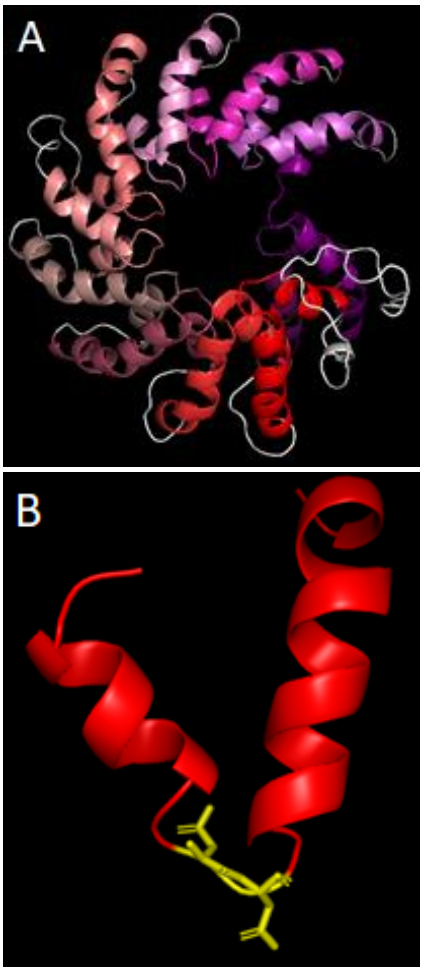


Figure 1.6. Structure of a TALE repeat.
(A) Protein structure prediction of an array of 11 TALE Repeats from *Xanthomonas oryzae* avirulence protein. In different colours we observe the different repeats. (B) Image of a single TALE Repeat. In yellow is highlighted the RVD region
The structures were obtained using Phyre2 and the images was obtained using The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

TALENs are easier to design than either MN or ZF, given the smaller repeat size, which simplifies the modular assembly, and the ability of targeting specific nucleotides instead of the codon targeting, in contrast with ZFNs.

1.2.2.4. CRISPR/Cas.

In contrast with the mentioned genome editing tools that base their targeting in protein-DNA interactions, CRISPR/Cas (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR Associated protein) will rely on ribonucleoprotein-DNA interactions for the targeting, which will lead to the DSB induction by an endonuclease. There are up to six different CRISPR/Cas systems, classified depending on the CRISPR/Cas *loci*, using different Cas proteins and guiding RNA (Makarova et al., 2011). The most described and used has been the type II CRISPR system from *Streptococcus pyogenes*, based on the Cas9 endonuclease (Jinek et al., 2012).

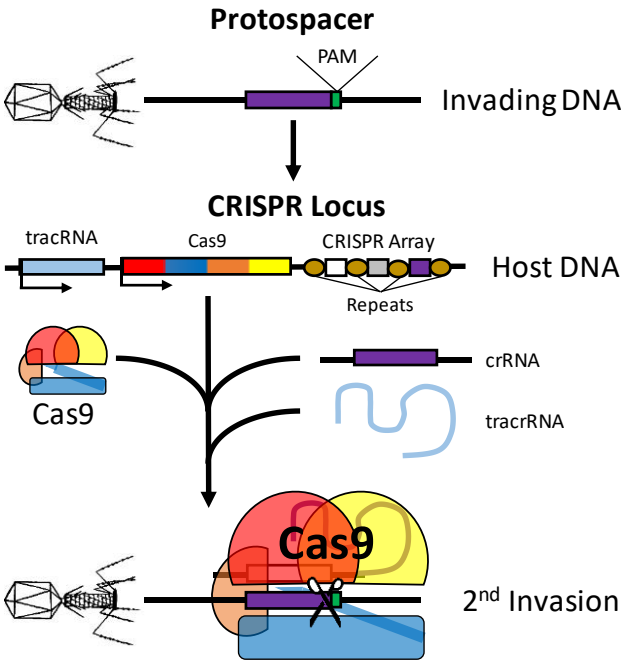


Figure 1.7. Diagram of the natural mechanism of viral resistance with CRISPR/Cas9.
Diagram showing the process of foreign DNA cleavage and insertion as spacers in the CRISPR Locus to use it as a guide for targeting the cleavage of the DNA introduced in a second viral invasion. Purple boxes represent viral sequences, red, blue, orange and yellow boxes represent the sequence of the different domains of the Cas9 protein, light blue boxes represent the tracrRNA sequence, dark yellow circles represent the CRISPR repeats and green boxes represent the PAM site. The colours of lines represent the expressed factors, each one correlating with the previously mentioned colour coding.

In its natural state, the CRISPR/Cas9 system will be the defence mechanism of a wide variety of organisms, from bacteria to archaea, against viral infection. In the natural system, a virus will introduce its DNA in the cell. This will be cleaved and integrated in the CRISPR locus. The integrated fragments will be called spacers and will be included between two repeats. Given a second infection of the same virus, these will be transcribed, forming the CRISPR RNA (crRNA). Together with a transactivating CRISPR RNA (tracrRNA), they will recruit and activate the Cas9 nuclease (Doudna and Charpentier, 2014). The targeting is dependent on homologies between the inserted DNA (spacer) and the DNA from a second infection (protospacer, Barrangou et al., 2007). For the system to work, a protospacer adjacent motif (PAM), usually a 5'-NGG-3', right next to the protospacer sequence is required (Gasiunas et al., 2012; Jinek et al., 2012). The existence of a PAM site will be essential to distinguish self and foreign DNA sequences to cleave (Marraffini and Sontheimer, 2010). This process points to the existence of an adaptative immune system in bacteria and archaea against DNA virus (Figure 1.7, Bortesi and Fischer, 2015).

The structure of the crRNA and tracrRNA complex allows the correct placement of the Cas9 around the target dsDNA (Figure 1.8A). For its use as a molecular biology tool, the crRNA and tracrRNA has been combined to form a synthetic RNA called the small guide RNA (sgRNA). The first component in this sgRNA will be the 20-nucleotide long spacer sequence required for homology recognition. In this sequence, the first 10-12 nucleotides upstream from the PAM site (seed sequence, Jinek et al., 2012; Cong et al., 2013; Sternberg et al., 2014) will be key for target

specificity (Semenova et al., 2011; Wiedenheft et al., 2011). Mismatches in this sequence can lead to off target effects (Pattanayak et al., 2013). The second main component will be tracrRNA (Trans-Activating CRISPR RNA), which will be in charge of supporting the recruitment of the Cas9 and act as a scaffolding for its correct positioning on the target DNA. The recruitment of the Cas9 is due to several contacts within the sgRNA sequence downstream from the PAM site (Figure 1.8A). The structure of the scaffold formed is based in a first anti-repeat in the tracrRNA that will match the repeat sequence in the crRNA, forming a first loop based on crRNA and tracrRNA sequence homology. After it, a series of stem loops will form in order to provide further scaffolding for the Cas9. The interactions between the Cas9 and the tracrRNA will come from direct contact. The most important will be between the repeat-anti-repeat duplex and the first stem loop, that will contact the α-bridge and the PAM interaction motif.

The additional loops will have fewer contacts, interacting with other parts of the PAM interaction motif and the RuvC nuclease domain. In the modified sgRNA version, the shortening of the sequence reduces the contacts between the Cas9 and the scaffold sequence. Several studies have shown that losing the other stem loops make the Cas9 activation less efficient, but do not completely inactivate DSB induction (Jinek et al., 2013), whereas losing the key repeat-anti-repeat and first stem loop blocks cleavage (Jinek et al., 2012). *In vivo* studies have shown that that additional loops help stabilize the RNA binding to the Cas9 and increase editing efficiency (Hsu et al., 2013; Mali et al., 2013; Nishimasu et al., 2014; Wright et al., 2015; Mekler et al., 2016).

The Cas9 protein, as showed in Figure 1.8B, is comprised of two lobes that have different functions. The α-helical domain is responsible for DNA recognition (REC lobe), while the NUC domain, comprised by an HNH-like nuclease and a RuvC like nuclease domains cleave the target strand (complementary to crRNA) and the opposite nontarget strand respectively (Gasiunas et al., 2012; Jinek et al., 2012; Chen et al., 2014). Both parts of the protein will be connected by the α-bridge between the first RuvC domain and the REC1 lobe and a second linker between the REC2 lobe and the second RuvC domain, completing the nuclease domain (Nishimasu et al., 2014). In addition to these lobes, the PAM interaction domain, required for correct recognition and cleavage of the target sequence (Figure 1.8C, Anders et al., 2014) is found as an elongated C-terminal domain.

In order to successfully search and bind to the target DNA, sgRNA must first bind to the Cas9 protein (Jinek et al., 2014; Jiang et al., 2015). This will lead to a conformation change in the protein that will enable it to bind and detach from the DNA as it searches for potential PAM sites (Sternberg et al., 2014; Singh et al., 2016). PAM recognition (Anders et al., 2014) and interaction with the guide RNA will be essential for protein activation (Jinek et al., 2012), with the protein remaining inactive as it scans the DNA. Once it finds a potential PAM site, the DNA will be melted, separating both strands and enabling RNA invasion (Anders et al., 2014; Sternberg et al., 2014; Szczelkun et al., 2014; Rutkauskas et al., 2015; Jiang et al., 2016; Palermo et al., 2016). Depending on the homology found between the guide RNA and the target DNA, the Cas9-sgRNA complex will remain long enough to unwind the DNA and bind the Cas9 to the cut site. If the sequence complementarity is not homologous, the Cas9-sgRNA complex will release itself from the DNA and look for another potential PAM site. Conformational changes in the Cas9 will help the rest of the homology region of the guide RNA to bind the target DNA (Nishimasu et al., 2014; Sternberg et al., 2014; Szczelkun et al., 2014). The annealing of the full homology region to the target DNA will induce more conformational changes in the Cas9 to allow the nuclease activity (Cencic et al., 2014; Josephs et al., 2015).

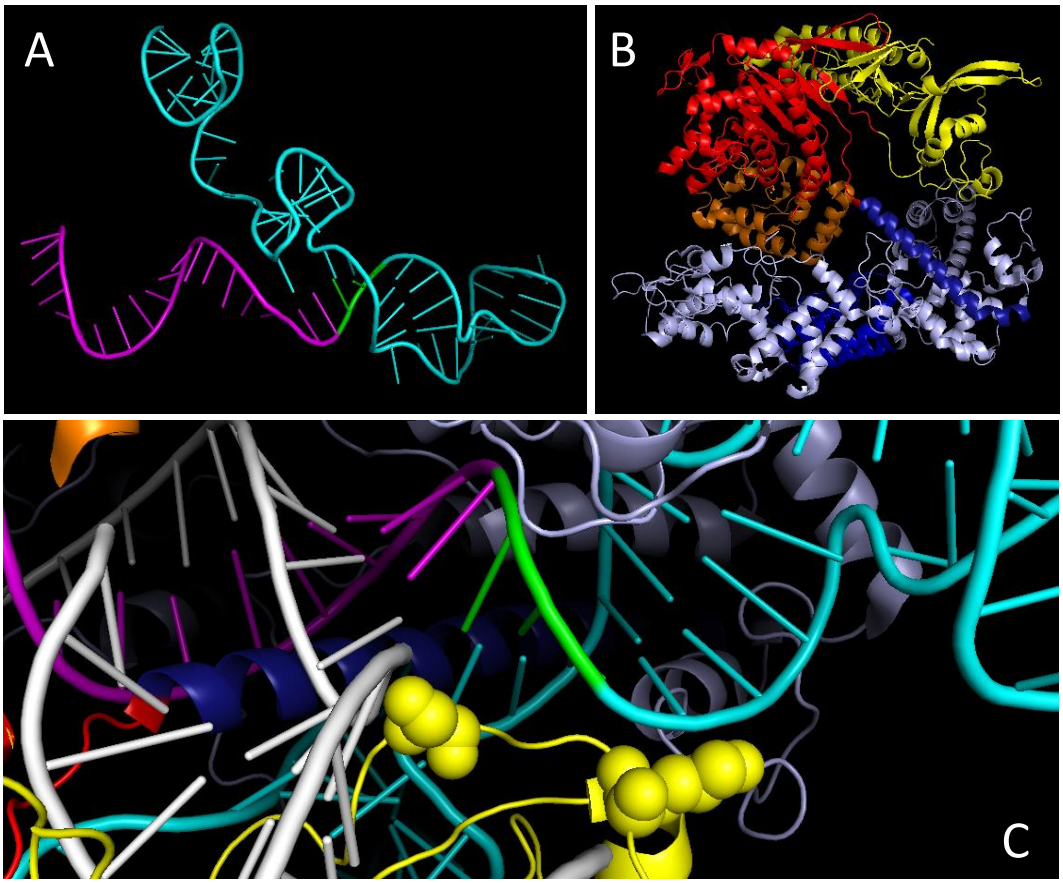


Figure 1.8. Structure of the different parts of the Cas9.
(A) gRNA structure. In magenta we find the homology region, in green the PAM site and in blue the RNA scaffold for Cas9 positioning. (B) Cas9 Structure. In blue we see the recognition domains (Dark blue for α-Bridge, light blue for REC1 and blue for REC2, in red the RuvC domain, orange the HNH domain and in yellow the PAM Interaction Domain. (C) Zoom of the Cas9 specific interaction with the PAM site. Highlighted in Spheres are the Lys1107 and Ser1109 responsible for the phosphate lock that guides this interaction. All sequences and structure were obtained from Anders et al. (2014), the domains are shown as defined in Nishimasu et al. (2014), and the images were obtained and analysed using The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

The non-target DNA strand that is displaced (R-Loop), will be the one that is read to find the PAM site, interacting with the R1333 and R1335 residues in the PAM interacting domain (Jiang et al., 2015, 2016). Following the 5'-NGG-3' recognition, the N nucleotide of the PAM will not interact with the Cas9, while the GG will be recognised and read by these arginine residues. The RNA-DNA hybrid will be placed between the REC and NUC lobes and recognised independently from the sequence by the Cas9, proving further specificity (O'Connell et al., 2014). Upstream from the PAM site, the phosphate lock loop (K1107-S1109) will separate the non-target and target strands in order to enable the formation of the RNA-DNA hybrid, rotating the non-target strand to make the target strand available for the seed sequence of the guide RNA to recognise its RNA (Figure 1.8C, Anders et al., 2014; Jiang et al., 2016; Palermo et al., 2016).

Once the Cas9 is activated by the sgRNA and in position for correct cleavage of the dsDNA, the RuvC and HNH domains will come in play. They create a blunt DSB 3 nucleotides upstream the PAM site (Gasiunas et al., 2012; Jinek et al., 2012).

The RuvC domain will be placed directly on the cut site, but the HNH domain will be placed further away from it (Figure 1.8). When the R-loop s formed and the Cas9 is in position, there will be a conformational change in the protein that will locate the HNH domain on the target strand, leaving the non-target strand caught in the R-Loop to be cleaved by the RuvC domain (Sternberg et al., 2015; Jiang et al., 2016). It is also suggested that the RuvC domain will act as an exonuclease in order to cut the non-target strand (Jinek et al., 2012). The HNH domain will become active, being able to cleave the target strand (Jiang et al., 2016; Palermo et al., 2016). After cleaving the DNA, the Cas9 will remain bound to the DNA until it is displaced by other DNA binding proteins (Sternberg et al., 2014).

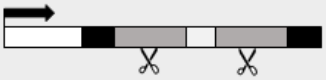
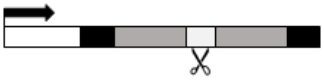
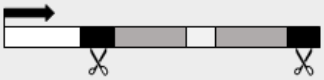
1.2.2.5. ODM.

This type of genome editing was based on directed mutagenesis of one or a few nucleotides using a homology template (Reviewed in Sauer et al., 2016a). This concept is not new, as it was first described in the 70s, and induces the modifications in DNA using a oligonucleotides (Hutchison et al., 1978). However, Oligonucleotide Directed Mutagenesis (ODM) was based on chimeraplasty, an RNA/DNA chimeric oligonucleotide, formed by DNA and 2'-O-methyl-modified RNA (Cole-Strauss et al., 1996). The chimeraplast finds a complementary sequence and the DNA repair mechanisms will identify and fix specific mismatches within that complementary sequence. The structure of a chimeraplast was based on the combination of DNA and RNA, which will fold due to homology between the lower strand (information strand) and the opposite strand. A chimeraplast is typically 68 and 88 base pairs long, requiring the addition of two loops that will stabilize the topology of the chain. The characterization and optimization of chimeraplasts designed for ODM has already been described (Metz et al., 2002).

1.2.3. CRISPR/Cas9 as a molecular biology tool: Adaptations and Modifications to the system.

The CRISPR/Cas9 system has been studied and optimized in some ways for its use as a molecular biology tool. As previously described, DSBs will induce modifications around the cut site, including insertion-deletion events, larger deletions or even rearrangements such as inversions or even translocations. However, the most common type of modification induced by CRISPR/Cas9 will be small indels in the target site, around 3 nucleotides upstream of the PAM site, where the Cas9 will induce the targeted DSBs. Depending on the location of the edit relative to a gene sequence, different outcomes can be obtained (Table 1.2). DSB formed in the coding sequence can lead to the formation of a non-synonymous mutation that could change the final protein, causing a change in gene function, introduce a premature stop codon which will lead to truncated protein expression or a frame shift causing complete or partial loss of function. Changes in the promoter sequence can provoke alterations in gene expression, increasing, reducing or even blocking all expression. One of the possible outcomes is a silent mutation that does not affect the expression or function of the gene.

Table 1.2. Different mutation sites produce different outcomes (Adapted from Krysan et al., 1999).

Region Affected	Location in the gene	Effect
Coding Sequence		Gene Disruption AA substitution
Intronic Sequence		Gene Disruption Gene Knock-Down AS Changes
Untranslated Regions		Gene Knock-Down Gene Knock-Out
Promoter Region		Gene Knock-Down Gene Knock-Out

AA: Aminoacid; AS: Alternative Splicing

All the changes previously described can be obtained by NHEJ repair after a DSB induced by genome editing leading to small indels. HR pathways can also be the basis for the introduction of DNA with these desired modifications, in addition to the targeted insertion of DNA fragments with different functionalities. As previously described, HR is based on the existence of macrohomologies between the sides of the DSB and a donor strand. This implies that providing a DNA fragment with the required homologies will be able to induce an insertion of the desired DNA (Carroll and Beumer, 2014). The length of the homology regions (Li et al., 2014; Chu et al., 2015), and the rational design of repair strands (Renaud et al., 2016) has helped the optimisation of these methods. However, it is still a rare pathway when compared to NHEJ (Frit et al., 2014). Aiming to increase the rates of HR repair, different approaches have been taken. HR has been favoured by impairing NHEJ, altering the function of key proteins such as LigIV (Forrester et al., 2011; Chu et al., 2015; Maruyama et al., 2015; Yu et al., 2015), RAD51 (Yu et al., 2011; Pinder et al., 2015) or Ku (Li et al., 2018b). At the same time, HR is restricted to the S and G2 phases of the cell cycle (Zhao et al., 2017). To circumvent this issue, timed delivery of the editing machinery and repair strands has been used to increase HR rates, inducing cellular arrest with chemicals and delivery of pre-assembled editing machinery (Lin et al., 2014; Yang et al., 2016). Modifications to the Cas9 have also been used to induce HR repair. By attaching CtIP, the protein responsible of resection prior to HR (Sartori et al., 2007) to the Cas9, it has been possible to induce HR repair at higher rates (Charpentier et al., 2018).

A classification for the use of Site Directed Nucleases (SDN) based on the type of repair and outcome obtained has also been established, using SDN-1 as indels caused by NHEJ, SDN-2 as single changes to the gene introduced via homologies, and SDN-3 as homology-directed gene insertions. This classification has been the basis of regulations of edited crops and the specifics around this classification will be explained in depth in Chapter 5.

The CRISPR/Cas9 system has been described in different organisms. These natural Cas-type endonucleases have different properties, such as alternative PAM site tolerance and different number of amino acids in their sequence. These Cas-type proteins have also been modified in order to enhance properties such as increased specificity or alterations on the PAM recognition of the targeted endonuclease (See Table 1.3).

In addition to the naturally occurring Cas9 variants and their modifications, codon optimisation has been used for ensuring correct expression in the foreign system. Codon optimisation has been described in plants and related organisms, such as cyanobacteria and algae (Campbell and Gowri, 1990; Sahoo et al., 2019). In the case of plants, the first codon-optimised Cas9 was described in Li et al. (2015a). More specifically, in *Gramineae*, following previous studies about gene composition (Wong et al., 2002), a different codon optimised Cas9 has been described (Ma et al., 2015b). However, Cas9 with different optimisations have been used for genome editing in plants, such as the non-optimised SpCas9, or human-optimised Cas9 (See Supplementary Table 1).

As previously described, Cas9 function relies on two different nuclease domains inducing breaks in each one of the strands of the genomic DNA. Targeted mutations of the Cas9 protein can lead to only one of the strands being cut, forming a nickase. The most common modifications are D10A for disrupting the RuvC domain function, and H840A for disrupting the function of the HNH domain in SpCas9 (Jinek et al., 2012). When disrupting both nuclease domains, a dead Cas9 was obtained (dCas9, Jinek et al., 2012).

In addition to specific mutations, protein engineering approaches have been used for altering specific functions of the Cas9 (Ding et al., 2016; Ribeiro et al., 2018). These have been based on the insertion of specific domains in the Cas9 sequence or the fusion of specific domains to a dCas9 that will be used to target said domain. The use of a dCas9 has allowed to use its targeting for directed modifications using other fused proteins. Examples of these are the control of gene expression using transcription activator domains (Cheng et al., 2013; Gilbert et al., 2013; Maeder et al., 2013; Perez-Pinera et al., 2013; Chavez et al., 2015), the use of light-sensitive expression inductors (Nihongaki et al., 2015; Polstein and Gersbach, 2015; Gao et al., 2016; Liu et al., 2016) and chemical inductors (Gao et al., 2016). The function of these domains would be to gather the transcription machinery in order to induce the expression of the gene in certain conditions. In addition to dCas9 fusion, transcription activation components have been added to the sgRNA, forming a three-component system (Konermann et al., 2015; Zalatan et al., 2015).

Other domains inducing modifications in the target DNA have also been attached to dCas9. By attaching cytidine (Komor et al., 2016; Rees et al., 2017) or adenine deaminase (Gaudelli et al., 2017) to a dCas9, base editors have been obtained. These induce C-T and A-G transitions respectively. To increase the versatility of the system, alternative dCas9 such as the one from *Staphylococcus aureus* have also been used to target base editors (Kim et al., 2017b). Epigenetic modifications have also been targeted using dCas9. Examples of these have been the use of cytosine DNA methyltransferase (Amabile et al., 2016; McDonald et al., 2016; Vojta et al., 2016; Stepper et al., 2017; Xiong et al., 2017), targeting CpG methylation, Lysine-specific histone demethylase (Hilton et al., 2015; Kearns et al., 2015) or histone methyltransferases (Amabile et al., 2016). In addition to targeted modifications, genome imaging has been possible by attaching fluorescent proteins to the dCas9 (Chen et al., 2013; Anton et al., 2014; Ma et al., 2015a).

In order to improve Cas9 targeting specificity, ZF or TALE domains have been fused to a Cas9 nuclease with a mutated PAM recognition site, significantly reducing off-target effects (Bolukbasi et al., 2015). Following the same principle as ZFNs and TALENs, two different dCas9 have been fused to FokI nucleases, obtaining a dimerization-dependent nuclease targeted by two sgRNA (Guilinger et al., 2014; Tsai et al., 2014), which implied a reduction in off target effects. A different approach was to reduce the amount of time the Cas9 was active. This was achieved by fusing the Cas9 to proteins targeting proteasomal degradation in absence of specific stabilizing proteins (Gautam et al., 2012; Maji et al., 2017). A different method for controlling

Cas9 activity has been the insertion of the intein 37R3-2, which would only allow Cas9 activity when the chemical 4-HT is present (Peck et al., 2011; Davis et al., 2015). A similar approach was used with the darkness-dependent dimerizing domains known as pdDronpa (Ando et al., 2004), which would impair the Cas9 function in presence of 500nm light (Zhou et al., 2017). A different approach to induce directed changes in a sequence was described in Anzalone et al. (2019), fusing a nickase with a reverse transcriptase and providing a modified sgRNA containing a primer binding site, a sequence to introduce a target-recognition sequence and using HR pathways to induce the insertion of the desired sequence.

In addition to different modifications of the Cas9 nuclease to increase the plasticity of the system or to target other protein functions, the sgRNA has also been modified, focusing on its optimal length and structure, focusing on the reduction of off-target effects and increasing on-target efficiency. Based on these optimisations, several tools have been developed for rational design of the sgRNA (See Chapter 4). At the same time, multiplex editing has allowed the modification of several targets at once by the use of more than one sgRNA at a time (Cong et al., 2013). The editing machinery can be delivered in DNA form, requiring the forementioned expression systems which may require integration of the DNA, or be delivered as a pre-assembled ribonucleoprotein complex (Kim et al., 2014). This approach is able to use a DNA free system that has been observed to reduce off-target effects caused by prolonged expression of the editing machinery (Wang et al., 2014; Lawrenson et al., 2015; Shan et al., 2015)

Besides the modifications to the system itself, regarding the molecular components necessary for CRISPR/Cas9 based editing, species specific optimizations are required, such as the isolation of a specific tissue that can accept the editing machinery and the method of delivery of said editing machinery. These factors will be reviewed in Chapter 2.

Table 1.3. Different available Cas9 natural variants and modified versions (Adapted from Ding et al., 2016; Pickar-Oliver & Gersbach, 2019).

Natural Species Cas9	Variant	Modifications	PAM	Features	References
Streptococcus pyogenes (SpCas9)			NGG	1,368 amino acids	Cong et al., 2013; Hsu et al., 2013; Hwang et al., 2013; Ran et al., 2013
		D1135E	NGG	Reduced NAG binding	Kleinstiver et al., 2015b
	37R3-2	37R3-2 intein inserted into Cas9	NGG	Increased specificity	Davis et al., 2015
		N497A-R661A-Q695A-Q926A	NGG	Increased specificity	Kleinstiver et al., 2016
	VRER	D1135V, G1218R, R1335E, T1337R	NGCG		Kleinstiver et al., 2015b
	EQR	D1135E, R1335Q, T1337R	NGAG		Kleinstiver et al., 2015b
	VQR	D1135V, R1335Q, T1337R	NGAN or NGNG		Kleinstiver et al., 2015b
	Cas9-HF1	N497A/R661A/Q695A/Q926A	NGG	Increased specificity	Kleinstiver et al., 2016
	eSpCas9 (1.0 and 1.1)	K810A/K1003A/R1060A	NGG	Increased specificity	Slaymaker et al., 2016
	xCas9-3.7c	A262T, R324L, S409I, E480K, E543D, M694I, E1219V	NG, GAA, GAT		Hu et al., 2018a
	HypaCas9f	N692A, M694A, Q695A, H698A	NGG	Increased specificity	Chen et al., 2017
	evoCas9g	M495V, Y515N, K526E, R661Q	NGG	Increased specificity	Casini et al., 2018
	HiFi Cas9e	R691A	NGG	Increased specificity (RNP delivery)	Vakulskas et al., 2018
Staphylococcus aureus (SaCas9)			NNGRRT or NNGRR(N)	1,053 amino acids	Kleinstiver et al., 2015b; Ran et al., 2015
			NAG or NGA or NNGGGT		Kleinstiver et al., 2015b; Steinert et al., 2015
	KKH		NNNRRT		Kleinstiver et al., 2015a
Francisella novicida (FnCas9)			YG		Hirano et al., 2016
	Cpf1		TTN		Ran et al., 2015
Streptococcus thermophiles (StCas9)			NNAGAAW	1,121 amino acids	Deveau et al., 2008; Cong et al., 2013
			NNAGAA or NNGGAA		Kleinstiver et al., 2015b; Steinert et al., 2015
Neisseria meningitides (NmCas9)			NNNNGATT	1,082 amino acids	Hou et al., 2013; Lee et al., 2016
Streptococcus canis (ScCas9)			NNG	1,375 amino acids	Chatterjee et al., 2018
Campylobacter jejuni (CjCas9)			NNNVRYM	984 amino acids	Yamada et al., 2017
Deltaproteobacteria and Planctomycetes (CasX)			TTCN	980 amino acids	Burstein et al., 2017

1.2.4. Examples of edited crops.

Genome editing has been described in many species using a wide variety of tissues and expression systems (Table 1.4, Martínez-Fortún et al., 2017). CRISPR/Cas9 in particular has been used in many species using different tissues and expression systems (see Supplementary Tables 1 and 2 and Chapter 2). Notably, one of the major hindrances to the CRISPR/Cas9 system, the possibility of observing off-target effects, has been studied and found to be lower than in humans (Peterson et al., 2016; Wolt et al., 2016; Peng et al., 2017; Hajiahmadi et al., 2019). However, they should be considered as a possibility for all experiments regarding genome editing. Given its common use, many of the improvements previously described have been adapted to plants, such as multiplex editing (Hashimoto et al., 2018; Zsögön et al., 2018; Armario Najera et al., 2019; Abdelrahman et al., 2021; Stuttman et al., 2021), HR-based insertion (Endo et al., 2006; Qi et al., 2013b; Hirohata et al., 2019; Endo et al., 2021; Nishizawa-Yokoi et al., 2020; Peng et al., 2020), DNA free editing (Svitashev et al., 2015; Woo et al., 2015; Zhang et al., 2016; Liang et al., 2017; Banakar et al., 2020; Kim et al., 2020b), epigenetic (Gallego-Bartolomé et al., 2018) and expression control (Qi et al., 2013a; Rodríguez-Leal et al., 2017; Zhang et al., 2018a; Pan et al., 2021), dCas9-mediated targeting (Lowder et al., 2015; Piatek et al., 2015) and base editing (Hess et al., 2017; Zhang and Gao, 2017; Hua et al., 2018; Kang et al., 2018; Li et al., 2018a; Tian et al., 2018; Yan et al., 2018; Mishra et al., 2020). The CRISPR/Cfp1 (Cas12a) system has also been adapted for its use in plants (Endo et al., 2016; Hu et al., 2017; Kim et al., 2017a; Tang et al., 2017; Xu et al., 2017; Yin et al., 2017; Schindele & Puchta, 2020), together with HR-mediated insertion (Begemann et al., 2017; Li et al., 2018c, 2018d) or expression control (Tang et al., 2017). This has led to improvement of many different traits in plants, such as yield (Miao et al., 2013; Li et al., 2016b; Zhang et al., 2018b; Miao et al., 2019), nutritional profile (Andersson et al., 2017; Jiang et al., 2017; Sun et al., 2017), abiotic (Shi et al., 2017; Liao et al., 2019; Zhang et al., 2019a; Joshi et al., 2020) and biotic stress resistance (Zhou et al., 2015; Chandrasekaran et al., 2016; Jia et al., 2016, 2017; Wang et al., 2016; Nekrasov et al., 2017; Peng et al., 2017; Zhang et al., 2017; Ortigosa et al., 2019; Zhang et al., 2020; Miao et al., 2019), fruit browning (Waltz, 2016), plant domestication (Klap et al., 2017; Soyk et al., 2017; Li et al., 2018f) and herbicide resistance (Li et al., 2015b, 2016a; Svitashev et al., 2015; Butler et al., 2016; Sun et al., 2016; Dong et al., 2020; Wang et al., 2019; Wu et al., 2020).

All the factors previously described point to the potential of CRISPR/Cas9 and other genome editing techniques to change crop development in a significant manner (Martínez-Fortún et al., 2017). Chapter 5 will focus on the comparison of modifications obtainable by conventional breeding techniques and the inclusion of edited crops in the market given the current legislation.

Table 1.4. Examples of genome editing in each of the major crops classified according to the UN ICC (United Nations Indicative Crop Classification) system (Adapted from Martínez-Fortún et al., 2017).

Crop Classification	Crop	Gene Editor	Target Sequence	Heritable edits observed	Reference
Cereals	Wheat (durum)	CRISPR/Cas9	*GASR7*, *DEP1*, *NAC2*, *PIN1* and *LOX2*	T2 generation	Zhang et al., 2016
	Wheat (bread)	CRISPR/Cas9	*MLO*	T2 generation	Wang et al., 2014
	Wheat (bread)	CRISPR/Cas9	*GASR7*	T0 generation only	Liang et al., 2017
	Maize	CRISPR/Cas9	*LIG1*, *Ms26*/*Ms45*, and *ALS1*/*ALS2*	T2 generation	Svitashev et al., 2015
		TALEN	*gl2*	T1 generation	Char et al., 2015
		ZFNs	*IPK1*	T2 generation	Shukla et al., 2009
		CRISPR/Cas9 and TALEN	*PDS*, *IPK1A*, *IPK* and *MRP4*	Protoplast editing (T0)	Liang et al., 2014
	Rice	CRISPR/Cas9	*SWEET11*	T2 generation	Ma et al., 2017
		TALEN	*SWEET14*	T2 generation	Li et al., 2012b
	Sorghum	CRISPR/Cas9	GFP	Transient expression	Jiang et al., 2013
	Barley	CRISPR/Cas9	*PM19*	T2 generation	Lawrenson et al., 2015
	Barley	TALEN	*HvPAPhy_a*	T0 generation	Wendt et al., 2013
Vegetables & melons	Cucumber	CRISPR/Cas9	*eIF4e*	T3 generation	Chandrasekaran et al., 2016.
	Tomato	CRISPR/Cas9	*AGO7*	T1 generation (Low fertility)	Brooks et al., 2014
	Watermelon	CRISPR/Cas9	*PDS*	Protoplast editing/ T0 generation	Tian et al., 2017
Fruits & nuts	Grapefruit	CRISPR/Cas9	*LOB1*	T0 generation	Jia et al., 2016
	Oranges	CRISPR/Cas9	*PDS*	T0 generation	Jia and Nian, 2014
	Grape	CRISPR/Cas9	*IdnDH*	Callous transformation (T0)	Ren et al., 2016
	Apple	CRISPR/Cas9	*PDS*	T0 generation	Nishitani et al., 2016
	Kiwi	CRISPR/Cas9	*PDS*	T0 generation	Wang et al., 2018a
	Strawberry	CRISPR/Cas9	*PDS*	T0 generation	Wilson et al., 2019
Oilseed crops	Soybean	CRISPR/Cas9 and TALEN	*PDS11*/*PDS18*	T0 generation (Lethal mutation)	Du et al., 2016
	Flax	ODM with CRISPR/Cas9 and TALEN	*EPSPS*	T1 generation	Sauer et al., 2016b
	Rapeseed	CRISPR/Cas9	*BolC.GA4*	T2 generation	Lawrenson et al., 2015
Root/tuber crops	Potato	CRISPR/Cas9	*GBSS*	Protoplast editing (T0)	Andersson et al., 2017
	Potato	TALEN	*ALS*	Protoplast editing (T0)	Nicolia et al., 2015
	Cassava	CRISPR/Cas9	*PDS*	T0 generation	Odipio et al., 2017
Beverage and spice crops	Coffee	CRISPR/Cas9	*PDS*	T0 generation	Breitler et al., 2018
	Cocoa	CRISPR/Cas9	*NPR3*	T0 generation	Fister et al., 2018
Leguminous crops	Medicago truncatula	CRISPR/Cas9	*PDS*	T0 generation	Meng et al., 2017, 2020
	Cowpea	CRISPR/Cas9	*SYMRK*	T0 generation	Ji et al., 2019
Sugar crops	Sugarcane	CRISPR/Cas9	*ALS*	T0 generation	Oz et al. 2021
Other crops	Cotton	CRISPR/Cas9	*MYB25*	T0 generation	Wang et al., 2018
	Populus	CRISPR/Cas9	*PDS*	T0 generation	Fan et al., 2015

1.3. Project Aims.

1. Find a platform for quick validation of promoters for CRISPR/Cas9 based editing in *Brachypodium*, selecting a tissue and method of transformation.
2. Create plasmids for the validation of the TaU6 promoter as a potential candidate for sgRNA expression for genome editing in *Brachypodium* using CRISPR/Cas9, using BdU6 and OsU3 as positive controls.
3. Insert sgRNA targeting endogenous (*BdDWF4*) and heterologous (HygR) genes in those plasmids.
4. Validate the use of the TaU6 promoter for the editing of endogenous and heterologous targets in *Brachypodium*.
5. Analyse the ability of already described models to predict editing efficiency in plants.
6. Create a simple model for the prediction of sgRNA efficiency in plants depending on the target sequence.
7. Review the range of genetic variation possible in conventional breeding.
8. Assess how these genetic changes could be obtained using CRISPR/Cas9.

Generally, the project is based on the use of CRISPR/Cas9 in the model plant *Brachypodium distachyon*. More specifically, a platform will be developed for the validation of sgRNA functionality, investigating different tissues and the possibility of the use of different transformation methods. This method will be tested with different endogenous and heterologous targets. On the other hand, determinants on the functionality of sgRNA, such as the promoter used for its expression and the target sequence will be studied. Finally, to put the different uses of CRISPR/Cas9 in the context of breeding techniques, the different outcomes obtained from conventional breeding processes will be reviewed.

1.4. References.

Abdelrahman, M., Wei, Z., Rohila, J. S., & Zhao, K. (2021). Multiplex genome-editing technologies for revolutionizing plant biology and crop improvement. *Frontiers in Plant Science*, 12, 721203.

Al-Dous, E. K., George, B., Al-Mahmoud, M. E., Al-Jaber, M. Y., Wang, H., Salameh, Y. M., Al-Azwani, E. K., Chaluvadi, S., Pontaroli, A. C., Debarry, J., Arondel, V., Ohlrogge, J., Saie, I. J., Suliman-Elmeer, K. M., Bennetzen, J. L., Kruegger, R. R., and Malek, J. A. (2011). *De novo* genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nature Biotechnology* 29, 521–527.

Alexandratos, N., and Bruinsma, J. (2012). World agriculture towards 2030/2050: the 2012 revision. *ESA Working paper* 12, 146.

Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P., et al. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301, 653–657.

Alves, S. C., Worland, B., Thole, V., Snape, J. W., Bevan, M. W., and Vain, P. (2009). A protocol for *Agrobacterium*-mediated transformation of *Brachypodium distachyon* community standard line Bd21. *Nature Protocols* 4, 638–649.

Amabile, A., Migliara, A., Capasso, P., Biffi, M., Cittaro, D., Naldini, L., and Lombardo, A. (2016). inheritable silencing of endogenous genes by hit-and-run targeted epigenetic editing. *Cell* 167, 219-232.e14.

An, D., Li, C., Zhou, Y., Wu, Y., and Wang, W. (2018). Genomes and transcriptomes of duckweeds. *Frontiers in Chemistry* 6, 230.

Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 513, 569–573.

Andersson, M., Turesson, H., Nicolia, A., Fält, A. S., Samuelsson, M., and Hofvander, P. (2017). Efficient targeted multiallelic mutagenesis in tetraploid potato (*Solanum tuberosum*) by transient CRISPR/Cas9 expression in protoplasts. *Plant Cell Reports* 36, 117–128.

Ando, R., Mizuno, H., and Miyawaki, A. (2004). Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting. *Science* 306, 1370–1373.

Anton, T., Bultmann, S., Leonhardt, H., and Markaki, Y. (2014). Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system. *Nucleus* 5, 163–172.

Anur, R. M., Mufithah, N., Sawitri, W. D., Sakakibara, H., and Sugiharto, B. (2020). Overexpression of sucrose phosphate synthase enhanced sucrose content and biomass production in transgenic sugarcane. *Plants* 9, 200.

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., and Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157.

Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361, eaar7191.

Araújo, P., Cesarino, I., Mayer, J. L. S., Ferrari, I. F., Kiyota, E., Sawaya, A. C. H. F., Leme, A. F. P., and Mazzafera, P. (2014). A model system to study the lignification process in *Eucalyptus globulus*. *Physiologia Plantarum* 152, 17–31.

Armario Najera, V., Twyman, R. M., Christou, P., and Zhu, C. (2019). Applications of multiplex genome editing in higher plants. *Current Opinion in Biotechnology* 59, 93–102.

Atallah, N. M., and Banks, J. A. (2015). Reproduction and the pheromonal regulation of sex type in fern gametophytes. *Frontiers in Plant Science* 6, 100.

Audebert, M., Salles, B., and Calsou, P. (2004). Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. *Journal of Biological Chemistry* 279, 55117–55126.

Audebert, M., Salles, B., Weinfeld, M., and Calsou, P. (2006). Involvement of polynucleotide kinase in a poly(ADP-ribose) polymerase-1-dependent DNA double-strand breaks rejoining pathway. *Journal of Molecular Biology* 356, 257–265.

Bablak, P., Draper, J., Davey, M. R., and Lynch, P. T. (1995). Plant regeneration and micropropagation of *Brachypodium distachyon*. *Plant Cell, Tissue and Organ Culture* 42, 97–107.

Banakar, R., Schubert, M., Collingwood, M., Vakulskas, C., Eggenberger, A. L., and Wang, K. (2020). Comparison of CRISPR-Cas9/Cas12a ribonucleoprotein complexes for genome editing efficiency in the rice phytoene desaturase (*OsPDS*) Gene. *Rice* 13, 4.

Banks, J. A., Nishiyama, T., Hasebe, M., Bowman, J. L., Gribskov, M., DePamphilis, C., et al. (2011). The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science* 332, 960–963.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712.

Begemann, M. B., Gray, B. N., January, E. E., Gordon, G. C., He, Y., Liu, H., Wu, X., Brutnell, T. P., Mockler, T. C., and Oufattole, M. (2017). Precise insertion and guided editing of higher plant genomes using Cpf1 CRISPR nucleases. *Scientific Reports* 7, 1–6.

Belfort, M., and Perlman, P. S. (1995). Mechanisms of intron mobility. *Journal of Biological Chemistry* 270, 30237–30240.

Belfort, M., and Roberts, R. J. (1997). Homing endonucleases: Keeping the house in order. *Nucleic Acids Research* 25, 3379–3388.

Betekhtin, A., Hus, K., Rojek-Jelonek, M., Kurczynska, E., Nibau, C., Doonan, J. H., and Hasterok, R. (2020). In vitro tissue culture in *Brachypodium*: Applications and challenges. *International Journal of Molecular Sciences* 21, 1037.

Birol, I., Raymond, A., Jackman, S. D., Pleasance, S., Coope, R., Taylor, G. A., et al. (2013). Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. *Bioinformatics* 29, 1492–1497.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326, 1509–1512.

Bolukbasi, M. F., Gupta, A., Oikemus, S., Derr, A. G., Garber, M., Brodsky, M. H., Zhu, L. J., and Wolfe, S. A. (2015). DNA-binding-domain fusions enhance the targeting range and precision of Cas9. *Nature Methods* 12, 1150–1156.

Borrill, P. (2020). Blurring the boundaries between cereal crops and model plants. *New Phytologist* 228, 1721–1727.

Bortesi, L., and Fischer, R. (2015). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnology Advances* 33, 41–52.

Bortiri, E., Coleman-Derr, D., Lazo, G. R., Anderson, O. D., and Gu, Y. Q. (2008). The complete chloroplast genome sequence of *Brachypodium distachyon*: sequence comparison and phylogenetic analysis of eight grass plastomes. *BMC research notes* 1, 61.

Bragg, J. N., Wu, J., Gordon, S. P., Guttman, M. E., Thilmony, R., Lazo, G. R., Gu, Y. Q., and Vogel, J. P. (2012). Generation and characterization of the Western Regional Research Center *Brachypodium* T-DNA insertional mutant collection. *PLoS ONE* 7, e41916.

Breitler, J. C., Dechamp, E., Campa, C., Zebral Rodrigues, L. A., Guyot, R., Marraccini, P., & Etienne, H. (2018). CRISPR/Cas9-mediated efficient targeted mutagenesis has the potential to accelerate the domestication of *Coffea canephora*. *Plant Cell, Tissue and Organ Culture* 134, 383–394.

Brooks, C., Nekrasov, V., Lipppman, Z. B., and Van Eck, J. (2014). Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. *Plant Physiology* 166, 1292–1297.

Brouwer, P., Bräutigam, A., Külahoglu, C., Tazelaar, A. O. E., Kurz, S., Nierop, K. G. J., van der Werf, A., Weber, A. P. M., and Schluepmann, H. (2014). *Azolla* domestication towards a biobased economy? *New Phytologist* 202, 1069–1082.

Brown, R. S., Sander, C., and Argos, P. (1985). The primary structure of transcription factor TFIIIA has 12 consecutive repeats. *FEBS Letters* 186, 271–274.

Brutnell, T. P. (2015). Model grasses hold key to crop improvement. *Nature Plants* 1, 15062.

Brutnell, T. P., Wang, L., Swartwood, K., Goldschmidt, A., Jackson, D., Zhu, X. G., Kellogg, E., and van Eck, J. (2010). *Setaria viridis*: A model for C4 photosynthesis. *Plant Cell* 22, 2537–2544.

Bugbee, B., and Koerner, G. (1997). Yield comparisons and unique characteristics of the dwarf wheat cultivar "USU-Apogee." *Advances in Space Research* 20, 1891–1894.

Burstein, D., Harrington, L. B., Strutt, S. C., Probst, A. J., Anantharaman, K., Thomas, B. C., Doudna, J. A., and Banfield, J. F. (2017). New CRISPR/Cas systems from uncultivated microbes. *Nature* 542, 237–241.

Butler, N. M., Baltes, N. J., Voytas, D. F., and Douches, D. S. (2016). Geminivirus-mediated genome editing in potato (*Solanum tuberosum* L.) using sequence-specific nucleases. *Frontiers in Plant Science* 7, 1045.

Campbell, W. H., and Gowri, G. (1990). Codon usage in higher plants, green algae, and cyanobacteria. *Plant Physiology* 92, 1–11.

Carroll, D., and Beumer, K. J. (2014). Genome engineering with TALENs and ZFNs: Repair pathways and donor design. *Methods* 69, 137–141.

Casini, A., Olivieri, M., Petris, G., Montagna, C., Reginato, G., Maule, G., Lorenzin, F., Prandi, D., Romanel, A., Demichelis, F., Inga, A., and Cereseto, A. (2018). A highly specific SpCas9 variant is identified by *in vivo* screening in yeast. *Nature Biotechnology* 36, 265–271.

Ceccaldi, R., Rondinelli, B., and D'Andrea, A. D. (2016). Repair pathway choices and consequences at the double-strand break. *Trends in Cell Biology* 26, 52–64.

Cencic, R., Miura, H., Malina, A., Robert, F., Ethier, S., Schmeing, T. M., Dostie, J., and Pelletier, J. (2014). Protospacer adjacent motif (PAM)-distal sequences engage CRISPR/Cas9 DNA target cleavage. *PLoS ONE* 9, e109213.

Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J. A., Somia, N. V., Bogdanove, A. J., and Voytas, D. F. (2011). Efficient design and assembly of custom TALEN and

other TAL effector-based constructs for DNA targeting. *Nucleic Acids Research* 39.

Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., Pearlsman, M., Sherman, A., Arazi, T., and Gal-On, A. (2016). Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Molecular plant pathology* 17, 1140–53.

Chang, C., Bowman, J. L., and Meyerowitz, E. M. (2016). Field Guide to Plant Model Systems. *Cell* 167, 325–339.

Char, S. N., Unger-Wallace, E., Frame, B., Briggs, S. A., Main, M., Spalding, M. H., Vollbrecht, E., Wang, K., and Yang, B. (2015). Heritable site-specific mutagenesis using TALENs in maize. *Plant Biotechnology Journal* 13, 1002–1010.

Charpentier, M., Khedher, A. H. Y., Menoret, S., Brion, A., Lamribet, K., Dardillac, E., Boix, C., Perrouault, L., Tesson, L., Geny, S., De Cian, A., Itier, J. M., Anegon, I., Lopez, B., Giovannangeli, C., and Concordet, J. P. (2018). CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair. *Nature Communications* 9, 1133.

Chatterjee, P., Jakimo, N., and Jacobson, J. M. (2018). Minimal PAM specificity of a highly similar SpCas9 ortholog. *Science Advances* 4, eaau0766.

Chavez, A., Scheiman, J., Vora, S., Pruitt, B. W., Tuttle, M., P R Iyer, E., et al. (2015). Highly efficient Cas9-mediated transcriptional programming. *Nature Methods* 12, 326–328.

Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G. W., Park, J., Blackburn, E. H., Weissman, J. S., Qi, L. S., and Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155, 1479–1491.

Chen, H., Choi, J., and Bailey, S. (2014). Cut site selection by the two nuclease domains of the Cas9 RNA-guided endonuclease. *Journal of Biological Chemistry* 289, 13284–13294.

Chen, J. S., Dagdas, Y. S., Kleinstiver, B. P., Welch, M. M., Sousa, A. A., Harrington, L. B., Sternberg, S. H., Joung, J. K., Yildiz, A., and Doudna, J. A. (2017). Enhanced proofreading governs CRISPR/Cas9 targeting accuracy. *Nature* 550, 407–410.

Chen, Z., Wen, F., Sun, N., and Zhao, H. (2009). Directed evolution of homing endonuclease I-SceI with altered sequence specificity. *Protein Engineering, Design and Selection* 22, 249–256.

Cheng, A. W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T. W., Rangarajan, S., Shivalila, C. S., Dadon, D. B., and Jaenisch, R. (2013). Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Research* 23, 1163–1171.

Chevalier, B. S., Kortemme, T., Chadsey, M. S., Baker, D., Monnat, R. J., and Stoddard, B. L. (2002). Design, activity, and structure of a highly specific artificial endonuclease. *Molecular Cell* 10, 895–905.

Chevalier, B. S., and Stoddard, B. L. (2001). Homing endonucleases: Structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Research* 29, 3757–3774.

Choulika, A., Perrin, A., Dujon, B., and Nicolas, J. F. (1994). The yeast I-Sce I meganuclease induces site-directed chromosomal recombination in mammalian cells. *Comptes Rendus de l'Academie des Sciences - Serie III* 317, 1013–1019.

Choulika, A., Perrin, A., Dujon, B., and Nicolas, J. F. (1995). Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* 15, 1968–1973.

Chu, V. T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., and Kühn, R. (2015). Increasing

the efficiency of homology-directed repair for CRISPR/Cas9-induced precise gene editing in mammalian cells. *Nature Biotechnology* 33, 543–548.

Clough, S. J., and Bent, A. F. (1998). Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant Journal* 16, 735–743.

Coen, E. S., and Meyerowitz, E. M. (1991). The war of the whorls: Genetic interactions controlling flower development. *Nature* 353, 31–37.

Cole-Strauss, A., Yoon, K., Xiang, Y., Byrne, B. C., Rice, M. C., Gryn, J., Holloman, W. K., and Kmiec, E. B. (1996). Correction of the mutation responsible for sickle cell anemia by an RNA-DNA oligonucleotide. *Science* 273, 1386–1389.

Cong, L. Le, Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 1766–1769.

Coomey, J. H., Sibout, R., and Hazen, S. P. (2020). Grass secondary cell walls, *Brachypodium distachyon* as a model for discovery. *New Phytologist* 227, 1649–1667.

Cove, D. J., Perroud, P. F., Charron, A. J., McDaniel, S. F., Khandelwal, A., and Quatrano, R. S. (2009). The moss *Physcomitrella patens*: A novel model system for plant development and genomic studies. *Cold Spring Harbor Protocols* 4, pdb.emo115.

D'Ambrosio, C., Stigliani, A.L., and Giorio, G. (2018). CRISPR/Cas9 editing of carotenoid genes in tomato. *Transgenic Research* 27, 367–378.

D'Hont, A., Denoeud, F., Aury, J. M., Baurens, F. C., Carreel, F., Garsmeur, O., et al. (2012). The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 488, 213–217.

Daley, J. M., and Wilson, T. E. (2005). Rejoining of DNA double-strand breaks as a function of overhang length. *Molecular and Cellular Biology* 25, 896–906.

Dalgaard, J. Z., Klar, A. J., Moser, M. J., Holley, W. R., Chatterjee, A., and Saira Mian, I. (1997). Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family. *Nucleic Acids Research* 25, 4626–4638.

Dalmais, M., Antelme, S., Ho-Yue-Kuang, S., Wang, Y., Darracq, O., D'Yvoire, M. B., Cézard, L., Légée, F., Blondet, E., Oria, N., Troadec, C., Brunaud, V., Jouanin, L., Höfte, H., Bendahmane, A., Lapierre, C., and Sibout, R. (2013). A TILLING platform for functional genomics in *Brachypodium distachyon*. *PLoS ONE* 8, e65503.

Davis, K. M., Pattanayak, V., Thompson, D. B., Zuris, J. A., and Liu, D. R. (2015). Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nature Chemical Biology* 11, 316–318.

Della-Maria, J., Zhou, Y., Tsai, M. S., Kuhnlein, J., Carney, J. P., Paull, T. T., and Tomkinson, A. E. (2011). Human Mre11/human Rad50/Nbs1 and DNA ligase IIIα/XRCC1 protein complexes act together in an alternative nonhomologous end joining pathway. *Journal of Biological Chemistry* 286, 33845–33853.

Deveau, H., Barrangou, R., Garneau, J. E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D. A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *Journal of Bacteriology* 190, 1390–1400.

Difilippantonio, M. J., Zhu, J., Chen, H. T., Meffre, E., Nussenzweig, M. C., Max, E. E., Ried, T., and

Nussenzweig, A. (2000). DNA repair protein Ku80 suppresses chromosomal aberrations and malignant transformation. *Nature* 404, 510–514.

Ding, Y., Li, H., Chen, L.-L. L., and Xie, K. (2016). Recent Advances in Genome Editing Using CRISPR/Cas9. *Frontiers in Plant Science* 7, 703.

Dong, H., Wang, D., Bai, Z., Yuan, Y., Yang, W., Zhang, Y., Ni, H., and Jiang, L. (2020). Generation of imidazolinone herbicide resistant trait in Arabidopsis. *PLoS ONE* 15, e0233503.

Doudna, J. A., and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR/Cas9. *Science* 346, 1258096.

Doyon, J. B., Pattanayak, V., Meyer, C. B., and Liu, D. R. (2006). Directed evolution and substrate specificity profile of homing endonuclease I-SceI. *Journal of the American Chemical Society* 128, 2477–2484.

Draper, J., Mur, L. A. J., Jenkins, G., Ghosh-Biswas, G. C., Bablak, P., Hasterok, R., and Routledge, A. P. M. (2001). *Brachypodium distachyon*. A new model system for functional genomics in grasses. *Plant Physiology* 127, 1539–1555.

Du, H., Zeng, X., Zhao, M., Cui, X., Wang, Q., Yang, H., Cheng, H., and Yu, D. (2016). Efficient targeted mutagenesis in soybean by TALENs and CRISPR/Cas9. *Journal of Biotechnology* 217, 90–97.

Duan, X., Gimble, F. S., and Quiocho, F. A. (1997). Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity. *Cell* 89, 555–564.

Dujon, B., Beifort, M., Butow, R. A., Jacq, C., Lemieux, C., Perlman, P. S., and Vogt, V. M. (1989). Mobile introns: definition of terms and recommended nomenclature. *Gene* 82, 115–118.

Eichten, S. R., Stuart, T., Srivastava, A., Lister, R., and Borevitz, J. O. (2016). DNA methylation profiles of diverse *Brachypodium distachyon* align with underlying genetic diversity. *Genome Research* 26, 1520–1531.

Ellenberger, T., and Tomkinson, A. E. (2008). Eukaryotic DNA ligases: Structural and functional insights. *Annual Review of Biochemistry* 77, 313–338.

Endo, M., Iwakami, S. and Toki, S. (2021). Precision genome editing in plants via gene targeting and subsequent break-induced single-strand annealing. *Plant Biotechnology Journal* 19, 563-574.

Endo, A., Masafumi, M., Kaya, H., and Toki, S. (2016). Efficient targeted mutagenesis of rice and tobacco genomes using Cpf1 from *Francisella novicida*. *Scientific Reports* 6, 38169.

Endo, M., Ishikawa, Y., Osakabe, K., Nakayama, S., Kaya, H., Araki, T., Shibahara, K. I., Abe, K., Ichikawa, H., Valentine, L., Hohn, B., and Toki, S. (2006). Increased frequency of homologous recombination and T-DNA integration in *Arabidopsis* CAF-1 mutants. *EMBO Journal* 25, 5579–5590.

Epinat, J. C., Amould, S., Chames, P., Rochaix, P., Desfontaines, D., Puzin, C., Patin, A., Zanghellini, A., Pâques, F., and Lacroix, E. (2003). A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Research* 31, 2952–2962.

Fairall, L., Schwabe, J. W. R., Chapman, L., Finch, J. T., and Rhodes, D. (1993). The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature* 366, 483–487.

Falter, C. and Voigt, C.A. (2016). Improving biomass production and saccharification in *Brachypodium distachyon* through overexpression of a sucrose-phosphate synthase from sugarcane. *Journal of Plant Biochemistry and Biotechnology* 25, 311–318

Fan, D., Liu, T., Li, C., Jiao, B., Li, S., Hou, Y., & Luo, K. (2015). Efficient CRISPR/Cas9-mediated targeted mutagenesis in *Populus* in the first generation. *Scientific Reports*, 5, 12217.

Fister, A. S., Landherr, L., Maximova, S. N., & Guiltinan, M. J. (2018). Transient expression of CRISPR/Cas9 machinery targeting TcNPR3 enhances defense response in *Theobroma cacao*. *Frontiers in Plant Science* 9, 268.

Fitzgerald, T. L., Powell, J. J., Schneebeli, K., Hsia, M. M., Gardiner, D. M., Bragg, J. N., McIntyre, C. L., Manners, J. M., Ayliffe, M., Watt, M., Vogel, J. P., Henry, R. J., and Kazan, K. (2015). *Brachypodium* as an emerging model for cereal-pathogen interactions. *Annals of Botany* 115, 717–731.

Flick, K. E., Jurica, M. S., Monnat, R. J., and Stoddard, B. L. (1998). DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-Ppol. *Nature* 394, 96–101.

Forrester, N. A., Sedgwick, G. G., Thomas, A., Blackford, A. N., Speiseder, T., Dobner, T., Byrd, P. J., Stewart, G. S., Turnell, A. S., and Grand, R. J. A. (2011). Serotype-specific inactivation of the cellular DNA damage response during adenovirus infection. *Journal of Virology* 85, 2201–2211.

Frit, P., Barboule, N., Yuan, Y., Gomez, D., and Calsou, P. (2014). Alternative end-joining pathway(s): Bricolage at DNA breaks. *DNA Repair* 17, 81–97.

Gallego-Bartolomé, J., Gardiner, J., Liu, W., Papikian, A., Ghoshal, B., Kuo, H. Y., Zhao, J. M. C., Segal, D. J., and Jacobsen, S. E. (2018). Targeted DNA demethylation of the Arabidopsis genome using the human TET1 catalytic domain. *Proceedings of the National Academy of Sciences of the United States of America* 115, E2125–E2134.

Gao, H., Smith, J., Yang, M., Jones, S., Djukanovic, V., Nicholson, M. G., West, A., Bidney, D., Falco, S. C., Jantz, D., and Lyznik, L. A. (2010). Heritable targeted mutagenesis in maize using a designed endonuclease. *Plant Journal* 61, 176–187.

Gao, Y., Xiong, X., Wong, S., Charles, E. J., Lim, W. A., and Qi, L. S. (2016). Complex transcriptional modulation with orthogonal and inducible dCas9 regulators. *Nature Methods* 13, 1043–1049.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 109, E2579–E2586.

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., and Liu, D. R. (2017). Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage. *Nature* 551, 464–471.

Gautam, A. K., Balakrishnan, S., and Venkatraman, P. (2012). Direct ubiquitin independent recognition and degradation of a folded protein by the eukaryotic proteasomes-origin of intrinsic degradation signals. *PLoS ONE* 7, e34864.

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A., Lim, W. A., Weissman, J. S., and Qi, L. S. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154, 442–451.

Gorbunova, V., and Levy, A. A. (1999). How plants make ends meet: DNA double-strand break repair. *Trends in Plant Science* 4, 263–269.

Greenup, A., Peacock, W. J., Dennis, E. S., and Trevaskis, B. (2009). The molecular biology of seasonal flowering-responses in Arabidopsis and the cereals. *Annals of Botany* 103, 1165–1172.

Gu, J., Lu, H., Tippin, B., Shimazaki, N., Goodman, M. F., and Lieber, M. R. (2007). XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps. *EMBO Journal* 26, 1010–1023.

Guilinger, J. P., Thompson, D. B., and Liu, D. R. (2014). Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nature Biotechnology* 32, 577–582.

Guillon, F., Bouchet, B., Jamme, F., Robert, P., Quéméner, B., Barron, C., Larré, C., Dumas, P., and Saulnier, L. (2011). *Brachypodium distachyon* grain: Characterization of endosperm cell walls. *Journal of Experimental Botany* 62, 1001–1015.

Gyenes-Hegyi, Z., Pók, I., Kizmus, L., Zsubori, Z., Nagy, E., and Marton, L. C. (2002). Plant height and height of the main ear in maize (*Zea mays* L.) at different locations and different plant densities. *Acta Agronomica Hungarica* 50, 75–84.

Hajiahmadi, Z., Movahedi, A., Wei, H., Li, D., Orooji, Y., Ruan, H., and Zhuge, Q. (2019). Strategies to increase on-target and reduce off-target effects of the CRISPR/Cas9 system in plants. *International Journal of Molecular Sciences* 20, 3719.

Hammarsten, O., DeFazio, L. G., and Chu, G. (2000). Activation of DNA-dependent protein kinase by single-stranded DNA ends. *Journal of Biological Chemistry* 275, 1541–1550.

Hashimoto, R., Ueta, R., Abe, C., Osakabe, Y., and Osakabe, K. (2018). Efficient multiplex genome editing induces precise, and self-ligated type mutations in tomato plants. *Frontiers in Plant Science* 9, 916.

Hay, A. S., Pieper, B., Cooke, E., Mandáková, T., Cartolano, M., Tattersall, A. D., et al. (2014). *Cardamine hirsuta*: A versatile genetic system for comparative studies. *Plant Journal* 78, 1–15.

Heath, P. J., Stephens, K. M., Monnat, R. J., and Stoddard, B. L. (1997). The structure of I-CreI, a group I intron-encoded homing endonuclease. *Nature Structural Biology* 4, 468–476.

Hess, G. T., Tycko, J., Yao, D., and Bassik, M. C. (2017). Methods and applications of CRISPR-mediated base editing in eukaryotic genomes. *Molecular Cell* 68, 26–43.

Hilton, I. B., D'Ippolito, A. M., Vockley, C. M., Thakore, P. I., Crawford, G. E., Reddy, T. E., and Gersbach, C. A. (2015). Epigenome editing by a CRISPR/Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature Biotechnology* 33, 510–517.

Hirano, H., Gootenberg, J. S., Horii, T., Abudayyeh, O. O., Kimura, M., Hsu, P. D., Nakane, T., Ishitani, R., Hatada, I., Zhang, F., Nishimasu, H., and Nureki, O. (2016). Structure and engineering of *Francisella novicida* Cas9. *Cell* 164, 950–961.

Hirohata, A., Sato, I., Kaino, K., Iwata, Y., Koizumi, N. and Mishiba, K. I. (2019). CRISPR/Cas9-mediated homologous recombination in tobacco. Plant Cell Reports 38, 463-473.

Holland, C. K., and Jez, J. M. (2018). Arabidopsis: the original plant chassis organism. *Plant Cell Reports* 37, 1359–1366.

Hong, S. Y., Sun, B., Straub, D., Blaakmeer, A., Mineri, L., Koch, J., Brinch-Pedersen, H., Holme, I. B., Burow, M., Jørgensen, H. J. L., Albà, M. M., and Wenkel, S. (2020). Heterologous microProtein expression identifies LITTLE NINJA, a dominant regulator of jasmonic acid signaling. *Proceedings of the National Academy of Sciences of the United States of America* 117, 26197–26205.

Hong, W. J., Kim, Y. J., Chandran, A. K. N., and Jung, K. H. (2019). Infrastructures of systems biology that facilitate functional genomic study in rice. *Rice* 12, 15.

Hopfner, K. P., Karcher, A., Shin, D. S., Craig, L., Arthur, L. M., Carney, J. P., and Tainer, J. A. (2000). Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell* 101, 789–800.

Hou, Z., Zhang, Y., Propson, N. E., Howden, S. E., Chu, L. F., Sontheimer, E. J., and Thomson, J. A. (2013).

Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proceedings of the National Academy of Sciences of the United States of America* 110, 15644–15649.

Houbaviy, H. B., Usheva, A., Shenk, T., and Burley, S. K. (1996). Cocrystal structure of YY1 bound to the adeno-associated virus P5 initiator. *Proceedings of the National Academy of Sciences of the United States of America* 93, 13577–13582.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G., and Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology* 31, 827–832.

Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., Zeina, C. M., Gao, X., Rees, H. A., Lin, Z., and Liu, D. R. (2018a). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* 556, 57–63.

Hu, S., Hu, X., Hu, J., Shang, L., Dong, G., Zeng, D., Guo, L., and Qian, Q. (2018b). Xiaowei, a new rice germplasm for large-scale indoor research. *Molecular Plant* 11, 1418–1420.

Hu, X., Wang, C., Liu, Q., Fu, Y., and Wang, K. (2017). Targeted mutagenesis in rice using CRISPR/Cpf1 system. *Journal of Genetics and Genomics* 44, 71–73.

Hua, K., Tao, X., Yuan, F., Wang, D., and Zhu, J. K. (2018). Precise A·T to G·C base editing in the rice genome. *Molecular Plant* 11, 627–630.

Huang, P., Shyu, C., Coelho, C. P., Cao, Y., and Brutnell, T. P. (2016). *Setaria viridis* as a model system to advance millet genetics and genomics. *Frontiers in Plant Science* 7, 1781.

Hutchison, C. A., Phillips, S., Edgell, M. H., Gillam, S., Jahnke, P., and Smith, M. (1978). Mutagenesis at a specific position in a DNA sequence. *Journal of Biological Chemistry* 253, 6551–6560.

Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., Peterson, R. T., Yeh, J. R. J., and Joung, J. K. (2013). Efficient genome editing in zebrafish using a CRISPR/Cas system. *Nature Biotechnology* 31, 227–229.

Ibarra-Laclette, E., Lyons, E., Hernández-Guzmán, G., Pérez-Torres, C. A., Carretero-Paulet, L., Chang, T. H., et al. (2013). Architecture and evolution of a minute plant genome. *Nature* 498, 94–98.

Idziak, D., Robaszkiewicz, E., and Hasterok, R. (2015). Spatial distribution of centromeres and telomeres at interphase varies among *Brachypodium* species. *Journal of Experimental Botany* 66, 6623–6634.

Ishizaki, K., Nishihama, R., Yamato, K. T., and Kohchi, T. (2016). Molecular genetic tools and techniques for *Marchantia polymorpha* research. *Plant and Cell Physiology* 57, 262–270.

Izawa, T., and Shimamoto, K. (1996). Becoming a model plant: The importance of rice to plant science. *Trends in Plant Science* 1, 95–99.

Ji, J., Zhang, C., Sun, Z., Wang, L., Duanmu, D., and Fan, Q. (2019). Genome Editing in Cowpea *Vigna unguiculata* Using CRISPR/Cas9. *International Journal of Molecular Sciences* 20, 2471.

Jia, H., and Nian, W. (2014). Targeted genome editing of sweet orange using Cas9/sgRNA. *PLoS ONE* 9, e93806.

Jia, H., Orbovic, V., Jones, J. B., and Wang, N. (2016). Modification of the PthA4 effector binding elements in Type I CsLOB1 promoter using Cas9/sgRNA to produce transgenic Duncan grapefruit alleviating XccΔpthA4:dCsLOB1.3 infection. *Plant Biotechnology Journal* 14, 1291–301.

Jia, H., Zhang, Y., Orbović, V., Xu, J., White, F. F., Jones, J. B., and Wang, N. (2017). Genome editing of the disease susceptibility gene *CsLOB1* in citrus confers resistance to citrus canker. *Plant Biotechnology Journal* 15, 817–823.

Jiang, F., Taylor, D. W., Chen, J. S., Kornfeld, J. E., Zhou, K., Thompson, A. J., Nogales, E., and Doudna, J. A. (2016). Structures of a CRISPR/Cas9 R-loop complex primed for DNA cleavage. *Science* 351, 867–871.

Jiang, F., Zhou, K., Ma, L., Gressel, S., and Doudna, J. A. (2015). A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* 348, 1477–1481.

Jiang, W. Z., Henry, I. M., Lynagh, P. G., Comai, L., Cahoon, E. B., and Weeks, D. P. (2017). Significant enhancement of fatty acid composition in seeds of the allohexaploid, *Camelina sativa*, using CRISPR/Cas9 gene editing. *Plant Biotechnology Journal* 15, 648–657.

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B., and Weeks, D. P. (2013). Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic acids research* 41, e188.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptative bacterian immunity. *Science* 337, 816–822.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. *eLife* 2, e00471.

Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., Kaplan, M., Iavarone, A. T., Charpentier, E., Nogales, E., and Doudna, J. A. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 343, 1247997.

Josephs, E. A., Kocak, D. D., Fitzgibbon, C. J., McMenemy, J., Gersbach, C. A., and Marszalek, P. E. (2015). Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. *Nucleic Acids Research* 43, 8924–41.

Joshi, R. K., Bharat, S. S., and Mishra, R. (2020). Engineering drought tolerance in plants through CRISPR/Cas genome editing. *3 Biotech* 10, 400.

Jurica, M. S., Monnat, R. J., and Stoddard, B. L. (1998). DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-CreI. *Molecular Cell* 2, 469–476.

Kane, N. C., Gill, N., King, M. G., Bowers, J. E., Berges, H., Gouzy, J., Bachlava, E., Langlade, N. B., Lai, Z., Stewart, M., Burke, J. M., Vincourt, P., Knapp, S. J., and Rieseberg, L. H. (2011). Progress towards a reference genome for sunflower. *Botany* 89, 429–437.

Kang, B. C., Yun, J. Y., Kim, S. T., Shin, Y. J., Ryu, J., Choi, M., Woo, J. W., and Kim, J. S. (2018). Precision genome engineering through adenine base editing in plants. *Nature Plants* 4, 427–431.

Kearns, N. A., Pham, H., Tabak, B., Genga, R. M., Silverstein, N. J., Garber, M., and Maehr, R. (2015). Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nature Methods* 12, 401–403.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* 10, 845–858.

Khan, M. A., and Stace, C. A. (1999). Breeding relationships in the genus *Brachypodium* (*Poaceae*: *Pooideae*). *Nordic Journal of Botany* 19, 257–269.

Kim, C. M. (2020a). Increasing plant yield by modulating root hair development in *Brachypodium*

distachyon. *Korean Journal of Agricultural Science* 47, 305–313.

Kim, H., Choi, J., and Won, K.-H. (2020b). A stable DNA-free screening system for CRISPR/RNPs-mediated gene editing in hot and sweet cultivars of *Capsicum annuum*. *BMC Plant Biology* 20, 449.

Kim, H., Kim, S. T., Ryu, J., Kang, B. C., Kim, J. S., and Kim, S. G. (2017a). CRISPR/Cpf1-mediated DNA-free plant genome editing. *Nature Communications* 8, 14406.

Kim, S., Kim, D., Cho, S. W., Kim, J., and Kim, J. S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Research* 24, 1012–1019.

Kim, Y. B., Komor, A. C., Levy, J. M., Packer, M. S., Zhao, K. T., and Liu, D. R. (2017b). Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nature Biotechnology* 35, 371–376.

Klap, C., Yeshayahou, E., Bolger, A. M., Arazi, T., Gupta, S. K., Shabtai, S., Usadel, B., Salts, Y., and Barg, R. (2017). Tomato facultative parthenocarpy results from *SlAGAMOUS-LIKE 6* loss of function. *Plant Biotechnology Journal* 15, 634–647.

Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., and Joung, J. K. (2016). High-fidelity CRISPR/Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529, 490–495.

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Topkar, V. V., Zheng, Z., and Joung, J. K. (2015a). Broadening the targeting range of *Staphylococcus aureus* CRISPR/Cas9 by modifying PAM recognition. *Nature Biotechnology* 33, 1293–1298.

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V, Nguyen, N. T., Zheng, Z., Gonzales, A. P. W., Li, Z., Peterson, R. T., Yeh, J. R. J., Aryee, M. J., and Joung, J. K. (2015b). Engineered CRISPR/Cas9 nucleases with altered PAM specificities. *Nature* 523, 481–485.

Koch, C. A., Agyei, R., Galicia, S., Metalnikov, P., O'Donnell, P., Starostine, A., Weinfeld, M., and Durocher, D. (2004). Xrcc4 physically links DNA end processing by polynucleotide kinase to DNA ligation by DNA ligase IV. *EMBO Journal* 23, 3874–3885.

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424.

Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O. O., Barcena, C., Hsu, P. D., Habib, N., Gootenberg, J. S., Nishimasu, H., Nureki, O., and Zhang, F. (2015). Genome-scale transcriptional activation by an engineered CRISPR/Cas9 complex. *Nature* 517, 583–588.

Kouzai, Y., Shimizu, M., Inoue, K., Uehara-Yamaguchi, Y., Takahagi, K., Nakayama, R., Matsuura, T., Mori, I. C., Hirayama, T., Abdelsalam, S. S. H., Noutoshi, Y., and Mochida, K. (2020). *BdWRKY38* is required for the incompatible interaction of *Brachypodium distachyon* with the necrotrophic fungus *Rhizoctonia solani*. *The Plant Journal* 104, 995.

Krämer, U. (2015). Planting molecular functions in an ecological context with *Arabidopsis thaliana*. *eLife* 4, e06100.

Krasileva, K. V., Vasquez-Gross, H. A., Howell, T., Bailey, P., Paraiso, F., Clissold, L., Simmonds, J., Ramirez-Gonzalez, R. H., Wang, X., Borrill, P., Fosker, C., Ayling, S., Phillips, A. L., Uauy, C., and Dubcovsky, J. (2017). Uncovering hidden variation in polyploid wheat. *Proceedings of the National Academy of Sciences of the United States of America* 114, E913–E921.

Krysan, P. J., Young, J. C., and Sussman, M. R. (1999). T-DNA as an insertional mutagen in Arabidopsis. *Plant Cell* 11, 2283–2290.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 35, 1547–1549.

Kuromori, T., Hirayama, T., Kiyosue, Y., Takabe, H., Mizukado, S., Sakurai, T., Akiyama, K., Kamiya, A., Ito, T., and Shinozaki, K. (2004). A collection of 11 800 single-copy Ds transposon insertion lines in Arabidopsis. *Plant Journal* 37, 897–905.

Kuromori, T., Sugimoto, E., and Shinozaki, K. (2021). Brachypodium *BdABCG25* is a homolog of Arabidopsis *AtABCG25* involved in the transport of abscisic acid. *FEBS Letters* 595, 954–959.

Lambowitz, A. M., and Belfort, M. (1993). Introns as mobile genetic elements. *Annual Review of Biochemistry* 62, 587–622.

Larré, C., Penninck, S., Bouchet, B., Lollier, V., Tranquet, O., Denery-Papini, S., Guillon, F., and Rogniaux, H. (2010). *Brachypodium distachyon* grain: Identification and subcellular localization of storage proteins. *Journal of Experimental Botany* 61, 1771–1783.

Laudencia-Chingcuanco, D. L., and Vensel, W. H. (2008). Globulins are the main seed storage proteins in *Brachypodium distachyon*. *Theoretical and Applied Genetics* 117, 555–563.

Lawrenson, T., Shorinola, O., Stacey, N., Li, C., Østergaard, L., Patron, N., Uauy, C., and Harwood, W. (2015). Induction of targeted, heritable mutations in barley and *Brassica oleracea* using RNA-guided Cas9 nuclease. *Genome biology* 16, 258.

Lee, C. M., Cradick, T. J., and Bao, G. (2016). The *Neisseria meningitidis* CRISPR/Cas9 system enables specific genome editing in mammalian cells. *Molecular Therapy* 24, 645–654.

Lee, M. B., Jeon, W. B., Kim, D. Y., Bold, O., Hong, M. J., Lee, Y. J., Park, J. H., and Seo, Y. W. (2011). *Agrobacterium*-mediated transformation of *Brachypodium distachyon* inbred line Bd21 with two binary vectors containing hygromycin resistance and GUS reporter genes. *Journal of Crop Science and Biotechnology* 14, 233–238.

Li, C., Rudi, H., Stockinger, E. J., Cheng, H., Cao, M., Fox, S. E., Mockler, T. C., Westereng, B., Fjellheim, S., Rognli, O. A., and Sandve, S. R. (2012a). Comparative analyses reveal potential uses of *Brachypodium distachyon* as a model for cold stress responses in temperate grasses. *BMC Plant Biology* 12, 65.

Li, C., Zong, Y., Wang, Y., Jin, S., Zhang, D., Song, Q., Zhang, R., and Gao, C. (2018a). Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biology* 19, 59.

Li, G., Jain, R., Chern, M., Pham, N. T., Martin, J. A., Wei, T., Schackwitz, W. S., Lipzen, A. M., Duong, P. Q., Jones, K. C., Jiang, L., Ruan, D., Bauer, D., Peng, Y., Barry, K. W., Schmutz, J., and Ronald, P. C. (2017). The sequences of 1504 mutants in the model rice variety Kitaake facilitate rapid functional genomic studies. *Plant Cell* 29, 1218–1231.

Li, G., Liu, D., Zhang, X., Quan, R., Zhong, C., Mo, J., Huang, Y., Wang, H., Ruan, X., Xu, Z., Zheng, E., Gu, T., Hong, L., Li, Z., Wu, Z., and Yang, H. (2018b). Suppressing Ku70/Ku80 expression elevates homology-directed repair efficiency in primary fibroblasts. *International Journal of Biochemistry and Cell Biology* 99, 154–160.

Li, J. F., Zhang, D., and Sheen, J. (2015a). Targeted plant genome editing via the CRISPR/Cas9 technology. *Methods in Molecular Biology (Clifton, N.J.)* 1284, 239–255.

Li, J. J., Meng, X., Zong, Y., Chen, K., Zhang, H., Liu, J., Li, J. J., and Gao, C. (2016a). Gene replacements and insertions in rice by intron targeting using CRISPR/Cas9. *Nature Plants* 2, 16139.

Li, K., Wang, G., Andersen, T., Zhou, P., and Pu, W. T. (2014). Optimization of genome engineering approaches with the CRISPR/Cas9 system. *PLoS ONE* 9, e105779.

Li, M., Li, X., Zhou, Z., Wu, P., Fang, M., Pan, X., Lin, Q., Luo, W., Wu, G., and Li, H. (2016b). Reassessment of the Four Yield-related Genes *Gn1a*, *DEP1*, *GS3*, and *IPA1* in Rice Using a CRISPR/Cas9 System. *Frontiers in Plant Science* 7, 377.

Li, S., Li, J., Zhang, J., Du, W., Fu, J., Sutar, S., Zhao, Y., and Xia, L. (2018c). Synthesis-dependent repair of Cpf1-induced double strand DNA breaks enables targeted gene replacement in rice. *Journal of Experimental Botany* 69, 4715–4721.

Li, S., Zhang, X., Wang, W., Guo, X., Wu, Z., Du, W., Zhao, Y., and Xia, L. (2018d). Expanding the scope of CRISPR/Cpf1-mediated genome editing in rice. *Molecular Plant* 11, 995–998.

Li, A. X., Jia, S. G., Yobi, A., Ge, Z. X., Sato, S. J., Zhang, C., Angelovici, R., Clemente, T. E., and Holding, D. R. (2018e). Editing of an alpha-kafirin gene family increases digestibility and protein quality in sorghum. *Plant Physiology* 177, 1425–1438.

Li, J., Jiao, G., Sun, Y., Chen, J., Zhong, Y., Yan, L., Jiang, D., Ma, Y., and Xia, L. (2020). Modification of starch composition, structure and properties through editing of TaSBEIIa in both winter and spring wheat varieties by CRISPR/Cas9. Plant Biotechnology Journal 19, 937-951.

Li, T., Liu, B., Spalding, M. H., Weeks, D. P., and Yang, B. (2012b). High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nature Biotechnology* 30, 390–392.

Li, T., Yang, X., Yu, Y., Si, X., Zhai, X., Zhang, H., Dong, W., Gao, C., and Xu, C. (2018f). Domestication of wild tomato is accelerated by genome editing. *Nature Biotechnology* 36, 1160–1163.

Li, Z., Liu, Z. Bin, Xing, A., Moon, B. P., Koellhoffer, J. P., Huang, L., Ward, R. T., Clifton, E., Falco, S. C., and Cigan, A. M. (2015b). Cas9-guide RNA directed genome editing in soybean. *Plant Physiology* 169, 960–970.

Liang, L., Deng, L., Nguyen, S. C., Zhao, X., Maulion, C. D., Shao, C., and Tischfield, J. A. (2008). Human DNA ligases I and III, but not ligase IV, are required for microhomology-mediated end joining of DNA double-strand breaks. *Nucleic Acids Research* 36, 3297–3310.

Liang, Z., Chen, K., Li, T., Zhang, Y., Wang, Y., Zhao, Q., Liu, J., Zhang, H., Liu, C., Ran, Y., and Gao, C. (2017). Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. *Nature Communications* 8, 14261.

Liang, Z., Zhang, K., Chen, K., and Gao, C. (2014). Targeted mutagenesis in *Zea mays* using TALENs and the CRISPR/Cas system. *Journal of Genetics and Genomics* 41, 63–68.

Liao, S. Y., Qin, X. M., Luo, L., Han, Y., Wang, X., Usman, B., Nawaz, G., Zhao, N., Liu, Y. G., and Li, R. B. (2019). CRISPR/Cas9-induced mutagenesis of *Semi-Rolled Leaf1,2* confers curled leaf phenotype and drought tolerance by influencing protein expression patterns and ROS scavenging in rice (*Oryza sativa* L.). *Agronomy* 9, 728.

Lin, S., Staahl, B. T., Alla, R. K., and Doudna, J. A. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* 3, e04766.

Liu, K. I., Ramli, M. N. Bin, Woo, C. W. A., Wang, Y., Zhao, T., Zhang, X., Yim, G. R. D., Chong, B. Y., Gowher, A., Chua, M. Z. H., Jung, J., Lee, J. H. J., and Tan, M. H. (2016). A chemical-inducible CRISPR/Cas9 system for rapid control of genome editing. *Nature Chemical Biology* 12, 980–987.

Lu, Q., Guo, F., Xu, Q., Cang, J., Lu, Q., Guo, F., Xu, Q., and Cang, J. (2020). LncRNA improves cold resistance of winter wheat by interacting with miR398. *Functional Plant Biology* 47, 544–557.

Lowder, L. G., Zhang, D., Baltes, N. J., Paul, J. W., Tang, X., Zheng, X., Voytas, D. F., Hsieh, T. F., Zhang, Y., and Qi, Y. (2015). A CRISPR/Cas9 toolbox for multiplexed plant genome editing and transcriptional regulation. *Plant Physiology* 169, 971–985.

Ma, H., Naseri, A., Reyes-Gutierrez, P., Wolfe, S. A., Zhang, S., and Pederson, T. (2015a). Multicolor CRISPR labeling of chromosomal loci in human cells. *Proceedings of the National Academy of Sciences of the United States of America* 112, 3002–3007.

Ma, L., Zhang, D., Miao, Q., Yang, J., Xuan, Y., and Hu, Y. (2017). Essential role of sugar transporter *OsSWEET11* during the early stage of rice grain filling. *Plant and Cell Physiology* 58, 863–873.

Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., et al. (2015b). A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Molecular Plant* 8, 1274–1284.

Ma, Y., Lu, H., Tippin, B., Goodman, M. F., Shimazaki, N., Koiwai, O., Hsieh, C. L., Schwarz, K., and Lieber, M. R. (2004). A biochemically defined system for mammalian nonhomologous DNA end joining. *Molecular Cell* 16, 701–713.

MacKinnon, K. J. M., Cole, B. J., Yu, C., Coomey, J. H., Hartwick, N. T., Remigereau, M. S., Duffy, T., Michael, T. P., Kay, S. A., and Hazen, S. P. (2020). Changes in ambient temperature are the prevailing cue in determining *Brachypodium distachyon* diurnal gene regulation. *New Phytologist* 227, 1709–1724.

Maeder, M. L., Linder, S. J., Cascio, V. M., Fu, Y., Ho, Q. H., and Joung, J. K. (2013). CRISPR RNA-guided activation of endogenous human genes. *Nature Methods* 10, 977–979.

Maji, B., Moore, C. L., Zetsche, B., Volz, S. E., Zhang, F., Shoulders, M. D., and Choudhary, A. (2017). Multidimensional chemical control of CRISPR/Cas9. *Nature Chemical Biology* 13, 9–11.

Majka, J., Alford, B., Ausio, J., Finn, R. M., and McMurray, C. T. (2012). ATP hydrolysis by RAD50 protein switches MRE11 enzyme from endonuclease to exonuclease. *Journal of Biological Chemistry* 287, 2328–2341.

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J. M., Wolf, Y. I., Yakunin, A. F., van der Oost, J., and Koonin, E. V (2011). Evolution and classification of the CRISPR/Cas systems. *Nature reviews. Microbiology* 9, 467–77.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., and Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science* 339, 823–826.

Marraffini, L. A., and Sontheimer, E. J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* 463, 568–571.

Martin, R. C., Kronmiller, B. A., and Dombrowski, J. E. (2020). Transcriptome analysis of responses in *Brachypodium distachyon* overexpressing the *BdbZIP26* transcription factor. *BMC Plant Biology* 20, 1–18.

Martínez-Fortún, J., Phillips, D. W., and Jones, H. D. (2017). Potential impact of genome editing in world agriculture. *Emerging Topics in Life Sciences* 1, 117–133.

Maruyama, T., Dougan, S. K., Truttmann, M. C., Bilate, A. M., Ingram, J. R., and Ploegh, H. L. (2015). Increasing the efficiency of precise genome editing with CRISPR/Cas9 by inhibition of nonhomologous end joining. *Nature Biotechnology* 33, 538–542.

McCaw, M. E., Wallace, J. G., Albert, P. S., Buckler, E. S., and Birchler, J. A. (2016). Fast-flowering mini-maize: Seed to seed in 60 days. *Genetics* 204, 35–42.

McConnell Smith, A., Takeuchi, R., Pellenz, S., Davis, L., Maizels, N., Monnat, R. J., and Stoddard, B. L. (2009). Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. *Proceedings of the National Academy of Sciences of the United States of America* 106, 5099–5104.

McDonald, J. I., Celik, H., Rois, L. E., Fishberger, G., Fowler, T., Rees, R., Kramer, A., Martens, A., Edwards, J. R., and Challen, G. A. (2016). Reprogrammable CRISPR/Cas9-based system for inducing sites-pecific DNA methylation. *Biology Open* 5, 866–874.

Mekler, V., Minakhin, L., Semenova, E., Kuznedelov, K., and Severinov, K. (2016). Kinetics of the CRISPR/Cas9 effector complex assembly and the role of 3′-terminal segment of guide RNA. *Nucleic Acids Research* 44, 2837–2845.

Meng, Y., Hou, Y., Wang, H., Ji, R., Liu, B., Wen, J., Niu, L., and Lin, H. (2017). Targeted mutagenesis by CRISPR/Cas9 system in the model legume *Medicago truncatula*. *Plant Cell Reports* 36, 371–374.

Meng, Y., Wang, C., Yin, P., Zhu, B., Zhang, P., Niu, L., and Lin, H. (2020). "Targeted mutagenesis by an optimized *Agrobacterium*-delivered CRISPR/Cas9 system in the model legume *Medicago truncatula*." in *The Model Legume Medicago truncatula*, ed. de Bruijn, F. (Hoboken, NJ: John Wiley & Sons) 1015–1018.

Metz, R., DiCola, M., Kurihara, T., Bailey, A., Frank, B., Roecklein, B., and Blaese, M. (2002). Mode of action of RNA/DNA oligonucleotides: Progress in the development of gene repair as a therapy for α1-antitrypsin deficiency. *Chest* 121, 91S-97S.

Miao, C., Wang, Z., Zhang, L., Yao, J., Hua, K., Liu, X., Shi, H., and Zhu, J.K. (2019). The grain yield modulator miR156 regulates seed dormancy through the gibberellin pathway in rice. *Nature Communications* 10, 3822.

Miao, J., Guo, D., Zhang, J., Huang, Q., Qin, G., Zhang, X., Wan, J., Gu, H., and Qu, L. J. (2013). Targeted mutagenesis in rice using CRISPR/Cas system. *Cell Research* 23, 1233–1236.

Michael, T. P. (2014). Plant genome size variation: Bloating and purging DNA. *Briefings in Functional Genomics and Proteomics* 13, 308–317.

Miller, J. C., Tan, S., Qiao, G., Barlow, K. A., Wang, J., Xia, D. F., et al. (2011). A TALE nuclease architecture for efficient genome editing. *Nature Biotechnology* 29, 143–150.

Miller, J., McLachlan, A. D., and Klug, A. (2001). Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *Journal of Trace Elements in Experimental Medicine* 14, 157–169.

Ming, R., VanBuren, R., Wai, C. M., Tang, H., Schatz, M. C., Bowers, J. E., et al. (2015). The pineapple genome and the evolution of CAM photosynthesis. *Nature Genetics* 47, 1435–1442.

Mishra, R., Joshi, R. K. and Zhao, K. (2020). Base editing in crops: current advances, limitations and future implications. *Plant Biotechnology Journal* 18, 20-31.

Miyao, A., Tanaka, K., Murata, K., Sawaki, H., Takeda, S., Abe, K., Shinozuka, Y., Onosato, K., and Hirochika, H. (2003). Target site specificity of the Tos17 retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. *Plant Cell* 15, 1771–1780.

Mockler, T. C., Michael, T. P., Priest, H. D., Shen, R., Sullivan, C. M., Givan, S. A., Mcentee, C., Kay, S. A., and Chory, J. (2007). The DIURNAL project: Diurnal and circadian expression profiling, model-based pattern matching, and promoter analysis. *Cold Spring Harbor Symposia on Quantitative Biology* 72, 353–363.

Moscou, M. J., and Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science* 326, 1501–1501.

Mosquna, A., Katz, A., Decker, E. L., Rensing, S. A., Reski, R., and Ohad, N. (2009). Regulation of stem cell maintenance by the Polycomb protein FIE has been conserved during land plant evolution. *Development* 136, 2433–2444.

Müller, L. M., Campos-Soriano, L., Levesque-Tremblay, V., Bravo, A., Daniels, D. A., Pathak, S., Park, H. J., and Harrison, M. J. (2020). Constitutive Overexpression of *RAM1* Leads to an Increase in Arbuscule Density in *Brachypodium distachyon*. *Plant Physiology* 184, 1263–1272.

Mussolino, C., Morbitzer, R., Lütge, F., Dannemann, N., Lahaye, T., and Cathomen, T. (2011). A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Research* 39, 9283–9293.

Myburg, A. A., Grattapaglia, D., Tuskan, G. A., Hellsten, U., Hayes, R. D., Grimwood, J., et al. (2014). The genome of *Eucalyptus grandis*. *Nature* 510, 356–362.

Nam, B. E., Park, Y. J., Gil, K. E., Kim, J. H., Kim, J. G., and Park, C. M. (2020). Auxin mediates the touch-induced mechanical stimulation of adventitious root formation under windy conditions in *Brachypodium distachyon*. *BMC Plant Biology* 20, 1–15.

Nekrasov, V., Wang, C., Win, J., Lanz, C., Weigel, D., and Kamoun, S. (2017). Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. *Scientific Reports* 7, 482.

Nicolia, A., Proux-Wéra, E., Åhman, I., Onkokesung, N., Andersson, M., Andreasson, E., and Zhu, L.-H. (2015). Targeted gene mutation in tetraploid potato through transient TALEN expression in protoplasts. *Journal of Biotechnology* 204, 17–24.

Nihongaki, Y., Yamamoto, S., Kawano, F., Suzuki, H., and Sato, M. (2015). CRISPR/Cas9-based photoactivatable transcription system. *Chemistry and Biology* 22, 169–174.

Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156, 935–49.

Nishitani, C., Hirai, N., Komori, S., Wada, M., Okada, K., Osakabe, K., Yamamoto, T., and Osakabe, Y. (2016). Efficient genome editing in apple using a CRISPR/Cas9 system. *Scientific Reports* 6, 31481.

Nishizawa-Yokoi, A., Mikami, M. and Toki, S. (2020). A universal system of CRISPR/Cas9-mediated gene targeting using all-in-one vector in plants. *Frontiers in Genome Editing* 2, 604289.

Niu, Y., Tenney, K., Li, H., and Gimble, F. S. (2008). Engineering Variants of the I-SceI Homing endonuclease with strand-specific and site-specific DNA-nicking activity. *Journal of Molecular Biology* 382, 188–202.

Nolte, R. T., Conlin, R. M., Harrison, S. C., and Brown, R. S. (1998). Differing roles for zinc fingers in DNA recognition: Structure of a six-finger transcription factor IIIA complex. *Proceedings of the National Academy of Sciences of the United States of America* 95, 2938–2943.

Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y. C., Scofield, D. G., et al. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature* 497, 579–584.

O'Connell, M. R., Oakes, B. L., Sternberg, S. H., East-Seletsky, A., Kaplan, M., and Doudna, J. A. (2014). Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* 516, 263–266.

Odipio, J., Alicai, T., Ingelbrecht, I., Nusinow, D. A., Bart, R., & Taylor, N. J. (2017). Efficient CRISPR/Cas9 genome editing of phytoene desaturase in cassava. *Frontiers in Plant Science* 8, 1780.

Okano, Y., Aono, N., Hiwatashi, Y., Murata, T., Nishiyama, T., Ishikawa, T., Kubo, M., and Hasebe, M. (2009). A polycomb repressive complex 2 gene regulates apogamy and gives evolutionary insights into early land plant evolution. *Proceedings of the National Academy of Sciences of the United States of America* 106, 16321–16326.

Orlowski, J., Boniecki, M., and Bujnicki, J. M. (2007). I-Ssp6803I: The first homing endonuclease from the PD-(D/E)XK superfamily exhibits an unusual mode of DNA recognition. *Bioinformatics* 23, 527–530.

Ortigosa, A., Gimenez-Ibanez, S., Leonhardt, N., and Solano, R. (2019). Design of a bacterial speck resistant tomato by CRISPR/Cas9-mediated editing of *SlJAZ2*. *Plant Biotechnology Journal* 17, 665–673.

Osakabe, Y., and Osakabe, K. (2015). Genome editing with engineered nucleases in plants. *Plant and Cell Physiology* 56, 389–400.

Oz, M. T., Altpeter, A., Karan, R., Merotto, A., & Altpeter, F. (2021). CRISPR/Cas9-mediated multi-allelic gene targeting in sugarcane confers herbicide tolerance. *Frontiers in Genome Editing* 3, 15.

Pacher, M., Schmidt-Puchta, W., and Puchta, H. (2007). Two unlinked double-strand breaks can induce reciprocal exchanges in plant genomes via homologous recombination and nonhomologous end joining. *Genetics* 175, 21–29.

Pan, C., Wu, X., Markel, K., Malzahn, A. A., Kundagrami, N., Sretenovic, S., Zhang, Y., Cheng, Y., Shih, P. M. and Qi, Y. (2021). CRISPR-Act3.0 for highly efficient multiplexed gene activation in plants. *Nature Plants* 7, 942–953.

Palermo, G., Miao, Y., Walker, R. C., Jinek, M., and McCammon, J. A. (2016). Striking plasticity of CRISPR/Cas9 and key role of non-target DNA, as revealed by molecular simulations. *ACS Central Science* 2, 756–763.

Palmbos, P. L., Wu, D., Daley, J. M., and Wilson, T. E. (2008). Recruitment of *Saccharomyces cerevisiae* Dnl4-Lif1 complex to a double-strand break requires interactions with Yku80 and the Xrs2 FHA domain. *Genetics* 180, 1809–1819.

Parinov, S., Sevugan, M., Ye, D., Yang, W. C., Kumaran, M., and Sundaresan, V. (1999). Analysis of flanking sequences from Dissociation insertion lines: A database for reverse genetics in Arabidopsis. *Plant Cell* 11, 2263–2270.

Parker, D., Beckmann, M., Enot, D. P., Overy, D. P., Rios, Z. C., Gilbert, M., Talbot, N., and Draper, J. (2008). Rice blast infection of *Brachypodium distachyon* as a model system to study dynamic host/pathogen interactions. *Nature Protocols* 3, 435–445.

Pattanayak, V., Lin, S., Guilinger, J. P., Ma, E., Doudna, J. A., and Liu, D. R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnology* 31, 839–843.

Pavletich, N. P., and Pabo, C. O. (1991). Zinc finger-DNA recognition: Crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 252, 809–817.

Pavletich, N. P., and Pabo, C. O. (1993). Crystal structure of a five-finger GLI-DNA complex: New perspectives on zinc fingers. *Science* 261, 1701–1707.

Peck, S. H., Chen, I., and Liu, D. R. (2011). Directed evolution of a small-molecule-triggered intein with improved splicing properties in mammalian cells. *Chemistry and Biology* 18, 619–630.

Pellicer, J., Fay, M. F., and Leitch, I. J. (2010). The largest eukaryotic genome of them all? *Botanical Journal of the Linnean Society* 164, 10–15.

Peng, F., Zhang, W., Zeng, W., Zhu, J.K. & Miki, D. (2020). Gene targeting in *Arabidopsis* via an all-in-one strategy that uses a translational enhancer to aid Cas9 expression. *Plant Biotechnology Journal* 18, 892–894.

Peng, A., Chen, S., Lei, T., Xu, L., He, Y., Wu, L., Yao, L., and Zou, X. (2017). Engineering canker-resistant plants through CRISPR/Cas9-targeted editing of the susceptibility gene *CsLOB1* promoter in citrus. *Plant Biotechnology Journal* 15, 1509–1519.

Perdiguero, P., Venturas, M., Cervera, M. T., Gil, L., and Collada, C. (2015). Massive sequencing of *Ulmus minor*'s transcriptome provides new molecular tools for a genus under the constant threat of Dutch elm disease. *Frontiers in Plant Science* 6, 1–12.

Perez-Pinera, P., Kocak, D. D., Vockley, C. M., Adler, A. F., Kabadi, A. M., Polstein, L. R., Thakore, P. I., Glass, K. A., Ousterout, D. G., Leong, K. W., Guilak, F., Crawford, G. E., Reddy, T. E., and Gersbach, C. A. (2013). RNA-guided gene activation by CRISPR/Cas9-based transcription factors. *Nature methods* 10, 973–6.

Pernstich, C., and Halford, S. E. (2012). Illuminating the reaction pathway of the FokI restriction endonuclease by fluorescence resonance energy transfer. *Nucleic Acids Research* 40, 1203–1213.

Peterson, B. A., Haak, D. C., Nishimura, M. T., Teixeira, P. J. P. L., James, S. R., Dangl, J. L., and Nimchuk, Z. L. (2016). Genome-wide assessment of efficiency and specificity in CRISPR/Cas9 mediated multiple site targeting in Arabidopsis. *PLoS ONE* 11, e0162169.

Petrik, D. L., Tryfona, T., Dupree, P., and Anderson, C. T. (2020). BdGT43B2 functions in xylan biosynthesis and is essential for seedling survival in *Brachypodium distachyon*. *Plant Direct* 4, e00216.

Piatek, A., Ali, Z., Baazim, H., Li, L., Abulfaraj, A., Al-Shareef, S., Aouida, M., and Mahfouz, M. M. (2015). RNA-guided transcriptional regulation *in planta* via synthetic dCas9-based transcription factors. *Plant Biotechnology Journal* 13, 578–589.

Pickar-Oliver, A., and Gersbach, C. A. (2019). The next generation of CRISPR/Cas technologies and applications. *Nature Reviews Molecular Cell Biology* 20, 490–507.

Pinder, J., Salsman, J., and Dellaire, G. (2015). Nuclear domain "knock-in" screen for the evaluation and identification of small molecule enhancers of CRISPR-based genome editing. *Nucleic Acids Research* 43, 9379–9392.

Polstein, L. R., and Gersbach, C. A. (2015). A light-inducible CRISPR/Cas9 system for control of endogenous gene activation. *Nature Chemical Biology* 11, 198–200.

Portwood, J. L., Woodhouse, M. R., Cannon, E. K., Gardiner, J. M., Harper, L. C., Schaeffer, M. L., et al. (2019). MaizeGDB 2018: The maize multi-genome genetics and genomics database. *Nucleic Acids Research* 47, D1146–D1154.

Powell, J. J., Carere, J., Sablok, G., Fitzgerald, T. L., Stiller, J., Colgrave, M. L., Gardiner, D. M., Manners, J. M., Vogel, J. P., Henry, R. J., and Kazan, K. (2017). Transcriptome analysis of Brachypodium during fungal pathogen infection reveals both shared and distinct defense responses with wheat. *Scientific Reports* 7, 1–14.

Prieto, J., Redondo, P., Padró, D., Arnould, S., Epinat, J. C., Pâques, F., Blanco, F. J., and Montoya, G. (2007). The C-terminal loop of the homing endonuclease I-CreI is essential for site recognition, DNA binding and cleavage. *Nucleic Acids Research* 35, 3262–3271.

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P. and Lim, W. A., (2013a). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.

Cell 152, 1173–83.

Qi, Y., Zhang, Y., Zhang, F., Baller, J. A., Cleland, S. C., Ryu, Y., Starker, C. G., and Voytas, D. F. (2013b). Increasing frequencies of site-specific mutagenesis and gene targeting in Arabidopsis by manipulating DNA repair pathways. *Genome Research* 23, 547–554.

Qiu, Y. L., and Yu, J. (2003). *Azolla*-a model organism for plant genomic studies. *Genomics Proteomics Bioinformatics* 1, 15–25.

Ran, F. A., Cong, L., Yan, W. X., Scott, D. A., Gootenberg, J. S., Kriz, A. J., Zetsche, B., Shalem, O., Wu, X., Makarova, K. S., Koonin, E. V., Sharp, P. A., and Zhang, F. (2015). *In vivo* genome editing using *Staphylococcus aureus* Cas9. *Nature* 520, 186–191.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013). Genome engineering using the CRISPR/Cas9 system. *Nature Protocols* 8, 2281–2308.

Rancour, D. M., Marita, J. M., and Hatfield, R. D. (2012). Cell wall composition throughout development for the model grass *Brachypodium distachyon*. *Frontiers in Plant Science* 3, 266.

Rees, H. A., Komor, A. C., Yeh, W. H., Caetano-Lopes, J., Warman, M., Edge, A. S. B., and Liu, D. R. (2017). Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nature Communications* 8, 1–10.

Ren, C., Liu, X., Zhang, Z., Wang, Y., Duan, W., Li, S., and Liang, Z. (2016). CRISPR/Cas9-mediated efficient targeted mutagenesis in Chardonnay (*Vitis vinifera* L.). *Scientific Reports* 6, 32289.

Renaud, J. B., Boix, C., Charpentier, M., De Cian, A., Cochennec, J., Duvernois-Berthet, E., et al. (2016). Improved genome editing efficiency and flexibility using modified oligonucleotides with TALEN and CRISPR/Cas9 nucleases. *Cell Reports* 14, 2263–2272.

Reyes-Chin-Wo, S., Wang, Z., Yang, X., Kozik, A., Arikit, S., Song, C., et al. (2017). Genome assembly with *in vitro* proximity ligation data and whole-genome triplication in lettuce. *Nature Communications* 8, 14953.

Riballo, E., Woodbine, L., Stiff, T., Walker, S. A., Goodarzi, A. A., and Jeggo, P. A. (2009). XLF-Cernunnos promotes DNA ligase IV-XRCC4 re-adenylation following ligation. *Nucleic Acids Research* 37, 482–492.

Ribeiro, L. F., Ribeiro, L. F. C., Barreto, M. Q., and Ward, R. J. (2018). Protein Engineering Strategies to Expand CRISPR/Cas9 Applications. *International Journal of Genomics* 2018, 1652567.

Rivera-Contreras, I. K., Zamora-Hernández, T., Huerta-Heredia, A. A., Capataz-Tafur, J., Barrera-Figueroa, B. E., Juntawong, P., and Peña-Castro, J. M. (2016). Transcriptomic analysis of submergence-tolerant and sensitive *Brachypodium distachyon* ecotypes reveals oxidative stress as a major tolerance factor. *Scientific Reports* 6, 1–15.

Ro, D. K., Paradise, E. M., Quellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J., Chang, M. C. Y., Withers, S. T., Shiba, Y., Sarpong, R., and Keasling, J. D. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 440, 940–943.

Rodríguez-Leal, D., Lemmon, Z. H., Man, J., Bartlett, M. E., and Lippman, Z. B. (2017). Engineering quantitative trait variation for crop improvement by genome editing. *Cell* 171, 470-480.e8.

Roh, J., Moon, J., Lee, Y. E., Park, C. H., and Kim, S. K. (2021). Seed-specific expression of Arabidopsis *AtCYP85A2* produces biologically active brassinosteroids such as castasterone and brassinolide to improve grain yield and quality in seeds of *Brachypodium distachyon*. *Frontiers in Plant Science* 12, 639508.

Rosen, L. E., Morrison, H. A., Masri, S., Brown, M. J., Springstubb, B., Sussman, D., Stoddard, B. L., and Seligman, L. M. (2006). Homing endonuclease I-CreI derivatives with novel DNA target specificities. *Nucleic Acids Research* 34, 4791–4800.

Rowlands, H. (2021). A transgenic approach to testing functional divergence of stress-responsive *GROWTH REGULATORY FACTOR* genes across flowering plants. *UVM Student Research Conference*.

Rutkauskas, M., Sinkunas, T., Songailiene, I., Tikhomirova, M. S., Siksnys, V., and Seidel, R. (2015). Directional R-loop formation by the CRISPR/Cas surveillance complex cascade provides efficient off-target site rejection. *Cell Reports* 10, 1534–1543.

Sablok, G., Budak, H., and Ralph, P. J. (2017). *Brachypodium genomics. Methods and Protocols.* New York, NY : Springer Science.

Sade, N., Weng, F., Tajima, H., Zeron, Y., Zhang, L., Wilhelmi, M. D. M. R., Day, G., Peleg, Z., and Blumwald, E. (2020). A cytoplasmic receptor-like kinase contributes to salinity tolerance. *Plants* 9, 1–12.

Sahoo, S., Das, S. S., and Rakshit, R. (2019). Codon usage pattern and predicted gene expression in *Arabidopsis thaliana*. *Gene: X* 2, 100012.

Salinas, J., and Sanchez-Serrano, J. (2006). Arabidopsis Protocols. Totowa, NJ: Humana Press.

Sallmyr, A., Tomkinson, A. E., and Rassool, F. V (2008). Up-regulation of WRN and DNA ligase lllά in chronic myeloid leukemia: Consequences for the repair of DNA double-strand breaks. *Blood* 112, 1413–1423.

Sartori, A. A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J., and Jackson, S. P. (2007). Human CtIP promotes DNA end resection. *Nature* 450, 509–514.

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., Kato, T., Nakao, M., et al. (2008). Genome structure of the legume, *Lotus japonicus*. *DNA Research* 15, 227–239.

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., and Tabata, S. (1999). Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Research* 6, 283–290.

Sato, S., and Tabata, S. (2006). *Lotus japonicus* as a platform for legume research. *Current Opinion in Plant Biology* 9, 128–132.

Sauer, N. J., Mozoruk, J., Miller, R. B., Warburg, Z. J., Walker, K. A., Beetham, P. R., Schöpke, C. R., and Gocal, G. F. W. (2016a). Oligonucleotide-directed mutagenesis for precision gene editing. *Plant Biotechnology Journal* 14, 496–502.

Sauer, N. J., Narváez-Vásquez, J., Mozoruk, J., Miller, R. B., Warburg, Z. J., Woodward, M. J., Mihiret, Y. A., Lincoln, T. A., Segami, R. E., Sanders, S. L., Walker, K. A., Beetham, P. R., Schöpke, C. R., and Gocal, G. F. W. (2016b). Oligonucleotide-mediated genome editing provides precision and function to engineered nucleases and antibiotics in plants. *Plant Physiology* 170, 1917–1928.

Schillaci, M., Arsova, B., Walker, R., Smith, P. M. C., Nagel, K. A., Roessner, U., and Watt, M. (2021). Time-resolution of the shoot and root growth of the model cereal *Brachypodium* in response to inoculation with *Azospirillum* bacteria at low phosphorus and temperature. *Plant Growth Regulation* 93, 149–162.

Schindele, P. and Puchta, H. (2020) Engineering CRISPR/LbCas12a for highly efficient, temperature-tolerant plant gene editing. *Plant Biotechnology Journal* 18, 1118–1120.

Scholthof, K. B. G., Irigoyen, S., Catalan, P., and Mandadi, K. K. (2018). *Brachypodium*: A monocot grass

model genus for plant biology. *Plant Cell* 30, 1673–1694.

Scholthof, K. B. G. (2020). *Brachypodium* and plant viruses: entwined tools for discovery. *New Phytologist* 227, 1676–1680.

Schwarz-Sommer, Z., Davies, B., and Hudson, A. (2003). An everlasting pioneer: The story of Antirrhinum research. *Nature Reviews Genetics* 4, 655–664.

Segal, D. J., Crotty, J. W., Bhakta, M. S., Barbas, C. F., and Horton, N. C. (2006). Structure of Aart, a designed six-finger zinc finger peptide, bound to DNA. *Journal of Molecular Biology* 363, 405–421.

Seligman, L. M., Chisholm, K. M., Chevalier, B. S., Chadsey, M. S., Edwards, S. T., Savage, J. H., and Veillet, A. L. (2002). Mutations altering the cleavage specificity of a homing endonuclease. *Nucleic Acids Research* 30, 3870–3879.

Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B., Van Der Oost, J., Brouns, S. J. J., and Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proceedings of the National Academy of Sciences of the United States of America* 108, 10098–10103.

Sessions, A., Burke, E., Presting, G., Aux, G., McElver, J., Patton, D., et al. (2002). A high-throughput Arabidopsis reverse genetics system. *Plant Cell* 14, 2985–2994.

Settles, A. M., Holding, D. R., Tan, B. C., Latshaw, S. P., Liu, J., Suzuki, M., et al. (2007). Sequence-indexed mutations in maize using the UniformMu transposon-tagging population. *BMC Genomics* 8, 116.

Shan, Q., Zhang, Y., Chen, K., Zhang, K., and Gao, C. (2015). Creation of fragrant rice by targeted knockout of the *OsBADH2* gene using TALEN technology. *Plant Biotechnology Journal* 13, 791–800.

Shaw, A. J., Schmutz, J., Devos, N., Shu, S., Carrell, A. A., and Weston, D. J. (2016). The *Sphagnum* genome project. A new model for ecological and evolutionary genomics. *Advances in Botanical Research* 78, 167–187.

Shi, J., Gao, H., Wang, H., Lafitte, H. R., Archibald, R. L., Yang, M., Hakimi, S. M., Mo, H., and Habben, J. E. (2017). ARGOS8 variants generated by CRISPR/Cas9 improve maize grain yield under field drought stress conditions. *Plant Biotechnology Journal* 15, 207–216.

Shukla, V. K., Doyon, Y., Miller, J. C., Dekelver, R. C., Moehle, E. A., Worden, S. E., et al. (2009). Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature* 459, 437–441.

Shulaev, V., Sargent, D. J., Crowhurst, R. N., Mockler, T. C., Folkerts, O., Delcher, A. L., et al. (2011). The genome of woodland strawberry (*Fragaria vesca*). *Nature Genetics* 43, 109–116.

Sibout, R., Proost, S., Hansen, B. O., Vaid, N., Giorgi, F. M., Ho-Yue-Kuang, S., et al. (2017). Expression atlas and comparative coexpression network analyses reveal important genes involved in the formation of lignified cell wall in *Brachypodium distachyon*. *New Phytologist* 215, 1009–1025.

Silva, G. H., Belfort, M., Wende, W., and Pingoud, A. (2006). From monomeric to homodimeric endonucleases and back: engineering novel specificity of LAGLIDADG enzymes. *Journal of Molecular Biology* 361, 744–754.

Silva, G., Poirot, L., Galetto, R., Smith, J., Montoya, G., Duchateau, P., and Paques, F. (2011). Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Current Gene Therapy* 11, 11–27.

Simsek, D., Brunet, E., Wong, S. Y. W., Katyal, S., Gao, Y., McKinnon, P. J., Lou, J., Zhang, L., Li, J., Rebar, E. J., Gregory, P. D., Holmes, M. C., and Jasin, M. (2011). DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. *PLoS Genetics* 7, e1002080.

Singh, D., Sternberg, S. H., Fei, J., Doudna, J. A., and Ha, T. (2016). Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nature Communications* 7, 12778.

Singh, R., Ong-Abdullah, M., Low, E. T. L., Manaf, M. A. A., Rosli, R., Nookiah, R., et al. (2013). Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature* 500, 335–339.

Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science* 351, 84–88.

Smith, R. A., Cass, C. L., Petrik, D. L., Padmakshan, D., Ralph, J., Sedbrook, J. C., and Karlen, S. D. (2021). Stacking *AsFMT* overexpression with *BdPMT* loss of function enhances monolignol ferulate production in *Brachypodium distachyon*. *Plant Biotechnology Journal* 19, 1878–1886.

Somerville, C., and Koornneef, M. (2002). A fortunate choice: The history of *Arabidopsis* as a model plant. *Nature Reviews Genetics* 3, 883–889.

Soreng, R. J., Peterson, P. M., Romaschenko, K., Davidse, G., Zuloaga, F. O., Judziewicz, E. J., Filgueiras, T. S., Davis, J. I., and Morrone, O. (2015). A worldwide phylogenetic classification of the *Poaceae* (*Gramineae*). *Journal of Systematics and Evolution* 53, 117–137.

Soyk, S., Müller, N. A., Park, S. J., Schmalenbach, I., Jiang, K., Hayama, R., Zhang, L., Van Eck, J., Jiménez-Gómez, J. M., and Lippman, Z. B. (2017). Variation in the flowering gene *SELF PRUNING 5G* promotes day-neutrality and early yield in tomato. *Nature Genetics* 49, 162–168.

Steinert, J., Schiml, S., Fauser, F., and Puchta, H. (2015). Highly efficient heritable plant genome engineering using Cas9 orthologues from *Streptococcus thermophilus* and *Staphylococcus aureus*. *Plant Journal* 84, 1295–1305.

Stepper, P., Kungulovski, G., Jurkowska, R. Z., Chandra, T., Krueger, F., Reinhardt, R., Reik, W., Jeltsch, A., and Jurkowski, T. P. (2017). Efficient targeted DNA methylation with chimeric dCas9-Dnmt3a-Dnmt3L methyltransferase. *Nucleic Acids Research* 45, 1703–1713.

Sternberg, S. H., Lafrance, B., Kaplan, M., and Doudna, J. A. (2015). Conformational control of DNA target cleavage by CRISPR/Cas9. *Nature* 527, 110–113.

Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., and Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507, 62–67.

Stoddard, B. L. (2011). Homing endonucleases: From microbial genetic invaders to reagents for targeted DNA modification. *Structure* 19, 7–15.

Strable, J., and Scanlon, M. J. (2009). Maize (*Zea mays*): A Model Organism for Basic and Applied Research in Plant Biology. *Cold Spring Harbor Protocols* 2009, pdb.emo132-pdb.emo132.

Strain, E., Hass, B., and Banks, J. A. (2001). Characterization of mutations that feminize gametophytes of the fern *Ceratopteris*. *Genetics* 159, 1271–1281.

Sturzenegger, A., Burdova, K., Kanagaraj, R., Levikova, M., Pinto, C., Cejka, P., and Janscak, P. (2014). DNA2 cooperates with the WRN and BLM RecQ helicases to mediate long-range DNA end resection in human cells. *Journal of Biological Chemistry* 289, 27314–27326.

Stuttmann, J., Barthel, K., Martin, P., Ordon, J., Erickson, J. L., Herr, R., Ferik, F., Kretschmer, C., Berner, T., Keilwagen, J., Marillonnet, S. and Bonas U. (2021). Highly efficient multiplex editing: one-shot generation of 8× *Nicotiana benthamiana* and 12× *Arabidopsis* mutants. *The Plant Journal* 106, 8-22.

Sun, N., Liang, J., Abil, Z., and Zhao, H. (2012). Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. *Molecular BioSystems* 8, 1255–1263.

Sun, Y., Jiao, G., Liu, Z., Zhang, X., Li, J., Guo, X., Du, W., Du, J., Francis, F. F., Zhao, Y., and Xia, L. (2017). Generation of high-amylose rice through CRISPR/Cas9-mediated targeted mutagenesis of starch branching enzymes. *Frontiers in Plant Science* 8, 298.

Sun, Y., Zhang, X., Wu, C., He, Y., Ma, Y., Hou, H., Guo, X., Du, W., Zhao, Y., and Xia, L. (2016). Engineering herbicide-resistant rice plants through CRISPR/Cas9-mediated homologous recombination of acetolactate synthase. *Molecular Plant* 9, 628–631.

Sun, Z., Li, N., Huang, G., Xu, J., Pan, Y., Wang, Z., Tang, Q., Song, M., and Wang, X. (2013). Site-specific gene targeting using transcription activator-like effector (TALE)-based nuclease in *Brassica oleracea*. *Journal of Integrative Plant Biology* 55, 1092–1103.

Sussman, D., Chadsey, M., Fauce, S., Engel, A., Bruett, A., Monnat, R., Stoddard, B. L., and Seligman, L. M. (2004). Isolation and characterization of new homing endonuclease specificities at individual target site positions. *Journal of Molecular Biology* 342, 31–41.

Svitashev, S., Young, J. K., Schwartz, C., Gao, H., Falco, S. C., and Cigan, A. M. (2015). Targeted mutagenesis, precise gene editing, and site-specific gene insertion in maize using Cas9 and guide RNA. *Plant Physiology* 169, 931–945.

Symington, L. S., and Gautier, J. (2011). Double-strand break end resection and repair pathway choice. *Annual Review of Genetics* 45, 247–271.

Szabados, L., Kovács, I., Oberschall, A., Ábrahám, E., Kerekes, I., Zsigmond, L., Nagy, R., Alvarado, M., Krasovskaja, I., Gál, M., Berente, A., Rédei, G. P., Haim, A. Ben, and Koncz, C. (2002). Distribution of 1000 sequenced T-DNA tags in the Arabidopsis genome. *Plant Journal* 32, 233–242.

Szczelkun, M. D., Tikhomirova, M. S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V., and Seidel, R. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proceedings of the National Academy of Sciences of the United States of America* 111, 9798–9803.

Szövényi, P., Frangedakis, E., Ricca, M., Quandt, D., Wicke, S., and Langdale, J. A. (2015). Establishment of *Anthoceros agrestis* as a model species for studying the biology of hornworts. *BMC Plant Biology* 15, 98.

Tang, H., Krishnakumar, V., Bidwell, S., Rosen, B., Chan, A., Zhou, S., Gentzbittel, L., Childs, K. L., Yandell, M., Gundlach, H., Mayer, K. F., Schwartz, D. C., and Town, C. D. (2014). An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*. *BMC Genomics* 15, 312.

Tang, X., Lowder, L. G., Zhang, T., Malzahn, A. A., Zheng, X., Voytas, D. F., Zhong, Z., Chen, Y., Ren, Q., Li, Q., Kirkland, E. R., Zhang, Y., and Qi, Y. (2017). A CRISPR/Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nature Plants* 3, 17018.

Taylor, G. K., Petrucci, L. H., Lambert, A. R., Baxter, S. K., Jarjour, J., and Stoddard, B. L. (2012). LAHEDES: The LAGLIDADG homing endonuclease database and engineering server. *Nucleic Acids Research* 40, W110–W116.

The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815.

The Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485, 635–641.

Thole, V., Peraldi, A., Worland, B., Nicholson, P., Doonan, J. H., and Vain, P. (2012). T-DNA mutagenesis in *Brachypodium distachyon*. *Journal of Experimental Botany* 63, 567–576.

Thornton, L. E., Keren, N., Ohad, I., and Pakrasi, H. B. (2005). *Physcomitrella patens* and *Ceratodon purpureus*, mosses as model organisms in photosynthesis studies. *Photosynthesis Research* 83, 87–96.

Tian, S., Jiang, L., Cui, X., Zhang, J., Guo, S., Li, M., Zhang, H., Ren, Y., Gong, G., Zong, M., Liu, F., Chen, Q., and Xu, Y. (2018). Engineering herbicide-resistant watermelon variety through CRISPR/Cas9-mediated base-editing. *Plant Cell Reports* 37, 1353–1356.

Tian, S., Jiang, L., Gao, Q., Zhang, J., Zong, M., Zhang, H., Ren, Y., Guo, S., Gong, G., Liu, F., and Xu, Y. (2017). Efficient CRISPR/Cas9-based gene knockout in watermelon. *Plant Cell Reports* 36, 399–406.

Tran, Q.-G., Han, Y.-J., Hwang, O.-J., Hoang, Q. T. N., and Kim, J.-I. (2018). Exploring responses to light in the monocot model Plant, *Brachypodium distachyon*. *Korean Journal of Plant Resources* 31, 522–530.

Trujillo, K. M., Yuan, S. S. F., Lee, E. Y. H. P., and Sung, P. (1998). Nuclease activities in a complex of human recombination and DNA repair factors Rad50, Mre11, and p95. *Journal of Biological Chemistry* 273, 21447–21450.

Tsai, S. Q., Wyvekens, N., Khayter, C., Foden, J. A., Thapar, V., Reyon, D., Goodwin, M. J., Aryee, M. J., and Joung, J. K. (2014). Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nature Biotechnology* 32, 569–576.

Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313, 1596–1604.

Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S., and Gregory, P. D. (2010). Genome editing with engineered zinc finger nucleases. *Nature Reviews Genetics* 11, 636–646.

Vakulskas, C. A., Dever, D. P., Rettig, G. R., Turk, R., Jacobi, A. M., Collingwood, M. A., et al. (2018). A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nature Medicine* 24, 1216–1224.

Vanamee, É. S., Santagata, S., and Aggarwal, A. K. (2001). FokI requires two specific DNA sites for cleavage. *Journal of Molecular Biology* 309, 69–78.

Vandenbussche, M., Chambrier, P., Bento, S. R., and Morel, P. (2016). Petunia, your next supermodel? *Frontiers in Plant Science* 7, 72.

Vanneste, K., Sterck, L., Myburg, A. A., Van de Peer, Y., and Mizrachi, E. (2015). Horsetails are ancient polyploids: Evidence from *Equisetum giganteum*. *Plant Cell* 27, 1567–1578.

Verde, I., Abbott, A. G., Scalabrin, S., Jung, S., Shu, S., Marroni, F., et al. (2013). The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature Genetics* 45, 487–494.

Verelst, W., Bertolini, E., De Bodt, S., Vandepoele, K., Demeulenaere, M., Pè, M. E., and Inzé, D. (2013).

Molecular and physiological analysis of growth-limiting drought stress in *Brachypodium distachyon* leaves. *Molecular Plant* 6, 311–322.

Vogel, J. P., Garvin, D. F., Leong, O. M., and Hayden, D. M. (2006a). *Agrobacterium*-mediated transformation and inbred line development in the model grass *Brachypodium distachyon*. *Plant Cell, Tissue and Organ Culture* 84, 199–211.

Vogel, J. P., Garvin, D. F., Mockler, T. C., Schmutz, J., Rokhsar, D., Bevan, M. W., et al. (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463, 763–768.

Vogel, J. P., Gu, Y. Q., Twigg, P., Lazo, G. R., Laudencia-Chingcuanco, D., Hayden, D. M., Donze, T. J., Vivian, L. A., Stamova, B., and Coleman-Derr, D. (2006b). EST sequencing and phylogenetic analysis of the model grass *Brachypodium distachyon*. *Theoretical and Applied Genetics* 113, 186–195.

Vojta, A., Dobrinic, P., Tadic, V., Bockor, L., Korac, P., Julg, B., Klasic, M., and Zoldos, V. (2016). Repurposing the CRISPR/Cas9 system for targeted DNA methylation. *Nucleic Acids Research* 44, 5615–5628.

Voytas, D. F. (2013). Plant genome engineering with sequence-specific nucleases. *Annual Review of Plant Biology* 64, 327–350.

Walker, J. R., Corpina, R. A., and Goldberg, J. (2001). Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* 412, 607–614.

Waltz, E. (2016). Gene-edited CRISPR mushroom escapes US regulation. *Nature* 532, 293.

Wang, W., Wu, Y., Shi, R., Sun, M., Li, Q., Zhang, G., Wu, J., Wang, Y., and Wang, W. (2020a). Overexpression of wheat α-mannosidase gene *TaMP* impairs salt tolerance in transgenic *Brachypodium distachyon*. *Plant Cell Reports* 39, 653–667.

Wang, G., Yu, G., Hao, Y., Cheng, X., Zhao, J., Sun, S., and Wang, H. (2020b). Molecular dissection of *TaLTP1* promoter reveals functional Cis-elements regulating epidermis-specific expression. *International Journal of Molecular Science*s 21, 2261.

Wang, N., Song, N., Tang, Z., Wang, X., Kang, Z., Dai, L., and Wang, B. (2020c). Constitutive Expression of Arabidopsis Senescence Associated Gene 101 in *Brachypodium distachyon* Enhances Resistance to *Puccinia brachypodii* and *Magnaporthe oryzae*. *Plants* 9, 1316.

Wang, F., Wang, C., Liu, P., Lei, C., Hao, W., Gao, Y., Liu, Y.-G., and Zhao, K. (2016). Enhanced rice blast resistance by CRISPR/Cas9-targeted mutagenesis of the ERF transcription factor gene *OsERF922*. *PloS ONE* 11, e0154027.

Wang, H., Rosidi, B., Perrault, R., Wang, M., Zhang, L., Windhofer, F., and Iliakis, G. (2005). DNA ligase III as a candidate component of backup pathways of nonhomologous end joining. *Cancer Research* 65, 4020–4030.

Wang, M., Wu, W., Wu, W., Rosidi, B., Zhang, L., Wang, H., and Iliakis, G. (2006). PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. *Nucleic Acids Research* 34, 6170–6182.

Wang, M., Wang, Z., Mao, Y., Lu, Y., Yang, R., Tao, X., Zhu, J.K. (2019). Optimizing base editors for improved efficiency and expanded editing scope in rice. Plant Biotechnology Journal 17, 1697–1699.

Wang, Z., Wang, S., Li, D., Zhang, Q., Li, L., Zhong, C., Liu, Y., & Huang, H. (2018a). Optimized paired-sgRNA/Cas9 cloning and expression cassette triggers high-efficiency multiplex genome editing in kiwifruit. *Plant Biotechnology Journal* 16, 1424.

Wang, P., Zhang, J., Sun, L., Ma, Y., Xu, J., Liang, S., Deng, J., Tan, J., Zhang, Q., Tu, L., Daniell, H., Jin, S., and Zhang, X. (2018b). High efficient multisites genome editing in allotetraploid cotton (*Gossypium hirsutum*) using CRISPR/Cas9 system. *Plant Biotechnology Journal* 16, 137–150.

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., and Qiu, J. L. (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nature Biotechnology* 32, 947–951.

Wang, Z., Wu, X., Ren, Q., Chang, X., Li, R., and Jing, R. (2010). QTL mapping for developmental behavior of plant height in wheat (*Triticum aestivum* L.). *Euphytica* 174, 447–458.

Watson, A., Ghosh, S., Williams, M. J., Cuddy, W. S., Simmonds, J., Rey, M. D., et al. (2018). Speed breeding is a powerful tool to accelerate crop research and breeding. *Nature Plants* 4, 23–29.

Watt, M., Schneebeli, K., Dong, P., and Wilson, I. W. (2009). The shoot and root growth of Brachypodium and its potential as a model for wheat and other cereal crops. *Functional Plant Biology* 36, 960–969.

Wendt, T., Holm, P. B., Starker, C. G., Christian, M., Voytas, D. F., Brinch-Pedersen, H., and Holme, I. B. (2013). TAL effector nucleases induce mutations at a pre-selected location in the genome of primary barley transformants. *Plant Molecular Biology* 83, 279–285.

Weng, X., Wang, L., Wang, J., Hu, Y., Du, H., Xu, C., Xing, Y., Li, X., Xiao, J., and Zhang, Q. (2014). *Grain Number, Plant Height, and Heading Date7* is a central regulator of growth, development, and stress response. *Plant Physiology* 164, 735–747.

Westfall, P. J., Pitera, D. J., Lenihan, J. R., Eng, D., Woolard, F. X., Regentin, R., et al. (2012). Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proceedings of the National Academy of Sciences of the United States of America* 109, E111–E118.

Wiedenheft, B., Van Duijn, E., Bultema, J., Waghmare, S., Zhou, K., Barendregt, A., Westphal, W., Heck, A., Boekema, E., Dickman, M., and Doudna, J. A. (2011). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proceedings of the National Academy of Sciences of the United States of America* 108, 10092–10097.

Williams-Carrier, R., Stiffler, N., Belcher, S., Kroeger, T., Stern, D. B., Monde, R. A., Coalter, R., and Barkan, A. (2010). Use of Illumina sequencing to identify transposon insertions underlying mutant phenotypes in high-copy Mutator lines of maize. *Plant Journal* 63, 167–177.

Williams, R. S., Moncalian, G., Williams, J. S., Yamada, Y., Limbo, O., Shin, D. S., Groocock, L. M., Cahill, D., Hitomi, C., Guenther, G., Moiani, D., Carney, J. P., Russell, P., and Tainer, J. A. (2008). Mre11 dimers coordinate DNA end bridging and nuclease processing in double-strand-break repair. *Cell* 135, 97–109.

Wilson, F., Harrison, K., Armitage, A. D., Simkin, A. J., & Harrison, R. J. (2018). CRISPR/Cas9-mediated mutagenesis of phytoene desaturase in diploid and octoploid strawberry. *Plant Methods* 15, 45.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G. V., and Provart, N. J. (2007). An "electronic fluorescent pictograph" Browser for exploring and analyzing large-scale biological data sets. *PLoS ONE* 2, 1–12.

Wolfe, S. A., Grant, R. A., Elrod-Erickson, M., and Pabo, C. O. (2001). Beyond the "recognition code": Structures of two Cys2His2 zinc finger/TATA box complexes. *Structure* 9, 717–723.

Wolfe, S. A., Nekludova, L., and Pabo, C. O. (2000). DNA recognition by Cys2His2 zinc finger proteins. *Annual Review of Biophysics and Biomolecular Structure* 29, 183–212.

Wolt, J. D., Wang, K., Sashital, D., and Lawrence-Dill, C. J. (2016). Achieving plant CRISPR targeting that limits off-target effects. *The Plant Genome* 9, 1–8.

Wong, G. K. S., Wang, J., Tao, L., Tan, J., Zhang, J., Passey, D. A., and Yu, J. (2002). Compositional gradients in *Gramineae* genes. *Genome Research* 12, 851–856.

Woo, J. W., Kim, J., Kwon, S. Il, Corvalán, C., Cho, S. W., Kim, H., Kim, S. G., Kim, S. T., Choe, S., and Kim, J. S. (2015). DNA-free genome editing in plants with preassembled CRISPR/Cas9 ribonucleoproteins. *Nature Biotechnology* 33, 1162–1164.

Wright, A. V., Sternberg, S. H., Taylor, D. W., Staahl, B. T., Bardales, J. A., Kornfeld, J. E., and Doudna, J. A. (2015). Rational design of a split-Cas9 enzyme complex. *Proceedings of the National Academy of Sciences of the United States of America* 112, 2984–2989.

Wu, C. A., Lowry, D. B., Cooley, A. M., Wright, K. M., Lee, Y. W., and Willis, J. H. (2008a). *Mimulus* is an emerging model system for the integration of ecological and genomic studies. *Heredity* 100, 220–230.

Wu, Y., Wang, W., Li, Q., Zhang, G., Zhao, X., Li, G., Li, Y., Wang, Y., and Wang, W. (2020). The wheat E3 ligase *TaPUB26* is a negative regulator in response to salt stress in transgenic *Brachypodium distachyon*. *Plant Science* 294, 110441.

Wu, D., Topper, L. M., and Wilson, T. E. (2008b). Recruitment and dissociation of nonhomologous end joining proteins at a DNA double-strand break in *Saccharomyces cerevisiae*. *Genetics* 178, 1237–1249.

Wu, J., Chen, C., Xian, G., Liu, D., Lin, L., Yin, S., Sun, Q., Fang, Y., Zhang, H., and Wang, Y. (2020). Engineering herbicide-resistant oilseed rape by CRISPR/Cas9-mediated cytosine base-editing. *Plant Biotechnology Journal* 18, 1857–1859

Wullschleger, S. D., Weston, D. J., Difazio, S. P., and Tuskan, G. A. (2013). Revisiting the sequencing of the first tree genome: *Populus trichocarpa*. *Tree Physiology* 33, 357–364.

Xiong, T., Meister, G. E., Workman, R. E., Kato, N. C., Spellberg, M. J., Turker, F., Timp, W., Ostermeier, M., and Novina, C. D. (2017). Targeted DNA methylation in human cells using engineered dCas9-methyltransferases. *Scientific Reports* 7, 6732.

Xu, R., Qin, R., Li, H., Li, D., Li, L., Wei, P., and Yang, J. (2017). Generation of targeted mutant rice using a CRISPR/Cpf1 system. *Plant Biotechnology Journal* 15, 713–717.

Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., Ni, P., et al. (2011). Genome sequence and analysis of the tuber crop potato. *Nature* 475, 189–195.

Yamada, M., Watanabe, Y., Gootenberg, J. S., Hirano, H., Ran, F. A., Nakane, T., Ishitani, R., Zhang, F., Nishimasu, H., and Nureki, O. (2017). Crystal structure of the minimal Cas9 from *Campylobacter jejuni* reveals the molecular diversity in the CRISPR/Cas9 systems. *Molecular Cell* 65, 1109-1121.e3.

Yan, F., Kuang, Y., Ren, B., Wang, J., Zhang, D., Lin, H., Yang, B., Zhou, X., and Zhou, H. (2018). Highly efficient A·T to G·C base editing by Cas9n-guided tRNA adenosine deaminase in rice. *Molecular Plant* 11, 631–634.

Yaneva, M., Kowalewski, T., and Lieber, M. R. (1997). Interaction of DNA-dependent protein kinase with DNA and with Ku: Biochemical and atomic-force microscopy studies. *EMBO Journal* 16, 5098–5112.

Yang, D., Scavuzzo, M. A., Chmielowiec, J., Sharp, R., Bajic, A., and Borowiak, M. (2016). Enrichment of G2/M cell cycle phase in human pluripotent stem cells enhances HDR-mediated gene repair with

customizable endonucleases. *Scientific Reports* 6, 21264.

Yang, H., Chang, F., You, C., Cui, J., Zhu, G., Wang, L., Zheng, Y., Qi, J., and Ma, H. (2015). Whole-genome DNA methylation patterns and complex associations with gene structure and expression during flower development in Arabidopsis. *The Plant Journal* 81, 268–281.

Yang, R., Jarvis, D. E., Chen, H., Beilstein, M. A., Grimwood, J., Jenkins, J., Shu, S. Q., Prochnik, S., Xin, M., Ma, C., Schmutz, J., Wing, R. A., Mitchell-Olds, T., Schumaker, K. S., and Wang, X. (2013). The reference genome of the halophytic plant *Eutrema salsugineum*. *Frontiers in Plant Science* 4, 46.

Yoon, J. S., and Seo, Y. W. (2021). *BdERF96* interacts with *BdASR1* to specifically respond to drought and oxidative stress in *Brachypodium distachyon*. *Journal of Plant Biochemistry and Biotechnology* 30, 287–296.

Yoon, J. S., Kim, J. Y., Kim, D. Y., and Seo, Y. W. (2021). A novel wheat ASR gene, *TaASR2D*, enhances drought tolerance in *Brachypodium distachyon*. *Plant Physiology and Biochemistry* 159, 400–414.

Yin, X., Biswal, A. K., Dionora, J., Perdigon, K. M., Balahadia, C. P., Mazumdar, S., Chater, C., Lin, H. C., Coe, R. A., Kretzschmar, T., Gray, J. E., Quick, P. W., and Bandyopadhyay, A. (2017). CRISPR/Cas9 and CRISPR/Cpf1 mediated targeting of a stomatal developmental gene *EPFL9* in rice. *Plant Cell Reports* 36, 745–757.

Young, N. D., Debellé, F., Oldroyd, G. E. D., Geurts, R., Cannon, S. B., Udvardi, M. K., et al. (2011). The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* 480, 520–524.

Yu, C., Liu, Y., Ma, T., Liu, K., Xu, S., Zhang, Y., Liu, H., La Russa, M., Xie, M., Ding, S., and Qi, L. S. (2015). Small molecules enhance CRISPR genome editing in pluripotent stem cells. *Cell Stem Cell* 16, 142–147.

Yu, S., Song, Z., Luo, J., Dai, Y., and Li, N. (2011). Over-expression of RAD51 or RAD54 but not RAD51/4 enhances extra-chromosomal homologous recombination in the human sarcoma (HT-1080) cell line. *Journal of Biotechnology* 154, 21–24.

Zaidi, M. A., O'Leary, S. J. B., Gagnon, C., Chabot, D., Wu, S., Hubbard, K., Tran, F., Sprott, D., Hassan, D., Vucurevich, T., Sheedy, C., Laroche, A., Gleddie, S., and Robert, L. S. (2020). A triticale tapetal non-specific lipid transfer protein (nsLTP) is translocated to the pollen cell wall. *Plant Cell Reports* 39, 1185–1197.

Zalatan, J. G., Lee, M. E., Almeida, R., Gilbert, L. A., Whitehead, E. H., La Russa, M., Tsai, J. C., Weissman, J. S., Dueber, J. E., Qi, L. S., and Lim, W. A. (2015). Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* 160, 339–350.

Zhang, H., Si, X., Ji, X., Fan, R., Liu, J., Chen, K., Wang, D., and Gao, C. (2018a). Genome editing of upstream open reading frames enables translational control in plants. *Nature Biotechnology* 36, 894–900.

Zhang, J., Li, C., Wu, C., Xiong, L., Chen, G., Zhang, Q., and Wang, S. (2006). RMD: a rice mutant database for functional analysis of the rice genome. *Nucleic Acids Research* 34, D745–D748.

Zhang, X., and Paull, T. T. (2005). The Mre11/Rad50/Xrs2 complex and non-homologous end-joining of incompatible ends in *S. cerevisiae*. *DNA Repair* 4, 1281–1294.

Zhang, Y., Bai, Y., Wu, G., Zou, S., Chen, Y., Gao, C., and Tang, D. (2017). Simultaneous modification of three homoeologs of *TaEDR1* by genome editing enhances powdery mildew resistance in wheat. *Plant Journal* 91, 714–724.

Zhang, Y., and Gao, C. (2017). Recent advances in DNA-free editing and precise base editing in plants.

Emerging Topics in Life Sciences 1, 161–168.

Zhang, Y., Li, D., Zhang, D., Zhao, X., Cao, X., Dong, L., Liu, J., Chen, K., Zhang, H., Gao, C., and Wang, D. (2018b). Analysis of the functions of *TaGW2* homoeologs in wheat grain weight and protein content traits. *Plant Journal* 94, 857–866.

Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., Qiu, J.-L., and Gao, C. (2016). Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nature Communications* 7, 12617.

Zhang, J.S., Zhou, Z.Y., Bai, J.J., Tao, X.P., Wang, L., Zhang, H., and Zhu, J.K. (2020). Disruption of MIR396e and MIR396f improves rice yield under nitrogen-deficient conditions. *National Science Review* 7, 102–112.

Zhang, J., Zhang, H., Botella, J.R., and Zhu, J.K. (2018c). Generation of new glutinous rice by CRISPR/Cas9-targeted mutagenesis of the *Waxy* gene in elite rice varieties. *Journal of Integrative Plant Biology* 60, 369–375.

Zhang, A. N., Liu, Y., Wang, F. M., Li, T. F., Chen, Z. H., Kong, D. Y., Bi, J. G., Zhang, F. Y., Luo, X. X., Wang, J. H., Tang, J. J., Yu, X. Q., Liu, G. L., and Luo, L. J. (2019). Enhanced rice salinity tolerance via CRISPR/Cas9-targeted mutagenesis of the *OsRR22* gene. *Molecular. Breeding* 39, 47.

Zhao, L., Bonocora, R. P., Shub, D. A., and Stoddard, B. L. (2007). The restriction fold turns to the dark side: A bacterial homing endonuclease with a PD-(D/E)-XK motif. *EMBO Journal* 26, 2432–2442.

Zhao, X., Wei, C., Li, J., Xing, P., Li, J., Zheng, S., and Chen, X. (2017). Cell cycle-dependent control of homologous recombination. *Acta Biochimica et Biophysica Sinica* 49, 655–668.

Zhou, J., Peng, Z., Long, J., Sosso, D., Liu, B., Eom, J. S., Huang, S., Liu, S., Vera Cruz, C., Frommer, W. B., White, F. F., and Yang, B. (2015). Gene targeting by the TAL effector PthXo2 reveals cryptic resistance gene for bacterial blight of rice. *Plant Journal* 82, 632–643.

Zhou, X. X., Fan, L. Z., Li, P., Shen, K., and Lin, M. Z. (2017). Optical control of cell signaling by single-chain photoswitchable kinases. *Science* 355, 836–842.

Zimin, A., Stevens, K. A., Crepeau, M. W., Holtz-Morris, A., Koriabine, M., Marçais, G., Puiu, D., Roberts, M., Wegrzyn, J. L., de Jong, P. J., Neale, D. B., Salzberg, S. L., Yorke, J. A., and Langley, C. H. (2014). Sequencing and assembly of the 22-Gb loblolly pine genome. *Genetics* 196, 875–890.

Zsögön, A., Čermák, T., Naves, E. R., Notini, M. M., Edel, K. H., Weinl, S., Freschi, L., Voytas, D. F., Kudla, J., and Peres, L. E. P. (2018). *De novo* domestication of wild tomato using genome editing. *Nature Biotechnology* 36, 1211–1216.

Chapter 2. Analysis of transformation methods for CRISPR/Cas9-based editing and screening of mutations.

Chapter 2. Analysis of transformation methods for CRISPR/Cas9-based editing and screening of mutations.

2.1 Introduction.

Given its status as a model crop, different techniques have been developed for the introduction of foreign DNA material into *Brachypodium distachyon*. These techniques differ depending on the mechanisms of DNA delivery, which will be consequent with the target tissue. Broadly, two different types of transformation mechanisms exist. Physicochemical methods will take advantage of the properties of the DNA and cell membrane and wall to deliver the DNA into the cell. These include techniques such as heat shock, electroporation, PEG-mediated transformation, microinjection, silicon carbide whiskers, or biolistic delivery. On the other hand, biological methods are based on the interaction of organisms to deliver the DNA. Examples of these are the use of *Agrobacterium* in plants or, to a certain extent, the use of viruses for the delivery of the desired DNA.

A different classification can be made depending on the ability of the DNA to remain in the cell and be passed on to further generations. Replication and vertical transference of the delivered DNA will require its integration into the host genome. On the other hand, if integration does not occur, the DNA can be expressed for a period of time but will be lost after subsequent cell divisions.

In this section, different techniques optimized for *Brachypodium* will be described, together with the different outcomes and the potential of these types of transformation for genome editing assays.

2.1.1. Different tissues are transformable using different techniques.

Classically, the objective of plant transformation, meaning the introduction of desired genetic material, has been to obtain a plant that carries a specific DNA insertion which will be stably inherited. In Arabidopsis, the floral dip technique has been the most widely used, given the simplicity of the protocol and the lack of tissue culture steps (Clough and Bent, 1998). The first protocol involving floral dip in cereals was described in wheat (Hess et al., 1990; Agarwal et al., 2009; Zale et al., 2009), and the first protocol involving floral dip in *Brachypodium* has already been registered (Peng and Zhaoqing, 2020) although its common use in the laboratory is yet to come.

In *Brachypodium*, one of the most common techniques uses embryogenic callus for transformation. This process starts with the isolation of an embryo, followed by the induction of embryogenic callus, transformation of the tissue, and plant regeneration. Friable embryogenic callus (FEC), which can be regenerated into a full plant, have been obtained from immature (Draper et al., 2001; Vogel and Hill, 2008) and mature embryos (Bablak et al., 1995; Jeon et al., 2010; Ozdemir and Budak, 2018; Yu et al., 2019). FEC induction relies on *in vitro* culture in specific media supplemented with the derived auxin 2,4-Dichlorophenoxyacetic acid (2,4-D, Bablak et al., 1995; Christiansen et al., 2005; Hammami et al., 2011; Lee et al., 2011). Callus starts appearing after a week of culture (Hammami et al., 2011). However, to obtain a larger quantity of FEC to transform, it can be grown for up to 9 weeks before it starts losing its regeneration potential (Păcurar et al., 2008; Betekhtin et al., 2018). It is also important to note that the callus should be periodically sub-divided and placed in fresh callus induction media to ensure callus keeps on developing. Depending on the specific protocol and the step of the process, sub-culturing steps can be done at different points in time (see previous references).

Regarding transformation of the FEC, different methods have been used. *Agrobacterium*-mediated transformation has been used, with several protocols described through the years. The first protocol, (Vogel et al., 2006), is based on the co-cultivation of *Agrobacterium* and pieces of FEC in callus induction media supplemented with acetosyringone to help transformation. After that, the transformed callus is transferred to a plate with timentin for a week to remove the *Agrobacterium* and later to a plate containing the selection marker desired, in this case hygromycin resistance, until the non-transformed callus was removed. The transformed pieces of callus are split and transferred

to a plate with the regeneration media. Different optimizations of the process have been proposed, such as the addition of $CuSO_4$ to the co-culture media or the drying of the callus after the co-culture, as proposed in Bragg et al. (2012), which led to the development of a T-DNA library. Alternative protocols, have also been described (Vunsh, 2018; Huize et al., 2020; Lihong et al., 2019; Xiaoquan & Suzhen, 2013; Xingrong, 2016). The main repository for *Brachypodium* resources, JGI, has also collected transformation protocols. Additional protocols that attempt to reduce the tissue culture requirements have also been developed, such as the microinjection-based protocol proposed by Zombori et al. (2011) or the co-cultivation of seeds (Fursova et al., 2012).

With the description of different protocols, key features of *Brachypodium* transformation have been detailed. Examples of these are the use of hygromycin for selection of transformed callus (Vogel et al., 2006) or the use of the maize Ubiquitin promoter for heterologous expression rather than the common CaMV 35S promoter (Lee et al., 2011; Bragg et al., 2012).

Another way that has been used to deliver DNA into embryogenic calli is the use of biolistic methods (Draper et al., 2001; Christiansen et al., 2005). In this physical method, the DNA is attached to gold microcarriers, using $CaCl_2$ and spermidine. Then, the gold is washed and placed on a macrocarrier. This is placed in the bombardment device, which places a rupture disc in line with the macrocarrier and the target tissue. Then, under high vacuum conditions, the rupture disc is subjected to increasing pressure using helium. When the rupture disc breaks due to the pressure, the helium accelerates the microcarriers towards the target tissue. To avoid physical damage caused by the macrocarrier colliding with the target, a protective mesh is placed in the projectiles path. After transformation, the process of obtaining the final product is equivalent to the one described in case of *Agrobacterium*-mediated transformation, except for the removal of the bacteria using timentin.

Electroporation has also been used to successfully generated stable transformants in maize after partial removal of the cell wall (D'Halluin et al., 1992; Pescitelli and Sukhapinda, 1995). An electrophoretic mechanism of transformation has also been used to deliver DNA into barley seed (Ahokas, 1989) or orchid protocorms (Griesbach, 1994). An *in planta* method involving microinjection into rye tillers has also been described (De La Peña et al., 1987). A different type of microinjection is the use of silica carbide whiskers to induce damage in callus tissue and inject the DNA. This approach has been used in cotton (Asad et al., 2008) and maize (Thompson et al., 1995) and different types of whiskers have been used in tobacco (Mizuno et al., 2005). However, these methods have not been described in *Brachypodium*.

The use of callus as a source of transformable plant material to obtain transgenic plants has been the most common tool for obtaining transgenic *Brachypodium*. However, the use of *in vitro* techniques for callus induction, maintenance and transformation makes these protocols complicated and laborious (Betekhtin et al., 2020).

Protoplasts are defined as plant cells whose cell wall has been removed (Fowke and Constabel, 1983). This removal is done by enzymatic digestion and leaves the cell membrane exposed, allowing transformation methods that would not be available if the cell wall is intact. The cell wall plays a role in the maintenance of osmotic pressure in the cell. Therefore, during the protoplast isolation process, the cells require a fine tuning of the osmotic pressure during the cell wall digestion, debris removal and the final solution they are in before the transformation. This is achieved by adding specific components that will preserve this osmotic pressure. The composition of the solutions differs from one species to another, but these principles remain the same.

In case of *Brachypodium*, two main types of protoplast isolation have been described. A sucrose-based protocol (Jung et al., 2014, 2015) and a mannitol based protocol (Hong et al., 2012). Both of them are based on the same principles, however, there are certain differences in the digestion solution composition. The sucrose-based protocol relies on the separation of the protoplasts using a sucrose gradient. This implies that after overlaying the digestion solution with the cleaning solution, a density gradient is formed, with the digestion solution at the bottom, the protoplasts at an interphase, and the cleaning solution on the top. On the other hand, protoplast separation in the mannitol-based protocols is done by centrifugal steps. For more details on these protocols and the advantages and disadvantages of each one of them, see materials and methods and discussion.

As previously introduced, the removal of the cell wall makes the use of different techniques for transformation possible. Examples of these are the case of electroporation (Fromm et al., 1985; Riggs and Bates, 1986; Zhang et al., 1988; Bates, 1989; Miao and Jiang, 2007) or heat shock (Göhring et al., 2014). Physical methods such as microinjection have also been used in barley (Holm et al., 2000) petunia (Griesbach, 1987) and tobacco protoplasts (Crossway et al., 1986). Also, other mechanisms such as bioactive beads (Sone et al., 2002; Liu et al., 2004; Murakawa et al., 2008a) and liposomes (Murakawa et al., 2008b) have been used to facilitate DNA delivery into tobacco protoplasts using PEG and electroporation techniques. However, the specific conditions for the application of these tools in *Brachypodium* have yet to be described. PEG mediated transformation is one of the most common mechanisms of protoplast transformation (Bart et al., 2006; Yoo et al., 2007; Mazarei et al., 2008) and was the transformation method of choice in the *Brachypodium* protocols previously outlined. Plant regeneration from *Brachypodium* protoplasts has yet to be described. However, other plants of the *Poaceae* group such as maize (Mitchell and Petolino, 1991) and rice (Abdullah et al., 1986) have been regenerated from isolated protoplasts.

Different types of transformations have also been used in plants, such as pollen tube transformation described in rice (Duan and Chen, 1985; Luo and Wu, 1989), cotton (Zhang et al., 2009), maize (Yang et al., 2009), and soybean (Liu et al., 2009), but have not been described in *Brachypodium*. Together with the previously described techniques, it is clear that transformation of *Brachypodium* is possible. However, certain techniques that have been described in other species are yet to be optimised in this model plant but have the potential of being applied to it.

2.1.2. Transient vs. stable transformants: different objectives, different outcomes.

The methods of introducing DNA into the plant cell have been described above. However, another factor to consider will be the fate of the delivered DNA. After delivery, the DNA can integrate in the genome of the host, generating a stable transformant. On the other hand, the DNA can remain in an extra-chromosomal location for some time. This can lead to transient expression of the delivered DNA. The fate of the delivered DNA will be determined by the insertion mechanism and the different features included in the sequence.

When the DNA is delivered as a circular plasmid, the integration process has been found to rely on concatenation of plasmid copies (Riggs and Bates, 1986). However, the insertion mechanism is still not fully understood.

In case of biolistic delivery, the plasmid can be delivered in its circular form. However, protocols using linearized DNA with the minimum cassettes required increased integration rates and made the integration patterns simpler. These minimal transgene cassettes have been used in wheat (Yao et al., 2006), maize (Lowe et al., 2009), rice (Fu et al., 2000), millet (O'Kennedy et al., 2011) and sugarcane (Jackson et al., 2013). Although biolistic delivery of DNA has been used in different species, it has been found to cause genomic rearrangements and overall damage to the genome (Liu et al., 2019). The

effects of these complications can lead to silencing of the delivered transgenes (Kohli et al., 1999; Sparks and Jones, 2004), aberrant expression (Srivastava et al., 1996; Stoger et al., 1998) and complete loss of the delivered DNA (Agrawal et al., 2005).

T-DNA insertion has been linked to DSB repair mechanisms, observing an increased integration when the DNA was damaged using X-ray radiation (Köhler et al., 1989) or homing endonucleases (Salomon and Puchta, 1998). Studies regarding the junctions between T-DNA and host DNA found patterns related to DSB repair mechanisms (Gheysen et al., 1991; Mayerhofer et al., 1991; Ohba et al., 1995; Gorbunova and Levy, 1997; Takano et al., 1997; Salomon and Puchta, 1998). These junction analyses also revealed that only a few homologies were found at the junction site, implying HDR was not involved in T-DNA insertion, but nonhomologous recombination and "illegitimate recombination" (see previous references). This implies that NHEJ mechanisms will be responsible for the insertion of T-DNA (Singer, 2018) and, by extension, delivered linear DNA. Regarding insertion rates, it has been observed that the frequency of integration is relatively low (Narasimhulu et al., 1996; Maximova et al., 1998; De Buck et al., 2000; Ghedira et al., 2013), which also fits with the description of NHEJ as the main mechanism for T-DNA insertion.

If the DNA is not integrated, it will remain in an extra-chromosomal location for a period of time. When the required genetic features (promoter, coding sequence and, if needed, terminator) are found in the delivered DNA, expression has been observed 3 hours after delivery, reaching a maximum level of expression between 18 and 48 hours (Abel and Theologis, 1994) and being able to be expressed for 10 days (Werr and Lörz, 1986). This has also been found after *Agrobacterium*-mediated transformation when integration does not occur (Johansen and Carrington, 2001).

In summary, different techniques have been developed for the introduction of DNA in *Brachypodium distachyon*. However, the fate of this DNA in terms of integration or transient expression is not easily predictable. Depending on the purpose of the transformation, different techniques can be used and different outcomes can be expected. Also, as the literature suggests, other methods that have not been described for this plant but for others can be optimised and applied. This implies that *Brachypodium* transformation is possible, but the optimization of some methods and implementation of other tools will be of interest in both research and crop development.

2.1.3. Screening methods.

Transgenic plants carrying the editing machinery will stably express the required genes and pass them onto further generations. Given the nature of genome editing, which relies on protein function to modify the genomic DNA, once editing has occurred there would be no need for the editing machinery to be in the system. The alterations in the genome induced by the editing machinery will be passed on to the following generations as part of the genomic DNA. This highlights the importance of transient expression regarding genome editing, since it would provide the desired modifications without having to count on random insertion of the editing machinery in the genomic DNA. Different techniques have been described for the detection of edits. These techniques may be dependent on the tissue transformed, the efficiency of the transformation and the efficiency of the editing itself. Many different tissues and techniques have been used for the introduction of the editing machinery in plants, as well as different methods for screening possible edits (see Supplementary Table 2).

The effects of the genome editing machinery will be observed in the DNA sequence. Analysing directly the genomic DNA using sequencing methods is an option. Sanger sequencing could be used, comparing the sequence of the treated tissue with non-edited tissue and observing the differences. Tools such as "Tracking Indels by Decomposition" (TIDE) allow the amplification of the DNA and analysis of aberrant signals in the sequence to identify indels and their significance respect to the

target site (Brinkman et al., 2014). However, the quality of the PCR product and the sequencing reaction are critical for the TIDE analysis to work.

On the other hand, Deep Sequencing techniques are able to use a pool of DNA containing both the edited and non-edited amplicons. The sensitivity of these Next Generation Sequencing (NGS)-based tools allows finding rare amplicons, saving the time required for cloning and single copy sequencing. However, the technique is error prone and costly when compared to others (Mali et al., 2013; Tsai et al., 2014). The data obtained from NGS sequencing can be analysed using different online tools (Güell et al., 2014), AGEseq (Xue and Tsai, 2015), BATCH-GE (Boel et al., 2016), CrispRVariants (Lindsay et al., 2016), CRISPResso (Pinello et al., 2016) or Cas-analyzer (Park et al., 2017).

After any editing experiment, it will be advisable to sequence the target region to determine the result of said editing. However, given the costs related to sequencing large numbers of individuals, it is useful to know which specimen have been edited prior to obtaining the sequence. A set of pre-screening methods have been proposed to successfully detect edits prior to knowing the actual edited sequence of the target gene.

A simple technique based on PCR amplification is the Annealing at Critical Temperature PCR (ACT-PCR, Hua et al., 2017). This technique is based on the differences in primer annealing temperature given mismatches in the primer sequence. With higher temperatures, primer annealing will not tolerate mismatches that lower temperatures would allow. If the sgRNA targets the 3' end of the primer, which is more sensitive to mismatches, the PCR will only amplify the non-edited DNA, allowing the finding of modifications in the sequence.

A different property of the mismatch is the ability of ssDNA to migrate in a non-denaturing electrophoresis, for example Polyacrylamide Gel Electrophoresis (PAGE) given conformational changes (Dong and Zhu, 2005). This is the basis of PCR single stranded conformational polymorphism (SSCP, Hayashi and Yandell, 1993; Inazuka *et al.*, 1997).

This property is also shared by dsDNA with mismatches. By denaturing the DNA and renaturing it, dsDNA will be formed carrying the mismatches. The migration differences found in Denaturing High Performance Liquid Chromatography (DHPLC, Underhill et al., 1997; Yu et al., 2006) can be used to find edits. PAGE has also been used using the same principle (Zhu et al., 2014).

Real-time PCR has also been used for screening edits. More specifically, melting analysis. This screening method is based on fluorescent probes intercalating into dsDNA. After PCR, a High-Resolution Melting (HRM) analysis is made (Wittwer et al., 2003). As the temperature increases beyond the melting point of dsDNA, differences in the sequence will change melting temperature. Detecting edits due to variations in dsDNA melting curves has been possible in zebrafish and plants (Thomas et al., 2014; D'Agostino et al., 2016; Samarut et al., 2016; Denbow et al., 2017).

Droplet Digital PCR (ddPCR) is able to detect DNA repair at the target site (Miyaoka et al., 2016a, 2016b). With a single PCR reaction, three different fluorescent probes will detect three different types of amplicon: one for a homologous sequence in the genome to find similar sequences to the target, another one to find the intact target sequence, and a third one that can be used for HDR tracking, recognizing the donor sequence.

Another sequence dependant feature that has been used for edit screening has been the use of restriction enzymes. These will cut the DNA in a sequence dependant manner, therefore, modifying a restriction site and blocking the recognition of a restriction enzyme is a simple method for edit screening (PCR-RE). The application of restriction enzymes also implies that the DNA can be pre-

treated to cut the non-edited DNA, leaving only the modified DNA (RE-PCR). This has proved useful for edit screening (Xie and Yang, 2013).

A different type of endonuclease, such as the T7 and Surveyor Nucleases, have been used in finding edits (Vouillot et al., 2015). These endonucleases are able to recognise mismatches rather than specific sequence patterns. These mismatches are created by denaturing and reannealing and detected using these endonucleases and running an agarose gel.

Each one of these screening methods previously described will come with advantages and disadvantages, regarding features such as costs, sensitivity, type of data provided and ability to describe guide efficiencies (Germini et al., 2018). Different advantages and disadvantages regarding screening choice will be discussed in following sections.

2.2. Hypothesis and objectives of the chapter.

The available protocols described for the introduction of genetic material in *Brachypodium distachyon*, are all based on tissue culture and a tissue dependant type of transformation. Tissue culture methods have been proven to be lengthy, sensitive and costly. In the context of genome editing experiments, the main objective of transformation has been to obtain a stable transformant carrying the editing machinery. We hypothesize that this is not compulsory for successful editing, and that transient expression in the tissues can be used with the purpose of rapid sgRNA validation. With this hypothesis in mind, different methods have been compared for the introduction of DNA in the system and the possibility of sgRNA validity screenings has been assessed.

2.3. Materials and Methods.

2.3.1. Plant material and growth conditions.

For the experiments regarding embryo size (measured as scutellum area) and percentage of embryos showing nuclear targeted GFP signalling after biolistic delivery, an inbred *Brachypodium distachyon* variety (Bd21) was used (de Souza et al. 2018). For the experiments regarding cell size and spot counts after biolistic delivery, in addition to protoplast isolations, a Bd21-3 line was used (Fuchs, 2019). Both varieties were sown and grown in the same conditions. Plants were sown in Levington M2 Compost and kept until germination at 21.5 °C with a short-day photoperiod. Once they germinated, they were transferred to a Polysec Chamber (Polysec Coldrooms, U.K.) and kept until needed at 21.5 °C with the same modified short-day photoperiod (8.5 hours of light and 15.5 hours of darkness).

2.3.2. Oligonucleotides and primers.

All oligonucleotides used were designed in the platform Benchling and obtained from Sigma-Aldrich (Germany). As recommended by the manufacturer, oligonucleotides were diluted to a concentration of 100 µm and kept at -20 °C. Further dilutions of these were done in Ambion Nuclease-Free Water Non DEPC Treated (Thermo Scientific, U.S.A.) and kept at -20 °C as well. For the specific sequences, see Supplementary Table 3.

2.3.3. Thermocompetent *Escherichia coli* preparation and transformation.

The thermocompetent *Escherichia coli* (from now on *E. coli*) used for plasmid amplification were prepared in the lab. A 5 ml culture of already thermocompetent *E.coli* was incubated for 16 hours at 37 ºC shaking at 180 rpm in a SI500 Shaking Incubator (Stuart Equipment, U.K.). All the *E. coli* cultures described were done using this incubator. This culture was refreshed in a 1:100 dilution and left growing at 37 °C and shaken at 180 rpm until the O.D. reached 0.55-0.6. Then, it was put in ice for 10 minutes and divided in 50 ml Falcon tubes. The cells were collected by centrifugation at 5000 rpm for 10 minutes at 4 °C in a Beckman Coulter Allegra X-12 Series Centrifuge (Beckman Coulter, U.S.A.) and resuspended gently in 20 ml of Tbf1 (Table 2.1). The resuspended cells were kept on ice for 5 minutes. After a second collection step using centrifugation at 5000 rpm for 5 minutes at 4 °C, the cells were resuspended gently in 2 ml of Tbf2 (Table 2.1). After incubating the tubes on ice for 15 minutes, 140 µl of DMSO were added to each 2 ml of cells. Finally, the cell solutions were divided in 100 µl aliquots, frozen in liquid nitrogen and stored at -80 °C until they were used.

Table 2.1. Competent cell preparation solutions composition.

Tbf1	Tbf2
30 mM CH_3CO_2K	10 mM MOPS Buffer
50 mM $MnCl_2 \cdot 4\ H_2O$	
100 mM KCl	10 mM KCl
10 mM $CaCl_2$	100 mM $CaCl_2$
15% Glycerol	
pH adjusted to 5.8 using CH_3CO_2H	pH adjusted to 6.5 using KOH

The transformation of *E. coli* was done using thermic shock. Firstly, one of the thermocompetent cell aliquots was thawed on ice. Then, 200 ng of the desired plasmid DNA (5 µl in case of a ligation) were added to the cells, which were kept on ice for 20-25 minutes before the thermic shock. For this, the cells were kept at 42 °C for 45 seconds. After the thermic shock the cells were put on ice for 2 minutes before adding 250 µl of SOC Medium (Fisher Scientific, U.S.A.) and incubating them for an hour at 37 °C shaking at 180 rpm using an incubator. After the incubation we plated 100 µl of the transformed

cells on a plate containing LB and the pertinent antibiotics. Ampicillin was used with a final concentration of 1 µg/ml in both liquid cultures and plates. Kanamycin was used in a 0.5 µg/ml concentration in both liquid cultures and plates.

2.3.4. Plasmids used in this chapter.

The different plasmids used were pActH2BGFP, pHvDGV and pDGV-H2BGFP. Plasmid maps were generated using SnapGene Viewer (Insightful Science, U.S.A.).

2.3.4.1. pActH2BGFP.

pActH2BGFP contains a GFP fused to the H2B *Arabidopsis thaliana* nuclear targeting signal guided by a rice Actin 1 promoter (OsAct1p). This promoter includes a 6xHis tag for antibody-based screening. The *A. thaliana* H2B-like coding sequence has been used successfully in wheat to target expression to the nucleus (Sparks and Jones, 2016). Upstream of the OsAct1p, the lac promoter and operator is placed to ensure that the transgene is not expressed in bacteria. At the same time, upstream of this sequence is a CAP binding site, which will activate the transcription of the lac operator in presence of cyclic Adenosyl Mono-Phosphate (cAMP). It also contains a ColE1 replication origin and an Ampicillin Resistance cassette whose expression is regulated by an AmpR promoter. However, this plasmid it is not designed to be used in conjunction with *Agrobacterium tumefaciens*, and lacked a plant antibiotic selection cassette, limiting its usefulness for producing stable transformants. However, its small size makes it useful for co-bombardment (Figure 2.1).

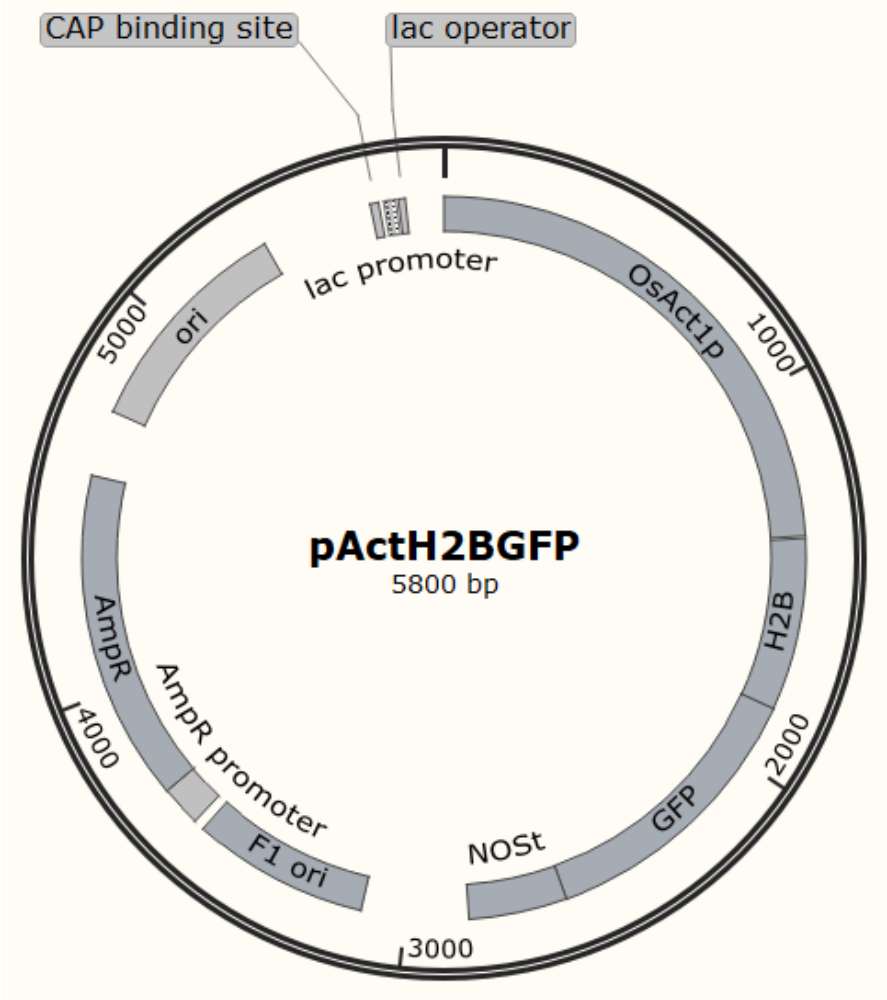


Figure 2.1. pAct H2B GFP Plasmid Map.
Map of the pActH2BGFP plasmid showing the key features contained in it.

2.3.4.2. pHvDGV.

pHvDGV was first described as an effective plasmid for genome editing in barley (Lawrenson et al., 2015). This plasmid contains the necessary elements for bacterial selection and replication, as well as a T-DNA containing the elements necessary for plant antibiotic selection and genome editing. It contains the *Streptococcus pyogenes* Cas9 gene under the transcriptional regulation of a maize Ubiquitin 1 promoter (ZmUbi1p), that contains the first intron of the gene to enhance gene expression. It also contains two independent wheat U6 promoters (TaU6p) to control the expression of two different guide RNA cassettes. To help guide insertion screening, which will be done by Golden Gate cloning, a LacZ is present in the first guide insertion site and an RFP present at the second insertion site, allowing colour screening of the guide insertion process (Figure 2.2). The process of sgRNA assembly will be further explained in Chapter 3.

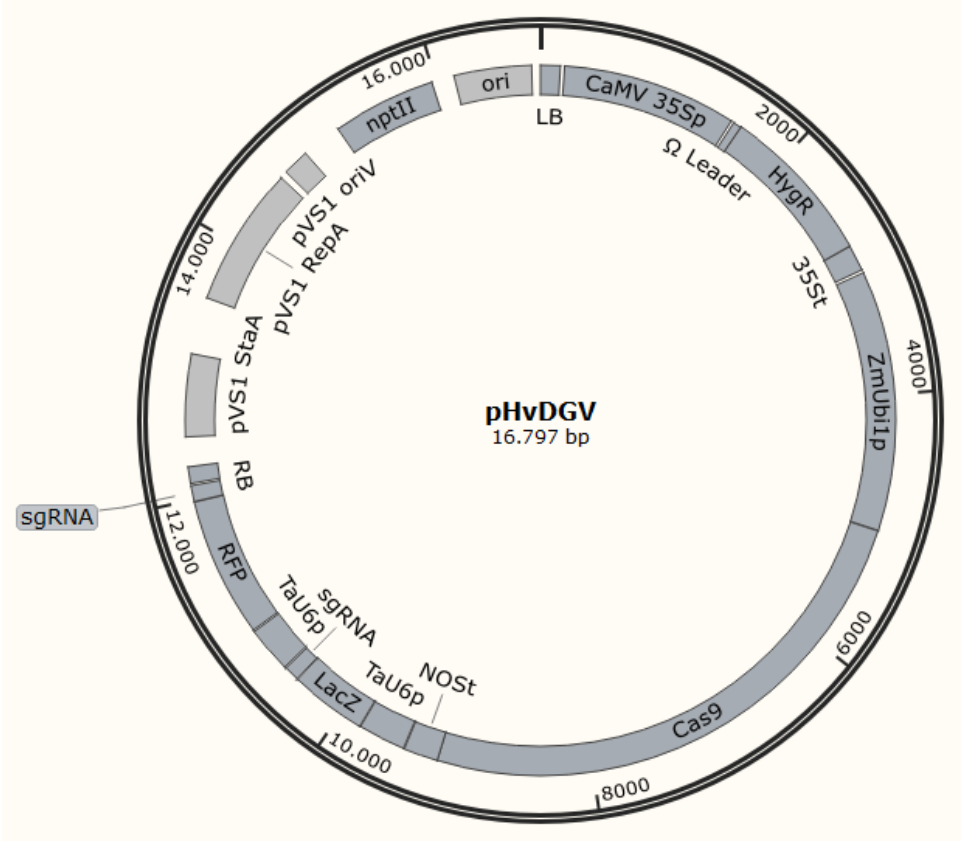


Figure 2.2. pHvDGV Plasmid Map.
Map of the pHvDGV plasmid showing its key features .

.

2.3.4.3. pDGV-H2BGFP.

To obtain a plasmid compatible with *Agrobacterium* mediated transformation that also possesses the simple screening provided by the nuclear targeted GFP in the pActH2BGFP and to investigate the efficacy of the ZmUbi1p in *Brachypodium,* the Cas9 gene was swapped for the H2BGFP, developing the pDGV-H2BGFP. The plasmid will contain the backbone of the pHvDGV, including all the elements previously described (Figure 2.3).

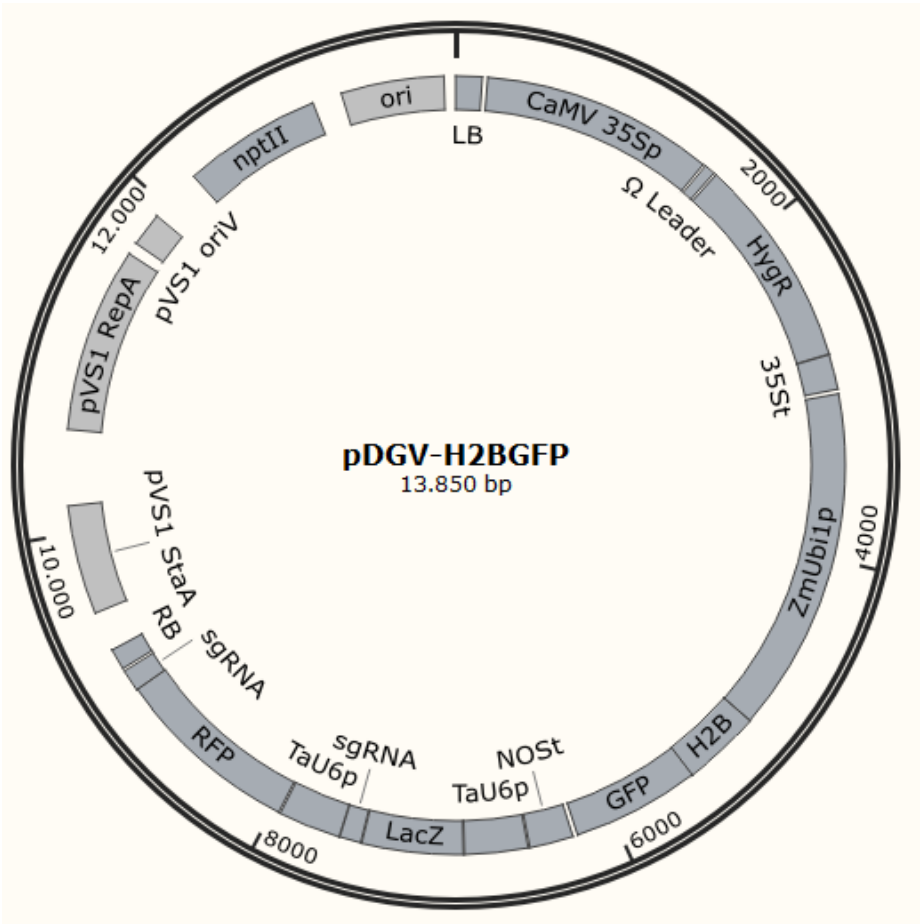


Figure 2.3. pDGV-H2BGFP Plasmid Map.
Plasmid map of the pDGV-H2BGFP showing the key features

PCR amplification of the H2BGFP region of pActH2BGFP used Phusion Green Hot Start II High-Fidelity PCR Master Mix (Thermo Scientific, U.S.A.) following the protocol provided by the manufacturer, using the pActH2BGFP Rv GG and pActH2BGFP Fw GG primers (see Supplementary Table 3). The HvDGV Ubi Rv GG and HvDGV NOS Fw GG primers were used to amplify the HvDGV backbone, excluding the Cas9 CDS (see Supplementary Table 3). The PCR reactions introduced overhangs allowing their subsequent assembly using Golden Gate Assembly. The PCR products were separated in a 1% agarose electrophoresis gel (Agarose from Bioline, U.K.) ran for 20 minutes in an electrophoretic tray (Embitec, U.S.A.) using TAE Buffer. Both amplicons were isolated from an agarose gel using the QIAQuick Gel Extraction Kit (Qiagen, Germany) following the protocol provided by the manufacturer. The purified products were used in a Golden Gate ligation using the T4 DNA Ligase and the restriction enzyme SapI (both from New England Biolabs, U.S.A.). The reaction was set up using 1 µl of T4 ligase, 1 µl of T4 ligase buffer, 0.5 µl of SapI restriction enzyme, 150 ng of the pDGV amplicon and 37.5 ng of the pActH2BGFP amplicon with a final volume of 10 µl. The cycles used for the Golden Gate ligation are shown in Table 2.2. Five µl of this ligation were transformed into thermocompetent *E. coli* via thermic shock.

Table 2.2. Cycles used for Golden Gate Ligation

37 °C	37 °C	16 °C	50 °C	80 °C
20s	3m	4m	5m	5m
26 Cycles				

The colonies containing the plasmid were screened using primers specific for the H2B-GFP (H2BGFP PCR Scr Fw and GFP qPCR Rv, see Supplementary Table 3) and the MyTaq™ Mix (Bioline, U.K.). To allow colony PCR screening, the protocol suggested by the manufacturer was modified, using a mix of MyTaq™ Mix (25 µl), nuclease free water (23 µl) and primers (1 µl of each primer at a 20 µM concentration). A plastic pipette tip was used to touch the *E. coli* colony of interest and the PCR mixture was mixed with that same pipette, mixing the colony and the PCR mixture. The PCR cycles were done as usual but increasing the first denaturation time to 10 minutes.

The positive colonies were inoculated in a 5 ml LB culture, grown over night and used for plasmid isolation using the QIAprep® Spin Miniprep Kit from Qiagen following the protocol suggested by the manufacturer with minor modifications. These plasmids were sequenced using primers contained in the HvDGV backbone aiming to amplify the H2B-GFP starting with the maize Ubi1 promoter or the NOS terminator (dCas9 Check Fw and dCas9 Check Rv, see Supplementary Table 3).

All sequencing reactions were done in-house on the ABI 3730 DNA analyser. In both plasmid and amplicon, the sequencing reactions were prepared with 100-250 ng of DNA, 1 µl of the desired primers in a 10 µM concentration in a final volume of 7 µl. The chromatogram analysis and export as a file in FASTA format was done using DNA Sequence Assembler (v4 (2013), Heracle BioSoft, www.DnaBaser.com).

2.3.5. Protoplast Isolations.

2.3.5.1. Mannitol-Based.

The protocol used was based on the ones described in Hong et al. (2012) and Shan et al. (2013) with minor modifications.

Plants were grown for 21-26 days as previously described. Before the isolation, the stock solutions were prepared. The Digestion Solution (0.6 M Mannitol, 10 mM $CaCl_2$, 20 mM KCl, 20 mM MES-KOH pH5.7) was sterilized using an autoclave. The Digestion Solution was kept at room temperature, and the enzymes were added prior to the isolation. The enzyme composition was 2% (w/v) Cellulase (Duchefa, Netherlands), 0.4% (w/v) Macerozyme (Duchefa, Netherlands) and 0.1% BSA (Sigma-Aldrich, Germany). The W5 solution contained 0.037% KCl (w/v), 0.9% NaCl (w/v), 1.84% $CaCl_2$(w/v), 2mM MES-KOH pH5.7. The solution was filter-sterilized and kept at room temperature.

For each protoplast isolation, primary leaf material from 3 different plants grown in the same pot, amounting to 1 g was cut and chopped in a sterile glass Petri dish, before adding 20 ml of Digestion Solution with the enzymes added. As specified in the protocols, no sterilising solution was applied to the leaves before the addition of the digestion solution. The solution was vacuum infiltrated into the chopped leaves for 30 minutes at -0.9 bar. After the vacuum infiltration, the digestion was incubated at 30 °C in the dark with gentle shaking at 35 rpm over-night. After the incubation, before starting the protoplast cleaning process, the solution was shaken at 50 rpm for 30 minutes. The solution was then filtered using a series of meshes of decreasing pore size (100 µm-64 µm-20 µm) into a 50 ml Falcon tube. These meshes were wetted with W5 solution before the filtering. After filtering the digestion solution, 10 ml of W5 solution were passed through the mesh. The protoplasts were cleaned twice by

centrifuging the mixture at 10 g at 4 °C for 20 minutes in a Beckman Coulter Allegra X-12 Series Centrifuge (Beckman Coulter, U.S.A). The final resuspension was done in 10 ml of W5 solution and the protoplasts were kept on ice for one hour to let them precipitate to the bottom of the tube. The supernatant was removed until there were 3 ml of protoplast solution and a sample was taken for the viability analysis.

2.3.5.2. Sucrose-Based.

The protocol used was based on the one described in Jung et al. (2015) with minor modifications.

Plants were grown as previously described for 21-26 days. Before the isolation, the stock solutions were prepared. The Digestion Solution (750 mM sucrose, 20 mM $CaCl_2$, 40 mM KCl, 20 mM MES-KOH pH 5.7) was filter sterilised, and the stock stored at -20^{o}C. Before use, the enzymes were added. The enzyme composition was 1.5% cellulase (Duchefa, Netherlands) and 1% macerozyme (Duchefa, Netherlands). The W5 solution contained 0.037% KCl (w/v), 0.9% NaCl (w/v), 1.84% $CaCl_2$ (w/v), 2mM MES-KOH pH 5.7. The solution was filter-sterilized and kept at room temperature.

For each protoplast isolation, 0.4 g of primary leaf material obtained from 3 different plants grown in the same pot was cut and immersed in digestion solution. As specified in the protocol, no sterilising solution was applied to the leaves before the addition of the digestion solution. The leaves were chopped into approximately 2 mm pieces using a pair of scalpels. The cut tissue was transferred to a 200 ml beaker with 20 ml of Digestion Solution. This was incubated during 1 h at 30 °C and then incubated in the dark for 18-20 h at room temperature with gentle shaking at 35 rpm. After the incubation, mixture was collected in a 50 ml Falcon centrifuge tube, sieving them through 8 layers of cheesecloth pre-wetted with W5 solution. After filtering the solution to remove the biggest debris, the cloth was washed again with 10 ml of W5 into the same Falcon tube. After the wash, the solution was overlaid with 5 ml of W5 solution by spreading the solution on the top of the digestion mix with a Pasteur pipette. This mix was left for 1 h at room temperature to allow the protoplasts to float to the interphase between the W5 solution and the digestion solution. After 1 h, the interphase was collected using a Pasteur pipette and transferred to another Falcon tube containing 20 ml of W5 solution. This mixture was centrifuged for 5 minutes at 100 g in a Beckman Coulter Allegra X-12 Series Centrifuge (Beckman Coulter, U.S.A). The residual enzyme solution was removed and the pellet was cleaned once using 10 ml of W5 solution. This was centrifuged for 5 minutes at 60g. After removing the supernatant, the protoplasts were resuspended in 3 ml of W5 solution. This final protoplast solution was analysed for viability tests using FDA staining.

2.3.6. FDA Staining of protoplasts.

The protocol followed for FDA staining of protoplasts for viability screenings was described in Huang et al. (1986). A stock solution of 5 mg/ml of Fluorescein Diacetate (FDA, Sigma-Aldrich, U.S.A.) in acetone was prepared. This stock was kept at -20 °C. Before the screenings, the working solution was prepared by diluting 20 µl of this stock in 1 ml of 0.65 M Mannitol. Then, 10 µl of this working solution were mixed with 10 µl of the final protoplast solution, mixing gently with a pipette. The mixture was put in a hemocytometer and observed under an epifluorescence microscope Leica DM6000B (Leica, Germany) for the viability counts. From each protoplast isolation, 2 measurements were made.

2.3.7. Embryo Isolation.

Immature seeds were taken from 3 different *B. distachyon* plants grown for around 70 days in the same pot. The pericarp was removed from the seed before the sterilization process. The seed were sterilized by submerging in 70 % ethanol for 1 minute and washed once with sterile water. This was followed by an incubation for 5 minutes in 7% calcium hypochlorite and 3 washes with sterile water. The seeds were placed in a sterile Petri dish and under a stereoscope the immature embryos were removed with forceps sterilized using a flame and 64 embryos were placed on a callus proliferation medium (BdCPM) plate in an 8x8 isometric grid in the centre of the plate within a 2 cm^2 square. This medium consisted of 4.3 g/l of Murashige and Skoog Salts, 30 g/l of sucrose 0.6 mg/l $CuSO_4$ and 2 g/l of Phytagel. Then the pH was adjusted to 5.8 with KOH and the volume adjusted to 990 ml. After autoclaving the medium, 2.5 mg/l of 2,4-dichlorophenoxyacetic acid (2,4-D) and 10 ml of M5 Vitamin mix (100x) containing 0.04 g Nicotinic Acid, 0.05 g of Thiamine-HCl, 4 g of Cysteine, 0.2 g of Glycine and 0.04 g of Pyridoxine-HCl with the pH adjusted to 8.5 using KOH. These embryos were placed in an Economy Incubator (Gallenkamp, Germany) at 25°C in the dark. From each pot of 3 plants, 3 different plates with 64 embryos each were prepared.

2.3.8. Biolistic delivery of DNA into embryos.

Embryos were bombarded using a Bio-Rad PDS-1000 (Bio-Rad, U.S.A.) following the protocol described in Sparks and Jones (2014) adapted for *B. distachyon*. Gold particle stocks were prepared by washing 20 mg of 0.6 µm gold microcarriers (Biorad, U.S.A.) with 1 ml of 100% ethanol. The gold particles were mixed with a vortex for 2 minutes before centrifuging for 5 s at full speed on a tabletop centrifuge and removing the supernatant. This was repeated three times. After the ethanol washes, the gold particles were washed with 1 ml of sterile distilled water following the same steps as the ethanol washes. Finally, the washed gold particles were resuspended in sterile distilled water with a final volume of 1 ml and aliquoted in 50 µl aliquots in 0.5 ml Eppendorf tubes. The aliquots were kept at -20 °C until they were used.

Prior to the bombardment, 25 µl 0.1 M spermidine aliquots were prepared from a 1 M stock solution in sterile water. Both stock and aliquots were kept at -20°C until the bombardment. Also, a 2.5 M $CaCl_2$ solution in sterile distilled water was prepared and kept at room temperature.

Since the suggested plasmid concentration for biolistic delivery was 1 µg/µl, the protocol combining the QIAPrep™ Spin Miniprep Kit from Qiagen and the protocol described in Kotchoni et al. (2003) was used. Starting with the same 5 ml cultures of *E. coli* containing the desired plasmid, they were grown for 16 hours at 37 °C and shaking at 180 rpm. Then, they were centrifuged using a Beckman Coulter Allegra X-12 Series Centrifuge (Beckman Coulter, U.S.A.). Then, they were resuspended in 250 µl of P1 buffer and transferred to a 1.5 ml Eppendorf Tube. After this, 250 µl of P2 buffer were added and mixed by inverting the tube 4-5 times. Then, the samples were kept at room temperature for 5 minutes, before adding 350 µl of N3 buffer. Then, the solutions were mixed and kept in ice for 10 minutes. To precipitate the debris, the samples were centrifuged for 10 minutes at 13000 rpm at 4 °C in a centrifuge and 650 µl of supernatant were transferred to a new tube. Then, 650 µl of isopropanol were added to the supernatant previously obtained and these were kept for an hour at -20 °C. After this, in order to precipitate the DNA, the samples were centrifuged for 20 minutes at 13000 rpm at 4 °C. After removing the supernatant, the pellet was washed twice with 500 µl of 70% ethanol. After removing the ethanol from the last wash, the samples were dried under the flow hood and the DNA was resuspended in the volume of Ambion Nuclease-Free Water Non DEPC Treated (Thermo Scientific, U.S.A.) necessary to obtain a 1 µg/µl concentration.

Before particle bombardment, one of the 50 µl aliquots of gold particles and one of the 25 µl 0.1 M spermidine aliquots were thawed on ice. The gold aliquot was resuspended with a vortex for 1 minute. 5 µg of the desired plasmid DNA were added to the gold particles and mixed with the vortex for 1 minute. Then, 50 µl of $CaCl_2$ and 20 µl of 0.1 M spermidine were added to the cap of the gold particle tube and mixed using a vortex. The gold was centrifuged for 5 seconds and the supernatant was removed. The gold was washed once in 150 µl of 100% ethanol, scraping the sides of the tube and applying the vortex to fully resuspend it. After the resuspension, it was centrifuged again at full speed for 5 seconds and the supernatant was removed. The gold particles were resuspended in 85 µl of 100% ethanol scraping the sides of the tube and using the vortex. The tube was closed, sealed with Parafilm and kept in ice until it was added to the macro-carriers. Macro-carriers and 650 psi rupture discs were obtained from Bio-Rad.

Prior to embryo bombardment, all surfaces of the gun and the laminar flow hood around it were sterilized using 100% ethanol. The Macro-carriers, rupture discs and equipment necessary for the preparation of each shot were sterilized using 100% ethanol as well before each bombardment. The settings of the gun were set up as described in Sparks and Jones (2014), and operated following the protocol described in it. After bombardment, the embryos were kept in the same incubator in the conditions previously described until they were screened using fluorescence microscopy.

2.3.9. Toluidine Blue Staining of *Brachypodium* embryos.

For cell size experiments, embryos were stained using Toluidine Blue. After embryo isolation the samples were fixed by incubating overnight in the dark at 4 °C in 100 µl of FAA fixation solution (10% Formaldehyde, 50% Ethanol and 5% Acetic Acid in distilled water). After that, the samples were decoloured using one wash with 75% Ethanol and another wash with 100% Ethanol. Then, an excess of Toluidine blue solution was added (0.01% Toluidine Blue in Phosphate Buffer 0.1 M pH 6.8) and the embryos were put under 0.9 bar of negative pressure for vacuum infiltration. Finally, embryos were washed three times with sterile distilled water and images of them were obtained using the Leica DM6000B (Leica, Germany) microscope. Images were analysed using the ImageJ software (N.I.H., U.S.A).

2.3.10. Fluorescence microscopy of transformed embryos.

In order to determine the efficiency of different transformation methods, fluorescent imaging was used. Two days after transformation, the embryos were placed on a microscope slide (Thermo Scientific, U.S.A.), observed and imaged using a Leica DM6000B (Leica, Germany) epifluorescence microscope. The images were analysed using the LAS-AF and LAS-X Software (Leica, Germany).

2.3.11. RE-PCR optimization.

To test the limits of the RE-PCR method for finding low-represented amplicons, an *in vitro* approach was used. Two different Hygromycin resistance cassettes (HygR) found in the pHvDGV and the pBract204 plasmids (Figure 2.4) were used. The HygR sequence in the pBract204 plasmid contains two different AdeI restriction sites, while the HygR in the pHvDGV does not have any. A diagram showing the relative locations of these restriction sites is shown in Figure 2.10A. By making a mixture of these two plasmids based on the plasmid copy number, it was possible to introduce known proportions of HygR sequences that differed in those AdeI restriction sites. The relation studied was 1:6000.

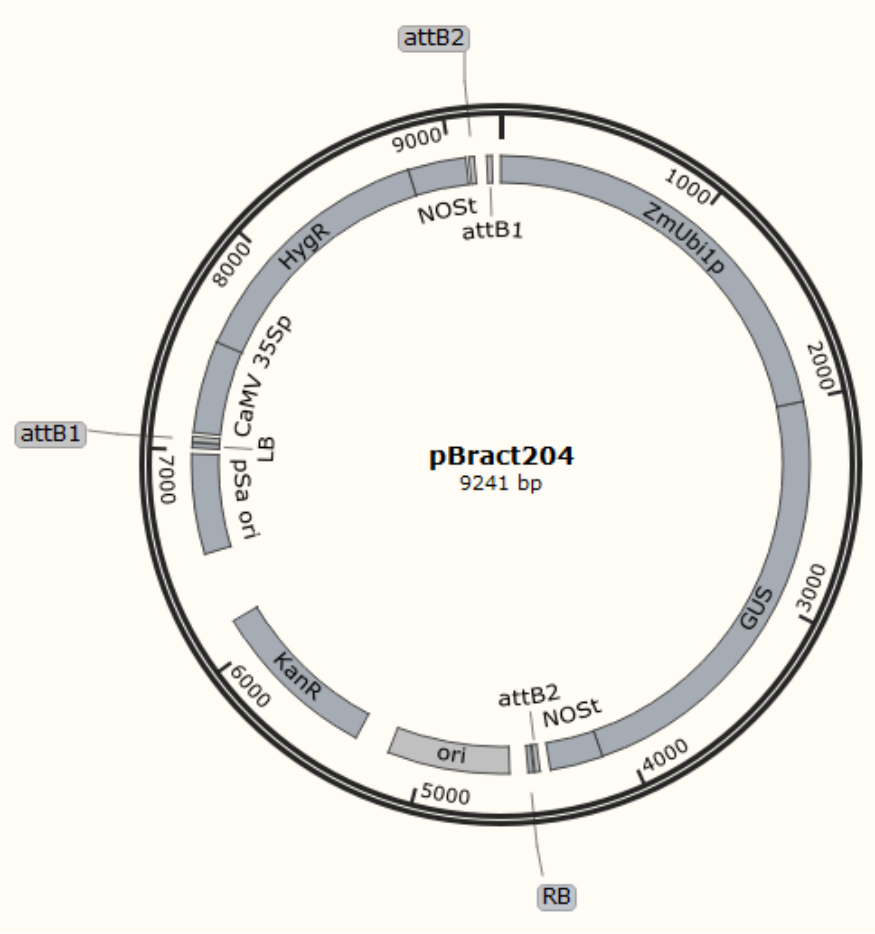


Figure 2.4. pBract204 Plasmid Map.
Plasmid map of the pBract204 showing the key features

First, a negative control where there was no restriction prior to the PCR reaction was performed. A PCR reaction was set up using the 1:6000 mixture of plasmids described previously in order to amplify the HygR sequence from both of them. The reaction was done using the MyTaq™ Mix from Bioline and the primers HygR CDS Fw and HygR CDS Rv (see Supplementary Table 3) following the recommendations of the manufacturer. To ensure enough DNA was obtained from the PCR reaction, 35 amplification cycles were applied. The PCR product obtained was separated using Agarose gel electrophoresis and the desired bands were isolated using the QIAQuick Gel Extraction Kit from Qiagen following the manual provided by the manufacturer. These bands were cloned into the pCR™2.1 plasmid provided in the TA Cloning™ Kit (Invitrogen, U.S.A.) following the instructions of the manufacturer. The ligation product was transformed into thermocompetent *E. coli* cells and plated on LB Agar plates with Kanamycin as the antibiotic resistance. The X-Gal/IPTG mixture was used for blue-white selection by supplementing the plate with an X-Gal (Melford, U.K.) 40 ml/ml and IPTG (Melford, U.K.) 40 mg/ml, which was spread using a sterile Digralsky spreader. After the over-night incubation, 10 white colonies were inoculated into different liquid LB Kanamycin cultures and after the over-night incubation the plasmids were isolated using the QIAprep® Spin Miniprep Kit from Qiagen. The plasmids were treated with AdeI restriction enzyme from Thermo Scientific following the instructions of the manufacturer. The restriction products were separated using Agarose Gel electrophoresis to observe if the amplicons cloned contained the restriction sites or they did not.

The RE-PCR assays started with the digestion of 200 ng of the 1:6000 plasmid dilution with AdeI, using 0.5 µl of the restriction enzyme in a final volume of 10µl. The restriction as done at 37 °C for 15 minutes with an inactivation step of 5 minutes at 65 °C. Then, the components required for PCR amplification of the HygR sequence, 25 µl of MyTaq™ Mix from Bioline, 13 µl of nuclease-free water and 1 µl of each primer in a 20 µM concentration (as specified by the manufacturer), were added. After the PCR reaction, the bands were separated using Agarose gel electrophoresis and the desired bands were isolated as previously described. Again, 200 ng of the DNA isolated from the gel were treated with AdeI following the same protocol. Some of the isolated DNA had to be saved for its cloning into the pCR™2.1 plasmid for the amplicon identification. After the digestion, the components necessary for the PCR amplification of the sequence of interest were added. Again, the PCR products were separated using electrophoretic techniques and the desired band was isolated as previously described and cloned into the pCR™2.1 plasmid for amplicon identification.

To ensure that the amplification happened after subsequent RE-PCR cycles, nested primers were designed. The first PCR amplification was done using the primers described in the negative control. The ones for the second RE-PCR reaction were HygR Nd Fw and HygR Nd Rv (See Supplementary Table 3).

2.3.12. Plotting and Statistical Analysis.

All graphical plots and statistical analysis performed in this chapter were done using GraphPad version 5.0.0 for Windows (GraphPad Software, San Diego, California USA, www.graphpad.com).

The Gaussian distribution analysis for embryo sizes in 10 week old plants was performed using the least squares fitting method using GraphPad, using the standard parameters from this tool.

The differences in scutellum area found in embryos right after the isolation and two days after the isolation in Bd21-3 and Bd21 were analysed using a non-parametric unpaired t-test, obtaining the P-values for a two-tailed test with 95% confidence intervals. The same analysis was used for the comparison of the cell areas found in the scutellum and axis of embryos.

The correlation between scutellum area and spot counts found after bombardment with pDGV-H2BGFP was analysed using a linear regression test using GraphPad, using the standard parameters from this tool.

2.4. Results.

2.4.1. Leaf Protoplasts as a potential platform for DNA transformation experiments.

Two different methods were used for protoplast isolation from leaf material. The first protocol used was based on centrifuge steps for protoplast separation after enzymatic digestion. Given this type of separation, mannitol was used to maintain osmotic pressure in the solution, therefore, this was named as "Mannitol-based Protocol". The protocol was based on Hong et al. (2012) and Shan et al. (2013), which isolated protoplast from *B. distachyon* leaves, taking in account other protocols described for other plants, such as rice (Shan et al., 2013; Jabnoune et al., 2015) or Arabidopsis (Zhai et al., 2009). The second protocol separated the protoplast using a sucrose gradient following the enzymatic digestion step, hence named "Sucrose-based Protocol". The protoplasts would be found at the interphase between the debris from the digestion solution and the cleaning solution. This protocol was taken from Jung et al. (2015).

The Mannitol-based Protocol successfully released protoplasts from leaf material (Figure 2.5A), which were shown to be viable after FDA staining (Figure 2.5B). When counting the total amount of protoplasts using a haemocytometer, the Mannitol-based Protocol showed high yields of total protoplasts, reaching a maximum of 3.6 million protoplasts per ml and averaging 2.3 million protoplasts per ml among all the replicates, with a minimum of 0.24 million protoplasts per ml. However, the viability of these protoplasts differed, showing a maximum of 82% and an average of 52.8% when putting together all the different replicates, having a minimum of 17% viable protoplasts (Figure 2.5C). As we observe from the results, even when the yield and viability within the technical replicates taken from the same biological replicate is similar, the variability of the yields and viability when observing the biological replicates independently implied a lack of consistency, which makes this isolation protocol less reliable as a source of easily transformable tissue.

The Sucrose-based Protocol was also successful in releasing alive protoplasts (Figure 2.5D, E). However, it yielded a lower number of protoplasts, obtaining a maximum of 1.45 million protoplasts per ml and averaging 0.86 million protoplasts per ml, with a minimum of 0.22 million protoplasts per ml. Nonetheless, the viability percentages that this type of isolation provided was higher than the ones obtained with the Mannitol-based protocol, obtaining a maximum of 96.4% and an average of 80.27% of viable protoplasts, and a minimum of 47% viability (Figure 2.5F). Like the Mannitol-based Protocol, the Sucrose-based Protocol was somewhat inconsistent in terms of protoplast density and viability.

In addition to these consistency issues, when kept over-night at room temperature, the viability of the protoplasts was reduced to 0% in both the Mannitol-based protocol (Figure 2.6A) and Sucrose-based protocol (Figure 2.6B). In contrast, a reduced protoplast viability was found when the protoplasts were kept over-night at 4 °C (Figure 2.6C). However, having to keep the protoplasts in cold conditions might affect their ability to express introduced DNA. The results obtained show that this protocol does not meet our requirements for considering it as optimal for the purposes explained.

In conclusion, both methods were successful at isolating viable leaf protoplast but the decrease in viabilities after only a few hours made them unsuitable for transformation experiments to optimise gRNA design. The potential reasons and amendments to the protocols used will be discussed later on.

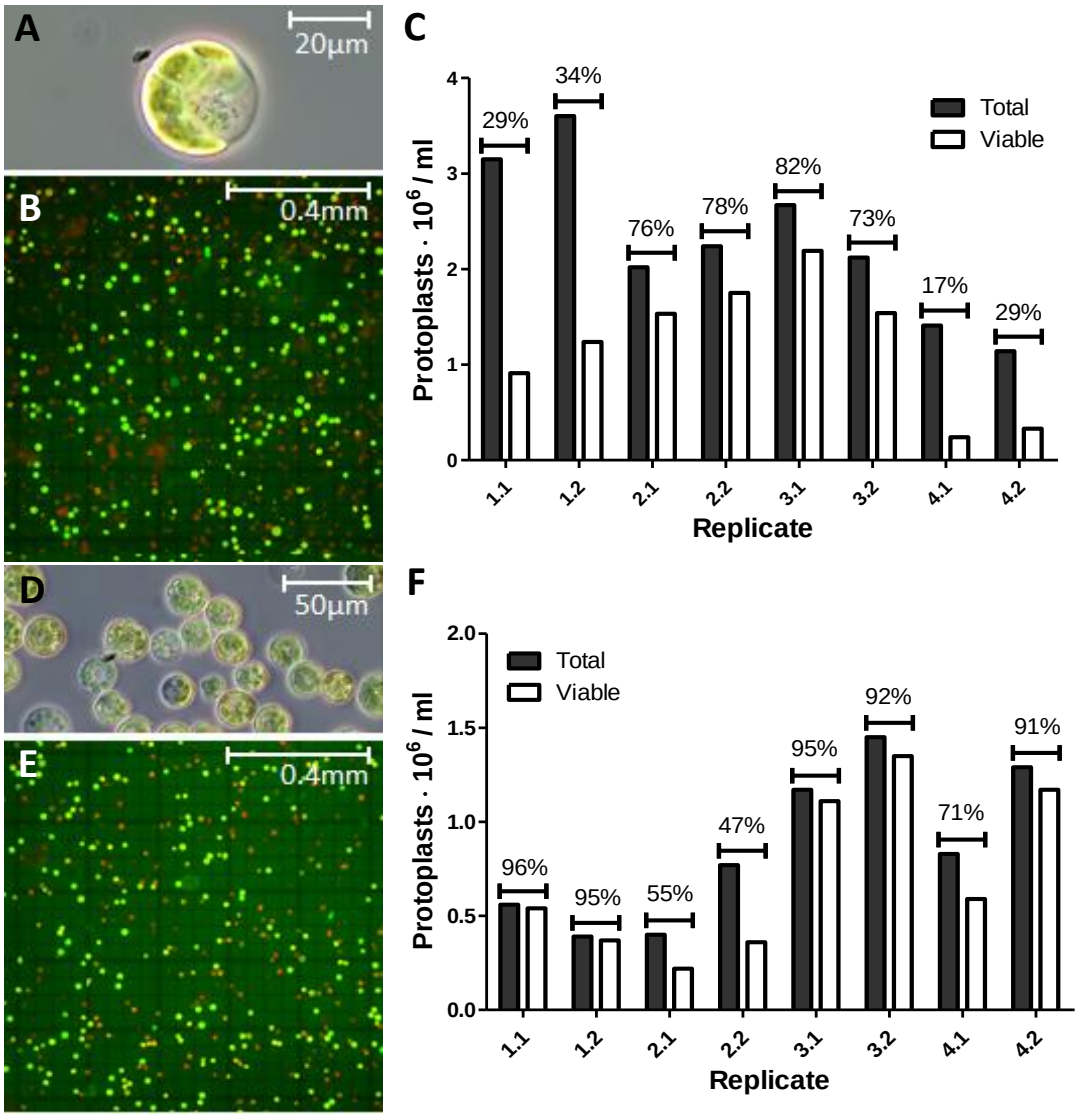


Figure 2.5. Both protoplast isolation protocols successfully released viable protoplasts.
Both Mannitol and Sucrose based protocols released viable protoplasts. (A) Protoplast isolated using the Mannitol-based Protocol under an optic microscope. (B) Fluorescent image of an FDA staining of a protoplast suspension obtained using the Mannitol-based Protocol. The viable protoplasts are green and non-viable in red. (C) Summary of the total protoplast yields, and alive protoplast yields of each one of the biological replicates (1, 2 and 3) and the technical replicates (i.e. 1.1 and 1.2) obtained using the Mannitol-based Protocol. Above each of the replicates, the percentage of viable protoplasts in the sample is shown. (D) Protoplasts isolated using the Sucrose-based Protocol under an optic microscope. (E) Fluorescent image of an FDA staining of a protoplast solution obtained using the Sucrose-based Protocol. Viable protoplasts are shown in red and the non-viable ones in green. (F) Summary of the total protoplast yields, and alive protoplast yields of each one of the biological replicates (1, 2 and 3) and the technical replicates (i.e. 1.1 and 1.2) obtained using the Sucrose-based Protocol. Above each of the replicates, the percentage of viable protoplasts in the sample is shown.

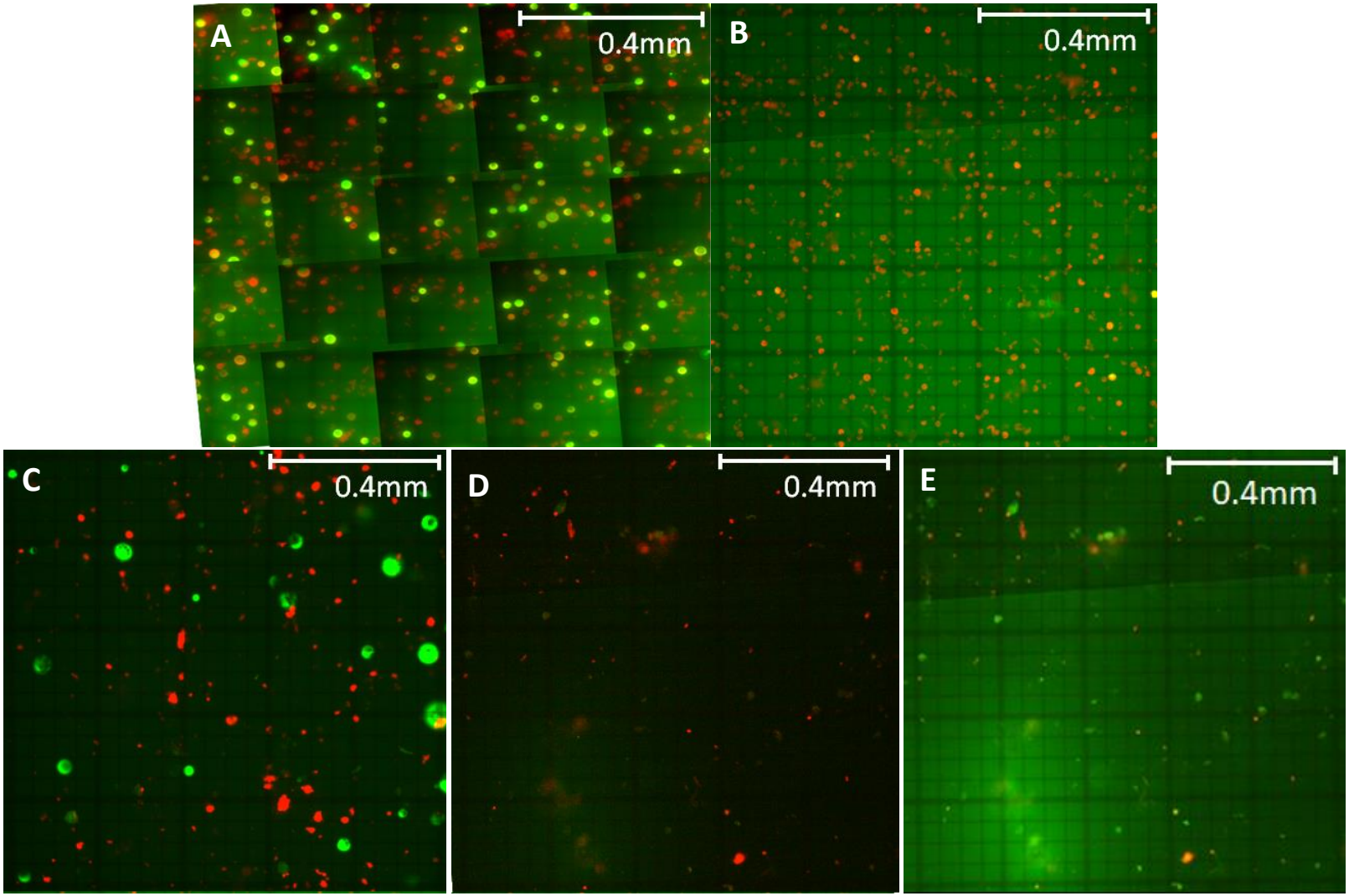


Figure 2.6. Protoplast viability is reduced after one day.
Fluorescent image of an FDA staining of a protoplast solution obtained using different protocols at different times. (A) Mannitol-based Protocol after the isolation. (B) Mannitol-based Protocol after one day at room temperature. (C) Sucrose-based Protocol after the isolation. (D) Sucrose-based Protocol after one day at room temperature. (E) Sucrose based protocol after one day at 4˚C.

2.4.2. Biolistic delivery of DNA into embryos.

Immature embryos are a different type of transformable tissue that can be used in transient expression studies. Given the characteristics of the tissue, two different methods for transformation are available: biolistic delivery and *Agrobacterium*-mediated transformation. In this chapter, the biolistic method was implemented and the efficiencies of the delivery of plasmid DNA were assessed. The possibility of using *Agrobacterium*-mediated transformation will be later discussed.

Brachypodium distachyon immature embryos obtained from 10 weeks old plants were bombarded with pActH2BGFP. After the bombardment, they were incubated for 2 days to ensure the embryos had enough time to express the delivered DNA. Then, they were screened under a fluorescence microscope. A clear difference was observed between embryos expressing the nuclear targeted GFP from the plasmid (GFP positive, GFP^+) and the ones that did not (GFP negative, GFP^-) (Figure 2.7A, B). To ensure that the distribution of GFP positive and negative embryos was random and not compromised by gold clumping, 64 embryos were distributed in a grid within the target area of the plate and the location of GFP positive and negative embryos was recorded. When there was no gold clumping, the distribution of GFP positive and GFP negative embryos was random (Figure 2.7C). This did not happen when the gold was clumped (Figure 2.7D).

Once classified as GFP positive or negative, the scutellum area of each embryo was measured. When the data was represented as a scatter plot, a threshold was observed, below which embryos do not express the bombarded plasmid, as no embryos with a scutellum area below 0.3 mm^2 were GFP positive (Figure 2.7E).

When representing this data as the number of embryos of a specific size found to be positive and negative, a distribution close to the Gaussian distribution is found in the GFP positive group (Figure 2.7F). The fitness of the data to this distribution was confirmed as well by the statistical analysis performed, which showed an r^2 value of 0.9 and a $S_{y,x}$ of 5.063 using 22 degrees of freedom. On the other hand, the GFP negative group did not show this fitness, showing an r^2 value of 0.33, $S_{y,x}$ of 2.68 with 23 degrees of freedom. From this distribution the scutellum area of 10 weeks old plants could also be observed. The majority of embryos had a scutellum area between 1.2 and 1.7 mm^2 two days after isolation, and a vast majority of positive embryos had a scutellum area around 1.4 and 1.5 mm^2, with the negative embryos being found all around the distribution but accumulating on the left tail of the distribution. These findings indicated an optimal size for biolistic delivery. However, it cannot be guaranteed that an embryo with a specific scutellum area will be transformed, given that GFP negative embryos are found through the whole distribution. Regarding the embryos with a scutellum size below the average of the distribution, the number of GFP positive embryos only surpasses the number of negative embryos when observing embryos with a scutellum size of 0.8 mm^2, observing more than 70% of embryos with a scutellum size of 0.9 mm^2 to be GFP positive. On the right side, the lack of replicates of high scutellum areas made it impossible to derive a maximum size threshold.

After observing an optimal size for embryos to take up and express a plasmid delivered via biolistic methods, the efficiency of this delivery was measured as number of cells transformed. For these experiments, a different ecotype was used. These ecotypes did not have any known genetical differences, and the sizes of these embryos after the isolation were not significantly different (Figure 2.8C).

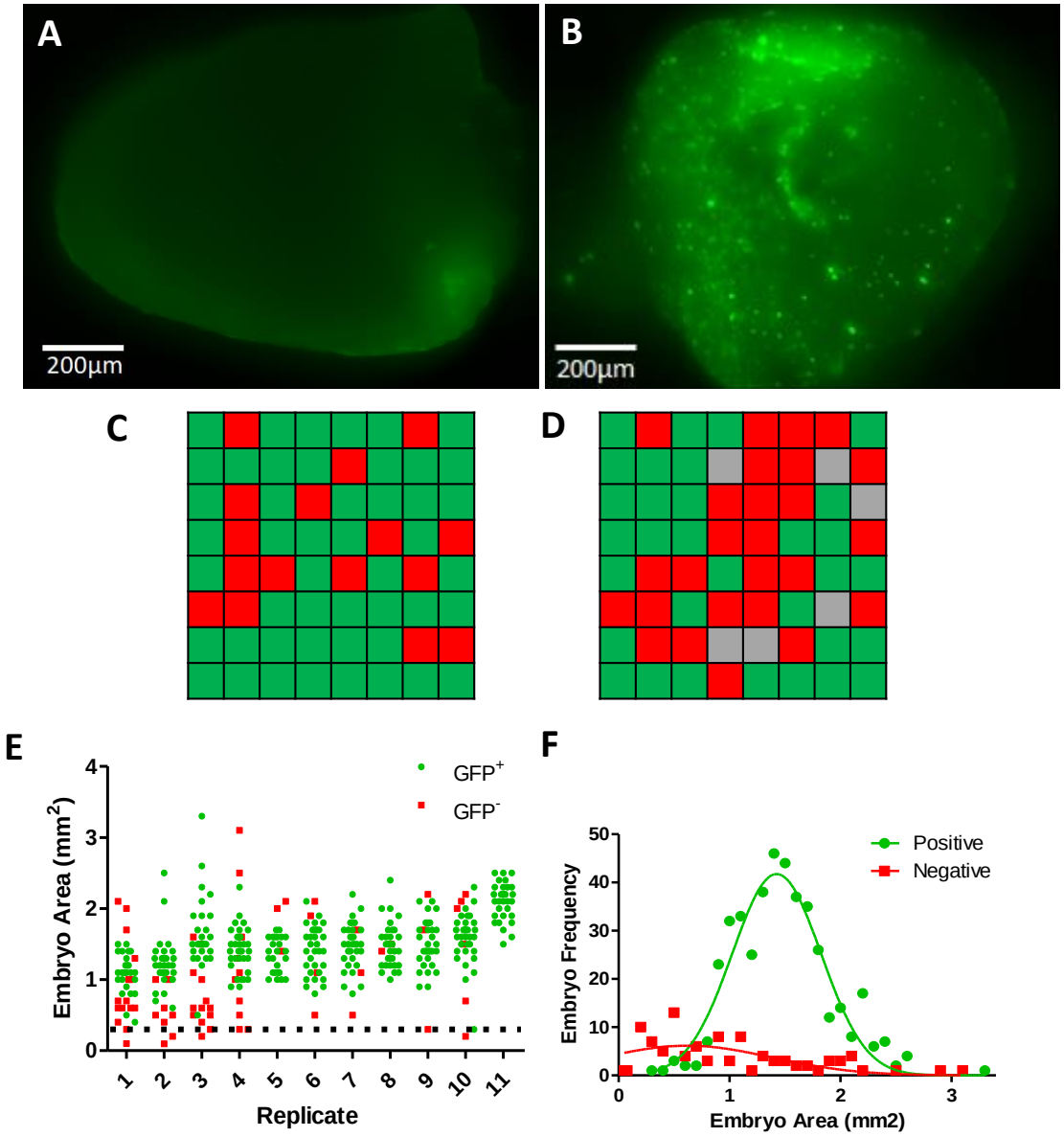


Figure 2.7. Scutellum area measurements after biolistic delivery of a plasmid containing GFP suggest an optimal embryo size for biolistic delivery of DNA.

(A) Picture of a control Bd21 embryo isolated from a 10 week old plant. (B) Picture of a GFP positive Bd21 embryo isolated from a 10 week old plant and bombarded with pActH2BGFP. (C) Diagram of the distribution of GFP positive and negative embryos after transformation on a plate without gold clumping. In green are represented the GFP positive embryos and in red the GFP negative embryos. (D) Diagram of the distribution of GFP positive and negative embryos after transformation on a plate without gold clumping. In green are represented the GFP positive embryos, in red the GFP negative embryos and in grey the ones that could not be screened. (E) Scatter plot displaying the size of embryos that showed GFP expression 2 days after bombardment with pActH2BGFP. In green are shown the GFP positive embryos and in red the GFP negative ones. A dotted line below shows the threshold below which no GFP positive embryos are found. (F) Frequency plot of the GFP positive and GFP negative embryos found after bombardment combining all replicates. Dots represent the counts in intervals of 0.1 mm^2 of scutellum area. Green dots show the number of GFP positive embryos with a specific scutellum area. A green line shows the best fitting Gaussian distribution for the data. Red dots show the number of GFP negative embryos with a specific scutellum area. A red line shows the best fitting Gaussian distribution for the data.

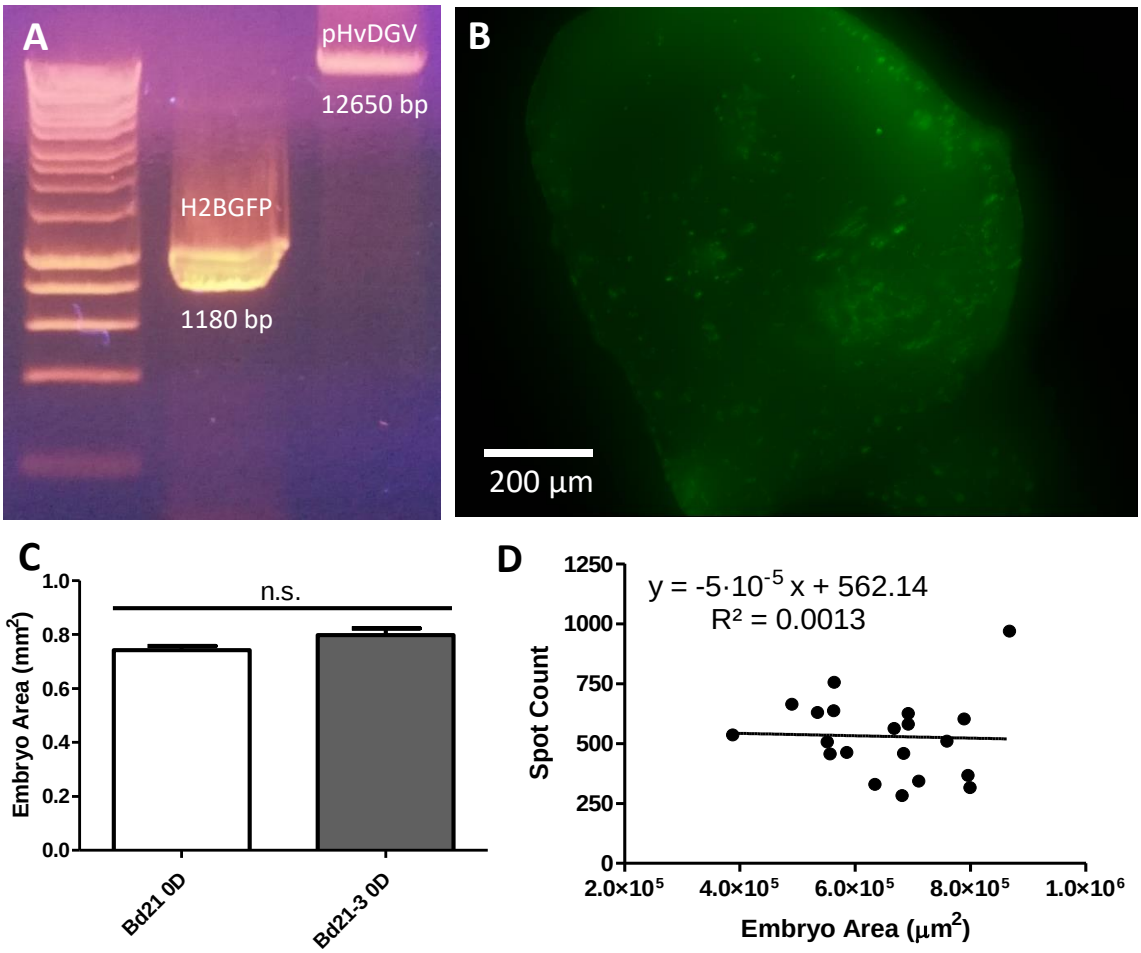


Figure 2.8. Number of cells transformed does not correlate with embryo size.
No correlation was found between scutellum area and spot count in *Brachypodium* embryos bombarded with pDGV-H2BGFP. (A) Bands amplified for the development of the pDGV-H2BGFP plasmid. From left to right are shown a 1 kb ladder, the band amplifying the GFP cassette from the pActH2BGFP plasmid and the band amplifying the pHvDGV backbone. These amplicons were obtained as described in the Materials and Methods of this chapter. (B) Picture of a GFP positive Bd21-3 embryo isolated from a 10 week old plant and bombarded with pDGV-H2BGFP. (C) Average scutellum areas of 10 week old embryos from Bd21 and Bd21-3 measured after the isolation. (D) Scatter plot of 20 different biological replicates of individual embryos bombarded with DGV-H2BGFP, showing the scutellum area after two days on the X axis and the amount of fluorescent nuclear spots found in the embryos.

After biolistic delivery and GFP screening, images of the GFP positive embryos were taken and the number of fluorescent spots found in the scutellum was counted (Figure 2.8B). These experiments implied the successful bombardment of the pDGV-H2BGFP plasmid into the Bd21-3 ecotype. The construction of the pDGV-H2BGFP was successful (Figure 2.8A) and showed GFP expression in bombarded embryos (Figure 2.8B). The scutellum area of the Bd21-3 ecotype and the Bd21 ecotype used in previous experiments was not significantly different (P-value of 0.0737 in an unpaired t-test, Figure 2.8C), making it possible to compare the experiments previously described and the experiments relating to spot counts. The results obtained showed an average of 530 spots per GFP positive embryo (Standard Deviation = 166.29; Number of observations = 20, Figure 2.8D), with a minimum of 283 spots and a maximum of 969. It was expected to observe a positive correlation between scutellum area and number of fluorescent spots, given that a larger scutellum would have more area to be affected by the bombardment. When calculating the best linear fit to the data, a high dispersion from

it is observed (r^2 of 0.02333, $S_{y,x}$ of 170.7), as well as a slope not significantly different from 0 after an F-test (slope of -0.00004953 ± 0.0003243 with a P value of 0.8803).

To put the spot count results in the context of the three-dimensional embryo, a model of the embryo was developed. This model would help understand the feasibility of using this tissue as a platform for sgRNA screenings in CRISPR/Cas9 genome editing. The model attempted to estimate the number of cells transformed in the context of the full embryo. This model was based on three assumptions. The first one approximates the shape of the embryo as semi-spherical (Figure 2.9A, B, C). The second assumes the cells in the embryo to be spherical instead of the usual polyhedral shape of the common stem or leaf cell (Figure 2.9D). Finally, that all cells in the embryo have the same size as the ones measured in the scutellum. This assumption was made despite the fact that cells in the axis and scutellum were found to be significantly different (P value < 0.05 in a two-tailed t-test, Figure 2.9E). However, to simplify the calculations, all cells were assumed to have the same area. Knowing that cells in the axis are larger than the ones found in the scutellum, this will lead to an over-estimation of the number of cells in the embryo. With these assumptions, the volume of the embryo and number of cells within an embryo can be estimated, and the percentage cells that can be transformed using biolistic methods in *B. distachyon* embryos be determined.

To relate the model of the embryo to the bombardments that have been shown previously, the cell area of embryos was measured both after the isolation and after 2 days of incubation in the same conditions embryos would be incubated after bombardment. These measurements were obtained from microscopy images of fixed embryos that had been stained with Toluidine Blue.

When cell size was measured after the isolation, the average cell area of each one of the biological replicates differed, with an average cell size among the replicates of 22.303 μm^2 (SD = 8.84; No. = 10), with cell sizes ranging from 89.46 μm^2 to 3.72 μm^2 (Figure 2.9F). When measured 3 days after the isolation, the average cell area among the replicates increased to 51.78μm^2 (SD = 13.76, No. = 5), ranging from 120.31 μm^2 to 8.41 μm^2 (Figure 2.9G). It was also observed that cell sizes observed did not correlate with their respective scutellum area.

At the same time, assuming the embryo is semi-spherical, the volume of the embryo can be calculated from the scutellum areas measured. Additionally, assuming cell areas represent the middle of the cell, cell volumes can be calculated. After assuming cells in the scutellum represent all cells in the embryo, the number of cells in an embryo can be calculated. The numerical results obtained are shown in Table 2.3.

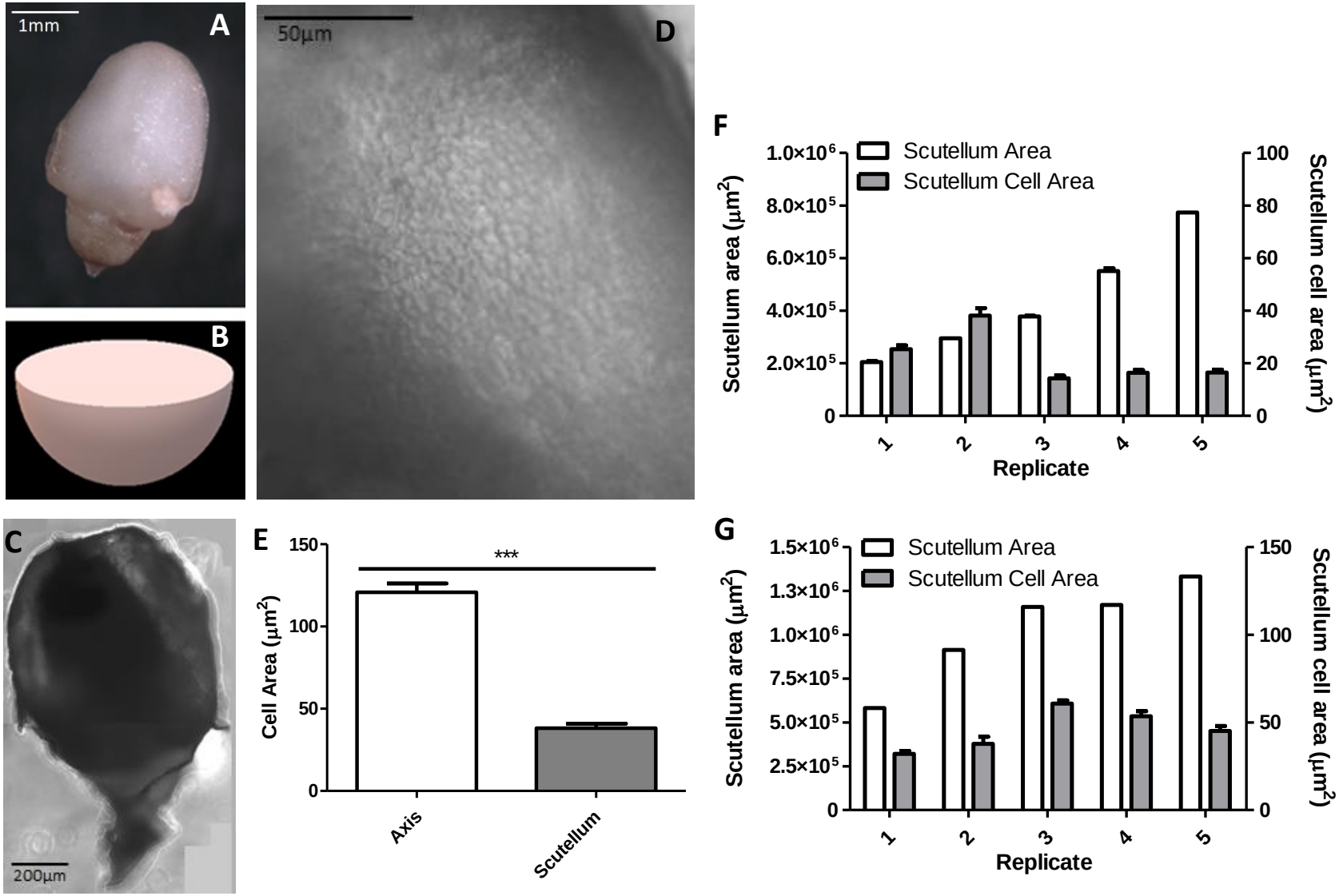


Figure 2.9. Cell sizes were measured after toluidine blue staining of the embryos and used to develop the physical shape model of the embryo.
(A) Picture of the scutellum of an embryo isolated from a 10 week old plant. (B) Model of the embryo as a semi-sphere with the scutellum facing up. (C, D) Images of an embryo obtained with the microscope with different magnifications after Toluidine Blue Staining. First one was used for calculating scutellum areas, while the second was used to calculate cell areas. (E) Comparison of the cell areas found in different parts of the same embryo, showing a significant difference between them. (F) Scutellum and cell areas measured after fixating the tissue right after isolating the embryo. In the left Y axis the Scutellum areas are represented, while the right Y axis represents cell areas. (G) Scutellum and cell areas measured after fixationing of the tissue 3 days after isolating the embryo. In the left Y axis the Scutellum areas are represented, while the right Y axis represents cell areas.

Table 2.3. Data obtained from the calculations leading to the proposal of the physical model of the embryo.

	0 Days	*3 Days*	*3 Days/0 Days*
Average Scutellum Area (μm^2)	$4.47 \cdot 10^5$	$10.31 \cdot 10^5$	2.31
Standard Deviation	$1.92 \cdot 10^5$	$2.92 \cdot 10^5$	
Number of observations	10	5	
Scutellum Radius (μm)	377.22	572.99	1.52
Embryo Volume (μm^3)	$112.42 \cdot 10^6$	$394.01 \cdot 10^6$	3.51
Average Cell Area (μm^2)	23.79	45.86	1.93
Standard Deviation	13.86	22.07	
Number of observations	490	250	
Cell Radius (μm)	2.75	3.82	1.52
Cell Volume (μm^3)	87.31	233.64	2.68
Cell Number	$1.29 \cdot 10^6$	$1.69 \cdot 10^6$	1.31

When comparing the differences between scutellum and cell area after the isolation and after 3 days of incubation, scutellum area grows to be 2.3 times bigger after 3 days, whereas cell area grows to 1.9 times bigger. At the same time, the calculated embryo volume is 3.5 times bigger after 3 days, while the cell volume grows 2.6 times bigger. When calculating the number of cells in embryos after the isolation and 3 days later, the numbers are similar, with 1.2 million cells per embryo after isolation and 1.7 million cells per embryo after 3 days. These results imply that the embryonic growth is mostly due to cell expansion.

Putting the results obtained for the number of transformed cells in the context of the physical model of the embryo proposed, the biolistic delivery protocol is estimated to transform an average of 1 in $3 \cdot 10^4$ cells.

2.4.3. RE-PCR cycles could be helpful for GE screenings.

In the context of the embryo, it was estimated that biolistic method introduced the foreign DNA to approximately 1 in 30,000 cells. Additionally, it was observed that the embryo develops by means of expansion rather than division, keeping this ratio at least until two days after the isolation of the embryo. This implies that, even with 100% editing efficiency, the edited DNA will be in an excess pool of non-edited DNA. An approach that has been used circumvent this problem has been the use of Restriction Enzyme-PCR (RE-PCR) techniques. In the context of genome editing, a site with a known restriction is selected for modification, with the restriction site lost following successful editing. Then, after applying the restriction reaction, the edited DNA would not be cleaved, while the non-edited would be cut. Following restriction, PCR would be used to amplify the edited DNA for further analysis.

To simulate the mixture of edited and non-edited DNA, an *in vitro* approach using plasmid DNA was applied. Both pHvDGV and pBract204 contain a HygR sequence, and after sequence comparisons, it was found that the one in pBract204 contains two AdeI restriction sites and the one in pHvDGV does not have any of them (Figure 2.10A). This provided a good model to test the sensitivity of mutation detection via restriction and PCR. Two sets of nested primers were designed for the amplification of the area around these two sites to simulate an editing screening with known relationships between edited and non-edited amplicons.

The RE-PCR assay was performed using a dilution of pHvDGV and pBract204 in a 1:6000 copy relation. Out of that mixture, PCR amplicons of HygR were obtained and cloned into the pCR™2.1 plasmid,

which had a third restriction site. The plasmid was introduced in *E. coli* and specific colonies were selected for the evaluation of the amplicons introduced in the plasmid. The RE-PCR tests were performed by applying AdeI restriction enzyme to the plasmid DNA mixture, amplifying using PCR and isolating the obtained band. This band was then be cloned in a pCR™2.1 plasmid or go onto the next RE-PCR cycle by restricting the PCR product isolated and amplifying it afterwards. Ten colonies obtained after cloning the amplicons obtained after no RE-PCR (RE-PCR 0), one (RE-PCR 1), two (RE-PCR 2) and three RE-PCR cycles (RE-PCR 3) into the pCR™2.1 were digested using AdeI (Figure 2.10B).

As expected, the colonies with the amplicon amplified without any previous restriction PCR (RE-PCR 0) showed that the plasmid was cut in three different sites, implying that all the amplicons came from the pBract204 plasmid. The differences between band size between plasmids was due to the TA cloning introducing the amplicon in a non-directional manner. On the other hand, when applying one RE-PCR cycle (RE-PCR 1), one out of nine colonies showed the HygR amplicon from the pHvDGV plasmid, given the fact that the plasmid was linearized but not cut further. After applying a second RE-PCR cycle (RE-PCR 2), one out of ten colonies showed the non-restricted band. The reduction in percentage could be due to the low number of replicates studied. After applying another RE-PCR cycle (RE-PCR 3), the number of pHvDGV amplicons increased to four out of nine. The progressive increase in the amplicons observed lead to the assumption that the results recorded in case of the RE-PCR 2 should be between 1 and 4 out of 10. The results obtained show that sequential RE-PCR cycles favour the appearance of low-represented amplicons, implying that its use for finding edited amplicons when targeting a restriction site could be possible.

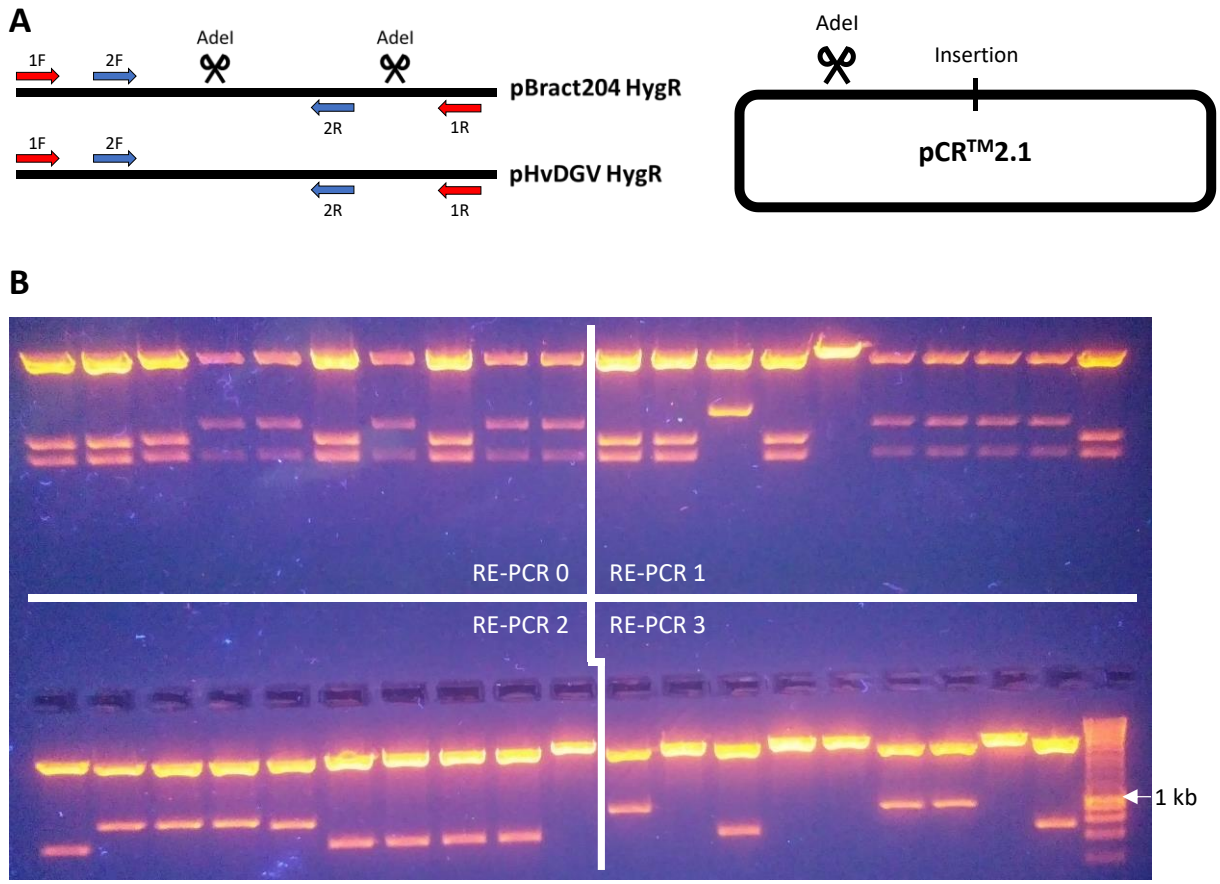


Figure 2.10. RE-PCR Cycles facilitate the finding of low-represented amplicons.
(A) Schematic representation of the two different amplicons together with the pCR™2.1 plasmid. The arrows labelled as 1 and 2 represent the primers used for the amplification of the HygR sequence, with the forward and reverse primers labelled as F and R respectively. 1F represents the HygR CDS Rv primer, 1R represents the HygR CDS Fw primer, 2F represents the HygR Nd1 Fw primer and 2R represents the pBHygR Rv primer (See Supplementary Table 3 for the specific sequences. Tagged as AdeI are the location of the restriction sites used in this test. On the pCR™2.1 plasmid is shown the insertion site. (B) Agarose gel showing the results of the RE-PCR assay. In the top left quadrant are shown the restrictions of 10 plasmids obtained after cloning the HygR amplicons after no RE-PCR (RE-PCR 0), in the top right quadrant are shown the ones after one restriction cycle (RE-PCR 1), in the bottom left quadrant the ones obtained after two RE-PCR cycles (RE-PCR 2) and in the bottom right quadrant the ones obtained after three RE-PCR cycles (RE-PCR 3). In the bottom left run a 1kb ladder is shown.

2.5. Discussion.

2.5.1. Leaf protoplasts as a platform for Genome Editing screenings.

The results obtained showed that protoplast isolation in *B. distachyon* is possible. It also showed that the yields of these isolations are high enough in both total quantity and initial viability percentage to allow further transformation using different methods, mainly PEG-mediated transformation.

The yields described in the original references the protocols used were based on were different from the ones experimentally obtained. The original mannitol-based protocol, described in Hong et al. (2012), reporter a yield of $1.7 \cdot 10^7$ protoplasts per ml. However, the weight of initial material was not specified in the protocol, making a correlation between weight and yield impossible. At the same time, this yield came with a reported viability of 90% after isolation, which was "moderately maintained" after 24 hours, but decreased to around 60% after 48 hours. Consequently, the effects of transient expression were screened after 24 hours. The original sucrose-based protocols yielded $5 \cdot 10^6$ - $5 \cdot 10^7$ protoplasts per ml (Jung et al., 2014) and 0.5 - $1.0 \cdot 10^6$ protoplasts per ml (Jung et al., 2015). Both protocols used 0.2 g of leaf tissue as the starting material. In the first protocol, all transient expression measurements were made 18 hours after transformation and the viability of these protoplasts was not reported. In case of the second protocol, the viability is not specified either, but the protocol recommends screening the protoplasts for transient expression after 24 to 96 hours, implying that the viability is still high. The results obtained experimentally in this chapter show a yield and viability that is high enough for transformation, $0.5 \cdot 10^6$ protoplasts per ml in case of the Mannitol-based protocol and around $1 \cdot 10^6$ protoplasts per ml in case of the Sucrose-based protocol. In the published protocols, given the high number of protoplasts released, minor declines in viability do not imply that transformation is not possible. However, the protocols used in this chapter showed a severe decline in viability that made it impossible to reliably test the transformation methods.

When comparing the Sucrose-based protocol and the Mannitol based protocol, it was observed that in both cases we found cells alive after the isolation process, which implies that the use of different compounds to keep the osmotic pressure is not relevant to the final viability or the reduction of protoplast viability over two days. However, given the osmolyte differences between them, the salts in each one of the digestion solutions differed. To add to the hypothesis that osmotic pressure was not the reason behind the decline in protoplast viability, when protoplasts were kept in the dark at 4^{o}C, the drop in protoplast viability was not as drastic as the one found when keeping the protoplasts at room temperature. The transport of mannitol and sucrose through the membrane has been linked to plasmodesmata, vesicles and proton-substrate carriers (Salmon et al., 1995; Ayre, 2011). Given the characteristics of protoplasts, this transport might be hindered by a decrease in protein activity due to cold conditions would have stopped osmolyte transport, which would have meant that the decline in viability would have been as severe as it was at room temperature. However, for transformation purposes, keeping the protoplasts at such cold temperatures would hinder their protein activity and would imply the delivered DNA would not be expressed. The main difference between the protocols described and the ones that have been used were the plant growth conditions. In Hong et al. (2012), plants were grown in both *in vitro* conditions in MS Agar media and soil. In Shan et al. (2013) plants were grown in *in vitro* conditions in MS Agar media. In Jung et al. (2011), plants were grown in hydroponic conditions. In the experiments described in the results section, plants were grown in soil, starting with a small period of time in the glasshouse until they germinated, and then moved to the final growth room where they would grow until samples were collected. When comparing this process to the *in vitro* or hydroponic culture found in the other protocols, the sterile conditions differ from the protocols in the bibliography and the ones that have been implemented. This difference could be the reason why these differences in protoplast viabilities after 2 days have been found. Plant health has

been described as a key factor regarding protoplast isolation efficiency and viability (Jung et al., 2015). This might be affected by non-sterile growth conditions.

Another issue found with protoplasts was the low reproducibility of the protocol. All elements of the procedure were done with care, from using glass pipettes to avoid physical damage to the protoplasts, to the reduced centrifugal forces applied during the washing steps, including the preparation of all the solutions required. However, different outcomes were recorded. Different factors could have affected the number of protoplasts and viability independently of the protocol, such as plant health, as previously outlined. However, given the fine-tuning of the osmotic pressure required for the isolation of protoplasts, it cannot be discarded that minor variations in the composition of the solutions could have affected the results obtained from each isolation. However, the results obtained point to further refinements of the protocols being required for the use of protoplasts as a source of transformable tissue for molecular biology experiments in *Brachypodium distachyon* in our laboratory.

The use of protoplasts has been prominent in several plant species for editing screens focusing on transient expression on the tissue (see Chapter 4). In the case of *Brachypodium*, protoplasts have already been used as a platform for CRISPR/Cas9 screenings (Hus et al., 2020). In this case, an optimized version of a mannitol-based protocol is used for the isolation and transformation of protoplasts with the purpose of validating sgRNA activity prior to stable transformation of callus using *Agrobacterium*-mediated transformation. This protocol uses Bd21 grown in soil and adds a cleaning step prior to the chopping of the leaves. It yields over 10^6 protoplasts per ml and has reported the transfection of protoplasts obtained using this PEG mediated transformation. This suggests that the cleaning of the leaves may be crucial for the viability of the protoplasts, concurring with the contamination problems previously suggested.

Plant regeneration from protoplasts has been reported in different species related to *Brachypodium* such as maize (Mitchell and Petolino, 1991) and rice (Abdullah et al., 1986), but not in *Brachypodium* itself. The existence of these protocols adapted for other plants opens the possibility of transiently transforming protoplasts with CRISPR/Cas9 and obtaining transgene-free plants carrying the edits. However, these protocols have to be first developed and optimised for *Brachypodium*.

2.5.2. Embryos as a platform for Genome Editing screenings.

Typically, *Brachypodium* embryos have been used as the primary tissue from which callus tissue has been obtained (Ozdemir and Budak, 2018). This callus can be transformed using different techniques to obtain stably transformed plants to study a specific genotype or for crop development. Here, a protocol is proposed that skips the callus induction process and focuses on the ability of immature embryos to accept foreign DNA and express it transiently. This implies that *Brachypodium* immature embryos can be used as a target tissue for genome editing assays and sgRNA efficiency validation. In this section, different aspects regarding bombardment efficiency in *Brachypodium* immature embryos will be discussed.

Optimally, biolistic delivery implies an even distribution of the gold particles in the target object (in this case, the central part of the plate, a circle of 1-2 cm diameter). However, in sub-optimal conditions this might not be the case. In Sparks and Jones (2014), the importance of using the right DNA-gold ratio was highlighted. If the DNA coating of gold particles is sub-optimal, for example an excess of DNA is used, the gold particles would form clumps and their distribution would be uneven when bombarded. Gold clumping would imply that only certain areas of the plate are hit, observing that most of the transformed and non-transformed embryos appear in specific areas of the plate, in addition to higher physical damage of the target tissue hit by the larger particles. This distribution of transformed embryos in the plate was random, implying that there was not clumping. Therefore, it

could be assumed that embryo transformation did not depend on faulty delivery but the ability of each embryo to take the desired DNA.

Ismagul et al. (2018) proposed a modification of the biolistic delivery for wheat, using PEG instead of spermidine, was proposed to increase integration rates and increase transformation efficiency. It would be interesting to study the possibility of the use of this modification in *Brachypodium*. However, given the focus on transient expression of the experiments describing biolistic delivery in this chapter, it was not applied here.

To assess the efficiency of the transformation process, a fluorescence screen was used. In the embryo transformation experiments, two different plasmids were used, pActH2BGFP and pDGV-H2BGFP. Both of them have a GFP fused with an H2B nuclear localization signal and are under the transcriptional control of two strong constitutive promoters (Rice Actin1 in pActH2BGFP and Maize Ubiquitin 1 in pDGV-H2BGFP, see Chapter 3 for more information on other constitutive promoters). This implies that changes in nuclear targeted GFP signalling after transformation would be due to the ability of the embryo to take and express the delivered DNA, given the lack of nuclear targeted fluorescence in non-bombarded embryos. In Sparks and Jones (2014), the importance of negative controls where the embryos are bombarded with gold without a plasmid are highlighted. In addition to this control, another control where embryos were bombarded with a plasmid not containing any GFP cassettes would have been beneficial. However, the results described in Majid and Parveez (2007) show how bombardment with gold particles does not induce fluorescence in itself. Other papers regarding transformation using biolistic delivery do not mention this control either, including examples in *Setaria viridis* (Mookkan, 2018) and wheat (Ismagul et al., 2018; Hamada et al., 2017). However, the addition of negatives controls in this test would have been beneficial for the claims intended. Alternative methods to study the expression of the delivered genes could have been based on Real Time-quantitative PCR methods. However, the observation of fluorescence *in vivo* proves the ability to express the DNA delivered using biolistic methods.

In the biolistic delivery assays described in this chapter, immature embryos were obtained from seed obtained from 10 weeks old plants. However, not all the seeds in a specific plant will have the same age, thus, being in different maturation stages. Immature and mature embryos have been found to behave differently, to the point that obtaining callus tissue from it requires specific protocols (Vogel et al., 2006; Vogel and Hill, 2008; Alves et al., 2009; Hammami et al., 2011; Lee et al., 2011; Bragg et al., 2015; Ozdemir and Budak, 2018; Yu et al., 2019). Here, it is hypothesized that the maturity of the embryo will play a role in its ability to take and express the delivered DNA. To assess the maturity of the embryo, scutellum area was measured, inferring embryo size. Scutellum area was measured after observing if the embryo was transformed or not. It was observed that embryos with a scutellum area below 0.3 mm^2 did not show nuclear targeted GFP expression. This could imply that embryos that have not developed enough were not able to express the delivered DNA. It would also be possible that immature embryos are prone to damage from the delivery process, hindering their ability to express the delivered transgenes.

The bombarded embryos were examined, dividing them in transformed and non-transformed. When analysing the scutellum areas and observing the frequency of each scutellum size in each one of the groups defined, it was observed that most positive embryos had a scutellum area between 1.4 and 1.5 mm^2. These findings imply that there is an optimal maturity for embryo transformation.

Biolistic delivery into *Brachypodium* embryos has not been described, focusing on the use of embryogenic callus as a target. However, biolistic delivery into embryos has been described in wheat (Hamada et al., 2017; Ismagul et al., 2018). These papers use the term “mature seeds” and “grains

obtained 11-14 days after flowering" respectively. Both terms are in a way precise in the definition of the optimal material. Nonetheless, for refining transient expression in *Brachypodium* embryos, finding an optimal size for the embryos is a valuable addition. When observing the distribution of the non-transformed embryos, it was observed that besides the threshold found, there is no correlation between embryo size and the probability of them not being transformed, as non-transformed embryos did not show a clear distribution of size. Various implications come from these distributions. The first one, as previously stated, an optimal size for embryo scutellum was found for biolistic delivery. At the same time, the lack of a pattern regarding the non-transformed embryos besides the threshold found implies that embryos above the threshold that were not transformed can be due to technical variabilities rather than the embryos themselves. Finally, putting together all the embryos observed, the total distribution of embryos obtained from seeds from 10 week old plants was obtained.

It is notable that plant tissues have autofluorescence in different wavelengths (Müller et al., 2013). Embryonic tissue is not an exception, as some levels of autofluorescence were observed in the same channel used for screening the nuclear targeted GFP. In case of embryos smaller than the described expression threshold, the autofluorescence levels were lower than the ones observed in the rest. However, in bigger embryos the autofluorescence levels were higher. This implies that in larger embryos, false negatives and false positives could be recorded due to the higher autofluorescence found. The lack of autofluorescence in smaller embryos implies that the lack of nuclear targeted GFP signal is due to the developmental stage they are in. Autofluorescence has also been linked to different types of stress (García-Plazaola et al., 2015). In the studies described in this chapter, this could be physical stress, given the high vacuum conditions and high-speed delivery of the particles that biolistic techniques imply, or biological stress, as some plates showed bacterial contamination after prolonged periods of incubation. The experiments performed in this chapter are based on GFP signals, which could be masked by autofluorescence being detected in that specific channel. However, the results obtained point to the expression of the GFP being strong enough to bypass this issue, being an effective reporter for transformation in *Brachypodium* embryos. However, to confirm this controls with only gold and gold coated with other plasmids without any GFP cassettes would be necessary.. An alternative to the use of GFP would be the use of other fluorescent reporters, but the autofluorescence of the target tissue in different channels should be assessed first.

After finding a range of sizes with a high transformation efficiency using biolistic delivery in embryos, the number of fluorescent spots in transformed embryos was assessed. It would seem obvious that a larger embryo would be able to be hit by more gold particles. However, this correlation was not found. In a study performed in grapevines using biolistic delivery, it is stated that the average number of hits per plate would be around 8000 (Kikkert et al., 2005). Putting this in the context of the bombardments described in this chapter, targeting a 314 mm^2 area in the centre of the plate, it would mean that every mm^2 would be hit by around 25 particles. The spot counts described in this chapter shows an average of 529.95 spots per embryo, which had an average scutellum area of 0.65 mm^2, which implies 814 spots per mm^2. However, the bombardments described in that reference target a different plant tissue and were made in different conditions when compared with the ones described in this chapter, which calls for caution connecting both studies. This distribution of spots around the plate would still imply a correlation between scutellum area and spot counts, but given the small areas studied, error could mask this correlation. Another possibility would be that those hits actually exist, but the transformed cells were not able to express the delivered DNA due to the loss of the transiently delivered DNA or gene silencing.

To put the biolistic transformation of embryos in the context of genome editing experiments, it was necessary to estimate the relation between the transformed cells and the total number of cells that could be in an embryo. To do that, a physical model of the embryo was developed. Some assumptions were made in order to use data that could be obtained and to simplify the calculations, while overestimating the number of cells. This would mean that when running the actual editing assays, the relation between the edited cells and the non-edited cells would be higher than expected, ensuring that finding the desired edits was possible. It should be noted that these approximations may not fit the actual structures observed in the embryo, such as the semispherical shape instead of the embyo and axis structure, or the cells being spherical rather than the irregular shape observed in the pictures gathered. However, these approximations were necessary due to the need of regular shapes in order to do the calculations of the number of cells, as direct measurements of volume of the embryo or the cells in the embryo would not have been possible without them.

One of the assumptions was that scutellum cells represented the entire embryo, despite the fact that axis cells were found to be significantly different. At the same time, the percentage of cells that would belong to the axis and scutellum would be dependent on the developmental stage of the embryo and would have to be measured using microtomy techniques that were not available. Thus, it was decided that in case of assuming cells would only have one size, the smaller one should be chosen in order to overestimate the number of cells that an embryo has. It is also notable that when embryos start the germination process, the cells in the axis start elongating (Bewley and Black, 1994). This would distort the calculations of the model if the embryo started germinating.

The data collected suggested that cells expand rather than divide. This was fitting for the type of tissue studied, since embryos would start developing root and shoot tissue from specific areas of the embryo axis. When those places are not considered, differences in total volume would only be due to cell expansion rather than cell division.

The final objective of the model was to calculate the number of cells that can be transformed using biolistic methods. The initial over estimation of the number of cells in the embryo implies that the proportion of cells transformed has been underestimated. After the calculations, a ratio of transformed cells and non-transformed cells of 1:30,000 was found. These findings will be explored in following sections to assess the possibility of using this technique for obtaining edited cells in *Brachypodium.*

2.5.3. RE-PCR Screenings.

The method to screen for successful editing will depend on the resources available in each laboratory, given the different costs and requirements of each one of the techniques (Reviewed in Germini et al., 2018). It is also important to differentiate between screenings that simply look for edits and methods that aim to calculate the editing efficiency. The samples obtained from the editing experiments will also be important. One of the key factors will be the quality of the DNA isolated from the sample itself, since read quality can be an important issue in sequencing-based techniques. However, this depends on the kit used for amplicon isolation and the type of sequencing used. The percentage of edited amplicons in a sample will also be important. This might be related to the editing efficiency itself or plainly because not all cells were transformed.

In case of protoplast isolation, flow-sorting has been used as a simple manner for fast selection of transformed cells (Bargmann and Birnbaum, 2009, 2010; You et al., 2014; Petersen et al., 2019). However, the described transformation efficiencies described in the protocols used was high enough to disregard the possible effects of non-transformed cells in editing screenings. Studies using protoplasts as the target tissue for editing screenings have used a wide variety of screening

techniques, from PCR-RE to deep sequencing (See Supplementary Table 2). These techniques have been able to provide editing efficiencies and ultimately sequencing of the specific edits obtained. Traditionally embryo and callus transformation aimed to produce stable integration events, and often rely on an antibiotic selection step. However, this adds more *in vitro* culture time, which lengthens and increases the cost of the process. If selection does not occur, the DNA isolated from transformed embryos or callus will contain both edited and non-edited DNA. This implies that editing screenings might not be able to reliably provide editing efficiencies. At the same time, even after selection, escapees or simply transformed but non-edited cells will change the number of edited amplicons that will be observed. In experiments using transformation of callus or embryo, screenings have been made on the regenerated plant rather than directly on the transformed tissue. In this chapter, focusing on transient assays, RE-PCR has been found to be effective in finding edits in a pool of unedited sequences. This is a useful tool for sgRNA screenings and has already been successful in finding edits in some studies in plants (Jiang et al., 2013; Lawrenson et al., 2015). This would be applicable to both embryo and, by extension, to callus transformation. However, the actual ratio of transformed cells should be studied and the limits of RE-PCR should be investigated as well for adapting this method to callus screenings.

However, RE-PCR methods might not be suitable for obtaining editing efficiencies, since the removal of the non-edited DNA will imply both the non-transformed DNA and the transformed but non-edited. This technique will be useful for finding rare amplicons, looking for specific edits and validating the functionality of specific elements of the editing machinery that have been confirmed to function. To obtain efficiencies in these type of experiments, other strategies would be possible, such as the use of qPCR for screening the number of cells that contain the editing machinery and the ones that carry the edits.

Putting the results in the context of the physical model of the embryo, the RE-PCR assays were performed using a 1:6,000 ratio between the “edited” (pHvDGV HygR) and “wild type” (pBract204 HygR), in contrast with the 1:30,000 ratio of transformed cells using biolistic delivery. The choice of dilution was done as a proof of concept showing that subsequent RE-PCR cycles are able to increase the number of low-represented amplicons. However, the viability of the combined method of biolistic delivery in embryos and the use of RE-PCR cycles for the identification of edits will be assessed in Chapter 3.

In conclusion, the findings of this section show the potential of RE-PCR as a method for finding rare amplicons, more specifically in the context of genome editing. At the same time, it is important to highlight the limitations of these techniques and the possible amendments and further developments based on these findings.

2.5.4. Searching for a platform for quick sgRNA activity in *Brachypodium*.

The aim of this chapter was to develop a quick system for the validation of sgRNA activity in *Brachypodium*. For this purpose, the choice of tissue and the possibility of different types of edit screening have been discussed. Different possibilities for choice of tissue, transformation method and screening are available. However, each protocol will pose different challenges and advantages. In case of *Brachypodium*, different techniques have been described for both stable and transient expression (Betekhtin et al., 2020).

One of the key factors affecting the rapidness of the test will be the age of the plant required for tissue isolation. In this area, protoplasts have a clear advantage, requiring leaves from 3 week-old plants, when compared to embryos, over 10 weeks, and callus, 16 weeks. However, the possibility of rolling

sowing implies that after the first growing period, samples can be collected weekly. The protocols required for isolation and transformation add a few days to the time from seed to sample.

In addition to the time required to obtain each transformable tissue, the simplicity of the methods will affect the choice of protocol. In case of protoplast isolation and transformation, although it has been described in *Brachypodium* (Hus et al., 2020), the results obtained in this chapter point to a complicated process for the optimization of the protocol and its consistent use as a source of transformable tissue. Immature embryos have been found to be a more consistent option for obtaining a transformable tissue using a simple protocol. The use of callus and its maintenance poses a different challenge given the complications in callus growth from embryos and its growth after transformation (Alves et al., 2009).

Another factor to consider would be the transformation method, including both the simplicity of the method and the cost per transformed sample. As previously described, protoplast isolation relies on enzymatic digestion of the cell wall and purification either by sucrose gradient or by centrifugal steps. The cost of the enzymes (cellulase and macerozyme) can vary depending on the manufacturer, but add up to around £10 per isolation. The rest of the materials are of common use in the laboratory and could be disregarded. In case of biolistic delivery of embryos, the particle delivery system alone can cost up to £8,000. This implies that laboratories that may not have this equipment may find this protocol prohibitive. An alternative for embryo transformation using Agrobacterium-mediated transformation in *Brachypodium* has been described (Zombori et al., 2011), but its effectiveness should be assessed in similar terms to the ones studied in this chapter to be able to compare both protocols. In addition to the initial investment, the rupture discs, macrocarriers and gold particles add up to a cost of £35 per shot, making a low-cost alternative more interesting. In addition to the transformation process itself, the protocol described requires a specific media that is also used for callus induction and transformation using Agrobacterium (BdCPM). This media contains several vitamins that also increase the cost related to the experiment. The use of callus for *Agrobacterium*-mediated transformation will only imply the isolation of embryos and its culture *in vitro*, changing the plate every two weeks, prior to the application of *Agrobacterium*. Usually, callus is used for stable transformation of *Brachypodium*. However, it could be used to assess sgRNA viability as embryos were used for the same purpose by using RE-PCR.

2.6. Conclusions.

The results obtained in this chapter, together with the data referenced in the bibliography, point to the versatility of different tissues of *Brachypodium distachyon* to be transformed using different techniques. Both protocols proposed for protoplast isolation were able to produce alive protoplasts, but further refinement will be required for them to become a transformable tissue in our lab. On the other hand, a protocol using embryos has been developed and used for the production of tissue transiently expressing the delivered DNA. This protocol has the potential of being a platform for the screening of edits using RE-PCR methods, which have been proven to allow the finding of rare amplicons in this tissue.

2.7. References.

Abdullah, R., Cocking, E. C., and Thompson, J. A. (1986). Efficient plant regeneration from rice protoplasts through somatic embryogenesis. *Bio/Technology* 4, 1087–1090.

Abel, S., and Theologis, A. (1994). Transient transformation of Arabidopsis leaf protoplasts: a versatile experimental system to study gene expression. *Plant Journal* 5, 421–427.

Agarwal, S., Loar, S., Steber, C., and Zale, J. (2009). Floral transformation of wheat. *Methods in Molecular Biology* 478, 105–113.

Agrawal, P. K., Kohli, A., Twyman, R. M., and Christou, P. (2005). Transformation of plants with multiple cassettes generates simple transgene integration patterns and high expression levels. *Molecular Breeding* 16, 247–260.

Ahokas, H. (1989). Transfection of germinating barley seed electrophoretically with exogenous DNA. *Theoretical and Applied Genetics* 77, 469–472.

Alves, S. C., Worland, B., Thole, V., Snape, J. W., Bevan, M. W., and Vain, P. (2009). A protocol for *Agrobacterium*-mediated transformation of *Brachypodium distachyon* community standard line Bd21. *Nature Protocols* 4, 638–649.

Asad, S., Mukhtar, Z., Nazir, F., Hashmi, J. A., Mansoor, S., Zafar, Y., and Arshad, M. (2008). Silicon carbide whisker-mediated embryogenic callus transformation of cotton (*Gossypium hirsutum* L.) and regeneration of salt tolerant plants. *Molecular Biotechnology* 40, 161–169.

Ayre, B. G. (2011). Membrane-Transport Systems for Sucrose in Relation to Whole-Plant Carbon Partitioning. *Molecular Plant* 4, 377–394.

Bablak, P., Draper, J., Davey, M. R., and Lynch, P. T. (1995). Plant regeneration and micropropagation of *Brachypodium distachyon*. *Plant Cell, Tissue and Organ Culture* 42, 97–107.

Bargmann, B. O. R., and Birnbaum, K. D. (2009). Positive fluorescent selection permits precise, rapid, and in-depth overexpression analysis in plant protoplasts. *Plant Physiology* 149, 1231–1239.

Bargmann, B. O. R., and Birnbaum, K. D. (2010). Fluorescence activated cell sorting of plant protoplasts. *Journal of Visualized Experiments* 2010, 1673.

Bart, R., Chern, M., Park, C. J., Bartley, L., and Ronald, P. C. (2006). A novel system for gene silencing using siRNAs in rice leaf and stem-derived protoplasts. *Plant Methods* 2, 13.

Bates, G. W. (1989). Electroporation of protoplasts. *Journal of Tissue Culture Methods* 12, 121–126.

Betekhtin, A., Hus, K., Rojek-Jelonek, M., Kurczynska, E., Nibau, C., Doonan, J. H., and Hasterok, R. (2020). *In vitro* tissue culture in *Brachypodium*: Applications and challenges. *International Journal of Molecular Sciences* 21, 1037.

Betekhtin, A., Rojek, M., Nowak, K., Pinski, A., Milewska-Hendel, A., Kurczynska, E., Doonan, J. H., and Hasterok, R. (2018). Cell wall epitopes and endoploidy as reporters of embryogenic potential in *Brachypodium distachyon* callus culture. *International Journal of Molecular Sciences* 19, 3811.

Bewley, J. D., and Black, M. (1994). *Seeds*. New York : Plenum Press

Boel, A., Steyaert, W., De Rocker, N., Menten, B., Callewaert, B., De Paepe, A., Coucke, P., and Willaert, A. (2016). BATCH-GE: Batch analysis of Next-Generation Sequencing data for genome editing assessment. *Scientific Reports* 6, 30330.

Bragg, J. N., Anderton, A., Nieu, R., and Vogel, J. P. (2015). *Brachypodium distachyon*. *Methods in Molecular Biology (Clifton, N.J.)* 1223, 17–33.

Bragg, J. N., Wu, J., Gordon, S. P., Guttman, M. E., Thilmony, R., Lazo, G. R., Gu, Y. Q., and Vogel, J. P. (2012). Generation and Characterization of the Western Regional Research Center *Brachypodium* T-DNA Insertional Mutant Collection. *PLoS ONE* 7, e41916.

Brinkman, E. K., Chen, T., Amendola, M., and Van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Research* 42, e168.

Christiansen, P., Andersen, C. H., Didion, T., Folling, M., and Nielsen, K. K. (2005). A rapid and efficient transformation protocol for the grass *Brachypodium distachyon*. *Plant Cell Reports* 23, 751–758.

Clough, S. J., and Bent, A. F. (1998). Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant Journal* 16, 735–743.

Crossway, A., Oakes, J. V., Irvine, J. M., Ward, B., Knauf, V. C., and Shewmaker, C. K. (1986). Integration of foreign DNA following microinjection of tobacco mesophyll protoplasts. *MGG Molecular & General Genetics* 202, 179–185.

D'Agostino, Y., Locascio, A., Ristoratore, F., Sordino, P., Spagnuolo, A., Borra, M., and D'Aniello, S. (2016). A Rapid and Cheap Methodology for CRISPR/Cas9 Zebrafish Mutant Screening. *Molecular Biotechnology* 58, 73–78.

D'Halluin, K., Bonne, E., Bossut, M., De Beuckeleer, M., and Leemans, J. (1992). Transgenic maize plants by tissue electroporation. *Plant Cell* 4, 1495–1505.

De Buck, S., De Wilde, C., Van Montagu, M., and Depicker, A. (2000). Determination of the T-DNA transfer and the T-DNA integration frequencies upon cocultivation of *Arabidopsis thaliana* root explants. *Molecular Plant-Microbe Interactions* 13, 658–665.

De La Peña, A., Lörz, H., and Schell, J. (1987). Transgenic rye plants obtained by injecting DNA into young floral tillers. *Nature* 325, 274–276.

Denbow, C. J., Lapins, S., Dietz, N., Scherer, R., Nimchuk, Z. L., and Okumoto, S. (2017). Gateway-compatible CRISPR/Cas9 vectors and a rapid detection by high-resolution melting curve analysis. *Frontiers in Plant Science* 8, 1171.

de Souza, W. R., Martins, P. K., Freeman, J., Pellny, T. K., Michaelson, L. V., Sampaio, B. L., et al. (2018). Suppression of a single BAHD gene in *Setaria viridis* causes large, stable decreases in cell wall feruloylation and increases biomass digestibility. *New Phytologist* 218, 81-93.

Dong, Y., and Zhu, H. (2005). Single-strand conformational polymorphism analysis: basic principles and routine practice. *Methods in molecular medicine* 108, 149–157.

Draper, J., Mur, L. A. J., Jenkins, G., Ghosh-Biswas, G. C., Bablak, P., Hasterok, R., and Routledge, A. P. M. (2001). *Brachypodium distachyon*. A new model system for functional genomics in grasses. *Plant Physiology* 127, 1539–1555.

Duan, X. L., and Chen, S. B. (1985). Variation of the characters in rice (*Oryza sativa*) induced by foreign DNA uptake. *Scientia Agricultura Sinica* 3, 6–10.

Fowke, L. C., and Constabel, F. (1983). "Plant Protoplasts: An Experimental System for Cell Biologists," in *Plant Cell Culture in Crop Improvement*, eds. Sen, S.K. and Giles, K.L. (New York, NY: Plenum Press), 185–200.

Fromm, M., Taylor, L. P., and Walbot, V. (1985). Expression of genes transferred into monocot and dicot plant cells by electroporation. *Proceedings of the National Academy of Sciences of the United States of America* 82, 5824–5828.

Fu, X., Duc, L. T., Fontana, S., Bong, B. B., Tinjuangjun, P., Sudhakar, D., Twyman, R. M., Christou, P.,

and Kohli, A. (2000). Linear transgene constructs lacking vector backbone sequences generate low-copy-number transgenic plants with simple integration patterns. *Transgenic Research* 9, 11–19.

Fuchs, L. (2019). Manipulating Meiotic Recombination in Grasses: Development of a CRISPR/Cas9 Pipeline in *Lolium*. PhD Thesis. Aberystwyth: Aberystwyth University.

Fursova, O., Pogorelko, G., and Zabotina, O. A. (2012). An efficient method for transient gene expression in monocots applied to modify the *Brachypodium distachyon* cell wall. *Annals of Botany* 110, 47–56.

García-Plazaola, J. I., Fernández-Marín, B., Duke, S. O., Hernández, A., López-Arbeloa, F., and Becerril, J. M. (2015). Autofluorescence: Biological functions and technical applications. *Plant Science* 236, 136–145.

Germini, D., Tsfasman, T., Zakharova, V. V., Sjakste, N., Lipinski, M., and Vassetzky, Y. (2018). A Comparison of Techniques to Evaluate the Effectiveness of Genome Editing. *Trends in Biotechnology* 36, 147–159.

Ghedira, R., De Buck, S., Van Ex, F., Angenon, G., and Depicker, A. (2013). T-DNA transfer and T-DNA integration efficiencies upon *Arabidopsis thaliana* root explant cocultivation and floral dip transformation. *Planta* 238, 1025–1037.

Gheysen, G., Villarroel, R., and Van Montagu, M. (1991). Illegitimate recombination in plants: A model for T-DNA integration. *Genes and Development* 5, 287–297.

Göhring, J., Fulcher, N., Schilcher, K., Barta, A., and Jacak, J. (2014). Suitable transfection methods for single particle tracing in plant suspension cells. *Plant Methods* 10, 15.

Gorbunova, V., and Levy, A. A. (1997). Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. *Nucleic Acids Research* 25, 4650–4657.

Griesbach, R. J. (1987). Chromosome-mediated transformation via microinjection. *Plant Science* 50, 69–77.

Griesbach, R. J. (1994). An improved method for transforming plants through electrophoresis. *Plant Science* 102, 81–89.

Güell, M., Yang, L., and Church, G. M. (2014). Genome editing assessment using CRISPR Genome Analyzer (CRISPR-GA). *Bioinformatics* 30, 2968–2970.

Hamada, H., Linghu, Q., Nagira, Y., Miki, R., Taoka, N., and Imai, R. (2017). An *in planta* biolistic method for stable wheat transformation. *Scientific Reports* 7, 11443.

Hammami, R., Cuadrado, A., Friero, E., Jouve, N., Soler, C., and González, J. M. (2011). Callus induction and plant regeneration from immature embryos of *Brachypodium distachyon* with different chromosome numbers. *Biologia Plantarum* 55, 797–800.

Hayashi, K., and Yandell, D. W. (1993). How sensitive is PCR-SSCP? *Human Mutation* 2, 338–346.

Hess, D., Dressler, K., and Nimmrichter, R. (1990). Transformation experiments by pipetting *Agrobacterium* into the spikelets of wheat (*Triticum aestivum* L.). *Plant Science* 72, 233–244.

Holm, P. B., Olsen, O., Schnorf, M., Brinch-Pedersen, H., and Knudsen, S. (2000). Transformation of barley by microinjection into isolated zygote protoplasts. *Transgenic Research* 9, 21–32.

Hong, S. Y., Seo, P. J., Cho, S. H., and Park, C. M. (2012). Preparation of leaf mesophyll protoplasts for transient gene expression in *Brachypodium distachyon*. *Journal of Plant Biology* 55, 390–397.

Hua, Y., Wang, C., Huang, J., and Wang, K. (2017). A simple and efficient method for CRISPR/Cas9-induced mutant screening. *Journal of Genetics and Genomics* 44, 207–213.

Huang, C.-N.-N., Cornejo, M. J., Bush, D. S., and Jones, R. L. (1986). Estimating Viability of Plant Protoplasts Using Double and Single Staining. *Protoplasma* 135, 80–87.

Huize, C., Jiao, N. I. U., Meiting, D. U., Rong, H. A. N., Weizhong, L. I. U. (2020). Genetic transformation method of *Brachypodium distachyon*. CN Patent No 111,500,622. Beijing: China National Intellectual Property Administration.

Hus, K., Betekhtin, A., Pinski, A., Rojek-Jelonek, M., Grzebelus, E., Nibau, C., Gao, M., Jaeger, K. E., Jenkins, G., Doonan, J. H., and Hasterok, R. (2020). A CRISPR/Cas9-Based Mutagenesis Protocol for *Brachypodium distachyon* and Its Allopolyploid Relative, *Brachypodium hybridum*. *Frontiers in Plant Science* 11, 614.

Inazuka, M., Wenz, H. M., Sakabe, M., Tahira, T., and Hayashi, K. (1997). A Streamlined Mutation Detection System: Multicolor Post-PCR Fluorescence Labeling and Single-Strand Conformational Polymorphism Analysis by Capillary Electrophoresis. *Genome Research* 7, 1094–1103.

Ismagul, A., Yang, N., Maltseva, E., Iskakova, G., Mazonka, I., Skiba, Y., Bi, H., Eliby, S., Jatayev, S., Shavrukov, Y., Borisjuk, N., and Langridge, P. (2018). A biolistic method for high-throughput production of transgenic wheat plants with single gene insertions. *BMC Plant Biology* 18, 135.

Jabnoune, M., Secco, D., Lecampion, C., Robaglia, C., Shu, Q., and Poirier, Y. (2015). An Efficient Procedure for Protoplast Isolation from Mesophyll Cells and Nuclear Fractionation in Rice. *Bio-Protocol* 5, e1412.

Jackson, M. A., Anderson, D. J., and Birch, R. G. (2013). Comparison of *Agrobacterium* and particle bombardment using whole plasmid or minimal cassette for production of high-expressing, low-copy transgenic plants. *Transgenic Research* 22, 143–151.

Jeon, W. B., Lee, M. B., Kim, D. Y., Hong, M. J., Lee, Y. J., and Seo, Y. W. (2010). Efficient Phosphinothricin Mediated Selection of Callus Derived from *Brachypodium* Mature Seed. *Korean Journal of Breeding Science* 42, 351–356.

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B., and Weeks, D. P. (2013). Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic acids research* 41, e188.

Johansen, L. K., and Carrington, J. C. (2001). Silencing on the spot. Induction and suppression of RNA silencing in the *Agrobacterium*-mediated transient expression system. *Plant Physiology* 126, 930–938.

Jung, H.-I., Gayomba, S. R., Yan, J., and Vatamaniuk, O. K. (2014). *Brachypodium distachyon* as a model system for studies of copper transport in cereal crops. *Frontiers in plant science* 5, 236.

Jung, H. Il, Yan, J., Zhai, Z., and Vatamaniuk, O. K. (2015). Gene functional analysis using protoplast transient assays. *Methods in Molecular Biology* 1284, 433–452.

Jung, H. Il, Zhai, Z., and Vatamaniuk, O. K. (2011). Direct transfer of synthetic double-stranded RNA into protoplasts of *Arabidopsis thaliana*. *Methods in molecular biology (Clifton, N.J.)* 744, 109–127.

Kikkert, J. R., Vidal, J. R., and Reisch, B. I. (2005). Stable transformation of plant cells by particle bombardment/biolistics. *Methods in molecular biology (Clifton, N.J.)* 286, 61–78.

Köhler, F., Cardon, G., Pöhlman, M., Gill, R., and Schieder, O. (1989). Enhancement of transformation rates in higher plants by low-dose irradiation: Are DNA repair systems involved in the

incorporation of exogenous DNA into the plant genome? *Plant Molecular Biology* 12, 189–199.

Kohli, A., Gahakwa, D., Vain, P., Laurie, D. A., and Christou, P. (1999). Transgene expression in rice engineered through particle bombardment: Molecular factors controlling stable expression and transgene silencing. *Planta* 208, 88–97.

Kotchoni, S. O., Gachomo, E. W., Betiku, E., and Shonukan, O. O. (2003). A home made kit for plasmid DNA mini-preparation. *African Journal of Biotechnology* 2, 86–87.

Lawrenson, T., Shorinola, O., Stacey, N., Li, C., Østergaard, L., Patron, N., Uauy, C., and Harwood, W. (2015). Induction of targeted, heritable mutations in barley and *Brassica oleracea* using RNA-guided Cas9 nuclease. *Genome biology* 16, 258.

Lee, M. B., Jeon, W. B., Kim, D. Y., Bold, O., Hong, M. J., Lee, Y. J., Park, J. H., and Seo, Y. W. (2011). Agrobacterium-mediated transformation of *Brachypodium distachyon* inbred line Bd21 with two binary vectors containing hygromycin resistance and GUS reporter genes. *Journal of Crop Science and Biotechnology* 14, 233–238.

Lihong, C., Gang, H., Yan, G. E., Jie, K. E., Lun, L. I. (2019). Transgenic plant and preparation method thereof. CN Patent No 110,438,147. Beijing: China National Intellectual Property Administration.

Lindsay, H., Burger, A., Biyong, B., Felker, A., Hess, C., Zaugg, J., Chiavacci, E., Anders, C., Jinek, M., Mosimann, C., and Robinson, M. D. (2016). CrispRVariants charts the mutation spectrum of genome engineering experiments. *Nature Biotechnology* 34, 701–702.

Liu, H., Kawabe, A., Matsunaga, S., Murakawa, T., Mizukami, A., Yanagisawa, M., Nagamori, E., Harashima, S., Kobayashi, A., and Fukui, K. (2004). Obtaining transgenic plants using the bio-active beads method. *Journal of Plant Research* 117, 95–99.

Liu, J., Nannas, N. J., Fu, F. F., Shi, J., Aspinwall, B., Parrott, W. A., and Dawe, R. K. (2019). Genome-scale sequence disruption following biolistic transformation in rice and maize. *Plant Cell* 31, 368–383.

Liu, J., Su, Q., An, L., and Yang, A. (2009). Transfer of a minimal linear marker-free and vector-free smGFP cassette into soybean via ovary-drip transformation. *Biotechnology Letters* 31, 295–303.

Lowe, B. A., Prakash, N. S., Way, M., Mann, M. T., Spencer, T. M., and Boddupalli, R. S. (2009). Enhanced single copy integration events in corn via particle bombardment using low quantities of DNA. *Transgenic Research* 18, 831–840.

Luo, Z. xun, and Wu, R. (1989). A simple method for the transformation of rice via the pollen-tube pathway. *Plant Molecular Biology Reporter* 7, 69–77.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., and Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science* 339, 823–826.

Maximova, S. N., Dandekar, A. M., and Guiltinan, M. J. (1998). Investigation of *Agrobacterium*-mediated transformation of apple using green fluorescent protein: High transient expression and low stable transformation suggest that factors other than T-DNA transfer are rate-limiting. *Plant Molecular Biology* 37, 549–559.

Mayerhofer, R., Koncz-kalman, Z., Nawrath, C., Bakkeren, G., Crameri, A., Angelis, K., Redei, G. P., Schell, J., Hohn, B., and Koncz, C. (1991). T-DNA integration: A mode of illegitimate recombination in plants. *EMBO Journal* 10, 697–704.

Mazarei, M., Al-Ahmad, H., Rudis, M. R., and Stewart, C. N. (2008). Protoplast isolation and transient gene expression in switchgrass, *Panicum virgatum* L. *Biotechnology journal* 3, 354–9.

Miao, Y., and Jiang, L. (2007). Transient expression of fluorescent fusion proteins in protoplasts of suspension cultured cells. *Nature Protocols* 2, 2348–2353.

Mitchell, J. C., and Petolino, J. F. (1991). Plant Regeneration from Haploid Suspension and Protoplast Cultures from Isolated Microspores of Maize. *Journal of Plant Physiology* 137, 530–536.

Miyaoka, Y., Berman, J. R., Cooper, S. B., Mayerl, S. J., Chan, A. H., Zhang, B., Karlin-Neumann, G. A., and Conklin, B. R. (2016a). Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Scientific Reports* 6, 23549.

Miyaoka, Y., Chan, A. H., and Conklin, B. R. (2016b). Using digital polymerase chain reaction to detect single-nucleotide substitutions induced by genome editing. *Cold Spring Harbor Protocols* 2016, 688–692.

Mizuno, K., Takahashi, W., Beppu, T., Shimada, T., and Tanaka, O. (2005). Aluminum borate whisker-mediated production of transgenic tobacco plants. *Plant Cell, Tissue and Organ Culture* 80, 163–169.

Müller, S. M., Galliardt, H., Schneider, J., George Barisas, B., and Seidel, T. (2013). Quantification of Förster resonance energy transfer by monitoring sensitized emission in living plant cells. *Frontiers in Plant Science* 4, 413.

Murakawa, T., Kajiyama, S., and Fukui, K. (2008a). Improvement of bioactive bead-mediated transformation by concomitant application of electroporation. *Plant Biotechnology* 25, 387–390.

Murakawa, T., Kajiyama, S., Ikeuchi, T., Kawakami, S., and Fukui, K. (2008b). Improvement of transformation efficiency by bioactive-beads-mediated gene transfer using DNA-lipofectin complex as entrapped genetic material. *Journal of Bioscience and Bioengineering* 105, 77–80.

Narasimhulu, S. B., Deng, X. B., Sarria, R., and Gelvin, S. B. (1996). Early transcription of *Agrobacterium* T-DNA genes in tobacco and maize. *Plant Cell* 8, 873–886.

O'Kennedy, M. M., Stark, H. C., and Dube, N. (2011). Biolistic-mediated transformation protocols for maize and pearl millet using pre-cultured immature zygotic embryos and embryogenic tissue. *Methods in molecular biology (Clifton, N.J.)* 710, 343–354.

Ohba, T., Yoshioka, Y., Machida, C., and Machida, Y. (1995). DNA rearrangement associated with the integration of T-DNA in tobacco: an example for multiple duplications of DNA around the integration target. *Plant Journal* 7, 157–164.

Ozdemir, B. S., and Budak, H. (2018). Application of tissue culture and transformation techniques in model species *Brachypodium distachyon*. *Methods in molecular biology (Clifton, N.J.)*. 1667, 289–310.

Păcurar, D. I., Thordal-Christensen, H., Nielsen, K. K., and Lenk, I. (2008). A high-throughput *Agrobacterium*-mediated transformation system for the grass model species *Brachypodium distachyon* L. *Transgenic Research* 17, 965–975.

Park, J., Lim, K., Kim, J. S., and Bae, S. (2017). Cas-analyzer: An online tool for assessing genome editing results using NGS data. *Bioinformatics* 33, 286–288.

Peng, L., Zhaoqing, C. (2020). Method for rapid and efficient genetic transformation of *Brachypodium distachyon* by inflorescence dip dyeing. CN Patent No 110,669,781. Beijing: China National Intellectual Property Administration.

Pescitelli, S. M., and Sukhapinda, K. (1995). Stable transformation via electroporation into maize Type II callus and regeneration of fertile transgenic plants. *Plant Cell Reports* 14, 712–716.

Petersen, B. L., Möller, S. R., Mravec, J., Jørgensen, B., Christensen, M., Liu, Y., Wandall, H. H., Bennett, E. P., and Yang, Z. (2019). Improved CRISPR/Cas9 gene editing by fluorescence activated cell sorting of green fluorescence protein tagged protoplasts. *BMC Biotechnology* 19, 36.

Pinello, L., Canver, M. C., Hoban, M. D., Orkin, S. H., Kohn, D. B., Bauer, D. E., and Yuan, G. C. (2016). Analyzing CRISPR genome-editing experiments with CRISPResso. *Nature Biotechnology* 34, 695–697.

Riggs, C. D., and Bates, G. W. (1986). Stable transformation of tobacco by electroporation: evidence for plasmid concatenation. *Proceedings of the National Academy of Sciences of the United States of America* 83, 5602–5606.

Salmon, S., Lemoine, R., Jamai, A., Bouché-Pillon, S., and Christophe Fromont, J. (1995). Study of sucrose and mannitol transport in plasma-membrane vesicles from phloem and non-phloem tissues of celery (*Apium graveolens* L.) petioles. *Planta* 197, 76–83.

Salomon, S., and Puchta, H. (1998). Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. *EMBO Journal* 17, 6086–6095.

Samarut, É., Lissouba, A., and Drapeau, P. (2016). A simplified method for identifying early CRISPR-induced indels in zebrafish embryos using High Resolution Melting analysis. *BMC Genomics* 17, 547.

Shan, Q., Wang, Y., Chen, K., Liang, Z., Li, J., Zhang, Y. Y., Zhang, K., Liu, J., Voytas, D. F., Zheng, X., Zhang, Y. Y., and Gao, C. (2013). Rapid and efficient gene modification in rice and *Brachypodium* using TALENs. *Molecular Plant* 6, 1365–1368.

Singer, K. (2018). "The mechanism of T-DNA integration: Some major unresolved questions," in *Agrobacterium Biology: from Basic Science to Biotechnology*, ed. Gelvin, S.B. (Gewerbestrasse, Springer International), 287-317.

Sone, T., Nagamori, E., Ikeuchi, T., Mizukami, A., Takakura, Y., Kajiyama, S., Fukusaki, E. I., Harashima, S., Kobayashi, A., and Fukui, K. (2002). A novel gene delivery system in plants with calcium alginate micro-beads. *Journal of Bioscience and Bioengineering* 94, 87–91.

Sparks, C. A., and Jones, H. D. (2004). "Transformation of Wheat by Biolistics," in *Transgenic Crops of the World: Essential Protocols*, ed. Curtis, I.S. (Dordrecht: Springer), 19–34.

Sparks, C. A., and Jones, H. D. (2014). Genetic transformation of wheat via particle bombardment. *Methods in Molecular Biology* 1099, 201–218.

Sparks, C. A., and Jones, H. D. (2016). "Wheat biotechnology: recent developments and future trends.," in *Biotechnology of major cereals*, ed. Jones, H. D. (Wallingford: CABI International), 88–103.

Srivastava, V., Vasil, V., and Vasil, I. K. (1996). Molecular characterization of the fate of transgenes in transformed wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics* 92, 1031–1037.

Stoger, E., Williams, S., Keen, D., and Christou, P. (1998). Molecular characteristics of transgenic wheat and the effect on transgene expression. *Transgenic Research* 7, 463–471.

Takano, M., Egawa, H., Ikeda, J. E., and Wakasa, K. (1997). The structures of integration sites in transgenic rice. *Plant Journal* 11, 353–361.

Thomas, H. R., Percival, S. M., Yoder, B. K., and Parant, J. M. (2014). High-throughput genome editing and phenotyping facilitated by high resolution melting curve analysis. *PLoS ONE* 9, e114632.

Thompson, J. A., Drayton, P. R., Frame, B. R., Wang, K., and Dunwell, J. M. (1995). Maize

transformation utilizing silicon carbide whiskers: a review. *Euphytica* 85, 75–80.

Tsai, S. Q., Wyvekens, N., Khayter, C., Foden, J. A., Thapar, V., Reyon, D., Goodwin, M. J., Aryee, M. J., and Joung, J. K. (2014). Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nature Biotechnology* 32, 569–576.

Underhill, P. A., Jin, L., Lin, A. A., Qasim Mehdi, S., Jenkins, T., Vollrath, D., Davis, R. W., Cavalli-Sforza, L. L., and Oefner, P. J. (1997). Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Research* 7, 996–1005.

Vogel, J., and Hill, T. (2008). High-efficiency *Agrobacterium*-mediated transformation of *Brachypodium distachyon* inbred line Bd21-3. *Plant Cell Reports* 27, 471–478.

Vogel, J. P., Garvin, D. F., Leong, O. M., and Hayden, D. M. (2006). *Agrobacterium*-mediated transformation and inbred line development in the model grass *Brachypodium distachyon*. *Plant Cell, Tissue and Organ Culture* 84, 199–211.

Vouillot, L., Thélie, A., and Pollet, N. (2015). Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3: Genes, Genomes, Genetics* 5, 407–415.

Vunsh, R. (2018). Methods for functional transgenics: Development of highly efficient transformation protocol in *Brachypodium* and its suitability for advancing *Brachypodium* transgenics. *Methods in molecular biology (Clifton, N.J.)*, 1667, 101–117.

Werr, W., and Lörz, H. (1986). Transient gene expression in a *Gramineae* cell line - A rapid procedure for studying plant promoters. *MGG Molecular & General Genetics* 202, 471–475.

Wittwer, C. T., Reed, G. H., Gundry, C. N., Vandersteen, J. G., and Pryor, R. J. (2003). High-Resolution Genotyping by Amplicon Melting Analysis Using LCGreen. *Clinical Chemistry* 49, 853–860.

Xiaoquan, Q. I., Suzhen, Z. (2013). Culture medium for cultivating transgenic plant. CN Patent No 103,320,377. Beijing: China National Intellectual Property Administration.

Xie, K., and Yang, Y. (2013). RNA-guided genome editing in plants using a CRISPR/Cas system. *Molecular Plant* 6, 1975–83.

Xingrong, W. U. (2016). Methods for plant transformation. U.S. Patent No 2016/0090602. Washington, DC: U.S. Patent and Trademark Office.

Xue, L. J., and Tsai, C. J. (2015). AGEseq: Analysis of genome editing by sequencing. *Molecular Plant* 8, 1428–1430.

Yang, A., Su, Q., and An, L. (2009). Ovary-drip transformation: A simple method for directly generating vector- and marker-free transgenic maize (*Zea mays* L.) with a linear GFP cassette transformation. *Planta* 229, 793–801.

Yao, Q., Cong, L., Chang, J. L., Li, K. X., Yang, G. X., and He, G. Y. (2006). Low copy number gene transfer and stable expression in a commercial wheat cultivar via particle bombardment. *Journal of Experimental Botany* 57, 3737–3746.

Yoo, S.-D., Cho, Y.-H., and Sheen, J. (2007). Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nature Protocols* 2, 1565–1572.

You, M. K., Lim, S.-H., Kim, M.-J., Jeong, Y. S., Lee, M.-G., and Ha, S.-H. (2014). Improvement of the fluorescence intensity during a flow cytometric analysis for rice protoplasts by localization of a green fluorescent protein into chloroplasts. *International journal of molecular sciences* 16, 788–804.

Yu, B., Sawyer, N. A., Chiu, C., Oefner, P. J., and Underhill, P. A. (2006). DNA mutation detection using denaturing high-performance liquid chromatography (DHPLC). *Current Protocols in Human Genetics*, 48, 1-14.

Yu, G., Wang, J., Miao, L., Xi, M., Wang, Q., and Wang, K. (2019). Optimization of mature embryo-based tissue culture and *Agrobacterium*-mediated transformation in model grass *Brachypodium distachyon*. *International Journal of Molecular Sciences* 20, 5448.

Zale, J. M., Agarwal, S., Loar, S., and Steber, C. M. (2009). Evidence for stable transformation of wheat by floral dip in *Agrobacterium tumefaciens*. *Plant Cell Reports* 28, 903–913.

Zhai, Z., Jung, H. Il, and Vatamaniuk, O. K. (2009). Isolation of protoplasts from tissues of 14-day-old seedlings of *Arabidopsis thaliana*. *Journal of Visualized Experiments*, 1149.

Zhang, H. M., Yang, H., Rech, E. L., Golds, T. J., Davis, A. S., Mulligan, B. J., Cocking, E. C., and Davey, M. R. (1988). Transgenic rice plants produced by electroporation-mediated plasmid uptake into protoplasts. *Plant Cell Reports* 7, 379–384.

Zhang, H., Zhao, F., Zhao, Y., Guo, C., Li, C., and Xiao, K. (2009). Establishment of transgenic cotton lines with high efficiency via pollen-tube pathway. *Frontiers of Agriculture in China* 3, 359–365.

Zhu, X., Xu, Y., Yu, S., Lu, L., Ding, M., Cheng, J., Song, G., Gao, X., Yao, L., Fan, D., Meng, S., Zhang, X., Hu, S., and Tian, Y. (2014). An efficient genotyping method for genome-modified animals and human cells generated with CRISPR/Cas9 system. *Scientific Reports* 4, 1–8.

Zombori, Z., Szécsényi, M., and Györgyey, J. (2011). Different approaches for *Agrobacterium*-mediated genetic transformation of *Brachypodium distachyon*, a new model plant for temperate grasses. *Acta Biologica Szegediensis* 55, 193–195.

Chapter 3. Validation of heterologous promoters for sgRNA expression in *B. distachyon.*

3.1 Introduction.

3.1.1. Editing efficiencies will be affected by the expression system used.

When using CRISPR/Cas9 in genome editing experiments based on the introduction of plasmid DNA to express the editing machinery, different factors can lead to divergences in editing efficiency. In the general introduction, several modifications to the CRISPR/Cas9 system were introduced, together with the optimization of the system for its use as a lab tool including Cas9 optimizations and sgRNA optimization. As discussed in Chapter 2, the method used for the delivery of the editing machinery is a major determinant of the type of edit screening, which translates in the description of editing efficiency. When delivering the editing machinery as DNA, the editing machinery will have to be expressed in the host organism to be able to induce the edits. Proper expression of any introduced heterologous gene will depend on regulatory sequences (Smale and Kadonaga, 2003; Venter and Botha, 2010), including promoters, enhancers and/or silencers (Timko et al., 1985; Guilfoyle, 1997; Singh, 1998; Butler and Kadonaga, 2002). In case of the Cas9, constitutive promoters are usually chosen due to their strong and consistent expression thorough the plant (Potenza et al., 2004). These include CaMV 35S (Jones et al., 1985; Odell et al., 1985; Fluhr et al., 1986; Battraw and Hall, 1990; Benfey et al., 1990), Maize Ubiquitin 1 (Callis et al., 1990; Christensen et al., 1992; Christensen and Quail, 1996) and Rice Actin 1 promoters (McElroy et al., 1990). These promoters, whose main function is to express proteins, will use the RNA polymerase II for the transcription of the RNA (Singh, 1998; Smale and Kadonaga, 2003; Potenza et al., 2004). Different strong constitutive promoters have been used for the expression of CRISPR/Cas9 (See Supplementary Table 1). Additionally, other promoters such as tissue specific (Wang et al., 2015; Yan et al., 2015) or inducible (Nandy et al., 2019; Barone et al., 2020; Wang et al., 2020) have been used. In the case of *Brachypodium*, other endogenous constitutive promoters have been studied (Coussens et al., 2012), which in the future could become of common use in crops. The effects of different expression systems for Cas9 expression has been studied and differences between systems have been observed (Mikami et al., 2015), being a factor that can be optimized in different species to maximize editing efficiency.

It has also been observed that the expression of the sgRNA will be key for the editing efficiency, both on and off target (Jinek et al., 2013). Usually, the expression of the sgRNA has been obtained using promoters that operate by complexing with the RNA polymerase III (Jinek et al., 2012, see Supplementary Table 1). In case of plants, several promoters have been used for the expression of the sgRNA, with the U6 and U3 promoter being the most common (Gao and Zhao, 2014; Yoshioka et al., 2015, see Supplementary Table 1). Originally, these genes control the constitutive expression of a small nuclear RNA involved in splicing and ribosome function (Kunkel et al., 1986; Kunkel and Pederson, 1988; Paule and White, 2000; Nabavi and Nazar, 2008). These promoters contain short internal termination sites, which supports the expression of short RNA (Arimbasseri et al., 2013). These types of promoters have already been used for the production of RNAi (Nie et al., 2010). However, there are some limitations to the use of these promoters, such as the requirement of a G at the beginning of the target sequence in case of the U6 promoter or an A in case of the U3 or the impossibility of their use *in vitro* due to the lack of RNA polymerase III (Zhang et al., 2017). The use of RNA polymerase II dependant promoters is not recommended for sgRNA expression due to RNA maturation process related to their transcription and their transport to the cytosol (Zhang et al., 2017). In addition to the general restrictions of the RNA polymerase II dependant promoters, it has been observed that foreign promoters of this type tend to have a reduced expression in heterologous systems (Bortesi and Fischer, 2015). In Čermák et al. (2017), the use of polycistronic, meaning several transcripts

guided by the same promoter and processed post-transcriptionally (Aeby et al., 2010; Nissim et al., 2014; Tang et al., 2016) to obtain the individual sgRNA was proposed, being based on a Pol II type promoter given the longer RNA expressed. This study proved the efficiency of the method in tomato, tobacco, *Medicago truncatula*, wheat and barley. A similar approach has also been used for Arabidopsis (Hui et al., 2019).

As shown in Supplementary Table 1, several U6 and U3 promoters have been described in different plant species. However, given the general restrictions described in the previous paragraph, the effectiveness of each promoter in each species should be assessed individually. An example of these type of studies is the one described in Ma et al. (2015), where different U6 and U3 promoters from rice (Shan et al., 2013a, 2013b; Xie and Yang, 2013; Zhou et al., 2014) and U3 and U6 promoters from Arabidopsis (Waibel and Filipowicz, 1990) were used for genome editing in rice. A different example would be Du et al. (2016), where the use of the endogenous soybean U6 promoter meant an increase in editing efficiency when comparing it to the use of the Arabidopsis U6 promoter. Similar findings have been described when comparing potato (Andersson et al., 2017) or cotton (Long et al., 2018) promoters in their respective species with the use of Arabidopsis RNA polymerase III promoters.

In case of *Brachypodium*, different promoters have been used, such as the endogenous *Brachypodium* U6 and rice U6 promoters (van der Schuren et al., 2018) or the rice U3 promoter (Raissig et al., 2016; Hus et al., 2020). However, given its potential as a model for cereals, the intercompatibility of different promoters among different cereals could open the way for other expression systems for the family.

3.1.2. Genes related to plant domestication are an interesting target for CRISPR/Cas9 studies in *Brachypodium*.

Several genes have been described to affect traits related to the use of plants in controlled conditions, affecting traits such as flowering time (Yan et al., 2004), height (Ashikari et al., 2005; Lawit et al., 2010; Corvalán and Choe, 2017; Xu et al., 2017), seed yield (Ashikari et al., 2005; Zalewski et al., 2012), dormancy (Gu et al., 2010; Chono et al., 2015; Sato et al., 2016) and shattering (Dong and Wang, 2015).

The rice *CYTOKININ OXIDASE-DEHYDROGENASE 2* (*OsCKX2*) has been found to affect grain production (Ashikari et al., 2005). This gene, part of the Gn1 QTL, has a role in cytokinin degradation, a phytohormone whose accumulation in the reproductive meristems has been linked to increases in grain number in cereals (Chen et al., 2020). In Ashikari et al. (2005), sequence analysis of different high performance rice varieties (Habataki, Koshikari and 5150) showed deletions in *OsCKX2*. The effect of this gene on grain yield was confirmed by over expressing the gene, leading to reduced yield, and by silencing using RNAi, leading to an increase in grain yield. This gene has been identified in other cereals (Mameaux et al., 2012), including *Brachypodium* (Shan et al., 2013a).

The rice variety “Habataki” also has a smaller size, which has been associated with the *GIBBERELLIN 20 OXIDASE* (*OsGA20ox*, Ashikari et al., 2005). This gene has a role in gibberellin synthesis, which has been related to plant growth (Brain, 1959). Genes involved in this pathway have been studied in different grasses (Pearce et al., 2015). The case of the barley gene *SEMI-DWARF 1* (*HvSDW1*), with the same function as *OsGA20ox*, is remarkable, where two semi dwarf varieties, “Baudin” and “Diamant”, showed a deletion in the first exon of the gene (Xu et al., 2017).

The "Habataki" variant also shows reduced height. This has been related to a deletion in the *GIBERELLIN 20 OXIDASE* gene (GA20ox2, Ashikari et al., 2005). Similar findings have been observed in barley, comparing tall barley varieties and the semi-dwarves "Baudin" and "Diamant" (Xu et al., 2017). This gene has also been described in *Brachypodium* (*BdGA20ox*), linking the reduction of its expression in drought conditions to growth reduction of the plant (Verelst et al., 2013).

Another gene affecting plant height is *DWARF 4,* which was described first in Arabidopsis (*AtDWF4,* Azpiroz et al., 1998; Choe et al., 1998). The effect of this gene on Brassinosteroid synthesis, ultimately leading to variations in cell elongation, has been reported and confirmed by the recovery of normal height when providing brassinolides to *dwf4* mutants (Azpiroz et al., 1998). The ortholog of this gene in *Brachypodium* has been described and confirmed by complementation with *AtDWF4* (Corvalán and Choe, 2017).

The control of flowering time, more specifically the dependance of some plants on vernalization to be able to flower, is also an interesting trait for plant domestication (Dennis and Peacock, 2009). In case of *Brachypodium*, several genes have been described to have an effect on the vernalization requirement (Woods et al., 2017). The expression of *VERNALIZATION 3* (*BdVRN3*), with a similar role as the Arabidopsis florigen *FLOWERING LOCUS T* (*AtFT*, Corbesier et al., 2007; Zeevaart, 2008), is repressed by *VERNALIZATION 2* (*BdVRN2*), whose expression in turn is repressed by *VERNALIZATION 1* (*BdVRN1*). *REPRESSOR OF VERNALIZATION 1* (*BdRVR1*) will repress the expression *BdVRN1*, which ultimately implies the repression of *BdVRN3*. In Woods et al. (2017), two *rvr1* mutants were studied, observing the loss of the vernalization requirement given the activation of the florigen. The role of *VRN2* in vernalization has also been described in diploid wheat (*TmVRN2*, Yan et al., 2004).

3.1.3. Genomic environment could affect editing efficiencies.

Proper Cas9 function will rely on the interaction of the sgRNA and Cas9 complex successfully binding to the target DNA, as stated in the general introduction. However, different regions of the genomic DNA are differently accessible to proteins depending different factors, such as DNA compaction. The most basic level of compaction is due to histones forming nucleosomes with around 147 bp wrapped around an octamer of H2A, H2B, H3, and H4 histones (Luger et al., 1997, 2012; Kouzarides, 2007). These are able to interact with other factors forming other structures during interphase (Happel and Doenecke, 2009; Bian and Belmont, 2012). During interphase, the genetic material in the nucleus can be divided in euchromatin and heterochromatin, with low and high levels of condensation respectively (Filion et al., 2010; Doenecke, 2017). The centromeres will be located in the high density heterochromatin (Pidoux and Allshire, 2005). The levels of compaction of the DNA will be dynamic thorough the cell cycle (Hémonnot et al., 2016), leaving different parts of the DNA open for its reading at different times.

The effect of chromatin dynamics on genome editing has been studied in humans and using *in vitro* studies, observing that CRISPR/Cas9 function is not impaired by cytosine methylation, but nucleosomes are able to block it, observing also a correlation of increased gene transcription and editing efficiency (Chen et al., 2016, 2017; Isaac et al., 2016; Daer et al., 2017; Jensen et al., 2017; Verkuijl and Rots, 2019). The effects of chromatin compaction on DSB repair pathway choice have also been studied (Janssen et al., 2019). These results have been observed *in vitro* and in human cells, but not using plant genomes as the target DNA.

An interesting platform for the study of sgRNA in different locations of the genome without using different target sequences would be the use of T-DNA insertions. As discussed in the

general introduction, Bragg et al. (2012) described a series of T-DNA lines with inserts in different positions in the genome with different objectives (Figure 3.1), including gene trapping (pJJ2LBP and pJJLBP2), activation tags (pJJ2LBA) and insertional mutagenesis (pOL001, pJJH, pJJB, pJJ2LB, and pJJB2LB). All these inserts share the same antibiotic resistance for its selection, HygR, providing a known sequence that can be found in different parts of the genome. The use of sgRNA targeting HygR in different T-DNA insertion lines could be used as a platform to assess editing efficiencies in different positions of the genome.

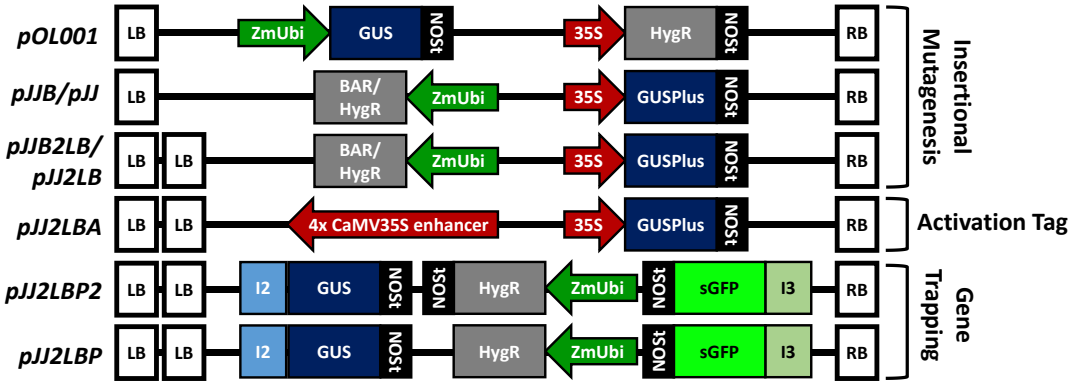


Figure 3.1. Different T-DNA were used for the development of the T-DNA library described in Bragg et al. (2012).

Diagrams of the different T-DNA regions of the plasmids used to obtain the library described in Bragg et al. (2012). In the left column, the names of the plasmids the T-DNA belong to are shown. The brackets on the right column show the different effects of the T-DNA insertions in the plant. In white boxes are shown the borders of the T-DNA (LB or RB). Green arrows represent the maize Ubiquitin promoter with an intron (ZmUbi). Red arrows represent CaMV 35S promoters (35S) and regulatory sequences (4xCaMV35S enhancer). Blue boxes represent the β-glucuronidase coding sequences (GUS and GUSPlus). Green boxes represent the GFP coding sequence. Grey boxes represent the antibiotic resistance gene, either hygromycin phosphotransferase (hygromycin resistance, HygR) or phosphinothricin acetyltransferase (BAR). The blue boxes labelled I2 represent the rice tubulin intron sequence and I3 the rice tubulin intron. The black boxes labelled as NOSt represent the NOS terminator sequence. Image adapted from Bragg et al. (2012).

3.2. Hypothesis and objectives of the chapter.

The first objective of this chapter was to identify possible target genes for genome editing in *Brachypodium*. These were chosen upon their relevance in the domestication of other cereals. Then, sgRNA were designed targeting the coding regions of these genes. One of the targets, *BdDWF4*, was selected for assessing the efficiency of different promoters expressing the sgRNA in *Brachypodium*. Additionally, the Hygromycin resistance cassette was also chosen as a target for editing. Finally, a T-DNA library-based assay was proposed for the study of the effects of chromosome location and compaction of the genomic DNA on editing efficiencies.

3.3. Materials and Methods.

3.3.1. Plasmids used in this chapter.

Different plasmids were used in this chapter. The plasmids pHvDGV and pActH2BGFP have been described in Chapter 2.

The pDON2 plasmid (Figure 3.2) was described in van der Schuren et al. (2018) and was obtained from the author. The plasmid contains the elements necessary for its replication in *E. coli*, together with a Kanamycin resistance cassette. It also contains the rrnB T1 and T2 sequences. In addition to that, it contains the attP1 and 2 sites, enabling Gateway Cloning. Between those sites, two independent promoters (*Brachypodium distachyon* U6, BdU6p and *Oryza sativa* U6, OsU6p) control the expression of two different guide RNA cassettes.

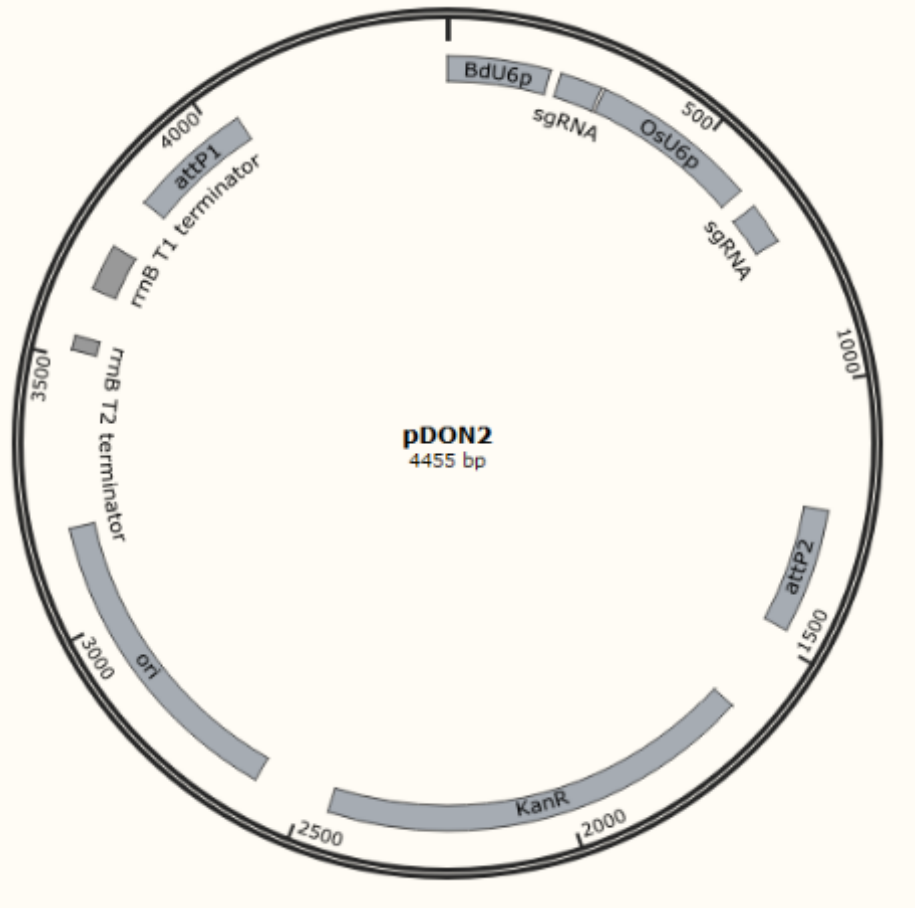


Figure 3.2. Plasmid map of the pDON2 plasmid described in van der Schuren et al. (2018).
Plasmid map of the pDON2 showing the key features.

The pENTR plasmid (Figure 3.3) was originally described in (Hus et al., 2020). It is similar to the pDON2 plasmid, containing the same elements but with a *Oryza sativa* U3 promoter controlling the sgRNA expression, in addition to the elements required for Gateway cloning, used for the insertion of the assembled sgRNA cassette into its original destination plasmid, pOsCas9 (Hus et al., 2020).

Figure 3.3. Plasmid map of the pENTR plasmid described in Hus et al. (2020).
Plasmid map of the pENTR plasmid showing the key features.

3.3.2. Single Guide Vectors (SGVs) ligations.

Two different Single Guide Vectors (SGV) were developed in this chapter in order to study the differences in sgRNA promoter efficiency. They were both assembled using the "backbone" of the pHvDGV, meaning the whole plasmid without the TaU6::sgRNA cassettes, which was swapped by the BdU6::sgRNA cassette from the pDON plasmid, forming the pBdSGV (Figure 3.4A), or the OsU3::sgRNA from the pENTR plasmid, forming the pOsSGV (Figure 3.4B).

This was done by specific amplification of the desired sequences and adding overhangs that allowed the use of Golden Gate ligations using SapI restriction enzyme. The amplification of the pHvDGV "backbone" was done using the PCR primers HvDGV gRNA GG Fw Bf and HvDGV gRNA GG Rv Bf (Supplementary Table 3) and the amplification of the sgRNA cassettes and promoters of the pDON2 and pENTR was done using pDON BdgRNA GG Fw Bf and pDON BdgRNA GG Rv Bf for the first one and pENTR OsgRNA GG Fw Bf and pDON BdgRNA GG Rv Bf for the latter (see Supplementary Table 3). The amplification of the bands was done using the Phusion Green Hot Start II High-Fidelity PCR Master Mix (Thermo Scientific, U.S.A.) following the manufacturer's protocol. The bands obtained were separated using agarose gel electrophoresis and isolated from said gel using the QIAquick® Gel Extraction Kit (Qiagen, Germany).

Given the overhangs introduced in the primers, Golden Gate assembly was possible. To do this, 150 ng of the backbone were mixed with 20.69 ng of the sgRNA cassette, together with 1 µl of T4 ligase (NEB, U.S.A.), 1 µl of 10x T4 ligase buffer, 1 µl of Sap I and a final volume of 10 µl. Five µl of the ligated plasmid was then transformed into *E. coli* using thermic shock and the

transformed cells were plated in LB Kanamycin medium. The insertion was confirmed using colony PCR with the primers Dual Guide Check Fw and Dual Guide Check Rv (See Supplementary Table 3), given the differences in size between the sgRNA cassettes in pHvDGV, pBdSGV and pOsSGV. The plasmids showing the expected amplicon were sent to sequence with the same primers for further confirmation of the correct insertion.

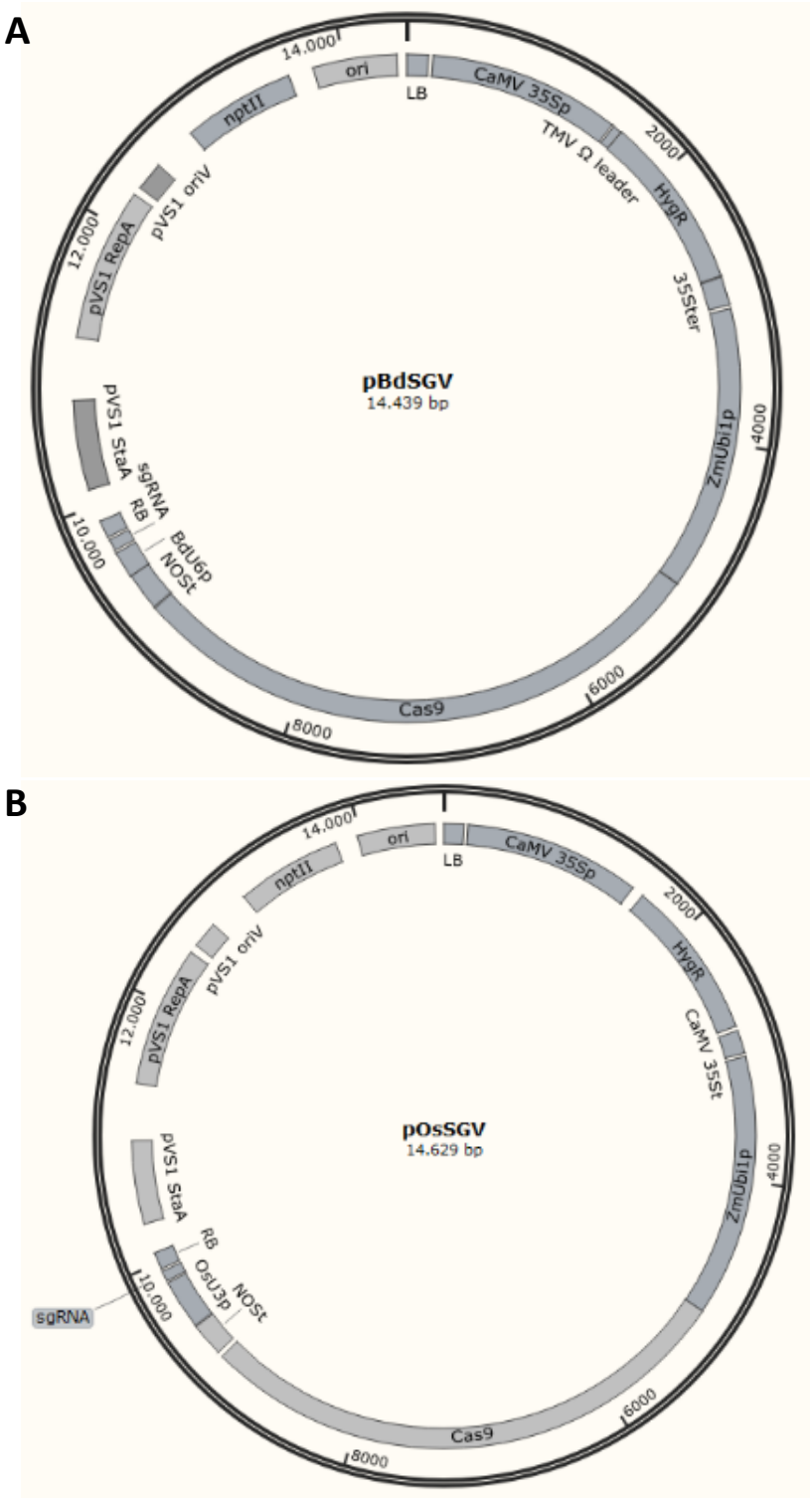


Figure 3.4. Plasmid maps of the SGV plasmids developed.
Plasmid map of the (A) BdSGV and (B) OsSGV showing the key features

3.3.3. sgRNA insertion in the plasmids of interest.

3.3.3.1. sgRNA design.

Different endogenous genes were used for sgRNA design: *BdCKX2*, *BdGA20ox2*, *BdRVR1*, *BdVRN2* and *BdDWF4*. sgRNA were also designed targeting the HygR cassette described in pBract204 and pJJ2LB (Bragg et al., 2012). The sequences of the endogenous genes were obtained from the references mentioned in the introduction of this chapter. The Hygromycin sequences used were obtained from the pBract204 and the HygR sequence described in Bragg et al. (2012), found in all the lines with hygromycin resistance described in said paper.

The gene sequences were uploaded to the online platform Benchling and potential target sites were identified using the integrated tool available, selecting single guide design, 20 base pairs of guide length and the standard NGG PAM site. Although the genome selected was *Brachypodium distachyon* v1.0, no annotation was used to specify coding sequences. All guides targeting restriction sites were collected in different tables (Supplementary Tables 4-10). It was considered that a target site contained a restriction site if a restriction enzyme recognition site was found within the first 2 nucleotides in the 3' end of the target site. To facilitate the screening using RE-PCR, unique restriction sites were targeted within a pre-designed amplicon. The final selection of targets used is specified in Table 3.1. Additionally, the HygR cassettes contained in pJJ2LB and pBract204 were aligned using the Benchling platform using the MAAFT Algorithm (Katoh and Standley, 2013). The analysis regarding the alignment of the different sgRNA scaffolds was performed using the same tool and parameters.

3.3.3.2. Insertion of the targeting sequences in the sgRNA cassette using Golden Gate assembly.

The protocol for sgRNA assembly was based on Golden Gate cloning. The process depended on overhangs designed to be specific to each one of the combinations of promoter and sgRNA sequence (Figure 3.5A). The oligonucleotides containing the gRNA flanked by the required overhangs were obtained from Thermo Scientific (U.S.A.), ordering both the direct and the complementary strands. The oligonucleotides were combined in a 2 µM dilution, which was stored at -20 ^{o}C until used. The Golden Gate ligation was set up with 200 ng of the destination plasmid, 2 µl of the oligonucleotide mix, 1 µl of T4 ligase (NEB, U.S.A.) 1 µl of 10x T4 ligase buffer, 0.5 µl of BsaI restriction enzyme (NEB, U.S.A.), in a final volume of 10 µl in a 0.2 ml PCR tube. The mixture was placed in a thermocycler and the Golden Gate digestion/ligation was ran with the cycles described in Figure 3.5B. Five µl of the ligation were transformed in *E. coli* using thermic shock, and the cells were plated on an LB Agar plate with Kanamycin. The obtained single colonies were then inoculated in 5 ml of LB medium with the pertinent antibiotic. Then, plasmid from those cultures was isolated using the QIAPrep™ Spin MiniPrep Kit (Qiagen, Germany), following the protocol described in the kit, which were then sent for sequencing using the specific primers NOSt Seq Fw and Dual Guide Check Rv (See Supplementary Table 3).

In the case of pHvDGV, the Golden Gate ligation implied swapping a LacZ cassette for the desired sgRNA. This made possible the use of white/blue selection. The plate used after transformation was supplemented with an X-Gal (Melford, U.K.) 40 ml/ml and IPTG (Melford, U.K.) 40 mg/ml, which was spread using a sterile Digralsky spreader. In addition to white/blue selection, the differences in base pairs between the sgRNA inserted and the LacZ cassette made possible to screen using colony PCR with Guide Check Fw and Dual Guide Check Rv (See Supplementary Table 3). The plasmid pHvDGV has another colour-based reporter in a second insertion site, an RFP, which gave the cells a pink tone when LacZ selection was not successful. For insertion of a second sgRNA, a similar process to the previously described was used, but with oligonucleotides

with different overhangs for directed insertion upon digestion using Esp3I (Figure 3.5A). After confirmation of the first insertion using white/blue selection, colony PCR and sequencing, 200 ng of the plasmid with the first insertion was used in a second Golden Gate ligation with the same mixture as previously described but using 1 µl of Esp3I (Thermo Scientific, U.S.A.). The ligation cycles used were the same as previously described. Then, *E. coli* was transformed as previously described and the transformed cells were plated on LB Kanamycin plates. In this case, the white/blue selection was not applied. However, the removal of the RFP implied the pink colouring did not occur. The difference in base pairs between the RFP and the inserted oligonucleotide allowed colony PCR screening using the same primers used for the previous colony PCR. The correct assembly of the guides would give the bands shown in Figure 3.7C. The correct assembly of the sgRNA cassette was confirmed using sequencing with NOSt Seq Fw and LacZ Rv after the first insertion and NOSt Seq Fw and Dual Guide Check Rv after the second insertion (See Supplementary Table 3).

After the insertion process was complete and confirmed using sequencing techniques, glycerol stocks were obtained to simplify the process of obtaining more plasmid.

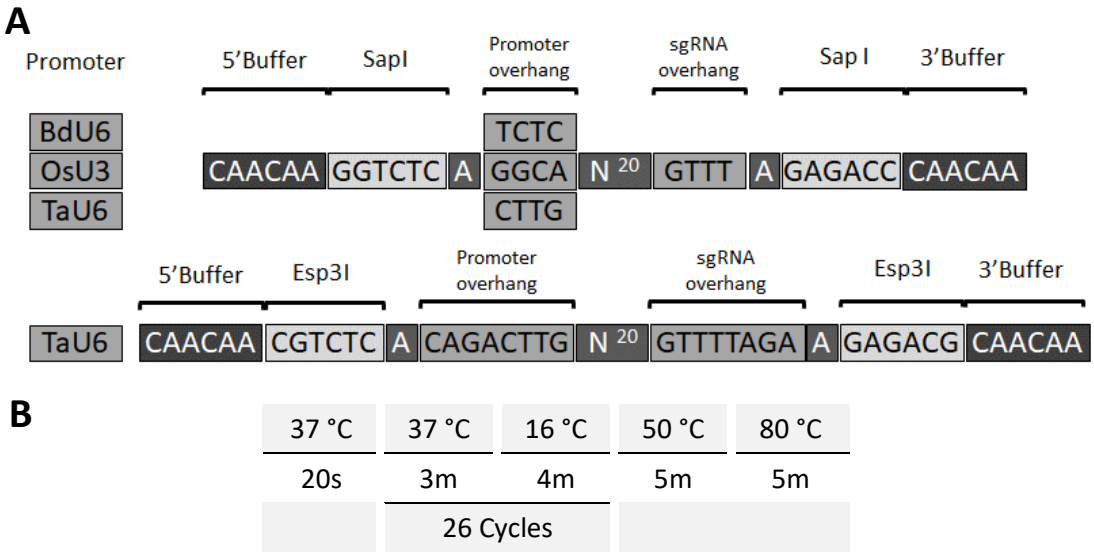


Figure 3.5. Details on the Golden Gate ligation for sgRNA assembly.
(A) Overhang design outline for the different promoters studied. The first three represent the overhangs necessary for the insertion of target sequences in the insertion sites of the SGV plasmids (BdU6 and OsU3) and the first insertion site of the pHvDGV (TaU6). The second type of overhang is specific for the TaU6 promoter found in the second insertion site of the pHvDGV plasmid. (B) Outline of the cycles used for Golden Gate assembly of the target sequences in all plasmids. The same cycles were used for the insertion in the SGV and both sites of the pHvDGV. Cycles required for Golden Gate ligation using BsaI and Esp3I.

3.3.4. Biolistic transformation.

Two different *Brachypodium* lines were used for the experiments described in this chapter. Bd21-3 was used for the experiments regarding the endogenous gene *BdDWF4*, whereas the line Bd21-3 GUS was used for the experiments regarding the Hygromycin resistance gene. Both plants were grown in the same conditions as described in Chapter 2. After 10 weeks, embryos isolated from these genotypes were transformed using biolistic delivery. The protocol used for biolistic transformation was as described in Chapter 2. The desired plasmids were isolated using the high-concentration protocol described in Chapter 2 and the DNA was diluted to a

concentration of 1 µg/µl. To obtain a quick method for screening the transformed embryos, the plasmids containing the editing machinery were co-bombarded with the pActH2BGFP. The total DNA added to the gold particles had to be 5 in µg total. Therefore, 3.5 µg of the plasmid containing the editing machinery and 1.5 µg of pActH2BGFP were used. Bd21-3 embryos were transformed with each one of the plasmids containing the editing machinery (pHvDGV, pBdSGV and pOsSGV) carrying guides targeting *BdDWF4* and the marker plasmid pActH2BGFP, whereas Bd21-3 GUS embryos were transformed with the plasmids containing the editing machinery carrying guides targeting HygR and pActH2BGFP.

The fluorescent screening of transformed embryos was done as described in Chapter 2. Positive embryos were collected in a 2ml Safe-Lock Eppendorf tube and snap frozen using liquid nitrogen. The samples were kept at -80 ºC until the samples were processed.

3.3.5. gDNA isolation.

gDNA was isolated from positive embryos using the DNeasy® Plant Mini Kit (Qiagen, Germany) following the protocol described by the manufacturer with minor modifications. Around 150 embryos were ground using a MM301 Mixer Mill Pulverizer Mixer (Retsch, Germany). The tube adapters were placed in the -80ºC freezer 30 minutes before grinding. One steel bead was placed inside each tube and the tubes were shaken at 30 shakes per second for 2 minutes. Then, 200 µl of the AP1 buffer were added and mixed by inversion with the homogenised tissue, prior to removing the beads with sterile forceps. Then, another 200 µl of AP1 buffer and 4 µl of RNase A were added to each tube. The samples were incubated in a pre-heated water bath at 65 ºC for 10 minutes, inverting the tubes once every 3 minutes. The tubes were placed in ice and 130 µl of AP2 buffer were added, incubating them in ice for 5 minutes. Then, the samples were centrifuged at maximum speed for 5 minutes in a tabletop centrifuge. The supernatant was pipetted into a QIAshredder spin column, which was then centrifuged for 2 minutes at maximum speed using a tabletop centrifuge. The flow-through was transferred to another tube and mixed with 1.5 volumes of AP3 buffer. 650 µl of the mixture were placed in a DNeasy Mini spin column, which was centrifuged for 1 minute at around 6000 g. This step was repeated until all the mixture was processed. The column was washed twice with 500 µl of AW2, centrifuging the first time at 6000 g and at maximum speed the second time. Finally, the elution was done applying 50 µl of sterile distilled water, leaving it for 5 minutes at room temperature before centrifuging the column at 6000 g for 1 minute.

The final tubes containing the DNA were kept in the freezer at -20 ºC until they were analysed.

3.3.6. Edit screening.

The screening of embryos co-bombarded with the editing machinery and the pActH2BGFP was performed as outlined in Chapter 2 with some modifications. Two RE-PCR cycles were applied for the screening of the samples. The band obtained after the second RE-PCR cycle was isolated and sequenced. The amplification and sequencing of the target genes was performed using nested primers. In case of *BdDWF4*, the primers used were BdDWF4 Fw and BdDWF4 Rv for the first amplification, BdDWF4 Nd1 Fw and BdDWF4 Nd1 Rv for the second amplification and BdDWF4 Nd2 Fw and BdDWF4 Nd2 Rv for the sequencing of the isolated band (See Supplementary Table 3). In case of the Hygromycin resistance cassette, the primers pBHygR Fw and pBHygR Rv were used for the first amplicon, HygR CDS Fw and HygR Nd Rv2 for the second amplicon and HygR Nd Fw and HygR Nd Rv3 for sequencing (See Supplementary Table 3). The protocol for each RE-PCR cycle was similar to the one described in Chapter 2, starting with 200 ng of genomic DNA, applying the enzyme recognising the restriction site targeted by the sgRNA

(BglI for *BdDWF4* and AdeI for HygR, both from Thermo Scientific, U.S.A.) for 30 minutes at 37 ^{o}C, applying afterwards the deactivation step with the required temperature for 10 minutes (65 ^{o}C for AdeI and 80 ^{o}C for BglI). The PCR after the restriction reaction was set up as described in Chapter 2, using the primers previously mentioned, together with the band isolation protocol and posterior RE-PCR cycles. As outlined in Chapter 2, to ensure enough DNA was obtained from the PCR reaction, 35 amplification cycles were applied.

The bands obtained after the second RE-PCR cycle were sequenced as described in Chapter 2 using the primers previously mentioned.

3.3.7. T-DNA line selection.

To study the effects of different levels of compactness in the chromatin on CRISPR/Cas9 function while minimizing the differences in efficiency due to different sgRNA sequences, different *Brachypodium* lines with the same T-DNA insertion were selected from the lines described in Bragg et al. (2012). For the analysis, the first three chromosomes of *Brachypodium* were analysed for their genetic density and T-DNA insertion frequency. The data for this analysis was collected from Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html). Using the Jbrowse tool, the transcript and T-DNA insertion feature densities were obtained and graphed. From this graph, different regions of interest in the chromosomes were selected for a more detailed analysis.

Regarding the T-DNA selection, it was decided to focus on the third chromosome of *Brachypodium*. Bragg et al. (2012) tabulated the different T-DNA insertion sites and the position of these insertions around described genes in the T-DNA lines that were obtained. Firstly, the table containing the collection of T-DNA was filtered for the chromosome of interest. A second filter regarding the location of the insert relative to a gene was applied, excluding all lines with insertions not classified as intergenic, meaning within 1000 bp of a known coding sequence. Additionally, lines involved in gene trapping (pJJ2LBP and pJJLBP2) or activation tagging (pJJ2LBA) were excluded to avoid unintended effects in the original gene function of *Brachypodium*. Finally, lines that only appeared once in the database, implying single insertion, were considered for the final selection depending on the feature studied.

3.4. Results.

3.4.1. SGV and sgRNA design and assembly.

To be able to compare different promoters and their ability to drive the expression necessary to induce edits in *Brachypodium*, three different plasmids were used. pBdSGV, based on the previously described pHvDGV but swapping the sgRNA cassettes for the one described in the pDON plasmid in van der Schuren et al. (2018), containing a BdU6 promoter for sgRNA expression, pOsSGV, again using the backbone of pHvDGV and the sgRNA cassette of the pENTR plasmid described in Hus et al. (2020), containing an OsU3 promoter, and the original pHvDGV containing both TaU6 promoters as described in Lawrenson et al. (2015). The tracrRNA part of the different cassettes was analysed as well by aligning the regions in each one of the plasmids of interest. The sequences of the tracrRNA (scaffold) contained in the three plasmids were aligned with the three sgRNA described in Jinek et al. (2013) (Figure 3.6A, Supplementary Figure 1). It was observed that the sequences contained in pBdSGV and pHvDGV (both sgRNA cassettes) had the initial part of the sgRNA described as v1.0 in this reference with additional loops, whereas the sequence contained in pOsSGV had the initial part of the v2.1 described in the reference with additional loops. The key first loop described in the three sgRNA proposed in Jinek et al. (2013), as described in Briner et al. (2014), were observed in the sgRNA contained in the three plasmids used in this chapter (Figure 3.6B).

Both SGV were obtained by amplification of the sgRNA cassettes and pHvDGV backbone with Golden Gate compatible overhangs (Figure 3.4 and 3.7A). These plasmids were then ready for the assembly of the sgRNA containing the targeting elements. Different sequences were analysed in the search of potential target sites for CRISPR/Cas9. sgRNA were designed for *BdCKX2*, *BdGA20ox2*, *BdRVR1*, *BdVRN2*, *BdDWF4*, and HygR (both pBract204 and the Hygromycin resistance cassette described in the T-DNA lines collected in Bragg et al. (2012)), targeting restriction sites within a designed amplicon (Supplementary Tables 4-10). Out of all the potential target genes, it was decided to focus on *BdDWF4* and HygR, and a few target sequences were selected, searching for a balance between on-target efficiency, off-target effect and cost of the restriction enzyme required for the RE-PCR-based screening proposed (Table 3.1). It should be noted that the HygR cassettes described in pJJ2LB and pBract204 had differences (Supplementary Figure 2). Therefore, it was decided to choose a spacer whose sequence would be able to target both versions (HygR1) to use as a control in pBract204 and another one that targeted exclusively pJJ2LB (HygR2). The sgRNA were assembled in each one of the plasmids using Golden Gate assembly. In case of pHvDGV, due to the swapping of marker genes, colony PCR methods were available for the screening of insertions (Figure 3.7B), as well as colour selection in the transformed *E. coli* colonies (Figure 3.7C). However, the SGV do not have this type of selection available, requiring the sequencing of the colonies to confirm the insertion of the desired target sequences in the sgRNA cassette. Using these type of screenings, all different plasmids of interest were obtained and amplified using *E. coli*, preparing them for the biolistic transformation of embryos.

Table 3.1. Target sequences for CRISPR/Cas9 based genome editing in *Brachypodium* embryos.

Target Gene	*ID*	*Sequence (3'→5') (Restriction site)*	*Restriction enzyme*	*On-target score*	*Off-target score**
BdDWF4	BdDWF4-1	GGTTATGGAC**GCCATCATGG(C)**	BglI	71.56	46.22
	BdDWF4-4	GGACACCACCCCCT**TCATGA**	PagI	65.26	48.23
HygR	HygR1	CAACGTGA**CACCCTGTG**CA	AdeI	62.45	99.26
	HygR4**	GCAGGATCGCCAC**GACTC**C	SchI	49.83	97.62

*Off target scores calculated using the *Brachypodium distachyon* (v1.0) included in Benchling
**The target sequence refers to the Hygromycin resistance cassette described in Bragg et al. (2012)

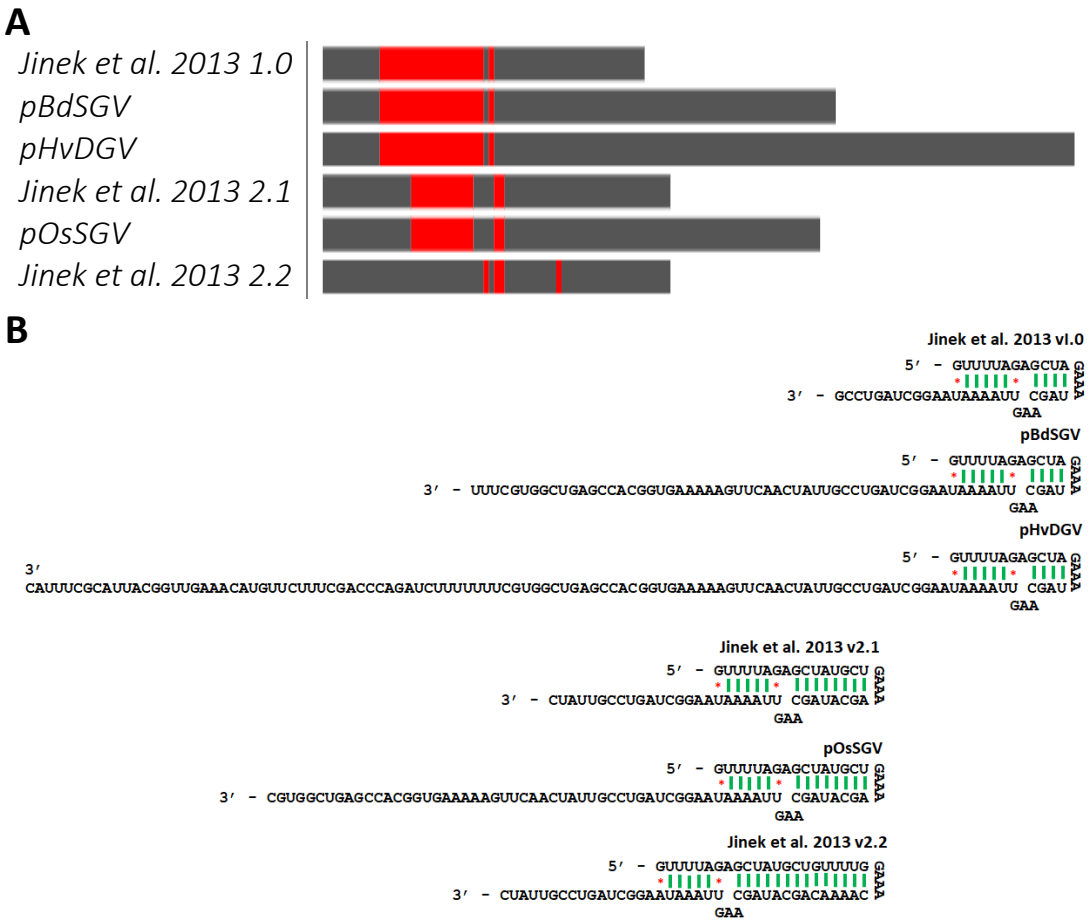


Figure 3.6. The analysis of the sgRNA scaffold in the plasmids used in this chapter shows similarities with described functional sgRNA.
(A) Diagram of the alignment of the three tracrRNA found in the plasmids used in this chapter with the sgRNA described in Jinek et al. (2013). The alignment is represented with the sequences in common in grey and the gaps in the alignment in red. The image was obtained from Benchling. (B) Diagram showing the first loop in all sgRNA studied and the sgRNA described in Jinek et al. (2013). All sequences are shown from 5' to 3'. Nucleotides shown in vertical represent the tetraloop, lines in green show nucleotide homology between the crRNA and tracrRNA in the lower and upper stem of the sgRNA, and asterisks in red represent mismatches. Shown with an offset below the sequence are the nucleotides in the bulge of the stem loop.

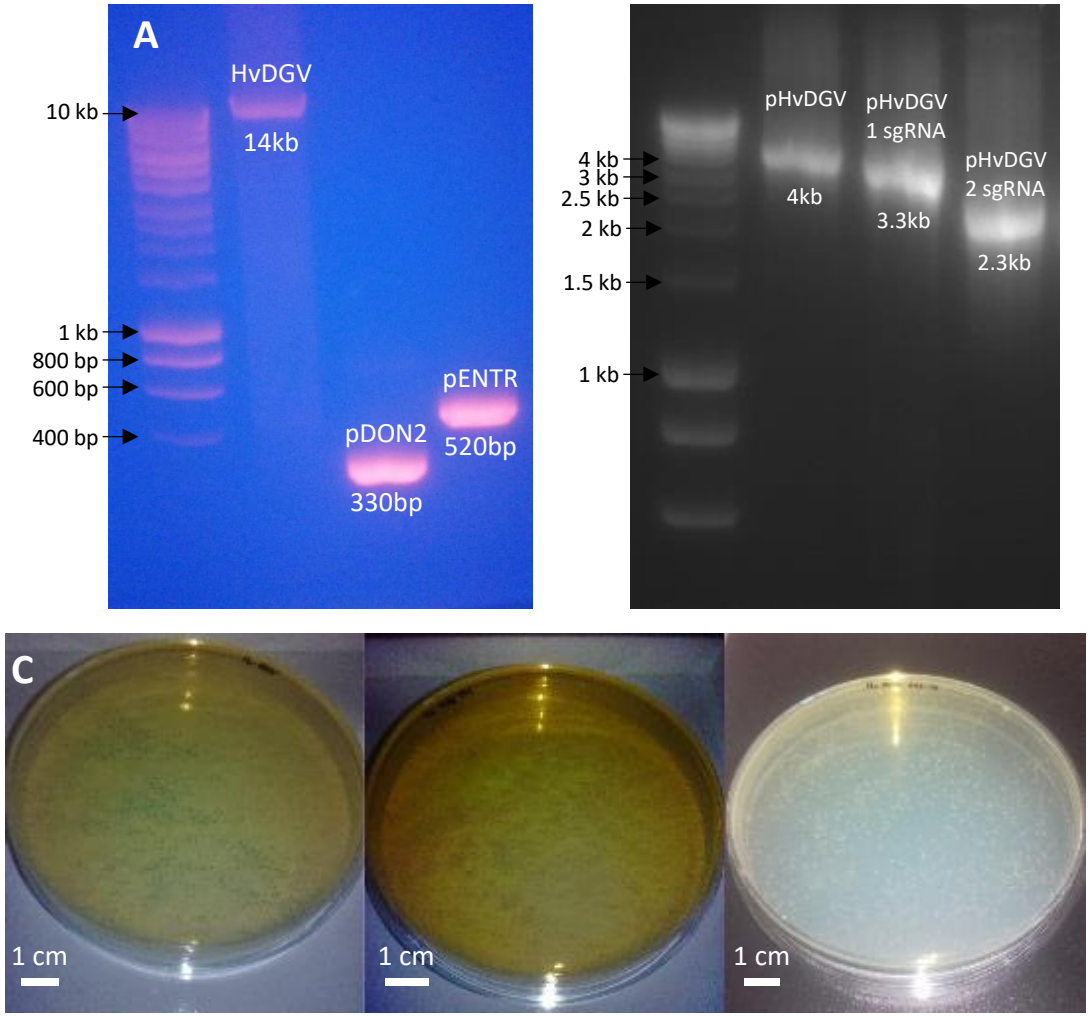


Figure 3.7. Successful assembly of the plasmids used in this chapter, together with the insertion of the desired sgRNA.
(A) SGV assembly, from left to right, 1kb ladder, HvDGV backbone, pDON sgRNA cassette, pENTR sgRNA cassette. (B) Example of the amplicons observed during the HvDGV sgRNA insertion. From left to right, 1kb ladder, amplicon of the sgRNA cassettes including LacZ and RFP, amplicon of the sgRNA cassettes after the swapping of LacZ for the first sgRNA, amplicon of the sgRNA cassettes after the swapping of both LacZ and RFP for both sgRNA. (C) Plates showing the colour selection of the HvDGV. From left to right, plasmid with no insertions, plasmid with an insertion in the first sgRNA cassette, and plasmid with both insertions.

3.4.2. Edit Screening.

A pilot experiment was carried out to look for evidence of editing activity using the constructs described in the previous section and two different *Brachypodium* ecotypes, Bd21-3 and Bd21-3 GUS. Targeted GFP fluorescence was observed in *Brachypodium* embryos which were co-bombarded with the plasmid DNA encoding the editing machinery and the plasmid pActH2BGFP for the visual selection of the transformed embryos, implying that GFP positive embryos would have both plasmids introduced in them (Figure 3.8A). The target genes were *BdDWF4* and the Hygromycin resistance cassette found in pBract204 (Table 3.1). The plasmids containing sgRNA targeting *BdDWF4* were co-bombarded into Bd21-3 plants, whereas the plasmids containing sgRNA targeting HygR were co-bombarded into transgenic Bd21-3 containing the T-DNA described in the pBract204 plasmid. The target genes were amplified from the genomic DNA

isolated from the transformed embryo samples (Figure 3.8B). The amplified products were then screened using RE-PCR. Gel clean-up protocols were used for the isolation of the relevant bands during the RE-PCR process to eliminate non-specific products seen in the control PCRs. After the first RE-PCR cycle, all the samples targeting *BdDWF4* in Bd21-3 were able to generate amplicons of the intended size, implying that these samples could have been successfully edited. For the samples where HygR was targeted in the Bd21-3 GUS line, , amplicons were produced when the guides were expressed using pBdSGV and pOsSGV, but not when using pHvDGV., One of the samples of pOsSGV targeting *BdDWF4* in Bd21-3 embryos was lost during band isolation after the second RE-PCR cycle. The remaining bands were then isolated and sequenced after the second RE-PCR cycle.

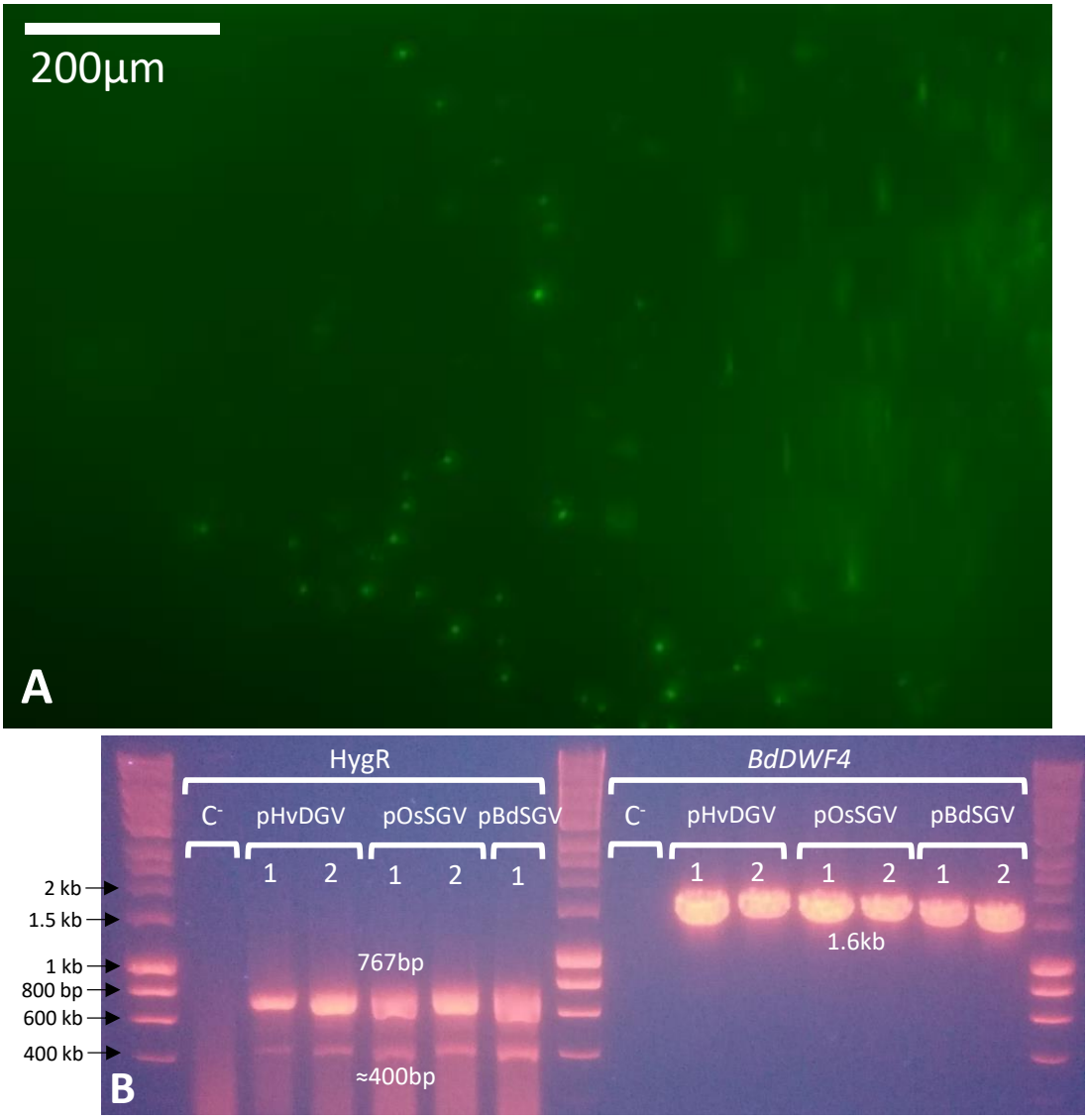


Figure 3.8. The editing machinery was successfully introduced in *Brachypodium* embryos prior to the screening of editing in the selected targets.
(A) Image of a Bd21-3 embryo co-bombarded with pActH2BGFP and pHvDGV containing sgRNA targeting HygR. (B) Control bands of the co-bombarded samples. The lanes 2-7 were amplified using HygR specific primers, whereas lanes 9-15 were amplified using *BdDWF4* specific primers. The lanes labelled C^- represent the negative control without any genomic DNA added to the PCR reaction. For all DNA length markers (runs 1, 8 and 16) a 1kb ladder was used.

Chromatograms were visually compared to controls to look for evidence of multiple sequences obtained after RE-PCR, which could be linked to *in vivo* generated edits (Figure 3.9). Peaks containing mixed signals were observed in all the recovered samples. In some cases where peaks were broader rather than discrete, or disproportionately high, these may reflect poor DNA quality, for example as a result of gel clean-up steps. However, for most samples there appear to be the expected specific ambiguities around the restriction enzyme used for the relevant RE-PCR (GCCNNNNNGGC for *BdDWF4* targets and CACNNNGTG for HygR), with subsequent sequence maintained in register, i.e. without evidence for indels in the recovered sequence. Strikingly, this is not the case for the two samples transformed with pBdSGV targeting *BdDWF4*, where sequences appear to be highly heterogeneous (Figure 3.9A). This pattern is not consistent with short indels caused by NHEJ DSB repair within the intended target site but could be due to larger edits caused by other mechanisms of DSB repair.

In case of the sample obtained after transformation with pOsSGV targeting the *BdDWF4* gene, a modification is observed in the target sequence, appearing further downstream of the PAM than expected in short indels linked to NHEJ DSB repair, which are often, but not always, found between 0 and 5 nucleotides downstream from the PAM site. In case of the samples obtained from pHvDGV, the consensus sequence shared most of the sequence with the control, including the original target sequence, implying that no modification has been done in the target site. However, both replicates show different peaks in similar positions as pOsSGV, in addition to differences outside of the target site. When comparing how diverse the mixture of amplicons sequenced is in each sample, it is clear that amplicons obtained after bombardment with pBdSGV had the most diverse, followed by pOsSGV and lastly pHvDGV, implying a possible difference in editing efficiency related to the use of different sgRNA cassettes. As previously pointed, the results obtained in *BdDWF4* editing show that modifications when comparing the control with the samples transformed with the editing machinery are obtained, but these cannot be directly linked to genome editing.

When observing the samples transformed with the editing machinery targeting the Hygromycin resistance gene (Figure 3.8B), the consensus sequences obtained from both the sample obtained after transforming with pBdSGV and pOsSGV showed no modifications in the target site. However, as observed in *BdDWF4* editing, background noise is observed in the samples, particularly around the target site, which could be related to targeted modifications in the sequence occurring, although these would occur at a lower level than the case of *BdDWF4* targeting. In this case, the modifications observed were more severe in the sample transformed with pOsSGV than the one transformed with pBdSGV, pointing to a higher editing efficiency in the first one than the latter.

Overall, the results described regarding the mixture of different amplicons obtained after the bombardment of the editing machinery point to the ability of the constructs pBdSGV, pOsSGV and pHvDGV to induce targeted alterations in the target gene *BdDWF4*. However, in case of HygR targeting, all the amplicons obtained from pHvDGV were restricted after the first RE-PCR cycle and the heterogeneity in the amplicons obtained after two RE-PCR cycles did not match the one observed in *BdDWF4*. These results could be due to targeted modifications induced by the introduced editing machinery, albeit the effect of these constructs cannot be categorically

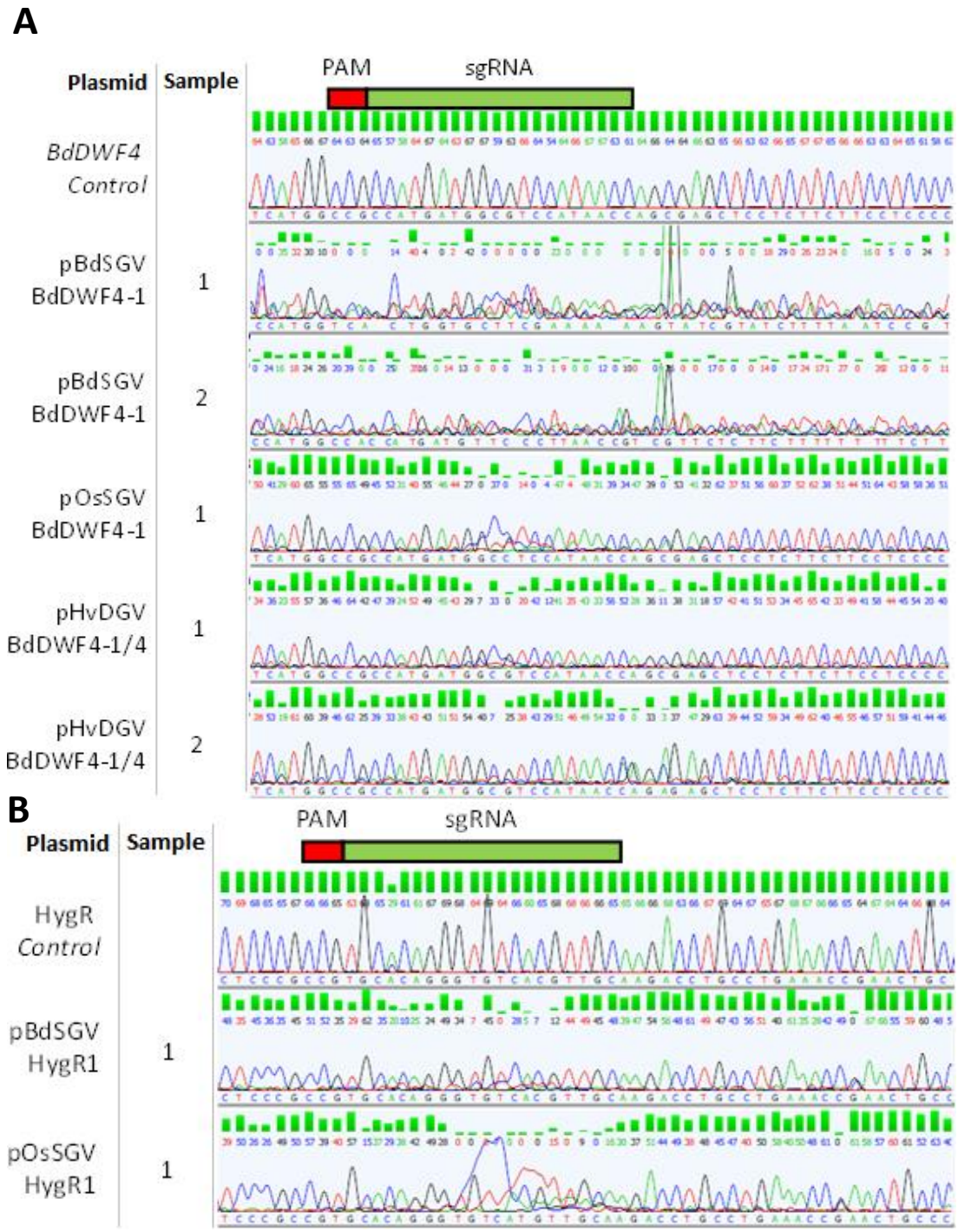


Figure 3.9. Chromatogram analysis of the bands obtained after RE-PCR shows differences between the control band and the ones obtained from co-bombarded samples.

(A) Comparison of the chromatograms obtained from bands isolated after amplification of *BdDWF4* from Bd21-3 and bands obtained from the co-bombarded samples after two cycles of RE-PCR. Under the column named Plasmid, the different plasmids used for each one of the samples are specified. Under the column named Sample, the different biological replicate number is shown (1 for first replicate, 2 for second replicate). In a diagram above the chromatograms are highlighted the PAM site and target sequence from the sgRNA studied. (B) Comparison of the chromatograms obtained from bands isolated after amplification of HygR from Bd21-3 GUS and bands obtained from the co-bombarded samples after two cycles of RE-PCR. Under the column named Plasmid, the different plasmids used for each one of the samples are specified. Under the column named Sample, the different biological replicate number is shown

confirmed with the data obtained. Further interpretations and improvements on this analysis will be discussed later in this chapter.

This provides promising material for future experiments, firstly to validate the results by including a wider range of controls and by cloning products to identify specific potential editing events, and then subsequently to compare the impact of empirical conditions and effects of altering constructs and host genotypes. Unfortunately, while such experiments were intended, problems with lab access cut this work short. These will be outlined in the discussion of this chapter.

3.4.3. T-DNA line selection for the assessment of genomic environment on CRISPR/Cas9 efficiency in plants.

The effects of the genomic environment on editing efficiencies has been reported in other organisms, but not in plants. A platform was designed for the study of these effects in genome editing efficiencies in plants using a single sgRNA targeting different positions in the genome using the Hygromycin resistance cassette included in the T-DNA insertion lines described in Bragg et al. (2012).

Firstly, the gene density and T-DNA insertion frequency was analysed in the first three chromosomes of *Brachypodium* (Supplementary Figures 3, 4, Figure 3.10). It was observed that in the three chromosomes studied, a reduction in gene density and T-DNA insertion frequency correlated with the centromeric locations described in Vogel et al. (2010). From this analysis, several positions of interest were observed along the genome, with different distances from the centromere, implying different compactness of the genome, and different number of transcripts. For selecting the T-DNA lines of interest, as outlined in the Materials and Methods, the library was filtered to obtain only lines related to insertional mutagenesis that had single inserts in intergenic regions, implying that the genes inserted are unique in the genome and that the insertion does not disrupt normal *Brachypodium* gene function.

In case of the location in the genome, four locations were considered, centromeric, mid arm, telomeric and true telomeric, being from closest to furthest from the centromere. The lines selected were collected in Table 3.2.

Another location in the third chromosome was selected, given the differences in transcript frequency observed between the 9^{th} and 10^{th} Mb in the third chromosome of *Brachypodium*, observing 105 and 29 transcripts described in each area respectively. The T-DNA lines selected were collected in Table 3.3.

The selection of T-DNA could be the basis for testing the ability of CRISPR/Cas9 to induce edits in different positions in the genome.

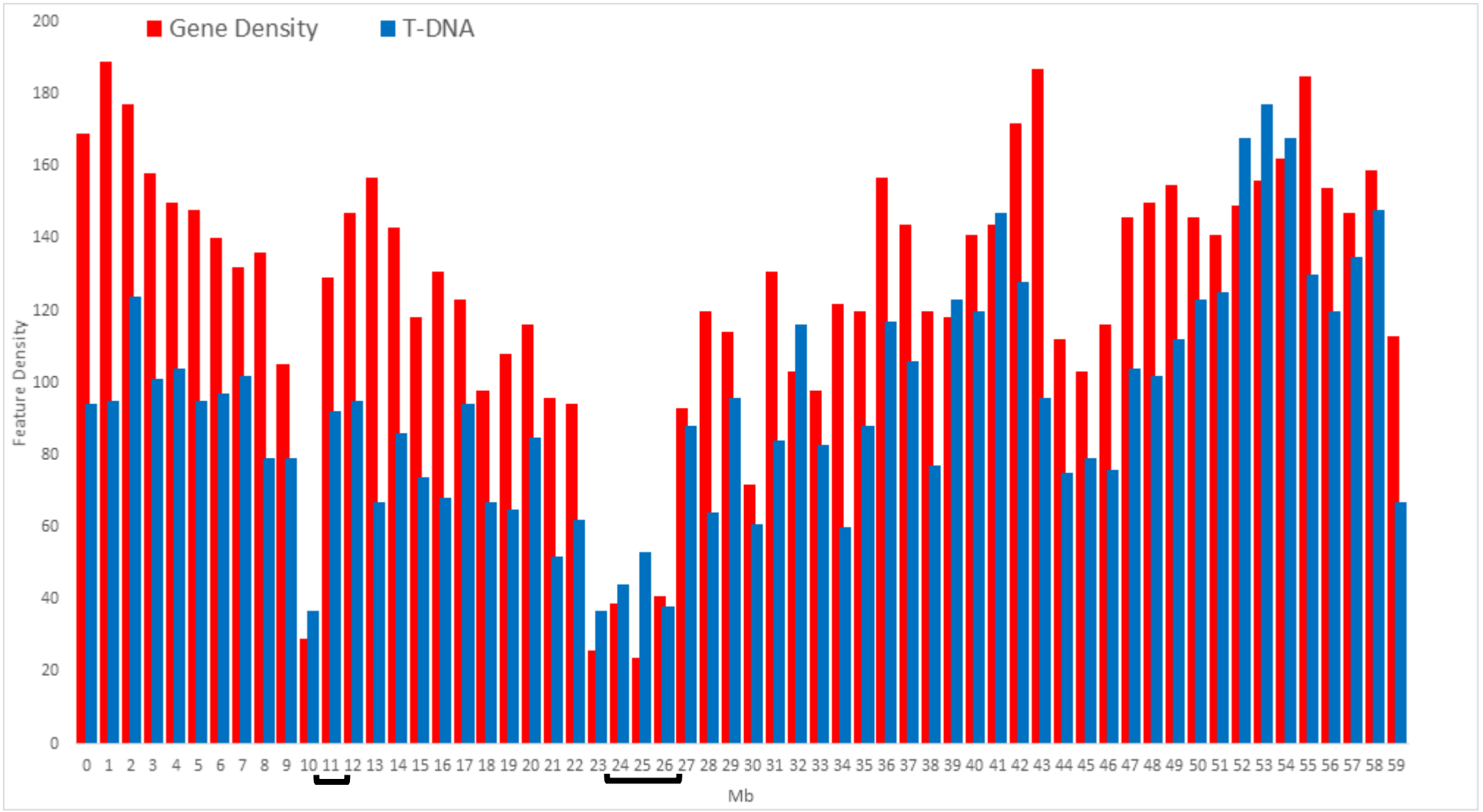


Figure 3.10. Gene density and T-DNA insertion frequencies vary along the third chromosome in *Brachypodium*.
Histogram showing the feature frequency of transcripts (red) and T-DNA insertions (blue) in the third Chromosome of *Brachypodium distachyon*. Data showed as feature frequency in each one of the megabases of the chromosome. Black brackets represent the centromeric repeats identified in Vogel et al. (2010).

Table 3.2. *Brachypodium* T-DNA Lines selected for testing the effects of chromosomic location in editing efficiencies.

Centromere					
Chrom.	Location (bp)	Line	Class	Strand	Insert
Bd3	23084724	JJ7899	intergenic	-	pJJ2LBA
Bd3	23538835	JJ15708	intergenic	.	pJJ2LBA
Bd3	24429423	JJ25299	intergenic	+	pJJ2LBA
Bd3	24462496	JJ18996	intergenic	.	pJJ2LBA
Arm					
Chrom.	Location (bp)	Line	Class	Strand	Insert
Bd3	12041735	JJ20152	intergenic	.	pJJ2LBA
Bd3	12053999	JJ26834	intergenic	-	pJJ2LBA
Bd3	12088525	JJ23136	intergenic	.	pJJ2LBA
Bd3	12288962	JJ9102	intergenic	.	PJJ2LBA
Bd3	12367263	JJ16683	intergenic	.	pJJ2LBA
Bd3	12375275	JJ20796	intergenic	.	pJJ2LBA
Bd3	12687586	JJ26824	intergenic	+	pJJ2LBA
Bd3	12746670	JJ27956	intergenic	-	pJJ2LBA
Telomere					
Chrom.	Location (bp)	Line	Class	Strand	Insert
Bd3	57119611	JJ16462	intergenic	.	pJJ2LBA
Bd3	57265134	JJ19	intergenic	+	pOL001
Bd3	57539421	JJ6646	intergenic	.	PJJ2LBP2
Bd3	57542228	JJ21603	intergenic	-	pJJ2LBA
Bd3	57648870	JJ16129	intergenic	.	pJJ2LBA
Bd3	57682412	JJ26890	intergenic	-	pJJ2LBA
Bd3	57844468	JJ28425	intergenic	+	pJJ2LBA
True Telomere					
Chrom.	Location (bp)	Line	Class	Strand	Insert
Bd3	184708	JJ9320	intergenic	-	pJJ2LBA
Bd3	400185	JJ28557	intergenic	-	pJJ2LBA
Bd3	571013	JJ4728	intergenic	.	PJJ2LB
Bd3	1677355	JJ27009	intergenic	+	pJJ2LBA
Bd3	1918007	JJ20698	intergenic	.	pJJ2LBA
Bd3	2199541	JJ26990	intergenic	-	pJJ2LBA

Table 3.3. *Brachypodium* T-DNA Lines selected for testing the effects of gene density in editing efficiencies.

High gene density					
Chromosome	Location (bp)	Line	Class	Strand	Insert
Bd3	9083874	JJ16728	intergenic	.	pJJ2LBA
Bd3	9282732	JJ8322	intergenic	.	pJJ2LBA
Bd3	9324983	JJ24944	intergenic	-	pJJ2LBA
Bd3	9457047	CRC754	intergenic	+	pJJ2LBA
Low gene density					
Chrom.	Location (bp)	Line	Class	Strand	Insert
Bd3	10487220	JJ25603	intergenic	+	pJJ2LBA
Bd3	10493826	JJ114	intergenic	.	POL001
Bd3	10874928	JJ22482	intergenic	.	pJJ2LBA
Bd3	10094359	JJ21969	intergenic	.	pJJ2LBA

3.5. Discussion.

3.5.1. Plasmid assembly and editing system analysis.

The first objective of this chapter was to obtain a series of plasmids for the validation of different promoters for the expression of sgRNA in *Brachypodium*. These plasmids were based on the cloning of sgRNA cassettes present in the pDON2 and pENTR plasmids in the pHvDGV, given that the correct expression of the Cas9 by the ZmUbi1p has been described in other references (See Supplementary Table 1) and was validated during the experiments regarding H2BGFP expression using the pDGV-H2BGFP (See Chapter 2). The cloning of the sgRNA cassettes was performed by amplifying with Golden Gate compatible overhangs and ligation of the fragments, as described in the Materials and Methods section of this chapter. This type of insertion poses some advantages when comparing it with other techniques such as conventional digestion ligation, which would leave the restriction sites due to the use of type II restriction enzymes, as Gateway cloning (Reece-Hoyes and Walhout, 2018), which requires specific *att* sequences to allow the swapping of the toxic *ccdB* gene for the desired sequence, or Gibson assembly (Gibson et al., 2009), which would require amplification steps with big overhangs (around 20 nucleotides) for correct homology-based insertion, which can lead to complications in the PCR steps. Instead, Golden Gate (Engler et al., 2008) allows a scarless insertion in one reaction without requiring large overhangs to guide it.

The original references describing the pDON2 and pENTR plasmids use different cloning mechanisms for the introduction of the sgRNA targeting sequences into the sgRNA cassette. In case of the pDON2, as described in van der Schuren et al. (2018) the mechanism of introduction is based on Digestion-Ligation steps, including the annealing of the primers using Touchdown techniques, gel isolation of the pDON2 plasmid after the correct restriction, and ligation. Afterwards, using Digestion-ligation, the sgRNA cassette is cloned into the p5Cas, requiring further gel isolation steps. As described in the Materials and Methods section of this chapter, the pDON2 plasmid contains two different sgRNA cassettes. One of them, containing a BdU6 promoter, requires restriction with the type IIS enzyme BsaI. This enzyme has an optimal restriction temperature of 37 ^{o}C, allowing Golden Gate assembly. However, the second sgRNA cassette, with a OsU6 promoter, requires the use of another type IIS restriction enzyme, BtgZI, with an optimal restriction temperature of 60 ^{o}C. This temperature overlaps with the deactivation temperature of the T4 ligase required for Golden Gate assembly, meaning this type of assembly was not possible. The insertion of the sgRNA targeting sequence into the pENTR is based on Golden Gate assembly in the pENTR plasmid and transfer of the sgRNA cassette into the pOsCas9 plasmid using Gateway Cloning. As described in the Materials and Methods of this section, the insertion of the sgRNA targeting sequence in the pENTR sgRNA cassette requires the use of BsaI. In case of pHvDGV, the insertion of sgRNA targeting sequences is based on Golden Gate assembly directly in the plasmid containing the Cas9. This direct assembly into the final plasmid implies an easier cloning process. In addition to that, and in contrast with the other systems, this Golden Gate ligation involves the swapping of selective markers for the sgRNA targeting sequence, facilitating the screening of positive colonies after ligation with visual methods or colony PCR. This easy assembly process was found as well in the SGV described in this chapter.

However, improvements could be made in the SGV systems. Given the strategy used for the introduction of the BdU6::sgRNA and OsU3::sgRNA cassettes as entire cassettes, the SGV plasmids do not include the visual or colony PCR screenings, which could be included to facilitate

the screening of correct sgRNA targeting sequences insertions as with pHvDGV. Additionally, several sgRNA cassettes could be introduced in the plasmid to allow multiplex editing.

As previously described, the sgRNA cassettes were cloned in their entirety into the pHvDGV backbone, which implies that the sgRNA scaffold contained in the original plasmids was also included. As Figure 3.7 shows, there are differences in the sequences. However, when comparing the sequences with the ones described in Jinek et al., (2013), it is observed that the elements essential for Cas9 function are included in all scaffolds, in addition to extra loops. It has been described that variations in sgRNA length do not block Cas9 function (Jinek et al., 2013) but may affect efficiency (Hsu et al., 2013; Mali et al., 2013; Nishimasu et al., 2014; Wright et al., 2015; Mekler et al., 2016), therefore, differences in the sgRNA scaffold sequence may imply differences in editing efficiency. However, the proper folding of the sgRNA could be analysed to ensure these additional loops are functional using tools such as RNAfold (Lorenz et al., 2011). However, given the possible differences in efficiency could be related to the sgRNA scaffold rather than the promoter used, it would be advisable to use a common sgRNA scaffold for the validation of the different promoters used.

The results in this chapter also show the ability of the protocol proposed in Chapter 2 of introducing plasmid DNA through biolistic methods into *Brachypodium* embryos. This was observed on the basis of targeted GFP expression when comparing the bombarded and non-bombarded embryos. The results obtained point to the ability of co-bombarding using the desired plasmids, including the editing machinery used in this chapter. However, this should be confirmed by isolating the DNA of the bombarded embryos and identifying the introduced editing machinery through PCR methods.

3.5.2. *In vivo* editing experiments.

The results obtained from initial pilot experiment suggest that the protocol described produced low levels of editing in the target sequence after transformation with some of the constructs. This resulted in the modification of the targeted enzyme recognition site, which allowed the recovery of PCR products following RE-PCR. For the specific combination of pBdSGV carrying the sgRNA targeting *BdDWF4* in the Bd21-3 ecotype, highly heterogeneous products were recovered, suggesting a more aggressive effect of the transformation resulting in multiple large indel events. These results would concur with the ones obtained in van der Schuren et al., (2018), the publication which described the ability of the BdU3 promoter of expressing sgRNA at a level sufficient for genome editing in *Brachypodium*. However, there are a number of caveats to the initial results described in this chapter which would need to be addressed in subsequent experiments (intended but unfortunately unable to be carried out as laboratory access was limited by COVID-19 measures).

The results obtained in this chapter point to the ability of the three different constructs (pBdSGV, pOsSGV and pHvDGV) to express the editing machinery and produce edits, albeit these would need confirmation with further replicates and the controls omitted. The ability of the BdU6 and OsU3 promoters to express functional in Brachypodium, contained in the pBdSGV and pOsSGV respectively, has been described in van der Schuren et al., (2018) and Hus et al., (2020). The ability to express enough sgRNA for editing in *Brachypodium* of the TaU6 promoter described in Lawrenson et al., (2015), which showed the lowest variation in the amplicons produced after two RE-PCR cycles, has not been published to the date of writing. However, the correct expression of the sgRNA could be extrapolated from the modifications obtained using all 3 plasmids if editing was confirmed. Further confirmation of the correct expression of sgRNA could

be obtained using quantitative Real Time PCR (qRT-PCR) techniques, as described in references such as Ma et al., (2015) or Ren et al., (2016), among others. This would confirm that the three plasmids are able to produce the machinery necessary for editing in *Brachypodium*. In case of the Cas9 expression, the use of the ZmUbi1 promoter has been widely reported as a constitutive promoter for expression in *Brachypodium* (Callis et al., 1990; Christensen et al., 1992; Christensen and Quail, 1996) and the results obtained in Chapter 2 regarding the targeted GFP expression in the nucleus after bombardment with the pDGV-H2BGFP should imply that the Cas9 is also being expressed. However, another experiment should focus on the investigation of the expression levels of the Cas9 and sgRNA cassettes to establish a relationship between the possible editing efficiency and the strength of the promoters.

The results obtained after RE-PCR cycles observed in the chromatograms shown in Figure 3.9 show modifications in the samples transformed with the editing machinery in the target genes. These modifications manifested as an heterogeneous sequence obtained form the sequencing of a pool of amplicons obtained from the same PCR product, as shown in Figure 3.8B. During the RE-PCR assays it was observed that the HygR amplification produced both the desired band and an unspecific amplicon. The experimental procedure proposed in this chapter uses gel clean-up for the isolation of the band of interest and discarding the unspecific amplicon observed. This was a crucial step to ensure the amplicon analysed in further RE-PCR cycles was the correct one. However, the gel-clean up process implies several steps that might affect the quality of the final DNA obtained. Firstly, gel clean-up protocols imply the running of an agarose gel and the identification of the bands using the UV light. This type of radiation can damage the DNA in the gel, meaning that the quality of the DNA obtained might be reduced. On the other hand, the process implies the cutting of the band from the gel, which if not done carefully can lead to contamination from other DNA bands. Finally, the gel clean-up process also implies the use of chemicals to be able to remove the agarose and intercalating agent of choice for the visualization of the DNA under the UV light. These chemicals can also affect the DNA and the purity of the final sample obtained, having an effect on the final quality of the DNA to be sequenced. In future experiments, it is of crucial importance to avoid these unspecific amplicons altogether, either by designing new primers that do not generate an unspecific amplicon or optimising the melting temperature of the primers to ensure only the bands of interest are produced. A negative control using the primers targeting the HygR cassette and the genomic DNA obtained from non-transgenic Bd21-3 plants would have been useful in ensuring no other amplicons were obtained from this initial amplification, together with further optimisation or redesign of the primers to avoid this amplicon. In this experiment, the gel clean-up was necessary due to the non-specific bands obtained, but in future experiments this should be avoided if possible.

As described in Chapter 2, the use of RE-PCR has its advantages and limitations. In this case, it has been possible to observe modifications in the targeted genes, but an estimation of the editing efficiency would not have been possible due to the restriction of the non-edited DNA, meaning that even with an estimation using platforms such as Synthego ICE (Synthego Performance Analysis, ICE Analysis. 2019. v2.0. Synthego), which provides a percentage of each one of the specific amplicons contained in a pool of amplicons, this estimation would be skewed towards the edited amplicons. The chromatograms obtained from the samples were ran through this platform. However, due to the differences between the control band and the samples screened, the analysis failed. This could be due to a pool of amplicons too complex to analyse or due to low DNA quality sequenced. On the other hand, the RE-PCR protocol used implied the sequencing of all the edited amplicons, obtaining all the different edits contained in the sample

in a single DNA sample. Future experiments should also focus on the insertion of the specific amplicons in plasmids via cloning and the sequencing of the specific modified amplicons. This would have made possible to prove the mixture of amplicons obtained contained edited sequences rather than background noise from contamination and the sequencing of several amplicons could have provided information on the types of modifications induced. Future experiments should focus on, in addition to the optimisation of the protocol itself, finding the specific edits obtained. However, as previously described, RE-PCR protocols would remove the non-edited amplicons, meaning that even after obtaining the specific edits, the estimation of the editing efficiency would not be accurate.

The results obtained in the chromatogram analysis point to several amplicons being sequenced at the same time. Different amplicons could be due to CRISPR/Cas9 activity in the co-bombarded embryos. Although typical CRISPR/Cas9 induced edits result in indels around the cut site, meaning between 0 and 5 base pairs downstream of the PAM site, larger modifications have also been observed (Chen et al., 2019; Sansbury et al., 2019; Schmidt et al., 2019). These type of modifications could explain the results obtained from the transformations with pBdSGV targeting *BdDWF4*. On the other hand, indels in the target site would manifest in more severe modifications along the sequence downstream of the PAM site, as different indels in different amplicons would lead to distortions in the chromatogram obtained. These type of modifications would explain the results observed in the samples transformed with pOsSGV targeting *BdDWF4* and HygR. Finally, the results observed in the samples transformed with pHvDGV targeting *BdDWF4* and the sample transformed with pBdSGV targeting HygR, the strongest peaks found in the chromatogram corresponded to the sequence observed in the control, but minor peaks could be observed in the background. These could be either background noise, implying that no modification was obtained, or that the modified amplicons were still underrepresented in the pool, despite the RE-PCR procedure. On the other hand, as previously stated, the modifications were observed either in the entire amplicon (pBdSGV targeting *BdDWF4*) or outside the expected cut site (all other samples), which could be related to poor quality of the DNA sequenced rather than targeted modifications induced by the editing machinery. To ensure the different amplicons obtained are due to the effects of the editing machinery and not due to the screening process, different negative controls would have been necessary, including bombardment with blank plasmids without a targeting sgRNA sequence, ensuring the bombardment and RE-PCR process is not causing the modifications, together with a negative control using the whole RE-PCR process on non-transformed embryos, ensuring the RE-PCR process proposed was not causing damage to the amplified DNA causing the alterations observed in the chromatograms. On the other hand, in order to assess the modification patterns induced by the editing machinery introduced, it is crucial to obtain the specific amplicons obtained using the cloning methods previously mentioned.

When comparing the different constructs assembled and their capacity of producing different amplicons of the target sequence, pBdSGV was able to induce the most severe changes observed when targeting *BdDWF4*, but the chromatogram showed smaller changes when targeting HygR. On the other hand, pOsSGV was able to induce severe changes in HygR and background-type changes in *BdDWF4*. pHvDGV was able to induce background-type changes in *BdDWF4* but was not able to produce any observable modifications in HygR . The results point to a difference in editing capacity depending on the target, meaning these may be sub-optimal. However, the on-target scores provided in Table 3.1 suggest that the predicted efficiency on *BdDWF4* was higher than the one for HygR, which correlates with the results observed. When focusing on each target, generally pBdSGV and pOsSGV were consistently able to induce edits, whereas pHvDGV

was not able to do so in all the targets studied. This points to the TaU6 promoter not expressing the sgRNA as efficiently as the other plasmids. It should be noted that both the BdU6 and OsU3 promoters have been used for genome editing in *Brachypodium* (Raissig et al., 2016; van der Schuren et al., 2018; Hus et al., 2020) but TaU6 has not. Given that negative results are rarely published, this may have also been found in the bibliography but not reported.

The results obtained are promising towards a *in vivo* sgRNA screening platform in Brachypodium. Despite the promising results, the lack of controls makes impossible to extract concrete and confirmed results from the experiments described in this chapter. Future experiments should focus on the addition of biological replicates and the inclusion of controls in different steps of the protocol.

The first set of controls should focus on the transformation process, including a series of bombardments with the gold particles without the plasmid DNA coating to confirm that the gold particles did not produce any type of fluorescence in the embryos. On the other hand, bombardments with the plasmids of interest (pBdSGV, pOsSGV and pHvDGV) without any targeting sequences would be required to confirm that the modifications are indeed induced by the editing machinery introduced and not by stress induced by the bombardment and tissue culture process. These samples should be treated as the samples processed in the pilot experiments, including the RE-PCR process, in order to assess the possible effects of these variables on the results obtained.

Another improvement to the protocol future experiments should focus on would be the elimination of gel clean-up steps. These are a useful tool in order to eliminate possible unspecific amplicons, however these usually imply a reduction in the quality of the amplicon isolated, which might interfere with the sequencing of the amplicons as a pool. A simple solution would be the re design of the primers to eliminate the unspecific amplicon or to run new gradient PCR to obtain a melting temperature with which only the desired amplicons are obtained.

On the other hand, the final aim of the experiments should be a series of specific sequenced amplicons containing edits in the target sequence. In order to obtain these, the pool of amplicons obtained after the RE-PCR process should be cloned in plasmids, which would then be sequenced after an amplification using *E. coli.*

5.3.3. Selection of T-DNA insertion lines to test the effects of the genomic environment on editing efficiency.

The final objective of this chapter was to design a platform to study the possible effects of the chromosomal environment on the ability of CRISPR/Cas9 to induce edits. The features studied would be the distance from centromeric sequences and gene density, which were firstly analyzed. The centromeric positions in the *Brachypodium* chromosomes have already been described (Vogel et al., 2010), basing the analysis on the data already published. On the other hand, gene density was analysed, obtaining histograms representing the number of transcripts found in each megabase of each chromosome. A depression in gene density was observed in the regions where the centromeres have been previously located, which correlates with the findings described in Li et al., (2018). In these same histograms, the number of T-DNA insertions found in the library described in Bragg et al., (2012) in each megabase of the chromosome is also represented. The T-DNA insertion frequencies follow a trend similar to gene density. However, as suggested in Kim et al., (2007), there should be no correlation in T-DNA insertion frequencies and the proximity to centromeric sequences. Therefore, this trend would be related to the selective techniques utilised in the obtention of the T-DNA library itself. Once the analysis was

finished, it was decided to select regions of interest to study each one of the features. In case of gene density, it was decided to select the 9th and 10th megabase of the third chromosome, obtaining a large difference in transcripts observed in a short space in the genome. In this chromosome in particular, Vogel et al., (2010) observed centromeric repeats in around the 11th megabase, but no secondary constriction has been described in this region (Lusinska et al., 2018), therefore it would be possible to compare the editing efficiencies obtained disregarding the possible effects of major chromosomic constrictions.

On the other hand, for studying the alterations caused by different compaction levels in different parts of the chromosome, different positions in the third chromosome of *Brachypodium* were selected. In this case, it was not possible to fix the number of transcripts found in each megabase. Depending on the results obtained regarding gene density, the results would have to be put in the context of the gene density observed in the region of interest.

The experiment proposed is based on the alterations of editing efficiency dependant on the genomic environment using a single sgRNA. As the results in this chapter suggest, editing efficiency is highly dependent on the target sequence (see also Chapter 4). This platform was developed as a means of controlling the variations that the target sequence could have on editing efficiency. However, it has been observed that T-DNA insertions have also implied modifications in the chromatin landscape (Jupe et al., 2018). This may imply that differences in editing efficiency found could be affected by these variations. A balance between both approaches, trying to reduce the variability caused by target sequence and avoiding the distortion caused by unintended alterations in the chromatin that could be induced by T-DNA insertion. On the other hand, the tests should focus on differences in editing efficiency, which would exclude the validating methods proposed in the chapter. As described in Chapter 2, other methods based on protoplast transformation may be more suitable for this experiment than the embryo-based protocol utilized in this chapter.

3.6. Conclusions

The results obtained in this chapter point to the ability of the method described in Chapter 2, involving embryo transformation followed by RE-PCR, to find edits in *Brachypodium* embryos. The results also suggest that the BdU6 and OsU3 promoters are able to express correctly the sgRNA, and point to TaU6 having a lower efficiency in *Brachypodium* than the other two. However, further experiments including the controls proposed in the discussion should be used in order to corroborate these findings. Finally, a platform for the study of the genomic environment on the ability of CRISPR/Cas9 to induce edits has been proposed.

3.7. References.

Aeby, E., Ullu, E., Yepiskoposyan, H., Schimanski, B., Roditi, I., Mühlemann, O., and Schneider, A. (2010). tRNASec is transcribed by RNA polymerase II in *Trypanosoma brucei* but not in humans. *Nucleic Acids Research* 38, 5833–5843.

Andersson, M., Turesson, H., Nicolia, A., Fält, A. S., Samuelsson, M., and Hofvander, P. (2017). Efficient targeted multiallelic mutagenesis in tetraploid potato (*Solanum tuberosum*) by transient CRISPR-Cas9 expression in protoplasts. *Plant Cell Reports* 36, 117–128.

Arimbasseri, A. G., Rijal, K., and Maraia, R. J. (2013). Transcription termination by the eukaryotic RNA polymerase III. *Biochimica et Biophysica Acta* 1829, 318–330.

Ashikari, M., Sakakibara, H., Lin, S., Yamamoto, T., Takashi, T., Nishimura, A., Angeles, E. R., Qian, Q., Kitano, H., and Matsuoka, M. (2005). Cytokinin Oxidase Regulates Rice Grain Production. *Science* 309, 741–745.

Azpiroz, R., Wu, Y., Locascio, J. C., and Feldmann, K. A. (1998). An Arabidopsis brassinosteroid-dependent mutant is blocked in cell elongation. *Plant Cell* 10, 219–230.

Barone, P., Wu, E., Lenderts, B., Anand, A., Gordon-Kamm, W., Svitashev, S., and Kumar, S. (2020). Efficient Gene Targeting in Maize Using Inducible CRISPR/Cas9 and Marker-free Donor Template. *Molecular Plant* 13, 1219–1227.

Battraw, M. J., and Hall, T. C. (1990). Histochemical analysis of CaMV 35S promoter-β-glucuronidase gene expression in transgenic rice plants. *Plant Molecular Biology* 15, 527–538.

Benfey, P. N., Ren, L., and Chua, N. H. (1990). Tissue-specific expression from CaMV 35S enhancer subdomains in early stages of plant development. *EMBO Journal* 9, 1677–1684.

Bian, Q., and Belmont, A. S. (2012). Revisiting higher-order and large-scale chromatin organization. *Current Opinion in Cell Biology* 24, 359–366.

Bortesi, L., and Fischer, R. (2015). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnology Advances* 33, 41–52.

Bragg, J. N., Wu, J., Gordon, S. P., Guttman, M. E., Thilmony, R., Lazo, G. R., Gu, Y. Q., and Vogel, J. P. (2012). Generation and Characterization of the Western Regional Research Center *Brachypodium* T-DNA Insertional Mutant Collection. *PLoS ONE* 7, e41916.

Brain, P. W. (1959). Effects of gibberellins on Plant Growth and Development. *Biological Reviews* 34, 37–77.

Briner, A. E., Donohoue, P. D., Gomaa, A. A., Selle, K., Slorach, E. M., Nye, C. H., Haurwitz, R. E., Beisel, C. L., May, A. P., and Barrangou, R. (2014). Guide RNA functional modules direct Cas9 activity and orthogonality. *Molecular Cell* 56, 333–339.

Butler, J. E. F., and Kadonaga, J. T. (2002). The RNA polymerase II core promoter: A key component in the regulation of gene expression. *Genes and Development* 16, 2583–2592.

Callis, J., Raasch, J. A., and Vierstra, R. D. (1990). Ubiquitin Extension Proteins of *Arabidopsis thaliana*. *The Journal of Biological Chemistry* 265, 12486–12493.

Čermák, T., Curtin, S. J., Gil-Humanes, J., Čegan, R., Kono, T. J. Y., Konečná, E., Belanto, J. J., Starker, C. G., Mathre, J. W., Greenstein, R. L., and Voytas, D. F. (2017). A multipurpose toolkit to enable advanced genome engineering in plants. *Plant Cell* 29, 1196–1217.

Chen, L., Zhao, J., Song, J., and Jameson, P. E. (2020). Cytokinin dehydrogenase: a genetic target

for yield improvement in wheat. *Plant Biotechnology Journal* 18, 614–630.

Chen, W., McKenna, A., Schreiber, J., Haeussler, M., Yin, Y., Agarwal, V., Noble, W. S., and Shendure, J. (2019). Massively parallel profiling and predictive modeling of the outcomes of CRISPR/Cas9-mediated double-strand break repair. *Nucleic Acids Research* 47, 7989–8003.

Chen, X., Liu, J., Janssen, J. M., and Gonçalves, M. A. F. V. (2017). The Chromatin Structure Differentially Impacts High-Specificity CRISPR-Cas9 Nuclease Strategies. *Molecular Therapy - Nucleic Acids* 8, 558–563.

Chen, X., Rinsma, M., Janssen, J. M., Liu, J., Maggio, I., and Gonçalves, M. A. F. V. (2016). Probing the impact of chromatin conformation on genome editing tools. *Nucleic Acids Research* 44, 6482–6492.

Choe, S., Dilkes, B. P., Fujioka, S., Takatsuto, S., Sakurai, A., and Feldmann, K. A. (1998). The *DWF4* gene of Arabidopsis encodes a cytochrome P450 that mediates multiple 22α-hydroxylation steps in brassinosteroid biosynthesis. *Plant Cell* 10, 231–243.

Chono, M., Matsunaka, H., Seki, M., Fujita, M., Kiribuchi-Otobe, C., Oda, S., Kojima, H., and Nakamura, S. (2015). Molecular and genealogical analysis of grain dormancy in japanese wheat varieties, with specific focus on *MOTHER OF FT AND TFL1* on chromosome 3A. *Breeding Science* 65, 103–109.

Christensen, A. H., and Quail, P. H. (1996). Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. *Transgenic Research* 5, 213–218.

Christensen, A. H., Sharrock, R. A., and Quail, P. H. (1992). Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. *Plant Molecular Biology* 18, 675–689.

Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., Giakountis, A., Farrona, S., Gissot, L., Turnbull, C., and Coupland, G. (2007). FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. *Science* 316, 1030–1033.

Corvalán, C., and Choe, S. (2017). Identification of brassinosteroid genes in *Brachypodium distachyon*. *BMC Plant Biology* 17, 5.

Coussens, G., Aesaert, S., Verelst, W., Demeulenaere, M., De Buck, S., Njuguna, E., Inzé, D., and Van Lijsebettens, M. (2012). *Brachypodium distachyon* promoters as efficient building blocks for transgenic research in maize. *Journal of Experimental Botany* 63, 4263–4273.

Daer, R. M., Cutts, J. P., Brafman, D. A., and Haynes, K. A. (2017). The Impact of Chromatin Dynamics on Cas9-Mediated Genome Editing in Human Cells. *ACS Synthetic Biology* 6, 428–438.

Dennis, E. S., and Peacock, W. J. (2009). Vernalization in cereals. *Journal of Biology* 8, 57.

Doenecke, D. (2017). "Chromatin Dynamics During the Cell Cycle," in *Chromatin Regulation and Dynamics* ed. Göndör, A. (Boston, MA: Academic Press), 117–139.

Dong, Y., and Wang, Y. Z. (2015). Seed shattering: From models to crops. *Frontiers in Plant Science* 6, 1–13.

Du, H., Zeng, X., Zhao, M., Cui, X., Wang, Q., Yang, H., Cheng, H., and Yu, D. (2016). Efficient targeted mutagenesis in soybean by TALENs and CRISPR/Cas9. *Journal of Biotechnology* 217, 90–97.

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* 3, e3647.

Filion, G. J., Filion, G. J., van Bemmel, J. G., van Bemmel, J. G., Braunschweig, U., Braunschweig, U., et al. (2010). Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. *Cell* 143, 212–224.

Fluhr, R., Kuhlemeier, C., Nagy, F., and Chua, N. H. (1986). Organ-specific and light-induced expression of plant genes. *Science* 232, 1106–1112.

Gao, Y., and Zhao, Y. (2014). Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *Journal of Integrative Plant Biology* 56, 343–349.

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* 6, 343–345.

Gu, X. Y., Liu, T., Feng, J., Suttle, J. C., and Gibbons, J. (2010). The *qSD12* underlying gene promotes abscisic acid accumulation in early developing seeds to induce primary dormancy in rice. *Plant Molecular Biology* 73, 97–104.

Guilfoyle, T. J. (1997). "The Structure of Plant Gene Promoters," in *Genetic Engineering. Genetic Engineering (Principles and Methods)*, ed. J. K. Setlow (Boston, MA: Springer), 15–47.

Happel, N., and Doenecke, D. (2009). Histone H1 and its isoforms: Contribution to chromatin structure and function. *Gene* 431, 1–12.

Hémonnot, C. Y. J., Ranke, C., Saldanha, O., Graceffa, R., Hagemann, J., and Köster, S. (2016). Following DNA Compaction During the Cell Cycle by X-ray Nanodiffraction. *ACS Nano* 10, 10661–10670.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G., and Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology* 31, 827–832.

Hui, L., Zhao, M., He, J., Hu, Y., Huo, Y., Hao, H., Hao, Y., Zhu, W., Wang, Y., Xu, M., and Fu, A. (2019). A simple and reliable method for creating PCR-detectable mutants in Arabidopsis with the polycistronic tRNA–gRNA CRISPR/Cas9 system. *Acta Physiologiae Plantarum* 41.

Hus, K., Betekhtin, A., Pinski, A., Rojek-Jelonek, M., Grzebelus, E., Nibau, C., Gao, M., Jaeger, K. E., Jenkins, G., Doonan, J. H., and Hasterok, R. (2020). A CRISPR/Cas9-Based Mutagenesis Protocol for *Brachypodium distachyon* and Its Allopolyploid Relative, *Brachypodium hybridum*. *Frontiers in Plant Science* 11, 614.

Isaac, R. S., Jiang, F., Doudna, J. A., Lim, W. A., Narlikar, G. J., and Almeida, R. (2016). Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. *eLife* 5, e13450.

Janssen, J. M., Chen, X., Liu, J., and Gonçalves, M. A. F. V. (2019). The Chromatin Structure of CRISPR/Cas9 Target DNA Controls the Balance between Mutagenic and Homology-Directed Gene-Editing Events. *Molecular Therapy - Nucleic Acids* 16, 141–154.

Jensen, K. T., Fløe, L., Petersen, T. S., Huang, J., Xu, F., Bolund, L., Luo, Y., and Lin, L. (2017). Chromatin accessibility and guide sequence secondary structure affect CRISPR/Cas9 gene editing efficiency. *FEBS Letters* 591, 1892–1901.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptative Bacterian Immunity.

Science 337, 816–822.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. *eLife* 2, e00471.

Jones, J. D., Dunsmuir, P., and Bedbrook, J. (1985). High level expression of introduced chimaeric genes in regenerated transformed plants. *The EMBO journal* 4, 2411–2418.

Jupe, F., Rivkin, A. C., Michael, T. P., Zander, M., Motley, S. T., Sandoval, J. P., Slotkin, R. K., Chen, H., Castanon, R., Nery, J. R., and Ecker, J. R. (2018). The complex architecture and epigenomic impact of plant T-DNA insertions. *PLoS genetics* 15, e1007819.

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30, 772–780.

Kim, S. I., Veena, and Gelvin, S. B. (2007). Genome-wide analysis of *Agrobacterium* T-DNA integration sites in the Arabidopsis genome generated under non-selective conditions. *Plant Journal* 51, 779–791.

Kouzarides, T. (2007). Chromatin Modifications and Their Function. *Cell* 128, 693–705.

Kunkel, G. R., Maser, R. L., Calvet, J. P., and Pederson, T. (1986). U6 small nuclear RNA is transcribed by RNA polymerase III. *Proceedings of the National Academy of Sciences of the United States of America* 83, 8575–8579.

Kunkel, G. R., and Pederson, T. (1988). Upstream elements required for efficient transcription of a human U6 RNA gene resemble those of U1 and U2 genes even though a different polymerase is used. *Genes & development* 2, 196–204.

Lawit, S. J., Wych, H. M., Xu, D., Kundu, S., and Tomes, D. T. (2010). Maize DELLA proteins dwarf plant8 and dwarf plant9 as modulators of plant development. *Plant and Cell Physiology* 51, 1854–1868.

Lawrenson, T., Shorinola, O., Stacey, N., Li, C., Østergaard, L., Patron, N., Uauy, C., and Harwood, W. (2015). Induction of targeted, heritable mutations in barley and *Brassica oleracea* using RNA-guided Cas9 nuclease. *Genome biology* 16, 258.

Li, Y., Zuo, S., Zhang, Z., Li, Z., Han, J., Chu, Z., Hasterok, R., and Wang, K. (2018). Centromeric DNA characterization in the model grass *Brachypodium distachyon* provides insights on the evolution of the genus. *Plant Journal* 93, 1088–1101.

Long, L., Guo, D. D., Gao, W., Yang, W. W., Hou, L. P., Ma, X. N., Miao, Y. C., Botella, J. R., and Song, C. P. (2018). Optimization of CRISPR/Cas9 genome editing in cotton by improved sgRNA expression. *Plant Methods* 14, 85.

Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., and Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology* 6, 26.

Luger, K., Dechassa, M. L., and Tremethick, D. J. (2012). New insights into nucleosome and chromatin structure: An ordered state or a disordered affair? *Nature Reviews Molecular Cell Biology* 13, 436–447.

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260.

Lusinska, J., Majka, J., Betekhtin, A., Susek, K., Wolny, E., and Hasterok, R. (2018). Chromosome identification and reconstruction of evolutionary rearrangements in *Brachypodium distachyon*, *B. stacei* and *B. hybridum*. *Annals of Botany* 122, 445–459.

Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., et al. (2015). A Robust CRISPR/Cas9 System for Convenient, High-Efficiency Multiplex Genome Editing in Monocot and Dicot Plants. *Molecular Plant* 8, 1274–1284.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., and Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science* 339, 823–826.

Mameaux, S., Cockram, J., Thiel, T., Steuernagel, B., Stein, N., Taudien, S., Jack, P., Werner, P., Gray, J. C., Greenland, A. J., and Powell, W. (2012). Molecular, phylogenetic and comparative genomic analysis of the cytokinin oxidase/dehydrogenase gene family in the Poaceae. *Plant Biotechnology Journal* 10, 67–82.

McElroy, D., Zhang, W., Cao, J., and Wu, R. (1990). Isolation of an efficient actin promoter for use in rice transformation. *Plant Cell* 2, 163–171.

Mekler, V., Minakhin, L., Semenova, E., Kuznedelov, K., and Severinov, K. (2016). Kinetics of the CRISPR/Cas9 effector complex assembly and the role of 3'-terminal segment of guide RNA. *Nucleic Acids Research* 44, 2837–2845.

Mikami, M., Toki, S., and Endo, M. (2015). Comparison of CRISPR/Cas9 expression constructs for efficient targeted mutagenesis in rice. *Plant Molecular Biology* 88, 561–572.

Nabavi, S., and Nazar, R. N. (2008). U3 snoRNA promoter reflects the RNA's function in ribosome biogenesis. *Current Genetics* 54, 175–184.

Nandy, S., Pathak, B., Zhao, S., and Srivastava, V. (2019). Heat-shock-inducible CRISPR/Cas9 system generates heritable mutations in rice. *Plant Direct* 3, e00145.

Nie, L., Thakur, M. Das, Wang, Y., Su, Q., Zhao, Y., and Feng, Y. (2010). Regulation of U6 Promoter Activity by Transcriptional Interference in Viral Vector-Based RNAi. *Genomics, Proteomics and Bioinformatics* 8, 170–179.

Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156, 935–49.

Nissim, L., Perli, S. D., Fridkin, A., Perez-Pinera, P., and Lu, T. K. (2014). Multiplexed and Programmable Regulation of Gene Networks with an Integrated RNA and CRISPR/Cas Toolkit in Human Cells. *Molecular Cell* 54, 698–710.

Odell, J. T., Nagy, F., and Chua, N. H. (1985). Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. *Nature* 313, 810–812.

Paule, M. R., and White, R. J. (2000). Transcription by RNA polymerases I and III. *Nucleic Acids Research* 28, 1283–1298.

Pearce, S., Huttly, A. K., Prosser, I. M., Li, Y. D., Vaughan, S. P., Gallova, B., Patil, A., Coghill, J. A., Dubcovsky, J., Hedden, P., and Phillips, A. L. (2015). Heterologous expression and transcript analysis of gibberellin biosynthetic genes of grasses reveals novel functionality in the *GA3ox* family. *BMC Plant Biology* 15, 130.

Pidoux, A. L., and Allshire, R. C. (2005). The role of heterochromatin in centromere function. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360, 569–579.

Potenza, C., Aleman, L., and Sengupta-Gopalan, C. (2004). Targeting transgene expression in research, agricultural, and environmental applications: Promoters used in plant transformation. *In Vitro Cellular and Developmental Biology - Plant* 40, 1–22.

Raissig, M. T., Abrash, E., Bettadapur, A., Vogel, J. P., and Bergmann, D. C. (2016). Grasses use an alternatively wired bHLH transcription factor network to establish stomatal identity. *Proceedings of the National Academy of Sciences of the United States of America* 113, 8326–8331.

Reece-Hoyes, J. S., and Walhout, A. J. M. (2018). Gateway recombinational cloning. *Cold Spring Harbor Protocols* 2018, 1–6.

Ren, C., Liu, X., Zhang, Z., Wang, Y., Duan, W., Li, S., and Liang, Z. (2016). CRISPR/Cas9-mediated efficient targeted mutagenesis in Chardonnay (*Vitis vinifera* L.). *Scientific reports* 6, 32289.

Sansbury, B. M., Hewes, A. M., and Kmiec, E. B. (2019). Understanding the diversity of genetic outcomes from CRISPR/Cas generated homology-directed repair. *Communications Biology* 2, 458.

Sato, K., Yamane, M., Yamaji, N., Kanamori, H., Tagiri, A., Schwerdt, J. G., Fincher, G. B., Matsumoto, T., Takeda, K., and Komatsuda, T. (2016). Alanine aminotransferase controls seed dormancy in barley. *Nature Communications* 7, 11625.

Schmidt, C., Pacher, M., and Puchta, H. (2019). Efficient induction of heritable inversions in plant genomes using the CRISPR/Cas system. *Plant Journal* 98, 577–589.

Shan, Q., Wang, Y., Chen, K., Liang, Z., Li, J., Zhang, Y. Y., Zhang, K., Liu, J., Voytas, D. F., Zheng, X., Zhang, Y. Y., and Gao, C. (2013a). Rapid and efficient gene modification in rice and *Brachypodium* using TALENs. *Molecular Plant* 6, 1365–1368.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J. J., Qiu, J. L., and Gao, C. (2013b). Targeted genome modification of crop plants using a CRISPR/Cas system. *Nature Biotechnology* 31, 686–688.

Singh, K. B. (1998). Transcriptional regulation in plants: The importance of combinatorial control. *Plant Physiology* 118, 1111–1120.

Smale, S. T., and Kadonaga, J. T. (2003). The RNA polymerase II core promoter. *Annual Review of Biochemistry* 72, 449–479.

Tang, X., Zheng, X., Qi, Y., Zhang, D., Cheng, Y., Tang, A., Voytas, D. F. F., and Zhang, Y. (2016). A Single Transcript CRISPR/Cas9 System for Efficient Genome Editing in Plants. *Molecular Plant* 9, 1088–1091.

Timko, M. P., Kausch, A. P., Castresana, C., Fassler, J., Herrera-Estrella, L., Van Den Broeck, G., Van Montagu, M., Schell, J., and Cashmore, A. R. (1985). Light regulation of plant gene expression by an upstream enhancer-like element. *Nature* 318, 579–582.

van der Schuren, A., Voiniciuc, C., Bragg, J., Ljung, K., Vogel, J., Pauly, M., and Hardtke, C. S. (2018). Broad spectrum developmental role of *Brachypodium* AUX1. *New Phytologist* 219, 1216–1223.

Venter, M., and Botha, F. C. (2010). Synthetic promoter engineering. *Plant Developmental Biology* 2, 393–414.

Verelst, W., Bertolini, E., De Bodt, S., Vandepoele, K., Demeulenaere, M., Pè, M. E., and Inzé, D. (2013). Molecular and physiological analysis of growth-limiting drought stress in *Brachypodium distachyon* leaves. *Molecular Plant* 6, 311–322.

Verkuijl, S. A., and Rots, M. G. (2019). The influence of eukaryotic chromatin state on CRISPR/Cas9 editing efficiencies. *Current Opinion in Biotechnology* 55, 68–73.

Vogel, J. P., Garvin, D. F., Mockler, T. C., Schmutz, J., Rokhsar, D., Bevan, M. W., et al. (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463, 763–768.

Waibel, F., and Filipowicz, W. (1990). U6 snRNA genes of Arabidopsis are transcribed by RNA polymerase III but contain the same two upstream promoter elements as RNA polymerase ll-transcribed U-snRNA genes. *Nucleic Acids Research* 18, 3451–3458.

Wang, X., Ye, L., Ursache, R., and Mähönen, A. P. (2020). An inducible genome editing system for plants. *Nature Plants* 6, 766–772.

Wang, Z. P., Xing, H. L., Dong, L., Zhang, H. Y., Han, C. Y., Wang, X. C., and Chen, Q. J. (2015). Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in Arabidopsis in a single generation. *Genome Biology* 16, 144.

Woods, D. P., Ream, T. S., Bouché, F., Lee, J., Thrower, N., Wilkerson, C., and Amasino, R. M. (2017). Establishment of a vernalization requirement in *Brachypodium distachyon* requires REPRESSOR OF VERNALIZATION1. *Proceedings of the National Academy of Sciences of the United States of America* 114, 6623–6628.

Wright, A. V., Sternberg, S. H., Taylor, D. W., Staahl, B. T., Bardales, J. A., Kornfeld, J. E., and Doudna, J. A. (2015). Rational design of a split-Cas9 enzyme complex. *Proceedings of the National Academy of Sciences of the United States of America* 112, 2984–2989.

Xie, K., and Yang, Y. (2013). RNA-Guided genome editing in plants using a CRISPR/Cas system. *Molecular Plant* 6, 1975–1983.

Xu, Y., Jia, Q., Zhou, G., Zhang, X. Q., Angessa, T., Broughton, S., Yan, G., Zhang, W., and Li, C. (2017). Characterization of the *sdw1* semi-dwarf gene in barley. *BMC Plant Biology* 17, 11.

Yan, L., Loukoianov, A., Blechl, A., Tranquilli, G., Ramakrishna, W., SanMiguel, P., Bennetzen, J. L., Echenique, V., and Dubcovsky, J. (2004). The Wheat *VRN2* Gene Is a Flowering Repressor Down-Regulated by Vernalization. *Science* 303, 1640–1644.

Yan, L., Wei, S., Wu, Y., Hu, R., Li, H., Yang, W., and Xie, Q. (2015). High-Efficiency Genome Editing in Arabidopsis Using YAO Promoter-Driven CRISPR/Cas9 System. *Molecular Plant* 8, 1820–1823.

Yoshioka, S., Fujii, W., Ogawa, T., Sugiura, K., and Naito, K. (2015). Development of a mono-promoter-driven CRISPR/Cas9 system in mammalian cells. *Scientific Reports* 5, 18341.

Zalewski, W., Orczyk, W., Gasparis, S., and Nadolska-Orczyk, A. (2012). *HvCKX2* gene silencing by biolistic or *Agrobacterium*-mediated transformation in barley leads to different phenotypes. *BMC Plant Biology* 12, 206.

Zeevaart, J. A. (2008). Leaf-produced floral signals. *Current Opinion in Plant Biology* 11, 541–547.

Zhang, T., Gao, Y., Wang, R., and Zhao, Y. (2017). Production of Guide RNAs *in vitro* and *in vivo* for CRISPR Using Ribozymes and RNA Polymerase II Promoters. *Bio-Protocol* 7, e2148.

Zhou, H., Liu, B., Weeks, D. P., Spalding, M. H., and Yang, B. (2014). Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. *Nucleic Acids Research* 42, 10903–10914.

Chapter 4. Developing a model for plant sgRNA efficiency prediction.

4.1. Introduction.

As introduced in Chapter 1, different factors that affect CRISPR/Cas function have been described. Given the nature of the sgRNA, a short RNA sequence, several features in its sequence have been described to have an impact on both the editing rates and accuracy of CRISPR/Cas9 induced edits. As expected, given its natural role for Cas9 targeting, off-target effects have been linked to the sgRNA sequence (Hsu et al., 2013). At the same time, its sequence has been also been linked to on-target efficiency (Mandal et al., 2014). Underlying features in the sequence, such as nucleotide content, sgRNA structure and melting temperatures, have been related to sgRNA activity (Liu et al., 2016). These features have been put in the context of CRISPR/Cas9 experiment design and have been implemented into activity prediction models and sgRNA design software (reviewed in Hanna and Doench, 2020). With time, sgRNA design tools have evolved from simple PAM finders to elaborate systems that predict on-target activity, identifying potential off-target sites, and predict the outcome obtained from editing using a specific sgRNA. This chapter will review different models used for sgRNA design and evaluate their ability of efficiency prediction when comparing on-target scores with the actual *in vivo* efficiency described in the literature. Then, a specific simplified model will be developed for the accurate prediction of on-target efficiency based on features found in the target sequence and its surroundings.

As previously introduced, sgRNA design tools will identify potential sgRNA target sequences. The user can provide a gene sequence, which will be aligned with the target genome if it is available in the database to ensure correct targeting in the desired species. Some tools are able to function without using a reference genome, requiring only a DNA sequence to find the possible target sites in said DNA sequence, or directly analysing pre-made user-selected guides. However, some tools require the alignment of the target sequences with a genomic sequence or for the user to specify if the sgRNA will be targeting a coding or non-coding sequence. As described in the general introduction, the first requirement of CRISPR/Cas9 is to have a PAM site. Depending on the tool, different alternative PAM sites are included as a possibility for potential target identification. Off-target scores will calculate the likelihood of the proposed sgRNA to induce cuts in sites different from the intended one. These calculations will be based on the target genome and possible alignments of the sgRNA spacer with it. On-target scores will predict the efficiency of the sgRNA when cutting in the intended site. Calculations of the on-target scores will be based on previous results defining features found in highly active sgRNA or inactive sgRNA and using them to predict the on-target activity. It is remarkable how different tools share algorithms (reviewed in Hanna and Doench, 2020). However, depending on the tool, different combinations of on and off-target predictors can be found in different platforms and interfaces. It is also remarkable that different algorithms have been obtained from data gathered from experiments with a similar approach.

4.1.1. On-target effect scores.

As previously introduced, on-target scores intend to predict editing efficiencies, described as probability of the CRISPR/Cas9 system to induce edits in the intended target site. Specific on-target prediction tools and algorithms have been developed for sgRNA design.

In Doench et al. (2014), a library-based experiment was able to assess the effects of sgRNA spacer sequence for both on-target and off-target effects. This library, composed of 1,841 sgRNA targeting mouse and human cell surface marker genes, was developed and introduced in cell lines using lentiviral vectors. With the results obtained using antibody labelling and Florescence Activated Cell Sorting (FACS), an algorithm for on-target efficiency calculation was developed.

This was based on ranking the sgRNA with different efficiencies, classifying the sgRNA in high and low efficiency, and comparing the nucleotide features between both groups. Then, each nucleotide in each position was treated as a binary variable (as its appearance or not), allowing independent weighting of each variable. This independent weighting, meaning giving a numeric value to a binary variable is known in machine learning as featurization. This study established the 30mer as the sequence of interest, including the 3 nucleotides preceding the PAM site, the PAM site itself, the 20 nucleotide spacer region of the sgRNA and the 4 nucleotides 5' from the homology region. This was also done for each two consecutive nucleotides within the guide. The GC content was also used as a variable, observing a non-linear relationship between GC content and efficiency, which concurred with previous references (Gagnon et al., 2014; Wang et al., 2014). A Support Vector Machine (SVM, Cervantes et al., 2020) was trained to generate the values corresponding to each feature. Using the values obtained, a score can be calculated by adding together the weights from each feature occurring in the sgRNA together with an intercept and obtaining the final value using a logistic regression. This was later defined as the "Rule set 1" and was coded in Python. More details on the specific values, formulas and calculations are provided in the Materials and Methods of this chapter.

In addition to the analysis previously described, Fusi et al. (2015) focused on the optimization of the *in silico* analysis, identifying guide features leading to an efficient gene knock-out. Two different sets of sgRNA were studied. Firstly, the same dataset used in Doench et al. (2014), named FC after the Fluorescence Cytometry, was analysed. Another set of 2,549 sgRNAs, named RES after the drug Resistance Assays, were also used in the development of their model. Fusi et al. (2015) focused on the optimisation of statistical analyses used to compare models, as well as the model development itself. Upon reanalysing the data from Doench et al. (2014), other features in addition to the ones described in the original reference were found to significantly contribute to a better prediction. This implied that the proposed model would use "order 1 features", meaning single nucleotides, "order 2 features", meaning dinucleotide features and combinations of nucleotides in specific positions and GC content in the 20mer sequence, which were already included in Doench et al. (2014). In addition to the features previously described, total nucleotide and dinucleotide counts, disregarding the position in the 30mer, location of the targeted cut in the coding sequence of a gene, and melting temperature of the 30mer and different parts of the 20mer were also considered. Other features were tested but were found to be redundant or not to significantly increase prediction accuracy. A notable case is the existence of microhomologies, which was found to be predictive on its own but did not significantly increase prediction accuracy, concurring with (Bae et al., 2014a).

The combination of the data and features found in Doench et al. (2014) and Fusi et al. (2015) led to the model described in Doench et al. (2016). Firstly, 3 different sgRNA libraries, GeCKO (Sanjana et al., 2014; Shalem et al., 2014), Koike-Yusa (Koike-Yusa et al., 2014) and Wang (Wang et al., 2014), all of them targeting human genes, were analysed using the Rule Set 1 described in Doench et al. (2014). Two other libraries developed in this study were analysed as well, Avana and Asiago, targeting human and mouse genes respectively. After this analysis, the machine learning algorithm was optimised, as well as the use of other features not included in the Rule Set 1 such as total nucleotide count in the 30mer, location of the target sequence in the context of the gene, other position dependent and independent dinucleotide features and the Tm of the 30mer and specific parts of it. The whole algorithm and values were denominated as "Azimuth 2.0", which is used by different sgRNA design tools. For more details regarding this tool, see Materials and Methods.

Moreno-Mateos et al. (2015) described another library-based study, in this case using 1,280 sgRNA targeting zebrafish genes. This study focused on sgRNA stability related to its nucleotide composition and its effect on Cas9 loading. It was found that G-rich and A-depleted sgRNA had a higher stability, which had a negative effect on the ensemble free energies (EFEs). This study also suggested that G enrichment may have a role in protecting the 5' end from exonuclease degradation. To test the effect of nucleotide composition and sgRNA efficiencies, the 30mer defined in Doench et al. (2014) was expanded to include 6 nucleotides outwards from the homology sequence and PAM site. This study used the nucleotide composition of the 20% most efficient sgRNAs to develop its model. With the results obtained, the sgRNA scoring method CRISPRscan was developed.

An unpublished model used the methods proposed in Fusi et al. (2015) and Doench et al. (2016) and the dataset published in Moreno-Mateos et al. (2015), developing a novel scoring method named Azimuth *in vitro*. This score has been included in the online tool CRISPOR (Concordet and Haeussler, 2018).

A different model was proposed in Chari et al. (2015), where CRISPR/Cas activity was studied in bacteria. This method was developed by introducing sgRNA targeting 1,400 genes using lentiviral infection. The sgRNA were divided in high and low activity in different Cas systems, comparing nucleotide frequencies between both groups. To develop a model, an SVM approach was taken. The model was tested *in silico* using the library described in Doench et al. (2014), observing certain correlation between them, but this correlation was hindered by the differences in the experimental process. This study also found no correlation between sgRNA specificity and activity.

Wang et al. (2014) focused on the description of off-target effects of CRISPR/Cas9 in human cells, observing almost no cleavage in potential off target sites found via sgRNA spacer sequence homology. Notably, most of these potential target sites were found in non-coding regions, pointing to a lack of relevance regarding undesired gene function alterations. It also studied the possibility of testing all sgRNA in a library simultaneously, using 73,151 sgRNA targeting 7,114 different genes, and was able to detect activity of almost all sgRNA using sgRNA barcode sequencing. In this study, the use of essential genes for human sgRNA screenings was also investigated, focusing on ribosomal genes, which other studies have defined as essential and useful for CRISPR/Cas9 screenings (Koike-Yusa et al., 2014; Shalem et al., 2014). The effect on sgRNA efficiency of the appearance of different nucleotide features was also investigated. It was observed that editing was impaired by a high or low GC content. Additionally, a lower knock-out efficiency was obtained when targeting the end of the coding sequence, whereas the targeting of the transcribed strand had a higher knock-out efficiency. The importance of the 3' end of the sgRNA (proximal to the PAM site) regarding Cas9 loading, favouring A and G, was also highlighted. An SVM was trained using the features described, which would be adapted to provide the algorithm named Wu Score.

Xu et al. (2015) first analysed different published datasets based on sgRNA libraries in (Gilbert et al., 2014; Koike-Yusa et al., 2014; Shalem et al., 2014; Wang et al., 2014; Zhou et al., 2014; Konermann et al., 2015). The main aim of their analysis was to find genes essential for the survival of an organism that could be used for large scale sgRNA screenings using negative selection. Reiterating previous findings, the study shows the usefulness of ribosomal genes for sgRNA screening in humans based on negative selection. Thus, it divided the sgRNA of interest in a "ribosomal" set, a "non-ribosomal" set and "mESC" set, referring to sgRNA targeting genes found to be essential in mouse embryonic stem cells. These were classified as efficient and non-

efficient sgRNA, and then analysed each one of the groups in each one of the datasets. The 40 bp sequence considered for the feature analysis consisted of the targeted sequence, and the flanking 3' and 5' region. It is remarkable how different sets favoured different features. An example of this is the favouring of a G at the 5' end of the homology sequence in the "ribosomal" and "non-ribosomal set", but this favouring was not observed in the "mESC" set. The features that had a significant impact on sgRNA efficiency were selected and analysed using the linear regression penalty model Elastic-Net (Zou and Hastie, 2005), obtaining feature descriptions based on the presence or absence of specific nucleotides in specific positions. This led to the description of the algorithm contained in the Spacer Scoring for CRISPR (SSC) software. The prediction ability of this model was also tested in non-essential genes, providing a method for scoring any type of sgRNA, and the use of positive screens rather than negative selection screenings. This approach was also applied to CRISPR interference (CRISPRi, Gilbert et al., 2013; Qi et al., 2013) and CRISPR activation (CRISPRa, Cheng et al., 2013; Gilbert et al., 2014; Tanenbaum et al., 2014; Chavez et al., 2015; Konermann et al., 2015; Zalatan et al., 2015) experiments, further expanding the potential of this model.

WU-CRISPR (Wong et al., 2015) provided different scores for on-target efficiency. The work was based on the re-evaluation of the studies done in Doench et al. (2014), using their dataset. It also used different factors that have been described to affect sgRNA efficiency. It first focused on the structure of the sgRNA, which was analysed *in silico* using the RNA-fold server (Lorenz et al., 2011, 2016), analysing the accessibility of different nucleotides and the importance of proper folding of the tracrRNA element in the sgRNA. In this analysis, sgRNA stability was predicted using the GC count (a higher GC count would imply a higher stability) but observed that a high GC content had a negative impact in editing efficiencies. At the same time, this analysis also found that non-functional sgRNA had a high thermodynamic stability of the RNA/DNA duplex formed between the sgRNA and the target. Specific patterns that may hinder the expression of the sgRNA or make the synthesis of the oligonucleotide difficult were also identified. The model also included the total nucleotide frequencies in the gRNA and position specific features. All features were put together and an SVM was trained to develop the final WU-CRISPR algorithm. The performance of the model was validated using a Receiver Operating Characteristic (ROC) curve analysis. The model was tested comparing it to the results described in Chari et al. (2015), observing the cross-compatibility of the model. In this study, the performance of the algorithms proposed in the original reference, Doench et al. (2014) and Xu et al. (2015), was also tested, observing how WU-CRISPR was more efficient in selecting functional sgRNA.

Stemmer et al. (2015) provided another sgRNA scoring method. This method led to the description of the online tool CCTop, which provides an off-target scoring method based on the likelihood of stable sgDNA/DNA heteroduplex formation, counting as well on the position of specific mismatches in the sequence. Then, all possible sgRNA in a given sequence are ranked depending on their off-target probability and the possible effects on the off-target sequence, counting as well on the distance between the intended target site and the off-target sites. Being a method highly dependent on available genomic sequences, human, mouse, medaka, zebrafish, stickleback, cavefish, *Caenorhabditis elegans*, Drosophila and Arabidopsis genomes have been included in the system. The prediction performance of this tool was measured *in vivo* using an introduced GFP in the Medaka genome.

This overview of different sgRNA design and scoring tools shows how, from similar approaches and results, different values predicting the efficiency of sgRNA are given. However, it has been found that not all models are accurate when extrapolating efficiencies of species others than

the ones they were trained on (Gagnon et al., 2014; Varshney et al., 2015; Haeussler et al., 2016). This was highlighted as well in Naim et al. (2020), where a comparative analysis between sgRNA scoring methods in plants was performed. The first thing drawn to attention was the need of some tools of a reference genome to obtain sgRNA which minimise the possibility of off-target effects. Depending on the tool, the genomes available may be predetermined, which makes these tools obsolete for designing guides for organisms whose genomes have not been included in their databases. At the same time, it found discrepancies between the efficiencies described *in vivo* and the efficiencies predicted by *in silico* methods. Besides the problems with sgRNA design, the on-target scores obtained from them did not correlate with the *in vivo* efficiencies described in plant-focused experiments. The conclusions of this paper highlight the need of plant-specific sgRNA design tools, given the discrepancies between *in silico* and *in vivo* efficiencies.

4.1.2. Off-target effect calculations and additional tools.

As described previously, off-target scores are defined as the probability of a sgRNA to target a Cas9 induced cut at a genome site different to one for which the gRNA was designed. Different papers have analysed off-target effects differently, focusing on different features in the target sequence (Wu et al., 2014; Singh et al., 2015), leading to different types of calculations for off-target scores.

One of the first off-target predictors was described in Hsu et al. (2013). This study examined sgRNA efficiency in human cells which lead to the description of the optimal tracrRNA part of a truncated sgRNA. The study also targeted a human gene with 57 different sgRNAs, analysing the effects of a possible mismatch in the homology sequence of a sgRNA. The results obtained were used to develop an algorithm providing a probability of an off-target site being targeted. This relied on a given sgRNA, which can target similar sequences in the genome with some mismatches. Different mismatches were found to have different effects on the possibility of the sgRNA binding to the off-target site, implying that not all off-target sites will be favoured equally. With this, the possibility of the CRISPR/Cas9 system targeting specific off-target sites can be ranked depending on the mismatches between the intended target site and a specific sequence.

One of the first tools to implement these type of calculations, CRISPR-P 1.0 (Lei et al., 2014), uses this algorithm. It starts identifying the possible off target sites using the BatMis algorithm (Tennakoon et al., 2012) to find all potential off target sites with a maximum of 6 mismatches. Then the possible off-target sites are scored using the Hsu et al. (2013) algorithm. The guides with the lowest number of potential off target sites are ranked and they are provided to the user. This type of calculation is also provided in other tools, such as Benchling and has been added to the R package Bioconductor in the tool CRISPRseek (Zhu et al., 2014).

Further iterations of CRISPR-P, CRISPR-P 2.0 (Liu et al., 2017a) and CRISPR-Local (Sun et al., 2019), integrate a different method, based on the Cutting Frequency Determination (CFD, Doench et al., 2016) to describe off-target effects. This tool is based on the work done in a previous paper, Doench et al. (2014), which at the same time was based on the method outlined by Hsu et al. (2013). In Doench et al. (2016), the off-target score is based on the percent activity values given possible mismatches in specific points of the sequence, which are gathered in a penalty matrix. This off-target penalty score is used in tools such as Broad Institute GPP sgRNA Designer (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design). These two papers also describe and refine, respectively, an on-target score calculation method.

CasOFFinder (Bae et al., 2014b) proposed a different tool for off-target predictions. The tool requires a reference genome and a proposed sgRNA provided by the user. Possible off-target effects of the sgRNA are identified, starting with PAM degeneracies, tolerating a NAG besides the common NGG PAM site. The entire genome of the target species will be scanned with the proposed guide followed by an NRG PAM site. Then, the proposed sgRNA sequence was compared with all the possible sgRNA found in the genome, delivering a ranked list of sgRNA with different number of mismatches in the sequence, obtaining the most likely off target effects first and least likely last. Following the same process, CasDesigner (Park et al., 2015) will rank the possible sgRNA in a given sequence that would have the least off-target effects. Both tools are based on the same algorithm and computing process. However, CasOFFinder can predict the possible off-target effects in a user given sgRNA, while CasDesigner proposes sgRNA with the least possible off-target effects. CasOFFFinder also uses the algorithm described in Bae et al. (2014a), which focused on the DSB repair via microhomology end joining (MMEJ). This implies that around the cut site, microhomologies will interact with each other, causing a deletion that will likely change the reading frame of the gene and lead to a knock-out of the resulting protein. In a similar manner, Shen et al. (2018) provides an algorithm capable of predicting the frequency of different products after editing at a specific target site. From a sgRNA library tested in mouse cells it was observed that most repair resulted from microhomology based deletions, microhomology-less deletions and single base insertions. Using these data, a machine learning model, inDelphi, was trained to predict the frequency of these outcomes at a specific target site. However, this algorithm has not been related to any tool to our knowledge.

In E-CRISP (Heigwer et al., 2014), a Bowtie algorithm is used to predict off target effects. They provide three different scores. First, a specificity score, calculated by subtracting from an initial 100 specificity score one point per possible off target site. The tool also provides different scores that can be classified as on-target scores, besides the outcome prediction using the microhomology scores described in Bae et al. (2014a). These scores will be explained in the following sections.

As described in the last section, off-target effects are predicted based on possible alignments of the proposed sgRNA with a reference genome counting on different mismatches having different effects on the binding of the sgRNA to an off-target site. However, the effect of off target sites targeting random non-coding sequences has been overstated, having a reduced effect in gene function in general (Doench et al., 2014; Ho et al., 2015). Nevertheless, the study of off target effects is key to reduce non-intended modifications in the genome using CRISPR/Cas9.

4.2. Objectives of the chapter.

Several tools have been developed for sgRNA design, providing predictions of on and off-target effects. However, discrepancies have been found between the efficiencies found *in vivo* using specific sgRNA and the ones predicted by these tools. In this chapter, different factors related to the gRNA sequence will be studied in a dataset obtained solely from plant-based experiments with different targets. These data were used to develop a model to estimate *in vivo* efficiencies depending on the gRNA sequence, and a simple method for scoring sgRNA with different sequences. To compare the performance of this new tool, the predicted efficiencies will be compared with the other sgRNA design tools and the scores they provide.

4.3. Materials and Methods.

4.3.1. Dataset collection and general analysis.

All the sgRNA analysed, used to develop the model and implemented in the adapted tools are summarised in the Supplementary Table 11. The dataset is referred to as the Plant Protoplast Dataset (PP dataset). The dataset was, as the title implies, collected from published protoplast-based experiments, focusing on the spacer sequences described in each reference. Using nucleotide blast (BLASTn, NIH-NCBI, U.S.A.), the sequence of the complete target gene was obtained. The 30mer sequence, as defined in Doench et al. (2014), was collected in an Excel file. When the sequence was not found in BLAST, the reference was searched for the required sequence and it was copied manually. To facilitate the development of the Excel functions, the 30mers collected were divided, obtaining one nucleotide per column and aligning all the 30mers using the PAM site.

The general features of the dataset were analysed using Excel and GraphPad. In case of the general nucleotide composition in a positional dependent manner, the percentages of different nucleotides in different positions were obtained using Excel. In case of GC content, positional independent nucleotide composition and Tm, the data was collected in Excel and the analysis was performed in GraphPad. The relationship between positional independent nucleotide composition or Tm with the average editing efficiency was analysed using linear regression, focusing on the fitness to the linear regression (r^2 and $S_{y,x}$) and the significance of the slope being different from 0 was assessed using an F Test, which was calculated using the integrated tool in GraphPad. In case of the GC content, a Gaussian Fit was used, focusing on the fitness to the non-linear regression (r^2 and $S_{y,x}$), also calculated using GraphPad.

4.3.2. Performance analysis of available online tools.

To test the correlation between the predicted efficiencies provided by available online tools and the experimental *in vivo* efficiencies described in the collected dataset, the online tool CRISPOR was used (Concordet and Haeussler, 2018). To avoid bias towards certain plant species, the system was set without a reference genome. In this tool, several existing algorithms are used to provide different scores for sgRNA efficiency, which were then compared using a linear regression analysis in GraphPad. This tool provides scores using different algorithms, including Doench et al. (2014) (Doench14), different versions of Doench et al. (2016) (OldDoench16, Doench16 and Azimuth in vitro), Moreno-Mateos et al. (2015) (Moreno-Mateos), Chari et al. (2015) (Chari), Wang et al. (2014) (Wang), Xu et al. (2015) (Xu), Wong et al. (2015) (Wu-CRISPR) and Stemmer et al. (2015) (CCTop).

To assess the correlation between predicted and *in vivo* described efficiency, the fitness of the data to the linear fit was observed, using the r^2 value and the $S_{y,x}$, and the slope of the best linear fit of the data. An F Test was used for assessing the significance of the slope being different from 0, which was calculated using the integrated tool in GraphPad.

4.3.3. Adaptation of existing models to Excel.

Three different sgRNA scoring methods were adapted to Excel to use as a proof of concept for the developed model. All of them focus on features in the homology sequence of the sgRNA and its surroundings and its effect on the on-target efficiency of the sgRNA. After adapting all three tools in Excel, the sequences analysed in the dataset collected were analysed using linear regression to assess the correlation between the predicted efficiencies and the *in vivo* efficiencies described. The linear regression analysis was performed as previously described.

4.3.3.1. Doench et al. 2014.

The work done in Doench et al. (2014) lead to the featurization of specific nucleotides in specific positions in the 30mer defined. The values given to the different features were obtained from the reference and collected in an Excel table. For the adaptation to the Excel coding, the 30mer was considered as the 3 nucleotides before the PAM site, the PAM site and the 24 nucleotides after it. This did not change in sgRNA with a spacer sequence longer or shorter than 20 nucleotides, obtaining a 30mer no matter the length of the spacer.

Having the features defined as nucleotides in specific positions and combinations of them, a simple command was used to develop a matrix of scores given when these nucleotides appeared in said position(s). Given the simplicity of the method, the matrix was developed in the same sheet as the sequence breakdown. The code was written as follows:

$$= IF(AF2 = \mathrm{G};\ Features!\$B\$3;\ 0)$$

Where *AF2* represents the 2nd nucleotide before the PAM site and *Features!$B$3* represents the value given for a G in that specific position.

This would give the value shown in the sheet "Features" when the nucleotide found in the 2nd nucleotide before the PAM site was a G. If it was not, the value given was 0.

When considering a combination of two nucleotides in consecutive positions, the code was written as follows:

$$= IF(AND(AF2 = \mathrm{G};\ AE2 = \mathrm{T});\ Features!\$B\$42;\ 0)$$

Where *AF2* represents the 2nd nucleotide before the PAM site and *AE2* represents the 3rd nucleotide before the PAM site. *Features!B42* represents the score defined for the appearance of a G and a T in the 2nd and 3rd nucleotides before the PAM site. This would give the value shown in the sheet "Features" when the nucleotide found in the 2nd nucleotide before the PAM site was a G and the 3rd nucleotide was a T. If it was not, the value given was 0.

The GC count was included in these calculations by counting the number of G or C in the 30mer sequence. The GC content was calculated using the following formulas:

$$= COUNT.IFS(L2{:}AO2; \mathrm{G})$$

$$= COUNT.IFS(L2{:}AO2; \mathrm{C})$$

$$= SUM(DK2{:}DL2)$$

Where *L2:AO2* represents the 20mer sequence and *DK2:DL2* represent the cells where the G and C counts were performed.

Doench et al. (2014) establishes the use of two different features depending on the GC content, having one for GC content over 10 and a different one for GC content under 10. In the adapted code this was represented as follows:

$$= IF(DM2 > 10;\ Features!\$B\$74;\ Features!\$B\$73)$$

Where *DM2* represents the GC content, *Features!$B$74* represents the score given for a GC content over 10 and *Features!$B$73* for GC content under 10.

To calculate the final scores, the formula provided in Doench et al. (2014) was used as follows:

$$g(sj) = int + \sum_i w_{ij}$$

$$f(sj) = \frac{1}{1 + e^{-g(sj)}}$$

Where *g(sj)* represents the addition of all the features analysed, obtained by the summation of the specific scores and the intercept (0.5976361542940) and *f(sj)* represents the final score, calculated as a function of the logistic regression.

4.3.3.2. Moreno-Mateos et al. 2015.

As previously described, Moreno-Mateos et al. (2015) uses a similar approach to Doench et al. (2014). However, the scores given to each feature are different. Given the similarities and differences of both methods, the feature scores given in Moreno-Mateos et al. (2015) were used with the algorithms used for Doench et al. (2014). The values given to the different features were obtained from the reference and collected in an Excel table.

Moreno-Mateos et al. (2015) defines its features around a 35mer instead of the 30mer described in Doench et al. (2014). The sequences were obtained and treated the same way as the previous analysis. A similar matrix was developed and the scores were calculated using the formulas described in Doench et al. (2014).

4.3.3.3. Doench et al. 2016.

Doench et al. (2016) expanded the feature list to 581, but works with the same 30mer defined in Doench et al. (2014). The values given to the different features were obtained from the reference and collected in an Excel table. It includes features similar to the ones described in Doench et al. (2014), but it also adds some others that required a different treatment. Given the extended list of features, they were divided in Simples (Single nucleotide features), Combined (2 nucleotide features, consecutive or non-consecutive), and General (nucleotide counts and T_m).

The single nucleotide and combined features were obtained in the matrix the same way they were obtained in the Doench et al. (2014) analysis.

The nucleotide counts were obtained as follows:

$$= COUNT.IF('sgRNA\ seq'!I2{:}AB2;\ \mathrm{A}) * General!\$D\$4$$

Where *'sgRNA seq'I2:AB2* represents the whole 30mer sequence and *General!$D$4* represents the score defined for each A in the sequence. This example would count the number of A found in the 30mer. Then, it would multiply the number of A found in the sequence per the feature score given.

The calculations regarding the GC content were performed as follows:

$$= IF\big(ADD\big(COUNT.IF('sgRNA\ seq'!I2{:}AB2;\ \mathrm{G}) + COUNT.IF('sgRNA\ seq'!I2{:}AB2;\ \mathrm{C})\big) > 10; General!\$B\$4; 0\big)$$

Where *'sgRNA seq'!I2:AB2* represents the 20 nucleotide spacer sequence and *General!$B$4* represents the score defined for GC contents higher than 10. This would give the value given in the bibliography if the addition of the count of G and C in the 20 nucleotide spacer sequence exceeded 10. If this did not happen, the value given back would be 0.

Another example of these is the number of times a combination of nucleotides appears in the 30mer sequence. To implement this count in the Excel coding, the following function was used:

$$= CONCATENATE(I2; J2)$$

$$= COUNT.IF('sgRNA\ seq'!AJ2{:}BB2; \mathrm{AA}) * General!\$I\$4$$

Where *I2;J2* represent consecutive nucleotides, specifically the 4th and 3rd nucleotide before the beginning of the homology region and *'sgRNA seq'!AJ2:BB2* represents the 30mer as consecutive nucleotides. This example would count the number of times AA was found in the 30mer sequence and multiply it by the feature score given in the reference.

The extended feature list includes the Tm of the 30mer and specific regions of the 30mer calculated using the "*Biopython_staluc*" function (Le Novère, 2002; Cock et al., 2009). Given the lack of a defined Tm threshold to apply the values described, this feature was left out.

Among the features described in the original reference, the location in a coding sequence was used as well. This would imply giving a positive value to sgRNA targeting a coding sequence and negative values to sgRNA targeting non-coding sequence. Since one of the aims of the chapter was to use the tool without the need of a reference genome, these features were not counted on.

The final scores were calculated using the formulas described in Doench et al. (2014).

4.3.4. CRISP-nt Plant Score.

To develop a model that predicted sgRNA efficiency in plants, an approach similar to the one described in the introduction was applied. This new tool will be based on the contribution of each nucleotide in each position to the efficiency of a guide. For the development of the model, the PP Dataset was used (See Supplementary Table 11). Following the guidelines from Doench et al. (2014), the 30mer sequence was analysed. For the calculation of the features related to the appearance of nucleotides in specific positions, the 30mer sequence then was divided in each one of the 4 possible nucleotides. If the nucleotide appeared in the sequence at that site, the editing efficiency described *in vivo* was added. This would imply that a nucleotide in a specific site would contribute to the related feature score in the model with the efficiency described *in vivo*. For the nucleotides not appearing at that site, a 0 was added, as they do not contribute to the editing efficiency described. After dividing the sequence into individual nucleotides, the nucleotide weights were calculated by adding the contribution of each nucleotide (it being the editing efficiency described *in vivo* or 0) and dividing by the possible efficiency (100% times the total count of sgRNA). The scores for each one of the single nucleotide features are shown in the Supplementary Table 12. The code to make these calculations was written as follows, starting with the division of the sgRNA in its nucleotide features:

$$= IF(sgRNA!I2 = \mathrm{A}; Scores!\$D\$2; 0)$$

Where *sgRNA!II2* represents the -4 position in the 30mer and Scores!$D$2 represents the efficiency described *in vivo*. Then, the calculation of the weight of each nucleotide to the model:

$$= \frac{SUM(A^{-4})}{\$AL\$1 * 100}$$

Where A^{-4} represents the summation of all the efficiencies found in the -4 position with adenines and *$AL$1* represents the total count of sgRNA introduced in the model.

To obtain a score regarding the nucleotide composition of each sgRNA, a one-hot encoding-based platform was developed in Microsoft Excel, giving back the score calculated by the model depending on the sgRNA sequence. This meant that a categorical variable, in this case the appearance of a nucleotide in a specific position, was given a numerical value, the feature score previously obtained. This facilitated the calculation of nucleotide related scores. To relativize the nucleotide related score, an "optimal" sgRNA was obtained by using the highest scores for each nucleotide in each position. To obtain a final score for each guide, the nucleotide features were added and the score is expressed as a percentage relative to that optimal guide. To normalize the data, before that relativization, the minimum score was calculated the same way the optimal sgRNA score was and subtracted to the maximum and specific sgRNA score.

The effect of GC content on guide efficiency was also studied. The GC content of the 20mer was calculated using the same code as described in the adaptation of Doench et al. (2014). Given the over-representation of low-efficiency guides, the *in vivo* efficiencies for each GC content were averaged. This was plotted and analysed using a Gaussian fit. The mean of this Gaussian fit, 12.19, was assumed as the optimal GC content. Then, the values given to each GC content were calculated as the deviation from this optimal GC content, applying a penalty to deviations below and above the optimal GC content. This was obtained using the following formula:

$$GC_{score} = 1 - \left|\frac{12.19 - GC_{Count}}{12.19}\right|$$

and coded as follows:

$$= 1 - ABS\left(\frac{12{,}19 - \left(COUNT.IF(sgRNA!M2{:}AF2;\,\mathrm{G}) + COUNT.IF(sgRNA!M2{:}AF2;\,\mathrm{C})\right)}{12{,}19}\right)$$

Where *sgRNA!M2:AF2* represents the 20mer sequence.

As suggested in Doench et al. (2016), the Tm of the 30mer and specific parts of this 30mer was calculated. In the original reference, the "*Biopython_staluc*" function (Le Novère, 2002; Cock et al., 2009) was used for this calculation. For its implementation in Excel, the T_m was calculated using the modified Marmur estimation (Marmur and Doty, 1962) was used as follows:

$$Tm = 2^{o}C\ (A + T) + \ 4^{o}C\ (G + C)$$

The score measuring the contribution of the Tm to the final efficiency was calculated by dividing the maximum Tm of the sequence studied (4 ^{o}C per nucleotide) by the actual Tm of the fragment studied. These features are counted on a preliminary model but not included in the final model.

From the model described in Doench et al. (2016), given that the total number of features added up to 1, the contribution of each feature can be relativized. With this, the weights of each one of the types of features studied can be relativized and a final score can be obtained for the CRISPRnt-Plant model. The weights used are summarised in Table 4.1.

Table 4.1. Relativization of the feature weights of the Doench et al. (2016) model and application to CRISPRnt-Plant.

Features	*Doench et al. 2016*	*CRISPRnt-Plant*
*Nucleotide Related Features**	0.74951076*	0.86621951
Tm (5mer end)	0.04275325	0.04941049
Tm (8mer middle)	0.0354462	0.04096564
Tm (5mer start)	0.00721575	0.00833934
Tm (30mer)	0.02121435	0.02451771
GC Content	0.00912624	0.01054731
Peptide Related Features	0.13473345	-
Total Contribution	1	1

* Including position dependent and independent features except GC content.

With this, the total scores obtained for each feature were relativized to add to 100. The final score was calculated by adding the features regarding nucleotide composition, GC content and melting temperature together.

The performance of the CRISPRnt-Plant model was analysed using linear regression as previously described.

4.3.5. Robustness tests.

Two different tests were performed to test the robustness of the CRISPRnt-Plant model.

The first test was based on random sampling of the PP dataset. This implied taking the 151 30mer dataset and dividing it randomly in groups of 10. The calculations to obtain the scores of the different position dependant nucleotide features were performed in each one of the groups. These scores were used to calculate the predicted efficiencies of the 151 30mers. It should be noted that the GC content and Tm calculations were made with the formulas described in the original method. Then, the predicted efficiencies were calculated using the scores obtained from each one of the groups. The predicted efficiencies were also calculated using the average scores among all the groups. The standard deviation was also calculated from those average scores to observe the variation of the scores depending on the sampling. The specific grouping done for this analysis was collected in Supplementary Table 13. The values obtained for the features in each one of the replicates and the average values are shown in Supplementary Table 14.

Secondly, a random removal test was performed. Ten random sgRNA were removed from the analysis (Supplementary Table 15). Then, the performance of the model was measured as the original version, using linear regression analysis. Those ten 30mers were then re analysed using the new features obtained (Supplementary Table 16), analysing the predictive ability of the model using linear regression.

4.4. Results

Each one of the sgRNA design tools gives a score for on-target and off-target effect. As outlined in the introduction, the scoring mechanisms for sgRNA effects differ from one another. In this section, the different methods will be compared with the results obtained in plants *in vivo*.

4.4.1. Dataset collection and general analysis

Firstly, guides used in papers describing genome editing in plants were collected from the bibliography, finding 151 sgRNA (Supplementary Table 11). For the analysis done in this chapter, only studies showing protoplast editing have been included, hence the name Plant Protoplast Database (PP Dataset). Guides were collected from different plant species, which showed differences in average guide efficiency and number of guides collected from said plant (Table 4.2). When focusing on the species represented in the dataset, the highest efficiency was described in *Brassica oleracea* (75.2%), while the lowest average efficiency was recorded in Arabidopsis (6.5%). It should be noted that these species only had one observation recorded in the dataset. The opposite was observed in maize and wheat, where the number of observations was 92 and 25 respectively, averaging 11.64 and 12.01. Due to the differences in number of observations between the species and the high Standard Error of the Mean (SEM), no conclusions about inter-species variations in editing efficiency were extrapolated from this analysis.

Table 4.2. The PP Dataset consists of sgRNA collected from different plants showing different reported editing efficiencies.

	Average Efficiency (%)	*N.O.*	*SEM*
Brassica oleracea	75.2	1	-
Brassica napus	56.8	1	-
Watermelon	46.85	2	4.75
Rice	32.76	9	7.69
Nicotiana tabacum	13.57	7	2.15
Wheat	12.01	25	2.67
Maize	11.64	92	1.44
Silene latifolia	10.34	2	10.05
Setaria italica	10.2	1	-
Potato	7.0	5	10.16
Arabidopsis	6.5	1	-
Bamboo	6.6	1	-

N.O. : Number of observations
SEM : Standard Error of the Mean

When analysing the distribution of efficiencies regardless of the species (Figure 4.1A), an overrepresentation of low efficiency sgRNA is observed, finding 56% of the total 151 sgRNA with a described efficiency of less than 10%. This was translated in an average efficiency of 14.7% when considering the whole dataset, with efficiencies ranging from 0.18 to 87.5%.

Using the spacer sequence described, the full gene sequence that each guide was targeting was obtained using BLAST methods and the 30mer sequence was recorded. Using the 30mer specified in Doench et al. (2014), a general analysis of nucleotide composition was performed. The frequency of each nucleotide in each position was calculated as a percentage (Figure 4.1B). If having different nucleotides in different positions of the 30mer did not have an effect in guide

function, the distribution would show around 25% of each nucleotide in each position, besides the NGG requirement of the PAM site. This distribution was not observed in some positions, with overrepresented and underrepresented nucleotides. To consider that a nucleotide was favoured or disfavoured in a position, it was decided it needed to have a deviation of 5% when compared to the standard 25%. These favoured and disfavoured nucleotides are summarized in Figure 4.1C. This simple analysis shows how the different sgRNA found in the dataset, although they have a high dispersion in the efficiencies described, have common nucleotides in specific positions and vary in others. The concurrencies and discrepancies of this analysis with different sgRNA efficiency predictive models and studies will be assessed in depth in the discussion of this chapter.

GC content and single nucleotide percentages in the spacer sequence have been used in predictive models of sgRNA efficiency (see introduction of this chapter). These were analysed in the collected dataset. Firstly, the GC content of each spacer sequence in the dataset was obtained and related with the average *in vivo* efficiency (Figure 4.2A). Consistent with previous observations, the dispersion of the described efficiencies in the PP Dataset is observed, as the high SEM in all average efficiencies suggests. However, a non-linear correlation between GC content and editing efficiency is hinted. When analysing the average efficiencies, this non-linear dependency was confirmed, as described in the bibliography. When analysing the Gaussian fit of the data, an optimal GC content of a spacer sequence, maximising efficiency, of 12.19 is implied. The goodness of fit of this distribution was proven by an r^2 value of 0.8325 with 9 degrees of freedom and an $S_{y,x}$ of 2.67. As previously described, the high dispersion of the data suggests that this measurement could not be accurate. However, the optimal GC content needed to be established for its use in the proposed model for sgRNA efficiency prediction.

Secondly, as suggested in Doench et al. (2016), the single nucleotide total counts were studied (Figure 4.2B). Using the same approach as the previous test, guide efficiency was plotted against the single nucleotide count. When observing the slopes of these correlations, it is notable how the T and C content do not have slopes significantly different from 0 (P values of 0.1177 and 0.9578 respectively). In contrast, both A and G contents have slopes significantly different from 0 (P values of 0.0005 and 0.0003 respectively). It is also notable that the slope of the A content shows a negative correlation and the G content shows a positive correlation with the editing efficiency described *in vivo*.

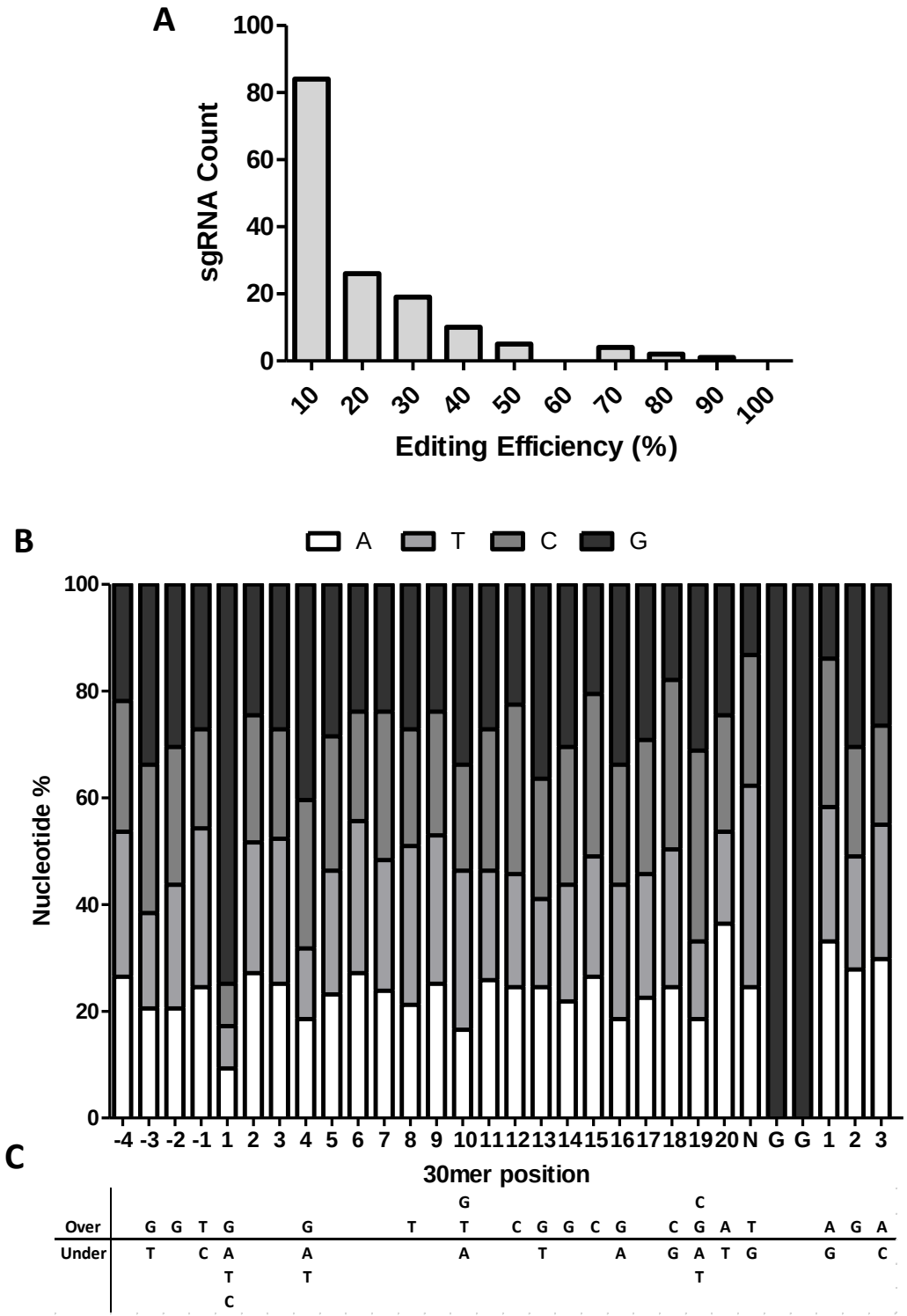


Figure 4.1. Nucleotide frequency analysis of the PP dataset shows favoured and disfavoured nucleotides in different positions.

(A) Distribution of the efficiencies described in the sgRNA collected in the dataset. Columns represent the number of sgRNA with frequencies described in each ten. Details on the number of observations are provided in Supplementary table 3.1. (B) Representation of the percentage of each nucleotide in each position of the 30mer. (C) Summary of the over and under-represented nucleotides in each position. The alignment with the figure above marks the specific position each cell is referring to.

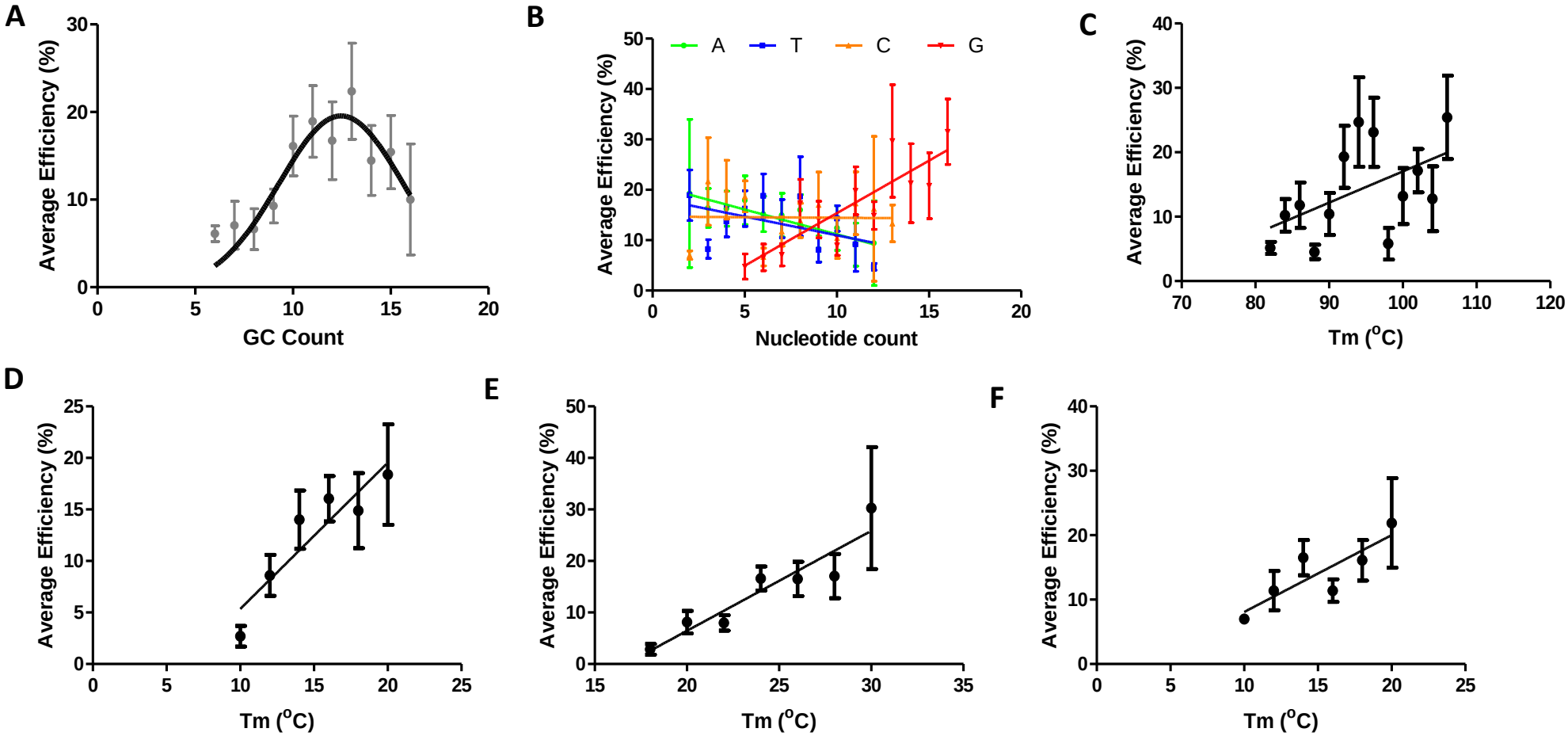


Figure 4.2. Analysis of different features suggested to affect sgRNA efficiency.
Different factors suggested to affect guide efficiency were analysed. (A) Plot of the average efficiency of sgRNA (Y axis) with different GC counts (X axis). The average editing efficiency and the SEM are shown in grey. The black line shows the best Gaussian fit to the data, calculated using GraphPad (see details in the text). (B) Plot of the average guide efficiencies and the nucleotide count of each single nucleotide found in the guide. The continuous lines show the best linear regression fit for the data. (C) Plot of the average efficiencies against the Tm of the analysed 30mer of those guides. The continuous line shows the best linear fit. (D) Plot of the average efficiencies against the Tm of the analysed 5mer proximal to the PAM site of those guides. The continuous line shows the best linear fit. (E) Plot of the average efficiencies against the Tm of the analysed 8mer proximal to the previous 5mer (away from the PAM site) of those guides. The continuous line shows the best linear fit. (F) Plot of the average efficiencies against the Tm of the analysed proximal to the 8mer (again away from the PAM site) of those guides. The continuous line shows the best linear fit.

Doench et al. (2016) also suggests the Tm of the 30mer and specific parts of the homology sequence play a role in guide efficiency. Following the same approach as the GC content, the Tm of interest were first calculated using the modified Marmur estimation (Marmur and Doty, 1962) as described in the Materials and Methods. Then, editing efficiency of guides with the same Tm of interest were averaged and plotted. It was observed that all 4 Tm of interest follow a linear correlation when plotted against the average efficiency of the guides with said Tm (Figure 4.2C, D, E, F). When considering the average efficiency related to the full 30mer Tm (Figure 4.2C), a high SEM is generally observed across different Tm. This dispersion of the data was observed as well in the statistical analysis of the best linear fit (r^2 0.2696, $S_{y,x}$ 6,494). Although a positive slope was observed (0.4850 ± 0.2407) the poor linear fit implied that it was not significantly different from 0 in an F test (P value 0.0690). On the other hand, the Tm analysis of the 5mer proximal to the PAM site (Figure 4.2D) showed an increase of the SEM with the increase of the Tm. However, a stronger correlation was observed (r^2 of 0.8469, $S_{y,x}$ of 2.524) with a positive slope significantly different from 0 (slope of 1.419 ± 0.3016, P value of 0.0093). Similar results were observed in the analysis of the 8mer in the centre of the homology region (Figure 4.2E), with an increase in variability with higher average efficiencies. As with the 5mer proximal to the PAM site, the linear fit was strongly supported by the (r^2 of 0.8750, $S_{y,x}$ of 3.470) and the slope was significantly different from 0 (slope of 1.940 ± 0.3279 with a P value of 0.0020). In a similar way, the final 5mer of the homology region (Figure 4.2F) showed a higher SEM with higher average efficiencies, with a good linear fit (r^2 of 0.7345, $S_{y,x}$ 3.002) and a slope significantly different from 0 (slope of 1.194 ± 0.3589 with a P value of 0.0292).

The results obtained from the linear regression analysis of the different Tm point to a positive correlation between increased Tm and editing efficiency in all 3 of the regions studied, except for the analysis of the full 30mer. Given the non-linear relation between GC content and guide efficiency and the dependency of the Tm calculations using the modified Marmur estimation of the GC content, a non-linear relationship was expected. These features were considered when developing a model fitting the dataset, as described in the Materials and Methods and as later outlined.

4.4.2. Performance analysis of online tools

To provide a performance analysis of the different methods available online, all 151 guides were analysed using CRISPOR (Concordet and Haeussler, 2018). To provide a graphical view of the data, it was plotted as an X-Y graph, showing the *in vivo* efficiencies as a percentage along the X axis and the predicted score on the Y axis (Figure 4.3). This plot was analysed using linear regression. Accurate predictions should relate 1 to 1 with the efficiency described *in vivo*, which would imply a slope that approximates 45 degrees. However, none of the methods used approached this slope. It should be noted how the dispersion of the data and the overrepresentation of low-efficiency guides could affect the accuracy of these models and their performance in this analysis, given the general lack of fitness to the linear regression proposed.

In case of "Doench14" (Figure 4.3A), the data did not fit adequately the linear regression proposed (r^2 of 0.0015, $S_{y,x}$ of 22.77). At the same time, the slope found in this linear regression was not significantly different from 0 (Slope of -0.05766 ± 0.1205 with a P value of 0.6331). When using the "OldDoench16" model (Figure 4.3B), having a slightly better linear fit is observed (r^2 of 0.02629, $S_{y,x}$ of 9.868) with a slope positive and significantly different from 0 (Slope of 0.1044 ± 0.05224 with a P value of 0.0475). "Doench16" (Figure 4.3C) shows further improvement on the fitness to the linear fit (r^2 of 0.03101, $S_{y,x}$ of 10.45) and an increase in the slope and, as expected, its significant difference from 0 (Slope of 0.1204 ± 0.05534 with a P value of 0.0311). The linear

fit of the scores obtained using "Moreno-Mateos" (Figure 4.3D) showed a lower fitness to the linear regression (r^2 0.01854 $S_{y,x}$ of 14.98), but the high dispersion of the data meant that even with a slope higher than the previous ones, this was not statistically significant (Slope of 0.1326 ± 0.07930 with a P value of 0.0966). The "Azimuth *in vitro*" model (Figure 4.3E) showed an increase in the performance when compared with the "Moreno-Mateos" score, finding a better linear fit (r^2 of 0.07089, $S_{y,x}$ of 13.03) and a significantly positive slope (Slope of 0.2318 ± 0.06896 with a P value of 0.0010). The score obtained in "Chari" (Figure 4.3F) showed a slightly better correlation than the previous ones (r^2 of 0.08144, $S_{y,x}$ of 26.13) with a slope significantly different from 0 (Slope of 0.5011 ± 0.1383 with a P value of 0.0004). The scores provided by "Wang" (Figure 4.3G) followed the same pattern as the previous ones, having a poor fitness to the linear regression with a high dispersion (r2 of 0.04833, $S_{y,x}$ of 16.81). In this case, the slope was found to be significantly different from 0 (Slope of 0.2439 ± 0.08898 with a P value of 0.0069). In case of "Xu" (Figure 4.3H), the scores obtained ranged from -2 to +2 instead of the 0-100 range the other tools provide, implying the absolute results may not be comparable with other tools. However, the linear regression showed similar results to the previous tools, finding a poor linear regression fitness (r2 of 0.01388, $S_{y,x}$ of 0.4007). In addition to that, the slope was found not to be significantly different from 0 (Slope of 0.003062 ± 0.002122 with a P value of 0.1511). The results obtained using "Wu-Crispr" (Figure 4.3I) followed the same pattern as the others, with a poor linear fit (r^2 of 0.004049, $S_{y,x}$ of 12.86) and a slope not significantly different from 0 (Slope of 0.04357 ± 0.1172 with a P value of 0.7124). Finally, the score obtained in "CCTop" (Figure 4.3J) did not show a good linear fit (r^2 of 0.02981, $S_{y,x}$ of 9.061), but had a slope significantly different from 0 (Slope of 0.1023 ± 0.04797 with a P value of 0.7124).

The linear regression analysis shows that generally, the fitness to the linear regression is poor. However, some of the models showed a slope significantly different from 0, although none of them approaches the expected slope of 1.

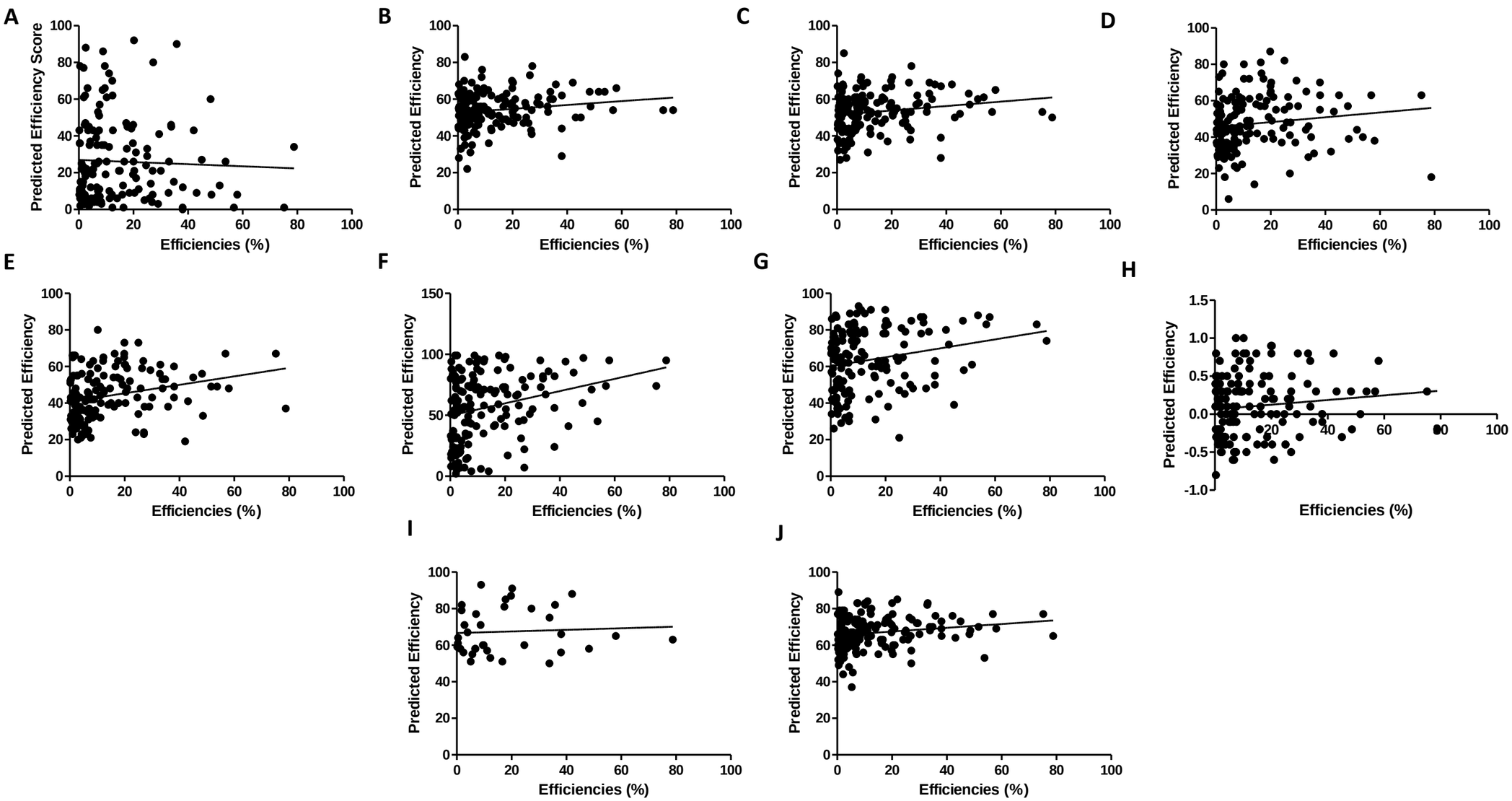


Figure 4.3. Predicted efficiencies and *in vivo* efficiencies do not correlate in a linear regression fit.
Linear regression of the predicted efficiency scores and the *in vivo* efficiencies described for each one of the analysed guides using different scoring methods available in CRISPOR. The scores were obtained using the scoring methods described in (A) Doench14, (B) OldDoench16, (C) Doench16, (D) Moreno-Mateos, (E) Azimuth in vitro, (F) Chari, (G) Wang, (H) Xu, (I) WU-CRISPR, (J) CCTop.

4.4.3. Adaptation of different methods in Excel

As proof of concept for coding the proposed model in Excel, different sgRNA efficiency scoring methods were adapted as described in the Materials and Methods section. Note that these adaptations might not show the same values as the ones obtained from the described online tools as some of the minor features were not included.

In case of Doench et al. (2014), it is notable that there is a high dispersion of the data. The linear fit of the data was not highly supported by the r^2 value and its standard error (r^2= 0.003231 with $S_{y,x}$=19.2). In addition to that, the slope does not differ significantly from 0, with a slope of -0.06588 ± 0.09480 with a P value of 0.4881 (Figure 4.4A).

When observing the regression obtained using the Moreno-Mateos et al. (2015) scores, the data shows a high variability in lower *in vivo* efficiencies, but a better fit when the predicted score is higher (Figure 4.4B). The values obtained for the regression were an r^2 of 0.06614 with a $S_{y,x}$ of 3.661. In this case the slope was significantly different from 0 (P value = 0.0014), with a slope of 0.05873 ± 0.01808. This approximates the expected regression, in a sense that guides with a higher described *in vivo* efficiency showed a higher predicted efficiency.

Similar results were obtained in the adapted Doench et al. (2016) scores (Figure 4.4C). The regression in this case showed an r^2 of 0,03894 and a $S_{y,x}$ of 1,537. The slope of the regression was positive and significantly different from 0 (0.01864 ± 0.007588 with a P value of 0,0152).

The adaptations of these algorithms showed deviations from the originally described versions but showed a similar lack of correlation between *in vivo* described efficiencies and predicted efficiencies.

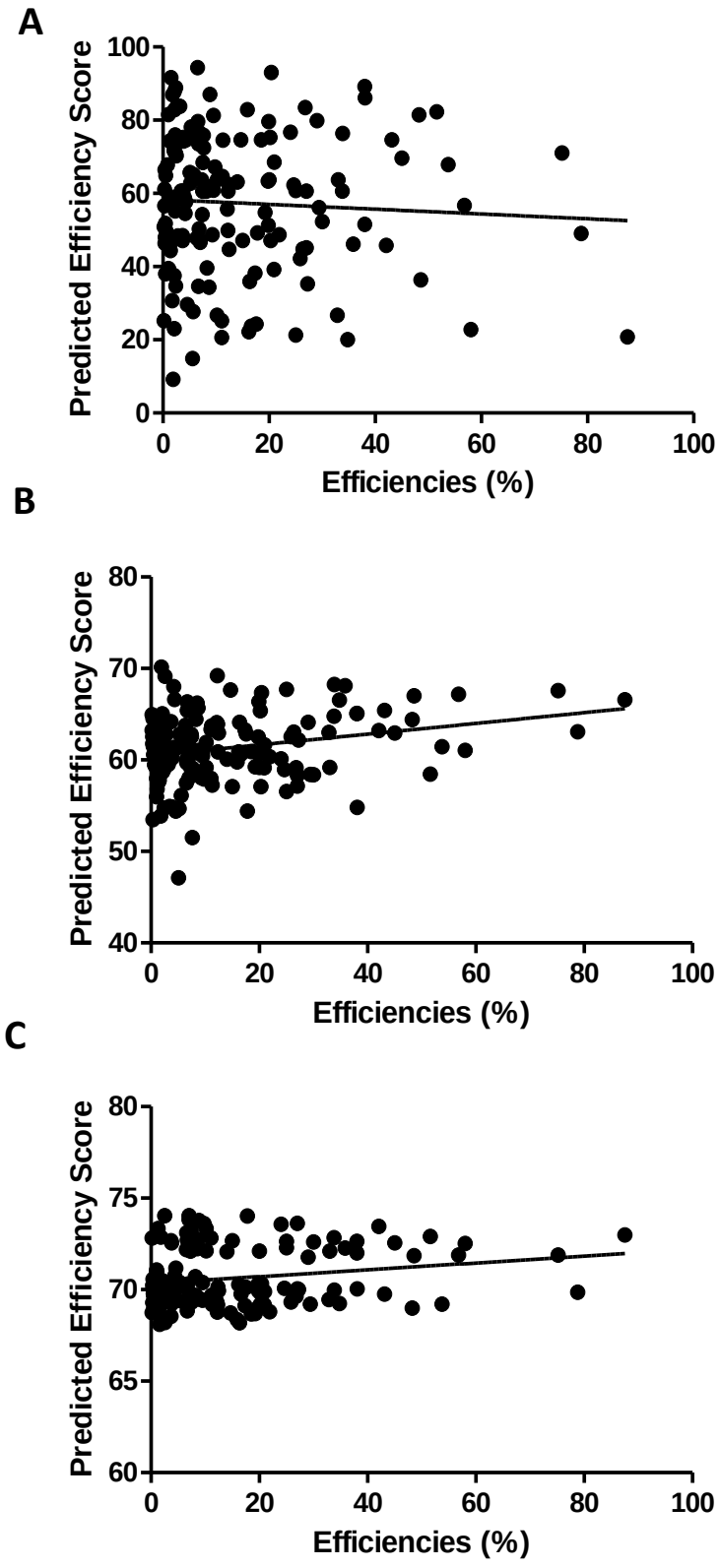


Figure 4.4. Predicted efficiencies and *in vivo* efficiencies do not correlate in a linear regression fit.

Linear regression of the predicted efficiency scores and the *in vivo* efficiencies described for each one of the analysed guides using different scoring methods. The scores were obtained using the adaptations of the scoring methods described in (A) Doench et al. (2014), (B) Moreno-Mateos et al. (2015), (C) Doench et al. (2016).

4.4.4. CRISPRnt-Plant

A model was developed based on the data obtained from the nucleotide analysis performed using the collection of 151 sgRNAs. The features included nucleotide appearance in each position, weighting each feature counting on the frequency of each nucleotide in each position and the described *in vivo* efficiency. It also included GC content, designing a feature to penalise deviations on the GC count above or below 12.19. Finally, it included the Tm calculations derived from different sections of the spacer sequence of the sgRNA.

In contrast with other tools, the single nucleotide percentages in the entire homology sequence and the Tm of the 30mer were not included in the model. The single nucleotide percentages were discarded as the high dispersion of the means measured by the SEM were considered too high to draw conclusions, even in the cases of A and G with slopes significantly different from 0. Under the same rationale, the Tm of the 30mer was left out of the final calculations given the high SEM recorded in the average efficiencies, which led to a slope not significantly different from 0. The features described were implemented into a one-hot encoding system in Excel, allowing the scoring of guides using this method given a user provided 30mer sequence. This scoring method was named CRISPnt-Plant.

It was observed that the feature scores utilised for the CRISPRnt-Plant model differed from the general percentage analysis of the PP dataset. To simplify this view, two heatmaps summarizing the nucleotide frequencies, as done in the PP Dataset general

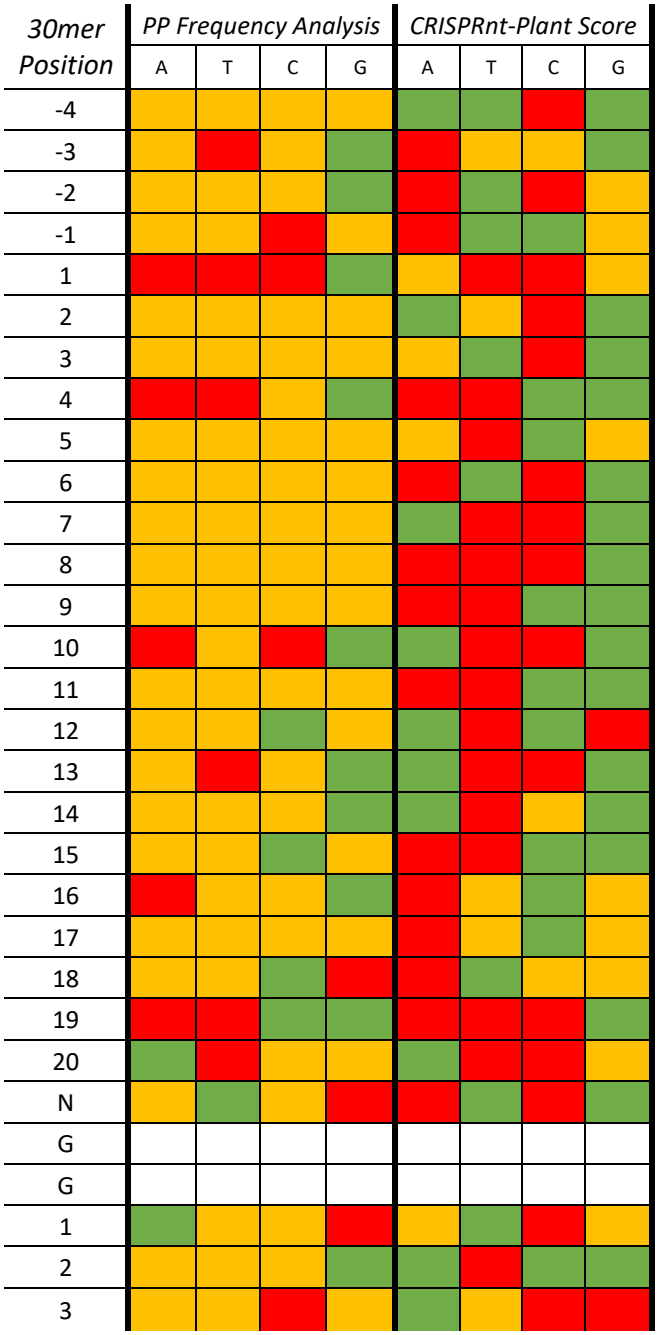


Figure 4.5. Comparing the nucleotide frequency analysis and the CRISPRnt-Plant Nucleotide Feature Scores highlights the independence of the model towards non-efficient and often used nucleotides.

Heatmap comparing the nucleotide frequency in the PP dataset in each position and the feature score given to each nucleotide in each position. In case of the frequency heatmap, in green are shown the nucleotides with higher than 30% appearance, in red the ones with lower than 20% appearance and in yellow the ones in between. In case of the feature score heatmap, in green are represented the features with a score higher than 0.15, in red the ones with a score lower than 0.14 and in yellow the ones in between. In white are shown the constant GG sites in the PAM site.

analysis, showing favoured, unfavoured and neutral features, and the value of the feature associated to said nucleotide are provided in Figure 4.5. In case of the heatmap of the nucleotide scores, similar groups were obtained by finding the average feature score and calculating 5% deviations from this average, having also favoured (above average plus 5% of the average), unfavoured (below average minus 5%) and neutral (between average plus 5% and average minus 5%). It is remarkable how there is certain correlation between the frequencies obtained from the general nucleotide analysis and the feature scores related to those nucleotides. However, the differences observed suggest that the model is able to include in the values both the frequency of nucleotide appearance and the efficiency related to said appearance.

Once the features of interest were calculated, the guide efficiency scoring method was tested representing the predicted scores against the *in vivo* efficiencies as an X-Y plot and finding a linear regression to fit the data (Figure 4.6). The model showed a positive slope of 0.3574 ± 0.03259 (r^2 = 0.4467, $S_{y,x}$ = 6.60), being significantly different from 0 with a P value of < 0.0001. As expected, the proposed model shows a better correlation with the data than the previously described models, since it was shaped by the same data it is describing. The intercept with the Y axis, which implies the minimum score given, was 50.19 ± 0.7196, which implies an over-estimation of sgRNA efficiency.

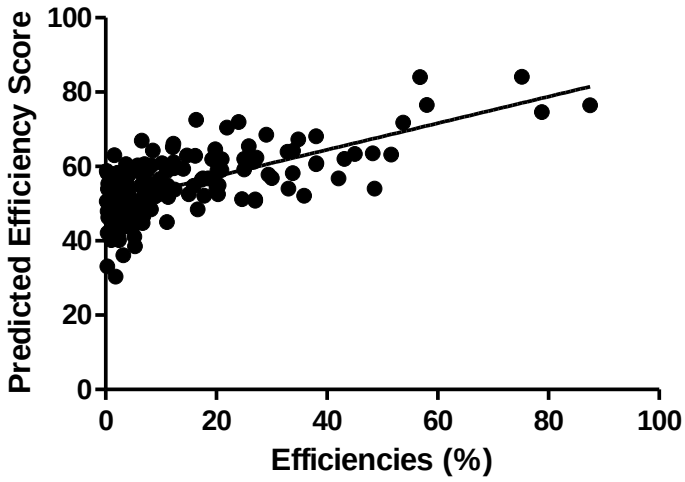


Figure 4.6. CRISPRnt-Plant performance was shown using linear regression analysis.
Linear regression of the efficiency scores obtained using the CRISP-NT method compared with the *in vivo* described efficiencies.

The performance of the CRISPRnt-Plant regression was compared with all the models studied in this chapter (Table 4.3). This comparison shows how CRISPRnt-Plant shows a better linear fit than any of the other tools used, which implies a better sgRNA efficiency prediction ability.

Table 4.3. Comparison of the statistical data obtained from the described, adapted and developed models.

Scoring Method	r^2	$S_{y,x}$	*Slope*	*P value*
Doench14	0.0015	22.77	-0.05766 ± 0.1205	0.6331
OldDoench16	0.02629	9.868	0.1044 ± 0.05224	0.0475
Doench16	0.03101	10.45	0.1204 ± 0.05534	0.0311
Moreno-Mateos	0.01854	14.98	0.1326 ± 0.07930	0.0966
Azimuth in vitro	0.07089	13.03	0.2318 ± 0.06896	0.0010
Chari	0.08144	26.13	0.5011 ± 0.1383	0.0004
Wang	0.04833	16.81	0.2439 ± 0.08898	0.0069
Xu	0.01388	0.4007	0.003062 ± 0.002122	0.1511
Wu-Crispr	0.004049	12.86	0.04357 ± 0.1172	0.7124
CCTop	0.02981	9.061	0.1023 ± 0.04797	0.0346
*Doench et al. 2014**	0.003231	19.2	0.06588 ± 0.09480	0.4881
*Doench et al. 2016**	0.03894	1.537	0.01864 ± 0.007588	0.0152
*Moreno-Mateos et al. 2015**	0.06614	3.661	0.05873 ± 0.01808	0.0014
CRISPRnt-Plant	0.4467	6.601	0.3574 ± 0.03259	<0.0001

*Results from the methods adapted in Excel.

4.4.5. Testing the robustness of the method

Two different analyses were performed to test the scoring method proposed, a random sampling analysis and a random removal test were used.

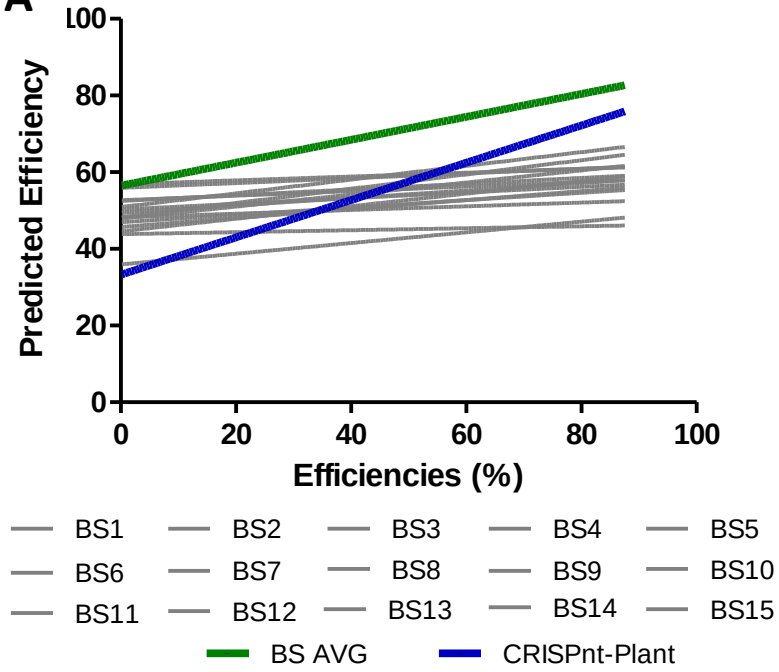
The first analysis performed was based on random sampling, dividing the 151 30mers in the PP dataset in 15 groups of 10 (Supplementary Table 13), and calculating the feature scores using the same methods as CRISPRnt-Plant using only the sgRNA in each group. Then, the predicted efficiency scores were calculated for all 151 sgRNA using the feature scores obtained from each group. The linear regression analysis was used to test the performance of each one of the repeats (Figure 4.7A, B). Of the 15 repeats, 9 showed a slope significantly different from 0 on the F test. This points to the robustness of the scoring method and its ability to discern low efficiency guides and high efficiency guides. However, the low r^2 value observed in all the repeats points to the same issues as the previous tests, a high dispersion of the data. The feature scores obtained from each one of the repeats were averaged, obtaining a new set of scores. The performance of the model using these average features was also analysed, using the same linear regression approach (Figure 4.7A, B). The performance of the average of the features of each group was better than the individual repeats, but it was not better than the original CRISPRnt-Plant score using 151 samples at once, finding a lower (however significant) positive slope and a lower r^2 value.

Using the features obtained from each one of the sampling groups, a simplified bootstrap approach was used, observing the variability of the feature scores in the different sampling

groups. The standard deviation of the mean of the feature score of each one of the features was observed. Not counting on the constant GG on the PAM site, where the standard deviation of the appearance of G was 0.04 due to the differences in the average efficiency in each group, the standard deviation ranged from 0.05 to 0.22. This range was divided in 3 equal tiers depending on the standard deviation, classifying the feature scores in low confidence (high standard deviation), high confidence (for low standard deviations) and non-conclusive (for standard deviations in the middle tier). To simplify the view, a heatmap was obtained (Figure 4.7C). The majority of the features (69 out of 120) were in the high confidence tier (standard deviation of less than 0.1), while only 8 were in the low confidence tier (standard deviation higher than 0.16). 43 were between these values. The fact that over 50% of the features were classified in the high confidence tier implies that, despite the high variation of efficiencies in the PP dataset, the features favoured and disfavoured are generally consistent between groups. At the same time, this analysis implies that the model used for the calculation of features is able to consistently give values that will accurately represent the efficiency of the guide they represent.

The random removal test implied 10 random sgRNA were removed from the 151 sgRNA used for the calculation of the feature scores (Supplementary Table 14). Using these scores, the predicted efficiencies of those 141 sgRNA were calculated. Then, the same linear regression analysis was performed (Figure 4.8A), observing almost no variation in the regression obtained, with a slope significantly different from 0 (Slope of 0.3635 ± 0.03291, with a P value of < 0.0001) and the fitness of the data to this regression (r^2 of 0.4675 with $S_{y,x}$ of 6.560). However, when analysing the guides that were not included in the feature score calculation, the expected significant positive correlation between the predicted efficiency and the *in vivo* described efficiency was not observed (Slope of 0.05689 ± 0.1669, P value of 0.7420, Figure 4.8B). At the same time, the linear fit was not statistically supported (r^2 of 0.01431 with $S_{y,x}$ of 5.764). Given that the model itself did not significantly change, it can be assumed that the deviations in predicting the efficiency of the guides left out is due to the limitations of the prediction model itself rather than variations due to the removal of the guides.

A



B

Scoring method	*Slope*	*P value*		*r²*	*Sy.x*
BS1	0.1170 ± 0.04152	0,0055	Significant	0,05061	8,409
BS2	0.09725 ± 0.03601	0,0077	Significant	0,04666	7,293
BS3	0.06259 ± 0.03286	0,0587	Not Significant	0,02377	6,654
BS4	0.1396 ± 0.04042	0,0007	Significant	0,07416	8,185
BS5	0.1640 ± 0.04165	0,0001	Significant	0,09423	8,434
BS6	0.02572 ± 0.04034	0,5248	Not Significant	0,002720	8,169
BS7	0.05944 ± 0.03487	0,0903	Not Significant	0,01913	7,061
BS8	0.1790 ± 0.03752	< 0.0001	Significant	0,1325	7,598
BS9	0.1861 ± 0.03535	< 0.0001	Significant	0,1569	7,158
BS10	0.07390 ± 0.04030	0,0687	Not Significant	0,02207	8,161
BS11	0.04993 ± 0.03315	0,1341	Not Significant	0,01500	6,713
BS12	0.04818 ± 0.03592	0,1818	Not Significant	0,01194	7,274
BS13	0.07492 ± 0.03531	0,0355	Significant	0,02933	7,151
BS14	0.09615 ± 0.04356	0,0288	Significant	0,03166	8,821
BS15	0.1571 ± 0.04104	0,0002	Significant	0,08958	8,311
BS AVG	**0.2992 ± 0.02995**	**< 0.0001**	**Significant**	**0,4012**	**6,065**
CRISP-Nt Plant	**0.3574 ± 0.03259**	**< 0.0001**	**Significant**	**0,4467**	**6,601**

C

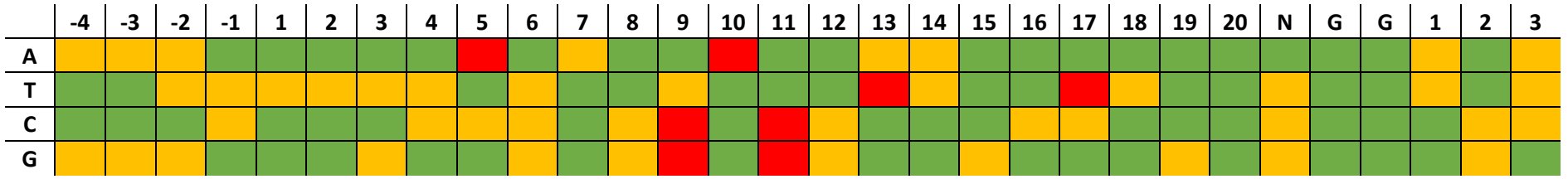


Figure 4.7. Random Sampling Analysis shows the robustness of the scoring method despite the reduced number of 30mers analysed.
(A) Linear regression analysis of the efficiency scores obtained using the different groups obtained from the random sampling (grey) and the score obtained using the average score (green). The linear regression analysis of the CRISPRnt-Plant scores is provided in blue. (B) Summary table of the linear regression analysis of the bootstrapped scores. BS represents each one of the groups analysed, whereas BS AVG shows the scores obtained from the average of the features obtained in each repeat. (C) Heatmap of the standard deviations of the average scores given to each nucleotide in each position of the 30mer. In green are represented the features in the high confidence tier, in red the ones in the low confidence tier and in yellow the ones that fell between the thresholds used.

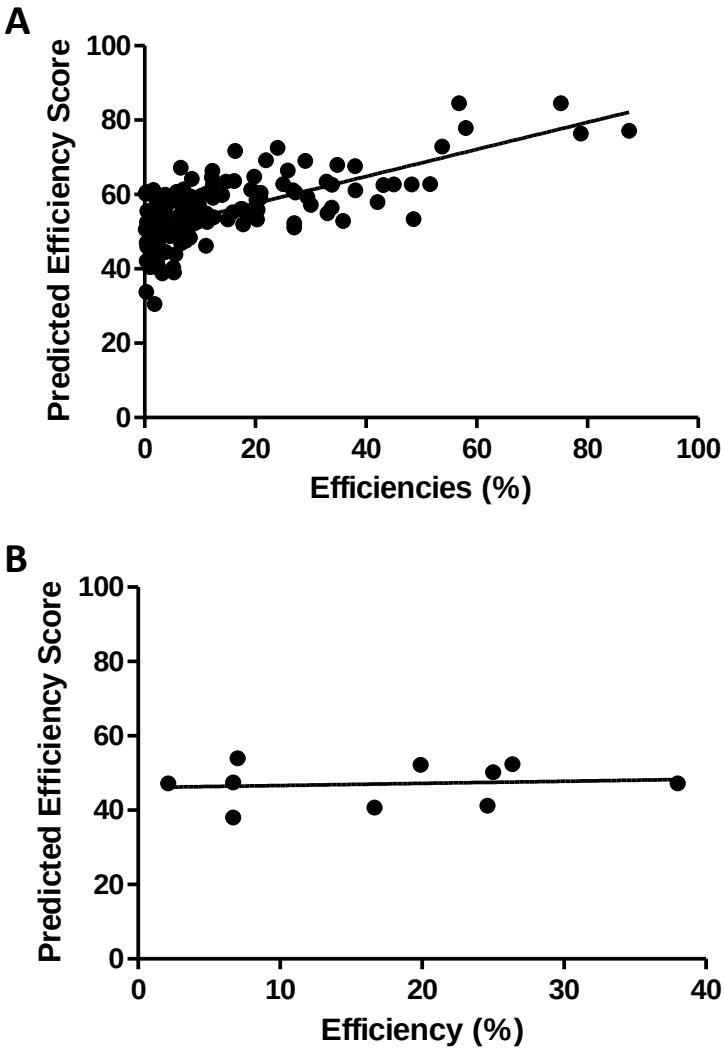


Figure 4.8. Random removal of sgRNA in the CRISPRnt-Plant showed the robustness of the model but highlighted the limitations of sgRNA efficiency prediction.
Linear regression analysis of the efficiency scores obtained using (A) the CRISP-NT method after removing 10 random guides compared with the *in vivo* described efficiencies and (B) the CRISPRnt-Plant scores from those 10 removed guides compared with the *in vivo* described efficiencies.

4.5. Discussion.

The results described in this chapter point to a lack of accuracy of the currently available editing efficiency predictors regarding plant-based studies. In order to provide an improved method for scoring sgRNA based on plant-related experiments, the CRISPRnt-Plant method was developed. The correlation analysis of the different scoring methods and the *in vivo* described efficiencies showed a better predictive performance by the new proposed method. The results highlight the need of plant-specific library-based studies focusing on sgRNA efficiency in order to accurately predict CRISPR/Cas9 editing efficiencies.

4.5.1. Forming a dataset from published sources.

The PP dataset was based on a collection of protoplast-based experiments describing editing efficiencies in different plants (Supplementary table 11). In contrast with other library-based experiments, the results were derived from a wide variety of species. This could have an effect in the variability of the dataset (Figure 4.1A).

The decision to restrict the dataset to protoplast-based studies aimed to reduce the possible variability derived from other methods of transformation. Protoplasts are a type of cell with, generally, a high and reproducible transformation efficiency, which would simplify the screening process. It should also be noted that protoplast-based experiments did not use the same type of screening (Supplementary Table 11), which could also lead to variations in efficiencies described, as introduced in Chapter 2. Protoplast-based experiments usually define sgRNA efficiency as the number of edited cells in a pool of total transformed cells. Efficiencies derived from embryo or callus-based methods were excluded as it is not easy to quantify the number of cells transformed. Usually, editing efficiency in embryos or callus is represented as number individual explant pieces showing editing out of the total transformed or the number that were successfully selected after the transformation.

All the protoplast experiments relied on the introduction of a plasmid into the system, which implies the expression of foreign genetic material. As discussed in Chapter 3, the selection of promoters for the expression of the Cas9 and sgRNA have a key role in editing efficiency (Liu et al., 2017b). In plants, optimisation of Cas9 expression also plays a role in editing rate, which may have a detrimental effect in less favoured species where specific Cas9 optimisation has not been studied. However, this would have included another variable to this study which the models would not have been able to handle. In the future, this type of studies should be performed on libraries with a common expression system for all the sgRNA studied, as references describing other models use.

Another factor to consider relating to editing efficiency is the structure of the sgRNA, as it is key to the correct targeting and loading of the Cas9. This could be a source of variability in editing efficiencies. In this study, however, the dataset used did not include the sgRNA scaffold. The possible effect of this factor in sgRNA efficiency has already been proven and introduced in other predictive models (See Introduction of this chapter). In the future, library-based studies would help solving this issue, since the variations of different sgRNA scaffolds used could be disregarded.

All these factors help explain the dispersion of editing efficiency data, which ranged from 0.18 to 87 %. At the same time, the lack of optimal systems for some of the plants may be a reason why there is an over-representation of low-efficiency guides in the dataset. The impact of this in the data obtained will be assessed later.

4.5.2. The bibliography described similar features to the ones observed in the dataset.

Different tools have described and identified a different set of features that affected sgRNA efficiency. With the general analysis of the PP Dataset and the description of the CRISPRnt-Plant model (Figure 4.1C, D), some of the same features were identified to correlate with sgRNA editing efficiency.

Moreno-Mateos et al. (2015) pointed to the enrichment in G of the first 14 nucleotides next to the PAM site in the homology sequence, while from the 15th to the 18th there was an enrichment for C. They also described a general depletion of A and T along the homology sequence except for the 9th and 10th nucleotides. The results derived in this study also found a similar trend, with a general underrepresentation of A and T and an overrepresentation of G and C along the homology sequence. This also concurs with Doench et al. (2014) and Gagnon et al. (2014) with defining the last position of the homology sequence as G rich. The analysis of the PP dataset also showed this a significant overrepresentation of G in this position. Doench et al. (2014) also pointed to a depletion of C in that position, which was also observed in the PP dataset. However, Moreno-Mateos et al. (2015), concurring with Doench et al. (2014), defines the variable nucleotide of the PAM site as GC rich, whereas an underrepresentation of G in that position in the PP dataset. It also shows an enrichment in G and depletion of C, while we find the opposite in the PP dataset.

Moreover, Chari et al. (2015) described an overrepresentation of G and under representation of T in the first nucleotide of the homology sequence adjacent to the PAM site. Analysis of the PP dataset revealed the same underrepresentation of T, but an overrepresentation of A in that position.

Xu et al. (2015), concurring with Wang et al. (2014), also found the first and second nucleotides of the homology sequence next to the PAM site to be GC rich, enhancing Cas9 loading. Analysis of the PP dataset concur with the GC enrichment in the second nucleotide upstream of the PAM, but do not show the same for the first nucleotide. Lower efficiencies with multiple Ts in the homology sequence next to the PAM site were also observed in the PP dataset. Xu et al. (2015) related this feature to lower sgRNA expression, concurring with Wu et al. (2014). This publication also pointed to increased efficiencies with a C in the third nucleotide, concurring with Cong et al. (2013). In general terms, Xu et al. (2015) found that nucleotides 5 to 12 showed A overrepresentation, whereas 14 to 17 showed G overrepresentation. This was not found in our results and contradicts other references previously described.

The main focus of Wong et al. (2015) was the addition of structural features to sgRNA design mechanisms. They detailed the different nucleotides in specific positions being essential for correct CRISPR/Cas9 function, combining the targeting sequence and the scaffolding sequence, to ensure the correct folding of the sgRNA and subsequent loading of Cas9. This paper also highlighted the differences in A count between functional and non-functional sgRNA, finding a significantly higher A count in functional sgRNA, in contrast with the findings of this study.

The observations made in the general analysis of the nucleotide content of the PP dataset and the features described in some of the tools available online are equivalent in some cases. This points to specific features that could be generally associated with sgRNA efficiency. However, the lack of correlation of others could be related to different features having a positive or negative impact in sgRNA efficiency in other taxa when compared to plants. This leads to the

suggestion that specific models and sgRNA designers may be able to improve sgRNA efficiency prediction for plant systems.

4.5.3. The tools available online are not able to reliably predict sgRNA efficiencies in the dataset studied.

The use of online resources for sgRNA design and efficiency prediction is of common use. These tools usually help in finding PAM sites within a given sequence, provide sequence features such as restriction sites that can be useful for screening edits, design primers for screening said edits, provide different off-target prediction mechanisms, and predict editing efficiencies using different algorithms. Several tools do include plant genomes, making both on and off-target calculations possible. In case of on-target scores, the process relies on algorithms, which, as stated in the introduction, use the results of library-based studies. These have been performed in human, mouse or zebrafish (See Introduction of this chapter), but there have been no sgRNA efficiency studies based on sgRNA libraries in plants. However, two systems have been defined as plant specific. The program CRISPR-P (Liu et al., 2017a) is supposed to be plant specific and includes several plant genomes, but the algorithm used is the one described in Doench et al. (2016). The CGAT program (Brazelton et al., 2015) also focused on plants, but details of the algorithm were not provided in the original reference. Different sgRNA library-based studies have been performed in plants such as rice (Meng et al., 2017; Lu et al., 2017), tomato (Jacobs et al., 2017) and soybean (Bai et al., 2020). On-target efficiency data from these studies could improve the predictive ability of already described methods if used as a training set, as well as improve the reliability of the CRISPRnt-Plant method proposed in this chapter.

The results obtained in this chapter challenge the notion of sgRNA efficiency prediction using the currently available algorithms, given a lack of correlation between the predicted efficiencies and the *in vivo* described efficiencies. Similar results were described in Naim et al. (2020). These authors were able to edit a transgene in *Nicotiana benthamiana*, reporting different editing efficiencies *in vivo*. These efficiencies were compared with scores obtained from different online tools. The same test was performed in endogenous genes in *N. benthamiana*, *Nicotiana tabacum*, *Arabidopsis thaliana*, *Musa acuminata* and *Oryza sativa*. This paper reported differences between the *in vivo* described efficiencies and the predicted efficiencies obtained from different online tools. The analysis of the PP dataset also found a lack of correlation between the described and predicted efficiencies. This lack of correlation may be due to the differences in sgRNA behaviour in the species the libraries were designed. The differences in sgRNA activity in a species dependent manner should be tested *in vivo*, testing the same sgRNA in different species. The analysis performed in Chari et al. (2015) and Xu et al. (2015), observing different algorithm performances when analysing different datasets, points to the same conclusions.

The accuracy of the predictive methods was tested by correlating the predicted and described efficiencies. This correlation was then analysed using linear regression. The rationale behind this test would be that an optimal sgRNA efficiency predictor would see equal values of *in vivo* described efficiency and predicted efficiency. This would be translated into a linear regression with a slope of 1 (45^{o}). It was observed that none of the models approached this regression (Figure 4.3), having some of them with slopes not significantly different from 0 after the statistical analysis using an F test. In addition to the lack of 1 to 1 correlation, the best linear fit found was not strongly supported by the r^2 value in any of the cases. When observing the data, no other type of regression consistent with the biological process studied (i.e. logistic regression)

fit the data either. This dispersion of the data implied that the models studied were not suitable for sgRNA efficiency prediction in the dataset studied.

4.5.4. Adaptation of different models in Excel.

Three of these tools were adapted and implemented in Excel: the first on-target predictor, described in Doench et al. (2014), and two of the most prominent models used, described Moreno-Mateos et al. (2015) and Doench et al. (2016), as defined in Haeussler et al. (2016). These were chosen for their relevance and for possibility of implementing the features in a one-hot encoding system.

In case of Doench et al. (2014), after developing the matrix of values, the addition of all the scores and the use of the formula described in the Materials and Methods provided the final score for each guide. The results obtained from the original version (Figure 4.3A) and the version adapted in Excel (Figure 4.4A) differ in the absolute values obtained. There was no correlation observed in the original version, but was present in the adapted version, having a higher r^2, lower $S_{y,x}$ and a slope significantly different from 0. The differences in the values obtained could be explained by differences in the application of the code or the final formula between the adapted and original algorithms. However, the implementation of this system as a proof of concept for one-hot encoding systems in Excel was successful.

The model described in Moreno-Mateos et al. (2015) also relied only on single nucleotide features and dinucleotide features, making it possible to implement them in the one-hot encoding used in Excel. The final calculations were not outlined in Moreno-Mateos et al. (2015). Given the similar data provided and the similar analysis used for the description of the features, the formula described Doench et al. (2014) for the calculation of the final score was used. When comparing the results obtained from the original reference (Figure 4.3D) and the ones obtained using the adapted model (Figure 4.4B), an improvement on r^2, $S_{y,x}$ and significance of the positive slope were obtained in the adaptation version. This implied that the formula used was not exactly the same as the original reference. However, the further implementation of one-hot encoding systems in Excel using this method was valuable for the development of our own model.

The model described in Doench et al. (2016) added additional features to the analysis. In addition to single and dinucleotide features, together with the GC content, the Tm of the different parts of the sgRNA were included to the model. Further combinations of nucleotides were added, such as appearance of specific nucleotides around the PAM site or the total nucleotide and dinucleotide counts along the 30mer. In addition to these, features regarding the position of the target relative to the coding sequence of a gene were added. Due to lack of detail about specific dinucleotide feature locations and Tm thresholds, these features were left out of the adapted model. Features related to target location in the gene were also excluded. This decision was taken given that the original analysis was made without the addition of a reference genome, therefore did not count on peptide positional features. In this case both models showed a slope significantly different from 0, but showed a similar fitness to the linear regression as the r^2 value shows. The differences between the adapted (Figure 4.4C) and original (Figure 4.3B, C) model implies that no conclusions regarding the predictive abilities of the model proposed in Doench et al. (2016) can be obtained from the adapted model. However, the adaptation further improved our abilities to introduce complex traits in the model developed in this chapter.

Generally, the adaptations were useful as a proof of concept, proving that one-hot encoding is possible using Excel, but the results obtained replicating existing algorithms point to caution when using the adapted methods for judging the performance of the original references.

4.5.5. The CRISPRnt-Plant model.

The results previously described point to a lack of accuracy of the models when predicting editing efficiencies in plants. Some of the features highlighted by these models aligned to the ones observed in the nucleotide analysis of the PP dataset. With this in mind, an alternative model was developed to predict sgRNA efficiencies in plant-based experiments, that included most of the features described in the published models.

Firstly, the PP dataset was not pre-filtered to separate high-efficiency and low-efficiency guides. This is a common process used in most of the models available online and has been used in the characterization of small interfering RNA (siRNA) with high efficiencies (Wang et al., 2009). Given the overrepresentation of low-efficiency sgRNA in the dataset, and the fact that all the guides used showed editing to some extent, it was decided to use the weighting system described in the Materials and Methods, focusing on the frequency of nucleotides in active sgRNA and the efficiency associated to them. In addition to that, models such as Doench et al. (2014) and Moreno-Mateos et al. (2015) filtered the features used to reduce the number of features used, simplifying the calculations, and to avoid the use of features with a non-significant effect on sgRNA efficiency. Instead of eliminating the features that would not significantly change the final scores, the CRISPRnt-Plant model provides final scores relative to the minimum and maximum possible score in an iterative manner. Every new sgRNA added to the model changing all the features in the model. However, the high number of features used for the single nucleotide features (120) meant that adding all possible dinucleotide features (480) would make the coding complicated. Given the use of all single nucleotide features, adding all dinucleotide features would have led to redundancies. This same rationale was behind excluding single nucleotide counts (e.g. the total number of A in the homology sequence).

As with other models, GC content was included as a feature. In the general analysis of the dataset, a non-linear relation was observed (Figure 4.2A), concurring with the bibliography. However, due to the high dispersion of the data, the optimal GC value described should be further refined. The GC content feature was designed to penalise deviations from the optimal, making necessary to provide a specific value instead of a range. Other models designed this feature to be based on two different scores given to GC contents higher or lower than a threshold, but the definition of these features is related to the method used to develop the model rather than manual data analysis.

The Tm of different parts of the gRNA were also included in the model. The general analysis of the dataset (Figure 4.2C, D, E, F), together with the findings of the published models, identified a linear correlation between sgRNA efficiency and Tm. The Tm was calculated using the modified Marmur estimation (Marmur and Doty, 1962). Tm scores were designed to be relative to a possible maximum Tm, applying a penalty to the sequences with lower Tm. In Doench et al. (2016), the Tm was calculated using the "*Biopython_staluc*" function (Le Novère, 2002; Cock et al., 2009). The use of more accurate Tm calculations may have benefitted the model. However, due to the limitations of the coding system used, the Marmur estimation was used.

Other features included in other tools, such as free energy of the sgRNA, availability of specific key nucleotides or location of the target site within the gene were excluded from the CRISPRnt-Plant model. Features affecting the structure of the sgRNA were not included due to the

limitations of the coding of choice, Excel. Features regarding coding sequence targeting were consciously left out, since the score the model would be focused on probability of editing rather than knock-out efficiency. Other features related to sgRNA expression have been also studied in the bibliography, such as the G at the 5' end required for U6 promoter usage (Wang et al., 2014), which have not been studied in this dataset.

The different types of features (i.e. nucleotide, GC content and Tm) were given a weighting value between 0 and 1. This meant that the contribution of each one of the types of feature was equal. To avoid specific features of relatively low importance providing most of the score, the relative contribution of each one of the types of features to the predicted score was calculated using the features described in Doench et al. (2016), as Table 4.3 shows. The model was developed in this manner as each one of the feature type was calculated independently.

Focusing on the nucleotide feature scores, meaning the values given to the appearance of a specific nucleotide on a specific position, it was observed that the scores differed from the general percentage analysis of the PP dataset (Figure 4.5). This shows how the model was successful in providing scores depending on the efficiency associated to the features without being affected by the number of guides showing a specific feature. On the other hand, a low-represented nucleotide found in a high-efficiency guide may imply that the feature has a high score, which may cause an outlier. This issue would be solved by increasing the number of sgRNA analysed.

When comparing the CRISPRnt-Plant model developed with the available and adapted tools, the statistical analysis performed shows a better correlation between the described and predicted efficiencies than any of the models used (Table 4.3). However, this could be due to the model being trained with the dataset it was analysing. Despite the positive results observed, the model was not capable of matching the predicted editing efficiencies 1 to 1. Further improvements on the dataset or the coding of the model in a language capable of deeper analysis would bring the CRISPRnt-Plant model closer to that.

4.5.6. Testing the CRISPRnt-Plant Model.

To test the robustness of the method, two different tests were performed.

The first re-sampling test was based on the Monte Carlo Simulation (Andrieu, 2004), which implies that random sampling of a population should infer the overall trend of the entire population. This approach has been used for centuries (Halton, 1970), being of common use in machine learning problems. In the random sampling analysis of this chapter, this trend was observed, with 9 out of 15 groups showing the positive slope observed when analysing the entire population (Figure 4.7A, B). The positive results confirm that the method for calculating the feature scores generally is capable of generating efficiency scores that are able to discriminate the low efficiency and high efficiency guides. The groups that did not show this trend could have been affected by the high density of low-efficiency guides within the group. The results point to the ability of the method used to obtain the feature scores to reflect the differences in the editing efficiency. However, 6 of the 15 groups did not show this significantly positive slope. As previously discussed, the CRISPRnt-Plant scoring method has a significantly positive slope, but it does not show the optimal 45^{o} slope. This might explain how different groups were able to keep the significantly positive slope and others did not. At the same time, the fact that some groups kept the significantly positive slope and others did not shows the sensitivity of the model to the 30mers used for the calculation of the features. As previously discussed, the model would

benefit from a higher number of samples analysed and a better distribution of the data covering several efficiencies.

A different analysis using the same groups was a modified bootstrap analysis. It is referred to as a modified bootstrap analysis since an actual bootstrap analysis would have implied random sampling of the 151 30mers of the PP dataset, being possible for them to be recorded more than once (Bootstrap with replacement, Diciccio and Romano, 1989; Gigli, 1996), and testing the possible outcomes obtained from different samplings of the same population. This particular type of test has been recommended for insufficient sample sizes (Adèr et al., 2008), which is one of the issues of concern of the analysis of the PP dataset. Given the low processing capacity of the Excel coding used, each one of the random samplings would have been done manually, being a task unrealistic for this project. However, a Python coded version would help run deeper statistical analysis based on random sampling in the future. The analysis performed in this chapter implied observing the standard deviation of the means of the feature scores calculated from each one of the groups from the random sampling test previously described (Figure 4.7C). It should be noted that the feature scores of the CRISPRnt-Plant model and the average scores of the repetitions are similar but not equal (Supplementary Table 14). The results obtained showed that generally the feature scores obtained had a standard deviation in the high-confidence tier established. However, some features were non-conclusive and others were in the low-confidence tier. This analysis would be the basis of specific features being discarded as their score may differ depending on the sampling. Similar approaches have been taken in other models such as Doench et al. (2014), where a Machine Learning approach was taken to reduce the amount of features and simplify the calculations of the model. However, due to this analysis being an adaptation of the bootstrap approach, the results were not included in the final CRISPRnt-Plant model. Once the Python coding is available and the bootstrap analysis is performed, this will be the basis of a simplification of the model discarding features that are not relevant for sgRNA efficiency prediction.

The final analysis was based on removing 10 random sgRNA from the calculation of the scores. The removal implied changes in the feature scores, which led to changes in the final performance of the model when analysed using the linear regression analysis. However, the general trend, slope and fitness to the linear regression of the data were still similar to the CRISPRnt-Plant model (Figure 4.8A). The results point to the robustness of the method. The second part of this test, the calculation of the scores of the removed individuals, did not show the expected results. The efficiencies of the sgRNA in the removed group ranged from 6.69 to 38%. This difference in efficiencies should have been enough to find a positive slope in the predicted efficiency analysis, but it was not found in the results (Figure 4.8B). These issues with predicted efficiency could be due to the feature scoring system not being able to provide a slope positive enough, which has been previously related to the high number of low efficiency guides distorting the analysis. The lack of predictive ability observed in this test implies the model will have to be further optimised.

4.6. Conclusions.

The results gathered in this chapter point to a lack of accuracy of on-target predictions in online available tools when comparing the efficiencies obtained in plants and the scores obtained from said tools. A new simple model that fits the data better has been developed based on the sequence features of the target site. Further refinements of this model, together with the addition of more data into it, could lead to the obtention of a plant-based CRISPR/Cas9 efficiency predictor with a better predicting capability than the ones currently available.

4.7. References.

Adèr, H. J., Hand, D. J., and Mellenbergh, G. J. (2008). *Advising on Research Methods: A Consultant's Companion*. Huizen, The Netherlands: Johannes van Kessel Publishing.

Andrieu, C. (2004). Monte Carlo methods for absolute beginners. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 3176, 113–145.

Bae, S., Kweon, J., Kim, H. S., and Kim, J. S. (2014a). Microhomology-based choice of Cas9 nuclease target sites. *Nature Methods* 11, 705–706.

Bae, S., Park, J., and Kim, J. S. (2014b). Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* 30, 1473–1475.

Bai, M., Yuan, J., Kuang, H., Gong, P., Li, S., Zhang, Z., Liu, B., Sun, J., Yang, M., Yang, L., Wang, D., Song, S., & Guan, Y. (2020). Generation of a multiplex mutagenesis population via pooled CRISPR-Cas9 in soya bean. Plant Biotechnology Journal 18, 721–731.

Brazelton, V. A., Zarecor, S., Wright, D. A., Wang, Y., Liu, J., Chen, K., Yang, B., and Lawrence-Dill, C. J. (2015). A quick guide to CRISPR sgRNA design tools. *GM crops & food* 6, 266–276.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., and Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing* 408, 189–215.

Chari, R., Mali, P., Moosburner, M., and Church, G. M. (2015). Unraveling CRISPR/Cas9 genome engineering parameters via a library-on-library approach. *Nature Methods* 12, 823–826.

Chavez, A., Scheiman, J., Vora, S., Pruitt, B. W., Tuttle, M., P R Iyer, E., et al. (2015). Highly efficient Cas9-mediated transcriptional programming. *Nature Methods* 12, 326–328.

Cheng, A. W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T. W., Rangarajan, S., Shivalila, C. S., Dadon, D. B., and Jaenisch, R. (2013). Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Research* 23, 1163–1171.

Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. a, Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., and De Hoon, M. J. L. (2009). Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422–1423.

Concordet, J. P., and Haeussler, M. (2018). CRISPOR: Intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Research* 46, W242–W245.

Cong, L. Le, Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., and Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* 339, 1766–1769.

Diciccio, T. J., and Romano, J. P. (1989). A Review of Bootstrap Confidence Intervals. *Journal of the Royal Statistical Society: Series B (Methodological)* 50, 338–354.

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J., and Root, D. E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR/Cas9. *Nature Biotechnology* 34, 184–191.

Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert,

B. L., Xavier, R. J., and Root, D. E. (2014). Rational design of highly active sgRNAs for CRISPR/Cas9-mediated gene inactivation. *Nature Biotechnology* 32, 1262–1267.

Fusi, N., Smith, I., Doench, J., and Listgarten, J. (2015). *In Silico* Predictive Modeling of CRISPR/Cas9 guide efficiency. *bioRxiv*, 021568.

Gagnon, J. A., Valen, E., Thyme, S. B., Huang, P., Ahkmetova, L., Pauli, A., Montague, T. G., Zimmerman, S., Richter, C., and Schier, A. F. (2014). Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS ONE* 9, e98186.

Gigli, A. (1996). Efficient bootstrap methods: A review. *Journal of the Italian Statistical Society* 5, 99–127.

Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E., Chen, Y., Whitehead, E. H., Guimaraes, C., Panning, B., Ploegh, H. L., Bassik, M. C., Qi, L. S., Kampmann, M., and Weissman, J. S. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* 159, 647–661.

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A., Lim, W. A., Weissman, J. S., and Qi, L. S. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154, 442–451.

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J. B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., Joly, J. S., and Concordet, J. P. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biology* 17, 148.

Halton, J. H. (1970). Retrospective and prospective survey of the Monte Carlo method. *SIAM Review* 12, 1–63.

Hanna, R. E., and Doench, J. G. (2020). Design and analysis of CRISPR/Cas experiments. *Nature Biotechnology* 38, 813–823.

Heigwer, F., Kerr, G., and Boutros, M. (2014). E-CRISP: Fast CRISPR target site identification. *Nature Methods* 11, 122–123.

Ho, T. T., Zhou, N., Huang, J., Koirala, P., Xu, M., Fung, R., Wu, F., and Mo, Y. Y. (2015). Targeting non-coding RNAs with the CRISPR/Cas9 system in human cell lines. *Nucleic Acids Research* 43, e17.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G., and Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology* 31, 827–832.

Jacobs, T. B., Zhang, N., Patel, D., & Martin, G. B. (2017). Generation of a collection of mutant tomato lines using pooled CRISPR libraries. Plant Physiology 174, 2023–2037.

Koike-Yusa, H., Li, Y., Tan, E. P., Velasco-Herrera, M. D. C., and Yusa, K. (2014). Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nature Biotechnology* 32, 267–273.

Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O. O., Barcena, C., Hsu, P. D., Habib, N., Gootenberg, J. S., Nishimasu, H., Nureki, O., and Zhang, F. (2015). Genome-scale transcriptional activation by an engineered CRISPR/Cas9 complex. *Nature* 517, 583–588.

Le Novère, N. (2002). MELTING, computing the melting temperature of nucleic acid duplex. *Bioinformatics* 17, 1226–1227.

Lei, Y., Lu, L., Liu, H. Y., Li, S., Xing, F., and Chen, L. L. (2014). CRISPR-P: A web tool for synthetic single-guide RNA design of CRISPR-system in plants. *Molecular Plant* 7, 1494–1496.

Liu, H., Ding, Y., Zhou, Y., Jin, W., Xie, K., and Chen, L. L. (2017a). CRISPR-P 2.0: An Improved CRISPR/Cas9 Tool for Genome Editing in Plants. *Molecular Plant* 10, 530–532.

Liu, X., Homma, A., Sayadi, J., Yang, S., Ohashi, J., and Takumi, T. (2016). Sequence features associated with the cleavage efficiency of CRISPR/Cas9 system. *Scientific reports* 6, 19675.

Liu, X., Wu, S., Xu, J., Sui, C., and Wei, J. (2017b). Application of CRISPR/Cas9 in plant biology. *Acta Pharmaceutica Sinica B* 7, 292–302.

Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., and Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology* 6, 26.

Lorenz, R., Hofacker, I. L., and Stadler, P. F. (2016). RNA folding with hard and soft constraints. *Algorithms for Molecular Biology* 11, 1–13.

Lu, Y., Ye, X., Guo, R., Huang, J., Wang, W., Tang, J., Tan, L., Zhu, J. kang, Chu, C., & Qian, Y. (2017). Genome-wide targeted mutagenesis in rice using the CRISPR/Cas9 system. Molecular Plant 10, 1242–1245.

Mandal, P. K., Ferreira, L. M. R., Collins, R., Meissner, T. B., Boutwell, C. L., Friesen, M., Vrbanac, V., Garrison, B. S., Stortchevoi, A., Bryder, D., Musunuru, K., Brand, H., Tager, A. M., Allen, T. M., Talkowski, M. E., Rossi, D. J., and Cowan, C. A. (2014). Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9. *Cell Stem Cell* 15, 643–652.

Marmur, J., and Doty, P. (1962). Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. *Journal of Molecular Biology* 5, 109–118.

Meng, X., Yu, H., Zhang, Y., Zhuang, F., Song, X., Gao, S., Gao, C., & Li, J. (2017). Construction of a genome-wide mutant library in rice using CRISPR/Cas9. Molecular Plant 10, 1238–1241.

Moreno-Mateos, M. A., Vejnar, C. E., Beaudoin, J. D., Fernandez, J. P., Mis, E. K., Khokha, M. K., and Giraldez, A. J. (2015). CRISPRscan: Designing highly efficient sgRNAs for CRISPR/Cas9 targeting in vivo. *Nature Methods* 12, 982–988.

Naim, F., Shand, K., Hayashi, S., O'Brien, M., McGree, J., Johnson, A. A. T., Dugdale, B., and Waterhouse, P. M. (2020). Are the current gRNA ranking prediction algorithms useful for genome editing in plants? *PLoS ONE* 15, e0227994.

Park, J., Bae, S., and Kim, J. S. (2015). Cas-Designer: A web-based tool for choice of CRISPR/Cas9 target sites. *Bioinformatics* 31, 4014–4016.

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P. and Lim, W. A., (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173–1183.

Sanjana, N. E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. *Nature Methods* 11, 783–784.

Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelsen, T. S., Heckl, D., Ebert, B. L., Root, D. E., Doench, J. G., and Zhang, F. (2014). Genome-scale CRISPR/Cas9 knockout screening in human cells. *Science* 343, 84–87.

Shen, M. W., Arbab, M., Hsu, J. Y., Worstell, D., Culbertson, S. J., Krabbe, O., Cassa, C. A., Liu, D. R., Gifford, D. K., and Sherwood, R. I. (2018). Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* 563, 646–651.

Singh, R., Kuscu, C., Quinlan, A., Qi, Y., and Adli, M. (2015). Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. *Nucleic Acids Research* 43, e118.

Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J., and Mateo, J. L. (2015). CCTop: An intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PLoS ONE* 10, e0124633.

Sun, J., Liu, H., Liu, J., Cheng, S., Peng, Y., Zhang, Q., Yan, J., Liu, H. J., and Chen, L. L. (2019). CRISPR-Local: A local single-guide RNA (sgRNA) design tool for non-reference plant genomes. *Bioinformatics* 35, 2501–2503.

Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S., and Vale, R. D. (2014). A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* 159, 635–646.

Tennakoon, C., Purbojati, R. W., and Sung, W. K. (2012). BatMis: A fast algorithm for k-mismatch mapping. *Bioinformatics* 28, 2122–2128.

Varshney, G. K., Pei, W., Lafave, M. C., Idol, J., Xu, L., Gallardo, V., Carrington, B., Bishop, K., Jones, M., Li, M., Harper, U., Huang, S. C., Prakash, A., Chen, W., Sood, R., Ledin, J., and Burgess, S. M. (2015). High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. *Genome Research* 25, 1030–1042.

Wang, T., Wei, J. J., Sabatini, D. M., and Lander, E. S. (2014). Genetic screens in human cells using the CRISPR/Cas9 system. *Science* 343, 80–84.

Wang, X., Wang, X., Varma, R. K., Beauchamp, L., Magdaleno, S., and Sendera, T. J. (2009). Selection of hyperfunctional siRNAs with improved potency and specificity. *Nucleic Acids Research* 37, e152.

Wong, N., Liu, W., and Wang, X. (2015). WU-CRISPR: Characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome Biology* 16, 218.

Wu, X., Kriz, A. J., and Sharp, P. A. (2014). Target specificity of the CRISPR/Cas9 system. *Quantitative Biology* 2, 59–70.

Xu, H., Xiao, T., Chen, C. H., Li, W., Meyer, C. A., Wu, Q., Wu, D., Cong, L., Zhang, F., Liu, J. S., Brown, M., and Liu, X. S. (2015). Sequence determinants of improved CRISPR sgRNA design. *Genome Research* 25, 1147–1157.

Zalatan, J. G., Lee, M. E., Almeida, R., Gilbert, L. A., Whitehead, E. H., La Russa, M., Tsai, J. C., Weissman, J. S., Dueber, J. E., Qi, L. S., and Lim, W. A. (2015). Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* 160, 339–350.

Zhou, Y., Zhu, S., Cai, C., Yuan, P., Li, C., Huang, Y., and Wei, W. (2014). High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* 509, 487–491.

Zhu, L. J., Holmes, B. R., Aronin, N., and Brodsky, M. H. (2014). CRISPRseek: A Bioconductor package to identify target-specific guide RNAs for CRISPR/Cas9 genome-editing systems. *PLoS ONE* 9, e108424.

Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic-net. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 67, 301–320.

Chapter 5. Traditional crop development and Genome Editing.

5.1. Introduction and regulatory background.

United Nations estimates the world population will reach almost 10 billion people by 2050. At the same time, the World Atlas of Desertification developed by the European Commission estimates that over 75% of the land was degraded by 2019 and that 90% could be degraded by 2050. This puts agriculture in the spotlight, given the importance of food crops in the food energy intake worldwide. This points to agricultural and crop development as one of the possible answers to these future needs and requirements. However, a balance between new techniques and varieties improving crop efficiency and their safety for consumption and the environment has to be reached.

With the description of recombinant DNA techniques, meaning plasmid construction for their introduction into foreign systems, several concerns about their potential uses started rising (Berg et al., 1974). An approach based on proportionality of the safety measures based on risk assessments was proposed in the Asilomar Conference in 1975, together with the conclusion that these techniques did not represent a risk inherently, but the potential threat should be assessed in a case by case basis (Berg et al., 1975). This led to the first regulations and guidelines for the use in the laboratory, such as the ones in the US in 1976 (reviewed in Talbot, 1983) or the UK in 1978 (Genetic manipulation: New guidelines for UK, 1978).

With the description of the first transgenic plants in 1983 (Herrera-Estrella et al., 1983), the regulations had to be modified since their use would not only be restricted to the laboratory, starting with the Coordinated Framework for Regulation of Biotechnology in the U.S.A., published in 1986, or the European regulations, regulating deliberate release (EU Council Directive 90/220/EEC, 1990), contained use of genetically modified microorganisms (EU Council 90/219/EEC Directive, 1990) or the use of pesticides related to said transgenics (EU Council Directive 91/414/EEC, 1991). With the foundation of the European Union in the Maastricht Treaty in 1992, the Precautionary Principle, included in Article 191(2) of the Treaty on the Functioning of the EU, was established. This principle is “relevant only in the event of potential risk, even if this risk cannot be fully demonstrated or quantified or its effects determined because of the insufficiency or inclusive nature of the scientific data” (EU Commission Communication, 2000). This publication also noted that the application of the precautionary principle should always be proportional, non-discriminatory, comparable to other measures in place, based on a cost-benefit analysis, subject to change, and facilitate future risk assessments based on scientific evidence. To this day, this principle has shaped the European Regulations referring to Genetically Modified Organisms (GMO). In 1997, the European Union released the EU Parliament and Council Regulation (EC) 258/97 concerning novel foods and novel food ingredients, stated that “novel food ingredients are subject to a single safety assessment [..], whereas in the case of novel foods and novel food ingredients which are substantially equivalent to existing foods or food ingredients a simplified procedure should be provided”. In 1998, the first GMO, MON810 maize, was approved for its cultivation in Europe, although an effective moratorium was put in place restricting the growth and market placement of this GMO and others.

In the year 2000, the Cartagena Protocol on Biosafety was released, regarding the “safe transfer, handling and use of living modified organisms”, meaning “any living organism that possesses a novel combination of genetic material obtained through the use of modern biotechnology”, which at this point included recombinant DNA technologies and cell fusion. This was initially adopted by the European Union and its members, but the acceptance of this definition was concluded in 2002 with the EU Council Decision 2002/628/EC. In 2001, the EU Parliament and

Council Directive 2001/18 Article 2 defined an organism as “any biological entity capable of replication or of transferring genetic material” and GMO as “an organism, with the exception of human beings, in which genetic material has been altered in a way that does not occur naturally by mating or natural recombination”. In this same directive, Annex IA when enumerating the techniques whose products would be considered GMO and the ones that would not be counted as GMO. This annex classifies all techniques using recombinant DNA techniques as GMO, including viral, bacterial or physical introduction of these “new combinations of genetic material”. It also defines that techniques such as *in vitro* fertilisation, conjugation, transduction, transformation and polyploidy induction are not considered GMO. In Annex IB it states that organisms obtained using mutagenesis are considered GMO but are exempt from the regulation. Organisms obtained from cell fusion are also exempt, but only if the product obtained could be obtained using traditional breeding methods. Part B of this Directive refers to the deliberate release of GMO for any purpose that does not imply placing on the market, adding the standard authorisation procedure, while Part C includes the procedure for placing in the market. In 2002, additional notes were added to this directive regarding guidance notes for the Environmental Risk Assessment for the deliberate release of GMO in the EU Commission Decision 2002/623/EC.

Based on this definition, several regulations have been released, including the EU Parliament and Council Regulation 178/2002 regarding General Food, which includes the foundation of the European Food Safety Authority (EFSA) and the centralization of the risk assessments, EU Parliament and Council Regulation EC 1946/2003, focused on the movements across borders of GMOs, EU Parliament and Council Regulation EC 1829/2003, focused on food and feed authorisation, EU Commission Regulation EC 1983/2003, focused on labelling and traceability, and EU Commission Regulation EC 641/2004, which expanded on the details about food and feed authorisation. In 2009, the EU Parliament and Council Directive 2009/41/EC established the guidelines for the authorisation of contained uses of GMO. In 2015, the EU Parliament and Council Directive 2015/412 gave the member states the power of restricting or prohibiting the cultivation of GMO within their own territory.

Through the years, 19 member states have banned the cultivation of GMOs and 62 regions within the EU have declared themselves GMO-free. Regarding the approval of GMOs for cultivation, only two varieties have been approved, including the MON810 maize and the Amflora Potato (Halford, 2019), the latter being the only one approved since 1998, with nearly all other authorisations withdrawn. However, the authorisation of the Amflora Potato was reversed by the EC and it has not been commercialised in the EU.

In 2018, upon revision of the use of genome editing techniques, the Court of Justice of the European Union stated that edited organisms were considered as GMO and therefore had to follow the same regulations as others under the same classification in the Judgement of the Case C-528/16 (CJEU, 2018). This would include any form of genome editing, excluding this type of mutagenesis from Annex B of the 2001/18 Directive, under the rationale of other random mutagenesis techniques having a history of safe use that genome editing techniques do not have. This position places the EU at odds with many other countries of the world that have adapted their own laws to regulate products of gene editing differently to genetically modified organisms (Schmidt et al. 2020a).

Argentina, for example, has passed new legislation to deregulate plants that although they have been obtained using New Plant Breeding Techniques, which include genome editing, as long as the resulting DNA sequence does not have a novel combination (Friedrichs et al., 2019). Other adjoining countries such as Brazil, Chile, Paraguay, and Uruguay signed a joint declaration to aim

for a similar legislation across these countries (Norero, 2018; Benítez Candia et al., 2020). Countries such as Ecuador, who banned the cultivation of GM crops, are also trying to harmonise the regulations across South America (Gatica-Arias, 2020). However, other South American Countries such as Venezuela and Peru have banned the commercial cultivation of GM crops (Turnbull et al., 2021). Chile, on the other hand, decided to focus on a case-by-case system, deciding on the regulation of specific products or varieties depending on the presence of foreign genetic material (Sánchez, 2020).

In case of Canada, the already existing regulation is based on the novelty of the trait obtained, in this case specifically regarding the Canadian environment and food and feed supply, with organisms with novel traits requiring a risk assessment prior to their commercialization and release (Smyth et al., 2017). This would mean that crops obtained using genome editing techniques would not be subject to said risk assessment if the combination of DNA is not novel. The regulation in the United States is based on a coordinated framework between the United States Department of Agriculture (USDA), specifically the Animal and Plant Health Inspection Service (APHIS) and the Food and Drug Administration (FDA). APHIS will focus the risk assessment on the risk of the genetically modified crop of becoming a pest. Depending on the risk assessment, the crop will either be regulated, meaning it will not have authorisation for cultivation and transport, or non-regulated. If the objective of the crop is to be used as food or feed, the FDA will have to produce a risk assessment of the product in question. With this regulation, several genetically modified crops have been authorised in the US, including a browning-resistant mushroom, herbicide-tolerant canola (Canola™), soybean oil (Calyno™) and waxy corn (Waltz, 2016; Lassoued et al., 2019).

A different approach was taken in countries such as Australia and New Zealand, where the regulations have been modified to not regulate plants obtained from SDN-1 type events as GMO, whereas SDN-2 and SDN-3 would be regulated as GMO due to the DNA insertion required (Friedrichs et al., 2019). Similarly, Japan has expressed an interest in reforming the regulations in order to assess differently different outcomes of genome editing techniques (Tsuda et al., 2019). In contrast, Russia decided to ban the cultivation of GM plants (Turnbull et al., 2021).

At the time of writing, in the UK, the Department for Environment, Food & Rural Affairs (DEFRA) is running a consultation to decide if organisms obtained from genome editing should continue be regulated as GMO even when the outcomes produced could be obtained using traditional breeding. However, in this consultation they do not define the meaning or scope of traditional breeding.

5.2. Objectives of the Chapter.

The emergence of gene editing as a breeding tool is challenging existing frameworks for food, feed and environmental safety around the world. The UK appears to be considering changing its laws to remove genome editing techniques from the definition of a GMO if the outcomes could be found naturally or could be achieved using traditional breeding techniques. However, the spectrum of these genetic changes has not been specified. In this chapter, different breeding mechanisms have been reviewed in an attempt to define the outcomes possible using current plant breeding methods and to compare and contrast these with the genetic changes known to result from gene editing.

Traditional breeding utilises genetic changes found in breeding lines, in wild germplasm or artificially induced. These modifications are described below under four overall categories

defined by the scale and type of the genetic change. Firstly, changes at the gene/DNA sequence level, including point mutations and small gene-level lesions such as insertions, deletions, inversions and duplications will be reviewed. Secondly, gross changes in the gene order or structure of individual chromosomes including major structural mutations and induced translocations. Thirdly, modifications of whole chromosome number including aneuploidy, polyploidy, haploidy. Finally, other methods for achieving inter-species hybridization by overcoming breeding barriers will be discussed.

Lastly, a comparison between the possible outcomes of genome editing techniques and the outcomes described from traditional breeding techniques will be discussed, highlighting where gene editing can produce genetic changes that are equivalent to traditional breeding.

5.3. Review on traditional breeding outcomes.

5.3.1. Genetic changes at the gene/DNA sequence level.

Mutations appear naturally in the genome at different rates in different species (Hodgkinson and Eyre-Walker, 2011). Arabidopsis, for example, shows a spontaneous mutation rate of 7×10^{-9} base substitutions per site per generation (Ossowski et al., 2010). On the other hand, rice has a reported mutation rate of 3.2×10^{-9} (Yang et al., 2015). The majority of these mutations have been observed to be short, being 99% of them shorter than 23 base pairs, including base pair substitutions and insertion-deletion events (The 3000 rice genomes Project, 2014). This mutation rate can be significantly increased using mutagens, which affect the DNA in different manners (Reviewed in Griffiths et al., 2000; Roychowdhury and Tah, 2013; Oladosu et al., 2016). The induction of genetic variability and appearance of new traits has been, and will continue to be, key for crop development (Novak and Brunner, 1992). These mutations arise spontaneously due to a range of causes including chemical instability of the bases, copying errors during DNA replication and miss-repair of DSB from naturally occurring mutagens. Meiosis has also been linked to mutagenic events. Recombination during meiotic crossovers has been linked with DSB induction by SPO11 proteins (Keeney et al., 1997; De Massy, 2013; Robert et al., 2016; Vrielynck et al., 2016). Some of the proteins involved in DSB repair described in the general introduction have also been associated with meiotic recombination, such as the MRE11 and RAD50 (Neale et al., 2005; Keeney and Neale, 2006; Uanschou et al., 2007). Usually, the DSB will be repaired by mechanisms similar to HR pathways, including resection of the DSB and invasion of a homologous strand (Lawrence et al., 2017). In Arabidopsis, these events occur between 100 and 200 times per meiotic nucleus, but only around 10 of them are solved by crossover mechanisms (Copenhaver et al., 1998; Giraut et al., 2011; Salomé et al., 2012; Wijnker et al., 2013), implying that the rest will be solved by conventional DSB repair mechanisms, which are mutagenic and will contribute to the background level of natural point and smaller mutations in a cell. Various studies have described the mutagenicity of this process in various organisms (Magni and Von Borstel, 1962; Perry and Ashworth, 1999; Lercher and Hurst, 2002; Pratto et al., 2014; Arbeithuber et al., 2015; Rattray et al., 2015).

One of the mechanisms of mutation induction is the addition of base analogs. The action of these mutagens will imply the inclusion of these analogs in the DNA sequence followed by the mispairing of these analogs with a base different from the original during replication. Examples of these are 5-bromouracil, an analog of thymine which when ionized pairs with guanine (Jacobs, 1969), or 2-amino-purine, analog of adenine which pairs with cytosine when protonated (Pitsikas et al., 2004).

A similar effect can be obtained using agents that modify an existent base and lead to a specific mispairing, such as alkylating agents. Ethyl-methanesulfonate (EMS) is one of the most common examples. This agent functions by adding an ethyl group to the nitrogenated bases (Kim et al., 2006). Although it is able to modify all 4 nucleotides, it has been observed a preference for guanine alkylation, making it pair with thymine instead of cytosine. This agent has a reported mutation rate of 1 mutation per 3.2 Mb in tomato (Minoia et al., 2010). Other examples of these agents are nitrosoguanidine (MNNG), with effects similar to EMS in both guanine and adenine residues (Goering and Pattee, 1971; Silverman et al., 2001; Wyatt and Pittman, 2006), or methyl-methanesulfonate (MMS), which induces the addition of a methyl group (Lundin et al., 2012). Another example of agents modifying directly the nitrogenated bases is the case of intercalating agents, such as proflavine, acridine orange and other chemicals referred to as ICR compounds. These usually cause single nucleotide indels.

Another pathway for mutagenesis is the induction of base damage. This mechanism blocks any type of base pairing, in contrast with the previous examples which cause mispairing. The lack of pairing leads to a stoppage in replication. The damage will lead to the inclusion of incorrect bases in the sequence. This effect has been described in plants treated with UV light (Stadler, 1928a, 1928b), associated mainly with CT transitions, among other effects such as transversions, frameshifts, duplications and deletions. UV radiation, another type of ionising radiation, has been observed to induce mutation rates 56 times higher than the usual somatic mutation rate (Kovalchuk et al., 2000).

Ionising radiation will also induce DNA damage. This damage has been associated with Reactive Oxygen Species (ROS), which will induce base damage and even strand breaks due to damage in the N-glycosidic bond between the nitrogenated base and the riboses. X-rays are an example of ionising radiation and its effects as a mutagen have been observed in plants (Stadler, 1930). Another example would be the use of γ-rays, which cause these short mutations in an frequency around between 7.5×10^{-9} and 9.8×10^{-6} (Li et al., 2016b).

Overall, it has been observed that the agents previously described are able to increase mutation rates 56 times in case of UV, 3 times in case of X-rays and 2 times in case of MMS (Kovalchuk et al., 2000), increasing the genetic variation in plant species and leading to a faster appearance of new traits.

The use of induced random mutagenesis has been prominent in crop development, with examples of varieties obtained using these methods being cultivated all over the world (Kharkwal and Shu, 2009). In case of rice, examples such as the “Zhefu 802” variety, obtained by irradiation using γ-rays of “Simei No.2”, or “Shwewartun”, obtained by irradiation of the “IR5” variety, have been widely sown in Asia (Ahloowalia et al., 2004). In barley, the “Diamant” variety, obtained by γ-ray irradiation of the “Valticky” variety, is one of the most important crops in Europe (Kharkwal and Shu, 2009). Another prominent barley variety, “Golden Promise”, has also been obtained from γ-ray irradiation of the “Maythorpe” barley variety (Forster, 2001). In North America, beans, with varieties such as the “Salinac” variety obtained using X rays on the “Navy” variety (Ahloowalia et al., 2004), lineseed, using EMS in the “Glenelg” variety to obtain “Zero”, wheat, using radio-induced mutagenesis on the “Salamanca” variety to obtain “Centauro” and “Bajio Plus” varieties, and soybean, using radio-induced mutagenesis on “Suaqui 86” to obtain “Hector” and “Esperanza”, are examples of well-established crops obtained using random mutagenesis (Green and Dribnenki, 1996).

Mutation breeding using external agents described above are well suited to annual crops that reproduce sexually, as well as for perennial, vegetatively-propagated crops. However, other approaches to generate novel, random genetic variation have been used. For example, prolonged periods of tissue culture are known to induce mutations. So-called somaclonal variations are thought to be newly induced mutations arising from the culture process. The causes of genetic instability during tissue culture are not well characterised but it is thought the cellular stresses experience by specialised cells undergoing genetic reprogramming during dedifferentiating contribute to this (Joyce et al., 2003). Although it cannot be excluded that some are actually pre-existing mutations already present in the explant tissue. Many somaclones have been released as commercial cultivars or varieties (Jain, 2001).

In addition to the induction of random variation, the selection of variation related to a specific trait is necessary in plant breeding. Marker-assisted breeding, which implies the identification of specific molecular markers and relating them to a trait of interest, has been a prominent approach in plant breeding (Collard and Mackill, 2008; Bhat et al., 2016). This has been attained by marker assisted selection, which will focus on finding DNA markers related to a trait and directly selecting the individuals carrying said set of markers (Collard and Mackill, 2008). This has led to the identification of several Quantitative Trait Loci (QTL) in several species (Camargo et al., 2018; Khan et al., 2021; Ashraf et al., 2012). However, with more complex traits, its potential does not surpass the one of conventional phenotyping (Zhao et al., 2014). On the other hand, Genomic Selection will focus on predictions of the breeding value of individuals within a population (Meuwissen et al., 2001). This will mean that from a combination of genetic markers and phenotype variations, a value inferring the potential of a new individual of having advantageous phenotypes within a population is obtained (Poland et al., 2012), helping the selection of possible new lines. This approach has been successfully applied in different crops such as rice (Spindel et al., 2015; Grenier et al., 2015), wheat (Heslot et al., 2013; Michel et al., 2016) or maize (Crossa et al., 2013; dos Santos et al., 2016), among others (Bhat et al., 2016). The combination of the different techniques previously described, including random mutations accelerated by different techniques and the identification of loci of interest to help the selection of specimens carrying advantageous traits has been a key tool in crop development.

5.3.2. Gross changes in the gene order or structure of individual chromosomes.

Many natural and artificial processes during traditional breeding lead to gross changes in chromosome structure or gene order. In addition to the specific mutations generated by DSB repair during meiosis, as previously described, the crossover events themselves result in natural translocations between homologous chromosomes that recombine parental alleles and forms the basis of most breeding activities. However, gross chromosomal translocations can also be induced artificially. The restricted genetic background of wheat has been a major barrier in breeding efforts. Intergenic translocation lines provide unique variation and introduce key traits that are missing in the species gene pool. Among the *Triticaceae* family, several chromosomes of different species have been described to contain beneficial traits such as biotic stress resistance (Zeller and Hsam, 1983) or yield potential (Rajaram et al., 1983). These species include *Aegilops*, *Agropyron*, *Triticum* and *Secale* (Friebe et al., 1996). However, some of these chromosomes carry other genes that are not advantageous. In order to reduce the number of genes transferred, Sears (Sears, 1956, 1993) proposed the translocation of chromosomic elements between *Aegilops* and wheat by irradiation of the pollen prior to crossing. This approach has been used for several translocations in wheat with rye, being the latter the most common source of translocations (Rabinovich, 1998). An example of these was the 1BL/1RS translocation, which has been incorporated into a significant number of commercial wheat

varieties. The short arm of rye chromosome 1 transferred much needed resistance to stem, stripe and leaf rusts, powdery mildew and increased yield (Zeller, 1973). The same chromosome fragment was also utilised in the 1AL/1RS translocation of wheat, providing stem rust, powdery mildew and greenbug resistance (Zeller and Hsam, 1983). Other rye chromosomes like 3R and 7R have also been used in wheat translocations but were of less commercial importance (Jung and Seo, 2014). Oat (*Avena sativa*) has also benefitted from this approach when receiving a translocation from *Avena barbata* (Aung and Thomas, 1978). In addition to the translocation *per se*, different modifications of the rye genome have been induced prior to the crossing with the irradiated pollen (Ren et al., 2012), focusing on an increase in variability of the chromosomic arms of interest to obtain new genetic material in the original species that can be transferred to the destination species.

5.3.3. Ploidy modifications.

In addition to utilising natural and induced genetic variation at the gene and chromosome level, traditional breeding also exploits changes in ploidy.

5.3.3.1 Aneuploids.

The term aneuploidy was first introduced in Täckholm (1922). With time, different types of aneuploidies have been defined, referring commonly to changes in chromosome number, such as additional or fewer than normal copies of specific chromosomes (Dyer et al., 1970; Khush, 1973; Sharma, 1990). The occurrence of aneuploidies has been linked to alterations in the chromosomic distribution during cell division, which in plants only occurs in meristematic tissue (mitosis) and reproductive tissue (meiosis). During the cellular cycle, the cell will duplicate its DNA and prepare for cell division. Prior to cell division, the chromosomes will condense, align in the metaphasic plate and be divided in two equal sets of chromosomes. The role of spindles, tubulin and microtubules in this process has been reported in depth (Deysson, 1968, 1975; Sharma, 1971; Bajer and Mole-Bajer, 1972; Bond and Chandley, 1983; Artvinli, 1987; Mazia, 1987). Alterations in the process of chromosome distribution will lead to aneuploidies in the daughter cells. The term vagrancy has been used for defining these alterations during mitosis, including metaphasic arrest, chromosome scattering during anaphase, asynchrony of the sets of chromosomes reaching the poles of the cell (forwards/laggards), or nondisjunction of the centromere leading to diplochromosomy, disturbed polarity due to split-spindles leading to aberrant numbers of chromosomes in the daughter cells, or binucleation processes. In case of alterations during meiosis, noncongression, meaning lack of alignment of bivalents during metaphase, or alterations in disjunction of bivalents, have been observed (Sharma, 1990).

Several processes have been linked to this spindle disfunction, such as the inhibition of the cell respiratory chain, causing a metaphasic arrest (Kihlman, 1966; Mann and Storey, 1966; Deysson, 1968; Sharma and Sharma, 1980; Sahu et al., 1981a, 1981b; Rost, 1984; Rost and Morrison, 1984). This has been referred to pseudo-aneugenicity. Direct alterations in spindle function (Paget and Walpole, 1958; Panda and Sharma, 1980) have been referred to as direct aneugenicity. When the distribution of chromosomes has not been affected but disruptions in the cytokinesis process have led to binucleation (Fourcade et al., 1963; Deysson, 1968; Sharma et al., 1978; de la Peña et al., 1981; Nandakumar et al., 1984), the term indirect aneugenicity has been used.

Aneuploidy has been found to occur naturally in several species (Table 5.1). Different factors have also been observed to induce these aneuploidies, such as agricultural chemicals, drugs, natural and industrial products, together with others such as soil status, pollution in water,

pesticides, plant toxins and pathogens, seed age and other environmental factors (Sharma, 1990; Fuchs et al., 2018). Aneuploidies have been linked with the appearance of natural polyploidies as a precursor (Xiong et al., 2011; Chester et al., 2012; Zhang et al., 2013; Wu et al., 2018) and have been observed to induce new variability and phenotypes in different plants (Torres et al., 2008; Veitia et al., 2008), both as an aneuploid (Sheltzer and Amon, 2011) or after recovering its euploidy (Henry et al., 2010; Gao et al., 2016). Aneuploidy has been used to induce variations in plants of commercial relevance such as wheat (Kimber, 1976) and maize (Makarevitch and Harris, 2010).

Table 5.1. Some plant genera with spontaneous aneuploidy (Adapted from Sharma, 1990).

Aegilops	*Avena*	*Collinsia*	*Humulus*	*Nicotiana*	*Pygaera*	*Triticum*
Agropyron	*Beta*	*Corchorus*	*Hyacinthus*	*Oenothera*	*Ribes*	*Tulipa*
Antirrhinum	*Canna*	*Crepis*	*Hyoscyamus*	*Oryza*	*Rosa*	*Verbena*
Haynaldia	*Capsicum*	*Dactylis*	*Lotus*	*Pennisetum*	*Secale*	*Zea*
Pharbitis	*Chrysanthemum*	*Datura*	*Lycopersicon*	*Petunia*	*Solanum*	
Arabidopsis	*Clarkia*	*Gossypium*	*Matthiola*	*Pisum*	*Sorghum*	
Arachis	*Clivia*	*Hordeum*	*Medicago*	*Prunus*	*Spinacea*	

5.3.3.2. Doubled haploids.

The focus of a breeding program is not always to induce additional variation. Techniques such as the use of double haploid lines have been utilised to obtain homozygous plants faster than using traditional breeding methods (Chen et al., 2011).

In natural conditions, plants can reproduce using an uncommon reproductive mechanism which implies that the genetic material of the embryo will only be provided by one of the parents (Schwander and Oldroyd, 2016). In nature, this can occur either by fusion of the reproductive cells followed by the elimination of one of the genomes (Rieger et al., 2012) or by fertilization of a non-nucleate egg (Pigneur et al., 2012). When the genetic material comes from pollen, the term androgenesis is used, whereas when the genetic material comes from the ovary, gynogenesis or parthenogenesis has been used (Kasha and Maluszynski, 2003; Schwander and Oldroyd, 2016). Generally, this implies that the offspring derived from this process will be haploid, although the common individual is usually diploid (Seguí-Simarro, 2010). In those cases, the genome of the haploid offspring will be doubled to form a diploid specimen, forming a completely homozygous plant (Kermicle, 1974). Androgenesis in plants has also been related to hybridization, given meiotic impairment and production of non-nucleate gametes (Waldman, 2008; Schwander and Oldroyd, 2016).

These events have been observed in different plants in a natural manner, such as *Cupressus dupreziana*, an African conifer (Pichot and El Maâtaoui, 2000; Pichot et al., 2000, 2008; Abdoun and Beddiaf, 2002), *Cupressus sempervirens* (Pichot et al., 2001, 2008; Burt and Trivers, 2006), maize (Kermicle, 1974) and tobacco (Goodsell, 1961; Burk, 1962; Chase, 1969). In case of hybrids, cases in pepper (*Capsicum frutescens,* Campos and Morgan, 1958), Chick pea (*Cicer arietinum*, Mallikarjuna et al., 2005), oilseed rape (*Brassica napus*, Chen and Heneen, 1989), tobacco (Clausen and Lammerts, 1929; Kostoff, 1929; Kehr, 1951; Burk, 1962; Seguí-Simarro, 2010), *Poa arachnifera* (Seguí-Simarro, 2010), *Solanum verrucosum* (Seguí-Simarro, 2010; Seguí-Simarro et al., 2011) and maize hybrids (Goodsell, 1961; Chase, 1969; Kermicle, 1974).

Different techniques have been used for double haploid production. Chase (1952) described a method based on parthenogenesis in maize. Parthenogenesis frequencies have been increased by the use of pollen irradiation before pollination, the use of seeds with twin embryos, sparse pollination, wide hybridizations or alien cytoplasm inductions in hybrids (Kasha and Maluszynski, 2003). Also based on parthenogenesis, crosses between wild diploid and cultivated tetraploid potato formed dihaploid potato through parthenogenesis (Chase, 1963). Wide hybridizations can lead to the loss of one of the sets of chromosomes. Kasha and Kao (1970) described a method of haploid production in barley using this strategy. Laurie and Bennett (1988) showed similar findings using maize, sorghum or millet to pollinize wheat. Barclay (1975) reported double haploids obtained from chromosome elimination after hybridization in wheat as well. Choo et al. (2011) reviewed different hybridization-based techniques for double haploid development in barley.

Other techniques for double haploid obtention imply the treatment and culture of specific tissues, namely reproductive tissues. Ovule culture has been prominent in sugar beet and onion, whereas semigamy occurs naturally in cotton (Zhang et al., 1999). A different strategy is the use of temperature-shock treatment in anthers (Seguí-Simarro, 2010), used in *Crepis tectorum* (Gerassimova, 1936) and *Antirrhinum majus* (Ehrensberger, 1948). *In vitro* anther culture has also been utilized as a way to artificially induce double haploids, a process implying the treatment with temperature shock and *in vitro* culture of the anthers, which often diploidize autonomously (Guha and Maheshwari, 1964, 1966; Wang et al., 2000). This process has been described in barley and wheat (Weyen, 2008; Devaux and Kasha, 2008), including anther culture (Clapham, 1973; Ouyang et al., 1973; Kao et al., 1991), pollen culture (Datta and Wenzel, 1987) and microspore culture (Datta and Wenzel, 1987; Hunter, 1988; Kao et al., 1991), among other species such as cucurbit plants (Dong et al., 2016).

A different approach has been proposed for the induction of haploid offspring by the modification of the centromere-specific histone 3 (CENH3, Ravi and Chan, 2010). This histone variant substitutes the histone H3 in centromeric locations (Jiang et al., 2003; McKinley et al., 2016). This role in the histone in the centromere implies that alterations in its function may cause disruptions in chromosome distribution (Allshire et al., 2008). This occurrence has been reported in *Arabidopsis thaliana* when crossing *cenh3* mutants with wild type plants, obtaining haploid offspring (Ravi and Chan, 2010). The loss of a complete genome has been related to a weakened centromeric function, which will mean that when competing with a wild type centromere, centromeres containing the mutant *cenh3* histone will not successfully operate and be lost (Ravi and Chan, 2010; Britt et al., 2016). Different modifications of CENH3 have been proposed to induce haploidy (Wang et al., 2019). This has led to this approach successfully creating haploid offspring in maize (Kelliher et al., 2016; Wang et al., 2021).

A different modification for haploid induction in maize has been the use of the Stock6 line (Coe, 1959). This line has been observed to have modifications in several QTL, pointing to the multigenic nature of haploid induction in plants (Prigge et al., 2012; Dong et al., 2013; Hu et al., 2016; Lashermes et al., 1988; Barret et al., 2008), being *qhir1* the most relevant to haploid induction (Wang et al., 2019), more specifically the *MATRILINEAL* (*MTL*) gene in this QTL, which has been found to contain alterations in the Stock6 line when compared to the wild type gene (Kelliher et al., 2017; Liu et al., 2017; Gilles et al., 2017). Similar results have been observed in rice (Yao et al., 2018) and given the close relationship of the *MTL* gene within the *Poaceae* family (Kelliher et al., 2017; Wang, 2001).

Other genes in maize such as *PHOSPHOLIPASE D3* (*PLD3*) have also been observed to cause haploid offspring when mutated (Li et al., 2021).

5.3.3.3 Polyploids.

Ploidy modifications have also been used to allow hybridizations that would not occur naturally. Changes in chromosome content involving full duplications of the genetic material occur naturally in plants at different rates, from 30 to 35 % (Stebbins, 1971) to 70 % (Masterson, 1994) and has been related to different factors such as temperature, herbivory, wounding, water deficit and nutrients shortage (Ramsey and Schemske, 1998). It has also been related to speciation events in plants (Otto and Whitton, 2000; Jiao et al., 2011; Renny-Byfield and Wendel, 2014).

Depending on the origin of the sets of chromosomes, the term autopolyploid has been used when all the genetic content comes from the same species, whereas allopolyploid is used for polyploids with sets of chromosomes of different species. Allopolyploids, also known as amphiploids, have been referred to as true allopolyploids when a hybridization event leads to lack of pairing and one daughter cell keeps both sets of chromosomes, or segmental allopolyploids when pairing is possible and a polyploid is obtained after doubling the genome after hybridization. The chromosomic mechanisms of polyploidization have been studied both in somatic cells (Ramsey and Schemske, 1998) and reproductive cells (Bretagnolle and Thompson, 1995; Ramanna and Jacobsen, 2003).

The effects and causes of polyploidy in plants have been widely described and reviewed (Stebbins Jr, 1947; de Wet and Harlan, 1975; Ramsey and Schemske, 1998; Soltis and Soltis, 1999, 2009; Otto and Whitton, 2000; Osborn et al., 2003; Ramsey and Ramsey, 2014; Sattler et al., 2016). However, different species can respond differently to polyploidisation (Dewey, 1979). One of the effects observed has been the “Gigas effect” (Sattler et al., 2016), meaning an increase in cell size and larger organs. This has been proven particularly useful for ornamental crop breeding (Schifino Wittmann and Dall’Agnol, 2003) It has also been related to higher tolerance to biotic and abiotic stresses (Levin, 2002). However, the increase in cell size usually implies a reduction in cell division (Stebbins, 1971), which also leads to a reduction in growth and late flowering (Levin, 2002), in addition to meiotic aberrations multivalent formation during the polyploidisation process that can also lead to infertility of the polyploids (Stebbins, 1971).

A different effect from polyploidy is genome redundancy, which implies that deleterious alleles can be “buffered” by other copies of the genome (Soltis and Soltis, 2000; Comai, 2005). This redundancy also implies the possibility of increasing gene variation without affecting essential genes (Adams and Wendel, 2005). Polyploidy has also been related to higher heterozygosity (Moody et al., 1993; Osborn et al., 2003), which in turn have been related to better performance in different species such as maize (Randolph, 1942), potato (Mendoza and Haynes, 1974) and alfalfa (Katepa-Mupondwa et al., 2002).

On a genetic level, other effects such as chromosome rearrangements, aneuploidies, point mutations, loss of duplicated genes, gene conversion have been observed (Renny-Byfield and Wendel, 2014). On an epigenetic level, variations in gene expression, modifications in chromatin compaction, RNA interference and dosage compensations have been observed (Osborn et al., 2003; Soltis et al., 2004).

Allopolyploidisation has been related to hybrid vigour, also known as heterosis, with increased biomass, stature, growth rate, fertility levels and stress tolerance when comparing the hybrid

and the original plant (Chen, 2010). Another role of polyploidy has been restoring the fertility of hybrid species, given the infertility due to the lack of homologous chromosomes (Olsen et al., 2006; Hegarty et al., 2008).

Allopolyploidy has also been used to facilitate for bridge crossing. This technique allows the transfer of genes between species where direct crossing is not possible due to differences in ploidy. An example of this was reported in Buckner et al. (1961), where gene transfer between Italian ryegrass and tall fescue, which are incompatible, was done by crossing the Italian ryegrass with meadow fescue and crossing that hybrid with the tall fescue. Similar studies have reported the usefulness of this technique in different species such as tobacco (Burk, 1967), cowpea (Fatokun, 2000), wheat (Chhuneja et al., 2007) and cotton (Ram, 2014).

Several species of commercial significance have been obtained through hybridizations using some of these methods (Table 5.2). Triticale (*Triticosecale*) is one of the most commercially relevant hybrids (Ayalew et al., 2018). Obtained as an allopolyploid between wheat (*Triticum aestivum*, 2n=42) and rye (*Secale cereale,* 2n=14), it has been developed in different polyploid versions, tetraploid (2n=28, Łapiński, 2002; Mcgoverin et al., 2011), hexaploid (2n = 42, Merker, 1975; Oettler, 2005) and octoploid (2n=56, Oettler et al., 1991; Furman et al., 1997; Oettler, 2005), being the hexaploid version the most commonly cultivated. This is an example of amphiploidy, where the genomes of both the parentals are included in the hybrid. To circumvent the sterility induced by this event, a doubling of the genome posterior to the hybridization is required. In comparison with the parentals, Triticale has positive properties from both parentals, having the adaptative prowess of rye with the yield and quality from wheat. However, to this day its more common usage is as forage. Another example is the tobacco hybrid *Nicotiana digluta*. *Nicotiana tabacum* (2n=48) has been crossed with *N. glutinosa* (2n=24). The sterility of the amphiploid hybrid obtained, *N. digluta* (2n=36), was overcame by chromosome doubling using colchicine. The hybrid has been backcrossed with *N. tabacum* to obtain a mosaic resistant line (Valleau, 1949). A similar approach was used for the synthesis of Raphannobrassica, obtained by crossing *Raphanus sativus* (2n=18) and *Brassica oleracea* (2n=18), either by crossing tetraploid forms of the parentals or by crossing and doubling the ploidy levels of the offspring using colchicine (McNaughton, 1973).

Table 5.2. Commercial polyploids (Adapted from Sattler et al., 2016).

Common Name	Origin	Ploidy level and chromosome number	References
Coffee	Natural Allopolyploidy	4x = 44	Clarindo and Carvalho, 2008
Leek	Natural Autopolyploidy	4x = 32	Levan, 1940
Peanut	Natural Allopolyploidy	4x = 40	Raina and Mukai, 1999
Banana	Natural Autopolyploidy	3x = 33	Dantas et al., 1997
Banana	Synthetic/natural Autopolyploidy/allopolyploidy	3x = 33; 4x = 44	Silva et al., 2001
Kiwifruit	Natural Autopolyploidy	4x = 116	Hopping, 1994
Tahiti Lime	Natural Allopolyploidy	3x = 27	Morton, 1987
Plums	Natural Allopolyploidy	6x = 48	Bennett and Leitch, 1995
Strawberry	Natural Allopolyploidy	8x = 56	Whitaker, 2011
Triploid Watermelon	Synthetic Autopolyploidy	3x = 33	Crow, 1994
Apple	Synthetic/natural Autopolyploidy/allopolyploidy	3x = 51; 4x = 68	Janick et al., 1996
Grape	Synthetic/natural Autopolyploidy/allopolyploidy	3x = 57; 4x = 76	Motosugi et al., 2002
Potato	Natural Autopolyploidy	4x = 48	Carputo et al., 2003
Sweet potato	Natural Autopolyploidy	6x = 90	Roullier et al., 2013
Yam	Natural Autopolyploidy	3x = 60; 4x = 80	Arnau et al., 2009
Triploid cassava 'Sree Harsha'	Synthetic Autopolyploidy	3x = 54	Sreekumari et al., 1999
Bread wheat	Natural Allopolyploidy	6x = 42	Haider, 2013
Durum wheat	Natural Allopolyploidy	4x = 28	Haider, 2013
Oat	Natural Allopolyploidy	6x = 42	Ansari and Thomas, 1983
Triticale	Synthetic Allopolyploidy	6x = 42; 8x = 56	Mergoum and Gómez-Macpherson, 2004
Tetraploid Rye	Synthetic Autopolyploidy	4x = 28	Müntzing, 1951
Tetraploid Westerwolds ryegrass	Synthetic Autopolyploidy	4x = 28	Armstrong, 1981
Alfalfa	Natural Autopolyploidy	4x = 42	Small and Jomphe, 1989
Tetraploid Italian ryegrass	Synthetic Autopolyploidy	4x = 28	Armstrong, 1981
Tetraploid Red clover	Synthetic Autopolyploidy	4x = 28	Levan, 1948
Cotton	Natural Allopolyploidy	4x = 52	Jiang et al., 1998
Tobacco	Natural Allopolyploidy	4x = 48	Leitch et al., 2008
Sugarcane	Natural Allopolyploidy	8x = 80	Premachandran et al., 2011
Triploid Sugar beet	Synthetic Autopolyploidy	3x = 27	Kinoshita and Takahashi, 1969
Chamomile	Synthetic Autopolyploidy	4x = 36	Das, 2015
Cranberry hibiscus	Natural Allopolyploidy	4x = 72	Contreras et al., 2009
Tetraploid Rhondodendron	Synthetic Autopolyploidy	4x = 52	Kehr, 1971
Snapdragons	Synthetic Autopolyploidy	4x = 32	Tolety and Sane, 2011
Lilies	Synthetic Autopolyploidy	3x = 36; 4x = 48	Zhou et al., 2011
Kobus magnolia	Synthetic Autopolyploidy	4x = 76	Parris et al., 2010
Triploid marigold	Synthetic Allopolyploidy	3x = 18	Jalil et al., 1974
Tulip	Synthetic/natural Autopolyploidy/allopolyploidy	3x = 36; 4x = 48	Marasek-Ciolakowska et al., 2012
Rose	Synthetic/natural Autopolyploidy/allopolyploidy	3x = 21; 4x = 28; 5x = 35; 6x = 42	Yokoya et al., 2000
Chrysanthemum	Synthetic/natural Autopolyploidy/allopolyploidy	4x = 36; 6x = 54; 8x = 72; 10x = 90	Liu et al., 2012b

Several techniques have been used for inducing polyploids in plants. For doubling induction in somatic cells, temperature has been used to induce polyploidy in maize (Randolph, 1932), rye and wheat (Dorsey, 1936). This same effect has been induced using colchicine, an inhibitor of spindle function (Blakeslee and Avery, 1937; Planchais et al., 2000). However, its affinity for plant tubulins has been found to be low (Dhooghe et al., 2011), pushing for the introduction of other compounds with similar effects. Examples of these are the dinitroaniline and phosphoric amide-based herbicides (Planchais et al., 2000). In some cases, somatic doubling is not possible (Dewitte et al., 2012), requiring sexual polyploidization (Ramanna and Jacobsen, 2003). This is based on the fusion of unreduced reproductive cells, having the effects observed in polyploids and benefitting from increased variability due to the possibility of recombination and high level of heterozygosity (Ramsey and Schemske, 1998; Peloquin et al., 1999). This technique has been facilitated by different treatments such as temperature, nitrous oxide, anti-tubulin agents, EMS, in addition to gene silencing using interference RNA or Viral-Induced Gene Silencing (VIGS, Dewitte et al., 2012). This approach has been used in different plants of commercial interest (Peloquin et al., 1999; Ramanna and Jacobsen, 2003). A summary of different plants obtained from polyploidizations is provided in Table 5.2.

5.3.3.4 Breaking the pre- and post-zygotic barriers to overcome natural species boundaries.

The further the parentals are from each other in genetic terms, the more interesting the mixture of gene pools will be. However, not all species are compatible for crossing due to different factors (Aleza et al., 2010). Pre-zygotic barriers refer to impairments in pollen function, which have been related to genic or ploidy differences. These impediments have been overcame using techniques such as direct pollination of the stylar, bud pollination, *in vitro* fertilization, protoplast fusion, growth hormone use, ploidy alteration prior to hybridization and bridging species use. On the other hand, post-zygotic barriers refer to failures in development after the hybrid has been formed, leading to hybrid inviability, sterility or breakdown, meaning the sterility of the F2 progeny due to disadvantageous recombination. These have been overcome using techniques based on backcrossing or embryo rescue. This implies that ploidy levels will play a key role in hybridization.

Protoplast fusion has been used to circumvent inter-species crossing incompatibility or obtaining heterozygous specimens from vegetatively propagated plants (Schieder, 1982), obtaining plants with different beneficial traits (Liu et al., 2005). The process will be based on somatic hybridization (Carlson et al., 1972), implying protoplast isolation of the species of interest, adhesion of the protoplasts and finally cell fusion (Constabel and Cutler, 1985). Cell fusion in plants can occur spontaneously but has been facilitated using different techniques (Withers and Cocking, 1972). After the fusion process, the genetic material of the cell can include the whole nuclear genomes from both parents, referred to a symmetric fusion, or include partially one of the parental genomes, referred to as asymmetric fusion and often obtained by irradiation treatment before the fusion (Grosser and Gmitter, 2011). This process has been facilitated with the use of different techniques. These include water-soluble polymers, such as polyvinyl-alcohol (Nagata, 1978), dextran, gelatin (Kameya, 1975), lectin (Glimelius et al., 1978) and PEG (Kao and Michayluk, 1974; Wallin et al., 1974); electric stimulation (Sencia et al., 1979; Zimmermann and Scheurich, 1981), liposomes (Nagata et al., 1979), and mineral salts (Michel, 1938; Power et al., 1970; Carlson et al., 1972; Binding, 1974; Melchers and Labib, 1974). After protoplast fusion, *in vitro* culture techniques will allow the development of a full plant. Several examples of hybrid formation across species, genera, tribe and families have been obtained

(Carlson et al., 1972; Pelletier et al., 1983; Xia, 2009; Ovcharenko et al., 2011). Wang et al. (2013) summarised several crops of commercial relevance developed over the years and grown in China (Table 5.3).

A different approach has been taken using bridge crossing. This technique implies the use of an intermediary species to transfer genes from two originally incompatible species. This approach has been used in wheat, where *Aegilops umbellullata* (2n=14) is crossed first with *Triticum dicoccoides* (2n=28) and the hybrid is crossed with *Triticum aestivum* (2n=42) (Rosyara et al., 2019). *Nicotiana* species also have used this approach, such as *N. sylvestris* (2n=24) being used to transfer nematode resistance from *N. repanda* (2n=48) to *N. tabacum* (2n=48) (Burk, 1967).

The techniques forementioned are necessary due to the impossibility of an embryo forming, usually due to a genetic imbalance after gamete fusion. However, in certain examples of wide crossing, even when an embryo is formed, it is not able to develop properly due to lack of endosperm development or long dormancy periods (Reed, 2005). The basis of embryo rescue will be to use *in* vitro culture techniques to help those embryos become a full plant. This process has had different aims, from studies in plant biology to breeding programs (Collins and Grosser, 1984; Sharma et al., 1996; Haslam and Yeung, 2011; Bridgen, 2019; Ramming, 2019). In particular, plants such as cassava (Lentini et al., 2020), where hand-made crosses may have a small offspring, requiring assurance that the next generation will be obtained (Yan et al., 2014), or hybridizations of cassava with castor bean (Baguma et al., 2019) or some *Aegilops* x *Triticum* crosses, where hybrid embryos are nurtured to maturity (Gill et al., 1981; Miller et al., 1987)], among other species (Collins and Grosser, 1984; Haslam and Yeung, 2011), embryo rescue has become a useful technique.

Table 5.3. Examples of hybrid crops obtained via protoplast fusion (Adapted from Wang et al., 2013).

Crop	Plant species	Protoplast fusion	Improvement of agronomical characteristics	References
Wheat	*Triticum aestivum* L.	*T. aestivum + Agropyron elongatum*	High HMW-GS, resistance to drought and salinity	Cheng et al., 2004; Feng et al., 2004; Wang et al., 2005a; Lei et al., 2006; Wang et al., 2008; Liu et al., 2012a
		T. aestivum + Avena sativa	Good adaptability, High protein and fat content in seed	Xiang et al., 2003a, 2003b
		T. aestivum + Bromus inermis	Tolerance to cold, drought, and diseases	Xiang et al., 1999
		T. aestivum + Haynaldia villosa	Resistance to diseases, high protein content	Xia et al., 1998
		T. aestivum + Leymus chinensis	Resistance to cold, drought, salinity and diseases	Xia and Chen, 1996; Chen et al., 2004
		T. aestivum + Psathyrostachys juncea	Tolerance to drought and salinity	Li and Xia, 2004
		T. aestivum + Setaria italica	High nutritional value and drought resistance	Xiang et al., 2004
Rapeseed	*Brassica napus* L.	*B. napus + Crambe abyssinica*	High erucic acid	Wang et al., 2003, 2004
		B. napus + Camelina sativa	High linolenic acid	Jiang et al., 2009
		B. napus + Isatis indigotica	Tolerance to diseases	Du et al., 2009
		B. napus + Orychophragmus violaceus	Oil quality	Zhao et al., 2008
		B. napus + Raphanus sativus	Novel CMS, resistance to clubroot	Wang et al., 2006
		B. napus + Sinapsis alba	Yellow seed, disease resistance, oil content, novel CMS	Wang et al., 2005b; Li et al., 2009, 2012
		B. napus + Sinapis arvensis	Novel CMS, tolerance to stem rot	Hu et al., 2004
Citrus	*Citrus unshiu*	*C. unshiu + Citrus sinensis*	CMS transfer	Guo et al., 2004; Cai et al., 2007, 2009; Wang et al., 2010
		C. unshiu + Citrus grandis		
		C. unshiu + Citrus reticulata		
	Citrus aurantium	*C. aurantium + Citrus sinensis*	Resistance to CTV	Liu and Deng, 2001
	Citrus reticulata	*C. reticulata + Poncirus trifoliata*	Resistance to CEV	Zhou et al., 2004a, 2004b, 2004c, 2005a, 2005b, 2005c
Cotton	*Gossypium hirsutum* L.	*G. hirsutum + Gossypium bickii*	Disease resistance	Sun et al., 2004, 2005, 2006; Yang et al., 2007a; Fu et al., 2009; Sun et al., 2011
		G. hirsutum + Gossypium davidsonii		
		G. hirsutum + Gossypium klotzschianum		
		G. hirsutum + Gossypium stockii		
Cabbage	*Brassica oleracea* L.	*B. oleracea + Brassica nigra*	Black rot resistance	Hao et al., 2008; Zhang et al., 2008a, 2008b; Wang et al., 2011
Potato	*Solanum tuberosum* L.	*S. tuberosum + Solanum chacoense*	Resistance to bacterial wilt	Cai et al., 2004; Li et al., 2005; Guo et al., 2010
Soybean	*Gycine. max* (L.) Merr.	*G. max + Glycine cyrtoloba*	Salinity resistance	Yang et al., 2006, 2007b
Sweet potato	*Ipomoea batatas*	*I. batatas + Ipomoea triloba*	Drought resistance	Liu et al., 1992, 1994; Yang et al., 2009
		I. batatas + Ipomoea lacunosa	Quality improvement	Liu et al., 1998; Zhang et al., 2002
		I. batatas + Ipomoea cairica	Virus resistance	Guo et al., 2006

5.4. Relating traditional breeding outcomes with genome editing.

As described in the general introduction, the mechanism of action of nuclease-based genome editing tools uses the induction of DSB in a target site. This can be solved by NHEJ, which usually implies small indel events in the target site, or the conservative HR repair, which will take an homologous strand and copy it to repair the DSB. Usually, when referring to the possible outcomes obtained from genome editing, the terms Site-Directed Nuclease (SDN) 1, 2 and 3 have been used (EFSA, 2012; Podevin et al., 2013). This classification is based on the outcome obtained from editing and the repair mechanisms that led to this particular outcome. SDN-1 will refer to minor indels in the target sequence, which are obtained usually through NHEJ pathways. On the other hand, similar indels that have been obtained by providing a HR strand with modifications when compared to the original sequence would be classified as SDN-2. The main difference between SDN-1 and SDN-2 is that SDN-1 will imply random modifications at the target site but SDN-2 allows choosing the specific modification (i.e. a nucleotide change leading to a different aminoacid being coded). An example of this type of use would be the targeted modification of a *EPSPS* gene to cause resistance to glyphosate in rice (Li et al., 2016a). Finally, SDN-3 will imply the insertion larger alterations, which include the introduction of foreign material. This case could be compared to targeted insertion of T-DNA, given the targeted DSB will be where the new DNA is inserted, whereas T-DNA insertions tend to be random (Kim et al., 2007). These outcomes are defined from the modifications induced by the editing machinery, but do not include the possible insertion in the host genome of the editing machinery in stable transformation events. To circumvent this issue, transient expression or ribonucleoprotein delivery imply the activity of the editing machinery in the transformed plant, but not in subsequent generations. Another way of removing the editing machinery would be to segregate out these genes using crosses.

As described in the first section of this chapter, the European Union considers any form of genome editing, it being SDN-1, SDN-2 or SDN-3, as genetically modified organisms. However, given the novelty of the technique, it has not been considered they have a history of safe use, and therefore cannot figure in the exempt organisms list, such as other GMO like crops obtained from random mutagenesis. On the other hand, the UK has stated that it will decide how crops obtained from genome editing are regulated upon scientific evidence and the comparison with different traditional breeding techniques. With this perspective in mind, it could be argued that many of the outcomes obtained through traditional breeding techniques could be replicated using genome editing tools.

The clearest equivalence would be the induction of variation in a gene pool using random mutagenesis and SDN-1 and SDN-2. Random mutagenesis techniques, with the exception of the use of base analogs, are based on the induction of base damage and/or DSB, which is then repaired in an erroneous manner inducing in-del events. This is the exact method nuclease-based genome editing utilises to induce targeted modifications. However, random mutagens, as the name suggests, induce random DSB events all along the genome, which afterwards have to be selected, removing disadvantageous traits from the modified individual. In case of targeted mutagenesis, this selection process would be simplified. However, as mentioned previously, the possible off-target effects of genome editing should not be disregarded. On the other hand, the simultaneous appearance of different DSB along the genome, which can lead to inversions or larger deletions within the chromosome (Schmidt et al., 2020b), could be also be obtained using genome editing techniques.

When considering the mixing of different gene pools, meiosis is a key process. This mixture will be based on the combination of genetic material through crossovers, which are also related to DSBs. As previously described, any modification coming from DSB repair could be directly related to SDN-1 and SDN-2. On the other hand, the combination of genetic material due to crossovers has also been the main reasoning behind large chromosomal translocations. As described in this chapter, SPO11 will be key for the location of these crossover events by inducing DSB where the crossover will occur. This protein has been combined with Cas9 and TALEN in order to target crossovers in specific locations (Sarno et al., 2017). The ability of genome editing tools to induce crossover events was described prior to the use of SPO11 (Peterson et al., 2016), and studies such as Beying et al. (2020) have been able to induce chromosome translocations in plants using CRISPR/Cas9. This type of outcome has been observed using traditional breeding techniques and could be reproducible using genome editing in a targeted manner.

Given the nature of nuclease-based genome editing techniques, any modification related to a specific gene or a process involving DSB induction could be stimulated using genome editing techniques. However, SDN-3 would function in a way closer to classic T-DNA based transgenesis, given the insertion of larger sequences, which cannot be achieved using traditional breeding methods.

Regarding larger changes in chromosomal content, such as aneuploidies and polyploidies, the techniques previously described use the disruption of the natural distribution of the chromosomic materials using chemicals. If these were to be replicated using genome editing techniques, it would have to be based on the direct modification of genes involved in this process. Several genes have been described to disrupt this process (De Storme and Mason, 2014; Tamayo-Ordóñez et al., 2016). A specific example would the Ph1 kinase, which has been observed to regulate the correct chromosome pairing in hexaploid wheat and wild relatives during crossing events (Rey et al., 2017). Modifications in these genes may imply an alteration of the distribution of the genetic material which could lead to aneuploidies or polyploidies. However, as with the chemical induction of these events, the modification of the genes could be lethal for the offspring. On the other hand, alterations in genes codifying for CENH3 or MTL have been proven to affect the production of haploid offspring. The targeting of these genes for SDN-1 based events would be another possible objective for facilitating breeding using genome editing techniques.

In addition to the techniques included in this chapter, based on genetic alterations at different levels, other techniques based on accelerating the breeding process based on growth conditions, such as speed breeding (Gosh et al., 2018), which do not involve genetic modificatios per se will also be valuable in combination with genetic techniques.

5.5. Conclusions.

The review of traditional breeding outcomes shows the extent of genetic change that are possible using traditional breeding methods, which have been successfully managed under the practices and legislation covering breeding, varietal registration and National listing. When compared with the possible outcomes of genome editing techniques, it is shown that, when used in a specific manner, these can be equivalent to the ones obtained by traditional breeding. Therefore, when used in a specific manner, genome editing outcomes can be indistinguishable from traditional breeding outcomes, implying that they should be placed under the same regulatory framework. This would imply that a product-based regulation would be able to

include the uses of genome editing techniques to the tool box of crop breeding upon the basis of their equivalence with said traditional breeding techniques.

5.6. References.

Abdoun, F., and Beddiaf, M. (2002). *Cupressus dupreziana* A. Camus: distribution, decline and regeneration on the Tassili n'Aijer, Central Sahara. *Comptes Rendus Biologies* 325, 617–627.

Adams, K., and Wendel, J. (2005). Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology* 8, 135–141.

Ahloowalia, B. S., Maluszynski, M., and Nichterlein, K. (2004). Global impact of mutation-derived varieties. *Euphytica* 135, 187–204.

Aleza, P., Juárez, J., Ollitrault, P., and Navarro, L. (2010). Polyembryony in non-apomictic *Citrus* genotypes. *Annals of Botany* 106, 533–545.

Allshire, R. C., & Karpen, G. H. (2008). Epigenetic regulation of centromeric chromatin: Old dogs, new tricks? *Nature Reviews Genetics* 9, 923–937.

Ansari, N., and Thomas, H. (1983). A study of homoeologous relationships in the cultivated oat *Avena sativa* (2n=6x=42). *Theoretical and Applied Genetics* 66, 303–305.

Arbeithuber, B., Betancourt, A. J., Ebner, T., and Tiemann-Boege, I. (2015). Crossovers are associated with mutation and biased gene conversion at recombination hotspots. *Proceedings of the National Academy of Sciences of the United States of America* 112, 2109–2114.

Armstrong, C. S. (1981). 'Grasslands Moata' Tetraploid Italian Ryegrass (*Lolium multiflorum* Lam.). *New Zealand Journal of Experimental Agriculture* 9, 337–341.

Arnau, G., Nemorin, A., Maledon, E., and Abraham, K. (2009). Revision of ploidy status of *Dioscorea alata* L. (*Dioscoreaceae*) by cytogenetic and microsatellite segregation analysis. *Theoretical and Applied Genetics* 118, 1239–1249.

Artvinli, S. (1987). Cytoskeleton, microtubules, tubulin and colchicine: review. *Cytologia* 52, 189–198.

Ashraf M., Akram N. A., Mehboob-Ur-Rahman and Foolad M. R. (2012). Marker-assisted selection in plant breeding for salinity tolerance. *Methods in Molecular Biology* 913, 305-33.

Aung, T., and Thomas, H. (1978). The structure and breeding behaviour of a translocation involving the transfer of mildew resistance from *Avena barbata* Pott. into the cultivated oat. *Euphytica* 27, 731–739.

Ayalew, H., Kumssa, T. T., Butler, T. J., and Ma, X. F. (2018). Triticale improvement for forage and cover crop uses in the southern great plains of the United States. *Frontiers in Plant Science* 9, 1130.

Baguma, J. K., Mukasa, S. B., Kawuki, R., Tugume, A. K., Buttibwa, M., Nalela, P., Eyokia, M., Oshaba, B., Ceballos, H., Lentini, Z., and Baguma, Y. (2019). Fruit set and plant regeneration in Cassava following interspecific pollination with castor bean. *African Crop Science Journal* 27, 99–118.

Bajer, A. S., and Mole-Bajer, J. (1972). "Spindle dynamics and chromosome movements," in *International Review of Cytology*, (San Diego, CA: Academic Press), 1–271.

Barclay, I. R. (1975). High frequencies of haploid production in wheat (*Triticum aestivum*) by chromosome elimination. *Nature* 256, 410–411.

Barret, P., Brinkmann, M., & Beckert, M. (2008). A major locus expressed in the male gametophyte with incomplete penetrance is responsible for in situ gynogenesis in maize. *Theoretical and Applied Genetics* 117, 581–594.

Bennett, M. D., and Leitch, I. J. (1995). Nuclear DNA amounts in angiosperms. *Annals of Botany* 76, 113–176.

Berg, P., Baltimore, D., Boyer, H. W., Cohen, S. N., Davis, R. W., Hogness, D. S., Nathans, D., Roblin, R., Watson, J. D., Weissman, S., and Zinder, N. D. (1974). Potential biohazards of recombinant DNA molecules. *Science* 185, 303.

Berg, P., Baltimore, D., Brenner, S., Roblin, R. O., and Singer, M. F. (1975). Summary statement of the Asilomar Conference on recombinant DNA molecules. *Proceedings of the National Academy of Sciences of the United States of America* 72, 1981–1984.

Beying, N., Schmidt, C., Pacher, M., Houben, A., and Puchta, H. (2020). CRISPR/Cas9-mediated induction of heritable chromosomal translocations in Arabidopsis. *Nature Plants* 6, 638–645.

Bhat, J. A., Ali S., Salgotra R. K., Mir Z. A., Dutta S., Jadon V., Tyagi A., Mushtaq M., Jain N., Singh P. K., Singh G. P. and Prabhu K. V. (2016) Genomic selection in the era of next generation sequencing for complex traits in plant Breeding. *Frontiers in Genetics* 7, 221.

Binding, H. (1974). Fusionsversuche mit isolierten protoplasten von *Petunia hybrida* L. *Zeitschrift für Pflanzenphysiologie* 72, 422–426.

Blakeslee, A. F., and Avery, A. G. (1937). Methods of inducing doubling of chromosomes in plants: By treatment with colchicine. *Journal of Heredity* 28, 393–411.

Bond, D. J., and Chandley, A. . (1983). Aneuploidy. *Oxford monographs on medical genetics,* 22, 431–432.

Bretagnolle, F., and Thompson, J. D. (1995). Gametes with the somatic chromosome number: mechanisms of their formation and role in the evolution of autopolyploid plants. *New Phytologist* 129, 1–22.

Bridgen, M. P. (2019). A Review of Plant Embryo Culture. *HortScience* 29, 1243–1246.

Britt, A. B., & Kuppu, S. (2016). CENH3: An emerging player in haploid induction technology. *Frontiers in Plant Science*, 7, 357.

Buckner, R. C., Hill, H. D., and Burrus, P. B. (1961). Some Characteristics of Perennial and Annual Ryegrass × Tall Fescue Hybrids and of the Amphidiploid Progenies of Annual Ryegrass × Tall Fescue. *Crop Science* 1, 75–80.

Burk, L. G. (1962). Haploids in genetically marked progenies of tobacco. *Journal of Heredity* 53, 222–226.

Burk, L. G. (1967). An interspecific bridge-cross: *Nicotiana repanda* through *N. sylvestris* to *N. tabacum*. *Journal of Heredity* 58, 215–218.

Burt, A., and Trivers, R. (2006). *Genes in conflict: the biology of selfish genetic elements*. Cambridge, MA: Harvard University Press.

Cai, X. D., Fu, J., Deng, X. X., and Guo, W. W. (2007). Production and molecular characterization of potential seedless cybrid plants between pollen sterile Satsuma mandarin and two seedy *Citrus* cultivars. *Plant Cell, Tissue and Organ Culture* 90, 275–283.

Cai, X., Fu, J., Chen, C., and Guo, W. (2009). Cybrid/hybrid plants regenerated from somatic

fusions between male sterile Satsuma mandarin and seedy tangelos. *Scientia Horticulturae* 122, 323–327.

Cai, X., Liu, J., and Xie, C. (2004). Mesophyll protoplast fusion of *Solanum tuberosum* and *Solanum chacoense* and their somatic hybrid analysis. *Acta Horticulturae Sinica* 34, 623–626.

Camargo A. V., Mackay I., Mott R., Han J., Doonan J. H., Askew K., Corke F., Williams K. and Bentley A.R. (2018) Functional Mapping of Quantitative Trait Loci (QTLs) Associated With Plant Performance in a Wheat MAGIC Mapping Population. *Frontiers in Plant Science* 9, 887.

Campos, F., and Morgan, D. T. (1958). Haploid Pepper From a Sperm. *Journal of Heredity* 49, 135–137.

Carlson, P. S., Smith, H. H., and Dearing, R. D. (1972). Parasexual Interspecific Plant Hybridization. *Proceedings of the National Academy of Sciences* 69, 2292–2294.

Carputo, D., Frusciante, L., and Peloquin, S. J. (2003). The role of 2n gametes and endosperm balance number in the origin and evolution of polyploids in the tuber-bearing *Solanums*. *Genetics* 163, 287–294.

Chase, S. S. (1952). Production of Homozygous Diploids of Maize from Monoploids. *Agronomy Journal* 44, 263–267.

Chase, S. S. (1963). Analytic Breeding in *Solanum Tuberosum* L. – a Scheme Utilizing Parthenotes and Other Diploid Stocks. *Canadian Journal of Genetics and Cytology* 5, 359–363.

Chase, S. S. (1969). Monoploids and monoploid-derivatives of maize (*Zea mays* L.). *The Botanical Review* 35, 117–168.

Chen, B. Y., and Heneen, W. K. (1989). Evidence for spontaneous diploid androgenesis in *Brassica napus* L. *Sexual Plant Reproduction* 2, 15–17.

Chen, J. F., Cui, L., Malik, A. A., and Mbira, K. G. (2011). *In vitro* haploid and dihaploid production via unfertilized ovule culture. *Plant Cell, Tissue and Organ Culture* 104, 311–319.

Chen, X. L., Xia, G. M., and Chen, H. M. (2004). Nuclear and cytoplasmic genome analysis of somatic hybrid of *Triticum aestivum* L. and *Leymus chinensis* (Trin.) Tzvel. *Journal of Plant Physiology and Molecular Biology* 30, 379–386.

Chen, Z. J. (2010). Molecular mechanisms of polyploidy and hybrid vigor. *Trends in Plant Science* 15, 57–71.

Cheng, A. X., Xia, G. M., Zhi, D. Y., and Chen, H. M. (2004). Intermediate fertile *Triticum aestivum* (+) *Agropyron elongatum* somatic hybrids are generated by low doses of UV irradiation. *Cell Research* 14, 86–91.

Chester, M., Gallagher, J. P., Symonds, V. V., Da Silva, A. V. C., Mavrodiev, E. V., Leitch, A. R., Soltis, P. S., and Soltis, D. E. (2012). Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). *Proceedings of the National Academy of Sciences of the United States of America* 109, 1176–1181.

Chhuneja, P., Kaur, S., Goel, R. K., Aghaee-Sarbarzeh, M., and Dhaliwal, H. S. (2007). "Introgression of Leaf Rust and Stripe Rust Resistance Genes from *Aegilops umbellulata* to Hexaploid wheat Through Induced Homoeologous Pairing," in *Wheat Production in Stressed Environments*, eds. H. Buck, J. Nisi, and N. Salomón (Springer Netherlands), 83–90.

Choo, T. M., Reinbergs, E., and Kasha, K. J. (2011). Use of Haploids in Breeding Barley. *Plant Breeding Reviews* 3, 219–252.

CJEU (2018). Court of Justice of the European Union - PRESS RELEASE N° 111/18. *CURIA - Press and Information*.

Clapham, D. (1973). Haploid *Hordeum* Plants from Anthers *In vitro*. *Journal of Plant Breeding* 69, 142–155.

Clarindo, W. R., and Carvalho, C. R. (2008). First *Coffea arabica* karyogram showing that this species is a true allotetraploid. *Plant Systematics and Evolution* 274, 237–241.

Clausen, R. E., and Lammerts, W. E. (1929). Interspecific Hybridization in *Nicotiana*. X. Haploid and Diploid Merogony. *The American Naturalist* 63, 279–282.

Coe, E. H. (1959). A line of maize with high haploid frequency. *The American Naturalist* 93, 381–382.

Collard, B. C., and Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363, 557–572.

Collins, G. B., and Grosser, J. W. (1984). "Culture of embryos," in *Cell Culture and Somatic Cell Genetics of Plants*, ed. I. K. Vasil (Cambridge: Academic Press), 241–257.

Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nature Reviews Genetics* 6, 836–846.

Constabel, F., and Cutler, A. J. (1985). "Protoplast fusion," in *Plant Protoplasts*, eds. L. C. Fowke and F. Constabel (New York, NY: CRC Press), 53–65.

Contreras, R. N., Ruter, J. M., and Hanna, W. W. (2009). An Oryzalin-induced Autoallooctoploid of *Hibiscus acetosella* "Panama Red." *Journal of the American Society for Horticultural Science* 134, 553–559.

Copenhaver, G., Browne, W., and Preuss, D. (1998). Assaying genome-wide recombination and centromere functions with Arabidopsis tetrads. *Proceedings of the National Academy of Sciences* 95, 247–252.

Crossa, J., Beyene, Y., Kassa, S., Pérez, P., Hickey, J. M., Chen, C., et al. (2013). Genomic prediction in maize breeding populations with genotyping-by-sequencing. *G3* 3, 1903–1926.

Crow, J. F. (1994). Hitoshi Kihara, Japan's pioneer geneticist. *Genetics* 137, 891–894.

Dantas, J. L. L., Shepherd, K., de Oliveira e Silva, S., and dos Santos Soares Filho, W. (1997). "Classificação botânica, origem, evolução e distribuição geográfica", in *A cultura da banana: aspectos técnicos, socioeconômicos e agroindustriais*, ed. Alves, E.J. (Brasília, DF: Embrapa SPI; Cruz das Almas: Embrapa-CNPMF), 27-34.

Das, M. (2015). *Chamomile: medicinal, biochemical, and agricultural aspects*. Boca Raton, FL: CRC Press, Taylor & Francis Group.

Datta, S. K., and Wenzel, G. (1987). Isolated microspore derived plant formation via embryogenesis in *Triticum aestivum* L. *Plant Science* 48, 49–54.

de la Peña, A., Puertas, M. J., and Merino, F. (1981). Bimeiosis induced by caffeine. *Chromosoma* 83, 241–248.

De Massy, B. (2013). Initiation of meiotic recombination: How and where? Conservation and

specificities among eukaryotes. *Annual Review of Genetics* 47, 563–599.

De Storme, N., and Mason, A. (2014). Plant speciation through chromosome instability and ploidy change: Cellular mechanisms, molecular factors and evolutionary relevance. *Current Plant Biology* 1, 10–33.

de Wet, J. M. J., and Harlan, J. R. (1975). Weeds and Domesticates : Evolution in the Man-Made Habitat. *Economic Botany* 29, 99–107.

Devaux, P., and Kasha, K. J. (2008). "Overview of Barley Doubled Haploid Production". in *Advances in Haploid Production in Higher Plants*, eds. Touraev, A., Forster, B. P., Jain, S. H. (Dordrecht, Springer), 47–63.

Dewey, D. R. (1979). Some applications and misapplications of induced polyploidy to plant breeding. *Basic life sciences* 13, 445–470.

Dewitte, A., Van, K., and Van, J. (2012). "Use of 2n Gametes in Plant Breeding," in *Plant Breeding*, ed. I. Y. Abdurakhmonov (Rijeka: IntechOpen), Ch. 4.

Deysson, G. (1968). Antimitotic Substances. *International Review of Cytology* 24, 99–148.

Deysson, G. (1975). "Microtubules and antimitotic substances", in *Proceedings of the International Symposium on Microtubules and Microtubule Inhibitors*, eds. Borgers, M. and De Brabander, M. (Amsterdam: North-Holland Publishing Co.; New York, NY: American Elsevier Publishing Co.), 89-119.

Dhooghe, E., van Laere, K., Eeckhaut, T., Leus, L., and van Huylenbroeck, J. (2011). Mitotic chromosome doubling of plant tissues *in vitro*. *Plant Cell, Tissue and Organ Culture* 104, 359–373.

Dong, Y. Q., Zhao, W. X., Li, X. H., Liu, X. C., Gao, N. N., Huang, J. H., Wang, W. Y., Xu, X. L., and Tang, Z. H. (2016). Androgenesis, gynogenesis, and parthenogenesis haploids in cucurbit species. *Plant Cell Reports* 35, 1991–2019.

Dong, X., Xu, X., Miao, J., Li, L., Zhang, D., Mi, X., Liu, C., Tian, X., Melchinge, A. E., & Chen, S. (2013). Fine mapping of *qhir1* influencing in vivo haploid induction in maize. *Theoretical and Applied Genetics* 126, 1713–1720.

Dorsey, E. (1936). Induced polyploidy in wheat and rye. *Journal of Heredity* 27, 155–160.

dos Santos, J. P. R., Pires, L. P. M., de Castro Vasconcellos, R. C., Pereira, G. S., Von Pinho, R. G., and Balestre, M. (2016). Genomic selection to resistance to *Stenocarpella maydis* in maize lines using DArTseq markers. *BMC Genetics* 17, 86.

Du, X. zhu, Ge, X. hong, Yao, X. cheng, Zhao, Z. gang, and Li, Z. yun (2009). Production and cytogenetic characterization of intertribal somatic hybrids between *Brassica napus* and *Isatis indigotica* and backcross progenies. *Plant Cell Reports* 28, 1105–1113.

Dyer, A. F., Jong, K., and Ratter, J. A. (1970). Aneuploidy: a redefinition. *Notes from the Royal Botanic Garden Edingurgh* 30, 177–182.

EFSA (2012). Scientific opinion addressing the safety assessment of plants developed using Zinc Finger Nuclease 3 and other Site-Directed Nucleases with similar function. *EFSA Journal* 10.

Ehrensberger, R. (1948). Versuche zur Auslösung von Haploidie bei Blütenpflanzen. *Biologisches Zentralblatt* 67, 537–546.

EU Commission Communication (2000). Communication from the Commission on the precautionary principle /* COM/2000/0001 final */.

EU Commission Decision (2002). 2002/623/EC: Commission Decision establishing guidance notes supplementing Annex II to Directive 2001/18/EC of the European Parliament and of the Council on the deliberate release into the environment of genetically modified organisms and repealing Counci. *Official Journal* L 200, 22–33.

EU Commission Regulation (2003). Commission Regulation (EC) No 1983/2003 implementing Regulation (EC) No 1177/2003 of the European Parliament and of the Council concerning Community statistics on income and living conditions (EU-SILC) as regards the list of target primary variables. *Official Journal* L 298, 34–85.

EU Commission Regulation (2004). Commission Regulation (EC) No 641/2004 on detailed rules for the implementation of Regulation (EC) No 1829/2003 of the European Parliament and of the Council as regards the application for the authorisation of new genetically modified food and feed, the n. *Official Journal* L 102, 14–25.

EU Council Decision (2002). 2002/628/EC: Council Decision concerning the conclusion, on behalf of the European Community, of the Cartagena Protocol on Biosafety. *Official Journal* L 201, 48–49.

EU Council Directive (1990a). Council Directive 90/219/EEC on the contained use of genetically modified micro-organisms. *Official Journal* L 117, 1–14.

EU Council Directive (1990b). Council Directive 90/220/EEC on the deliberate release into the environment of genetically modified organisms. *Official Journal* L 117, 15–27.

EU Council Directive (1991). Council Directive 91/414/EEC concerning the placing of plant protection products on the market. *Official Journal* L 230, 1–32.

EU Parliament and Council Directive (2001). Directive 2001/18/EC of the European Parliament and of the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC - Commission Declaration. *Official Journal* L 106, 1–39.

EU Parliament and Council Directive (2009). Directive 2009/41/EC of the European Parliament and of the Council on the contained use of genetically modified micro-organisms. *Official Journal* L 125, 75–97.

EU Parliament and Council Directive (2015). Directive (EU) 2015/412 of the European Parliament and of the Council amending Directive 2001/18/EC as regards the possibility for the Member States to restrict or prohibit the cultivation of genetically modified organisms (GMOs) in their territory Text w. *Official Journal* L 68, 1–8.

EU Parliament and Council Regulation (1997). Regulation (EC) No 258/97 of the European Parliament and of the Council concerning novel foods and novel food ingredients. *Official Journal* L 43, 1–6.

EU Parliament and Council Regulation (2002). Regulation (EC) No 178/2002 of the European Parliament and of the Council laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety. *Official Journal* L 31, 1–24.

EU Parliament and Council Regulation (2003a). Regulation (EC) No 1829/2003 of the European Parliament and of the Council on genetically modified food and feed. *Official Journal* L 268, 1–23.

EU Parliament and Council Regulation (2003b). Regulation (EC) No 1946/2003 of the European Parliament and of the Council on transboundary movements of genetically modified

organisms. *Official Journal* L 287, 1–10.

Fatokun, C. A. (2000). "Breeding cowpea for resistance to insect pests: attempted crosses between cowpea and *Vigna vexillata*.," in *Challenges and opportunities for enhancing sustainable cowpea production. Proceedings of the World Cowpea Conference III held at the International Institute of Tropical Ag*, eds. C. . Fatokun, S. A. Tarawali, B. B. Singh, P. M. Kormawa, and M. Tamo (Ibadan, Nigeria: IITA), 52–61.

Feng, D., Xia, G., Zhao, S., and Chen, F. (2004). Two quality-associated HMW glutenin subunits in a somatic hybrid line between *Triticum aestivum* and *Agropyron elongatum*. *Theoretical and Applied Genetics* 110, 136–144.

Forster, B.P. (2001). Mutation genetics of salt tolerance in barley: An assessment of Golden Promise and other semi-dwarf mutants. *Euphytica* 120, 317-328.

Fourcade, .M.F., Berger, C. A., and Witkus, E. R. (1963). Cytological effects of aminopyrine. *Caryologia* 16, 347-351.

Friebe, B., Jiang, J., Raupp, W. J., McIntosh, R. A., and Gill, B. S. (1996). Characterization of wheat-alien translocations conferring resistance to diseases and pests: Current status. *Euphytica* 91, 59–87.

Friedrichs, S., Takasu, Y., Kearns, P., Dagallier, B., Oshima, R., Schofield, J., & Moreddu, C. (2019). An overview of regulatory approaches to genome editing in agriculture. *Biotechnology Research and Innovation* 3, 208–220.

Fu, L. L., Yang, X. Y., Zhang, X. L., Wang, Z. W., Feng, C. H., Liu, C. X., Jiang, P. Y., and Zhang, J. L. (2009). Regeneration and identification of interspecific asymmetric somatic hybrids obtained by donor-recipient fusion in cotton. *Chinese Science Bulletin* 54, 3035–3044.

Fuchs, L. K., Jenkins, G., and Phillips, D. W. (2018). Anthropogenic impacts on meiosis in plants. *Frontiers in Plant Science* 9, 1429.

Furman, B. J., Qualset, C. O., Skovmand, B., Heaton, J. H., Corke, H., and Wesenberg, D. M. (1997). Characterization and analysis of North American Triticale Genetic Resources. *Crop Science* 37, 1951–1959.

Gao, L., Diarso, M., Zhang, A., Zhang, H., Dong, Y., Liu, L., Lv, Z., and Liu, B. (2016). Heritable alteration of DNA methylation induced by whole-chromosome aneuploidy in wheat. *New Phytologist* 209, 364–375.

Gatica-Arias, A. (2020). The regulatory current status of plant breeding technologies in some Latin American and the Caribbean countries. *Plant Cell, Tissue and Organ Culture* 141, 229–242.

Genetic manipulation: New guidelines for UK (1978). *Nature* 276, 104–108.

Gerassimova, H. (1936). Experimentally produced haploid plant in *Crepis tectorum*. *Biologitcheski Journal* 5, 895–900.

Gill, B. S., Waines, J. G., and Sharma, H. C. (1981). Endosperm abortion and the production of viable *Aegilops squarrosa* x *Triticum boeoticum* hybrids by embryo culture. *Plant Science Letters* 23, 181–187.

Gilles, L. M., Khaled, A., Laffaire, J., Chaignon, S., Gendrot, G., Laplaige, J., Bergès, H., Beydon, G., Bayle, V., Barret, P., Comadran, J., Martinant, J., Rogowsky, P. M., & Widiez, T. (2017). Loss of pollen-specific phospholipase NOT LIKE DAD triggers gynogenesis in maize. *The EMBO Journal* 36, 707–717.

Giraut, L., Falque, M., Drouaud, J., Pereira, L., Martin, O. C., and Mézard, C. (2011). Genome-wide crossover distribution in *Arabidopsis thaliana* meiosis reveals sex-specific patterns along chromosomes. *PLoS Genetics* 7, e1002354.

Glimelius, K., Wallin, A., and Eriksson, T. (1978). Concanavalin A Improves the Polyethylene Glycol Method for Fusing Plant Protoplasts. *Physiologia Plantarum* 44, 92–96.

Goering, R. V., and Pattee, P. A. (1971). Mutants of *Staphylococcus aureus* with increased sensitivity to ultraviolet radiation. *Journal of Bacteriology* 106, 157–161.

Goodman, R. M., Hauptli, H., Crossway, A., and Knauf, V. C. (1987). Gene transfer in crop improvement. *Science* 236, 48–54.

Goodsell, S. F. (1961). Male Sterility in Corn by Androgenesis. *Crop Science* 1, 227–228.

Ghosh, S., Watson, A., Gonzalez-Navarro, O. E., Ramirez-Gonzalez, R. H., Yanes, L., Mendoza-Suárez, M., Simmonds, J., Wells, R., Rayner, T., Green, P., Hafeez, A., Hayta, S., Melton, R. E., Steed, A., Sarkar, A., Carter, J., Perkins, L., Lord, J., Tester, M., Osbourn, A., Mosco, M. J., Nicholson, P., Harwood, W., Martin, C., Domoney, C., Uauy, C., Hazard, B., Wulff, B. H. and Hickey, L. T. (2018). Speed breeding in growth chambers and glasshouses for crop breeding and model plant research. *Nature Protocols* 13, 2944–2963.

Green, A. G., and Dribnenki, J. C. (1996). "Breeding and development of Linola™ (low linolenic flax)," in *The Third Meeting of the International Flax Breeding Group* (Rome: FAO), 145–150.

Grenier, C., Cao, T. V., Ospina, Y., Quintero, C., Châtel, M. H., Tohme, J., et al. (2015). Accuracy of genomic selection in a rice synthetic population developed for recurrent selection breeding. *PLoS ONE* 10, e0136594.

Griffiths, A. J., Miller, J. H., Suzuki, D. T., Lewontin, R. C., and Gelbart, W. M. (2000). "Induced mutations," in *An Introduction to Genetic Analysis*, ed. W. H. Freeman (New York: W. H. Freeman), Ch. 16.

Grosser, J. W., and Gmitter, F. G. (2011). Protoplast fusion for production of tetraploids and triploids: Applications for scion and rootstock breeding in citrus. *Plant Cell, Tissue and Organ Culture* 104, 343–357.

Guha, S., and Maheshwari, S. C. (1964). *In vitro* production of embryos from anthers of *Datura*. *Nature* 204, 497.

Guha, S., and Maheshwari, S. C. (1966). Cell division and differentiation of embryos in the pollen grains of *Datura in vitro*. *Nature* 212, 97–98.

Guo, J. M., Liu, Q. C., Zhai, H., and Wang, Y. P. (2006). Regeneration of plants from *Ipomoea cairica* L. protoplasts and production of somatic hybrids between *I. cairica* L. and sweet potato, *I. batatas* (L.) Lam. *Plant Cell, Tissue and Organ Culture* 87, 321–327.

Guo, W. W., Prasad, D., Cheng, Y. J., Serrano, P., Deng, X. X., and Grosser, J. W. (2004). Targeted cybridization in *Citrus*: Transfer of Satsuma cytoplasm to seedy cultivars for potential seedlessness. *Plant Cell Reports* 22, 752–758.

Guo, X., Xie, C., Cai, X., Song, B., He, L., and Liu, J. (2010). Meiotic behavior of pollen mother cells in relation to ploidy level of somatic hybrids between *Solanum tuberosum* and *S. chacoense*. *Plant Cell Reports* 29, 1277–1285.

Haider, N. (2013). The origin of the B-genome of bread wheat (*Triticum aestivum* L.). *Genetika* 49, 303–314.

Halford, N. G. (2019). Legislation governing genetically modified and genome-edited crops in Europe: the need for change. *Journal of the Science of Food and Agriculture* 99, 8–12.

Hao, W., Zhao, H., Sheng, X., Chen, B., and Liu, F. (2008). Biological characters of somatic hybrids between *Brassica oleracea* L. var. botrytis and *Brassica nigra*. *Acta Agriculturae Boreali-Sinica* 23, 100–104.

Haslam, T. M., and Yeung, E. C. (2011). "Zygotic embryo culture: an overview," in *Plant Embryo Culture: Methods and Protocols*, eds. T. A. Thorpe and E. C. Yeung (New York: Springer), 3–15.

Hegarty, M. J., Barker, G. L., Brennan, A. C., Edwards, K. J., Abbott, R. J., and Hiscock, S. J. (2008). Changes to gene expression associated with hybrid speciation in plants: Further insights from transcriptomic studies in *Senecio*. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363, 3055–3069.

Henry, I. M., Dilkes, B. P., Miller, E. S., Burkart-Waco, D., and Comai, L. (2010). Phenotypic consequences of aneuploidy in *Arabidopsis thaliana*. *Genetics* 186, 1231–1245.

Herrera-Estrella, L., Depicker, A., Van Montagu, M., and Schell, J. (1983). Expression of chimaeric genes transferred into plant cells using a Ti-plasmid-derived vector. *Nature* 303, 209–213.

Heslot, N., Rutkoski, J., Poland, J., Jannink, J. L., and Sorrells, M. E. (2013). Impact of marker ascertainment bias on genomic selection accuracy and estimates of genetic diversity. *PLoS ONE* 8, e74612.

Hodgkinson, A., and Eyre-Walker, A. (2011). Variation in the mutation rate across mammalian genomes. *Nature Reviews Genetics* 12, 756–766.

Hopping, M. E. (1994). Flow cytometric analysis of *Actinidia* species. *New Zealand Journal of Botany* 32, 85–93.

Hu, H., Schrag, T. A., Peis, R., Unterseer, S., Schipprack, W., Chen, S., Lai, J., Yan, J., Prasanna, B. M., Nair, S. K., Chaikam, V., Rotarenco, V., Shatskaya, O. A., Zavalishina, A., Scholten, S., Schön, C. C., & Melchinger, A. E. (2016). The genetic basis of haploid induction in maize identified with a novel genome-wide association method. *Genetics* 202, 1267–1276.

Hu, Q., Li, Y. C., Mei, D. S., Fang, X., Hansen, L. N., and Andersen, S. B. (2004). Establishment and identification cytoplasmic male sterility in *Brassica napus* by intergeneric somatic hybridization. *Scientia Agricultura Sinica* 37, 333–338.

Hunter, C. P. (1988). Plant regeneration from microspores of barley, *Hordeum vulgare*. Ph.D. Thesis, Ashford: Wye College, University of London

Jacobs, M. (1969). Studies on the genetic activity of thymidine-base analogue in *Arabidopsis thaliana*. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis* 7, 51–62.

Jain, S. M. (2001). Tissue culture-derived variation in crop improvement. *Euphytica* 118, 153-166.

Jalil, R., Khoshoo, T., and Pal, M. (1974). Origin, nature and limit of polyploidy in marigolds. *Current Science* 43, 777–779.

Janick, J., Cimmins, J., Brown, S., and Hemmat, M. (1996). "Apples," in *Fruit breeding, Volume I: tree and tropical fruits*, eds. J. Janick and J. Moore (New York: John Wiley & Sons), 1-77.

Jiang, J., Birchler, J. A., Parrott, W. A., & Dawe, R. K. (2003). A molecular view of plant

centromeres. *Trends in Plant Science* 8, 570–575.

Jiang, C. X., Wright, R. J., El-Zik, K. M., and Paterson, A. H. (1998). Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). *Proceedings of the National Academy of Sciences of the United States of America* 95, 4419–4424.

Jiang, J. J., Zhao, X. X., Tian, W., Li, T. B., and Wang, Y. P. (2009). Intertribal somatic hybrids between *Brassica napus* and *Camelina sativa* with high linolenic acid content. *Plant Cell, Tissue and Organ Culture* 99, 91–95.

Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., Tomsho, L. P., Hu, Y., Liang, H., Soltis, P. S., Soltis, D. E., Clifton, S. W., Schlarbaum, S. E., Schuster, S. C., Ma, H., Leebens-Mack, J., and Depamphilis, C. W. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100.

Joyce S.M., Cassells A.C., Jain, S.M. (2003). Stress and aberrant phenotypes in *in vitro* culture. *Plant Cell, Tissue and Organ Culture* 74, 103–121.

Jung, W. J., and Seo, Y. W. (2014). Employment of wheat-rye translocation in wheat improvement and broadening its genetic basis. *Journal of Crop Science and Biotechnology* 17, 305–313.

Kameya, T. (1975). Induction of hybrids through somatic cell fusion with dextran sulfate and gelatin. *The Japanese Journal of Genetics* 50, 235–246.

Kao, K. N., and Michayluk, M. R. (1974). A method for high-frequency intergeneric fusion of plant protoplasts. *Planta* 115, 355–367.

Kao, K. N., Saleem, M., Abrams, S., Pedras, M., Horn, D., and Mallard, C. (1991). Culture conditions for induction of green plants from barley microspores by anther culture methods. *Plant Cell Reports* 9, 595–601.

Kasha, K. J., and Kao, K. N. (1970). High frequency haploid production in barley (*Hordeum vulgare* L.). *Nature* 225, 874–876.

Kasha K.J., Maluszynski M. (2003) "Production of doubled haploids in crop plants. An introduction", in *Doubled Haploid Production in Crop Plants*, eds. Maluszynski M., Kasha K.J., Forster B.P., Szarejko I. (Dordrecht: Springer), 1-4..

Katepa-Mupondwa, F. M., Christie, B. R., and Michaels, T. E. (2002). An improved breeding strategy for autotetraploid Alfalfa (*Medicago sativa* L.). *Euphytica* 123, 139–146.

Keeney, S., Giroux, C. N., and Kleckner, N. (1997). Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* 88, 375–384.

Keeney, S., and Neale, M. J. (2006). Initiation of meiotic recombination by formation of DNA double-strand breaks: Mechanism and regulation. *Biochemical Society Transactions* 34, 523–525.

Kehr, A. (1971). A tetraploid *Rhododendron carolinianum*. *Journal of the American Rhododendron Society* 25, 1–11.

Kehr, A. E. (1951). Monoploidy in *Nicotiana*. *Journal of Heredity* 42, 107–112.

Kelliher, T., Starr, D., Wang, W., McCuiston, J., Zhong, H., Nuccio, M. L., & Martinf, B. (2016). Maternal haploids are preferentially induced by CENH3-tailswap transgenic complementation in maize. *Frontiers in Plant Science* 7, 414.

Kelliher, T., Starr, D., Richbourg, L., Chintamanani, S., Delzer, B., Nuccio, M. L., Green, J., Chen,

Z., McCuiston, J., Wang, W., Liebler, T., Bullock, P., & Martin, B. (2017). MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. *Nature* 542, 105–109.

Kermicle, J.L. (1974). "Origin of androgenetic haploids and diploids induced by the indeterminate gametophyte (ig) mutation in maize" in *Haploids in higher plants: advances and potential: Proceedings of the first International Symposium*, ed. Kasha, K.J. (Guelph: University of Guelph Press), 137.

Khan S. U., Saeed S., Khan M. H. U., Fan C., Ahmar S., Arriagada O., Shahzad R., Branca F., and Mora-Poblete F.. (2021) Advances and Challenges for QTL Analysis and GWAS in the Plant-Breeding of High-Yielding: A Focus on Rapeseed. *Biomolecules* 11, 1516.

Parris, J. K., Ranney, T. G., Knap, H. T., and Vance Baird, W. (2010). Ploidy levels, relative genome sizes, and base pair composition in *Magnolia*. *Journal of the American Society for Horticultural Science* 135, 533–547.

Kharkwal, M. C., and Shu, Q. Y. (2009). "The role of induced mutations in world food security," in *Induced Plant Mutations in the Genomics Era,* ed. Shu, Q.Y. (Rome: FAO), 33–38.

Khush, G. S. (1973). *Cytogenetics of aneuploids*. New York, NY: Academic Press.

Kihlman, B. . (1966). *Action of Chemicals on Dividing Cell*. Englewood Cliffs, NJ: Prentice-Hall.

Kim, S. I., Veena, and Gelvin, S. B. (2007). Genome-wide analysis of *Agrobacterium* T-DNA integration sites in the Arabidopsis genome generated under non-selective conditions. *Plant Journal* 51, 779–791.

Kim, Y. S., Schumaker, K. S., and Zhu, J. K. (2006). EMS mutagenesis of Arabidopsis. *Methods in molecular biology (Clifton, N.J.)* 323, 101–103.

Kimber, G. (1976). The use of aneuploids in studies of genetics, breeding, and evolution in wheat. *Basic life sciences* 8, 103–116.

Kinoshita, T., and Takahashi, M.-E. (1969). Studies in Polyploid Varieties of Sugar Beets : XIV. Use of Cytoplasmic Male Sterility in the Production of Triploid Hybrids, and their performance in Trials. *Journal of the Faculty of Agriculture, Hokkaido University*, 56, 171-186..

Kostoff, D. (1929). An Androgenic *Nicotiana* Haploid. *Zeitschrift für Zellforschung und Mikroskopische Anatomie* 9, 640–642.

Kovalchuk, I., Kovalchuk, O., and Hohn, B. (2000). Genome-wide variation of the somatic mutation frequency in transgenic plants. *EMBO Journal* 19, 4431–4438.

Łapiński, B. (2002). "Application of tetraploid Triticale in improvement of related crop species," in *Proceedings of the 5th International Triticale Symposium, Radzików, Poland, 30 June - 5 July, 2002. Volume I: oral presentations*, eds. Arseniuk, E. and Osiński, R. (Radzików: Plant Breeding and Acclimatization Institute), 71–77.

Lashermes, P., & Beckert, M. (1988). Genetic control of maternal haploidy in maize (*Zea mays* L.) and selection of haploid inducing lines. *Theoretical and Applied Genetics* 76, 405–410.

Lassoued, R., Phillips, P. W. B., Smyth, S. J., & Hesseln, H. (2019). Estimating the cost of regulating genome edited crops: expert judgment and overconfidence. *GM Crops and Food* 10, 44–62.

Laurie, D. A., and Bennett, M. D. (1988). The production of haploid wheat plants from wheat x maize crosses. *Theoretical and Applied Genetics* 76, 393–397.

Lawrence, E. J., Griffin, C. H., and Henderson, I. R. (2017). Modification of meiotic recombination by natural variation in plants. *Journal of Experimental Botany* 68, 5471–5483.

Lei, S., Shuang-yi, Z., Fang, C., and Guang-min, X. I. A. (2006). Screening and Localization of SSR Markers Related to Salt Tolerance of Somatic Hybrid Wheat Shanrong No. 3. *Scientia Agricultura Sinica* 39, 225–230.

Leitch, I. J., Hanson, L., Lim, K. Y., Kovarik, A., Chase, M. W., Clarkson, J. J., and Leitch, A. R. (2008). The ups and downs of genome size evolution in polyploid species of *Nicotiana* (*Solanaceae*). *Annals of Botany* 101, 805–814.

Lentini, Z., Restrepo, G., Buitrago, M. E., and Tabares, E. (2020). Protocol for Rescuing Young Cassava Embryos. *Frontiers in Plant Science* 11, 522.

Lercher, M. J., and Hurst, L. D. (2002). Human SNP variability and mutation rate are higher in regions of high recombination. *Trends in Genetics* 18, 337–340.

Levan, A. (1940). Meiosis of *Allium porrum*, a Tetraploid Species With Chiasma Localisation. *Hereditas* 26, 454–462.

Levan, A. (1948). Nordisk polyploidiforadling hos Jordbruksvaxter. *Nord JordbrForskn* 30, 468–490.

Levin, D. (2002). The role of chromosomal change in plant evolution. *Oxford Series in Ecology abd Evolution* 1, 241.

Li, A., Jiang, J., Zhang, Y., Snowdon, R. J., Liang, G., and Wang, Y. (2012). Molecular and cytological characterization of introgression lines in yellow seed derived from somatic hybrids between *Brassica napus* and *Sinapis alba*. *Molecular Breeding* 29, 209–219.

Li, A., Wei, C., Jiang, J., Zhang, Y., Snowdon, R. J., and Wang, Y. (2009). Phenotypic variation in progenies from somatic hybrids between *Brassica napus* and *Sinapis alba*. *Euphytica* 170, 289–296.

Li, C. L., and Xia, G. M. (2004). Asymmetric somatic hybridization between mixed wheat and *Psathyrostachys juncea*. *Chinese journal of biotechnology* 20, 610–614.

Li, J. J., Meng, X., Zong, Y., Chen, K., Zhang, H., Liu, J., Li, J. J., and Gao, C. (2016a). Gene replacements and insertions in rice by intron targeting using CRISPR/Cas9. *Nature Plants* 2, 16139.

Li, S., Zheng, Y.-C., Cui, H.-R., Fu, H.-W., Shu, Q.-Y. and Huang, J.-Z. (2016b). Frequency and type of inheritable mutations induced by γ rays in rice as revealed by whole genome sequencing. *Journal of Zhejiang University: Science B* 17, 905–915.

Li, Y., Lin, Z., Yue, Y., Zhao, H., Fei, X., E., L., Liu, C., Chen, S., Lai, J., & Song, W. (2021). Loss-of-function alleles of *ZmPLD3* cause haploid induction in maize. *Nature Plants* 7, 1579–1588.

Li, Y., Cai, X., Li, L., and Xie, C. (2005). Evaluation of resistance to *Rolstonia solanacearum* of potato somatic hybrids. *Chinese Potato* 19, 198–203.

Liu, C., Li, X., Meng, D., Zhong, Y., Chen, C., Dong, X., Xu, X., Chen, B., Li, W., Li, L., Tian, X., Zhao, H., Song, W., Luo, H., Zhang, Q., Lai, J., Jin, W., Yan, J., & Chen, S. (2017). A 4-bp insertion at *ZmPLA1* encoding a putative phospholipase A generates haploid induction in maize. *Molecular Plant* 10, 520–522.

Liu, C., Li, S., Wang, M., and Xia, G. (2012a). A transcriptomic analysis reveals the nature of salinity tolerance of a wheat introgression line. *Plant Molecular Biology* 78, 159–169.

Liu, J., and Deng, X. (2001). Production of somatic hybrid plants between sour orange and sweet orange via electrofusion for creation of CTV-resistant rootstock. *Acta Phytophysiologica Sinica* 27, 473–477.

Liu, J., Xu, X., and Deng, X. (2005). Intergeneric somatic hybridization and its application to crop genetic improvement. *Plant Cell, Tissue and Organ Culture* 82, 19–44.

Liu, P. L., Wan, Q., Guo, Y. P., Yang, J., and Rao, G. Y. (2012b). Phylogeny of the Genus *Chrysanthemum* L.: Evidence from Single-Copy Nuclear Gene and Chloroplast DNA Sequences. *PLoS ONE* 7, e48970.

Liu, Q., Kokubu, T., and Sato, M. (1992). Shoot regeneration from protoplast fusions of sweetpotato and its related species. *Japanese Journal Breeding* 42, 88–89.

Liu, Q., Mi, K., Zhou, H., Ma, B., and Zhai, H. (1998). Regeneration and identification of interspecific somatic hybrid plants between sweetpotato and *Ipomoea lacunose*. *Acta Agronomica Sinica* 24, 529–535.

Liu, Q., Wang, J., Li, W., and Zhou, H. (1994). Protoplast fusion and regeneration of interspecific somatic hybrid plants between sweet potato (*Ipomoea batatas* (L.) Lam.) and its related species. *Journal of Agricultural Biotechnology* 2, 85–90.

Lundin, C., North, M., Erixon, K., Walters, K., Jenssen, D., Goldman, A. S. H., and Helleday, T. (2012). Methyl methanesulfonate (MMS) produces heat-labile DNA damage but no detectable in vivo DNA double-strand breaks. *Nucleic Acids Research* 40, 5794–5794.

Magni, G. E., and Von Borstel, R. C. (1962). Different Rates of Spontaneous Mutation During Mitosis and Meiosis in Yeast. *Genetics* 47, 1097–1108.

Makarevitch, I., and Harris, C. (2010). Aneuploidy causes tissue-specific qualitative changes in global gene expression patterns in maize. *Plant Physiology* 152, 927–938.

Mallikarjuna, N., Jadhav, D., Clarke, H., Coyne, C., and Muehlbauer, F. (2005). Induction of Androgenesis as a Consequence of Wide Crossing in Chickpea. *SAT eJournal* 1, 1–3.

Mann, J. D., and Storey, W. B. (1966). Rapid Action of Carbamate Herbicides Upon Plant Cell Nuclei. *Cytologia* 31, 203–207.

Marasek-Ciolakowska, A., Ramanna, M. S., Arens, P., and Van Tuyl, J. M. (2012). Breeding and cytogenetics in the genus *Tulipa*. *Floriculture and Ornamental Biotechnology* 6, 90–97.

Masterson, J. (1994). Stomatal Size in Fossil Plants: Evidence for Polyploidy in Majority of Angiosperms. *Science* 264, 421–424.

Mazia, D. (1987). The Chromosome Cycle and the Centrosome Cycle in the Mitotic Cycle. *International Review of Cytology* 100, 49–92.

McGoverin, C. M., Snyders, F., Muller, N., Botes, W., Fox, G., and Manley, M. (2011). A review of Triticale uses and the effect of growth environment on grain quality. *Journal of the Science of Food and Agriculture* 91, 1155–1165.

McKinley, K. L., & Cheeseman, I. M. (2016). The molecular basis for centromere identity and function. *Nature Reviews Molecular Cell Biology* 17, 16–29.

McNaughton, I. H. (1973). Synthesis and sterility of *Raphanobrassica*. *Euphytica* 22, 70–88.

Melchers, G., and Labib, G. (1974). Somatic hybridisation of plants by fusion of protoplasts. *MGG Molecular & General Genetics* 135, 277–294.

Mendoza, H. A., and Haynes, F. L. (1974). Genetic basis of heterosis for yield in the autotetraploid potato. *Theoretical and Applied Genetics* 45, 21–25.

Mergoum, M., and Gómez-Macpherson, H. (2004). *Triticale improvement and production*. Rome: FAO.

Merker, A. (1975). Chromosome composition of hexaploid Triticale. *Hereditas* 80, 41–52.

Meuwissen T. H. E., Hayes B. J. and Goddard M. E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829

Michel, W. (1938). Über die experimentelle Fusion pflanzlicher Protoplasten. *Protoplasma* 30, 471–471.

Michel, S., Ametz, C., Gungor, H., Epure, D., Grausgruber, H., Löschenberger, F., et al. (2016). Genomic selection across multiple breeding cycles in applied bread wheat breeding. *Theoretical and Applied Genetics* 129, 1179–1189.

Miller, T. E., Reader, S. M., Ainsworth, C. C., and Summers, R. W. (1987). "The Introduction of a Major Gene for Resistance To Powdery Mildew of Wheat, *Erysiphe Graminis* F. Sp. *Tritici*, From *Aegilops speltoides* Into Wheat, *Triticum aestivum*," in *Proceedings of the EUCARPIA Conference Cereal Breeding Related to Integrated Cereal Production*, eds. M. L. Jorna and L. A. J. Shootmaker (Wageningen: Pudoc), 179–183.

Minoia, S., Petrozza, A., D'Onofrio, O., Piron, F., Mosca, G., Sozio, G., Cellini, F., Bendahmane, A., and Carriero, F. (2010). A new mutant genetic resource for tomato crop improvement by TILLING technology. *BMC Research Notes* 3, 69.

Moody, M. E., Mueller, L. D., and Soltis, D. E. (1993). Genetic variation and random drift in autotetraploid populations. *Genetics* 134, 649–657.

Morton, J.F. (1987). "Tahiti Lime," in *Fruits of warm climates*, (ed.) Morton, J.F. (Miami, FL: Florida Flair Books), 172-175.

Motosugi, H., Okudo, K., Kataoka, D., and Naruo, T. (2002). Comparison of growth characteristics between diploid and colchicine-induced tetraploid grape rootstocks. *Journal of the Japanese Society for Horticultural Science* 71, 335–341.

Müntzing, A. (1951). Cyto-Genetic Properties and Practical Value of Tetraploid Rye. *Hereditas* 37, 17–84.

Nagata, T. (1978). A novel cell-fusion method of protoplasts by polyvinyl alcohol. *Naturwissenschaften* 65, 263–264.

Nagata, T., Eibl, H., and Melchers, G. (1979). Fusion of Plant Protoplasts Induced by a Positively Charged Synthetic Phospholipid. *Zeitschrift fur Naturforschung - Section C Journal of Biosciences* 34, 460–462.

Nandakumar, L., Rangaswamy, N. S., and Krishnamurty, H. S. (1984). Isovitexin (C-glycosylflavone) from the legume *Rhynchosia minima* (L.) DC.-a new inhibitor of cytokinesis. *Caryologia* 37, 227–232.

Neale, M. J., Pan, J., and Keeney, S. (2005). Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature* 436, 1053–1057.

Novak, F. J., and Brunner, H. (1992). Plant breeding: Induced mutation technology for crop improvement. *IAEA Bulletin* 24, 25–33.

Oettler, G. (2005). The fortune of a botanical curiosity - Triticale: Past, present and future.

Journal of Agricultural Science 143, 329–346.

Oettler, G., Wehmann, F., and Utz, H. F. (1991). Influence of wheat and rye parents on agronomic characters in primary hexaploid and octoploid triticale. *Theoretical and Applied Genetics* 81, 401–405.

Oladosu, Y., Rafii, M. Y., Abdullah, N., Hussin, G., Ramli, A., Rahim, H. A., Miah, G., and Usman, M. (2016). Principle and application of plant mutagenesis in crop improvement: a review. *Biotechnology & Biotechnological Equipment* 30, 1–16.

Olsen, R. T., Ranney, T. G., and Viloria, Z. (2006). Reproductive behavior of induced allotetraploid x*Chitalpa* and in vitro embryo culture of polyploid progeny. *Journal of the American Society for Horticultural Science* 131, 716–724.

Osborn, T. C., Pires, J. C., Birchler, J. A., Auger, D. L., Chen, Z. J., Lee, H.-S., Comai, L., Madlung, A., Doerge, R. W., Colot, V., and Martienssen, R. A. (2003). Understanding mechanisms of novel gene expression in polyploids. *Trends in Genetics* 19, 141–147.

Ossowski, S., Schneeberger, K., Lucas-Lledó, J. I., Warthmann, N., Clark, R. M., Shaw, R. G., Weigel, D., and Lynch, M. (2010). The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* 327, 92–94.

Otto, S. P., and Whitton, J. (2000). Polyploid incidence and evolution. *Annual Review of Genetics* 34, 401–437.

Ouyang, T.-W., Hu, H., Chuang, C.-C., and Tseng, C.-C. (1973). Induction of Pollen Plants From Anthers of *Triticum Aestivum* L. Cultured in Vitro. *Scientia Sinica* 16, 79–90.

Ovcharenko, O., Momot, V., Cherep, N., Sheludko, Y., Komarnitsky, I., Rudas, V., and Kuchuk, N. (2011). Transfer of transformed *Lesquerella fendleri* (Gray) Wats. chloroplasts into *Orychophragmus violaceus* (L.) O.E. Schulz by protoplast fusion. *Plant Cell, Tissue and Organ Culture* 105, 21–27.

Paget, G. E., and Walpole, A. L. (1958). Some cytological effects of griseofulvin. *Nature* 182, 1320–1321.

Panda, B. B., and Sharma, C. B. S. R. (1980). Cytogenetic hazards from agricultural chemicals: 3. Monitoring the cytogenetic activity of trichlorfon and dichlorvos in *Hordeum vulgare*. *Mutation Research - Genetic Toxicology* 78, 341–345.

Pelletier, G., Primard, C., Vedel, F., Chetrit, P., Remy, R., Rousselle, and Renard, M. (1983). Intergeneric cytoplasmic hybridization in *Cruciferae* by protoplast fusion. *MGG Molecular & General Genetics* 191, 244–250.

Peloquin, S. J., Boiteux, L. S., and Carputo, D. (1999). Meiotic mutants in potato: Valuable variants. *Genetics* 153, 1493–1499.

Perry, J., and Ashworth, A. (1999). Evolutionary rate of a gene affected by chromosomal position. *Current Biology* 9, 987–989.

Peterson, B. A., Haak, D. C., Nishimura, M. T., Teixeira, P. J. P. L., James, S. R., Dangl, J. L., and Nimchuk, Z. L. (2016). Genome-wide assessment of efficiency and specificity in CRISPR/Cas9 mediated multiple site targeting in Arabidopsis. *PLoS ONE* 11, e0162169.

Pichot, C., and El Maâtaoui, M. (2000). Unreduced diploid nuclei in *Cupressus dupreziana* A. Camus pollen. *Theoretical and Applied Genetics* 101, 574–579.

Pichot, C., El Maâtaoui, M., Raddi, S., and Raddi, P. (2001). Surrogate mother for endangered

Cupressus. *Nature* 412, 39–39.

Pichot, C., Fady, B., and Hochu, I. (2000). Lack of mother tree alleles in zymograms of *Cupressus dupreziana* A. Camus embryos. *Annals of Forest Science* 57, 17–22.

Pichot, C., Liens, B., Rivera Nava, J. L., Bachelier, J. B., and El Maâtaoui, M. (2008). Cypress surrogate mother produces haploid progeny from alien pollen. *Genetics* 178, 379–383.

Pigneur, L. M., Hedtke, S. M., Etoundi, E., and Van Doninck, K. (2012). Androgenesis: A review through the study of the selfish shellfish *Corbicula* spp. *Heredity* 108, 581–591.

Pitsikas, P., Patapas, J. M., and Cupples, C. G. (2004). Mechanism of 2-aminopurine-stimulated mutagenesis in *Escherichia coli*. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis* 550, 25–32.

Planchais, S., Glab, N., Inzé, D., and Bergounioux, C. (2000). Chemical inhibitors: A tool for plant cell cycle studies. *FEBS Letters* 476, 78–83.

Podevin, N., Davies, H. V., Hartung, F., Nogué, F., and Casacuberta, J. M. (2013). Site-directed nucleases: A paradigm shift in predictable, knowledge-based plant breeding. *Trends in Biotechnology* 31, 375–383.

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., et al. (2012). Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome* 5, 103–113.

Power, J. B., Cummins, S. E., and Cocking, E. C. (1970). Fusion of isolated plant protoplasts. *Nature* 225, 1016–1018.

Pratto, F., Brick, K., Khil, P., Smagulova, F., Petukhova, G. V., and Camerini-Otero, D. (2014). Recombination initiation maps of individual human genomes. *Science* 346, 1256442.

Premachandran, M. N., Prathima, P. T., and Lekshmi, M. (2011). Sugarcane and polyploidy - A review. *Journal of Sugarcane Research* 1, 1–15.

Prigge, V., Xu, X., Li, L., Babu, R., Chen, S., Atlin, G. N., & Melchinger, A. E. (2012). New insights into the genetics of in vivo induction of maternal haploids, the backbone of doubled haploid technology in maize. *Genetics* 190, 781–793.

Rabinovich, S. V. (1998). Importance of wheat-rye translocations for breeding modern cultivars of *Triticum aestivum* L. *Euphytica* 100, 323–340.

Raina, S. N., and Mukai, Y. (1999). Genomic in situ hybridization in *Arachis* (*Fabaceae*) identifies the diploid wild progenitors of cultivated (*A. hypogaea*) and related wild (*A. monticola*) peanut species. *Plant Systematics and Evolution* 214, 251–262.

Rajaram, S., Mann, C. H. E., Qrtiz-Ferrara, G., and Mujeeb-Kazi, A. (1983). "Adaptation, stability and high yield potential of certain 1B/1R CIMMYT wheats," in *Proceedings of the Sixth International Wheat Genetics Symposium*, ed. Sakamoto, S. (Kyoto: Maruzen Co.), 613-621.

Ram, M. (2014). "Polyploidy and distant hybridization in plant breeding," in *Plant breeding methods*, ed. M. Ram (New Delhi: PHI Learning Pvt Ltd), 423–445.

Ramanna, M. S., and Jacobsen, E. (2003). Relevance of sexual polyploidization for crop improvement - A review. *Euphytica* 133, 3–8.

Ramming, D. W. (2019). The Use of Embryo Culture in Fruit Breeding. *HortScience* 25, 393–398.

Ramsey, J., and Ramsey, T. S. (2014). Ecological studies of polyploidy in the 100 years following its discovery. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369,

20130352.

Ramsey, J., and Schemske, D. W. (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* 29, 467–501.

Randolph, L. F. (1932). Some Effects of High Temperature on Polyploidy and Other Variations in Maize. *Proceedings of the National Academy of Sciences* 18, 222–229.

Randolph, L. F. (1942). The influence of heterozygosis on fertility and vigor in autotetraploid maize. *Genetics* 27, 163.

Rattray, A., Santoyo, G., Shafer, B., and Strathern, J. N. (2015). Elevated Mutation Rate during Meiosis in *Saccharomyces cerevisiae*. *PLoS Genetics* 11, e1004910.

Ravi, M., & Chan, S. W. L. (2010). Haploid plants produced by centromere-mediated genome elimination. *Nature* 464, 615–618.

Reed, S. M. (2005). "Embryo rescue," in *Plant Development and Biotechnology*, eds. N. Robert, R. N. Trigiano, and D. J. Gray (Boca Raton: CRC Press), 235–240.

Ren, T. H., Chen, F., Yan, B. J., Zhang, H. Q., and Ren, Z. L. (2012). Genetic diversity of wheat-rye 1BL.1RS translocation lines derived from different wheat and rye sources. *Euphytica* 183, 133–146.

Renny-Byfield, S., and Wendel, J. F. (2014). Doubling down on genomes: Polyploidy and crop plants. *American Journal of Botany* 101, 1711–1725.

Rey, M. D., Martín, A. C., Higgins, J., Swarbreck, D., Uauy, C., Shaw, P., and Moore, G. (2017). Exploiting the ZIP4 homologue within the wheat Ph1 locus has identified two lines exhibiting homoeologous crossover in wheat-wild relative hybrids. *Molecular Breeding* 37, 95.

Rieger, R., Michaelis, A., and Green, M. (2012). *Glossary of genetics and cytogenetics: classical and molecula*r. Berlin, NY: Springer Science and Business.

Robert, T., Nore, A., Brun, C., Maffre, C., Crimi, B., Bourbon, H. M., and De Massy, B. (2016). The TopoVIB-Like Protein Family Is Required for Meiotic DNA Double-Strand Break Formation. *Obstetrical and Gynecological Survey* 71, 350–351.

Rost, T. L. (1984). The comparative cell cycle and metabolic effects of chemical treatments of root tip meristems. III. Chlorsulfuron. *Cytologia* 49, 61–72.

Rost, T. L., and Morrison, S. L. (1984). The Comparative Cell Cycle and Metabolic Effects of Chemical Treatments on Root Tip Meristems II. Propham, chlorpropham, and 2,4-dinitrophenol. *Cytologia* 49, 61–72.

Rosyara, U., Kishii, M., Payne, T., Sansaloni, C. P., Singh, R. P., Braun, H. J., and Dreisigacker, S. (2019). Genetic Contribution of Synthetic Hexaploid Wheat to CIMMYT's Spring Bread Wheat Breeding Germplasm. *Scientific Reports* 9, 12355.

Roullier, C., Duputié, A., Wennekes, P., Benoit, L., Fernández Bringas, V. M., Rossel, G., Tay, D., McKey, D., and Lebot, V. (2013). Disentangling the Origins of Cultivated Sweet Potato (*Ipomoea batatas* (L.) Lam.). *PLoS ONE* 8, e62707.

Roychowdhury, R., and Tah, J. (2013). "Mutagenesis-a potential approach for crop improvement," in *Crop Improvement: New Approaches and Modern Techniques*, eds. Hakeem, K.R., Ahmad, P., Ozturk, M (Boston, MA: Springer), 149-187.

Sahu, R. K., Behera, B. N., and Sharma, C. B. S. R. (1981a). Effects of a fungicide Dexon and its

derivative on root meristems. *The Nucleus* 24, 60–65.

Sahu, R. K., Behera, B. N., and Sharma, C. B. S. R. (1981b). Systemic fungicides as pretreating chemicals for root meristem squashes. *Current Science* 50, 278—279.

Salomé, P. A., Bomblies, K., Fitz, J., Laitinen, R. A. E., Warthmann, N., Yant, L., and Weigel, D. (2012). The recombination landscape in *Arabidopsis thaliana* F 2 populations. *Heredity* 108, 447–455.

Sánchez, M. A. (2020). Chile as a key enabler country for global plant breeding, agricultural innovation, and biotechnology. *GM Crops and Food* 11, 130–139.

Sarno, R., Vicq, Y., Uematsu, N., Luka, M., Lapierre, C., Carroll, D., Bastianelli, G., Serero, A., and Nicolas, A. (2017). Programming sites of meiotic crossovers using Spo11 fusion proteins. *Nucleic Acids Research* 45, e164.

Sattler, M. C., Carvalho, C. R., and Clarindo, W. R. (2016). The polyploidy and its key role in plant breeding. *Planta* 243, 281–296.

Schieder, O. (1982). "Somatic Hybridization: a New Method for Plant Improvement," in *Plant Improvement and Somatic Cell Genetics*, eds. I. K. Vasil, W. R. Scowcroft, and K. J. Frey (New York, NY: Academic Press), 239–253.

Schifino Wittmann, M. T., and Dall'Agnol, M. (2003). Induçao De Poliploidia No Melhoramiento De Plantas. *Pesquisa Agropecuária Gaúcha* 9, 155–164.

Schmidt, S. M., Belisle, M., and Frommer, W. B. (2020a). The evolving landscape around genome editing in agriculture. *EMBO Reports* 21, e50680.

Schmidt, C., Schindele, P., and Puchta, H. (2020b). From gene editing to genome engineering: restructuring plant chromosomes via CRISPR/Cas. *aBIOTECH* 1, 21–31.

Schwander, T., and Oldroyd, B. P. (2016). Androgenesis: Where males hijack eggs to clone themselves. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371, 20150534.

Sears, E. R. (1956). "The transfer of leaf-rust resistance from *Aegilops umbellulata* to wheat," in *Genetics in plant breeding. Brook-haven Symposia in Biology*, eds. Smith, H.H., Konzak, C.F., Shapiro, S., Curtis, H.J., Koshland, M.E. (Upton, NY: Brookhaven National Laboratory), 1-22.

Sears, E. R. (1993). Use of radiation to transfer alien chromosome segments to wheat. *Crop Science* 33, 897–901.

Secretariat of the Convention on Biological Diversity (SCBD) (2000). *Cartagena Protocol on Biosafety to the Convention on Biological Diversity*. Montreal: Secretariat of the Convention on Biological Diversity.

Seguí-Simarro, J. M. (2010). Androgenesis Revisited. *Botanical Review* 76, 377–404.

Seguí-Simarro, J. M., Corral-Martínez, P., Parra-Vega, V., and González-García, B. (2011). Androgenesis in recalcitrant solanaceous crops. *Plant Cell Reports* 30, 765–778.

Sencia, M., Takeda, J., Abe, S., and Nakamura, T. (1979). Induction of cell fusion of plant protoplasts by electrical stimulation. *Plant and Cell Physiology* 20, 1441–1443.

Sharma, A. K., and Sharma, A. (1980). *Chromosome Technique. Theory and Practice*. London: Butterworths.

Sharma, C. B. S. . (1971). Cytological effects of chemicals on mitotic plant cells. *Journal of*

Scientific and Industrial Research 30, 571–587.

Sharma, C. B. S. R. (1990). Chemically induced aneuploidy in higher plants. *Mutagenesis* 5, 105–126.

Sharma, C. B. S. R., Sahu, R. K., and Panigrahi, S. (1978). Inhibition of cytokinesis by a bacterial toxin - thuringiensin-A. *Caryologia* 31, 89–93.

Sharma, D. R., Kaur, R., and Kumar, K. (1996). Embryo rescue in plants - A review. *Euphytica* 89, 325–337.

Sheltzer, J. M., and Amon, A. (2011). The aneuploidy paradox: Costs and benefits of an incorrect karyotype. *Trends in Genetics* 27, 446–453.

Silva, S. O., Souza Junior, M. T., Alves, É. J., Silveira, J. R. S., and Lima, M. B. (2001). Banana Breeding Program at Embrapa. *Crop Breeding and Applied Biotechnology* 1, 399–436.

Silverman, J. A., Oliver, N., Andrew, T., and Tongchuan, L. I. (2001). Resistance studies with daptomycin. *Antimicrobial Agents and Chemotherapy* 45, 1799–1802.

Small, E., and Jomphe, M. (1989). A synopsis of the genus *Medicago* (*Leguminosae*). *Canadian Journal of Botany* 67, 3260–3294.

Smyth, S. J. (2017). Canadian regulatory perspectives on genome engineered crops. *GM Crops and Food* 8, 35–43.

Soltis, D. E., and Soltis, P. S. (1999). Polyploidy: Recurrent formation and genome evolution. *Trends in Ecology and Evolution* 14, 348–352.

Soltis, D. E., Soltis, P. S., and Tate, J. A. (2004). Advances in the study of polyploidy since Plant speciation. *New Phytologist* 161, 173–191.

Soltis, P. S., and Soltis, D. E. (2000). The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Sciences of the United States of America* 97, 7051–7057.

Soltis, P. S., and Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annual Review of Plant Biology* 60, 561–588.

Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redoña, E., et al. (2015). Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genetics* 11, e1004982.

Sreekumari, M. T., Jos, J. S., and Nair, S. G. (1999). “Sree Harsha”: A superior triploid hybrid in Cassava. *Euphytica* 106, 1–6.

Stadler, L. J. (1928a). Genetic Effects of X-Rays in Maize. *Proceedings of the National Academy of Sciences* 14, 69–75.

Stadler, L. J. (1928b). Mutations in barley induced by X-rays and radium. *Science* 68, 186–187.

Stadler, L. J. (1930). Some Genetic Effects of X-Rays. *Journal of Heredity* 21, 3–20.

Stebbins, G. L. (1971). Adaptive Radiation of Reproductive Characteristics in Angiosperms, II: Seeds and Seedlings. *Annual Review of Ecology and Systematics* 2, 237–260.

Stebbins Jr, G. L. (1947). Evidence on Rates of Evolution from the Distribution of Existing and Fossil Plant Species. *Ecological Monographs* 17, 149–158.

Sun, Y., Liu, S., Wang, Y., Jones, B. J., Wang, H., and Zhu, S. (2011). An interspecific somatic hybrid between upland cotton (*G. hirsutum* L. cv. ZDM-3) and wild diploid cotton (*G. klotzschianum* A.). *Plant Cell, Tissue and Organ Culture* 106, 425–433.

Sun, Y., Nie, Y., Guo, X., Huang, C., and Zhang, X. (2006). Somatic hybrids between *Gossypium hirsutum* L. (4x) and *G. davidsonii* Kellog (2x) produced by protoplast fusion. *Euphytica* 151, 393–400.

Sun, Y., Zhang, X., Nie, Y., and Guo, X. (2005). Production of fertile somatic hybrids of *Gossypium hirsutum* + *G. bickii* and *G. hirsutum* + *G. stockii* via protoplast fusion. *Plant Cell, Tissue and Organ Culture* 83, 303–310.

Sun, Y., Zhang, X., Nie, Y., Guo, X., Jin, S., and Liang, S. (2004). Production and characterization of somatic hybrids between upland cotton (*Gossypium hirsutum*) and wild cotton (*G. klotzschianum* Anderss) via electrofusion. *Theoretical and Applied Genetics* 109, 472–479.

Täckholm, G. (1922). Zytologische Studien über die Gattung Rosa. *Acta Horti Bergiani* 7, 97–381.

Talbot, B. (1983). Development of the national institutes of health guideliness for recombinant DNA research. *Public Health Reports* 98, 361–368.

Tamayo-Ordóñez, M. C., Espinosa-Barrera, L. A., Tamayo-Ordóñez, Y. J., Ayil-Gutiérrez, B., and Sánchez-Teyer, L. F. (2016). Advances and perspectives in the generation of polyploid plant species. *Euphytica* 209, 1–22.

The 3000 rice genomes Project (2014). The 3,000 rice genomes project. *GigaScience* 3.

Tolety, J., and Sane, A. (2011). "*Antirrhinum*," in *Wild crop relatives: genomic and breeding resources plantation and ornamental crops.*, ed. C. Kole (New York: Springer-Verlag), 1–14.

Torres, E. M., Williams, B. R., and Amon, A. (2008). Aneuploidy: Cells losing their balance. *Genetics* 179, 737–746.

Tsuda, M., Watanabe, K. N., & Ohsawa, R. (2019). Regulatory Status of Genome-Edited Organisms Under the Japanese Cartagena Act. *Frontiers in Bioengineering and Biotechnology* 7, 387.

Turnbull, C., Lillemo, M., & Hvoslef-Eide, T. A. K. (2021). Global Regulation of Genetically Modified Crops Amid the Gene Edited Crop Boom – A Review. *Frontiers in Plant Science*, 12, 630396.

Uanschou, C., Siwiec, T., Pedrosa-Harand, A., Kerzendorfer, C., Sanchez-Moran, E., Novatchkova, M., Akimcheva, S., Woglar, A., Klein, F., and Schlögelhofer, P. (2007). A novel plant gene essential for meiosis is related to the human CtIP and the yeast COM1/SAE2 gene. *EMBO Journal* 26, 5061–5070.

Valleau, W. D. (1949). The genetics of mosaic resistance in *Nicotiana glutinosa*. *Journal of agricultural research* 78, 77–79.

Veitia, R. A., Bottani, S., and Birchler, J. A. (2008). Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. *Trends in Genetics* 24, 390–397.

Vrielynck, N., Chambon, A., Vezon, D., Pereira, L., Chelysheva, L., De Muyt, A., Mézard, C., Mayer, C., and Grelon, M. (2016). A DNA topoisomerase VI-like complex initiates meiotic recombination. *Science* 351, 939–943.

Waldman, A. S. (2008). Ensuring the fidelity of recombination in mammalian chromosomes. *BioEssays* 30, 1163–1171.

Wallin, A., Glimelius, K., and Eriksson, T. (1974). The induction of aggregation and fusion of *Daucus carota* protoplasts by polyethylene glycol. *Zeitschrift für Pflanzenphysiologie* 74, 64–80.

Wang, G. xiang, Tang, Y., Yan, H., Sheng, X. guang, Hao, W. wei, Zhang, L., Lu, K., and Liu, F. (2011). Production and characterization of interspecific somatic hybrids between *Brassica oleracea* var. Botrytis and *B. nigra* and their progenies for the selection of advanced pre-breeding materials. *Plant Cell Reports* 30, 1811–1821.

Wang, J., Jiang, J., and Wang, Y. (2013). Protoplast fusion for crop improvement and breeding in China. *Plant Cell, Tissue and Organ Culture* 112, 131–142.

Wang, J., Xiang, F., and Xia, G. (2005a). *Agropyron elongatum* chromatin localization on the wheat chromosomes in an introgression line. *Planta* 221, 277–286.

Wang, L., Pan, Z. Y., and Guo, W. W. (2010). Proteomic analysis of leaves from a diploid cybrid produced by protoplast fusion between Satsuma mandarin and pummelo. *Plant Cell, Tissue and Organ Culture* 103, 165–174.

Wang, N., Gent, J. I., & Dawe, R. K. (2021). Haploid induction by a maize *cenh3* null mutant. *Science Advances*, 7, eabe2299.

Wang, M. C., Peng, Z. Y., Li, C. L., Li, F., Liu, C., and Xia, G. M. (2008). Proteomic analysis on a high salt tolerance introgression strain of *Triticum aestivum*/*Thinopyrum ponticum*. *Proteomics* 8, 1470–1489.

Wang, M., Van Bergen, S., and Van Duijn, B. (2000). Insights into a key developmental switch and its importance for efficient plant breeding. *Plant Physiology* 124, 523–530.

Wang, S., Jin, W., & Wang, K. (2019). Centromere histone H3- and phospholipase-mediated haploid induction in plants. *Plant Methods* 15, 42.

Wang, X. (2001). Plant phospholipases. *Annual Review of Plant Biology* 52, 211–231.

Wang, Y. P., Snowdon, R. J., Rudloff, E., Wehling, P., Friedt, W., and Sonntag, K. (2004). Cytogenetic characterization and *fae1* gene variation in progenies from asymmetric somatic hybrids between *Brassica napus* and *Crambe abyssinica*. *Genome* 47, 724–731.

Wang, Y. P., Sonntag, K., and Rudloff, E. (2003). Development of rapeseed with high erucic acid content by asymmetric somatic hybridization between *Brassica napus* and *Crambe abyssinica*. *Theoretical and Applied Genetics* 106, 1147–1155.

Wang, Y. P., Sonntag, K., Rudloff, E., and Chen, J. M. (2005b). Intergeneric somatic hybridization between *Brassica napus* L. and *Sinapis alba* L. *Journal of Integrative Plant Biology* 47, 84–91.

Wang, Y. P., Sonntag, K., Rudloff, E., Groeneveld, I., Gramenz, J., and Chu, C. C. (2006). Production and characterization of somatic hybrids between *Brassica napus* and *Raphanus sativus*. *Plant Cell, Tissue and Organ Culture* 86, 279–283.

Weyen, J. (2008). "Barley and Wheat Doubled Haploids in Breeding," in *Advances in Haploid Production in Higher Plants*, 179–187.

Whitaker, V. M. (2011). Applications of molecular markers in strawberry. *Journal of Berry Research* 1, 115–127.

Wijnker, E., James, G. V., Ding, J., Becker, F., Klasen, J. R., Rawat, V., Rowan, B. A., de Jong, D. F., de Snoo, C. B., Zapata, L., Huettel, B., de Jong, H., Ossowski, S., Weigel, D., Koornneef, M.,

Keurentjes, J. J. B., and Schneeberger, K. (2013). The genomic landscape of meiotic crossovers and gene conversions in *Arabidopsis thaliana*. *eLife* 2, e01426.

Withers, L. A., and Cocking, E. C. (1972). Fine-structural studies on spontaneous and induced fusion of higher plant protoplasts. *Journal of Cell Science* 11, 59–75.

Wu, Y., Sun, Y., Sun, S., Li, G., Wang, J., Wang, B., Lin, X., Huang, M., Gong, Z., Sanguinet, K. A., Zhang, Z., and Liu, B. (2018). Aneuploidization under segmental allotetraploidy in rice and its phenotypic manifestation. *Theoretical and Applied Genetics* 131, 1273–1285.

Wyatt, M. D., and Pittman, D. L. (2006). Methylating agents and DNA repair responses: Methylated bases and sources of strand breaks. *Chemical Research in Toxicology* 19, 1580–1594.

Xia, G. (2009). Progress of chromosome engineering mediated by asymmetric somatic hybridization. *Journal of Genetics and Genomics* 36, 547–556.

Xia, G., and Chen, H. (1996). Plant regeneration from intergeneric somatic hybridization between *Triticum aestivum* L. and *Leymus chinensis* (trin.) Tzvel. *Plant Science* 120, 197–203.

Xia, G., Li, Z., Wang, S., Xiang, F., Liu, J., Chen, P., and Liu, D. (1998). Asymmetric somatic hybridization between haploid common wheat and UV- irradiated *Haynaldia villosa*. *Plant Science* 137, 217–223.

Xiang, F. N., Xia, G. M., Zhou, A. F., Chen, H. M., Huang, Y., and Zhai, X. L. (1999). Asymmetric somatic hybridization between wheat (*Triticum aestivum*) and *Bromus inermis*. *Acta Botanica Sinica* 41, 458–462.

Xiang, F., Xia, G., and Chen, H. (2003a). Asymmetric somatic hybridization between wheat (*Triticum aestivum*) and *Avena sativa* L. *Science in China, Series C: Life Sciences* 46, 243–252.

Xiang, F., Xia, G., and Chen, H. (2003b). Effect of UV dosage on somatic hybridization between common wheat (*Triticum aestivum* L.) and *Avena sativa* L. *Plant Science* 164, 697–707.

Xiang, F., Xia, G., Zhi, D., Wang, J., Nie, H., and Chen, H. (2004). Regeneration of somatic hybrids in relation to the nuclear and cytoplasmic genomes of wheat and *Setaria italica*. *Genome* 47, 680–688.

Xiong, Z., Gaeta, R. T., and Pires, J. C. (2011). Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. *Proceedings of the National Academy of Sciences of the United States of America* 108, 7908–7913.

Yan, H., Lu, L., Alzate, A., Ceballos, H., Hershey, C., Chen, S., and Li, K. (2014). Fruit, seed and embryo development of different cassava (*Manihot esculenta* Crantz) genotypes and embryo rescue. *African Journal of Biotechnology* 13, 1524–1528.

Yang, S., Wang, L., Huang, J., Zhang, X., Yuan, Y., Chen, J. Q., Hurst, L. D., and Tian, D. (2015). Parent-progeny sequencing indicates higher mutation rates in heterozygotes. *Nature* 523, 463–467.

Yang, X. Y., Zhang, X. L., Jin, S. X., Fu, L. L., and Wang, L. G. (2007a). Production and characterization of asymmetric hybrids between upland cotton Coker 201 (*Gossypium hirsutum*) and wild cotton (*G. klozschianum* Anderss). *Plant Cell, Tissue and Organ Culture* 89, 225–235.

Yang, Y., Guan, S., Zhai, H., He, S., and Liu, Q. (2009). Development and evaluation of a storage

root-bearing sweetpotato somatic hybrid between *Ipomoea batatas* (L.) Lam. and *I. triloba* L. *Plant Cell, Tissue and Organ Culture* 99, 83–89.

Yang, Y., Jiang, D. A., Xu, H. X., Yan, C. Q., and Hao, S. R. (2006). Cyclic electron flow around photosystem 1 is required for adaptation to salt stress in wild soybean species *Glycine cyrtoloba* ACC547. *Biologia Plantarum* 50, 586–590.

Yang, Y., Yan, C. Q., Cao, B. H., Xu, H. X., Chen, J. P., and Jiang, D. A. (2007b). Some photosynthetic responses to salinity resistance are transferred into the somatic hybrid descendants from the wild soybean *Glycine cyrtoloba* ACC547. *Physiologia Plantarum* 129, 658–669.

Yao, L., Zhang, Y., Liu, C., Liu, Y., Wang, Y., Liang, D., Liu, J., Sahoo, G., & Kelliher, T. (2018). *OsMATL* mutation induces haploid seed formation in indica rice. *Nature Plants* 4, 530–533.

Yokoya, K., Roberts, A., Mottley, J., Lewis, R., and Brandham, P. (2000). Nuclear DNA amounts in roses. *Annals of Botany* 85, 557–561.

Zeller, F. J. (1973). 1B/1R wheat-rye chromosome substitutions and translocations. in *Proceedings 4th International Wheat Genetics Symposium*, eds. E. R. Sears and L. M. S. Sears (Columbia: University of Missouri), 209–221.

Zeller, F. J., and Hsam, S. L. K. (1983). Broadening the genetic variability of cultivated wheat by utilizing rye chromatin. in *Proceedings 6th International Wheat Genetics Symposium*, ed. S. Sakamoto (Kyoto: Kyoto University), 161–173.

Zhang, B. Y., Liu, Q. C., Zhai, H., Zhou, H. Y., Zhang, D. P., and Wang, Y. (2002). Production of Fertile interspecific somatic hybrid plants between sweetpotato and its wild relative, *Ipomoea lacunose*. *Acta Horticulturae* 583, 81–85.

Zhang, H., Bian, Y., Gou, X., Zhu, B., Xu, C., Qi, B., Li, N., Rustgi, S., Zhou, H., Han, F., Jiang, J., Von Wettstein, D., and Liu, B. (2013). Persistent whole-chromosome aneuploidy is generally associated with nascent allohexaploid wheat. *Proceedings of the National Academy of Sciences of the United States of America* 110, 3447–3452.

Zhang, J., Stewart, J. M., and Turley, R. B. (1999). Analysis of semigamy expression in cotton (*Gossypium barbadense*). in *Proceedings of the 1999 Beltwide Cotton Conference*, eds. C. Richter, P. Dugger, and D. A. (Orlando: Beltwide Cotton Conferences), 446–447.

Zhang, L., Zhao, H., Bi, C., and Liu, F. (2008a). Development and Identification of Interspecific Somatic Hybrids Between Cauliflower and Black Mustard. *Chinese Bulletin of Botany* 25, 176–184.

Zhang, L., Zhao, H., and Liu, F. (2008b). Identification and analysis of asymmetric somatic hybrids between *B. oleracea* and *B. nigra*. *Journal of molecular cell biology* 41, 265–274.

Zhao, Y., Mette, M. F., Gowda, M., Longin, C. F. H., and Reif, J. C. (2014). Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. *Heredity* 112, 638–645.

Zhao, Z. G., Hu, T. T., Ge, X. H., Du, X. Z., Ding, L., and Li, Z. Y. (2008). Production and characterization of intergeneric somatic hybrids between *Brassica napus* and *Orychophragmus violaceus* and their backcrossing progenies. *Plant Cell Reports* 27, 1611–1621.

Zhou, K., Guo, W., Xia, R., Huang, R., and Hu, L. (2004a). Effects of somatic hybrids rootstocks on the growth of young tree and the activities of anti-oxidant enzymes in leaves and root of Navel Orange (*Citrus sinensis*). *Plant Physiol Commun* 40, 540–544.

Zhou, K., Guo, W., Xia, R., Huang, R., Hu, L., and Liu, F. (2005a). Some physiology and biochemical characters of five citrus rootstocks. *Journal of Huazhong Agricultural University* 24, 79–82.

Zhou, K., Guo, W., Xia, R., Wang, G., and Bao, H. (2005b). Effects of different kinds of rootstocks on the growth of *Citrus* young tree and the carbohydrates contents in leaves. *Journal of Tropical Subtropical Botany* 13, 17–20.

Zhou, K., Guo, W., Xia, R., Wang, G., and Bao, H. (2005c). Effects of two kinds of rootstocks on the growth of young citrus trees and the contents of mineral nutrients in leaves. *Acta Botanica Boreali-Occidentalia Sinica* 25, 293–298.

Zhou, K., Guo, W., Xia, R., Wang, G., and Shen, T. (2004b). Effects of two kinds of rootstock on growth and change of nutrient contents in leaf of young tree of Navel orange. *Journal of Plant Nutrition and Soil Science* 10, 657–662.

Zhou, K., Guo, W., Xia, R., Wang, G., and Shen, T. (2004c). Effects of two kinds of rootstocks on the growth of navel orange (*Citrus sinensis* Osbeck) young tree and the annual changes in the contents of carbohydrates in leaves. *Wuhan Journal of Botanical Research* 22, 355–358.

Zhou, S., Zhou, G., and Li, K. (2011). Euploid endosperm of triploid × diploid/tetraploid crosses results in aneuploid embryo survival in *Lilium*. *HortScience* 46, 558–562.

Zimmermann, U., and Scheurich, P. (1981). High frequency fusion of plant protoplasts by electric fields. *Planta* 151, 26–32.

Chapter 6. General discussion

6.1. Developing a transient expression platform for genome editing studies.

Genome editing techniques, particularly CRISPR/Cas9, have risen to become one of the most relevant tools in molecular biology, given the possibility of inducing targeted mutations with a simple system. However, its use as a tool for crop development is still in doubt in some countries given regulatory restrictions. *Brachypodium distachyon* has also risen as a model for the *Poaceae* family, one of the plant clades with the highest commercial value. Targeted modifications have been achieved using CRISPR/Cas9 in Brachypodium (Raissig et al., 2016; van der Schuren et al., 2018; Hus et al., 2020). However, further optimisations to the system can be achieved, such as the validation of new techniques for delivery of the editing machinery, new promoters with activity in Brachypodium or further refinement of the sgRNA design and target choice for optimal editing efficiency. The main objective of this thesis was to investigate these possible optimisations to increase the efficiency of CRISPR/Cas9 editing in Brachypodium.

One of the main objectives of the thesis was to develop a platform for the validation of promoters for the expression of sgRNA and, by extension, a possible means to assay sgRNA activity. Two different main platforms were considered, protoplasts and embryos. Protoplasts were isolated using two different protocols, but their maintenance and transformation was not reliable enough for its use as a routine assay in the lab due to inconsistencies in the quality of the starting plant material, as the results provided in Chapter 2 show. However, given their prominence as a transformable tissue (Supplementary Table 11), they could be considered as a possibility for the validation of sgRNA using transient expression in plants.

Due to the challenges of working with protoplasts, an alternative method was proposed: the use of immature embryos. Due to the type of target tissue, biolistic delivery was chosen as the method for rapid transformation. A physical model of the embryo was developed for further understanding of the transformation efficiency obtained and the possibility of their use for sgRNA validation. Given the transformation efficiency observed, it was necessary to utilise a RE-PCR assay to facilitate the identification of edited amplicons, given that they would be in a pool of non-transformed and non-edited DNA. These results were the basis for the techniques used in Chapter 3 for the validation of sgRNA promoters. Additionally, the ability of the ZmUbi1 promoter to express correctly the Cas9 was tested by the development of the pDGV-H2BGFP construct, confirming that the Cas9 will be expressed using the plasmids designed in Chapter 3. This plasmid will also make it possible to compare *Agrobacterium*-mediated transformation and biolistic delivery into Brachypodium embryos in the future.

In Chapter 3, the main focus was to perform a pilot experiment to test the ability of three different RNA Pol II promoters to induce edits in Brachypodium embryos and validating the protocol described in Chapter 2. Three different plasmids were assembled with the aim of obtaining equivalent plasmids only differing in the sgRNA cassette. Different possible targets were considered for the editing assays, including endogenous sequences and transgenes previously inserted into the genomic DNA of Brachypodium. Possible sgRNA were designed for each one of the genes of interest, focusing on *BdDWF4* and HygR. In case of HygR, two different versions of the transgene were analysed (in the plasmids pJJ2LB and pBract204 respectively), selecting one spacer that would target only pJJ2LB and another one that would be able to target both cassettes. Once the spacers were designed, they were successfully inserted in the plasmids containing the rest of the editing machinery, which had different promoters controlling the expression of the sgRNA.

The results obtained in the pilot experiment showed a mixture of amplicons obtained from the samples bombarded with DNA containing the elements necessary for the expression of the editing machinery. In particular, two of the promoters studied, BdU6 (van der Schuren et al., 2018) or OsU3 (Raissig et al., 2016; Hus et al., 2020), were able to consistently produce modifications in the target *BdDWF4,* with samples showing a diversity in chromatograms obtained from the amplicons produced after the RE-PCR protocols proposed. This points to an ability of inducing edits, concurring with the published work. However, TaU6 did not show the same results, which was surprising given the ubiquity of this system in other species (Lawrenson et al., 2015). Given the necessary elements were the same in all three systems, it can be assumed that the TaU6 promoter was the reason behind the lower editing efficiency observed. The results obtained point to the method developed in Chapter 2 being able to introduce the necessary DNA in *Brachypodium* embryos. On the other hand, the results point to the ability of the constructs designed to produce edits in *Brachypodium* and that these are observable using RE-PCR. This implies that, although TaU6 was not the best option for sgRNA expression in *Brachypodium*, the method described could be a platform for the study of other promoters. However, as described in Chapter 3, although the results presented point to the ability of these constructs to produce edits, these results will need confirmation by further repetitions and the inclusion of the necessary controls in order to categorically confirm that the combination of the protocol described in Chapter 2 to introduce DNA in *Brachypodium* embryos, together with the RE-PCR approach proposed and the constructs designed, is able to generate edits *in vivo*.

6.2. Developing a plant-based model for the prediction of editing efficiencies.

In addition to analysing the capabilities of each promoter to induce edits, the spacer sequence has been described to be a major factor regarding editing efficiency, to the extent that several algorithms have been described predicting editing efficiencies depending on the spacer sequence (See Chapter 4). However, upon reviewing the literature and analysing the spacers described with different algorithms, it was observed that the efficiencies reported in the bibliography did not correlate with the efficiencies predicted using different algorithms. These results correlate with the ones described in (Naim et al., 2020). Therefore, it was decided to use the database collected for this analysis to develop a model that had a better fit to the data than the tools available online, developing the CRISPRnt-Plant model. To allow simple modifications and addition of new guides, this was encoded in Excel using a one-hot encoding system. This model showed correlations when favouring different nucleotides in different positions when compared to other models, but differed in others (See Chapter 4). Upon analysing the performance of the model using the same analysis as the other tools, it was observed that CRISPRnt-Plant had a better correlation with the reported efficiencies than the other models available online. However, different statistical analysis of the model suggested that there is room for improvement. The results of this chapter point to the need for a plant-based analysis of sgRNA efficiency in order to develop an accurate model for the prediction of editing efficiency using CRISPR/Cas9.

6.3. Putting genome editing techniques in the context of conventional breeding.

Genome editing, and in particular CRISPR/Cas9, has the potential to be a useful tool for both research and crop development. However, its use in crop development may become limited due

to the application of existing regulations on genetically modified organisms in various countries. In July 2018, the ECJ ruled that organisms derived through a process of gene editing should be regulated as GMOs in the European Union. However, many politicians and others in the United Kingdom have openly disagreed with this ECJ ruling and have indicated that the UK should take a different approach. DEFRA set up a public consultation to gather views on whether organisms developed using genetic technologies such as gene editing should be regulated as genetically modified organisms even if their genetic changes could have been produced through traditional breeding. However, the range of outcomes that are possible through traditional breeding were not defined.

Plant breeding is based on harnessing genetic variation, either by selecting advantageous variation in breeding lines or wild populations or by inducing additional variation where it does not exist naturally. This is the basis of random mutagenesis techniques, which will induce many mutations from which the advantageous variations will be selected and isolated. These sequence variations can go from point mutations through larger in-del events to major inversions, duplications or deletions. In some cases, aneuploidies and chromosome translocations have also been used to introduce novel chromosomic content. Plant breeding has also deliberately or unintentionally generated a range of whole genome duplications. Autopolyploids arise by genome duplication within the same species whereas allopolyploids are generated when two or more distinct genomes combine but often remain distinct during meiosis. On the other hand, when the aim is to reduce the variability and facilitate the production of homozygous individuals with advantageous traits, double haploids have been used.

Novel genetic variation can also be generated through wide hybridisations that are normally partially or completely incompatible. To circumvent this issue, techniques such as protoplast fusion, bridge crossing or embryo rescue have been used.

The possible outcomes of genome editing have been summarised as SDN-1, referring to small indels caused mainly by NHEJ based repair of DSB events induced by the editing machinery, SDN-2, referring to small indels caused by HR based repair of DSB after providing a recombination template containing said indels, and SDN-3, referring to the insertion of larger foreign DNA which can be functional. When comparing the possible outcomes obtained from genome editing and the already in use conventional breeding techniques, it could be argued that SDN-3 relates to transgenesis rather than conventional breeding outcomes. However, SDN-1 and SDN-2 could be equivalent to the aforementioned techniques able to induce variation. Additionally, several targets could be used to replicate other outcomes observed in conventional breeding techniques. This implies that genome editing techniques could be equivalent to conventional breeding techniques, provided that the target modification does not imply the insertion of functional foreign material. Therefore, upon the logic of equivalence of genome editing techniques and traditional breeding techniques, it could be concluded that SDN-1 and SDN-2 are able to induce changes that would be equivalent on a genetic level, and therefore should not be classified as GMO under the new UK regulation. It therefore follows that regulating organisms resulting from SDN-1 and -2 edits as GMOs would not be proportional if they could be made using traditional breeding methods.

6.4. Future perspectives.

The results provided in this thesis also lead to other interesting lines of research.

Firstly, a protoplast isolation and transformation method could be adapted for the lab, based on the protocol described in Hus et al. (2020). As described in Chapter 2, there are advantages and disadvantages in both protoplast and embryo-based protocols. Having both of them available in the lab would be a great tool for the future. Additionally, a protocol using *Agrobacterium*-mediated transformation (as described in Zombori et al., 2011) would also be useful. Given the pDGV-H2BGFP plasmid assembled, comparisons between transformation efficiencies using biolistic and *Agrobacterium*-mediated transformation would be possible while using the same plasmid.

Regarding the experiments described in Chapter 2 and 3, further optimisations and controls are required for its use. The RE-PCR protocol must be optimised, focusing on the primers used for the initial amplification of the target genes and the isolation of the amplicons through protocols other than gel-clean up. Other controls focused on the bombardment process must also be performed, including a set of "blank" bombardments, with the gold without any DNA coating to confirm the fluorescence observed is due to the DNA introduced. To be able to confirm the possible modifications are due to the editing machinery and not other factors, another set of bombardments with the DNA containing the editing machinery without any targeting elements (sgRNA). The samples obtained from these bombardments should also be used as a control for the rest of the process, performing the RE-PCR process and ultimately sequencing the amplicons obtained. In addition to the controls necessary, future experiments will require the sequencing of the specific edits obtained through cloning of the amplicons obtained. Once optimised and corroborated, these techniques, together with other techniques that could be available, more promoters within the *Poaceae* family could be tested in Brachypodium. As outlined in Supplementary Table 2, other promoters such as maize U3 (Liang et al., 2014) and U6 (Svitashev et al., 2015) or rice U6 (Jiang et al., 2013) have been utilised for CRISPR/Cas9 in other species. Using the same approach as described in Chapter 3, the expression of these promoters could be tested. In addition to their capability of inducing edits, other factors such as sgRNA expression or sgRNA sequencing after being expressed would help understand the differences in each one of the cassettes and link them to the differences in editing efficiency.

Additionally, a platform for the investigation of gDNA environment on genome editing efficiencies has been proposed, using the analysis performed in Chapter 3 selecting T-DNA lines with the same unique insertion in different positions in the genome. These possible effects have been described in humans or *in vitro* (Chen et al., 2016, 2017; Isaac et al., 2016; Daer et al., 2017; Jensen et al., 2017; Verkuijl and Rots, 2019). However, this has not been studied in plants. Studying editing efficiency targeting the inserted transgenes would elucidate the effects of chromatin compaction on CRISPR/Cas9 function.

The CRISPRnt-Plant model could be further refined in different ways. Firstly, the same modelling approach could be used in a library-based dataset, which would imply the variation observed would be due to the different sequences used and not other factors that cannot be controlled when collecting a dataset from the published literature. Additionally, the CRISPRnt-Plant model could be coded in a different coding language. As described in Chapter 4, Excel poses some advantages due to the simplicity of the software, but also poses some limitations. By coding it in a different language such as Python, the addition of other sgRNA to the analysis and deeper statistical analysis could be performed.

Regarding Chapter 5 concerning the place of genome editing in the crop development toolbox, further communications between the scientific community and the pertinent authorities will be required to assess the potential advantages and risks and include them as a possibility. However, the results provided in Chapter 5, together with other communications from the scientific community will be essential for the understanding of these authorities and the general public of the pros and cons these types of techniques pose in future crop development programs.

6.5. References.

Chen, X., Liu, J., Janssen, J. M., and Gonçalves, M. A. F. V. (2017). The Chromatin Structure Differentially Impacts High-Specificity CRISPR-Cas9 Nuclease Strategies. *Molecular Therapy - Nucleic Acids* 8, 558–563.

Chen, X., Rinsma, M., Janssen, J. M., Liu, J., Maggio, I., and Gonçalves, M. A. F. V. (2016). Probing the impact of chromatin conformation on genome editing tools. *Nucleic Acids Research* 44, 6482–6492.

CJEU (2018). Court of Justice of the European Union - PRESS RELEASE N° 111/18. *CURIA - Press and Information*.

Daer, R. M., Cutts, J. P., Brafman, D. A., and Haynes, K. A. (2017). The Impact of Chromatin Dynamics on Cas9-Mediated Genome Editing in Human Cells. *ACS Synthetic Biology* 6, 428–438.

Hus, K., Betekhtin, A., Pinski, A., Rojek-Jelonek, M., Grzebelus, E., Nibau, C., Gao, M., Jaeger, K. E., Jenkins, G., Doonan, J. H., and Hasterok, R. (2020). A CRISPR/Cas9-Based Mutagenesis Protocol for *Brachypodium distachyon* and Its Allopolyploid Relative, *Brachypodium hybridum*. *Frontiers in Plant Science* 11, 614.

Isaac, R. S., Jiang, F., Doudna, J. A., Lim, W. A., Narlikar, G. J., and Almeida, R. (2016). Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. *eLife* 5, e13450.

Jensen, K. T., Fløe, L., Petersen, T. S., Huang, J., Xu, F., Bolund, L., Luo, Y., and Lin, L. (2017). Chromatin accessibility and guide sequence secondary structure affect CRISPR/Cas9 gene editing efficiency. *FEBS Letters* 591, 1892–1901.

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B., and Weeks, D. P. (2013). Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic acids research* 41, e188.

Lawrenson, T., Shorinola, O., Stacey, N., Li, C., Østergaard, L., Patron, N., Uauy, C., and Harwood, W. (2015). Induction of targeted, heritable mutations in barley and *Brassica oleracea* using RNA-guided Cas9 nuclease. *Genome biology* 16, 258.

Liang, Z., Zhang, K., Chen, K., and Gao, C. (2014). Targeted mutagenesis in *Zea mays* using TALENs and the CRISPR/Cas system. *Journal of Genetics and Genomics* 41, 63–68.

Naim, F., Shand, K., Hayashi, S., O'Brien, M., McGree, J., Johnson, A. A. T., Dugdale, B., and Waterhouse, P. M. (2020). Are the current gRNA ranking prediction algorithms useful for genome editing in plants? *PLoS ONE* 15, e0227994.

Raissig, M. T., Abrash, E., Bettadapur, A., Vogel, J. P., and Bergmann, D. C. (2016). Grasses use an alternatively wired bHLH transcription factor network to establish stomatal identity. *Proceedings of the National Academy of Sciences of the United States of America* 113, 8326–8331.

Svitashev, S., Young, J. K., Schwartz, C., Gao, H., Falco, S. C., and Cigan, A. M. (2015). Targeted mutagenesis, precise gene editing, and site-specific gene insertion in maize using Cas9 and guide RNA. *Plant Physiology* 169, 931–945.

van der Schuren, A., Voiniciuc, C., Bragg, J., Ljung, K., Vogel, J., Pauly, M., and Hardtke, C. S. (2018). Broad spectrum developmental role of Brachypodium AUX1. *New Phytologist* 219, 1216–1223.

Verkuijl, S. A., and Rots, M. G. (2019). The influence of eukaryotic chromatin state on

CRISPR/Cas9 editing efficiencies. *Current Opinion in Biotechnology* 55, 68–73.

Zombori, Z., Szécsényi, M., and Györgyey, J. (2011). Different approaches for *Agrobacterium*-mediated genetic transformation of *Brachypodium distachyon*, a new model plant for temperate grasses. *Acta Biologica Szegediensis* 55, 193–195.

7. Supplementary Materials

7. Supplementary Materials.

7.1. Supplementary figures.

```
pBract204         TTTGGAGAGGACGACCCCGATATGAAAAAGCCTGAACTCACCGCGACGTCTGTCGAGAAG 60
pJJ2LB            ------------------------------------------------------------ 60

pBract204         TTTCTGATCGAAAAGTTCGACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAA 120
pJJ2LB            ----------------------------------------------GAGGGCGAAGAA 120

pBract204         TCTCGTGCTTTCAGCTTCGATGTAGGAGGGCGTGGATATGTCCTGCGGGTAAATAGCTGC 180
pJJ2LB            TCTCGTGCTTTCAGCTTCGATGTAGGAGGGCGTGGATATGTCCTGCGGGTAAATAGCTGC 180

pBract204         GCCGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACTTTGCATCGGCCGCGCTCCCG 240
pJJ2LB            GCCGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACTTTGCATCGGCCGCGCTCCCG 240

pBract204         ATTCCGGAAGTGCTTGACATTGGGGAATTCAGCGAGAGCCTGACCTATTGCATCTCCCGC 300
pJJ2LB            ATTCCGGAAGTGCTTGACATTGGGGAGTTTAGCGAGAGCCTGACCTATTGCATCTCCCGC 300

pBract204         CGTGCACAGGGTGTCACGTTGCAAGACCTGCCTG-AACCGAACTGCCCGCTGTTCTGCAG 360
pJJ2LB            CGTGCACAGGGTGTCACGTTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTT------ 360

pBract204         GTAAATTTCTAGTTTTTCTCCTTCATTTTCTTGGTTAGGACCCTTTTCTCTTTTTATTTT 420
pJJ2LB            ------------------------------------------------------------ 420

pBract204         TTTGAGCTTTGATCTTTCTTTAAACTGATCTATTTTTTAATTGATTGGTTATGGTGTAAA 480
pJJ2LB            ------------------------------------------------------------ 480

pBract204         TATTACATAGCTTTAACTGATAATCTGATTACTTTATTTCGTGTGTCTATGATGATGATG 540
pJJ2LB            ------------------------------------------------------------ 540

pBract204         ATAACTGCAGCCGGTCGCGGAGGCCATGGATGCGATCGCTGCGGCCGATCTTAGCCAGAC 600
pJJ2LB            ----CTACAACCGGTCGCGGAGGCTATGGATGCGATCGCTGCGGCCGATCTTAGCCAGAC 600

pBract204         GAGCGGGTTCGGCCCATTCGGACCGCAAGGAATCGGTCAATACACTACATGGCGTGATTT 660
pJJ2LB            GAGCGGGTTCGGCCCATTCGGACCGCAAGGAATCGGTCAATACACTACATGGCGTGATTT 660

pBract204         CATATGCGCGATTGCTGATCCCCATGTGTATCACTGGCAAACTGTGATGGACGACACCGT 720
pJJ2LB            CATATGCGCGATTGCTGATCCCCATGTGTATCACTGGCAAACTGTGATGGACGACACCGT 720

pBract204         CAGTGCGTCCGTCGCGCAGGCTCTCGATGAGCTGATGCTTTGGGCCGA------------ 780
pJJ2LB            CAGTGCGTCCGTCGCGCAGGCTCTCGATGAGCTGATGCTTTGGGCCGAGGACTGCCCCGA 780

pBract204         ------------------------------------------------------------ 840
pJJ2LB            AGTCCGGCACCTCGTGCACGCGGATTTCGGCTCCAACAATGTCCTGACGGACAATGGCCG 840

pBract204         ------------------------------------------------------------ 900
pJJ2LB            CATAACAGCGGTCATTGACTGGAGCGAGGCGATGTTCGGGGATTCCCAATACGAGGTCGC 900

pBract204         ------------------------------------------------------------ 960
pJJ2LB            CAACATCTTCTTCTGGAGGCCGTGGTTGGCTTGTATGGAGCAGCAGACGCGCTACTTCGA 960

pBract204         ------------------------------------------------------------ 1020
pJJ2LB            GCGGAGGCATCCGGAGCTTGCAGGATCGCCACGACTCCGGGCGTATATGCTCCGCATTGG 1020

pBract204         ------------------------------------------------------------ 1080
pJJ2LB            TCTTGACCAACTCTATCAGAGCTTGGTTGACGGCAATTTCGATGATGCAGCTTGGGCGCA 1080

pBract204         ------------------------------------------------------------ 1140
pJJ2LB            GGGTCGATGCGACGCAATCGTCCGATCCGGAGCCGGGACTGTCGGGCGTACACAAATCGC 1140

pBract204         ------------------------------------------------------------ 1200
pJJ2LB            CCGCAGAAGCGCGGCCGTCTGGACCGATGGCTGTGTAGAAGTACTCGCCGATAGTGGAAA 1200

pBract204         ---- 1204
pJJ2LB            CCGA 1204
```

Supplementary Figure 1. Alignment of the two HygR cassettes used for sgRNA design.
Sequence alignment of the HygR cassettes contained in the plasmids pBract204 and pJJ2LB. The dashes represent the gaps in the alignment and the numbers at the right represent the position of the last nucleotide in the line relative to the first one in the alignment.

7. Supplementary Materials.

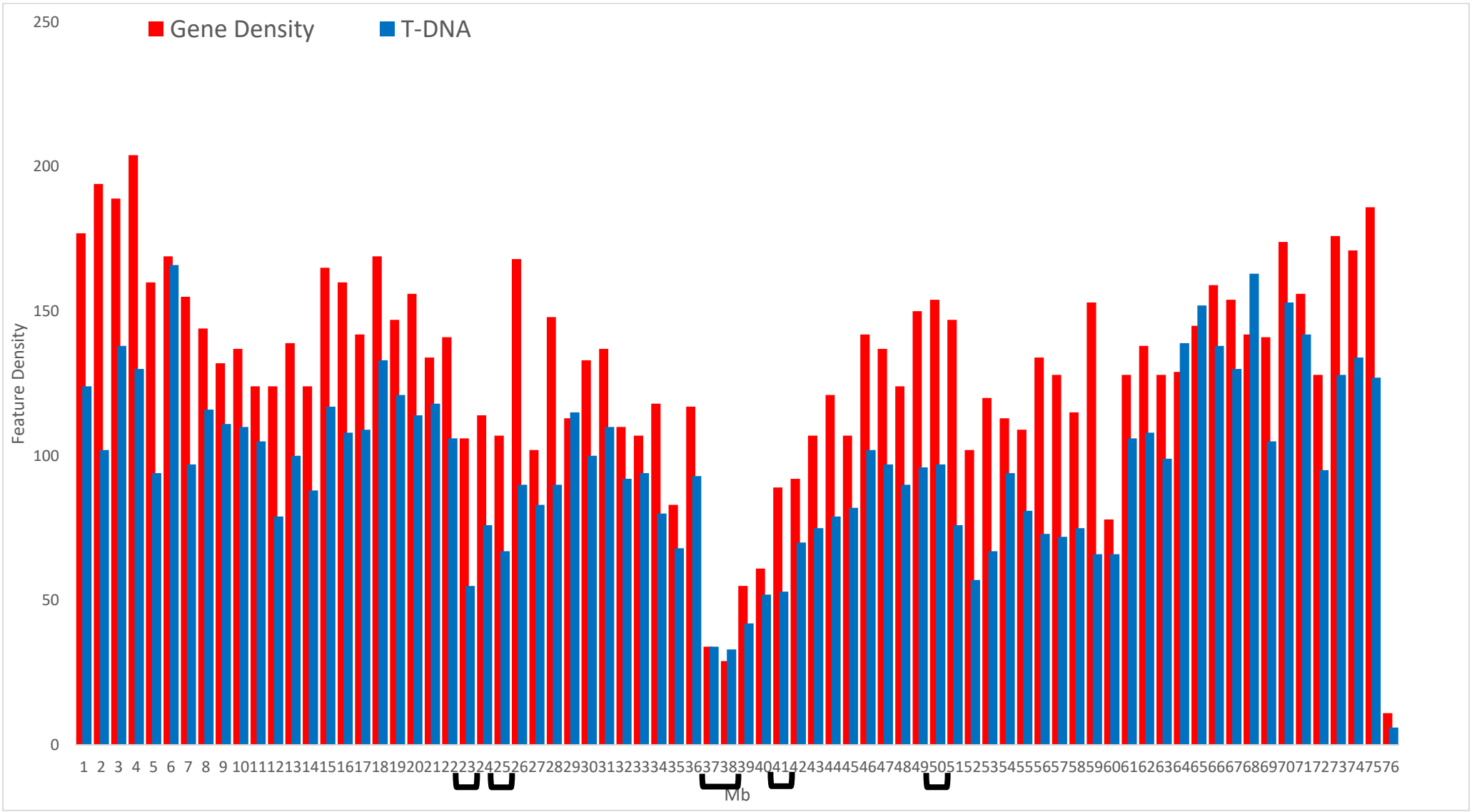
```
Jinek et al. 2013 v1.0   ----------------GTTTTAGAGCT--------AGAA----ATAGCAAGTTAAAATAA 60
pBdSGV                   ----------------GTTTTAGAGCT--------AGAA----ATAGCAAGTTAAAATAA 60
pHvDGV                    ----------------GTTTTAGAGCT--------AGAA----ATAGCAAGTTAAAATAA 60
Jinek et al, 2013 v2.1   ----------------GTTTTAGAGCTATGCTGAAAAGC----ATAGCAAGTTAAAATAA 60
pOsSGV                   ----------------GTTTTAGAGCTATGCTGAAAAGC----ATAGCAAGTTAAAATAA 60
Jinek et al. 2013 v2.2   GTTTTAGAGCTATGCTGTTTTGGAAAC--------AAAACAGCATAGCAAGTT-AAATAA 60

Jinek et al. 2013 v1.0   GGCTAGTCCG---------------------------------------------------- 120
pBdSGV                   GGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTT------------- 120
pHvDGV                   GGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCTAGACCCA 120
Jinek et al, 2013 v2.1   GGCTAGTCCGTTATC----------------------------------------------- 120
pOsSGV                   GGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC---------------- 120
Jinek et al. 2013 v2.2   GGCTAGTCCGTTATC----------------------------------------------- 120

Jinek et al. 2013 v1.0   ---------------------------------- 153
pBdSGV                   ---------------------------------- 153
pHvDGV                   GCTTTCTTGTACAAAGTTGGCATTACGCTTTAC 153
Jinek et al, 2013 v2.1   ---------------------------------- 153
pOsSGV                   ---------------------------------- 153
Jinek et al. 2013 v2.2   ---------------------------------- 153
```

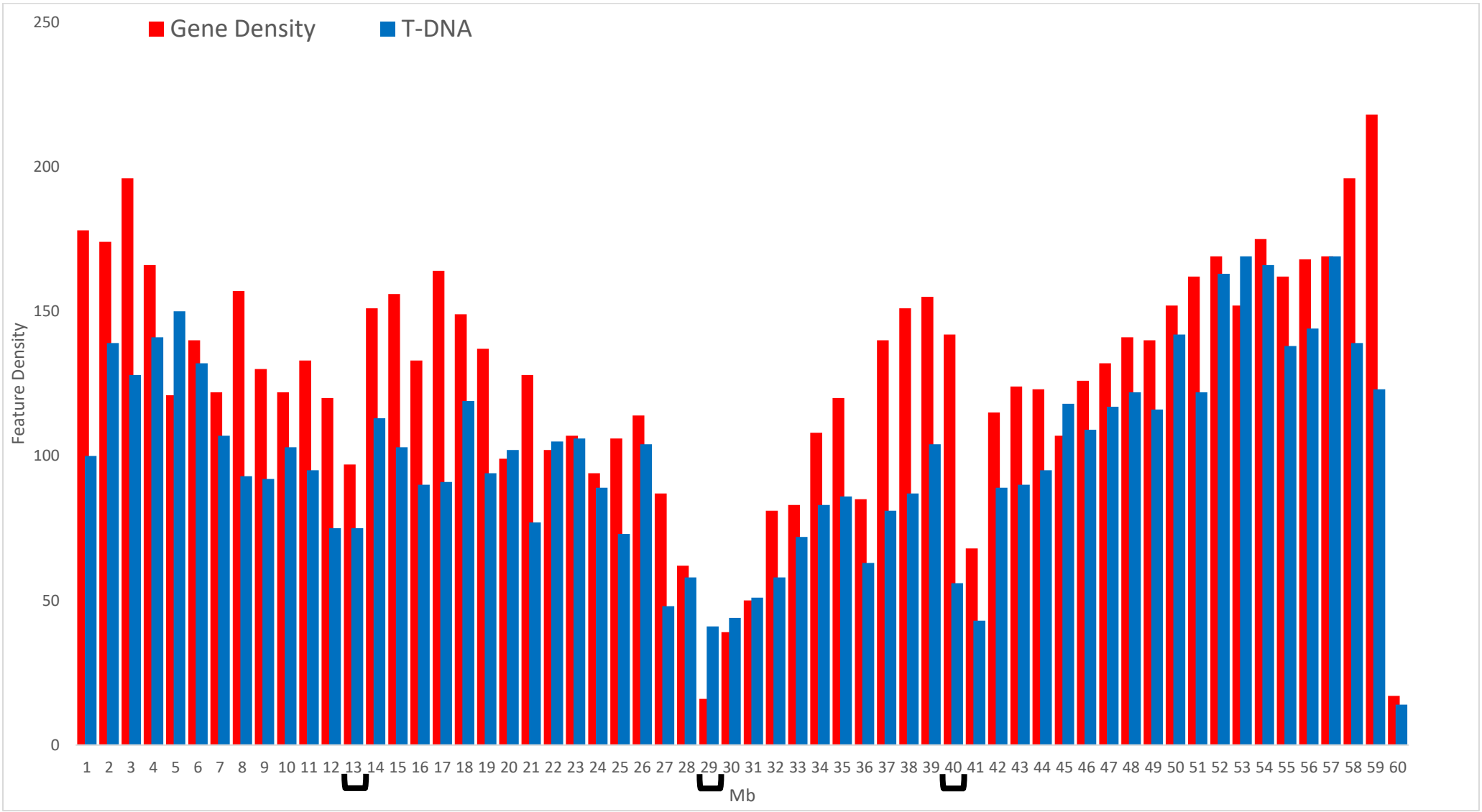
Supplementary Figure 2. Alignment of the different tracrRNA parts of the sgRNA cassettes of the plasmids used in Chapter 3.

Sequence alignment of the tracrRNA sequences contained in pBdSGV, pHvDGV and pOsSGV with the ones described in Jinek et al. 2013. The dashes represent the gaps in the alignment and the numbers at the right represent the position of the last nucleotide in the line relative to the first one in the alignment.



Supplementary Figure 3. Gene density and T-DNA insertion frequencies vary along the first chromosome in *Brachypodium*.
Histogram showing the feature frequency of transcripts (red) and T-DNA insertions (blue) in the first Chromosome of *Brachypodium*. Data showed as feature frequency in each one of the megabases of the chromosome. Black brackets represent the centromeric repeats identified in Vogel et al. (2010).

7. Supplementary Materials.



Supplementary Figure 4. Gene density and T-DNA insertion frequencies vary along the second chromosome in *Brachypodium*.
Histogram showing the feature frequency of transcripts (red) and T-DNA insertions (blue) in the second Chromosome of *Brachypodium*. Data showed as feature frequency in each one of the megabases of the chromosome. Black brackets represent the centromeric repeats identified in Vogel et al. (2010).

7.2. Supplementary tables.

Supplementary Table 1. Expression systems utilised in different CRISPR/Cas9 editing publications.

Reference	Species	Editing System		
		sgRNA Promoter	Cas9 Promoter	Cas9 Optimization
Shan et al., 2014	Rice	Os U3	2x35S	Rice optimised
	Wheat	Ta U6		
Zhang et al., 2016	Wheat	Ta U6	2x35S	Rice optimised
Liang et al., 2017	Wheat	Ta U6	2x35S	Rice optimised
Shan et al., 2013	Rice	Os U3	2x35S	Rice optimised
	Wheat	Ta U6		
Liang et al., 2014	Maize	Zm U3	2x35S	Rice optimised
Svitashev et al., 2015	Maize	Zm U6	Zm Ubiquitin 1	Maize optimised
Ma et al., 2017	Rice	Os U3B	Zm Ubiquitin 1	Rice optimised
Miao et al., 2013	Rice	Os U3B	Zm Ubiquitin 1	Rice optimised
Jiang et al., 2013	Arabidopsis	At U6-26	35S	*Chlamydomonas reindhartii* optimised
	Nicotiana benthamiana			
	Rice	Os U6	35S	*Streptococcus pyogenes* Cas9
				Rice optimised
	Sorghum	Os U6	Os Actin 1	Monocot Optimized
Lawrenson et al., 2015	Barley	Ta U6	Zm Ubiquitin 1	*Streptococcus pyogenes* Cas9
	Brassica oleracea	At U6-26	CsVMV	
Van den Schuren et al., 2018	*Brachypodium*	Bd U6	Zm Ubiquitin	*Brachypodium* optimised
Raissig et al., 2016	*Brachypodium*	Os U3B	Zm Ubiquitin 1	Rice optimised
Chandrasekaran et al., 2016	Cucumber	At U6	35S	Plant optimised
Brooks et al., 2014	Tomato	At U6	35S	Human optimised
Tian et al., 2017	Watermelon	At U6	2x35S	Maize optimised
Jia et al., 2016	Grapefruit	35S	35S	Human optimised
Jia and Nian, 2014	Orange	35S	35S	Human optimised
Ren et al., 2016	Grape	At U6	35S	*Streptococcus pyogenes* Cas9
Nishitani et al., 2016	Apple	At U6-1	2x35S	Plant optimised
Du et al., 2016	Soybean	At U6-26	Zm Ubiquitin	Dicot optimised
		Gm U6-16g-1		
Andersson et al., 2017	Potato	At U6 /	35S	Plant optimised
		St U6		
Wang et al., 2017	Rice	Os U6	Zm Ubiquitin	FnCpf1
				LbCpf1
Wang et al., 2018a	Grape	At U3B	2x35S	Plant optimised
		At U6-1		
		At U6-29		
		At U6d		
Ma et al., 2015	Rice	Os U6a	2x35S	Plant optimised
		Os U6b		
		Os U6c		
		Os U3		
	Arabidopsis	At U3b		
		At U6-29		
		At U3d		
		At U6-1		
Wang et al., 2018b	Kiwi	At U6	35S	*Streptococcus pyogenes* Cas9
		At U6 (PTG)		
Okuzaki et al., 2018	*Brassica napus*	At U6	Pc Ubiquitin 4-2	*Arabidopsis* optimised
Mikami et al., 2015	Rice	Os U3/Os U6	2x35S	Rice optimised
				Human optimised
			Pc Ubiquitin 4-2	*Arabidopsis* optimised
Fauser et al., 2014	Arabidopsis	At U6-26	Pc Ubiquitin 4-2	*Arabidopsis* optimised
Johansen et al., 2019	Potato	At U6	PPDK	*Streptococcus pyogenes* Cas9
		St U6		

7. Supplementary Materials.

Supplementary Table 1. Expression systems utilised in different CRISPR/Cas9 editing publications (cont.).

Reference	Species	Editing System		
		sgRNA Promoter	Cas9 Promoter	Cas9 Optimization
Lin et al., 2018	Bamboo	Os U3/Os U6	2x35S	Human optimised
	Setaria italica			
	Rice			
	Arabidopsis			
	Nicotiana tabacum			
	Tomato			
	Maize			
	Brassica oleracea			
	Brassica napus			
Xing et al., 2014	Maize	At U6-1	2x35S	Maize optimised
		At U6-29		
		At U6-26		
		Os U3		
	Arabidopsis	Ta U3		
		Ta U3b		
Hooghvorst et al., 2019	Melon	At U6	2x35S	Maize optimised
Arndell et al., 2019	Wheat	Ta U6	Zm Ubiquitin 1	Rice optimised
Cui et al., 2019	Wheat	Ta U6	35S	*Chlamydomonas reindhartii* optimised
				Plant codon optimised
Yuan et al., 2019	Peanut	Mt U6	2x35S	Human optimised
Jacobs et al., 2015	Soybean	Mt U6	2x35S	Human optimised
Bernard et al., 2019	Chicory	Ci U6	2x35S	Plant optimised
Hudzieczek et al., 2019	*Silene latifolia*	At U6	35s	*Arabidopsis* optimised
Zhang et al., 2018	Wheat	Ta U6	Zm Ubiquitin 1	Plant opt Cas9
		Ta U3		
Feng et al., 2016	Maize	Zm U3	2x35S	Human optimised
Feng et al., 2013	Arabidopsis	At U6-26	2x35S	Human optimised
	Rice	Os U6-2		
Gao et al., 2015	*Nicotiana tabacum*	At U6	2x35S	*Nicotiana* optimised
Zhu et al., 2016	Maize	Zm U6	Zm Ubiquitin 2	Maize optimised
Osakabe et al., 2016	Arabidopsis	At U6-1	2x35S	Fungal and plant codon optimised

Os : *Oryza sativa* (rice)
Ta : *Triticum aestivum* (wheat)
Zm : *Zea mays* (maize)
At : *Arabidopsis thaliana*
Bd : *Brachypodium distachyon*
35S : Cauliflower Mosaic Virus 35S
St : *Solanum tuberosum* (potato)
Gm : *Glycine max* (soybean)

PTG : Polycistronic tRNA-sgRNA cassette
Mt : *Medicago truncatula*
Ci : *Cichorium intybus* (chicory)
CsVMV : *Cassava Vein Mosaic Virus*
Pc : *Petroselinum crispum*
PPDK : Pyruvate phosphate dikinase
Fn : *Francisella novicida*
Lb : *Lachnospiraceae bacterium*

7. Supplementary Materials.

Supplementary Table 2. Tissues, transformation and screening methods used for CRISPR/Cas9 editing publications.

Reference	Species	Tissue	Transformation	Material	Screening
Shan et al., 2014	Rice	Protoplasts	PEG	Plasmid	PCR-RE*
	Wheat	Protoplasts		Plasmid	T7EI
	Rice	Callus	Biolistics	Plasmid	PCR-RE*
Zhang et al., 2016	Wheat	Protoplasts	PEG	Plasmid	PCR-RE
		Callus	Biolistics	Plasmid	
				Plasmid	
			Biolistics	RNA	
Liang et al., 2017	Wheat	Protoplasts	PEG	Plasmid	PCR-RE
			PEG	RNP	
		Embryos	Biolistics	Plasmid	Deep Sequencing
			Biolistics	RNP	
Shan et al., 2013	Rice	Protoplasts	PEG	Plasmid	PCR-RE
	Wheat			Plasmid	
	rice	Callus	Biolistics	Plasmid	PCR-RE
Liang et al., 2014	Maize	Protoplasts	PEG	Plasmid	PCR-RE/Enrichment PCR
Svitashev et al., 2015	Maize	Embryos	Biolistics	Plasmid	Deep Sequencing
				Plasmid (Cas9)/RNA (sgRNA)	Deep sequencing
				RNA (sgRNA)	Deep Sequencing
Ma et al., 2017	Rice	Callus	Agrobacterium	T-DNA	qPCR/Sequencing
Miao et al., 2013	Rice	Callus	Agrobacterium	T-DNA	Sequencing
		Callus	Biolistics	Plasmid	Sequencing
Jiang et al., 2013	Arabidopsis	Leaf	Agrobacterium	T-DNA	PCR-RE
	Nicotiana benthamiana	Leaf	Agrobacterium	T-DNA	PCR-RE
	Sorghum	Embryos	Agrobacterium	T-DNA	Phenotype
	Rice	Protoplasts	PEG	Plasmid	PCR-RE
Lawrenson et al., 2015	Barley	Embryos	Agrobacterium	T-DNA	Enrichment PCR/Sequencing
	Brassica oleracea	Cotyledon	Agrobacterium	T-DNA	Enrichment PCR/Sequencing
Van den Schuren et al., 2018	*Brachypodium*	Callus	Agrobacterium	T-DNA	Sequencing
Raissig et al., 2016	*Brachypodium*	Callus	Agrobacterium	T-DNA	Sequencing
Chandrasekaran et al., 2016	Cucumber	Callus	Agrobacterium	T-DNA	PCR-RE/Sequencing
Brooks et al., 2014	Tomato	Cotyledon	Agrobacterium	T-DNA	Sequencing
Woo et al., 2015	*Nicotiana attenuata*	Protoplasts	PEG	RNP	T7EI/Deep sequencing
	Arabidopsis				
	Rice				
	Lettuce				RFLP**
Tian et al., 2017	Watermelon	Protoplasts	PEG	Plasmid	PCR-RE/Sequencing
		Cotyledon	Agrobacterium	T-DNA	
Jia et al., 2016	Grapefruit	Leaves	Xcc-facilitated Agroinfiltration	T-DNA	Sequencing
		Epicotyl	Agrobacterium	T-DNA	
Jia and Nian, 2014	Orange	Leaves	Xcc-facilitated Agroinfiltration	T-DNA	Enrichment PCR/RE-PCR***
Ren et al., 2016	Grapes	Callus	Agrobacterium	T-DNA	Surveyor Assay/Sequencing
Nishitani et al., 2016	Apple	Leaves	Agrobacterium	T-DNA	Sequencing
Du et al., 2016	Soybean	Hairy Root	Agrobacterium****	T-DNA	PCR-RE/T7E1/Cruiser
		Cotyledon	Agrobacterium	T-DNA	PCR-RE/T7E1/Cruiser
Sauer et al., 2016	Arabidopsis	Protoplasts	PEG	ssODN	NGS
	Flax			ssODN and CRISPR/Cas9 (plasmid)	
Andersson et al., 2017	Potato	Protoplasts	PEG	Plasmid	Sequencing

Supplementary Table 2. Tissues, transformation and screening methods used for CRISPR/Cas9 editing publications (cont.).

Reference	Species	Tissue	Transformation	Material	Screening
Wang et al., 2018	Cotton	Callus	Agrobacterium	T-DNA	T7E1/Sequencing
Wang et al., 2018b	Kiwi	Callus	Agrobacterium	T-DNA	T7E1/Sequencing
Okuzaki et al., 2018	*Brassica napus*	Hypocotyls	Agrobacterium	T-DNA	PCR-RE/Sequencing
Kim et al., 2018	Wheat	Protoplasts	PEG	Plasmid	Enrichment PCR/T7E1/Sequencing
Johanssen et al., 2019	Potato	Protoplasts	PEG	Plasmid	IDAA/Sequencing

*This reference acknowledges the use of Enrichment PCR for the screening of edits.
** In regenerated callus.
***Enrichment PCR was used for the screening of edited samples, RE-PCR was used for efficiency assays.
****_Agrobacterium rhizogenes._

7. Supplementary Materials.

Supplementary Table 3. Primer table

Primer ID	Sequence (5' → 3')
pActH2BGFP Rv GG	GCTCTTCTCTGCCCAAATGTTTGAACGATCTGC
pActH2BGFP Fw GG	GCTCTTCCTGCACAATGGCGAAGGCAGATAAG
HvDGV Ubi Rv GG	TTGTTGGCTCTTCTGCAGAAGTAACACCAAACA
HvDGV NOS Fw GG	CAACAAGCTCTTCGCAGATCGTTCAAACATTTGG
H2BGFP PCR Scr Fw	TCTCAAGCAAAGCGATGGGA
GFP qPCR Rv	AAGAAGATGGTGCGCTCCTG
dCas9 Check Fw	ACGTGAGTTCAGCAACGGTCTA
dCas9 Check Rv	ACCGGCAACAGGATTCAATC
HygR CDS Fw	TCGGTTTCCACTATCGGCGAGT
HygR CDS Rv	GAGGGCGAAGAATCTCGTGCTTTC
HygR Nd1 Fw	TATCGGCACTTTGCATCGGC
HygR Nd1 Rv	ATTGCCGTCAACCAAGCTCT
HvDGV gRNA GG Fw Bf	CAACAAGCTCTTCGGACACGAAGTGATCCG
HvDGV gRNA GG Rv Bf	TTGTTGGCTCTTCAGCGTCGATCTAGTAACATAGATG
pDON BdgRNA GG Fw Bf	CAACAAGCTCTTCTCGCTACTTGGGCTGTTGCTCTCT
pDON BdgRNA GG Rv Bf	TTGTTGGCTCTTCTGTCAAAAAAAAGCACCGACTCG
pENTR OsgRNA GG Fw Bf	CAACAAGCTCTTCTCGCAAGGGATCTTTAAACATACGAACA
Dual Guide Check Fw	CTGGCCAGCCACTATGAAAAGC
Dual Guide Check Rv	TCCGTCGATCACTACGAAGTCG
NOSt Seq Fw	TTGGCAATAAAGTTTCTTAA
Dual Guide Check Rv	TCCGTCGATCACTACGAAGTCG
Guide Check Fw	ACGTGAGTTCAGCAACGGTCTA
Dual Guide Check Rv	TCCGTCGATCACTACGAAGTCG
NOSt Seq Fw	TTGGCAATAAAGTTTCTTAA
LacZ Rv	CCAGCTGCATTAATGAATCG
BdDWF4 Fw	CCCACAAGGCTACAACCAC
BdDWF4 Rv	CGGAGAAACTCTGTCGTCG
BdDWF4 Nd1 Fw	TATAAGAGCGCACCGCAACA
BdDWF4 Nd1 Rv	AGCATCGACCACTTTCCCAG
BdDWF4 Nd2 Fw	TTGCACAAAGCTTGCACAGA
BdDWF4 Nd2 Rv	CGGTATATCTTCCCGTACCGC
pBHygR Fw	TTTGGAGAGGACGACCCCGA
pBHygR Rv	TCGGCCCAAAGCATCAGCTC
HygR Nd Rv2	CGGTGTCGTCCATCACAGTT
HygR Nd Rv3	ATTGACCGATTCCTTGCGGT

7. Supplementary Materials.

Supplementary Table 4. sgRNA design targeting *BdCKX2.*

Primers	
Strand	*Sequence (5' → 3')*
Fw	CGCTCGAAGACACTCCCAACAAAG
Rv	GGCTACACGACAAACGCACGT

Spacers							
Strand	*Sequence (5' → 3')*	*PAM*	*Position*	*Specificity Score*	*Efficiency Score*	*Enzyme*	
Rv	TCGTGAGAAAACTGCTAAGC	AGG	182	49.28	52.91	BlpI	
Rv	TGCCGGCGCGTCGTCGCCCA	GGG	268	43.83	44.03	PasI	
Rv	GACGGGTACAGCACGGCTTG	AGG	372	47.69	58.19	BpuEI	
Fw	GTGCTGTACCCGTCGAGCCC	GGG	392	49.09	52.12	SrfI	
Rv	CCAGCGCCGCGATGTCGCCC	GGG	398	77.33	42.55	SrfI	
Fw	CGCCGGGGGCGAGCAGCTGT	GGG	616	45.11	47.72	PvuII	Exon 1
Rv	GGGGGCCGTGGCGGAAGGCC	TGG	734	29.69	34.13	StuI	
Rv	GTAGGACGTTGGAGATCTGG	GGG	752	46.36	65.43	PsuI	
Rv	TGTAGGACGTTGGAGATCTG	GGG	753	46.05	60.55	PsuI	
Rv	CTGTAGGACGTTGGAGATCT	GGG	754	48.50	32.84	PsuI	
Rv	GCTGTAGGACGTTGGAGATC	TGG	755	47.61	31.83	PsuI	
Rv	CGGTGATGACGTCGAGCTGT	AGG	770	47.49	59.81	SfcI	
Fw	GGAGATGGTGACGTGTTCGA	AGG	986	48.67	60.25	BstBI	
Fw	GATGGTGACGTGTTCGAAGG	AGG	989	48.75	62.42	BstBI	
Fw	GGAGCGCGCGGACCTGTTCG	AGG	1010	49.16	54.02	XagI	
Rv	CGCCCAGCACCGCCTCGAAC	AGG	1011	45.43	41.90	XagI	
Fw	GCGCGCGGACCTGTTCGAGG	CGG	1013	47.19	67.82	XagI	
Fw	GGCAGTTCGGGGTGATCACG	CGG	1051	47.83	52.59	BclI	
Fw	CATCCGCCTGGCACCGGCGC	CGG	1079	45.02	29.59	SgrAI	Exon 2
Rv	CTGGTCGCCCGTCAGGGACG	CGG	1127	48.07	59.55	DrdI	
Fw	GGGGCTCTTGGACTACGTCG	AGG	1193	48.93	57.85	PsyI	
Fw	GGGCTCTTGGACTACGTCGA	GGG	1194	49.19	60.09	PsyI	
Fw	GGCGTGCTCTACTGCCTCGA	GGG	1329	48.00	53.14	XhoI	
Rv	CGTAGTAGAGCGCGCCCTCG	AGG	1332	46.99	58.28	XhoI	
Rv	AACGTCGGACTCGCCGTCGG	CGG	1364	44.67	61.02	AhdI	
Rv	GGCAACGTCGGACTCGCCGT	CGG	1367	45.26	55.31	AhdI	

7. Supplementary Materials.

Supplementary Table 5. sgRNA design targeting *BdGA20ox2.*

Primers	
Strand	*Sequence (5' → 3')*
Fw	CAAGACAGCTCTCCTGCACACTC
Rv	ACAGAGCAGAGCGAGAACAGAGT

Spacers

Strand	*Sequence (5' → 3')*	*PAM*	*Position*	*Specificity Score*	*Efficiency Score*	*Enzyme*	
Rv	AGGTGGATACGGCAGCTGGG	AGG	163	48.32	69.55	PvuI	Exon 1
Rv	GTGAGGTGGATACGGCAGCT	GGG	166	48.35	58.31	PvuI	
Rv	AGTGAGGTGGATACGGCAGC	TGG	167	47.98	48.91	PvuI	
Rv	TCGCTGCAGTGAGGTGGATA	CGG	174	48.78	55.2	BfuI	
Rv	TGCTGGGAGCGGAAGGGCTG	AGG	236	45.72	46.67	BbvCI	
Fw	GCCGACCACGGCCGAGGAGC	TGG	379	42.69	32.35	BseRI	
Rv	CACTGGCATGCCCAGCTCCT	CGG	379	48.77	48.72	BseRI	
Fw	CCGACCACGGCCGAGGAGCT	GGG	380	45.77	43.58	BseRI	
Fw	GTGCTCCGGAAGAGCCGCAA	TGG	425	94.62	53.76	BglI	
Fw	CTCCGGAAGAGCCGCAATGG	CGG	428	92.76	56.02	BglI	
Rv	CGGCGTCGCCGCCGCCATTG	CGG	428	55.01	45	BglI	
Fw	CGGAAGAGCCGCAATGGCGG	CGG	431	80.92	39.9	BglI	
Rv	CCACCTGCGCCACGGCGCGC	CGG	458	74.16	33.99	AscI	
Fw	CGCGGCGCTCCGGCGCGCCG	TGG	460	54.37	48.87	AscI	
Fw	CCGGCGCGCCGTGGCGCAGG	TGG	469	62.1	44.1	AarI	
Rv	GGCCGCGTCCACGCCGTGCC	CGG	517	44.87	37.46	AdeI	
Fw	CCAGGTGTCCGGGCACGGCG	TGG	520	45.76	61.53	AdeI	
Rv	CGTCCAGCGCCGCGCGCGCC	AGG	539	44.08	21.01	MauBI	
Fw	GGACGCGGCCCTGGCGCGCG	CGG	541	37.97	39.83	MauBI	
Rv	GCTTGTCCGCGAGCGGCATC	CGG	578	47.76	43.86	LweI	
Fw	CCACGGCCAAGACTGCGCAG	AGG	715	48.63	67.77	NsbI	
Fw	GGAGCTGCTGGAGCTGAGCC	TGG	956	44.2	20.42	BlpI	Exon 2
Fw	GAGCTGCTGGAGCTGAGCCT	GGG	957	46.05	49.8	BlpI	
Fw	CGCGCTCACCATCCTCCTGC	AGG	1103	45.87	50.67	SbfI	
Rv	CGCCGACATCGTCCTGCAGG	AGG	1104	45.15	58.18	SbfI	
Rv	ACCCGCCGACATCGTCCTGC	AGG	1107	47.45	50.37	SbfI	
Fw	ATCCTCCTGCAGGACGATGT	CGG	1113	48.56	61.36	PsyI	
Fw	CTCCTGCAGGACGATGTCGG	CGG	1116	45.45	61.48	PsyI	
Rv	GGCGCCAGTCGCCGTCGACG	AGG	1137	39.56	57.23	SgrDI	
Rv	CGCCGGGGACGGGGCGGACG	GGG	1158	32.62	28.56	EciI	
Rv	GCGCCGGGGACGGGGCGGAC	GGG	1159	42.02	29.25	EciI	
Rv	GGCGCCGGGGACGGGGCGGA	CGG	1160	37.71	23.52	EciI	

7. Supplementary Materials.

Supplementary Table 6. sgRNA design targeting *BdRVR1.*

Primers	
Strand	*Sequence (5' → 3')*
Fw	TCAGGCACAACTTCCCCAAT
Rv	CTCCGATGCTCCCCTCACC

Spacers

Strand	*Sequence (5' → 3')*	*PAM*	*Position*	*Specificity Score*	*Efficiency Score*	*Enzyme*	
Fw	GAGAAATGGGAAGGCGCCGC	CGG	102	47.53	55.12	KasI	Exon 1
Rv	CTGCTTCTTGGCACCCGAGG	AGG	196	47.67	53.89	AvaI	
Rv	CGGCTGCTTCTTGGCACCCG	AGG	199	48.28	50.99	AvaI	
Rv	CGCGCTTGGGCTTGCGGCCG	CGG	365	44.52	23.27	EagI	
Rv	GCTCTCCTCGACGGCGCGCT	TGG	379	77.6	29.7	PauI	
Fw	GGGGATCCCGTAAAGGTCAC	CGG	503	49.71	55.95	BstEII	
Rv	GGCGCAGCTTACGAGTTCGA	AGG	568	49.62	66.02	BstBI	
Rv	GGCTTCTCGTTCTTCTGCTC	AGG	691	46.01	40.71	Bpu10I	Exon 2
Fw	GATGGGAGCTTATCAGTAAC	TGG	1324	47.18	46.33	CaiI	Exon 3
Fw	GCCAGAAGAGGCTGACAAAA	AGG	1362	46.86	32.48	CspCI	
Fw	CCAGAAGAGGCTGACAAAAA	GGG	1363	44.74	44.99	CspCI	
Fw	GTCAGTCATGCACAAATGTG	TGG	1458	42.64	64.95	AleI	
Fw	CAAGTCGAAAGGAACACCCT	GGG	1511	48.97	64.13	PasI	
Rv	CCTTTTGCACAATAAACCCA	GGG	1516	49.27	68.72	PasI	
Fw	TCCCAAATAAACGAGGCCTA	CGG	1712	48.34	59.26	StuI	
Fw	TCTAAAGCATTACGCCATCC	TGG	1824	49.82	48.17	BglI	
Rv	TACTTTGAATTTGGCCAGGA	TGG	1827	48.44	43.94	BglI	

Supplementary Table 7. sgRNA design targeting *BdVRN2*

Primers	
Strand	*Sequence (5' → 3')*
Fw	AGTCCAAATCTATGCACCAAGTGGA
Rv	GGGTGCAGGTGAAAAATGTTCCTCAG

Spacers

Strand	*Sequence (5' → 3')*	*PAM*	*Position*	*Specificity Score*	*Efficiency Score*	*Enzyme*	
Rv	GCTCATCGGGAGGCTGGGAC	TGG	839	48.91	43.27	BseYI	Exon 11
Fw	GAGCGCTTTTATGGACCTGA	AGG	869	49.12	56.87	AcuI	
Rv	ACCATCACACCAGTGCCTTC	AGG	873	49.21	37.49	AcuI	
Fw	ATTCAACATGGGGTCGACAG	TGG	1068	49.26	69.3	SacI	Exon 12
Fw	CGAAGCACCGTTGTCCAAGA	AGG	1101	48.82	60.01	XcmI	
Rv	GCCGTCCTCCATTGCCTTCT	TGG	1104	47.8	16.81	XcmI	
Fw	ACCGTTGTCCAAGAAGGCAA	TGG	1107	47.16	60.94	XcmI	
Fw	GTTGTCCAAGAAGGCAATGG	AGG	1110	47.34	70.96	XcmI	
Fw	GGAGGACGGCATGCCCACGT	CGG	1128	47.82	65.18	AjiI	
Rv	CAGAAACCGAAATCCGACGT	GGG	1130	49.65	72.09	AjiI	
Fw	GGATTTCGGTTTCTGCGTCA	AGG	1149	48.53	52.38	CseI	
Fw	TAGGAGAAGCCATTGGGCGA	TGG	1190	48.56	51.36	BtgZI	

7. Supplementary Materials.

Supplementary Table 8. sgRNA design targeting *BdDWF4.*

Primers	
Strand	*Sequence (5' → 3')*
Fw	CCCCACAAGGCTACAACCACC
Rv	ACGGAGAAACTCTGTCGTCGCA

Spacers							
Strand	*Sequence (5' → 3')*	*PAM*	*Position*	*Specificity Score*	*Efficiency Score*	*Enzyme*	
Rv	GGTTATGGACGCCATCATGG	CGG	313	46.22	71.56	BglI	Exon 1
Rv	GCTGGTTATGGACGCCATCA	TGG	316	48.61	52.35	BglI	
Rv	GAGGAAGAAGAGGAGCTCGC	TGG	334	37.07	53.61	SacI	
Rv	AGAAGGTGAGCAGGGCCAGG	AGG	365	44.9	66.66	EcoRIII	
Fw	CACCTTCTACACCACCACGG	TGG	391	44.79	73.17	BtgI	
Rv	GTGGCATTTCGCCACCGTGG	TGG	391	48.61	62.54	BtgI	
Fw	TTCCCCCGGGCGCCATCGGC	TGG	465	46.98	36.16	PflMI	
Rv	GCCGATGAAGGGCCAGCCGA	TGG	466	48.56	48.96	PflMI	
Fw	CCCTTCATCGGCGAGACCTT	CGG	488	49.34	38.63	BsaI	
Fw	TCATGGAGGAGCACATCGCG	CGG	549	46.95	73.83	BtgZI	
Fw	CCTGTTCGGGGACCGGACGG	TGG	595	48.48	56.94	DrdI	
Rv	CGTCCGCCGACACCACCGTC	CGG	596	44.59	55.33	DrdI	
Fw	CGGGGACCGGACGGTGGTGT	CGG	601	46.41	46.51	DrdI	
Fw	GGACCGGACGGTGGTGTCGG	CGG	604	45.12	56.58	DrdI	
Rv	ACAGCTTCCCTTCGTTCTGC	AGG	638	49.42	45.97	PstI	
Fw	GCGCAGCATCGGGGGGATAC	TGG	688	49.31	30.11	BfuI	
Fw	ACTGGGAAAGTGGTCGATGC	TGG	706	48.73	53.87	BmrI	
Fw	CCGCGCCGTGCTCCTCCCGG	AGG	796	40.7	61.46	PfoI	
Rv	TGTGGCGCTCGACCTCCGGG	AGG	797	47.97	69.6	PfoI	
Rv	GCGTGTGGCGCTCGACCTCC	GGG	800	48.03	46.59	XagI	
Rv	TTCGTGCTGGGCGGAGAAGA	CGG	859	43.85	48.86	BbsI	
Fw	GGACCCGGGGGAGGAGGAGA	CGG	1038	29.23	43.54	Esp3I	Exon 2
Fw	GGAGTACATCACCTTCATGA	AGG	1074	47.59	61.59	PagI	
Rv	GGACACCACCCCCTTCATGA	AGG	1074	48.23	65.26	PagI	
Fw	GGCCTACTGGAAGGCCCTCA	AGG	1131	48.91	55.33	SmoI	
Rv	CGTGTGGGTGTGTACCTTGA	GGG	1134	48.34	49.45	SmoI	

7. Supplementary Materials.

Supplementary Table 9. sgRNA design targeting the HygR contained in pBract204

Primers	
Strand	*Sequence (5' → 3')*
Fw	TTTGGAGAGGACGACCCCGA
Rv	TCGGCCCAAAGCATCAGCTC

Spacers						
Strand	*Sequence (5' → 3')*	*PAM*	*Position*	*Specificity Score*	*Efficiency Score*	*Enzyme*
Rv	AAACTTCTCGACAGACGTCG	CGG	47	97.92	70.64	BsaHI
Rv	ATGTCAAGCACTTCCGGAAT	CGG	244	97.31	39.24	Kpn2I
Rv	ACGTGACACCCTGTGCACGG	CGG	303	97.31	73.35	SduI
Fw	TTGCATCTCCCGCCGTGCAC	AGG	305	94.03	37.13	SduI
Fw	TGCATCTCCCGCCGTGCACA	GGG	306	86.59	51.46	SduI
Rv	GCAACGTGACACCCTGTGCA	CGG	306	99.26	62.45	AdeI
Rv	AAATTTACCTGCAGAACAGC	GGG	351	86.3	52.3	MspA1I
Fw	ATGATAACTGCAGCCGGTCG	CGG	554	97.59	52.65	Cfr10I
Rv	GGCCGCAGCGATCGCATCCA	TGG	569	92.02	49.58	BseGI
Fw	GGCCATGGATGCGATCGCTG	CGG	578	93.99	59.61	SfaAI
Rv	CCCGCTCGTCTGGCTAAGAT	CGG	590	96.53	51.39	DdeI
Fw	CCGATCTTAGCCAGACGAGC	GGG	601	95.16	58.05	MbiI
Rv	CCGATTCCTTGCGGTCCGAA	TGG	619	95.87	44.48	CpoI
Fw	TTCGGCCCATTCGGACCGCA	AGG	624	94.46	57.13	CpoI

Supplementary Table 10. sgRNA design targeting the HygR contained in pJJ2LB.

Primers	
Strand	*Sequence (5' → 3')*
Fw	TTTGGAGAGGACGACCCCGA
Rv	TCGGCCCAAAGCATCAGCTC

Spacers						
Strand	*Sequence (5' → 3')*	*PAM*	*Position*	*Specificity Score*	*Efficiency Score*	*Enzyme*
Rv	GCAACGTGACACCCTGTGCA	CGG	198	99.26	62.45	AdeI
Fw	CGCTGTTCTACAACCGGTCG	CGG	257	92.81	52	AgeI
Fw	GGCTATGGATGCGATCGCTG	CGG	281	95.52	58.2	SgfI
Rv	CCCGCTCGTCTGGCTAAGAT	CGG	293	96.53	51.39	DdeI
Rv	CCGATTCCTTGCGGTCCGAA	TGG	322	95.87	44.48	CpoI
Fw	TTCGGCCCATTCGGACCGCA	AGG	327	94.46	57.13	CpoI
Fw	TGAGCTGATGCTTTGGGCCG	AGG	467	96.21	42.49	NmeAIII
Rv	AATCCGCGTGCACGAGGTGC	CGG	492	95.46	47.38	AdeI
Fw	GACTGGAGCGAGGCGATGTT	CGG	576	93.23	34.33	BtgZI
Rv	CGGCCTCCAGAAGAAGATGT	TGG	608	88.88	58.99	AjuI
Fw	AGGTCGCCAACATCTTCTTC	TGG	613	86.31	35.83	AjuI
Fw	CTTCTTCTGGAGGCCGTGGT	TGG	626	90.57	39.55	BtgI
Fw	TTGCAGGATCGCCACGACTC	CGG	697	98.4	44.48	SchI
Rv	AGCATATACGCCCGGAGTCG	TGG	697	96.65	57.78	SchI
Fw	TGCAGGATCGCCACGACTCC	GGG	698	97.62	49.83	SchI
Fw	CTCTATCAGAGCTTGGTTGA	CGG	750	94.33	48.8	HincII
Rv	TTGTGTACGCCCGACAGTCC	CGG	821	98.87	52.63	FaqI

7. Supplementary Materials.

Supplementary Table 11. Plant Protoplast Dataset

Reference	*Species*	*Target*	*Screening*	*Efficiency (%)*	Spacer Sequence *(5' → 3') (PAM)*
Shan et al., 2014	Rice	*PDS*	PCR-RE	33	GTTGGTCTTTGCTCCTGCAG(AGG)
		DEP1		38	CGGCCGGCGGCGGATGCAGC(TGG)
	Wheat	*LOX2*		45	GTGCCGCGCGACGAGCTCTT(CGG)
Liang et al., 2017	Wheat	*GW2-A1*	PCR-RE/Deep Sequencing	35.87	CAGGATGGGGTATTTCTAG(AGG)
		GASR7		48.6	GTTGCCGTAGGTGCCCGG(CGG)
Shan et al., 2013	Rice	*PDS*	PCR-RE	20	GTTGGTCTTTGCTCCTGCAG(AGG)
				14	CGTCCAACCCATTCCTCTGC(AGG)
		BADH2		27	GCAGATCTTGCAGAATCCT(TGG)
		Os02g23823		30	GCTGAAAGATTTGACGTCCC(AGG)
		MPK2		38	GCGGCGGCCATGGCCATCA(CGG)
	Wheat	*MLO*		29	GGAGATTGGGTCCTGCGTGA(CGG)
Liang et al., 2014	Maize	*IPK1*	PCR-RE	11.1	GGGAGAAGGAGACGGATCCC(TGG)
Jiang et al., 2013	Rice	*OsSWEET14* (promoter)	RE-PCR	87.5	GAGCTTAGCACCTGGTTGGA(GGG)
Tian et al., 2017	Watermelon	PDS	RE-PCR	51.6	ATGCCGCTAGAGTGGTGCC(CGG)
				42.1	AATGGAGAACAGCATCTCG(AGG)
Andersson et al., 2017	Potato	GBSS	HRFA	7.0	GATATTAGAATCACATAGGG(TGG)
				3.8	TGTTGACAAGGGTGTTGAAT(TGG)
				6.8	GACAAGAAGATCCCTTTGAT(TGG)
Kim et al., 2018	Wheat	*DREB2*	PCR-RE/T7E1	6.7	GCAGGACGTCGACGAGGACT(CGG)
		ERF3		10.2	GCGAGGGGCAAGCACTACCG(CGG)
Johansen et al., 2019	Potato	*GBSS*	IDAA	58	GGTCCTTGGAGCAAAACTGG(TGG)
				24	AGGGCTGTTAACAAGCTTGA(TGG)
Lin et al., 2018	Bamboo	*PDS*	PCR-RE	6.6	GATGGAGATTGGTATGAGAC(CGG)
	Setaria italica	*PDS*		10.2	GTTGGTCTTTGCTCCTGCAG(AGG)
	Rice	*PDS*		7.3	GTTGGTCTTTGCTCCTGCAG(AGG)
	Arabidopsis	*PDS*		6.5	GGACTTTTGCCAGCCATGGT(CGG)
	Nicotiana tabacum	*PDS*		15.00	GCCGTTAATTTGAGAGTCCA(AGG)
	Tomato	*PDS*		3.7	GCTGTTAACTTGAGAGTCCA(AGG)
	Maize	*IPK*		0.2	TGGGAGAAGGAGACGGATCCC(TGG)
				1.1	GGTCGGCGGCGTGGTCGAGCT(CGG)
	Brassica oleracea	*GA4a*		75.2	GTTGAGAGGGGAGCCGGTGA(TGG)
	Brassica napus	*GA4a*		56.8	GTTGAGAGGGGAGCCGGTGA(TGG)
Xing et al., 2014	Maize	*HTK1*	PCR-RE	33.8	GGCTTCGTGCCAACCAACGA(GGG)
Hooghvorst et al., 2019	Melon	*PDS*	PCR Sequencing	25.00	AGTGAGATTGTGGGCGAT(GGG)
				25	ATCATCTATCTGTGGTCTA(GGG)
Arndell et al., 2019	Wheat	*EPSPS* Homeoalleles	TIDE	1.3	GAACTGCAATGCGGCCACTGA(CGG)
				9.8	GCTACTACAGCTGCCGTCAG(TGG)
				1.8	GATCAGTCTCCAAACATACCCA(TGG)
				7	GAAATAAGTATGAGATCCAT(GGG)
				17.8	AGTCGAAAAGGACGCCAAAG(AGG)
				2.5	GTAATGCTGGAACTGCAATG(CGG)

Supplementary Table 11. Plant Protoplast Dataset (cont.).

Reference	*Species*	*Target*	*Screening*	*Efficiency (%)*	Spacer Sequence *(5' → 3') (PAM)*
Cui et al., 2019	Wheat	*ABCC6*	Deep Sequencing	8.72	GTAATGCTGGAACTGCAATG(CGG)
				8.88	CACGCCGTCGAGATTACTGG(AGG)
		LTP9.4		8.55	AGTACTCACGGAGATCCAAG(GGG)
				7.45	GCCGTGCGTGGCGTACGTGA(CGG)
		NFXL1		7.4	AGTGCTGCTCCGGCGTGCAG(GGG)
				17.35	TGACTGGCACAACGCAAGGT(GGG)
Yuan et al., 2019	Peanut	*FAD2*	PCR Sequencing	34.78	GATGGAGTTGGTGTGCCGCA(AGG)
Hudzieczek et al., 2019	*Silene latifolia*	*AP3*	Deep Sequencing	0.29	AAACACTTCGTCGCGGTCGA(GGG)
				20.39	TTGTACCATTTCCTTAGCCC(CGG)
Zhang et al., 2018	Wheat	*Pinb*	Barcode sequencing	4.57	AGACACCGGGGCTAAGGAAA(TGG)
		Pinb		4.37	GATGTCACTAAGCAATAAATAA(AGG)
		Pinb		3.86	GGAGATGAATAGATGGGTTG(AGG)
		DA1		6.81	GCTTAGCTTCGGCCGCTCCTG(CGG)
		waxy		1.74	ACTCATGAGATGATGCATGCA(TGG)
		waxy		2.56	ATCTCGGCGCCGACGAACACG(AGG)
		waxy		2.41	AGGCGGCCTCGGCGACGTCCT(CGG)
Feng et al., 2016	Maize	*Zmzb7*	PCR-RE/RE-PCR	27	AGACCAGTACAAGGACGCCT(GGG)
		Hsg3		8.2	CAAGATGCTATGTATCAGC(TGG)
		Hsg4		4.1	ACCGATAAGCACAGCAGCTG(TGG)
		Hsg6		26.8	GATGTCTTCATCATGGATCC(AGG)
		Hsg7		16.65	ACGAGAGCTGCAGGCGGCCA(TGG)
		Hsg12		16.35	ACCCTTGTTGTCGGATCCAG(CGG)
Gao et al., 2014	*Nicotiana tabacum*	*PDS*	PCR-RE	20.3	CGGCGTGGCGCCGGAGCTCA(CGG)
				7.7	GCCGTTAATTTGAGAGTCCA(AGG)
				19.8	GAGATTGTTATTGCTGGTGC(AGG)
				6.5	GAGGCAAGAGATGTCCTAGG(TGG)
		PDR6		9.5	GCTGCATGGAAAGATGATGA(TGG)
				16.2	GTAGATGCAGAAGCTAGAGT(TGG)
Zhu et al., 2016	Maize	n.s.	CTAB/Surveyor/Deep Sequencing	2.29	GTGGAATCATCAAACCAGGA(AGG)
				18.5	GAGACTTGAGGATCTGTTCA(CGG)
				1.04	GGAGCGGCGCCGCGCCATAT(GGG)
				5.22	GCCCAATAGCAAGTATACTG(TGG)
				9.33	GCATAGTCTTGATATTATTC(CGG)
				1.41	GGTGTACTTGAGATCATGGA(AGG)
				53.75	GAAAGATTGCTCCCATGTAC(AGG)
				48.27	GAGGCGTTCTCCAGATGCCA(TGG)
				2.4	GAGAACGCCTCGCATGCCCA(CGG)
				0.48	GCCCTACTGAATCGATATAT(CGG)
				7.63	GTTGACAAACTCTTCACGTC(AGG)
				2.96	GATCCTGCTCAATCTTTACA(TGG)
				2.31	GGATTACCTACTAAGTATAC(GGG)
				1.83	GTTAGACCTGATCTAGAAGT(GGG)
				3.9	GCCAGTCAACAATAATTGTA(AGG)
				5.58	GATGCCATTTAATTCCTACA(TGG)
				2.4	GAACCCTAGAATTGGTACCT(TGG)

Supplementary Table 11. Plant Protoplast Dataset (cont.).

Reference	*Species*	*Target*	*Screening*	*Efficiency (%)*	Spacer Sequence *(5' → 3') (PAM)*
Zhu et al., 2016	Maize	n.s.	CTAB/Surveyor/Deep Sequencing	20.98	GTGTCATGTTTGAACCATAT(GGG)
				5.31	GACTCACTATTCTGATCATA(GGG)
				3.69	GCGGGGACTCCGAGCGCTGC(TGG)
				2.16	GCGTCCTCCCATCCTCGCCA(TGG)
				32.86	GAACGGCGCTCCAGCAGCTG(TGG)
				11.09	GGTCTACAACCCCTTTGCAG(CGG)
				3.18	GCAACACTACTGCTGCTAGC(AGG)
				15.92	GCGCCTGCACCTCCGGAGGT(TGG)
				0.97	GACGTGCGTGACGCCGGCCT(TGG)
				4.17	GCCGACGAGGACGAAGTCGA(GGG)
				2.15	GAGCCACAGTCGCAGCTGAC(TGG)
				12.28	GGCCAACGAGCGCTCCGTGC(AGG)
				14.7	GGAGGTCGACCTGCACGCGC(GGG)
				2.19	GCGCACACAGTTTCCAAAGC(TGG)
				9.53	GTACCACATTGTGGAGCGCG(AGG)
				78.83	GCTGCTACAGCAAGGCCTGA(TGG)
				12.36	GTTTGATTTGAAAGCGGGCG(CGG)
				1.6	GGCGCGTCCGCTGTGTTCGA(CGG)
				1	GCGCAGTCGTCACAAAGATC(TGG)
				24.62	GTGTGCGAGTTGGAGAAACA(TGG)
				5.72	GTGTGATATGCTGGATCAAA(AGG)
				29.4	GGGTTTCTTAGCGCGTGCAG(AGG)
				6.37	GTAGGCTCTTGCTGTCTAGT(TGG)
				2.44	GACCATAGCCAGGTTGAATC(CGG)
				43.12	GGAGGTTAGACGAGAGCTGC(AGG)
				5.1	GATCCTTATCTTGACTACCA(TGG)
				11.22	GTCTTGTTGGAGAGTGGTCA(TGG)
				20.18	GCGGTAGGAGAAGTGTCGCG(CGG)
				33.84	GAAGCCCTCGAGAGTCATGA(CGG)
				12.27	GAGCTGCTCACCGTGCTCGA(GGG)
				3.44	GGTGCTTTCTCATCAGTAAT(TGG)
				1.14	GTTCAGCTTTGCAGAATATT(CGG)
				12.14	GGGGTGGAGACGACTTCCAT(GGG)
				19.26	GAAGTCGTCTCCACCCCGGC(CGG)
				0.18	GCACTCCACTCCGTAGCACA(TGG)
				12.46	GAATGGAGACGACCTGGGCG(CGG)
				0.33	GGGCATTCGAGGACAGGTTA(AGG)
				0.38	GCATCCTATAACGACCTTTC(CGG)
				19.85	GCTGAAGGGGATGAAGAGCT(TGG)
				0.49	GTACTTGGCCGAGCTTCTCG(CGG)
				0.4	GCCGTCAAGTACATCGACCG(AGG)
				1.54	GACGCACCTGGCCATCGTCA(TGG)
				26.37	GTGGGTACTTGCGACACATG(TGG)
				4.31	GCTAAGCCATAAGGAGGGCC(TGG)
				27.24	GTACACTTCCATATCCATGG(CGG)

Supplementary Table 11. Plant Protoplast Dataset (cont.).

Reference	*Species*	*Target*	*Screening*	*Efficiency (%)*	Spacer Sequence *(5' → 3') (PAM)*
Zhu et al., 2016	Maize	n.s.	CTAB/Surveyor/Deep Sequencing	19.9	GAGGCGGTCCGTGATGACGA(CGG)
				17.59	GGACAATTGGTGAGGACATA(GGG)
				20.94	GTTAGGGTTCCCCGTAGCCA(TGG)
				21.94	GGCCGGAGCGCCACCTTCTG(CGG)
				25.85	GAGAGAGTGTCCCATGGGGC(TGG)
				2.12	GTGAATATAGTCCCACGACA(AGG)
				2.14	GTGGTTCAATACATGAGATA(AGG)
				8.29	GGAGGAGGACGATGAGCTCA(AGG)
				0.37	GAGAGCAACTGTTCAGGAGC(TGG)
				11.33	GCCGCTCTGGTTCACCGAGC(TGG)
				38.09	GTGCATCATGCGCACCAACA(TGG)
				2.79	TAGAGCTTAGTTGGCCATGG(TGG)
				0.51	TAATATGCAATCTGCTAGCA(TGG)
				2.09	ATCTTGATATTTGCTGGAGC(TGG)
				7.51	ACAGCGTGAGGGCATGCATC(TGG)
				0.84	ATTGCTCTTGGCGCTGCTCG(AGG)
				5.82	CTGCAGAGCTTTGTTGGCCA(TGG)
				1.89	CAGAAGTCCTGGAGGGGGCT(AGG)
				0.66	CATGAACTGTCTGAATCAGC(TGG)
				6.69	CTCCTGGCGTTCCCGCCACA(TGG)

n.s. : target gene not specified in the reference

7. Supplementary Materials.

Supplementary Table 12. Feature Values of the CRISPRnt Plant model.

		Nucleotide			
		A	T	C	G
Position in the 30 mer	-4	0.155	0.1553	0.1113	0.152
	-3	0.1056	0.1399	0.1386	0.1731
	-2	0.1322	0.196	0.1007	0.1481
	-1	0.0868	0.1535	0.2068	0.1413
	1	0.1455	0.0975	0.1176	0.1512
	2	0.1543	0.146	0.1152	0.1574
	3	0.1404	0.1625	0.0918	0.1672
	4	0.1001	0.1083	0.1527	0.1691
	5	0.1479	0.1156	0.161	0.1479
	6	0.1157	0.168	0.1255	0.1622
	7	0.1752	0.1339	0.1054	0.167
	8	0.0843	0.1325	0.1357	0.2088
	9	0.1169	0.1261	0.1706	0.1664
	10	0.1739	0.1316	0.1016	0.1644
	11	0.1039	0.1034	0.2047	0.1525
	12	0.1514	0.121	0.1581	0.1364
	13	0.1851	0.1152	0.1019	0.1546
	14	0.1624	0.1078	0.139	0.16
	15	0.1179	0.1109	0.1582	0.1916
	16	0.1171	0.141	0.1612	0.1487
	17	0.1199	0.1461	0.1681	0.1391
	18	0.0927	0.183	0.1495	0.1466
	19	0.1255	0.0984	0.1244	0.198
	20	0.1833	0.1009	0.1082	0.1466
	N	0.1206	0.1547	0.1329	0.1751
	G	0	0	0	0.1437
	G	0	0	0	0.1437
	1	0.1415	0.185	0.1111	0.1392
	2	0.157	0.081	0.1734	0.1551
	3	0.176	0.14	0.1043	0.1385

7. Supplementary Materials.

Supplementary Table 13. Grouping of the PP dataset for the Random Sampling Test.

Group	Efficiency	Spacer Sequence *(5' → 3')*
1	51.6	ATGCCGCTAGAGTGGTGCCCGG
	15.92	GCGCCTGCACCTCCGGAGGTTGG
	2.16	GCGTCCTCCCATCCTCGCCATGG
	5.31	GACTCACTATTCTGATCATAGGG
	1	GCGCAGTCGTCACAAAGATCTGG
	19.8	GAGGCAAGAGATGTCCTAGGTGG
	29	CCGTCACGCAGGACCCAATCTCC
	9.5	GTAGATGCAGAAGCTAGAGTTGG
	0.18	GCACTCCACTCCGTAGCACATGG
	0.33	GGGCATTCGAGGACAGGTTAAGG
2	16.65	ACCCTTGTTGTCGGATCCAGCGG
	0.66	CATGAACTGTCTGAATCAGCTGG
	2.4	GCCCTACTGAATCGATATATCGG
	48.6	CCGCCGGGCACCTACGGCAAC
	6.37	GTAGGCTCTTGCTGTCTAGTTGG
	38.09	GTGCATCATGCGCACCAACATGG
	17.8	GGTCGAAAAGGACGCCAAAGAGG
	25.00	AGTGAGATTGTGGGCGATGGG
	15.00	GCCGTTAATTTGAGAGTCCAAGG
	6.8	GACAAGAAGATCCCTTTGATTGG
3	27.24	GTACACTTCCATATCCATGGCGG
	12.14	GGGGTGGAGACGACTTCCATGGG
	7.7	GAGATTGTTATTGCTGGTGCAGG
	19.9	GAGGCGGTCCGTGATGACGACGG
	0.29	TTGTACCATTTCCTTAGCCCCGG
	7.0	GATATTAGAATCACATAGGGTGG
	27	CAAGATGCTATGTATCAGCTGG
	1.3	GAACTGCAATGCGGCCACTGACGG
	26.37	GTGGGTACTTGCGACACATGTGG
	20.3	GCCGTTAATTTGAGAGTCCAAGG
4	1.41	GAAAGATTGCTCCCATGTACAGG
	87.5	GAGCTTAGCACCTGGTTGGAGGG
	7.4	TGACTGGCACAACGCAAGGT
	1.89	CAGAAGTCCTGGAGGGGGCTAGG
	12.46	GAATGGAGACGACCTGGGCGCGG
	1.04	GCCCAATAGCAAGTATACTGTGG
	20.18	GCGGTAGGAGAAGTGTCGCGCGG
	2.12	GTGAATATAGTCCCACGACAAGG
	8.55	GCCGTGCGTGGCGTACGTGA
	4.37	GGAGATGAATAGATGGGTTG

Group	Efficiency	Spacer Sequence *(5' → 3')*
5	75.2	GTTGAGAGGGGAGCCGGTGATGG
	3.44	GGTGCTTTCTCATCAGTAATTGG
	5.1	GATCCTTATCTTGACTACCATGG
	3.69	GCGGGGACTCCGAGCGCTGCTGG
	26.8	ACGAGAGCTGCAGGCGGCCATGG
	2.09	ATCTTGATATTTGCTGGAGCTGG
	14	CGTCCAACCCATTCCTCTGCAGG
	3.70	GCTGTTAACTTGAGAGTCCAAGG
	32.86	GAACGGCGCTCCAGCAGCTGTGG
	1.6	GGCGCGTCCGCTGTGTTCGACGG
6	9.33	GGTGTACTTGAGATCATGGAAGG
	1.8	GATCAGTCTCCAAACATACCCATGG
	7	GAAATAAGTATGAGATCCATGGG
	1.1	GGTCGGCGGCGTGGTCGAGCTCGG
	0.4	GCCGTCAAGTACATCGACCGAGG
	3.9	GATGCCATTTAATTCCTACATGG
	4.57	GATGTCACTAAGCAATAAATAA
	42.1	CCTCGAGATGCTGTTCTCCATT
	27	GCAGATCTTGCAGAATCCTTGG
	2.5	GTAATGCTGGAACTGCAATGCGG
7	35.87	CCTCTAGAAATACCCCATCCTG
	2.96	GGATTACCTACTAAGTATACGGG
	24	AGGGCTGTTAACAAGCTTGATGG
	2.79	TAGAGCTTAGTTGGCCATGGTGG
	20.94	GTTAGGGTTCCCCGTAGCCATGG
	20	GTTGGTCTTTGCTCCTGCAGAGG
	2.56	AGGCGGCCTCGGCGACGTCCT
	5.82	CTGCAGAGCTTTGTTGGCCATGG
	0.84	ATTGCTCTTGGCGCTGCTCGAGG
	7.51	ACAGCGTGAGGGCATGCATCTGG
8	43.12	GGAGGTTAGACGAGAGCTGCAGG
	0.97	GACGTGCGTGACGCCGGCCTTGG
	53.75	GAGGCGTTCTCCAGATGCCATGG
	16.2	GTGGAATCATCAAACCAGGAAGG
	45	GTGCCGCGCGACGAGCTCTTCGG
	2.41	AGACCAGTACAAGGACGCCT
	1.83	GCCAGTCAACAATAATTGTAAGG
	34.78	TCGACCGCGACGAAGTGTTT
	4.31	GCTAAGCCATAAGGAGGGCCTGG
	0.38	GCATCCTATAACGACCTTTCCGG

7. Supplementary Materials.

Supplementary Table 13. Grouping of the PP dataset for the Random Sampling Test (cont.).

Group	Efficiency	Spacer Sequence *(5' → 3')*
9	20.98	GTGTCATGTTTGAACCATATGGG
	2.29	GAGACTTGAGGATCTGTTCACGG
	1.74	ATCTCGGCGCCGACGAACACG
	11.1	CCAGGGATCCGTCTCCTTCTCCC
	38	CGGCCGGCGGCGGATGCAGCTGG
	58	GGTCCTTGGAGCAAAACTGGTGG
	8.72	CACGCCGTCGAGATTACTGG
	5.72	GTGTGATATGCTGGATCAAAAGG
	17.59	GGACAATTGGTGAGGACATAGGG
	78.83	GCTGCTACAGCAAGGCCTGATGG
10	33	GTTGGTCTTTGCTCCTGCAGAGG
	7.63	GATCCTGCTCAATCTTTACATGG
	6.81	ACTCATGAGATGATGCATGCA
	6.60	GATGGAGATTGGTATGAGACCGG
	6.5	GCTGCATGGAAAGATGATGATGG
	2.19	GCGCACACAGTTTCCAAAGCTGG
	48.27	GAGAACGCCTCGCATGCCCACGG
	11.33	GCCGCTCTGGTTCACCGAGCTGG
	30	GCTGAAAGATTTGACGTCCCAGG
	5.22	GCATAGTCTTGATATTATTCCGG
11	16.35	CGGCGTGGCGCCGGAGCTCACGG
	8.2	ACCGATAAGCACAGCAGCTGTGG
	11.22	GTCTTGTTGGAGAGTGGTCATGG
	7.45	AGTGCTGCTCCGGCGTGCAG
	21.94	GGCCGGAGCGCCACCTTCTGCGG
	24.62	GTGTGCGAGTTGGAGAAACATGG
	8.29	GGAGGAGGACGATGAGCTCAAGG
	5.58	GAACCCTAGAATTGGTACCTTGG
	25	ATCATCTATCTGTGGTCTAGGG
	18.5	GGAGCGGCGCCGCGCCATATGGG
12	19.85	GCTGAAGGGGATGAAGAGCTTGG
	0.2	TGGGAGAAGGAGACGGATCCCTGG
	20.39	AGACACCGGGGCTAAGGAAATGG
	1.54	GACGCACCTGGCCATCGTCATGG
	9.53	GTACCACATTGTGGAGCGCGAGG
	33.80	GGCTTCGTGCCAACCAACGAGGG
	0.51	TAATATGCAATCTGCTAGCATGG
	2.44	GACCATAGCCAGGTTGAATCCGG
	3.18	GCAACACTACTGCTGCTAGCAGG
	6.50	GGACTTTTGCCAGCCATGGTCGG

Group	Efficiency	Spacer Sequence *(5' → 3')*
13	17.35	GATGGAGTTGGTGTGCCGCA
	0.37	GAGAGCAACTGTTCAGGAGCTGG
	19.26	GAAGTCGTCTCCACCCCGGCCGG
	9.8	GCTACTACAGCTGCCGTCAGTGG
	2.31	GTTAGACCTGATCTAGAAGTGGG
	25.85	GAGAGAGTGTCCCATGGGGCTGG
	29.4	GGGTTTCTTAGCGCGTGCAGAGG
	8.88	AGTACTCACGGAGATCCAAG
	2.15	GAGCCACAGTCGCAGCTGACTGG
	0.48	GTTGACAAACTCTTCACGTCAGG
14	7.30	GTTGGTCTTTGCTCCTGCAGAGG
	1.14	GTTCAGCTTTGCAGAATATTCGG
	4.1	GATGTCTTCATCATGGATCCAGG
	6.7	GCAGGACGTCGACGAGGACTCGG
	3.86	GCTTAGCTTCGGCCGCTCCTG
	10.2	GCGAGGGGCAAGCACTACCGCGG
	33.84	GAAGCCCTCGAGAGTCATGACGG
	3.8	TGTTGACAAGGGTGTTGAATTGG
	4.17	GCCGACGAGGACGAAGTCGAGGG
	14.7	GGAGGTCGACCTGCACGCGCGGG
15	38	GCGGCGGCCATGGCCATCACGG
	12.27	GAGCTGCTCACCGTGCTCGAGGG
	2.4	GACCCAATAGCTGGTATACACGG
	11.09	GGTCTACAACCCCTTTGCAGCGG
	0.49	GTACTTGGCCGAGCTTCTCGCGG
	12.28	GGCCAACGAGCGCTCCGTGCAGG
	56.8	GTTGAGAGGGGAGCCGGTGATGG
	2.14	GTGGTTCAATACATGAGATAAGG
	12.36	GTTTGATTTGAAAGCGGGCGCGG
	10.20	GTTGGTCTTTGCTCCTGCAGAGG

7. Supplementary Materials.

Supplementary Table 14. CRISPRnt-Plant Scores obtained from the Random Sampling Groups.

Group	Nt	30 MER POSITION																												
		-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	N	G	G	1	2	3
1	A	0.0282	0.0304	0.0033	0.026	0	0.2557	0.1289	0	0.0996	0.1256	0.198	0.0018	0.1263	0.2597	0.1049	0.207	0.0033	0.01	0.0171	0.0525	0.1592	0.0716	0	0.074	0.0033	0	0	0.2191	0.1154	0.3347
	T	0.15	0.1062	0.357	0.0484	0	0.095	0.516	0.0374	0.0018	0.1369	0.0812	0.0531	0.516	0.0216	0.1716	0.1263	0.0531	0.2515	0.0583	0.0531	0.357	0.1467	0.0221	0.1271	0.0809	0	0	0.0484	0.0216	0.0544
	C	0.2257	0.166	0.1068	0.29	0	0.0482	0.0531	0.0436	0.1896	0.1798	0.0275	0.1342	0.0117	0.0904	0.057	0.115	0.1202	0.0698	0.198	0.1699	0.0275	0.0216	0.1798	0.263	0.403	0	0	0.1177	0.0067	0
	G	0.1049	0.516	0.0484	0.2208	0.1348	0.1467	0.0784	0.2748	0	0.01	0.2567	0.244	0.1011	0.1943	0.2597	0.0033	0.2027	0.0531	0.3217	0.1701	0.084	0.3376	0.1856	0.198	0.0531	0.1348	0.1348	0.0216	0.2418	0.0422
2	A	0.0439	0.0066	0.0657	0.0153	0.1723	0.0374	0.1569	0.0683	0.1519	0.0695	0.1616	0.1943	0.214	0.0461	0.255	0.178	0.15	0.1938	0.0868	0	0.1943	0.1573	0.1374	0.2655	0.164	0	0	0.1874	0.2052	0.1804
	T	0.1306	0.15	0.0683	0.094	0.486	0.2952	0.0923	0.486	0.2153	0.2325	0.0637	0.0657	0.1903	0.1913	0.1587	0.0153	0.1569	0.486	0.066	0.0663	0.094	0.024	0	0.1015	0.1299	0	0	0.1546	0.1616	0.0462
	C	0.1665	0.2471	0.2194	0.3483	0.0066	0.1135	0.1022	0.1874	0	0.0637	0.2244	0.2749	0	0	0.1938	0.0995	0.1628	0.0683	0.2795	0.2772	0.2273	0.2675	0.2655	0.0066	0.2255	0	0	0.187	0.3809	0.486
	G	0.4335	0.1273	0.3155	0.2083	0.1561	0.178	0.4335	0.1176	0.1209	0.2681	0.1665	0.25	0.1462	0.2418	0.1209	0.3167	0.2197	0.1387	0.368	0.2	0	0.1591	0.1854	0.2768	0.25	0.1774	0.1774	0	0.0439	0.214
3	A	0.1735	0.1789	0.1365	0.0029	0.1415	0.0897	0.2712	0.1389	0.1377	0.27	0.1374	0.0851	0.0698	0.0894	0.2712	0	0.1666	0.2314	0.1809	0.0932	0.1804	0.2669	0.1214	0.1383	0.14	0	0	0	0.1599	0.166
	T	0.203	0.0907	0.2044	0.201	0.0029	0.1797	0.0698	0.008	0.1178	0.1534	0.2712	0.1828	0.1119	0.1849	0.0882	0.2046	0	0.1818	0.1001	0.1537	0.203	0.1208	0.2637	0.1214	0.2012	0	0	0.1655	0.0734	0.1357
	C	0.0712	0.2724	0.1622	0.1969	0	0.2365	0.108	0.2724	0.199	0.0961	0.0029	0.2637	0.2471	0.2357	0.0672	0.1121	0.0029	0.0703	0.183	0.2724	0.167	0.1316	0.103	0.1166	0.1218	0	0	0.106	0.199	0
	G	0.2192	0.1602	0.0939	0.1387	0.1723	0.1214	0.1328	0.1968	0.1822	0.1602	0.1325	0.1699	0.1214	0.013	0.2314	0.1125	0.1645	0.203	0	0.1597	0.04	0.0698	0.1502	0.202	0.1214	0.1492	0.1492	0.1214	0.1667	0.1427
4	A	0.0104	0.0189	0.4427	0.0147	0	0.2582	0.0641	0.0181	0.0236	0.0754	0.3403	0.0271	0.0931	0.875	0.0825	0.1027	0.0313	0	0.0328	0.074	0.0422	0.0212	0.0141	0.3272	0.0245	0	0	0.0201	0.0697	0.0992
	T	0.2546	0.074	0.0476	0.2837	0.074	0.0212	0	0.1246	0.3091	0.3133	0.0145	0.0177	0.0855	0.0313	0.0177	0	0.875	0.0854	0.1246	0.2753	0.875	0.0478	0.0271	0.0465	0.0104	0	0	0	0.0245	0.3403
	C	0.1335	0.0735	0.0718	0.1051	0.0189	0.0992	0.048	0.3198	0	0	0.0855	0.0465	0.447	0.0558	0.875	0.249	0.0585	0.0533	0.074	0.0534	0.2018	0.0104	0.0916	0.0141	0.1373	0	0	0.0834	0	0.0313
	G	0.0201	0.2312	0.0863	0.0476	0.172	0.0589	0.2792	0.1103	0.0694	0.0758	0.1065	0.3217	0.0123	0.1028	0.0763	0.0313	0.0992	0.3226	0.2849	0.0624	0.0513	0.2589	0.3448	0.0951	0.4745	0.1469	0.1469	0.2739	0.3874	0.0441
5	A	0.446	0.268	0.0335	0	0.1445	0.1898	0.3286	0.268	0.752	0.204	0.1974	0.044	0.0209	0	0.14	0.3515	0.1342	0.051	0.0357	0.3286	0.051	0.0277	0.0344	0.2248	0.0885	0	0	0.0581	0.2215	0.357
	T	0.1134	0.1052	0.3088	0.0308	0	0.3865	0.2029	0.0209	0.029	0.0408	0.0338	0.0277	0.1186	0.1052	0.0363	0.057	0.0872	0.016	0.0209	0.069	0.0291	0.3096	0.3286	0.0344	0.2131	0	0	0.3945	0.044	0.016
	C	0.0238	0.0885	0.037	0.446	0.14	0.114	0.0185	0.1732	0.0604	0	0.3286	0.1152	0.1112	0.076	0.1368	0.3286	0	0.2368	0.2628	0	0.0885	0.1401	0.1187	0.0659	0.016	0	0	0.1319	0.3286	0.0704
	G	0.268	0.273	0.2058	0.1401	0.1794	0.0635	0.1525	0.1753	0.2112	0.2309	0.268	0.5403	0.752	0.3453	0.752	0.037	0.2216	0.1676	0.016	0.1915	0.3424	0	0.1932	0.3286	0	0.1685	0.1685	0.268	0.0805	0.1288
6	A	0.153	0.421	0	0.0523	0	0.1767	0.1335	0.1217	0.0457	0.1444	0.1335	0.0249	0.2334	0.1697	0.0359	0.1097	0.0966	0.018	0.2017	0.1068	0.0214	0.032	0.0579	0.049	0.141	0	0	0.0997	0.0674	0.0748
	T	0.032	0.032	0.0788	0.2433	0.1887	0.018	0.0593	0.421	0.0421	0	0.1112	0.0524	0.1181	0.1043	0.0405	0	0.039	0.0363	0.018	0.2455	0.1341	0.2334	0.025	0.117	0.109	0	0	0.2455	0.2195	0.07
	C	0.0075	0.0862	0.2334	0.004	0	0.011	0.095	0.0457	0.039	0.0235	0.0592	0.144	0.0145	0	0.2334	0.137	0.0215	0.421	0.0368	0.032	0.1697	0.0905	0.1272	0	0.018	0	0	0.0403	0.1475	0.018
	G	0	0.0558	0.0454	0.0353	0.0404	0.1363	0	0.0331	0.234	0.223	0.011	0.1673	0.0145	0.0438	0.27	0.0581	0.216	0.17	0.025	0.0075	0	0.0522	0.0933	0.15	0.07	0.0997	0.0997	0.0457	0.0202	0.23
7	A	0	0.0418	0.019	0.1673	0.1078	0.0279	0.1545	0.1187	0.0582	0.1941	0.0582	0	0.0515	0.1348	0.24	0	0.2094	0.0926	0	0.2094	0.0288	0.0751	0.1961	0.1692	0.189	0	0	0.2793	0.1148	0.0582
	T	0.2072	0.24	0.1087	0.0582	0.0279	0.119	0.1393	0.0296	0.0296	0.1495	0.1539	0.1186	0.1375	0.1291	0.0431	0.1186	0.2	0.2084	0.142	0.1961	0.1328	0.1329	0.0751	0.0256	0.1221	0	0	0.1041	0.0371	0.1309
	C	0.0524	0.1987	0.1921	0.0876	0.2084	0.2169	0.0256	0.0582	0.1078	0.0268	0.0659	0.0296	0.0419	0.2094	0.1195	0.1367	0.1423	0.1042	0.0845	0.134	0.1474	0.1233	0.0754	0.0524	0.0256	0	0	0.0333	0.24	0.089
	G	0.0307	0.1148	0.1615	0.0751	0.1162	0.0984	0.1087	0.1513	0.1643	0.1142	0.2247	0.164	0.3587	0.0991	0.1336	0.0751	0.03	0.1187	0.1348	0.0418	0.1559	0	0.134	0.1487	0.0296	0.1233	0.1233	0.0984	0.163	0.1921
8	A	0.4844	0.0111	0.2672	0.1568	0.3478	0.2983	0.2017	0.0307	0.1843	0.0931	0	0.1511	0.0619	0.2175	0.0915	0.0619	0.3769	0.1585	0.2108	0	0.162	0	0	0.2664	0.2038	0	0	0.1942	0.3895	0.2044
	T	0	0.3248	0.0241	0.0235	0	0.306	0.0431	0.0038	0.0097	0.2248	0.2965	0.3031	0.0068	0.2475	0.3478	0	0.0183	0	0	0.2779	0.205	0.2175	0.1574	0.1613	0.1968	0	0	0.0829	0.0183	0.2175
	C	0.1288	0.0169	0.0583	0.4451	0.0241	0.0217	0.014	0.274	0.2539	0.1758	0.1303	0.1026	0.4451	0.0212	0.3769	0.2698	0	0.1788	0.0585	0.16	0.4312	0.2738	0.1536	0.1594	0.2269	0	0	0.297	0.3372	0.3478
	G	0.1359	0.3478	0.4844	0.0183	0.207	0.2277	0.3832	0.2851	0.2248	0.2601	0.0241	0.2299	0.4312	0.2692	0	0.4312	0.1464	0.259	0.3989	0.208	0.1536	0.0745	0.3137	0	0.186	0.2028	0.2028	0	0.0547	0.0183
9	A	0.1076	0	0.4056	0.0425	0	0.0551	0.1759	0.0667	0.1759	0.1384	0.4494	0.0572	0.4056	0.3453	0.0872	0.2348	0.3682	0.3899	0.2182	0.1942	0.2098	0.1576	0.1335	0.2611	0.0539	0	0	0.3785	0.0966	0.1929
	T	0.38	0.0572	0.48	0.4222	0.0174	0.1335	0.4619	0.1335	0	0.4637	0.2092	0.1316	0.1335	0.2098	0.1929	0.0572	0.0229	0.0872	0.1634	0.0572	0.0667	0.3376	0.1759	0.2098	0.4647	0	0	0.1136	0.0967	0.3186
	C	0.0572	0.2509	0.1013	0.1759	0.2336	0.4029	0.0872	0.2883	0.3447	0.0872	0.0174	0.5842	0.0523	0.0174	0.4085	0.58	0.064	0.0229	0.2098	0.4991	0.298	0.1105	0.0503	0.2453	0.0229	0	0	0.1759	0.3915	0.2348
	G	0.3904	0.2737	0.1181	0.2186	0.2778	0.3116	0.1561	0.4378	0.0617	0.1987	0.2336	0.1881	0.3116	0.2519	0.1827	0.2132	0.2186	0.2299	0.3582	0.2015	0	0	0.4589	0.2282	0.1929	0.2429	0.2429	0.0667	0.38	0.2336
10	A	0.3	0.1882	0.114	0.1537	0	0.2083	0.0522	0.2754	0.2142	0.1437	0.13	0.066	0.13	0.065	0.0707	0.0654	0	0.1799	0	0.045	0.0513	0.0705	0.198	0.173	0.315	0	0	0.1326	0.2963	0.129
	T	0.1937	0.0681	0.208	0.1127	0	0.1991	0.1675	0.0522	0.0681	0.1732	0.0586	0.2217	0.1311	0.2165	0.1451	0.1451	0.1024	0	0.1484	0.1528	0.1481	0.0586	0.0522	0	0.0689	0	0	0.2258	0.0474	0.2929
	C	0.0522	0.2161	0.0828	0	0.0681	0.1105	0.0907	0.0491	0.0849	0.2523	0.2217	0.1583	0.4827	0.0763	0.4827	0.33	0.298	0.1427	0.1667	0.1133	0.4827	0.3709	0.2318	0.1107	0.2003	0	0	0.0763	0.0897	0
	G	0.0849	0.1207	0.33	0.2754	0.1675	0	0.2523	0.1749	0.198	0.0602	0.2083	0.1444	0.0892	0.0676	0.1291	0.2744	0.1825	0.0681	0	0.2284	0.2217	0.0671	0.0667	0.33	0	0.1576	0.1576	0.066	0.1825	0.0775
11	A	0.1549	0.1141	0.0833	0.1435	0.0783	0.1529	0.1079	0	0.166	0.0829	0.1507	0.128	0.1665	0.0558	0.0833	0.0829	0.1379	0.2462	0.1232	0.1641	0.1623	0.2462	0.1298	0.171	0.0829	0	0	0	0.2385	0.1753
	T	0.082	0.0558	0.25	0.0558	0.25	0.1792	0.1623	0.1792	0.1122	0.1425	0.084	0.1811	0.0745	0.2481	0.2462	0.1529	0.0694	0.25	0.1122	0.1166	0.2347	0.1359	0.1838	0.1204	0.1241	0	0	0.1863	0.0808	0
	C	0.1534	0.2028	0.1646	0.1512	0.1635	0.082	0.1379	0.1722	0.1051	0.151	0.25	0.1298	0.1915	0.1061	0.1785	0.155	0.185	0.147	0.1621	0.185	0.1232	0.1363	0.1321	0	0.1915	0	0	0.096	0.134	0.082
	G	0	0.155	0.1606	0.1707	0.1503	0.1451	0.2049	0.1061	0.178	0.1722	0.1504	0.1553	0.1362	0.165	0.0829	0.1545	0.1836	0.1136	0.1566	0.1522	0.0896	0	0	0.1253	0.1698	0.1472	0.1472	0.2462	0.1228	0.1283

7. Supplementary Materials.

Supplementary Table 14. CRISPRnt-Plant Scores obtained from the Random Sampling Groups (cont.).

Group	Nt	30 MER POSITION																												
		-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	N	G	G	1	2	3
12	A	0.002	0.1248	0.0244	0.0185	0.2039	0.015	0.0802	0.0169	0.108	0.0686	0.0132	0.0953	0.0185	0.0036	0.1115	0.135	0.338	0.1393	0.1659	0.135	0.1415	0.0867	0.2039	0.1406	0.0636	0	0	0.0369	0.271	0
	T	0.0802	0.1716	0.065	0.135	0.0051	0.0953	0.1985	0.1716	0.2015	0.0315	0.065	0.1449	0.0554	0.0953	0.0185	0.1469	0.1045	0.0281	0.0199	0.0051	0.0329	0.0154	0.0244	0.1318	0.085	0	0	0.2012	0	0.1496
	C	0.184	0.0376	0.0697	0.0554	0	0.1152	0.1259	0.0972	0.0475	0.271	0.0866	0.0103	0.0244	0.1148	0.2015	0.0748	0.0164	0.2015	0.136	0.0236	0.0953	0.17	0.0633	0.0194	0.0447	0	0	0.065	0.0704	0.0655
	G	0.0244	0.0244	0.1805	0.1423	0.0963	0.1522	0.002	0.107	0.002	0	0.1805	0.1072	0.1615	0.1393	0.0792	0.0281	0.0958	0.0341	0.0169	0.1305	0.1097	0.091	0.1449	0.0953	0.338	0.0979	0.0979	0.1215	0.0285	0.0895
13	A	0.0635	0.1407	0.0666	0.0509	0.0888	0.13	0.1926	0.0944	0.0048	0.1192	0.0355	0.0297	0.0514	0.294	0.0231	0.0888	0.1926	0.1229	0.0134	0.0048	0.0231	0.0385	0.1256	0.1735	0.1574	0	0	0.014	0.0048	0
	T	0.0231	0	0.2484	0.1635	0	0.014	0.0776	0.294	0.2433	0.1603	0	0.2297	0.1635	0.1191	0.0048	0.0746	0.0043	0.0671	0.1737	0.294	0.0598	0	0.0048	0.0231	0.0954	0	0	0.2138	0.1166	0.0132
	C	0.1071	0.1571	0.0509	0.0215	0	0.098	0	0.0215	0.0694	0.067	0.1069	0.0606	0.095	0.0048	0.1427	0.1875	0.101	0.1471	0.0985	0.1191	0.1149	0.196	0.1735	0.0962	0.1926	0	0	0.0979	0.126	0.1407
	G	0.1827	0.0812	0.056	0.1362	0.1189	0.1914	0.1444	0.1236	0.1147	0	0.2082	0	0.14	0.0959	0.14	0.0215	0.1636	0	0.163	0.0958	0.1854	0.1302	0.1195	0.1603	0.056	0.1159	0.1159	0	0.1926	0.1418
14	A	0.039	0.067	0.0303	0.073	0	0.1897	0.1841	0.0703	0.0266	0.0523	0	0.0397	0.0923	0.0715	0.1607	0.067	0.1303	0.0719	0.0668	0.0114	0.1605	0.0387	0.0553	0.1901	0.057	0	0	0.0727	0.0771	0.0292
	T	0.1861	0.0377	0.2202	0.1078	0.0377	0.041	0.0403	0.0377	0.041	0.11	0.0398	0.1005	0.0505	0.0422	0.041	0.147	0.0553	0.041	0.188	0.0628	0.0266	0.1897	0.025	0.0387	0.0377	0	0	0	0	0.072
	C	0.0393	0.0414	0	0.0377	0.0386	0.0702	0.0417	0.0114	0.3384	0.1149	0.1124	0	0.13	0.107	0.147	0.0411	0.0692	0.11	0.0712	0.2427	0.0386	0.0805	0.07	0.094	0.1115	0	0	0.1303	0.0381	0.0852
	G	0.0719	0.1184	0.0733	0.0703	0.1027	0.0923	0.102	0.118	0.0775	0.0567	0.0719	0.1053	0.0417	0.1141	0.0455	0.1594	0.0944	0.0986	0.041	0.0499	0.0812	0	0.1757	0.0712	0.0944	0.0898	0.0898	0.0719	0.118	0.3384
15	A	0.1899	0.0721	0	0.0726	0	0.0734	0.0049	0	0.3454	0.0953	0.296	0.0662	0.0698	0.1227	0.175	0.2322	0.0725	0	0	0.0227	0.38	0.0227	0.1065	0.2232	0.0821	0	0	0.0645	0.1168	0.3454
	T	0.1124	0	0.2045	0.0734	0	0.164	0.2261	0.1236	0.065	0.0428	0.1236	0.0931	0.1128	0.0617	0	0.202	0.102	0.0945	0.0466	0.0726	0.0734	0.2689	0.0214	0	0.568	0	0	0.568	0.024	0.0491
	C	0.2007	0.0049	0.0638	0.2986	0	0	0.1756	0.0771	0.024	0.38	0.096	0	0.1692	0.1653	0.0951	0.0893	0.1169	0.225	0.2593	0.2085	0.0049	0.1119	0.1331	0.1228	0.1287	0	0	0.2514	0.0214	0.0638
	G	0.0049	0.2045	0.1679	0.0214	0.158	0.2046	0.0721	0.2679	0.2019	0.3454	0.1925	0.2689	0.568	0.2096	0.225	0.1228	0.2199	0.1759	0.0721	0.3458	0.1748	0.1236	0.2712	0.0854	0.1227	0.158	0.158	0.1228	0.2169	0.2048
AVG	A	0.1464	0.1122	0.1128	0.066	0.0857	0.1439	0.1491	0.0859	0.1662	0.1251	0.1534	0.0673	0.1203	0.1833	0.1288	0.1278	0.1605	0.127	0.0902	0.0961	0.1312	0.0875	0.1009	0.1898	0.1177	0	0	0.1171	0.163	0.1564
	T	0.1432	0.1009	0.1916	0.1369	0.0726	0.1498	0.1638	0.1415	0.099	0.1583	0.1071	0.1282	0.1337	0.1339	0.1035	0.0965	0.126	0.1222	0.0921	0.1399	0.1781	0.1493	0.0924	0.0839	0.1671	0	0	0.1803	0.0644	0.1271
	C	0.1069	0.1373	0.1076	0.1775	0.0601	0.116	0.0749	0.1394	0.1242	0.1259	0.121	0.1369	0.1642	0.0853	0.2477	0.1937	0.0906	0.1466	0.152	0.166	0.1745	0.149	0.1312	0.0911	0.1377	0	0	0.1259	0.1674	0.1143
	G	0.1314	0.1869	0.1685	0.1279	0.15	0.1419	0.1668	0.1786	0.136	0.145	0.1624	0.2038	0.2257	0.1568	0.1819	0.1359	0.1639	0.1435	0.1571	0.1497	0.1126	0.0909	0.1891	0.1663	0.1439	0.1475	0.1475	0.1016	0.16	0.1484

Supplementary Table 15. Spacers excluded in the Random Removal Test.

Reference	Species	Target	Efficiency (%)	Spacer Sequence *(5' → 3')*
Shan et al., 2013	Rice	*MPK2*	38	GCGGCGGCCATGGCCATCACGG
Andersson et al., 2017	Potato	*GBSS*	7.0	GATATTAGAATCACATAGGGTGG
Kim et al., 2018	Wheat	*TaDREB2*	6.7	GCAGGACGTCGACGAGGACTCGG
Hooghvorst et al., 2019	Melon	*PDS*	25.00	AGTGAGATTGTGGGCGATGGG
Feng et al., 2016	Maize	*Hsg7*	16.65	ACCCTTGTTGTCGGATCCAGCGG
Zhu et al., 2016	Maize	n.s.	24.6	GTGTGCGAGTTGGAGAAACATGG
			26.4	GTGGGTACTTGCGACACATGTGG
			19.9	GAGGCGGTCCGTGATGACGACGG
			2.12	GTGAATATAGTCCCACGACAAGG
			6.69	CTCCTGGCGTTCCCGCCACATGG

n.s. : Target gene not specified in the reference.

7. Supplementary Materials.

Supplementary Table 16. CRISPRnt-Plant Scores obtained after Random Removal of spacers.

Nt	30 MER POSITION																												
	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	N	G	G	1	2	3
A	0.1549	0.1014	0.131	0.0868	0.1439	0.1554	0.1394	0.1043	0.1517	0.1169	0.1782	0.0791	0.1172	0.1782	0.0966	0.1537	0.1883	0.155	0.1219	0.1065	0.106	0.0864	0.1191	0.1834	0.1234	0	0	0.1451	0.1521	0.1771
T	0.1529	0.1456	0.1992	0.1565	0.0975	0.1424	0.1649	0.1011	0.1126	0.1718	0.1339	0.1328	0.123	0.1243	0.0954	0.1097	0.1096	0.1078	0.1082	0.1422	0.1461	0.1778	0.0905	0.0961	0.1542	0	0	0.1877	0.0745	0.1381
C	0.1011	0.1353	0.0969	0.2003	0.1222	0.1151	0.0797	0.1545	0.16	0.1126	0.1063	0.1338	0.1634	0.0889	0.2047	0.1628	0.1068	0.1462	0.1507	0.1617	0.1659	0.148	0.1211	0.1082	0.1244	0	0	0.1022	0.1762	0.1072
G	0.1526	0.1697	0.1445	0.1364	0.1484	0.1512	0.1655	0.1635	0.1385	0.1613	0.1598	0.2106	0.1669	0.1672	0.1506	0.1294	0.145	0.1547	0.192	0.1472	0.1436	0.1455	0.2008	0.1448	0.1711	0.1416	0.1416	0.1339	0.152	0.1292

7.3. Supplementary References

Arndell, T., Sharma, N., Langridge, P., Baumann, U., Watson-Haigh, N. S., and Whitford, R. (2019). gRNA validation for wheat genome editing with the CRISPR/Cas9 system. *BMC Biotechnology* 19, 71.

Bernard, G., Gagneul, D., dos Santos, H. A., Etienne, A., Hilbert, J. L., and Rambaud, C. (2019). Efficient genome editing using CRISPR/Cas9 technology in chicory. *International Journal of Molecular Sciences* 20, 1155.

Cui, X., Balcerzak, M., Schernthaner, J., Babic, V., Datla, R., Brauer, E. K., Labbé, N., Subramaniam, R., and Ouellet, T. (2019). An optimised CRISPR/Cas9 protocol to create targeted mutations in homoeologous genes and an efficient genotyping protocol to identify edited events in wheat. *Plant Methods* 15, 119.

Fauser, F., Schiml, S., and Puchta, H. (2014). Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. *Plant Journal* 79, 348–359.

Feng, C., Yuan, J., Wang, R., Liu, Y., Birchler, J. A., and Han, F. (2016). Efficient Targeted Genome Modification in Maize Using CRISPR/Cas9 System. *Journal of Genetics and Genomics* 43, 37–43.

Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D. L., Wei, P., Cao, F., Zhu, S., Zhang, F., Mao, Y., and Zhu, J. K. (2013). Efficient genome editing in plants using a CRISPR/Cas system. *Cell Research* 23, 1229–1232.

Gao, J., Wang, G., Ma, S., Xie, X., Wu, X., Zhang, X., Wu, Y., Zhao, P., and Xia, Q. (2015). CRISPR/Cas9-mediated targeted mutagenesis in *Nicotiana tabacum*. *Plant Molecular Biology* 87, 99–110.

Hooghvorst, I., López-Cristoffanini, C., and Nogués, S. (2019). Efficient knockout of phytoene desaturase gene using CRISPR/Cas9 in melon. *Scientific Reports* 9, 1–7.

Hudzieczek, V., Cegan, R., Cermak, T., Bacovska, N., Machalkova, Z., Dolezal, K., Plihalova, L., Voytas, D., Hobza, R., and Vyskot, B. (2019). *Agrobacterium rhizogenes*-mediated transformation of a dioecious plant model *Silene latifolia*. *New Biotechnology* 48, 20–28.

Jacobs, T. B., LaFayette, P. R., Schmitz, R. J., and Parrott, W. A. (2015). Targeted genome modifications in soybean with CRISPR/Cas9. *BMC Biotechnology* 15, 16.

Johansen, I. E., Liu, Y., Jørgensen, B., Bennett, E. P., Andreasson, E., Nielsen, K. L., Blennow, A., and Petersen, B. L. (2019). High efficacy full allelic CRISPR/Cas9 gene editing in tetraploid potato. *Scientific Reports* 9, 17715.

Kim, D., Alptekin, B., and Budak, H. (2018). CRISPR/Cas9 genome editing in wheat. *Functional and Integrative Genomics* 18, 31–41.

Liang, Z., Zhang, K., Chen, K., and Gao, C. (2014). Targeted mutagenesis in *Zea mays* using TALENs and the CRISPR/Cas system. *Journal of Genetics and Genomics* 41, 63–68.

Lin, C.-S. S., Hsu, C.-T. T., Yang, L.-H. H., Lee, L.-Y. Y., Fu, J.-Y. Y., Cheng, Q.-W. W., Wu, F.-H. H., Hsiao, H. C. W. C.-W., Zhang, Y., Zhang, R., Chang, W.-J. J., Yu, C.-T. T., Wang, W., Liao, L.-J. J., Gelvin, S. B., and Shih, M.-C. C. (2018). Application of protoplast technology to CRISPR/Cas9 mutagenesis: from single-cell mutation detection to mutant plant regeneration. *Plant Biotechnology Journal* 16, 1295–1310.

Okuzaki, A., Ogawa, T., Koizuka, C., Kaneko, K., Inaba, M., Imamura, J., and Koizuka, N. (2018). CRISPR/Cas9-mediated genome editing of the fatty acid desaturase 2 gene in *Brassica napus*. *Plant Physiology and Biochemistry* 131, 63–69.

Osakabe, Y., Watanabe, T., Sugano, S. S., Ueta, R., Ishihara, R., Shinozaki, K., and Osakabe, K. (2016). Optimization of CRISPR/Cas9 genome editing to modify abiotic stress responses in plants. *Scientific Reports* 6, 26685.

Shan, Q., Wang, Y., Li, J., and Gao, C. (2014). Genome editing in rice and wheat using the CRISPR/Cas system. *Nature protocols* 9, 2395–410.

Wang, M., Mao, Y., Lu, Y., Tao, X., and Zhu, J. kang (2017). Multiplex Gene Editing in Rice Using the CRISPR/Cpf1 System. *Molecular Plant* 10, 1011–1013.

Wang, X., Tu, M., Wang, D., Liu, J., Li, Y., Li, Z., Wang, Y., and Wang, X. (2018a). CRISPR/Cas9-mediated efficient targeted mutagenesis in grape in the first generation. *Plant Biotechnology Journal* 16, 844–855.

Wang, Z., Wang, S., Li, D., Zhang, Q., Li, L., Zhong, C., Liu, Y., and Huang, H. (2018b). Optimized paired-sgRNA/Cas9 cloning and expression cassette triggers high-efficiency multiplex genome editing in kiwifruit. *Plant Biotechnology Journal* 16, 1424–1433.

Xing, H. L., Dong, L., Wang, Z. P., Zhang, H. Y., Han, C. Y., Liu, B., Wang, X. C., and Chen, Q. J. (2014). A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biology* 14, 327.

Yuan, M., Zhu, J., Gong, L., He, L., Lee, C., Han, S., Chen, C., and He, G. (2019). Mutagenesis of *FAD2* genes in peanut with CRISPR/Cas9 based gene editing. *BMC Biotechnology* 19, 24.

Zhu, J., Song, N., Sun, S., Yang, W., Zhao, H., Song, W., and Lai, J. (2016). Efficiency and Inheritance of Targeted Mutagenesis in Maize Using CRISPR/Cas9. *Journal of Genetics and Genomics* 43, 25–36.