Universitat Politècnica de València

Escola Tècnica Superior d'Enginyeria Agronòmica i del Medi Rural

Study of the precise genome editing in *Nicotiana benthamiana* using the prime editing system

Final thesis of Degree in Biotechnology

Author: Antonio Soriano Revert

External cotutor: Dr. Diego Orzáez Calatayud

External cotutor: Dr. Marta Vázquez Vilar

Academic tutor: Dr. Mª Pilar López Gresa

Academic year 2019 – 2020

Valencia, June 2020

Title: Study of the precise genome editing in *Nicotiana benthamiana* using the prime editing system

Summary:

Although many people are not aware of, plant breeding started at the very beginning of agriculture. A non-naturally occurring selection has always been stablished either by farmer's seed choice or induced plant crosses. During the last decades, an additional way of selecting plants for the desired traits has experienced an exponential growth. The rapid evolution of molecular biology-based techniques has boosted genetic engineering till unexpected limits.

Genetic engineering involves the use of biotechnology for direct manipulation of genes aiming, for instance, to transfer them across species boundaries, to knockout, to edit or to overexpress them. Among all possible benefits of plant genetic engineering to agriculture, it should be remarked its ability to increase crop production yields, nutritional content and pest resistance.

In general, gene editing has been carried out using specific nucleases that allow targeted insertions, deletions and precise sequence substitutions. To do so, those enzymes trigger double-strand breaks (DSBs) in the DNA that can be repaired either by non-homologous end joining (NHEJ) or by homology-directed repair (HDR), thus promoting the previously mentioned mutations. Traditionally, genetic engineers have used meganucleases, transcription activator–like effector nucleases (TALENs) and zinc finger nucleases (ZFNs), each of them presenting their own set of advantages and drawbacks.

In spite of the spectacular progress achieved with their application, an afterwards developed technology consisting of a bacterial CRISPR-associated protein 9 nuclease from *Streptococcus pyogenes* exceeded all their benefits. CRISPR-Cas system has completely revolutionized Genetic Engineering since it is an RNA-based nuclease that relies on base-paring rules between an engineered RNA and the desired target DNA site to be mutated rather than protein-DNA interaction needed for the previous nucleases.

However, scientific progress has gone a step further and new applications derived from CRISPR-Cas system have been developed. The most remarkable ones are base editing and its newly enhanced version prime editing. These techniques are able to induce direct genome edits without requiring DSBs or donor DNA templates. Specifically, prime editing is composed by a Cas9 endonuclease paired to a programmed reverse transcriptase able to both fuse to the target site and promote the desired edit with the information provided by a prime editing guide RNA.

The aim of this study is to design the proper tools for applying the recently published prime editing technique in *Nicotiana benthamiana*. From a deep bibliographical search, several potential candidates have been considered and analysed for future applications of this method. Among all of them, the Acetolactate Synthase and the Flowering Locus T-5 genes have been selected for designing the necessary guide RNAs and planning the experiments for testing the technique.

Key words:

Prime Editing; *Nicotiana benthamiana;* CRISPR-Cas; genetic engineering; *GoldenBraid*

Author: Antonio Soriano Revert

Location and date: Valencia, June 2020

External cotutor: Dr. Diego Orzáez Calatayud

External cotutor: Dra. Marta Vázquez Vilar

Academic tutor: Mª Pilar López Gresa

Título: Estudio del editado de genoma de precisión en *Nicotiana benthamiana* mediante el sistema "prime editing"

Resumen:

Aunque mucha gente no es consciente de ello, la mejora genética de plantas ha sido aplicada desde los inicios de la agricultura. Siempre se ha establecido una selección artificial, bien sea por la simple elección ejercida por los agricultores de las semillas más productivas o por la realización de cruces artificiales que no se dan en la naturaleza. Durante las últimas décadas, una forma de selección adicional a las ya existentes ha sufrido un crecimiento exponencial, promovida por la rápida evolución de las técnicas basadas en la biología molecular.

La ingeniería genética se basa en el uso de la biotecnología para la manipulación dirigida de genes con el propósito de, entre otros, superar los límites reproductivos entre especies diferentes transfiriéndolos de unas a otras, induciendo mutantes con genes silenciados o incluso sobreexpresándolos. Entre todos los aspectos positivos que la ingeniería genética de plantas aporta a la agricultura, merece destacar su capacidad de incrementar los rendimientos de producción, el contenido nutricional de los productos obtenidos o la resistencia a plagas.

En general, la edición genética se ha llevado a cabo utilizando nucleasas que permiten inserciones, deleciones y sustituciones específicas de secuencia. Para ello, estas enzimas desencadenan roturas de doble cadena (DSBs) del DNA que pueden ser reparadas, o bien mediante la unión de los extremos no homólogos (NHEJ), o bien mediante la reparación dirigida por homología (HDR), promoviendo así las mutaciones anteriormente mencionadas. Tradicionalmente, los ingenieros genéticos han utilizado meganucleasas, nucleasas de actividad similar a activador de transcripción (TALENs) y nucleasas con dedos de zinc (ZFNs).

A pesar del espectacular progreso que se ha conseguido gracias a su uso, una tecnología posteriormente desarrollada que consiste en una nucleasa Cas 9 asociada a CRISPR procedente de *Streptococcus pyogenes* ha mejorado incluso todas sus ventajas. El sistema CRISPR-Cas se basa en la simple complementariedad de bases entre un RNA previamente diseñado y la secuencia de DNA que se desea editar, sin considerar las interacciones proteína-DNA necesarias para las anteriores nucleasas.

Sin embargo, el progreso científico ha avanzado un paso más adelante desarrollándose así diferentes aplicaciones del sistema CRISPR-Cas. Las más remarcables son "base editing" y su nueva versión mejorada conocida como "prime editing". Dichas técnicas son capaces de inducir ediciones del genoma sin requerir DSBs o secuencias de DNA que actúen como molde. Específicamente, "prime editing" se compone de la endonucleasa Cas9 unida a una retrotranscriptasa programada con el fin de unirse a la secuencia diana e inducir la edición deseada a partir de la información proporcionada por el guía de RNA de "prime editing".

El propósito del presente estudio es desarrollar de las herramientas necesarias para aplicar la recientemente publicada técnica "prime editing" en *Nicotiana benthamiana.* Tras la pertinente búsqueda bibliográfica, se han identificado y analizado diferentes candidatos potencialmente útiles para futuros usos. Entre todos ellos, los genes de la Acetolactato Sintasa y los del *Flowering Locus T-5* han sido seleccionados para diseñar los guías de RNA y planificar los experimentos necesarios con el fin de probar la técnica.

Palabras clave:

"Prime editing"; *Nicotiana benthamiana*; CRISPR-Cas; ingeniería genética; *GoldenBraid.*

Autor: Antonio Soriano Revert

Localidad y fecha: Valencia, junio de 2020

Cotutor externo: Dr. Diego Orzáez Calatayud

Cotutor externo: Dra. Marta Vázquez Vilar

Tutora académica: Dra. Mª Pilar López Gresa

Agradecimientos

Simplemente dar las gracias a todas esas personas que, por poco que sea, han aportado su granito de arena para que este Trabajo de Fin de Grado sea una realidad. Compañeros de laboratorio, profesores, amigos y familiares que me habéis apoyado no solo durante este proyecto, sino a lo largo de los cuatro años de carrera. En especial a Diego, por haberme proporcionado un sitio en su súper laboratorio y un proyecto tan apasionante como lo es este. Y también a Marta, por haberme ayudado con TODO lo que he necesitado y mucho más. ¡No podía haber tenido más suerte con vosotros! Esperemos que no sea un adiós, sino más bien un hasta pronto. En fin, este trabajo cierra una etapa para dar lugar a otra nueva que deseo que sea igual o incluso mejor que la anterior. ¡Gracias a todos!

Table of contents

Figures' index

Tables' index

Nomenclatures and abbreviations

ALS: acetolactate synthase
AMP: ampicillin
Bp: base-pair
CAM: chloramphenicol
CaMV: cauliflower mosaic virus
Cas: CRISPR associate proteins
CRISPR: clustered regularly interspaced short palindromic repeats
DNA: deoxyribonucleic acid
DSBs: double strand breaks
EMS: ethyl methanesulfonate
FT: Flowering Locus T
FT5: flowering locus T-5
gRNA: normal guide RNA for CRISPR-Cas-based approaches
Indels: nucleotide insertions or deletions
IPTG: isopropyl β-d-1-thiogalactopyranoside
KAN: kanamycin
Kb: kilobase
M-MLV: Moloney murine leukemia virus
NLS: nuclear localization signal
PAM: protospacer adjacent motive
PAMPs: pathogen associated molecular patterns
PBS-RT: genome binding region composed of a primer binding site and a RT template
pDGB: GB destination plasmid
pegRNA: prime editing guide RNA
PS: protospacer
pUPD2: universal parts domesticator plasmid 2
RNA: ribonucleic acid
RT: reverse transcriptase
RYMT: *Rice Yellow Mottle Virus*
SAM: Shot Apical Meristem
sgRNAs: guide RNA specifically devoted for Cas9
SPE: spectinomycin
TALENs: transcription activator-like effector nucleases
TFL1: Terminal Flower 1
TILLING: Targeted Induced Local Lesions IN Genomes
TU: transcriptional unit
X-Gal: 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
ZFNs: zinc finger nucleases

1. Introduction

Human being has always pursued the selection of plant resources. From the very beginning of agriculture, our ancestors decided which seeds wanted to keep for the following season, depending on the quality of the mother plant products. It was just a matter of survival, but this is simply how evolution from wild relatives to productive crops has occurred (Harris, 1967). A crucial role in this history has been played by domestication. It has been described as a complex evolutionary process inducing morphological and physiological changes of animals and plants generated by their human use (Purugganan & Fuller, 2009). In fact, those variations allow to differentiate them from their wild relatives.

Among others, selection has been able to trigger the transition from wild teosinte to the already worldwide known corn. In this specific case, the domesticated corn has become a biological monstrosity, presenting huge and firmly attached corn grains that make it more fruitful and easy to harvest. However, it is unable to propagate without the human action. The less densely clustered and smaller teosinte grains allow its simpler dispersion of seeds for reproduction under natural conditions (Beadle, 1980). It is quite obvious that wild relatives are naturally selected with the only purpose of surviving, reproducing and avoiding predator attacks. In contrast, domesticated individuals are not that naturally fit, they are more susceptible to hazards and less prepared to their own survival. However, they present several traits of interest for humans, such as an increased production yield, a major fruit size or a reduced toxicity, that prompt their artificial selection.

It is worldwide known that human population is experiencing an increasing growth rate and food is becoming a limiting factor. Several organisations, for instance the United Nations (UN), have already predicted a not really promising horizon in 2050. By that time, agriculture should be able to feed an estimated population of almost 9.8 billion people in a sustainable way. This, paired to the diet shifts and the higher demand of biofuels, will require an increase of crop production yields without clearing more available land for agriculture, and improved crop breeding seems to be the only way to do so (Ray et al., 2013).

Traditional methods, as the previously mentioned selection of seeds, allowed domestication procedures to be applied from the very beginning. However, the rapid progress of molecular biology-based techniques has prompted the development of more advanced breeding methods. Among others, genome engineering is the most relevant for this study. It permits the direct manipulation of genes through biotechnology, not only providing a wider pool of genetic material but also decreasing the time needed to introgress multiple desirable traits in an elite genetic background (Hilder & Boulter, 1999). Therefore, genetic engineering can provide the necessary tools to face up the challenges proposed by modern agriculture, such as drug, heat and salinity tolerance (Mittler & Blumwald, 2010). Thus, crop productivity would increase satisfying future generations' needs.

Genetic engineering has already been applied since the 1970s as a methodology to insert new genetic elements in the genome of an organism. Derived from this branch of genetics, genome editing appeared around 2010 due to the discovery of programmable nucleases. The engineered enzymes at issue are able to directly induce DNA double-strand breaks (DSBs) at specific genomic sites. Those cuts can be repaired by the endogenous machinery of the cell either by non-homologous end joining (NHEJ) or homology-directed repair (HDR) at specific genomic locations (Sander & Joung, 2014). In plants, NHEJ is the preferred method, even in a higher proportion than in animal cells. Nevertheless, just in case a DNA template is present, the HDR path is triggered. As it could be predicted, cellular machinery is not perfect, so DSBs are error-

prone cleavages that can induce the desired targeted insertions, deletions and precise sequence substitutions in some cases (Figure 1).

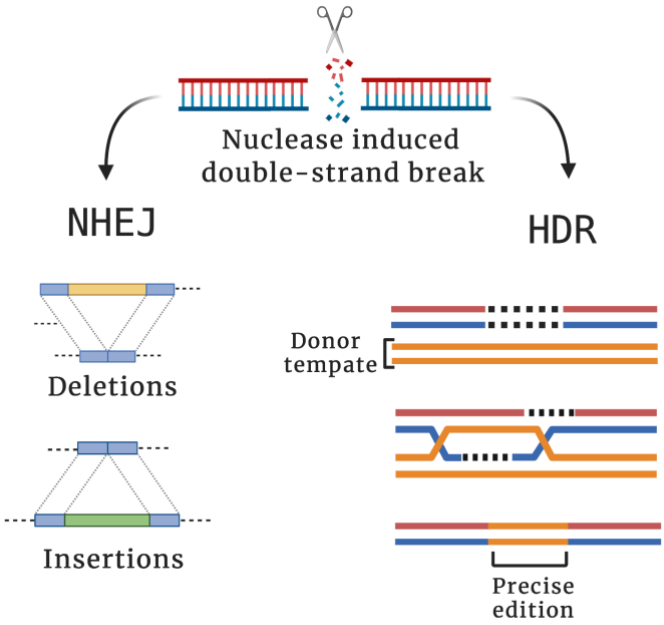

Figure 1. Repair mechanisms of DSBs. DSBs induced by programmable nucleases, such as ZFNs and TALENs, could be repaired either by NHEJ (inducing deletions and insertions) or by HDR (promoting a more precise edition of the genome, although with low efficiency and DNA template delivery difficulty)

Based on the mentioned cleavages, different techniques have emerged, each of them presenting its own advantages and drawbacks. Depending on the programmable nuclease used, they have been classified in different groups. Early applied methods, for instance Zinc Finger Nucleases (ZFNs) and Transcription Activator-Like Effector Nucleases (TALENs) were employed in first trials. Specifically, ZFNs resulted from the fusion of zinc finger proteins with the non-specific DNA cleavage domain from *Fok*I restriction endonuclease. Their site specificity depends on the zinc-finger domain design. Similarly, TALENs were originated from the combination of TALE proteins DNA binding domain and *Fok*I cleavage domain. They contain 33-35 amino acidic repeated domains that specify their target sequence to induce DSBs (Gaj et al., 2013). The main difference between this two is that ZFNs are able to recognize specifically from 9 to 18 base-pairs in triplets, whereas TALENs recognition is dictated per individual base-pair.

Although the already discussed endonucleases widened the range of genetic engineering applications, an afterwards discovered defense mechanism of bacteria and archaea revolutionized the field. It was named CRISPR-Cas (clustered regularly interspaced short palindromic repeats – CRISPR-associated proteins) and it conferred them sequence-specific resistance either to DNA or RNA (Van Der Oost et al., 2014). At genomic scale, it was described as a CRISPR array formed by several short direct repeats separated by the spacers (short variable DNA sequences) and flanked by the Cas genes. Indeed, there are variants of those Cas genes encoding different Cas proteins. When CRISPR-Cas system was discovered, the first proteins to be described were Cas9. Since that moment, they have covered almost all the applications, being the most commonly used Cas.

CRISPR together with Cas proteins confer adaptive immunity to bacteria in 3 steps. The first one is the adaptation, where foreign DNA fragments coming from invading plasmids or viruses are introduced into the CRISPR array as new spacers, thus conferring sequence specific defense against future infections of the same pathogens. Through the addition of new spacers that are bordered by the repeated sequences, the region is constantly built and enlarged to widen its range of action, being able to memorize the already fought biohazards. Second step is characterized by the expression of the CRISPR array. The genetic machinery is recruited to transcribe the Cas coding sequences and the corresponding guide RNAs involved in the defense. Finally, the interference stage takes place when the previously synthesized CRISPR RNAs serve as guide RNAs coupled to Cas proteins that target and cleave the desired genetic material (Makarova et al., 2015) (Figure 2).

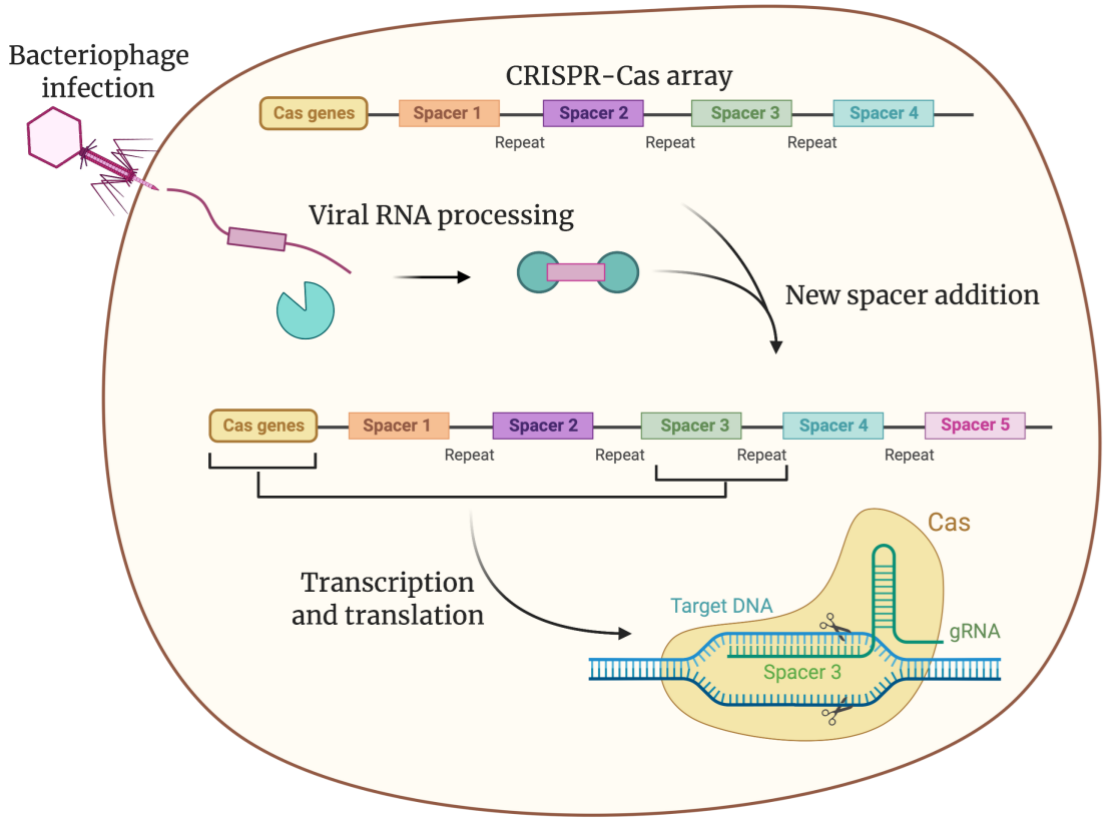

Figure 2. CRISPR immunization against foreign genetic elements in bacteria. CRISPR-Cas array is enlarged by the addition of a new spacer (spacer 5) derived from the bacteriophage infection. Transcription and translation of this CRISPR-Cas locus originate the CRISPR-Cas system paired to the corresponding gRNA.

Compared to the previously existing techniques, this system has been revolutionary since it only relies on base-paring rules between an engineered guide RNA and the desired target site to be cleaved rather than protein-DNA interaction needed for the previous nucleases. Moreover, in addition to the DSBs generation that could lead to insertions, deletions and point mutations, Cas proteins have been extensively engineered for gene activation, DNA methylation, histone modification or even tagged with fluorescent proteins in order to help with specific loci imaging. It should be mentioned that Cas9 proteins associated to CRISPR guide RNAs are just able to cut when the target sequence adjoins to a protospacer adjacent motive (PAM) at the 5’ end. Even though several of those PAMs have been reported, such as 5’-NAG, the most

common one is 5'NGG, considering that N accounts for any nitrogenous base. Hence, viral sequences located in the CRISPR locus cannot be cut since they are not next to a PAM (Sander & Joung, 2014).

It is clear that the already existing techniques, including CRISPR-Cas system, promote point mutations but in an inefficient way, typically inducing random indels (insertions and deletions). There have been reported several examples, even in plants, where those indels play a crucial role in the obtention of knockout genes, that is to say, genes whose expression is inhibited after the indel takes place. Although this fact could demonstrate their suitable implementation in breeding programmes, the ability to directly edit a single base-pair at researcher's will was still pending. Thus, genetic engineering technologies have progressed towards the specificity and the precise genome editing. New methodologies derived from the original CRISPR Cas system have been developed, improving the editing efficiency and accuracy. The most highlighted ones are base editing and its updated version prime editing.

Base editing arose as one of the most promising CRISPR-Cas9 applications. In particular, firstly developed base editors resulted from the fusion of a catalytically inactivated Cas9 nickase (nCas9) from *Streptococcus pyo*genes, the APOBEC1 cytidine deaminase and the uracil glycosylase inhibitor (UGI) of base excision repair. This complex retained the ability to, when programmed with the corresponding guide RNA, mediate the direct conversion from cytidine (C) to uridine (U), thereby effecting a C to T or G to A substitution (Kim et al., 2017; Komor et al., 2016). This new technique was firstly applied in human and murine cell lines, since its irreversible point mutations could serve to correct certain human diseases. Apart from an increased efficiency and a reduced error rate, base editing generates the desired mutation without requiring neither a DSB of the DNA nor a donor template for the edition (Kim et al., 2017; Komor et al., 2016). Among the main drawbacks found in primary base editors, it should be remarked their narrow capability to convert only cytidine to uridine, due to the unique presence of a cytidine deaminase in the structure. Moreover, the editing window width was around 5 nucleotides, so cytosines nearby to the desired point of mutation could be also modified prompting non-targeted editions. In addition, they just covered mammalian hosts.

Researchers wanted to overcome those problems and, just a few months later, some results were already reviewed. For instance, the editing window width of the firstly discovered base editors was reduced to 2 nucleotides, enhancing their accuracy (Kim et al., 2017). Furthermore, adenine base editors were described as a new approach able to replace A to G or T to C, thus covering the whole range of transition mutations for the disease-correcting or disease-supressing in human cells. In this case, a catalytically inactive version of the Cas9 was fused to a transfer RNA adenosine deaminase, instead of a cytidine deaminase, achieving quite promising efficiencies (Gaudelli et al., 2017). Additionally, other hosts rather than mammalian cells were explored, such as the application of cytidine base editors in *Oryza sativa, Triticum spp.* and *Zea mays* (rice, wheat or maize) (R. Qin et al., 2019; Zong et al., 2017), as well as the newly developed adenine base editors in *Arabidopsis thaliana, Brassica napus* and *Gossypium hirsutum* (Arabidopsis, canola and cotton) (Kang et al., 2018; L. Qin et al., 2020).

In October 2019, Anzalone et al. surprised the scientific community with the release of an updated version of base editing. This important group of researchers from Harvard University named it prime editing due to its precise and accurate genome editing ability. It is a precise genome editing method able to directly write genetic information on a specific DNA site. It is composed of a catalytically impaired Cas9 nickase (nCas9) protein fused to an engineered reverse transcriptase (RT) both coupled to a specifically designed guide RNA that drives the complex towards the target site and, in addition, encodes the desired edit. Since it is an

evolution of base editing approach, it does not require DSBs and neither donor DNA template (Anzalone et al., 2019).

This methodology is able to expand the editing scope, allowing to correct almost the 89% of known pathogenic human variants. Indeed, base editing was able to perform all four possible transition mutations (C to T, G to A, A to G and T to C), but prime editing can conduct all the base-to-base conversions, including the mentioned transitions and all the eight possible transversions (C to A, A to C, C to G, G to C, G to T, T to G, A to T and T to A). What is more, none other DSB-free method than prime editing has been reported to trigger targeted insertions and deletions, such as the 3 base-insertion needed to correct the most common case of cystic fibrosis or the 4-base deletion required to treat Tay-Sachs disease (Anzalone et al., 2019).

The importance of the technology relies on the special design of the prime editing guide RNA (pegRNA). As all the guides used in previous methodologies, pegRNAs are composed of a spacer sequence that directs the complex and hybridizes with the target DNA sequence. In this particular case, next to their 3' end, they have an additional genome binding region composed of a primer binding site and an RT template. This fragment encodes the desired edit and, with the aid of the RT, the mutation is introduced on the original DNA strand.

Specifically, the strand containing the PAM sequence is firstly nicked by the nCas9. The resulting 3'end binds the primer binding region of the pegRNA and it serves as initial point for the RT to retrotranscribe new DNA containing the desired edit. This enzyme uses the RT template part of the pegRNA as a mould. Here comes the first problem, since two different flaps are created, that is, a 5' non-edited flap and a 3' flap containing the mutation, the finally ligated one can be the non-edited one. Although the non-edited 5' flap could seem the preferred by the cell machinery to ligate since it is thermodynamically more favoured, there are some structure-specific enzymes (such as *FEN1*) prone to cleave 5' DNA flaps. Thus, the equilibrium is displaced towards the 3' flap. That is great, since the newly retrotranscribed fragment is ligated into the genomic DNA in most cases. However, here comes a second problem. A mismatch is induced between the edited and the non-edited strand. The preferential repair using the edited strand as a template can be induced by nicking the non-edited strand with an additional sgRNA, an approach already followed in base editing by the same researchers (Anzalone et al., 2019) (Figure 3).

Just as a reminder, prime editing requires a Cas9 nickase-RT fusion. Anzalone et al. used the Cas9 H840A nickase and the commercial Moloney murine leukemia virus (M-MLV) RT variant. Additionally, they tested three different approaches. Prime editor 1 (PE1) resulted from the simple fusion of Cas9 H840A with M-MLV RT. Afterwards, various mutations of the RT were considered, providing thermostability, enzyme processivity and DNA:RNA substrate affinity, thus obtaining prime editor 2 (PE2) and enhancing the editing efficiencies of PE1. However, best results were obtained with the two versions of prime editor 3 (PE3 and PE3b). The PE3 strategy incorporates to the PE2 Cas9 H840A-RT an additional single guide RNA (sgRNA) next to it in order to nick the non-edited strand. Thus, the induced DNA repair mechanism would use the edited strand as a template, permanently installing the desired mutation. The only difference among PE3 and PE3b is that PE3 nicks the original non-edited strand whereas PE3b only recognizes the target to nick once the edition has taken place, diminishing the probability of possible indels. This could be used even to nick the PAM avoiding future editions in the same site (Anzalone et al., 2019).

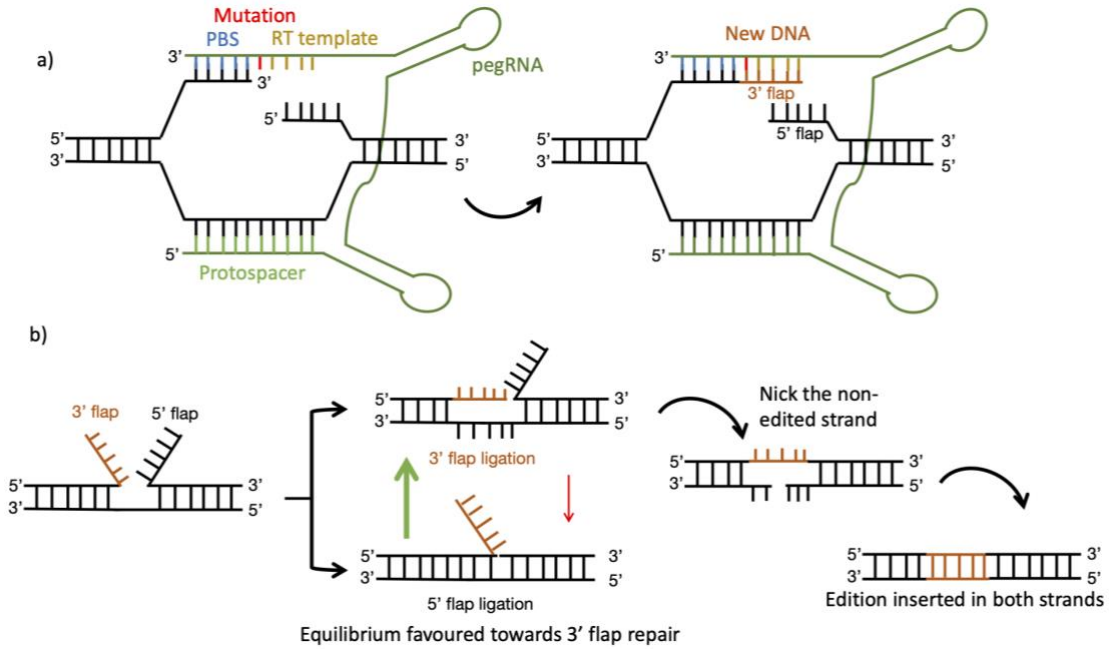

Figure 3. Prime editing strategy (adapted from Anzalone et al.). a) After the Cas9 induces a nick on the PAM strand, pegRNA is able to bind. Its PBS region of the 3' extension binds to the complementary DNA sequence and triggers the retrotranscription of a newly created DNA at the 3' end carrying the desired mutation (based on the RT template). **b)** When the guide is released, the equilibrium between 3' flap and 5' flap ligation is displaced towards the 3' one carrying the edition. Then, the nicking of the non-edited strands promotes the repair of this one, thus definitively inserting the edition in both DNA strands.

First trials were carried out with human HEK293T cells providing much lower off-target activity, fewer byproducts and higher efficiency than CRISPR Cas9 HDR based techniques. In March 2020, Lin and colleagues used codon optimization and included plant regulatory regions to adapt prime editors for their use in plants. In fact, they achieved successful insertions, deletions and point mutations in rice and wheat protoplasts, thus obtaining a 21,8% of prime-edited regenerated plants (Lin et al., 2020a).

Among other experiments performed by Lin et al, the efficiency of the M-MLV RT in plant hosts was compared to others, for instance the retron-derived RT (RT-retron) from *Escherichia coli* BL21 or the CaMV RT (RT-CaMV) from cauliflower mosaic virus 12. Obtained results suggested that M-MLV RT could be replaced by other RT, although the editing efficiency would be equal or lower. Furthermore, it was demonstrated that nicking the non-edited strand through the sgRNA does not increase efficiency in plants, so PE3b are worthless (Lin et al., 2020b).

All those investigations stablish the basis from which further research projects should start. As it has been reviewed, results can vary from one organism to the other, and there are not general rules to follow when dealing with this technique. Thus, prime editing system needs to be refined and several conditions should be tested for each specific host. Throughout this study, the scarce information about prime editing technology has been deeply analysed in order to design the proper tools for its application in *Nicotiana benthamiana.* From a bibliographical search, several potential candidates are proposed for future research and, among them, the Acetolactate Synthase (ALS) and the Flowering Locus T-5 (FT5) genes have been selected for designing the necessary guide RNAs and planning the experiments for testing the technique.

2. Objectives

The main objectives of this final thesis of Degree in Biotechnology are:

- To define the best strategy for prime editing in *N. benthamiana*.
- To adapt to the *GoldenBraid* cloning system all DNA elements required for the assembly of prime editing constructs to dicots.
- To design and assembly the required guide RNAs and DNA constructs for precise editing of FT5 and ALS in *Nicotiana benthamiana.*
- To search potential agronomically relevant targets for prime editing in plants.
- To envision future perspectives of the technique that would widen the range of genetic engineering applications,

3. Materials and methods

3.1. *Benchling* informatic tool

Benchling is an online platform that allowed to deal with all the informatic bioprocessing. This tool was used to store all the genetic sequences, reproduce the theoretical assemblies, design the desired guide RNAs and check the successfully obtained DNA constructs through sequence alignments. Furthermore, it provided a theoretical previsualization of the agarose electrophoresis gels corresponding to the PCR and restriction analysis assays, thus avoiding choosing the incorrect restriction enzymes.

3.2. Cloning vectors and antibiotic resistance

Throughout the experimental procedure, several cloning vectors were employed, each of them carrying different antibiotic resistance genes. All their features are summarized in Table 1.

Table 1. Cloning vectors with their respective GB number and antibiotic resistance

VECTOR	GB NUMBER	RESISTANCE
pUPD2	GB0307	Chloramphenicol
pDGB3_α1	GB0015	Kanamycin
pDGB3_ α2	GB0017	Kanamycin
pDGB3_ Ω1	GB0019	Spectinomycin
pDGB3_ Ω2	GB0021	Spectinomycin

3.3. Microbiological material:

Escherichia coli: all the cloning experiments were performed using the chemically competent *Escherichia coli* TOP10 strain (*ThermoFisher Scientific*). Bacteria were grown in lysogenic Luria-Bertani (LB) broth and plated in LB-agar solid media. In addition, different selection antibiotics were used, such as chloramphenicol (34 µg/mL) and kanamycin (50 µg/mL), depending on the transformation vector (Table 1). Liquid media were placed to grow at 37 degrees Celsius (°C) for 16 hours applying a constant agitation of 220 rpm. Plates were left also 16 hours in the same thermic chamber without shaking.

In order to select the colonies containing the desired plasmid, 0.5 mM Isopropyl β-d-1-thiogalactopyranoside (IPTG) that triggers lac operon transcription and 40 µg/mL 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) chromogenic substrate were added to the growth media. These components allowed the easy colony selection on which *GoldenBraid* relies. Cloning vectors are carrying a lac operon that encodes for β-galactosidase, an enzyme involved in the blue coloration of the colonies. Basically, bacteria containing the insert within the lac operon were white since this gene was not expressed whereas the ones without insert were able to produce β-galactosidase and presented a blue colour.

Agrobacterium tumefaciens: the strain C58 was used in this study to agroinfiltrate the desired constructs into *Nicotiana benthamiana* leaves. *A. tumefaciens* was grown both in LB broth and LB-agar plates with plasmid selection antibiotic (Table 1) and supplemented with rifampicin (50 µg/mL) for 48 hours at 28°C, with and without 200 rpm agitation respectively.

3.4. Plant material

The *in vivo* tests were performed with *Nicotiana benthamiana*, a close relative of *Nicotiana tabacum* (tobacco) which is very resistant due to its desertic origin. It is a model organism in plant research frequently used to express substances of interest benefiting from its reduced defense mechanisms against the small percentage of biotic stresses able to attack it. In addition, its genome has been recently published by the members of Newcotiana project, one of them being Dr. Diego Orzáez laboratory, whereby I am enrolled with.

3.5. Bacterial transformation

In order to transform ***E. coli*** with the desired construct, TOP 10 strain competent cells were used. Their preparation was performed through the *Mix & Go! E. coli* transformation kit. Specifically, 1,5 mL tubes containing -80°C cryopreserved aliquots of 70 to 100 µL were slowly defrosted in ice and mixed with 5 µL of the desired construct during 5 mins more in ice. Then 300 µL of SOC (Super Optimal broth with Catabolite repression) were added and the tube was precultured for 1 hour at 37°C with agitation. Afterwards, the precultured media was plated on LB-agar dishes containing the already mentioned compounds (IPTG, X-Gal and selection antibiotic specified in Table 1). After 16 hours, white colonies should be selected and picked for the following steps.

A. tumefaciens transformation was done through electroporation. 2 mL tubes containing 50 µL aliquot of electrocompetent C58 cells stored at -80°C were defrosted in ice and mixed with 1 µL of DNA. The mixture was transferred to electroporation cuvettes and exposed to 1440 volts (V) during 5 milliseconds (ms). Afterwards, 500 µL of SOC were added and the tubes were precultured for 2 hours 28°C under 200 rpm agitation. Finally, 50 µL of each transformation were plated on Petry dishes containing LB-Agar, the plasmid selection antibiotic (Table 1) and rifampicin.

3.6. PCR assays

Polymerase chain reactions (PCRs) were employed in order to obtain some cloning constructs. For that purpose, *Integrated DNA Technologies* platform was used to order the customized primers whereas *ThermoFisher Scientific Tm calculator* allowed to obtain the melting temperature of each of them.

In particular, the Master Mix provided by *ThermoFisher Scientific* was used, containing dNTPs, buffer and their "Phusion High Fidelity DNA Polymerase". Reagents and thermocycler conditions are summarized in Table 2. Afterwards, PCR products were purified through the *PCR clean-up and gel extraction kit* from *Macherey-Nage*

Table 2. PCR components and conditions. *Annealing temperature depended on the primers used. ** The number of cycles and the extension time varied depending on the length of the construct.

REAGENT	AMOUNT	CONDITIONS			
Mix (dNTPs, buffer and enzyme)	25 µL	**Step**	**Temp.**	**Time**	**Cycles**
DNA template	1 ng	Initial denaturation	98°C	30 s	1
Forward primer	2,5 µL	Denaturation	98°C	10 s	25 – 35 **
Reverse primer	2,5 µL	Annealing	X °C *	10 - 30 s	
H_2O (up to 50 µL)	19 µL	Extension	72°C	15 - 30 s/Kb **	
		Final extension	72°C	5 – 10 min	1
		Final hold	4°C	Hold	Hold

3.7. Restriction-ligation reactions

GoldenBraid system was employed for the assembly of the different genetic devices generated in this study. Although their exact design will be further on discussed in the results section, the different reagents and their concentrations, as well as the conditions needed for the ligation are stated in Table 3.

It is important to remark that initial pieces derived either from PCR products (linear DNA) or entry vectors (plasmid DNA), so different DNA concentrations were used for the reaction (60 or 20 fmol respectively). Moreover, thanks to the *GoldenBraid* cloning system, different DNA parts could be simultaneously combined in a single tube reaction, always taking into account that the restriction enzymes used were the proper ones. Regarding the conditions, just mention that the background removal step was performed at 50°C since the restriction enzyme is active at this temperature but the ligase isn't, so the remaining original plasmids in the tube were cut decreasing the final number of undesired blue colonies (Table 3).

Moreover, primer-dimers needed for some assemblies were obtained from initial primer dilutions at 100 µM, mixing 1 µL of each primer in 198 µL of water to get a final concentration of 1µM each and letting them anneal for 1 hour at room temperature.

Table 3. Restriction-ligation reactions components and conditions.

REAGENT	AMOUNT	CONDITIONS			
Buffer T4 Ligase	1,5 µL	**Step**	**Temp.**	**Time**	**Cycles**
BSA	1,5 µL	Restriction	37°C	3 min	35
T4 Ligase	0,8 µL	Ligation	16°C	4 min	
Restriction enzyme	0,8 µL	Background removal	50 °C	10 min	1
Initial vector / PCR product	20 fmol / 60 fmol	Inactivation	80°C	10 min	1
Destination vector	20 fmol	Final hold	16°C	Hold	Hold
H_2O (up to 15 µL)	X µL				

3.8. Plasmid extraction

After picking an individual colony and growing it for the corresponding time, the resulting 3 mL liquid culture was pelleted and a miniprep kit was used in order to extract the plasmid DNA. To do so, the *E.Z.N.A.® Plasmid Mini Kit from Omega Bio-tek Inc* was used in the case of *E. coli* transformed cells, whereas for the *A. tumefaciens* cells, the *QIAGEN QIAprep Spin Miniprep Kit* was needed. Sufficient DNA concentrations to carry out next steps were achieved just following the fabricant instructions.

3.9. Restriction enzyme analysis

After obtaining the purified plasmids, restriction enzyme assays were performed in order to check out the correct insertion of the desired fragment. Depending on the cloning vector and insert, different enzymes were used, letting them act for an hour at 37°C. Samples preparation was performed retaining the tubes in ice, thus avoiding the restriction to start. Components and conditions required for these assays are stated in Table 4.

Table 4. Restriction analysis components and conditions.

REAGENT	AMOUNT	CONDITIONS
DNA sample	400 ng	1 hour at 37°C
Commercial enzyme	0,5 µL	
10X Buffer	1,5 µL	
H_2O	Till a final volume of 15 µL	

3.10. Agarose gel electrophoresis

This technique allows to track and split DNA depending on its size. Thus, results from the restriction assay as well as all PCR products obtained in this study were verified with an agarose gel electrophoresis. Expected bands were stablished with the online informatic platform *Benchling* and compared with the ones obtained.

Gels containing 1% agarose were run with buffer TAE 1X (Tris-acetate EDTA) and ethidium bromide (EtBr) at 1/1000 dilution was added for DNA staining. Depending on the cast size, the running voltage could vary between 90 – 140 V. In addition, samples were mixed with *Orange Loading Dye* (2,5% Ficoll-400, 11mM EDTA, 3.3mM Tris- HCl, 0,017% SDS and 0,15% Orange G) both favouring their visualization and increasing their density in order to stay at the bottom of the gel without floating. *Gene Ruler DNA Ladder 1kb* and *100bp (ThermoFisher Scientific)* were employed to elucidate bands size and the results were finally observed through a UV light transilluminator and captured with the *GeneSnap* software.

3.11. Sequencing

Once the constructs had been correctly verified through the restriction analysis, they were sequenced in order to guarantee that the base-pair sequence aligned perfectly with the theoretical DNA sequence. All alignments were performed using *Benchling.* Sequencing was carried out at the Sequencing and Genetic Expression Analysis service of the "Instituto de Biología Molecular y Celular de Plantas" (IBMCP) through the *Sanger sequencing methodology* with the aid of an ABI 3130 XL Genetic Analyzer.

For that purpose, 10 µL of each sample (200 ng DNA / µL aprox.) and 5 µL per sample of forward primer (5 µM) were needed. It is important to consider that, depending on the insert length, reverse or even intermediate primers could be required to sequence the whole region considering that each assay can read optimally up to 500 - 600 nucleotides.

3.12. Bacterial cryopreservation and recovery

All designed constructs were stored in the *GoldenBraid* glycerol stocks and DNA sequences collections of the laboratory for further uses. Thus, the previously verified colonies were individually grown in liquid cultures containing the selection antibiotics (Table 1) during 16 hours at 37°C with agitation. 500 µL of the resulting broth were mixed with 500 µL of 50% glycerol in the cryovials. Each tube was named with a GB number and stored at the -80°C ultra-freezer. In addition, DNA sequences of all generated GB elements were uploaded to the *GoldenBraid* collection found at the GBCloning website (https://gbcloning.upv.es/search/features/).

In the same way, some stocks of the GB collection were used in this study. To recover them from the -80°C ultra-freezer conditions, a small amount of the cryovial was defrosted with the aid of a burning inoculation loop and grown in liquid media as explained above.

3.13. Agroinfiltration

As mentioned before, infiltration of *Nicotiana benthamiana* leaves was performed with C58 strain of *A. tumefaciens.* To do so, one colony per construct was picked and precultured in 5 mL liquid medium containing LB, the plasmid selection antibiotic (Table 1) and rifampicin at 28°C for 48 hours under agitation. By that time, 5 µL of the saturated broth were subcultured again in 5 mL of the same components. After 24 hours at 28°C under agitation, the *A. tumefaciens* cultures containing the desired construct were ready to be inoculated into the plants.

A. tumefaciens cultures were mixed with agroinfiltration buffer. This last was composed of 2-(N-morpholino) ethanesulfonic acid (MES) and magnesium chloride ($MgCl_2$). The exact proportions are detailed in Table 5.

In particular, to obtain the final agroinfiltration solution, subcultures were pelleted by spinning at 4500 rpm for 15 minutes. Then, the supernatant was poured off and cells were resuspended in 5 mL of agroinfiltration buffer. They were covered with aluminum foil in order to avoid light degradation of acetosyringone and left in agitation for 2 hours. After this incubation period, OD at 600 nm was measured and adjusted with agroinfiltration buffer to 0,1. Then, the agroinfiltration solution was ready to be inoculated into the abaxial surface of plant leaves with the aid of a 1 mL sterile syringe without needle.

Table 5. Agroinfiltration buffer mix for a final volume of 100 mL.

REAGENT	AMOUNT
10x MES (100 mM) at pH 5.6	10 mL
100x $MgCl_2$ (1 M)	1 mL
1000x acetosyringone (200 mM)	0,1 mL
Distilled H_2O	88,9 mL

4. Results

Prime editing is a really powerful tool that still remains to be exploited. Its recent discovery has widened the range of genetic engineering applications. Indeed, it has allowed, for the first time ever, to perform all the 8 possible transversions in addition to the already achievable 4 transitions. Consequently, it has permitted the correction of genetic disease mutations in human cells by base-to-base substitutions. Moreover, no other technology enabled to induce directed insertions and deletions which such a great yield of success. Due to the possibility of editing the genome information at user's wish, prime editing has a guaranteed promising future not only in biomedicine but also in plant biotechnology.

4.1. Search of potential candidate genes for prime editing in plants

A deep bibliographic search of recent publications in the fields of genetic engineering and plant breeding resulted in the identification of several potential genes that could be further studied as candidates for prime editing. Among them, several gain-of-function mutations have been reported in different plant species, for instance conferring herbicide tolerance, pathogen resistance, reduced flowering capability or increased fruit size.

4.1.1. Biotic stress resistance genes

Biotic hazards are one of the main causes of the agricultural annual losses. Pathogens are not only able to decrease production yields, but also to synthesize toxins affecting human health. In addition, they are constantly mutating, thus surpassing the applied disease control methods. Plants are able to naturally combat these attacks through several defense mechanisms. As a first barrier, they present a basal defense that is rapidly activated, without being specific for each single pathogen. This first line of defense is mainly mediated by the misnamed secondary metabolites, which in fact are crucial for plants lifespan, once the pathogen associated molecular patterns (PAMPs) are detected by plant receptors. Over the generations, evolution has refined plant defense mechanisms, but pathogenic strategies to overcome them have also progressed. Hence, new treatments are required (Gururani et al., 2012).

Here comes the role of gene mediated resistance. Plants' genome stores a huge set of resistance genes to be transcribed when a specific pathogen infection takes place. Indeed, resistance genes can be either dominant or recessive and their way of action varies from encoding proteins that inactivate toxins produced by pathogens to the commonly known gene-for-gene interactions (Hammond-Kosack & Kanyuka, 2007). The usage of plant resistance genes provides a great range of benefits, such as a reduction of the pathogen population without affecting other beneficial associations, a minor damage to the host plant and an environmentally friendly contribution through the diminished pesticide application by farmers (Corrion & Day, 2015) Due to the increasing necessity of new methodologies, prime editing seems to be the perfect system to directly introduce point mutations that confer resistance to crops. Therefore, several candidates were evaluated for their potential use with the emerging technology.

It has been demonstrated that translation initiation factors are involved in plant resistance against viruses. Specifically, both eukaryotic translation initiation factors 4E and 4F (eIF4E and eIF4F respectively) play a crucial role against potyviruses, the main viral pathogens attacking plants. Those viruses have a positive single-stranded RNA coupled to a viral-encoded protein (VPg) at the 5'-end and a poly-A tail at the 3'-end. Basically, this VPg needs to bind the translation initiation factor for the viral replication. Therefore, mutations in this factor that inhibit its binding to VPg without affecting its functionality might be interesting for plant breeding

(Robaglia & Caranta, 2006). Several groups have been searching for the desired editions. For instance, Piron et al used TILLING (Targeted Induced Local Lesions IN Genomes) with EMS (ethyl methanesulfonate) mutagen coupled to a gene-specific single-nucleotide detection system to create a mutant population of *Solanum lycopersicum* (tomato) and discover agronomic traits of interest related to potyvirus resistance. In tomato, it is known that mutations in eIF4E confer resistance to *Potato virus Y* (PVY) and *Tobacco etch virus* (TEV). After obtaining the mutant population, they inoculated the resulting plants with PVY and TEV plus a *Pepper mottled virus* (Pep-MoV-Texas variety) and screened for a mutant line resistant to all of them. By designing some primers for that genomic area, performing a western blot and obtaining the segregating generation, they confirmed that this mutation was affecting the splicing pattern of eIF4E. The homologous gene where the mutation was firstly identified was named *SI-eIF4E* and the specific edit was numbered as G1485A, indicating a G to A transition at position 1485 from the ATG start codon. Specifically, with this change, the splicing pattern of the mutant was altered, being exons 2 and 3 deleted, thus inhibiting VPg binding (Piron et al., 2010). Similarly, other researchers had already mapped translation initiation factors in *Oryza sativa* (rice) variety Gigante. In particular, Albar et al. found an amino acid change at locus Rymv1 containing an eIF(iso)4G gene that enhanced the recessive resistance to *Rice Yellow Mottle Virus* (RYMT). The polymorphism was located in exon 7, 925 nucleotides downstream from the start codon and it was an A to G transition that induced an amino acid substitution (Glu309Lys) (Albar et al., 2006). With prime editing these mutations could be easily reproduced, using this genetic resource to promote virus resistance in tomato, rice and other crops' breeding programmes, obviously with the corresponding modifications (Piron et al., 2010).

Following the same trend, several authors also investigated the gene-for-gene resistance of *Arabidopsis thaliana* (Arabidopsis) against different strains of bacterial pathogen *Pseudomonas syringae* expressing distinct avirulence (avr) genes. In this gene-for-gene relationship, two complementary genes are required: the resistance (R) gene encoded by the plant host and the avr gene of the bacterial or fungal pathogen. Resistance mechanisms of the plant are only activated when the resulting gene products are detected, following kind of a receptor-ligand model (Van Der Biezen & Jones, 1998). In most cases, plants induce the hypersensitive response (HR), a localized cell death in order to prevent the pathogen to spread, which is mediated by reactive oxygen species synthesis and ion fluxes (Heath, 2000). It was checked that the edition of several genes corresponding to the RPS region (resistance to *P. syringae*), such as RPS2 or RPS3, could enhance even more Arabidopsis resistance against this bacteria (Caicedo et al., 1999; Tao et al., 2000). Although the exact mutation still remains unknown, prime editing could be used to further investigate in this field.

Not only to combat pathogenic infections but also to enhance the ornamental features, gene-for-gene relationship was also studied by other laboratory group that discovered a specific mutation at the DND1 (Defense with No HR Death) locus in Arabidopsis. Obtained mutants were able to combat *P. syringae* attacks without presenting any HR, that is, without any cell death region characteristic of the infected leaves. Furthermore, those mutants upgraded their ability to combat a broad spectrum of different viral, fungal and bacterial hazards (Clough et al., 2000). An A to G point mutation taking place at Trp 290 created a stop codon, thus truncating the final structure of the DND1 protein. Prime editing could be useful to find the corresponding relative in ornamental plants and edit it, hence improving their visual quality.

4.1.2. Abiotic stress resistance genes

Apart from an enhanced resistance to pathogens, other gain-of function mutations could also be achieved with prime editing. For instance, UV light is an unexpected enemy of the plants. Although it seems contradictory since thanks to sun-light plants are able to convert inorganic

matter into organic compounds for their nutrition, high exposure to UV light is a remarkable abiotic stress causing cell death by impairing the metabolism, damaging DNA, proteins and lipids. As it has been discussed above, plants also present a natural response mediated by the secondary metabolism. Some flavonoids, which are phenolic compounds derived from this secondary metabolism, are able to act as plant pigments absorbing the excess of UV light. To do so, they shift the photosynthetic absorption peaks towards the UV region of the spectrum, thus avoiding photoinhibition (light-induced inhibition of the photosynthesis). In particular, anthocyanins are the ones involved in this process and their production is obviously regulated at genomic level. In 2009, a group of researchers was able to identify a mutation in the ELA1 (*enhanced level of anthocyanin*) locus responsible of the increased resistance against extreme UV light exposure firstly observed in one of their mutants. Moreover, an increased cold stress-tolerance and an age-dependant negative regulation of PAP1 (gene encoding a MYB transcription factor that plays an important role activating genes of the anthocyanin biosynthetic pathway) were achieved (Choi et al., 2009).

4.1.3. Genes involved in plant development

Furthermore, other traits of interest had been deeply analysed. In some cases, it could be beneficial to obtain dwarf plants. For that purpose, gibberellic acid (GA) levels should be regulated, since it is the main hormone controlling seed germination, plant elongation and flowering time. This regulation is exerted through DELLA proteins that interact with GA, so DELLA mutants are GA-insensitive since DELLA repressors are not active. Several specific mutations for wheat, rice, barley and Arabidopsis promoting those dwarf phenotypes have been described. A group from Germany wanted to verify the molecular basis of those mutations and, to do so, they worked with Arabidopsis. Several deletions and point mutations were carried out, obtaining quite promising results. It is known that transgenic lines required for the assays could be more easily obtained with prime editing tools (Willige et al., 2007).

To sum up, these studies are just some examples of the work that has been done identifying base-to-base or few nucleotide mutations that can result in beneficial traits summarized in Table 6. Prime editing could eventually reproduce these mutations in any plant species with homologous genes and good transformation amenability facilitating genetic engineering progress.

Table 6. Potential candidate genes for future experiments with prime editing

Trait of interest	Altered gene	Mutation	Plant species	Reference
Potyvirus resistance	eIF4E	G1485A	*Solanum lycopersicum*	(Piron et al., 2010)
Rice Yellow Mottle Virus resistance	eIF4G	G to A transition (Glu309Lys amino acid change)	*Oryza sativa*	(Albar et al., 2006)
Pseudomonas syringae resistance	RPS region	Unknown	*Arabidopsis thaliana*	(Caicedo et al., 1999; Tao et al., 2000)
Pseudomonas syringae resistance without HR	DND1	G to A (Trp290Stop)	*Arabidopsis thaliana*	(Clough et al., 2000)
Enhanced UV and cold resistance	ELA1	Unknown	*Arabidopsis thaliana*	(Choi et al., 2009)
Dwarf phenotype	DELLA	Unknown	*Arabidopsis thaliana*	(Willige et al., 2007)

4.2. Specific *N. benthamiana* genes considered in this study

As it has already been discussed in the above section, a broad spectrum of applications is waiting for the evolution of prime editing. In this final degree thesis, two target genes were selected as a proof of concept for testing the performance of the technique in *N. benthamiana*. The particular choices were made based on previous work done at the laboratory. Specifically, the selected genes had already been analysed and edited with CRISPR-Cas9 and Cas12a by other colleagues.

4.2.1. *Flowering Locus T-5*

On the one hand, *FLOWERING LOCUS T* (FT) was chosen. FT genes belong to the phosphatidyl ethanolamine-binding protein (PEBP) gene family. This family is conserved in a wide range of species, including monocots and dicots, and they are key controllers of plant flowering and architecture. Proteins from this family, such as FT5 and TERMINAL FLOWER 1 (TFL1), act in an opposite way, being TFL1 a repressor and FT an activator. Specifically, TFL1 is expressed in the shot apical meristem (SAM). It represses the transition to flowering and maintains the indeterminate growth of SAM by inhibiting floral meristem identity genes. In other words, it controls when and where flowers are made. The switch to flowering is a complex process necessary for plant adaptation to temperature and day-length and mainly controlled by the FT activator (Hanzawa et al., 2005).

In fact, genes encoding these proteins arose the genetic engineers' interest. Several gain-of-function studies have demonstrated that a single nucleotide edit, promoting an amino acid change, is enough to trigger the conversion from FT activator to TFL1 repressor. Although the whole protein is required to exert full function, both structures are similar enough to act on each other's pathway. A single nucleotide change is involved in the conversion of tyrosine (Tyr) to histidine (His) and vice versa. Specifically, an A to G transition is the key nucleotide involved in this swap. The mutation at issue was interesting for this study since it provided several advantageous features to the plants carrying it. For instance, late-flowering plants are really useful to avoid their reproduction through undesired crosses, thus increasing biosafety of the transgenic lines. In addition, it could also enhance biomass production, since no resources are destinated to flowering, permitting an increased production of substances of interest in the plant (proteins, antibiotics, chemical compounds…)

By using base editing, Kang et al engineered an adenine base editor capable of inducing this point mutation, thus obtaining late-flowering plants in Arabidopsis (Kang et al., 2018). In *N. tabacum* (tobacco) five FT genes, being FT4 and FT5 flowering activators while FT1, FT2 and FT3 are flowering repressors, have been described (Beinecke et al., 2018). A recent study showed the relevant role of FT5 in flowering time in tobacco (Schmidt et al., 2020). *N. benthamiana* is a close relative of tobacco and thus a similar function can be expected for FT5. Hence, it was decided to target FT5 in *N. benthamiana*. Therefore, pegRNAs were designed to introduce a mutation that would result in a Tyr85His change in gene Niben101Scf00863g13007.1-FT5 of *N. benthamiana*. The edition was directed to exon 2, position 250 away from the start inducing a T to C transition (complementary to the previously mentioned A to G change). Thus, a delay of flowering can be expected (Figure 4a).

4.2.2. Acetolactate synthase gene

On the other hand, pegRNAs were also designed to edit the acetolactate synthase (ALS) gene. A considerable family of herbicides have been described to block this enzyme. In fact, those herbicides were firstly commercialized in 1982 and their horribly regulated application boosted

the rapid emergence of a huge range of plants resistant to them. Molecular changes developed by chemical manufacturers permitted the release of up to 50 different ALS-inhibiting herbicides used for selective control of grass and broadleaf crops (Tranel & Wright, 2002).

As mentioned, this kind of herbicide inhibits the activity of ALS. Since this enzyme is directly involved in the biosynthesis of the branched-chain amino acids isoleucine, valine and leucine, the treated plants are unable to produce them and starvation of these amino acids causes death. In addition, other secondary effects have been described, such as the accumulation of 2-ketobutyrate, protein synthesis disruption and photosynthate transport inhibition (Tranel & Wright, 2002).

With respect to the resistant individuals, it is important to consider that the ALS gene is located at a highly variable region of the genome. Regardless the variation, several point mutations have been identified to confer herbicide resistance in different species. The closest relative to *N. benthamiana* is *N. tabacum* (tobacco) and interesting mutations have been described for its ALS. For instance, the tobacco transformation of a mutated gene coming from Arabidopsis conferred resistance to chlorsulfuron herbicide. This gene had been previously edited with a C to T transition at nucleotide 870 promoting a proline (Pro) to serine (Ser) substitution that caused a chlorsulfuron-resistant phenotype in Arabidopsis (Haughn et al., 1988). In the same way, two tobacco mutants were identified, one of them carrying two different mutations and the other just one. In the first case, a C to G transition at nucleotide 587 resulted in an alanine to proline substitution of amino acid 196 and a G to T change at nucleotide 1719 promoted a tryptophan to leucine substitution at amino acid 573. In the second one, nucleotide 588 experimented a C to A transition promoting a glutamine to proline change at position 196 in the enzyme. Both mutants were resistant to the sulfonylurea herbicide application (a herbicide family where chlorsulfuron is included (Lee et al., 1988).

For this study, the chlorsulfuron resistance mutation was interesting since this herbicide is not as much toxic as others that are being currently applied. In addition, it could test the performance of prime editing in a stable transformation since this herbicide could help to select the edited plants. To check the prime editing efficiency, the first described mutation was used. PegRNAs designed in this study were targeting a C to T transition, implying a proline to serine amino acidic change (Pro193Ser) at position 577 in exon 1 of the gene Niben101Scf02892g01012.1 of *N. benthamiana*, thus expecting a chlorsulfuron resistance (Figure 4b).

It is important to remark that the edited genes have been verified in the recently published version of the *N. benthamiana* genome. This release derives from a new sequencing of the genome performed by *Newcotiana* project and it is available for all groups involved in COVID-19 research (https://nbenth.com/annotator/index).

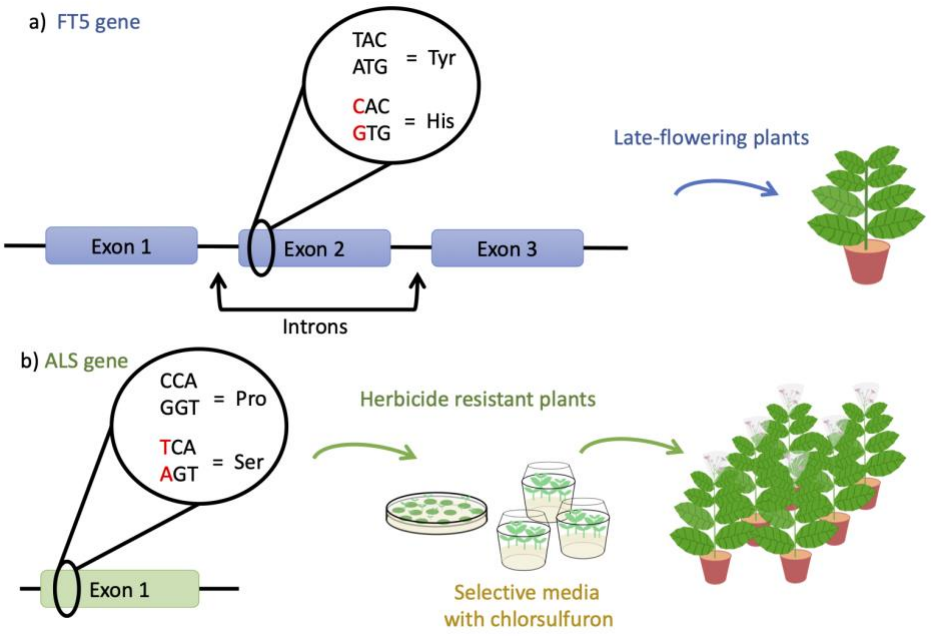

Figure 4. Potentially inserted editions in this study, specifically in the exon 2 of FT5 gene Niben101Scf00863g13007.1 (**a**) and exon 1 of ALS gene Niben101Scf02892g01012.1 (**b**)

4.3. Prime editing genome binding regions design

Once the target genes had already been decided, the corresponding guide RNAs (gRNAs) for this new technique were designed. The lack of experience designing pegRNAs in the host laboratory added to the low number of previous publications proposed an uncertain field where trial-and-error seemed to be the most convenient approach. Actually, it is known that there are lots of factors regulating the editing efficiency and even more if the assays are performed in different host species. Based on guidelines stated by the few published prime editing papers at the beginning of this thesis, the specific design of the pegRNAs for *N. benthamiana* was done as hereunder discussed.

As it has been mentioned in the introduction, protospacers (also known as spacers) are short DNA sequences that compose the CRISPR-Cas locus, each of them coming from a previous pathogenic infection (Figure 2). They are crucial elements in the final gRNA structure. Specifically, the commonly used CRISPR-Cas9 system relies on a gRNA composed of a 5' spacer, which targets the genomic region where the mutation is induced, and a Cas9 scaffold next to it. However, the pegRNA is composed of a 5' spacer followed by the Cas9 scaffold and, additionally, it includes on the 3' end a primer binding site (PBS) and an RT template (Figure 3). Therefore, it presents two regions that bind to the plant genome.

The first one is the typical CRISPR-Cas9 protospacer (PS) that targets the system towards the desired region of the genome. Previous experiments at the laboratory were used as starting point for their design. As other Cas9 approaches, these protospacers targeted a sequence located next to a PAM and obviously, the longer the length, the higher the recognition specificity, thus avoiding off-target edits. A constant length of 20 nucleotides was stablished for all of them, which was more than enough to specify a concrete region of the plant's genome, without binding to any undesired additional site. In that way, efficiency could be compared if this length is changed in future trials.

The second genome binding region is specific for prime editing (PBS-RT), so it had never been included in previous CRISPR-Cas experiments. It is composed of a PBS and an RT template containing the desired edit that will be employed by the RT to insert the mutation through a retrotranscription mechanism. Following the advice provided by Anzalone et al., a PBS of 11 – 15 nucleotides was stablished before the mutation point while the RT template containing the edition varied from 9 – 11 nucleotides.

There is a lack of a common repository and standard nomenclature for gRNAs in plants, making more difficult the comparison of their performance among publications. In a movement towards consensus nomenclature for accurate reporting, at D. Orzáez laboratory a set of standard rules were defined. Standard gRNAs were identified complying the laboratory rules, which follow the pattern *prefixGeneName_StrandPosition*, being *sg* the proper prefix for CRISPR-Cas9 guides, *c* or *nc* for the coding or non-coding strand respectively and *position* the exact position of the PAM – 3 nucleotide (Cas9 cutting site) with respect to the start codon and without considering the introns. However, for the prime editing PBS-RT genome binding region, a new naming approach was proposed since nobody had designed them before in the group. Hence, they were named depending on the protospacer they were paired to and the positions of the first and last edited nucleotides (only the first one for single base-pair mutations) with respect to the start codon. The resulting name followed the structure *sgGeneName_StrandPosition_PEMutPosition*. Obviously, there was no need to specify the strand they were going to anneal with because it was the opposite to the protospacer one. The proposed naming rules are graphically exemplified in Figure 5.

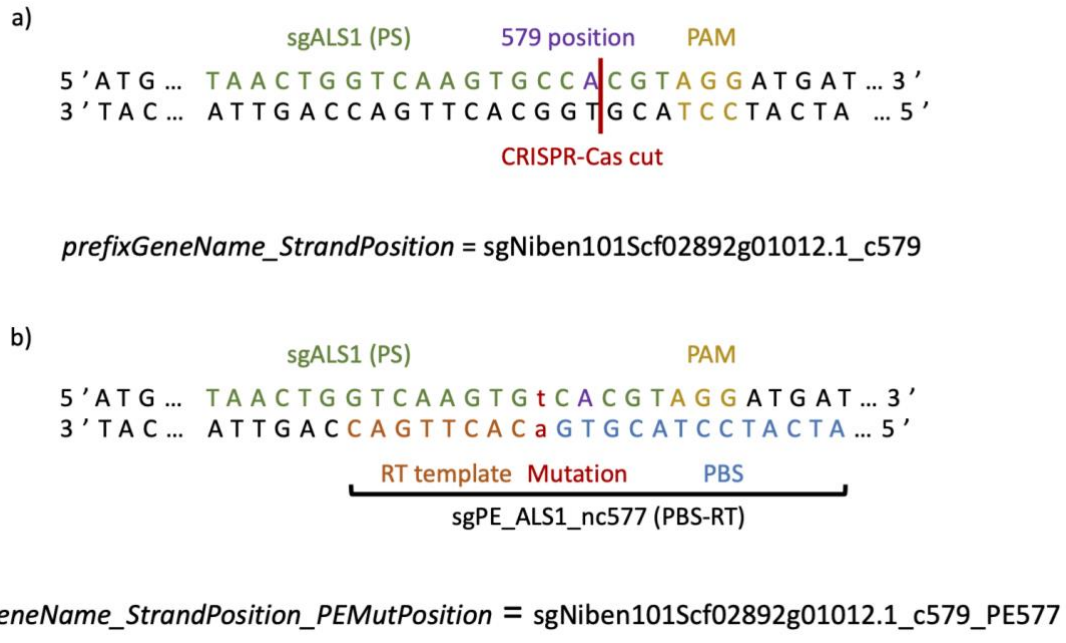

Figure 5. Naming rules for plant molecular biology gRNAs proposed at the laboratory for CRISPR-Cas (**a**) and Prime Editing (**b**) strategies, being exemplified by sgALS1 and sgPE_ALS_nc577 respectively

In the case of ALS, the desired mutation was a C to T transition at position 577. For that purpose, two different protospacers were used (sgALS1 and sgALS2) each of them binding to a nearby region of the coding or non-coding strand respectively. To induce the desired mutation, a specific PBS-RT was designed for each of them (sgPE_ALS1_nc577 and sgPE_ALS2_c577). Additionally, just to test the sequence specific edition of prime editing technique, two more PBS-RTs performing a 3 base-pair (bp) edition were designed, each one following the same design as for sgPE_ALS1_nc577 and sgPE_ALS2_c577 (sgPE_ALS1_nc577-579 and sgPE_ALS2_c577-579, respectively). Moreover, three extra protospacers annealing to the edited sequence were built: sgALSMut1 was carrying the desired 3 bp mutation and it would only bind once either sgPE_ALS1_nc577-579 or sgPE_ALS2_c577-579 had properly mutated the original genomic sequence, sgALSMut2 couldn't bind neither until the 3 bp edition had been done since this edition creates its PAM and gALSPreMut1-C would bind once the 1 bp mutation had been performed, either by the action of sgPE_ALS1_nc577 or sgPE_ALS2_c577.

With regard to the FT5 gene, a similar approach was followed. The desired mutation was a T to C transition at position 250. In the same way, two protospacers were designed (sgFT5.1 and sgFT5.2) and just one PBS-RT, that works properly in combination with both of them (sgPE_FT5_c251). In addition, a mutated protospacer was assembled, as before, that could only bind once the edition had taken place (sgFT5.1Mut). All the obtained sequences are exemplified in Table 7.

Table 7. Designed protospacers and genome binding regions both for ALS and FT5, including the sgRNAs and pegRNAs of each gene. In the case of the pegRNAs, the PBS is written in green, the edition in red and the RT template in purple (considering the edition as part of it).

NAME	NICKNAME	DESCRIPTION	SEQUENCE (5'-3')
sgNiben101Scf02892g01012.1_c579	sgALS1	Binds first ALS gene region	TAACTGGTCAAGTGCCACGT
sgNiben101Scf02892g01012.1_nc583	sgALS2	Binds second ALS gene region	TCAGTACCGATCATCCTACG
sgNiben101Scf02892g01012.1Mut_c579	sgALSMut1	Carries 3 bp edition	TAACTGGTCAAGTGtccCGT
sgNiben101Scf02892g01012.1Mut_nc584	sgALSMut2	3 bp edition creates its PAM	ATCAGTACCGATCATCCTAC
sgNiben101Scf02892g01012.1PreMut_c579	sgALSPreMut	Carries 1 bp edition	AAATCTAAGAGAAcACCTCCATTGG
sgNiben101Scf02892g01012.1_c579_PE577	sgPE_ALS1_nc577	Bound to gALS1 - it edits 1 bp	ATCATCCTACGTGaCACTTGAC
sgNiben101Scf02892g01012.1_c579_PE577-579	sgPE_ALS1_nc577-579	Bound to gALS1 - it edits 3 bp	CGATCATCCTACGggaCACTTGAC
sgNiben101Scf02892g01012.1_nc583_PE577	sgPE_ALS2_c577	Bound to gALS2 - it edits 1 bp	AACTGGTCAAGTGtCACGTAGGAT
sgNiben101Scf02892g01012.1_nc583_PE577-579	sgPE_ALS2_c577-579	Bound to gALS2 - it edits 3 bp	AACTGGTCAAGTGtccCGTAGGAT

sgNiben101Scf00863g13007.1-FT5_nc241	sgFT5.1	FT5 guide	AAAGAGATGCTAACCAATGG
sgNiben101Scf00863g13007.1-FT5_nc251	sgFT5.2	FT5 guide	TGGAGGTATTCTCTTAGATT
sgNiben101Scf00863g13007.1-FT5Mut_c255	sgFT5.1Mut	FT5 guide	ATCTAAGAGAAcACCTCCAT
sgNiben101Scf00863g13007.1-FT5_nc241_PE251 / sgNiben101Scf00863g13007.1-FT5_nc258_PE251	sgPE_FT5-C251	Bound to gFT5.1/gFT5.2 - it edits 1 bp	AAATCTAAGAGAAcACCTCCATTGG

4.4. Prime editing cloning with *GoldenBraid*

Since prime editing was a completely new technique in the laboratory, all the required DNA pieces should be adapted and refined in order to be used with a specific cloning method. That means they had to comply to a grammar stablished by the assembly method used for cloning. Thus, once the individual genome binding regions had been properly designed, they were adapted and assembled with all the other required DNA elements to compose the definitive pegRNAs.

In this particular case, the *GoldenBraid* (GB) assembly system was chosen. The aim of this method is to facilitate plant synthetic biology multigene engineering procedures by connecting small DNA components to obtain higher order modules and genetic devices. It is based on type IIS restriction enzymes and relies on the same principles as the *Golden Gate* cloning strategy. Compared to type II, which are able to cut within the sequence they recognize, type IIS enzymes cut a fixed number of nucleotides away from the restriction site. For that reason, they promote a higher assembly efficiency since from one cloning step to the other, the restriction sites are not conserved, thus avoiding cyclic cleavages. More in detail, when cutting, these enzymes leave the so-called overhangs, which are protruding nucleotides forming sticky ends at the edges of the cleaved DNA sequence. Those regions were used to define the GB grammar, allowing the combination of standard DNA parts (Patron et al., 2015). Furthermore, this system is modular, that is, the different DNA parts bordered with same overhangs can be easily exchanged and reused, thus widening the range of possibilities. Hence, with a couple of type IIS restriction enzymes (*BsmBI* and *BsaI*) and several entry and destination vectors (Table 1), all possible multipartite assemblies can be performed (Sarrion-Perdigones et al., 2011). It is important to remark that all the already available GB elements were extracted from the GB glycerol stocks and DNA sequences collection, stored at -80°C ultra-freezer mixed with glycerol and classified depending on their GB number. In fact, all the generated assemblies from this work were also uploaded to the online version of this collection found at https://gbcloning.upv.es/search/features/

As mentioned, prime editing elements required a new adaptation since the already available GB design schemes did not satisfy their needs. Hence, the crucial point of this study was to develop a new assembly strategy able to include all the additionally required elements for building a pegRNA. A new GB grammar specifically devoted to pegRNAs was developed. The pegRNA grammar along with the new DNA parts generated during this work will also be included in the set of tools of the *GoldenBraid* website devoted to genome editing (https://gbcloning.upv.es/tools/grna/).

Prime editing relies on the expression of two independent transcription units (TU), what in synthetic biology is defined as a genomic region composed of a promoter, a coding region and a terminator. In this particular case, one of them was encoding for the pegRNAs and the other

for the Cas9-RT enzymatic complex. Different assembly strategies were followed for their construction. Thus, they are independently exposed in the following sections.

4.4.1. Prime editing guide RNAs' transcriptional unit cloning

PegRNAs were encoded by the first TU. As it has already been discussed, their structure differed from traditional CRISPR-Cas gRNAs. Consequently, their assembly was also different, being one of the main purposes of this work to describe it. It is important to consider that the designed strategy to clone this transcriptional unit is based on 3 different elements. On the one hand, two standard GB parts, that is, two DNA elements cloned into the entry vector pUPD2 were created in this study and made available at the GB glycerol stocks collection so they can be reused in future assays (U6-26_tRNA and scaffold). On the other hand, a variable element that depends on the genomic region to edit. This last element is composed of two genome binding sequences (protospacer and PBS-RT) that should be designed at user's wish in the form of primer-dimers. The overall transcription unit presented the structure U6-26_tRNA_PS_Scaffold_PBS-RT. Throughout this section, the individual components are going to be analysed and the final assembly will also be described.

Like all TUs, a promoter was located at the first place. In this case, the Arabidopsis U6-26 RNA polIII promoter was used. This one is highly recommended for gRNAs expression and it has been widely used by the plant genetic engineers' community. It both ensures a precise transcription from the very beginning of the coding region, without any subsequent processing signal (cap, polyadenylation…) and provides high levels of transcripts' accumulation. However, the main drawback is that this promoter imposes a G as the first nucleotide to initiate transcription, causing problems in the design of the subsequent protospacer in some cases. Thus, an additional element providing the required G was added to the design. In this case the Gly tRNA was introduced, a non-functional sequence whose main purpose was to provide that G. After the transcription, this sequence is cleaved by the plant endogenous ribonucleases P and Z and eliminated releasing the clean protospacer and scaffold of the gRNA. To construct the first GB standard part, both U6-26 and tRNA were needed. Their sequence was extracted from GB2135 vector (GB glycerol stocks collection) and domesticated through a PCR performed with specific primers containing the desired overhangs. A GB restriction-ligation reaction was used to clone the purified PRC products into the entry vector pUPD2, thus obtaining the first constant part of the TU (Figure 7).

Domestication of the second GB standard part required for the final TU assembly did not differed that much from the previous one. In this case, the target element to be cloned was the scaffold, a specific sequence required by the Cas9 to properly bind to its corresponding gRNA. As before, a PCR was performed to amplify the scaffold sequence originally placed in vector GB2245 of the collection. Specially designed primers contained the corresponding overhangs for its afterwards cloning into pUPD2 entry vector, obtaining the second invariable element of the pegRNA final TU (Figure 7).

At this point, all the necessary GB standard parts had been domesticated. They were stored at the GB collection, being available for future studies. Then, the only remaining DNA pieces were the variable pegRNA parts. Those should be defined by the researcher depending on the gene to edit. On the basis of the selected targets in this study, two different genome binding regions were defined, whose sequences were presented in the previous section. Unlike other genome editing techniques, prime editing comprised two different genome binding regions for its pegRNAs. The protospacer (PS), which is commonly used in CRISPR-Cas assays to target and direct the complex towards a specific region of the genome and the newly developed region composed of a primer binding site and an RT-template (PBS-RT) containing the mutation. Each

of them was assigned a determined position in the final scheme, thus requiring specific overhangs for the GB assembly. Starting from Table 7 theoretical sequences, specific primers were designed including the desired overhangs. Since primers were partially complementary (except for the overhangs), the corresponding pairs were left to form primer-dimers in order to create the double-stranded DNA sequences with the required sticky ends for their further assembly (Figure 6).

On the one hand, the commonly used protospacers were flanked with the GTGC overhang conferred by the forward primer and the CAAA provided by the reverse one. An additional A was inserted right after the 5'-overhang of the coding strand. This nucleotide was part of the tRNA, so it was required for its proper release. An alternative that could have been followed was to use TGCA instead of GTGC as the forward primer overhang. However, it is a palindromic sequence (same nucleotides are obtained when reading from 5'-3' in one strand and from 3'-5' in the other) and it could cause a decrease of the restriction-ligation efficiency, since some protospacers spacers would bind in the correct sense and some others wouldn't. Therefore, the first approach was chosen.

On the other hand, the newly developed PBS-RT presented a CGGT overhang at their forward primer and a GCGA at their reverse one. In this case, a poly-A tail was inserted right before the overhang provided by the reverse primer in order to stop transcription, since this was the final element included into the TU assembly. It should be mentioned that right before the PBS-RT sequence, after the overhang provided by the forward primer, additional GC bases were introduced. That was a crucial step in order to avoid the GTGC overhang repetition, since it was already used for the correct insertion of the protospacers and, if repeated for these PBS-RTs, the GB assembly would miss the intermediate parts.

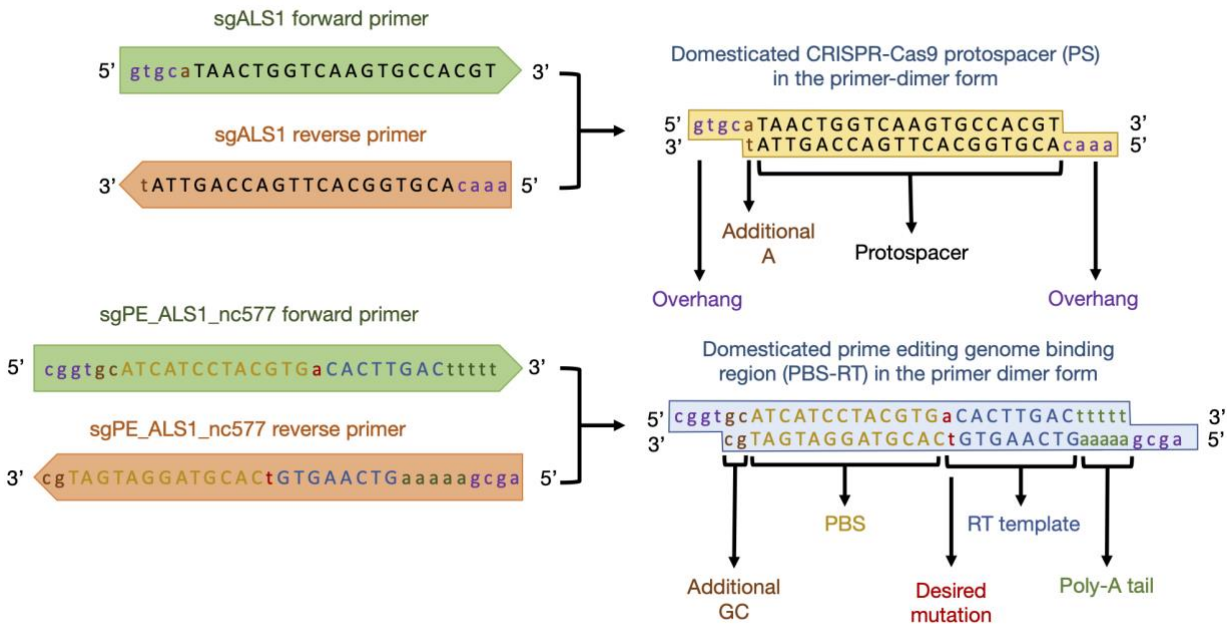

Figure 6. Specific primers designed and resulting primer-dimer sequences both for sgALS1 and sgPE_ALS1_nc577, a protospacer and a PBS-RT genome binding region respectively.

At that point, the GB standard parts for the construction of the definitive pegRNAs' TU were available in the circular entry plasmid pUPD2 (U6-26_tRNA and scaffold). Moreover, both variable genome binding regions targeting the specific gene to edit were also in the form of primer dimers with the corresponding overhangs. Hence, all of them were simply combined in a destination vector pDGB3_α1 taking profit of the *GoldenBraid* assembly system versatility. Via a single all-in-one-tube restriction-ligation reaction, the final guides were "scarless" assembled in the correct order, owing to a properly deliberated design of the individual pieces (Figure 7). When ligated, *E. coli* TOP10 competent cells were transformed and plated. From each Petry dish, two colonies were picked and one them was sequenced, achieving a perfect alignment in all the designed pegRNAs. Hence, the results showed a 100% assembly efficiency, confirming that the newly designed approach could successfully be used in further experiments.

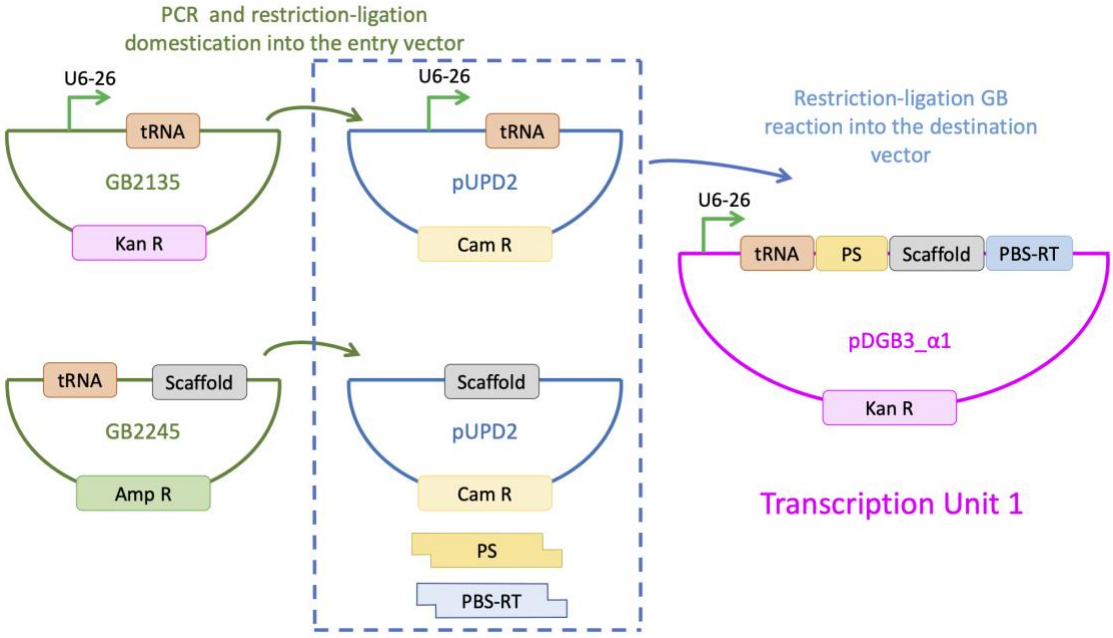

Figure 7. **Assembly of the first TU (pegRNA)** based on two GB standard parts (pUPD2:U6-26_tRNA and pUPD2_Scaffold) and two variable elements (PS and PBS-RT) designed depending on the region to edit in the form of primer-dimers.

4.4.2. Standard guide RNAs' transcriptional unit cloning

The protospacers played a key role in the designed prime editing strategy, being the directors of the enzymatic complex towards the place to mutate. Nevertheless, an additional experiment was suggested for them. They were domesticated using the common procedure described by Vazquez-Vilar and colleagues at their most recent version *GoldenBraid 3.0.* Due to the vast experience assembling those type of gRNAs, no modification was required during the process (Vazquez-Vilar et al., 2016). Thus, standard gRNAs apart from pegRNAs were used to test the efficiency of the Cas9 H840A nickase. The aim was to control that gRNAs were efficiently cutting at the targeted sites, thus discarding low cutting efficiencies of those gRNAs if pegRNAs were not showing the expected results.

Specifically, a two-steps ligation was required for their assembly. Initially, vector GB2245 was used to provide the required sequences into pUPD2 entry vector. This plasmid derived from the GB glycerol stocks and DNA sequences collection and it contained a tRNA sequence and a Cas9

scaffold. Due to the designed protospacers' overhangs, they were inserted between the two previously mentioned elements, obtaining a tRNA_PS_scaffold sequence without intermediate undesired nucleotides. Once into the pUPD2 entry vector, this three-elements fragment was introduced into a pDGB3_α1 destination vector through a second restriction-ligation reaction, including a U6-26 promoter coming from GB1001 right before the tRNA. As a result, the final gRNAs' TU (completely different from the pegRNAs' one) was assembled also in a pDGB3_α1 vector containing the sequence U6-26_tRNA_PS_Scaffold from 5' to 3' end (Figure 8). It should be mentioned that this TU was not required for the prime editing strategy, but it was assembled to perform the mentioned verification.

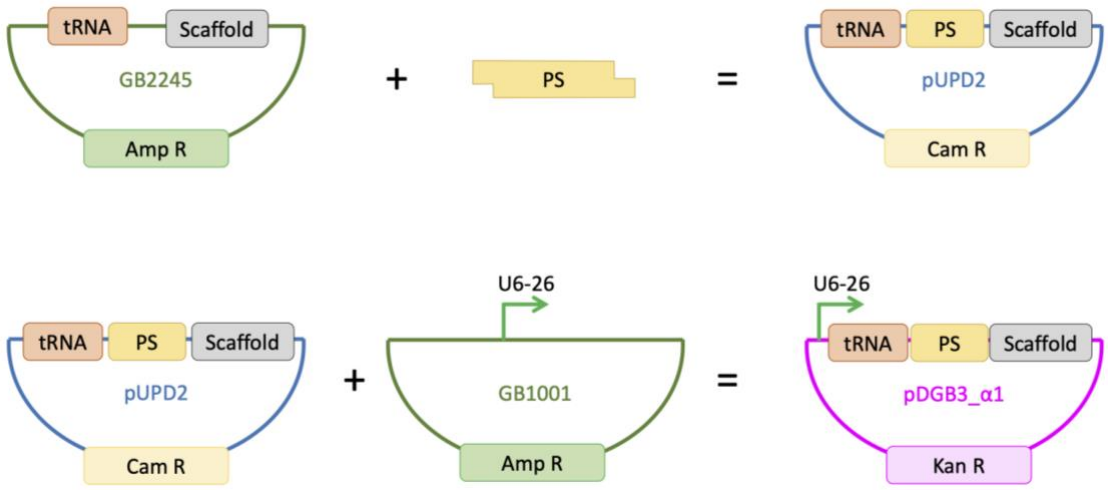

Figure 8. Assembly of the additional standard gRNA TU not required by prime editing strategy. The procedure described by Vazquez-Vilar et al. in 2016 was followed to obtain the final construct U6-26_tRNA_PS_Scaffold in pDGB3_α1 destination vector

In order to check the correct assembly of both steps, a restriction enzyme analysis was performed after each restriction-ligation reaction. For the first one, pUPD2 plasmids were cleaved with *BsaI* obtaining two different bands on the agarose gel: 184 bp and 2105 bp. Conversely, pDGB3_α1 vectors resulting from the second restriction-ligation reaction were cut with *EcoRI,* thus expecting to obtain 2 bands of 6345 bp and 438 bp. Correct assembly results are represented in Figure 9.

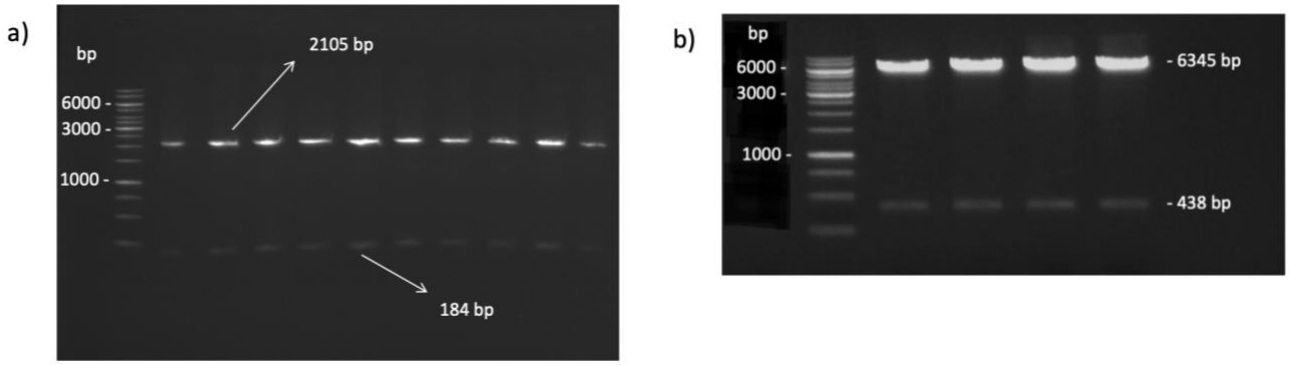

Figure 9. Agarose gel electrophoresis obtained for the restriction enzyme analysis of standard gRNAs, specifically of pUPD2:tRNA_PS_Scaffold (**a**) and pDGB3_α1:U6-26_tRNA_PS_Scaffold assemblies (**b**).

4.4.3. Cas-RT transcriptional unit cloning

Till this point, the definitive pegRNAs had been successfully cloned in order to compose the first TU necessary for prime editing. Nonetheless, additional elements were required for the proper functioning of the system. Indeed, all the enzymatic complex in charge of exerting the desired mutation was lacking. The assembly approach designed to obtain the second TU followed a completely different scheme that will be hereunder described. Actually, one of the main differences among prime editing and other genome editing methodologies was its novel combination of enzymes able to trigger editions through the retrotranscription pathway. For that purpose, it required a Cas9-RT fusion. Several studies have reported that Cas9 is composed of 2 different domains, one of them able to nick the strand where the enzyme binds and the other the complementary strand. Among all the available mutated variants of Cas9, the H840A was chosen for this strategy. The mutation carried by this nickase version allows it to cut only the complementary strand, so it was perfect for prime editing aim. Cas9 H840A was coupled to a commercial Moloney murine leukemia virus (M-MLV) RT variant, as exposed in the above sections. It is important to mention that the Cas9 had a codon optimization for humans, although it was properly expressed in *N. benthamiana*, whereas the RT sequence codons were optimized for plant uses with the *Integrated DNA Technologies* tool.

In order to assemble this TU, Cas9 H840A sequence was firstly obtained from GB1692. This vector contained the Cas9 H840A and a consecutive NLS (nuclear localization signal in order to be directed towards the nucleus to perform its function) at its 3'-end. However, Anzalone et al. allocated the NLS before the Cas9, at the 5' end. Thus, the already existing GB vector was not a hundred percent useful and the order of the DNA should be inverted. To do so, a huge forward primer containing the NLS and 20 additional nucleotides able to align within the beginning of Cas9 sequence was designed. It was ordered as an ultramer DNA oligo. On the other side, a reverse primer aligning exactly several base pairs before the 3' edge of Cas9 was used, thus getting rid of the subsequent NLS from the original vector. Obviously, the designed primers were carrying overhangs to proceed with the domestication. Once purified, PCR products were inserted into pUPD2 entry vector due to these identity parts.

On the other site, the M-MLV RT sequence was directly ordered as a gBlock also containing the desired overhangs. This means that a double-stranded DNA sequence was directly acquired, not only containing the RT sequence but also a linker on its 5'-end, reproducing exactly Anzalone et al. design. Since a chimeric protein was been expressed, the linker was required to confer enough flexibility for the correct functioning of the enzymatic subunits, without any steric effects. In the same way, the gBlock was also inserted into pUPD2, thus obtaining all the required elements for the second TU.

Like other TUs assembled for synthetic biology experiments, additional regulatory elements were required for its expression within the plant. In order to build it properly, a promoter and a terminator were missing. *GoldenBraid* glycerol stocks were used to obtain the remaining pieces. In detail, GB0030 and GB0037 provided the *CaMV* 35S promoter and the *A. tumefaciens* Tnos terminator respectively, the most widely used regulatory elements in plant biotechnology. Both were contained in pUPD2 entry vectors, so the final assembly could be easily carried out at this point. By means of a unique restriction-ligation reaction, the TU was successfully assembled into a pDGB3_α2 destination vector achieving the predicted sequence 35S_Cas9_NLS_Linker_RT_Tnos (Figure 10). In fact, the detailed domestication of these pieces was done through the *GoldenBraid* TU assembler tool found at https://gbcloning.upv.es/do/domestication/, being NLS_Cas9 a B3-B4 piece and Linker_RT a B5. Specific assembly guidelines were stated by Vazquez-Vilar and colleagues in 2017 (Vazquez-Vilar et al., 2017)

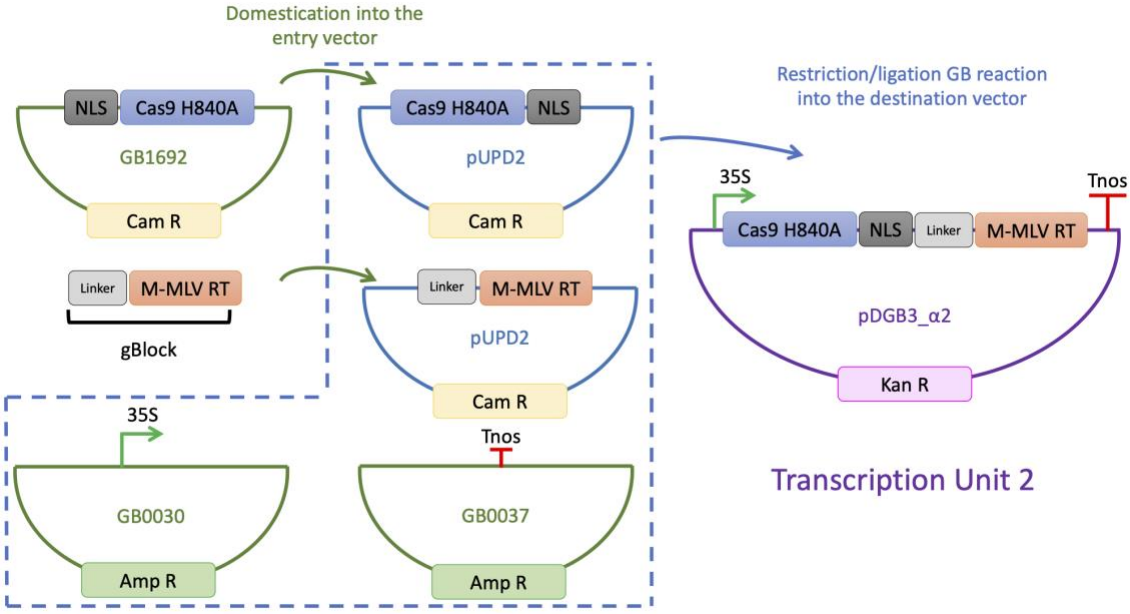

Figure 10. Assembly of the second TU (Cas9-RT) build from the construct 35S_Cas9_NLS_Linker_RT_Tnos into pDGB3_α2

In order to check the assembled construct, two independent colonies of the Cas9-RT transcriptional unit were digested with *BglII* and *BamHI.* In the case of the first enzyme, 3 bands of 937 bp, 3800 bp and 9527 bp were expected and for *BamHI*, 2 bands of 3776 bp and 10488 bp were predicted. The obtained results gel is shown in Figure 11, demonstrating the successful assembly.

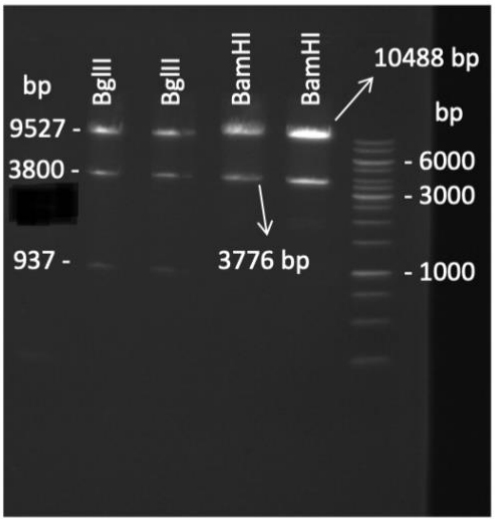

Figure 11. Agarose gel electrophoresis obtained for the restriction enzyme analysis of Cas9-RT TU. Two individual colonies were digested with *BglII* and *BamHI* obtaining the predicted bands, being the first and third columns colony 1 and second and forth columns colony 2.

4.5. Agroinfiltration experiment design

All the designed constructs aimed to be infiltrated into *N. benthamiana* plants for testing the performance of the assembled constructs. Hence, the obtained destination vectors (pDGB3_α1 with guide RNAs and pDGB3_α2 with Cas9-RT complex) were transformed into C58 *A. tumefaciens* cells, a bacterial strain used to further on infiltrate the plants. They were plated and afterwards individual colonies were grown in liquid culture to later on extract the corresponding plasmids. Restriction enzyme assays were used to check the proper insertion of the plasmids into the bacteria. At this point, the required pieces to be infiltrated into the plants were ready. *In vivo* test experimental design was thought to check all the possible combinations and compare them in order to achieve relevant results. In particular, the ALS gene was firstly selected for its edition.

An overall number of 12 5-weeks *N. benthamiana* plants were used, each of them agroinfiltrated with 4 different elements. To begin with, all solutions contained p19, a gene-silencing repressor encoded by tomato bushy stunt virus (TBSV) able to maintain high levels of transient expression and inhibit the post-transcriptional gene silencing (Qiu et al., 2002). Plants 1 to 4 were used for testing the previously described prime editing approach, involving the infiltration of a pegRNA, the corresponding mutated CRISPR-Cas guide RNA and the Cas9-RT fusion complex. To compare its editing efficiency with the standard Cas9 editing approach, plants 5 to 8 were infiltrated with equivalent guide RNAs but, in this case, using the common Cas9 for the edition. Furthermore, the editing efficiency of the Cas9-RT complex was tested in plants 8 to 12 via the introduction of only the pegRNA, thus allowing to compare the results with the ones obtained in prime editing approach (plants 1 to 4) using two guide RNAs. It is important to consider that plants 5 to 12 just required 3 elements, so an Agrobacterium strain carrying an empty plasmid (stuffer fragment) was co-infiltrated. This one is a non-functional sequence aimed to discard possible variations in editing efficiencies derived from the variable total number of Agrobacterium strains co-introduced. The specific elements infiltrated in each plant are summarized in Table 8. Because of the exceptional situation due to COVID-19, plant infiltrated samples were collected at 4 days post infiltration and stored at -80°C but genomic DNA couldn't be extracted in order to check the presence of the intended mutations. Genomic DNA extraction followed by targeted site PCR amplification and subsequent Sanger sequencing of the obtained PCR products will be done as soon as possible.

Table 8. Specific constructs agroinfiltrated in each of the 12 *N. benthamiana* plants used for the *in vivo* test of ALS target gene editing.

Constructs	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
sgPE_ALS1_nc577	X					X				X		
sgPE_ALS1_nc577-599			X									
sgPE_ALS2_c577		X						X				X
sgPE_ALS2_c577-599				X								
sgALS1_C					X				X			
sgALS2_NC							X				X	
sgALSMut1C												
sgALSMut2NC			X	X								
sgALSPreMut1C	X	X								X		
Cas9 H840A_M-MLV RT	X	X	X	X					X	X	X	X
Cas9					X	X	X	X				
P19	X	X	X	X	X	X	X	X	X	X	X	X
Stuffer fragment					X	X	X	X	X	X	X	X

5. General discussion and future perspectives

All over this final degree thesis, every single detail regarding the prime editing system has been deeply analysed. The main purpose was to describe the most convenient prime editing approach to be used in *Nicotiana benthamiana.* That implied to define the formerly unknown guidelines for adapting the required constructs to *GoldenBraid* assembly system and to test them *in planta*. Due to the exceptional situation, not all the objectives could be overcome on time. However, the most important part, which was the domestication of the DNA pieces, was successfully done, thus providing the basis for future studies of the technique in the laboratory. All the created vectors were stored in the *GoldenBraid* glycerol stocks and DNA sequences collection of the laboratory. Their specific GB numbers are compiled in Supplementary Table 3.

In this particular case, *GoldenBraid* assembly method was selected. Like *Golden Gate* and *MoClo* methods, GB is based on type IIS enzymes, what enables to choose the desired overhangs, not happening with previous systems based on type II restriction enzymes (*BioBricks*). In addition, they do not insert extra nucleotides between the assembled elements, promoting a scarless ligation of the TUs, what is crucial for fusion proteins and guide RNAs assemblies. Compared to *MoClo*, which is its direct competitor since both are devoted to plant molecular biology applications, GB reduces the total number of destination vectors used and favours the reusability of previously generated TUs and genetic devices but at cost of requiring more steps for larger assemblies (Casini et al., 2015).

Fortunately, not so many issues were found during the experimental work. All the DNA parts were correctly assembled. In fact, only two individual white colonies were picked per construct, just in case one of them was wrong. Contrary to what was expected, even for the innovative pegRNA cloning approach, all the analysed colonies aligned perfectly with the theoretically predicted constructs. In this way, it has been demonstrated that with a properly premeditated design, it is possible to insert two DNA elements (protospacer and PBS-RT in the current study) in the form of primer-dimers into a destination vector through a *GoldenBraid* all-in-one-tube restriction-ligation reaction, what had never been done before. Regarding the PCRs, all of them amplified the desired fragments with not so much trouble except for the NLS_Cas9 H840A. It took several attempts to finally get the optimal conditions for its amplification. The 4.2 Kb fragment was too large to be handled as the shorter ones. Nothing new, since it was already known from previous colleagues' experiments.

The main previous study from which this thesis emerged was reported by Anzalone et al. in 2019. However, during the experimental realization, Lin and colleagues published a plant optimized version of prime editing. Although some points had been equally followed, some others provided useful information to the thesis. For instance, they demonstrated that nicking the non-edited strand was worthless in plant hosts. It did not enhance its repair. It is clear that plant cellular machinery differs a lot from animals, so that should be the main reason why repair mechanisms are not favoured in the same way. This fact was not taken into account in the experimental design of this work but should be considered for further optimization. Apart from that, the rest of experiments described editing efficiencies, variable protospacer lengths and experimental temperatures that provide useful information for planning future experiments in plants. Considering that different hosts react in a distinct way to all those variables, further investigation is required in each case, that is why a valid approach for *Nicotiana benthamiana* was aimed to be evaluated in this thesis.

In spite of the large amount of research projects that are still pending to take full profit of all different possibilities provided by this technique, prime editing has already demonstrated to be a really powerful tool that will for sure play a crucial role in the future of genome editing field.

This newly developed gene editing approach overcomes the base editing technology since it allows to trigger all possible single-base-pair mutations and even programmed indels of several nucleotides length. Indeed, the mentioned mutations can be performed without the need of neither DSBs nor donor DNA templates. The absence of DSBs minimizes the possibilities of getting undesired mutations at the edition point and the lack of need of a donor DNA template simplifies the experimental design while increases the chances of getting the desired edition in plants, where the homology-directed repair pathway is less frequently used than in animals.

Never before, the ability to directly induce genome editions at user's wish with such ease had been that accessible. Hence, several genes have been proposed as possible candidates for its application, widening the currently available pool of targets. Maybe, the correct guidance of this method could help human being to face up all the emerging challenges, for instance feeding an increasing population, adapting agriculture to the changing environmental conditions and enhancing crops' production yields.

6. Conclusions

The main conclusions extracted from this study were:

- A prime editing strategy was successfully defined for its application in *Nicotiana benthamiana.*
- Standard rules for naming pegRNAs were proposed.
- All the required DNA elements, including the Cas9-RT TU, 6 pegRNAs and 8 standard gRNAs, were correctly cloned using the *GoldenBraid* assembly method. The corresponding *in vivo* test results are still pending in order to check the editing efficiency of the synthesized constructs.
- A set of candidate genes were proposed in order to further investigate the great possibilities of the prime editing technique.

7. Bibliography

7.1. Papers

ALBAR, L., BANGRATZ-REYSER, M., HÉBRARD, E., NDJIONDJOP, M. N., JONES, M., & GHESQUIÈRE, A. (2006). Mutations in the eIF(iso)4G translation initiation factor confer high resistance of rice to Rice yellow mottle virus. *Plant Journal*, *47*(3), 417–426. https://doi.org/10.1111/j.1365-313X.2006.02792.x

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., & LIU, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. In *Nature* (Vol. 576, Issue 7785). https://doi.org/10.1038/s41586-019-1711-4

BEADLE, G. W. (1980). The Ancestry of Corn. In *Scientific American* (Vol. 242, pp. 112–119). Scientific American, a division of Nature America, Inc. https://doi.org/10.2307/24966237

BEINECKE, F. A., GRUNDMANN, L., WIEDMANN, D. R., SCHMIDT, F. J., CAESAR, A. S., ZIMMERMANN, M., LAHME, M., TWYMAN, R. M., PRÜFER, D., & NOLL, G. A. (2018). The FT/FD-dependent initiation of flowering under long-day conditions in the day-neutral species *Nicotiana tabacum* originates from the facultative short-day ancestor *Nicotiana tomentosiformis*. *The Plant Journal*, *96*(2), 329–342. https://doi.org/10.1111/tpj.14033

CAICEDO, A. L., SCHAAL, B. A., & KUNKEL, B. N. (1999). Diversity and molecular evolution of the RPS2 resistance gene in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(1), 302–306. https://doi.org/10.1073/pnas.96.1.302

CASINI, A., STORCH, M., BALDWIN, G. S., & ELLIS, T. (2015). Bricks and blueprints: Methods and standards for DNA assembly. *Nature Reviews Molecular Cell Biology*, *16*(9), 568–576. https://doi.org/10.1038/nrm4014

CHOI, S., KWON, Y. R., HOSSAIN, M. A., HONG, S. W., LEE, B. HA, & LEE, H. (2009). A mutation in ELA1, an age-dependent negative regulator of PAP1/MYB75, causes UV- and cold stress-tolerance in Arabidopsis thaliana seedlings. *Plant Science*, *176*(5), 678–686. https://doi.org/10.1016/j.plantsci.2009.02.010

CLOUGH, S. J., FENGLER, K. A., YU, I. C., LIPPOK, B., SMITH, R. K., & BENT, A. F. (2000). The Arabidopsis dnd1 "defense, no death" gene encodes a mutated cyclic nucleotide-gated ion channel. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(16), 9323–9328. https://doi.org/10.1073/pnas.150005697

CORRION, A., & DAY, B. (2015). Pathogen Resistance Signalling in Plants. *ELS*, 1–14. https://doi.org/10.1002/9780470015902.a0020119.pub2

GAJ, T., GERSBACH, C. A., & BARBAS, C. F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. In *Trends in Biotechnology* (Vol. 31, Issue 7, pp. 397–405). Elsevier Current Trends. https://doi.org/10.1016/j.tibtech.2013.04.004

GAUDELLI, N. M., KOMOR, A. C., REES, H. A., PACKER, M. S., BADRAN, A. H., BRYSON, D. I., & LIU, D. R. (2017). Programmable base editing of T to G C in genomic DNA without DNA cleavage. *Nature*, *551*(7681), 464–471. https://doi.org/10.1038/nature24644

GURURANI, M. A., VENKATESH, J., UPADHYAYA, C. P., NOOKARAJU, A., PANDEY, S. K., & PARK, S. W. (2012). Plant disease resistance genes: Current status and future directions. *Physiological and Molecular Plant Pathology*, *78*, 51–65. https://doi.org/10.1016/j.pmpp.2012.01.002

HAMMOND-KOSACK, K. E., & KANYUKA, K. (2007). Resistance Genes (R Genes) in Plants . *ELS*. https://doi.org/10.1002/9780470015902.a0020119

HANZAWA, Y., MONEY, T., & BRADLEY, D. (2005). A single amino acid converts a repressor to an activator of flowering. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(21), 7748–7753. https://doi.org/10.1073/pnas.0500932102

HARRIS, D. R. (1967). New Light on Plant Domestication and the Origins of Agriculture: A Review. *Geographical Review*, *57*(1), 90. https://doi.org/10.2307/212761

HAUGHN, G. W., SMITH, J., MAZUR, B., & SOMERVILLE, C. (1988). Transformation with a mutant Arabidopsis acetolactate synthase gene renders tobacco resistant to sulfonylurea herbicides. *MGG Molecular & General Genetics*, *211*(2), 266–271. https://doi.org/10.1007/BF00330603

HEATH, M. C. (2000). Hypersensitive response-related death. In *Programmed Cell Death in Higher Plants* (pp. 77–90). Springer Netherlands. https://doi.org/10.1007/978-94-010-0934-8_6

HILDER, V. A., & BOULTER, D. (1999). Genetic engineering of crop plants for insect resistance - A critical review. In *Crop Protection* (Vol. 18, Issue 3, pp. 177–191). Elsevier BV. https://doi.org/10.1016/S0261-2194(99)00028-9

KANG, B. C., YUN, J. Y., KIM, S. T., SHIN, Y. J., RYU, J., CHOI, M., WOO, J. W., & KIM, J. S. (2018). Precision genome engineering through adenine base editing in plants. *Nature Plants*, *4*(7), 427–431. https://doi.org/10.1038/s41477-018-0178-x

KIM, Y. B., KOMOR, A. C., LEVY, J. M., PACKER, M. S., ZHAO, K. T., & LIU, D. R. (2017). Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nature Publishing Group*, *35*(4). https://doi.org/10.1038/nbt.3803

KOMOR, A. C., KIM, Y. B., PACKER, M. S., ZURIS, J. A., & LIU, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, *533*(7603), 420–424. https://doi.org/10.1038/nature17946

LEE, K. Y., TOWNSEND, J., TEPPERMAN, J., BLACK, M., CHUI, C. F., MAZUR, B., DUNSMUIR, P., & BEDBROOK, J. (1988). The molecular basis of sulfonylurea herbicide resistance in tobacco. *The EMBO Journal*, *7*(5), 1241–1248. https://doi.org/10.1002/j.1460-2075.1988.tb02937.x

LIN, Q., ZONG, Y., XUE, C., WANG, S., JIN, S., ZHU, Z., WANG, Y., ANZALONE, A. V., RAGURAM, A., DOMAN, J. L., LIU, D. R., & GAO, C. (2020a). Prime genome editing in rice and wheat. *Nature Biotechnology*. https://doi.org/10.1038/s41587-020-0455-x

LIN, Q., ZONG, Y., XUE, C., WANG, S., JIN, S., ZHU, Z., WANG, Y., ANZALONE, A. V., RAGURAM, A., DOMAN, J. L., LIU, D. R., & GAO, C. (2020b). Prime genome editing in rice and wheat. *Nature Biotechnology*. https://doi.org/10.1038/s41587-020-0455-x

MAKAROVA, K. S., WOLF, Y. I., ALKHNBASHI, O. S., COSTA, F., SHAH, S. A., SAUNDERS, S. J., BARRANGOU, R., BROUNS, S. J. J., CHARPENTIER, E., HAFT, D. H., HORVATH, P., MOINEAU, S., MOJICA, F. J. M., TERNS, R. M., TERNS, M. P., WHITE, M. F., YAKUNIN, A. F., GARRETT, R. A., VAN DER OOST, J., … KOONIN, E. V. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews Microbiology*, *13*(11), 722–736. https://doi.org/10.1038/nrmicro3569

MITTLER, R., & BLUMWALD, E. (2010). Genetic Engineering for Modern Agriculture: Challenges and Perspectives. *Annual Review of Plant Biology*, *61*(1), 443–462. https://doi.org/10.1146/annurev-arplant-042809-112116

PATRON, N. J., ORZAEZ, D., MARILLONNET, S., WARZECHA, H., MATTHEWMAN, C., YOULES, M., RAITSKIN, O., LEVEAU, A., FARRÉ, G., ROGERS, C., SMITH, A., HIBBERD, J., WEBB, A. A. R., LOCKE, J., SCHORNACK, S., AJIOKA, J., BAULCOMBE, D. C., ZIPFEL, C., KAMOUN, S., … HASELOFF, J. (2015). Standards for plant synthetic biology: a common syntax for exchange of <scp>DNA</scp> parts. *New Phytologist*, *208*(1), 13–19. https://doi.org/10.1111/nph.13532

PIRON, F., NICOLAÏ, M., MINOÏA, S., PIEDNOIR, E., MORETTI, A., SALGUES, A., ZAMIR, D., CARANTA, C., & BENDAHMANE, A. (2010). An induced mutation in tomato eiF4E leads to immunity to two potyviruses. *PLoS ONE*, *5*(6). https://doi.org/10.1371/journal.pone.0011313

PURUGGANAN, M. D., & FULLER, D. Q. (2009). The nature of selection during plant domestication. *Nature*, *457*(7231), 843–848. https://doi.org/10.1038/nature07895

QIN, L., LI, J., WANG, Q., XU, Z., SUN, L., ALARIQI, M., MANGHWAR, H., WANG, G., LI, B., DING, X., RUI, H., HUANG, H., LU, T., LINDSEY, K., DANIELL, H., ZHANG, X., & JIN, S. (2020). High-efficient and precise base editing of C•G to T•A in the allotetraploid cotton (Gossypium hirsutum) genome using a modified CRISPR/Cas9 system. *Plant Biotechnology Journal*, *18*(1), 45–56. https://doi.org/10.1111/pbi.13168

QIN, R., LIAO, S., LI, J., LI, H., LIU, X., YANG, J., & WEI, P. (2019). Increasing fidelity and efficiency by modifying cytidine base-editing systems in rice. *Crop Journal*, 1–7. https://doi.org/10.1016/j.cj.2019.04.007

QIU, W., PARK, J. W., & SCHOLTHOF, H. B. (2002). Tombusvirus P19-mediated suppression of virus-induced gene silencing is controlled by genetic and dosage features that influence pathogenicity. *Molecular Plant-Microbe Interactions*, *15*(3), 269–280. https://doi.org/10.1094/MPMI.2002.15.3.269

RAY, D. K., MUELLER, N. D., WEST, P. C., & FOLEY, J. A. (2013). Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLoS ONE*, *8*(6), e66428. https://doi.org/10.1371/journal.pone.0066428

ROBAGLIA, C., & CARANTA, C. (2006). Translation initiation factors: A weak link in plant RNA virus infection. *Trends in Plant Science*, *11*(1), 40–45. https://doi.org/10.1016/j.tplants.2005.11.004

SANDER, J. D., & JOUNG, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. In *Nature Biotechnology* (Vol. 32, Issue 4, pp. 347–350). Nature Publishing Group. https://doi.org/10.1038/nbt.2842

SARRION-PERDIGONES, A., FALCONI, E. E., ZANDALINAS, S. I., JUÁREZ, P., FERNÁNDEZ-DEL-CARMEN, A., GRANELL, A., & ORZAEZ, D. (2011). GoldenBraid: An iterative cloning system for standardized assembly of reusable genetic modules. *PLoS ONE*, *6*(7). https://doi.org/10.1371/journal.pone.0021622

SCHMIDT, F. J., ZIMMERMANN, M. M., WIEDMANN, D. R., LICHTENAUER, S., GRUNDMANN, L., MUTH, J., TWYMAN, R. M., PRÜFER, D., & NOLL, G. A. (2020). The Major Floral Promoter NtFT5 in Tobacco (Nicotiana tabacum) Is a Promising Target for Crop Improvement. *Frontiers in Plant Science*, *10*(January), 1–16. https://doi.org/10.3389/fpls.2019.01666

TAO, Y., YUAN, F., LEISTER, R. T., AUSUBEL, F. M., & KATAGIRI, F. (2000). Mutational analysis of the Arabidopsis nucleotide binding site-leucine-rich repeat resistance gene RPS2. *Plant Cell*, *12*(12), 2541–2554. https://doi.org/10.1105/tpc.12.12.2541

TRANEL, P. J., & WRIGHT, T. R. (2002). Resistance of weeds to ALS-inhibiting herbicides: what have we learned? *Weed Science*, *50*(6), 700–712. https://doi.org/10.1614/0043-

1745(2002)050[0700:rrowta]2.0.co;2

VAN DER BIEZEN, E. A., & JONES, J. D. G. (1998). Plant disease-resistance proteins and the gene-for-gene concept. *Trends in Biochemical Sciences*, *23*(12), 454–456. https://doi.org/10.1016/S0968-0004(98)01311-5

VAN DER OOST, J., WESTRA, E. R., JACKSON, R. N., & WIEDENHEFT, B. (2014). Unravelling the structural and mechanistic basis of CRISPR-Cas systems. In *Nature Reviews Microbiology* (Vol. 12, Issue 7, pp. 479–492). Nature Publishing Group. https://doi.org/10.1038/nrmicro3279

VAZQUEZ-VILAR, M., BERNABÉ-ORTS, J. M., FERNANDEZ-DEL-CARMEN, A., ZIARSOLO, P., BLANCA, J., GRANELL, A., & ORZAEZ, D. (2016). A modular toolbox for gRNA-Cas9 genome engineering in plants based on the GoldenBraid standard. *Plant Methods*, *12*(1). https://doi.org/10.1186/s13007-016-0101-2

VAZQUEZ-VILAR, M., QUIJANO-RUBIO, A., FERNANDEZ-DEL-CARMEN, A., SARRION-PERDIGONES, A., OCHOA-FERNANDEZ, R., ZIARSOLO, P., BLANCA, J., GRANELL, A., & ORZAEZ, D. (2017). GB3.0: a platform for plant bio-design that connects functional DNA elements with associated biological data. *Nucleic Acids Research*, *45*(4), 2196–2209. https://doi.org/10.1093/nar/gkw1326

WILLIGE, B. C., GHOSH, S., NILL, C., ZOURELIDOU, M., DOHMANN, E. M. N., MAIER, A., & SCHWECHHEIMER, C. (2007). The DELLA domain of GA INSENSITIVE mediates the interaction with the GA INSENSITIVE DWARF1A gibberellin receptor of Arabidopsis. *Plant Cell*, *19*(4), 1209–1220. https://doi.org/10.1105/tpc.107.051441

ZONG, Y., WANG, Y., LI, C., ZHANG, R., CHEN, K., RAN, Y., QIU, J. L., WANG, D., & GAO, C. (2017). Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nature Biotechnology*, *35*(5), 438–440. https://doi.org/10.1038/nbt.3811

7.2. URLs of interest

GoldenBraid 4.0: https://gbcloning.upv.es

N. benthamiana genome site: https://nbenth.com/annotator/index

ThermoFisher Scientific https://www.thermofisher.com/es/es/home.html

Benchling: https://www.benchling.com/

8. Supplementary information

Supplementary Table 1. DNA constructs from the Golden Braid collection used in this work. Samples were taken from the *Golden Braid* glycerol stocks and DNA sequences collection (AMP: ampicillin; CAM: chloramphenicol; KAN: kanamycin; SPE: spectinomycin).

GB NUMBER	NAME	RESISTANCE
GB0030	35S promoter	AMP
GB0037	pTnos terminator	AMP
GB0307	pUPD2	CAM
GB2135	pU6-26 tRNA	KAN
GB1692	Cas9	CAM
GB0019	pDGB3_ Ω1	SPE
GB0021	pDGB3_ Ω	SPE
GB0015	pDGB3_α1	KAN
GB0017	pDGB3_α2	KAN
GB1001	U6-26 promoter	AMP
GB2245	tRNA scaffold vector	AMP

Supplementary Table 2. DNA oligonucleotides used to assembly all the DNA constructs described in this work.

NICKNAME	DESCRIPTION	SEQUENCE (5'-3')
TS19NOV01	NLS-Cas9 H840A PCR FP	CTACGTCTCGCTCGAATGAAACGGACAGCCGACGGAAGCGAGTTCGAGT CACCAAAGAAGAAGCGGAAAGTCATGGACAAGAAGTACTCCATTGGG
TS19NOV02	NLS-Cas9 H840A PCR RP	TGACGTCTCGCTCACGAACCGTCTCCACCGAGCTGAGAGAG
TS19NOV03	Scaffold PCR FP	TCACCGTCTCGCTCGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG
TS19NOV04	Scaffold PCR RP	TGACCGTCTCGCTCAGCACCGACTCGGTGCC
TS19NOV05	U6-26_tRNA PCR FP	TCACCGTCTCGCTCGGGAGCATCTTCATTCTTAAGATATGAAGAT
TS19NOV06	U6-26_tRNA PCR RP	TGACCGTCTCGCTCAGCACCAGCCGGGAATCG
TS20JAN01	NLS_Cas9 sequence check RP1	GTCGACATCGCTGTTGTCTG
TS20JAN02	Scaffold (New OH) PCR RP	TGACCGTCTCGCTCAACCGACTCGGTGCCACTT
TS20JAN24	RT sequence check FP	GAATCCAACATCCTGATTTGATCC
TS20FEB01	sgALS1-c FP	gtgcaTAACTGGTCAAGTGCCACGT
TS20FEB02	sgALS1-c RP	aaacACGTGGCACTTGACCAGTTAt
TS20FEB03	sgALS2-nc FP	gtgcaTCAGTACCGATCATCCTACG
TS20FEB04	sgALS2-nc RP	aaacCGTAGGATGATCGGTACTGAt
TS20FEB05	sgALSMut1-c FP	gtgcaTAACTGGTCAAGTGtccCGT
TS20FEB06	sgALSMut1-c RP	aaacACGggaCACTTGACCAGTTAt
TS20FEB07	sgALSMut2-nc FP	gtgcaATCAGTACCGATCATCCTAC
TS20FEB08	sgALSMut2-nc RP	aaacGTAGGATGATCGGTACTGATt
TS20FEB09	sgALSPreMut-c FP	gtgcaTAACTGGTCAAGTGtCACGT
TS20FEB10	sgALSPreMut-c RP	aaacACGTGaCACTTGACCAGTTAt
TS20FEB11	sgPE_ALS1_nc577 FP	cggtgcATCATCCTACGTGaCACTTGACttttt
TS20FEB12	sgPE_ALS1_nc577 RP	agcgaaaaaGTCAAGTGtCACGTAGGATGATgc
TS20FEB13	sgPE_ALS1_nc577-579 FP	cggtgcCGATCATCCTACGggaCACTTGACttttt
TS20FEB14	sgPE_ALS1_nc577-579 RP	agcgaaaaaGTCAAGTGtccCGTAGGATGATCGgc
TS20FEB15	sgPE_ALS2_577FP	cggtgcAACTGGTCAAGTGtCACGTAGGATttttt
TS20FEB16	sgPE_ALS2_c577 RP	agcgaaaaaATCCTACGTGaCACTTGACCAGTTgc
TS20FEB17	sgPE_ALS2_c577-579 FP	cggtgcAACTGGTCAAGTGtccCGTAGGATttttt
TS20FEB18	sgPE_ALS2_c577.579 RP	agcgaaaaaATCCTACGggaCACTTGACCAGTTgc
TS20FEB19	NLS_Cas9 sequence check FP1	ATAAGGCTGACTTGCGGTTG
TS20FEB20	NLS_Cas9 sequence check FP2	ACTGCCTGAGAAGTACAAGG
TS20FEB21	NLS_Cas9 sequence check RP2	GTAAGCGACTGTAGGAGAATCG
TS20FEB22	NLS_Cas9 sequence check RP3	GTACTTGGTGTTCATGCGTG
TS20FEB23	sgFT5.1_nc FP	gtgcaAAAGAGATGCTAACCAATGG
TS20FEB24	sgFT5.1_nc RP	aaacCCATTGGTTAGCATCTCTTTt
TS20FEB25	sgFT5.2_nc FP	gtgcaTGGAGGTATTCTCTTAGATT
TS20FEB26	sgFT5.2_nc RP	aaacAATCTAAGAGAATACCTCCAt
TS20FEB26	sgFT5.1Mut_nc FP	gtgcaATCTAAGAGAAcACCTCCAT
TS20FEB28	sgFT5.1Mut_nc RP	aaacATGGAGGTgTTCTCTTAGATt
TS20FEB29	sgPE_FT5-c251 FP	cggtgcAAATCTAAGAGAAcACCTCCATTGGttttt
TS20FEB30	sgPE_FT5-c251 RP	agcgaaaaaCCAATGGAGGTgTTCTCTTAGATTTgc

Supplementary Table 3. List of constructs generated in this work. All DNA sequences can be searched by entering the corresponding GB Number at https://gbcloning.upv.es/search/features/. (AMP: ampicillin; CAM: chloramphenicol; KAN: kanamycin; SPE: spectinomycin).

GB NUMBER	NICKNAME	RESISTANCE	TYPE	DESCRIPTION
GB3260	pUPD2_U6-26_tRNA	CAM	Other	Domesticated promoter and tRNA for pegRNAs TU
GB3261	pUPD2_Scaffold	CAM	Other	Domesticated scaffold for pegRNA TU
GB3262	pUPD2_linker-RT	CAM	Other	Domesticated RT for Cas-RT TU
GB3263	pUPD2_NLS-Cas9	CAM	Other	Domesticated Cas9 H840A for Cas-RT TU
GB3345	pUPD2:tRNA_sgALS1-C_Scaffold	CAM	Other	tRNA_gRNA building block
GB3346	pUPD2:tRNA_sgALS2-NC_Scaffold	CAM	Other	tRNA_gRNA building block
GB3347	pUPD2:tRNA_sgALSMut1-C_Scaffold	CAM	Other	tRNA_gRNA building block
GB3348	pUPD2:tRNA_sgALSMut2-NC_Scaffold	CAM	Other	tRNA_gRNA building block
GB3349	pUPD2:tRNA_sgALSPreMut1-C_Scaffold	CAM	Other	tRNA_gRNA building block
GB3350	pUPD2:tRNA_sgFT5.1-NC_Scaffold	CAM	Other	tRNA_gRNA building block
GB3351	pUPD2:tRNA_sgFT5.2-NC_Scaffold	CAM	Other	tRNA_gRNA building block
GB3352	pUPD2:tRNA_sgFT5.1Mut-C_Scaffold	CAM	Other	tRNA_gRNA building block
GB3353	α1:U6-26_tRNA_sgALS1-C_Scaffold	KAN	TU	Standard gRNA TU
GB3354	α1:U6-26_tRNA_sgALS2-NC_Scaffold	KAN	TU	Standard gRNA TU
GB3355	α1:U6-26_tRNA_sgALSMut1-C_Scaffold	KAN	TU	Standard gRNA TU
GB3356	α1:U6-26_tRNA_sgALSMut2-NC_Scaffold	KAN	TU	Standard gRNA TU
GB3357	α1:U6-26_tRNA_sgALSPreMut1-C_Scaffold	KAN	TU	Standard gRNA TU
GB3358	α1:U6-26_tRNA_sgFT5.1-NC_Scaffold	KAN	TU	Standard gRNA TU
GB3359	α1:U6-26_tRNA_sgFT5.2-NC_Scaffold	KAN	TU	Standard gRNA TU
GB3360	α1:U6-26_tRNA_sgFT5.1Mut-C_Scaffold	KAN	TU	Standard gRNA TU

GB3361	α1:sgPE_ALS1_nc577	KAN	TU	pegRNA TU
GB3362	α1:sgPE_ALS1_nc577-579	KAN	TU	pegRNA TU
GB3363	α1:sgPE_ALS2_c577	KAN	TU	pegRNA TU
GB3364	α1:sgPE_ALS2_c577-579	KAN	TU	pegRNA TU
GB3365	α1:sgPE_FT5.1-C251	KAN	TU	pegRNA TU
GB3366	α1:sgPE_FT5.1-C251	KAN	TU	pegRNA TU
GB3367	α2:35S_Cas_RT_Tnos	KAN	TU	Cas-RT TU
GB3368	α1:25S_P19_Tnos	KAN	TU	P19 TU