Τμήμα Βιοχημείας και Βιοτεχνολογίας

Πανεπιστήμιο Θεσσαλίας

Πρόγραμμα Μεταπτυχιακών Σπουδών

‘Προηγμένες Πειραματικές και Υπολογιστικές Βιοεπιστήμες’

Έλεγχος της αποτελεσματικότητας γονιδιωματικής τροποποίησης της τεχνολογίας CRISPR-Cas9 σε γονίδια του Lotus japonicus

Διπλωματική εργασία

Βλέτσος Παναγιώτης

Λάρισα 2023

Department of Biochemistry and Biotechnology

University of Thessaly

Master of Science program

‘Advanced Experimental & Computational Biosciences’

Evaluating the genome editing efficiency of CRISPR-Cas9 technology in *Lotus japonicus* genes.

Master Thesis

Vletsos Panagiotis

Larissa 2023

Three-member committee:

Papadopoulou Kaliope

Professor of Plant Biotechnology

Mathiopoulos Konstantinos

Professor of Molecular Biology

Vasileiadis Sotirios

Assistant Professor of Microbial Molecular-Ecology/ Genomics

Contents

Abstact

The legume *Lotus japonicus* has been explored as a model for studying beneficial symbiotic plant-microbe interactions like nodulation and for its ability to synthesize an abundance of specialized metabolites, such as triterpenes. An intertwined association has been discovered between nodulation and triterpene metabolism during which, genes known for their role in the triterpene metabolism like *AMY2* and *CYP71D353* also seem to play a role in regulating nodulation. Another association was revealed between nodulation and LSK1, a member of a family of GSK3b/SHAGGY-like kinases. In order to study those two associations in detail, a CRISPR/Cas9 system optimized for *L. japonicus* was established. A Golden Gate-based cloning strategy was used to create the CRISPR/Cas9 system, resulting in the assembly of various binary vectors targeting each gene separately. A series of *Agrobacterium rhizogenes* mediated hairy root transformations were used to assess the efficiency of not only the CRISPR/Cas9 system, but of each individual guide RNA. This was achieved through the use of a T7 endonuclease to detect the mutations and to analyze sequencing results of both the type and number of mutations detected into the transformed roots by T7. After confirming the efficiency, the next step is the *Agrobacterium tumefaciens* mediated stable transformation to acquire the desired stable mutant lines, though the CRISPR/Cas9 system and then outcrossing them to remove the transgene. Our end goal is to utilize the CRISPR/Cas9-mediated genome edited seeds for functional analysis to better understand the interconnections between nodulation, triterpene metabolism (*AMY2*) and the GSK3b/SHAGGY-like kinases (*LSK1*).

1. Introduction

1.1 Lotus japonicus

Lotus japonicus is a wild legume native to the regions stretching from the East- to Central-Asia, including Japan, Korea, and China. The plant was first recognized in Kyoto, Japan, a few centuries ago. *L. japonicus* belongs to the family Leguminosae (Fabaceae), the third largest among the Angiosperms, containing a substantial number of crop plants and woody trees with significant agricultural value. Legumes are unique among all plant families due to their ability to participate in symbiotic interactions with nitrogen fixing rhizobial bacteria, allowing them to grow even in nitrogen-poor soils (Wang et al., 2018). Legumes store nitrogen in the nodules at their root structure, and when the plant is harvested, the root biomass breaks down, releasing the stored nitrogen and making it available to future crops. This interaction allows not only the legumes but also subsequent crops to have a renewable source of nitrogen, making legumes a key component in sustainable agricultural systems. In addition, legumes form symbiotic interactions with mycorrhizal fungi allowing for better absorption of phosphorus and other nutrients (Márquez & Stougaard, 2005).

L. japonicus is, alongside *Medicago truncatula*, a model organism for studying the biological processes of legumes and, in particular, their symbiosis with rhizobia and mycorrhiza*. L. japonicus* is a diploid and self-fertile plant with a small genome (~470Mbp), consisting of six chromosomes, high transformability, and a short life cycle (2-3 months), making it a suitable model organism. Apart from that*, L. japonicus* is also amenable to transformation by either *Agrobacterium tumefaciens* or *Agrobacterium rhizogenes*. The ecotypes usually used in research are Gifu and MG20 (Mijakogima), which were first collected from the Gifu prefecture and the Mijako island in Okinawa prefecture, respectively (Küster, 2013; Márquez & Stougaard, 2005).

1.1.1 Triterpenes

Triterpenes are one of the most numerous and diverse groups of natural products existing in plants. They are derived from a precursor molecule known as squalene, which undergoes various structural modifications to form different triterpene compounds (Figure 1). Triterpenes can exist in simple unmodified forms, but they often accumulate as conjugates with other macromolecules like carbohydrates, most notably as triterpene glycosides. Triterpene exhibit diverse biological activities, with some of them possessing medicinal properties, like betulinic acid or lupeol who possess anti-inflammatory and antioxidant activity. Despite their great diversity, the majority of triterpenes found in nature are cyclic triterpenes with 30 C-atoms and 1–5 ring systems. Hence, both simple and conjugated triterpenes seem to have a wide range of applications in many biotechnological sectors.

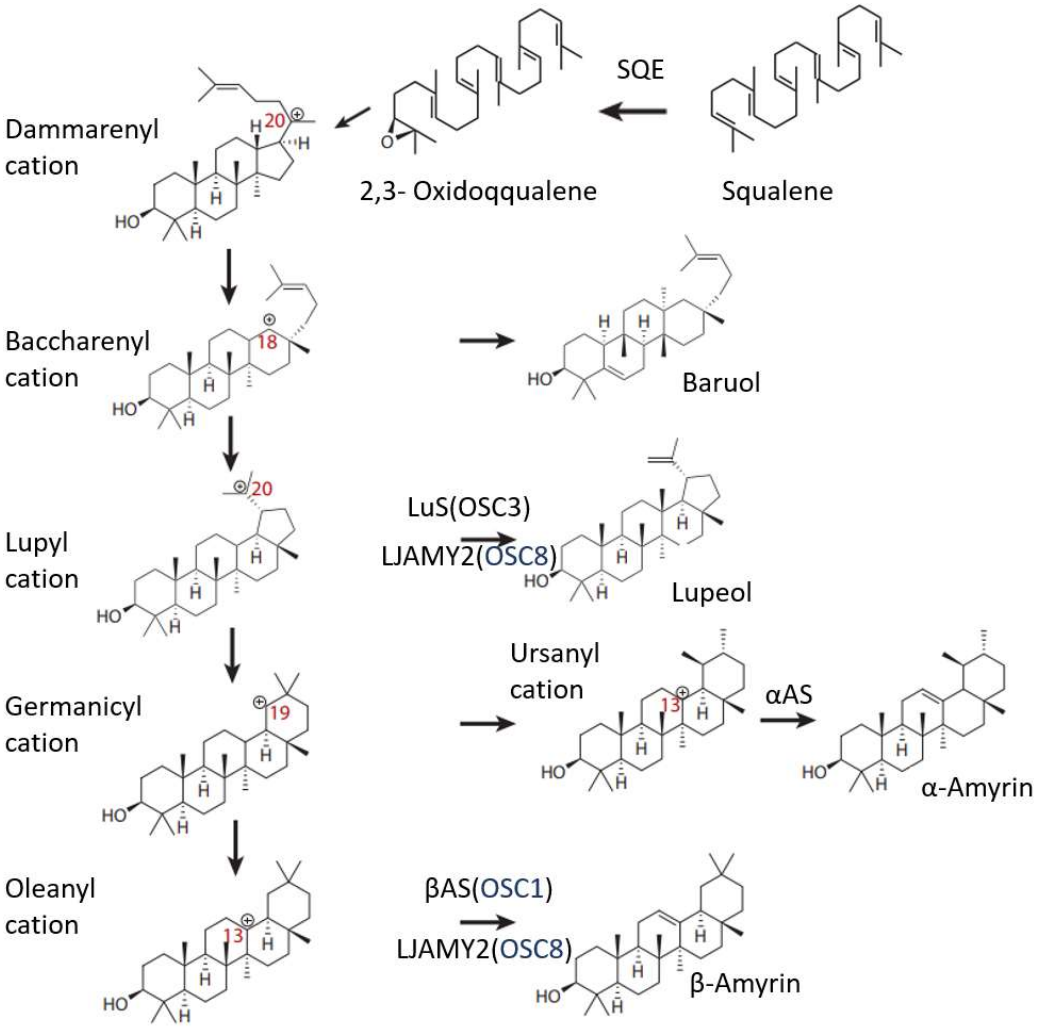

Figure 1. Cyclization of 2,3-oxidosqualene and the pathway for the biosynthesis of triterpenes such as lupeol, baruol, α-amyrin and β-amyrin. OSC3/LuS: Lupeol synthase, OSC1/AMY1/βAS: β-amyrin synthase, AMY2/OSC8: capable of synthesizing both lupeol and β amyrin

1.1.2 Symbiotic nitrogen fixation- *Mesorhizobium loti* Nodulation

The legume–rhizobia symbiosis occurs in 88% of all legume species (De Faria et al., 1989). This interaction is beneficial for both parties, as plants take up nitrogen in the form of either nitrate ($NO3^-$) or ammonium ($NH4^+$) and bacteria receive a carbon source from the plants' carbohydrates. The symbiotic nitrogen fixation offers an ecological alternative to the use of synthetic fertilizers in agriculture. The host plant initiates the symbiotic relationship by exuding flavonoid molecules from its roots that are then recognized by their compatible rhizobia partners. The bacteria are attracted to the roots by those flavonoids and are stimulated by them to produce Nod factor molecules, the main signaling molecules involved in the initiation of symbiosis between rhizobia and legume plants. Those factors trigger a complex signal cascade in the host plant, leading to the organogenesis of novel root organs called nodules. Nodules provide an ideal environment for biological nitrogen fixation (Ferguson et al., 2019). The model plant *L. japonicus* can form symbiotic relationships with the nitrogen fixing bacterium *Mesorhizobium loti.* The result of the symbiosis is the formation of nodules within 10 days post infection (Márquez & Stougaard, 2005).

1.1.3 Genes of interest

In *L. japoninus* there is a metabolic gene cluster created by three genes OSC8/LjAMY2, *CYP88D5* and *CYP71D353* (Figure 2). This gene cluster creates a new triterpene biosynthetic pathway with all three genes being tightly co-regulated during root and nodule development, indicating a possible role in both processes (Krokida et al., 2013). The synthesis of the triterpene lupeol in *L. japonicus* roots and nodules can be solely attributed to OSC3/LuS gene and might also affect nodule formation by regulating the ENOD40 gene expression (Delis et al., 2011). The cyclization of 2,3-oxidosqualene is catalyzed by enzymes known as OxidoSqualene Cyclases (OSCs), making it the deciding step between sterol and triterpene biosynthesis (Phillips et al., 2006; Sawai et al., 2014; Thimmappa et al., 2014). Eight OSC genes (OSC1–OSC8) have been identified in *L. japonicus*. *OSC1* and *OSC8* correspond to the previously reported *LjAMY1* and *LjAMY2* (Iturbe-Ormaetxe et al., 2003), with *OSC1* being a β-amyrin synthase while OSC8 is capable of synthesizing both lupeol and β amyrin (Figure 1). On the other hand, *OSC3* and *OSC5* were identified as lupeol synthase and cycloartenol synthase, respectively (Sawai et al., 2006). Although simple triterpenes, such as β-amyrin and lupeol are quite common in plants, the triterpene scaffolds are often further modified to more elaborate molecules by tailoring enzymes, with the Cytochrome (CYP) P450s–mediated oxygenation of the triterpene scaffold being a common modification. CYP71D353 for example catalyzes the formation of 20-hydroxybetulinic acid in a sequential three-step oxidation of 20-hydroxylupeol, a derivative from β-amyrin (Krokida et al., 2013).

Apart from the triterpene metabolism, it also seems that kinase enzymes are also involved in symbioses. The family of GSK3b/SHAGGY-like kinases (SKs) consists of signaling proteins that participate in metabolic regulation, hormone regulation and mediating biotic and abiotic stress responses in plants. In *L. japonicus, Lotus SHAGGY-like kinase 1* (*LSK1*) modulates the expression of the autoregulation of nodulation (AON) components, which is activated after rhizobium infection and suppresses nodulation in response to sufficient nitrate levels. Consequently, *LSK1* seems to negatively regulate nodulation (Garagounis et al., 2019).

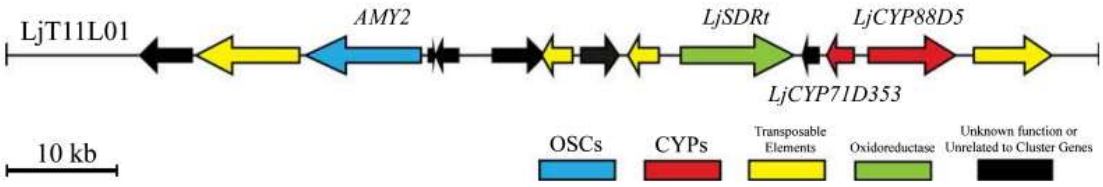

Figure 2. Visualization of genomic region on chromosome 3 *of Lotus japonicus* encompassing OSC8/AMY2 flanked by genes involved in triterpene metabolism in legumes such CYP.

1.2 CRISPR/Cas

1.2.1 Genome editing

Genome editing is a method used in reverse genetics that allows for site-specific modifications of genomic DNA through the use of endogenous DNA repair mechanisms (Khalil, 2020). This method is based on the processing and repair of double strand breaks (DSB). The DSBs can be repaired by either homologous recombination (HR) or non-homologous end joining (NHEJ) depending on the presence or absence of a donor template, respectively. The NHEJ pathway repairs the break by rejoining both ends, causing few base insertions or deletions, resulting in an indel or frameshift that causes gene disruption. The HR

pathway uses a donor template, which is then integrated into the target site, thus modifying the site of interest (Khalil, 2020; Wendy & Pederson, 2017).

During the last decade, the field of genome editing has been fast advancing, leading to the development of many genetic modification technologies and methods. A basic method is to induce the formation of DSB's through the use of endonucleases. There are four major technologies that make use of endonucleases: meganucleases (MegNs), zinc finger nucleases (ZFNs), transcription activator-like effector nuclease (TALENs), and clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein (Cas) (CRISPR/Cas). The ZFN and TALEN technologies recognize DNA sites hinged on the interactions between DNA and artificial proteins making them difficult to use, while the CRISPR/Cas system is based on DNA-RNA interactions. This fact, among other advantages such as easy prediction of off-target sites, multiplex targeting, and ease of designing targets, makes the CRISPR/Cas system much easier to use (Khalil, 2020).

1.2.2 Origin of CRISPR/Cas system

CRISPR-Cas are acquired immune systems used by bacteria and archaea to defend against invasion by foreign genetic elements such viruses. The system integrates fragments of the foreign DNA, mostly of mobile genetic elements (MGE), in unique spacer sequences that are inserted into CRISPR arrays embedded in the microbial genomes (Lee et al., 2015). Upon an encounter with similar MGEs, the transcripts generated by those spacer sequences are able to recognize the MGE sequence and used to guide the Cas nucleases to the target site, which results in the inactivation of the latter. A complete CRISPR/Cas system consists of a CRISPR array, partially palindromic repeated sequences separated by the spacers, and the adjoining cluster of Cas genes, which encodes the adaption and effector modules. The adaptation module consists of the Cas1 and Cas2 proteins, among others, forming a complex responsible for integrating the foreign DNA into the array. The effector module acts as a sequence-specific nuclease and is far more diverse compared to the adaption module among the different CRISPR/Cas systems. E.g., all CRISPR-Cas systems are divided into two classes depending on the effector module and each class can be further divided into three types each (Figure 3). Class 1 systems have multi-subunit effector complexes comprising several Cas proteins, whereas Class 2 systems possess a single multidomain effector protein (Koonin & Makarova, 2019; Lee et al., 2015).

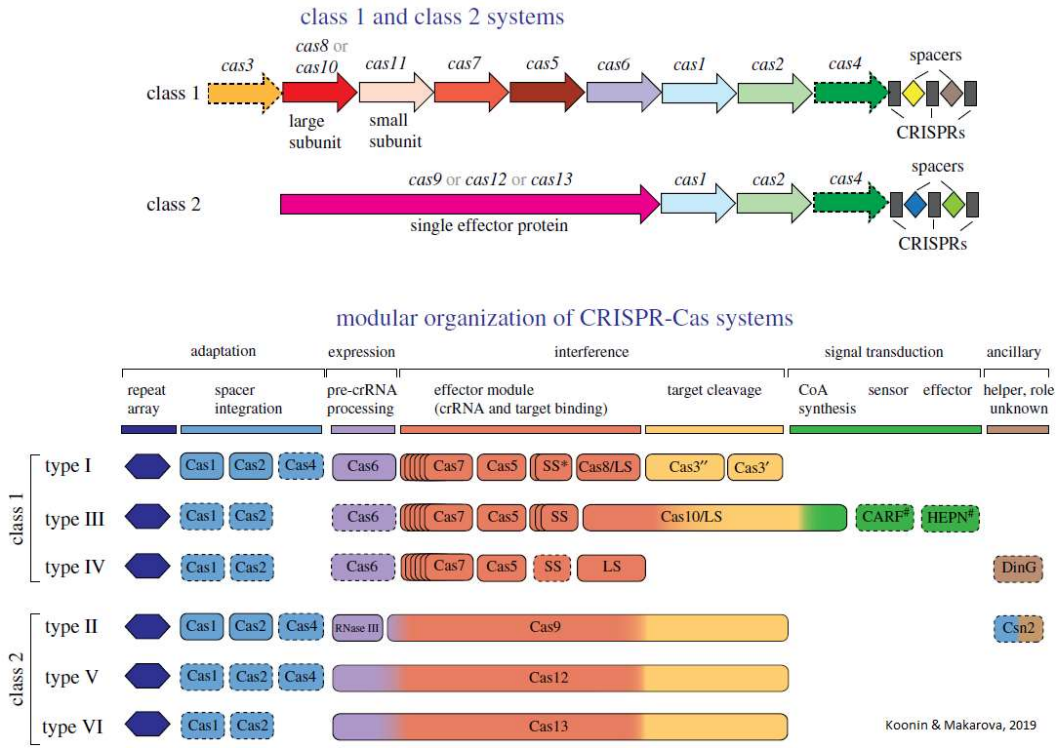

Figure 3 Class 1 and class 2 CRISPR-Cas systems and types. The common feature among all types are the adaption proteins cas1 and cas2 with the classification depending on the effector module: multiprotein effectors(Class 1) and the single protein effectors (Class 2) (Koonin & Makarova, 2019)

The acquired immune defense of the CRISPR/Cas system can be separated into three separate yet intertwined stages:

1. Adaption, in which Cas proteins binds into PAM (Protospacer-Adjacent Motif) motif though theirs DNA binding domains and introduce DSBs into the targeted DNA molecule in the region upstream or downstream of the PAM motif. The released segment (protospacer) is then inserted into the CRISPR array between two repeats, becoming a new spacer sequence.
2. Expression and processing of the pre-crRNA (pre-CRISPR RNA), in which the whole CRISPR array is first transcribed into a long transcript, the pre-crRNA, and then processed into the mature crRNA by either Cas or other proteins. The mature crRNA contains a CRISPR array repeat-derived region, which helps stabilize the structure of the molecule and be involved in the interactions with the cas proteins, and a spacer-derived sequence responsible for guiding the cas protein into the target DNA.
3. Interference, where the crRNA acts as a guide RNA (gRNA) for the Cas proteins to recognize the protospacer in the MGEs, which are then cleaved and inactivated by the Cas proteins.

One of the best characterized systems is the CRISPR/Cas9, otherwise known as Class 2 type II system. Cas9 protein forms a ribonucleoprotein (RNP) with a trans-activating CRISPR RNA (tracrRNA), that acts as a scaffold for the binding of Cas9 and the crRNA, while the crRNA is complementary to the target DNA and is the one that guides Cas9 to cleave the target DNA. Furthermore, the PAM sequence that exists in the target DNA, is

required for Cas9 protein's binding to the target. The Cas9 protein possesses a RuvC-like and a HNH nuclease family domain that cleave the anti-sense and sense strand, respectively, thus creating a double stranded break (DSB), (Wan et al., 2021; Xu & Li, 2020).

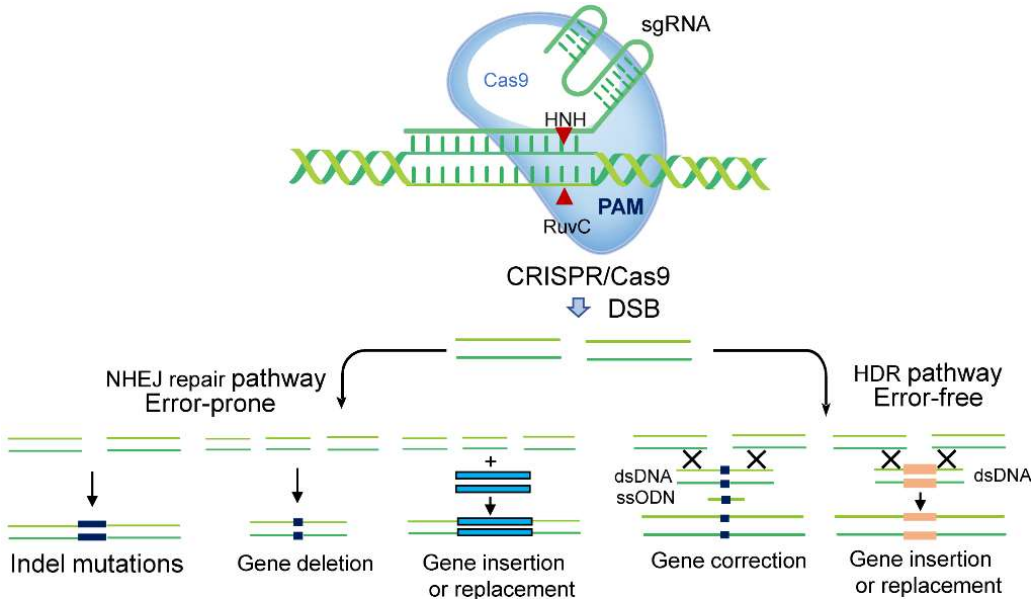

Figure 4 The applications of CRISPR-Cas9 system in genome editing The genome modifications depending on the double-strand break (DSB) repair pathways: the nonhomologous end joining (NHEJ) and the homology directed repair (HDR) pathway (Wan et al., 2021).

1.3 Golden Gate cloning

There are several cloning technologies that rely on the use of DNA restriction enzymes to cut a vector and a DNA fragment in a site-specific manner, so that they may be easily joined together by a DNA ligase. Many of these technologies share the same drawback, that despite being efficient to use, they leave behind the cleavage site sequence, which might also affect the expressed protein. To combat this drawback, several technologies have been developed, with Golden Gate cloning being one of them. Golden Gate is based on the use of type IIS restriction enzymes, which can cut outside of their recognition sequence and hence create non-palindromic overhangs. This also means that those enzymes can generate a considerable number of different overhangs, for example BsaI can generate 256 different four-base pair overhangs. The end result is that, by carefully designing the overhangs, it is possible to ligate multiple fragments in a single reaction without adding any unwanted sequences (Figure 5). Due to this fact the process has been described as 'seamless' and an 'one-pot, one-step cloning' (Engler et al., 2008, 2009).

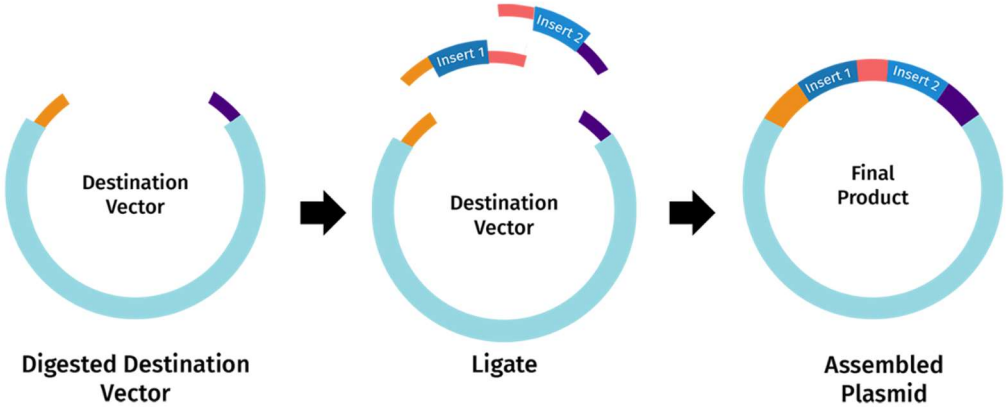

Figure 5 Visualization of the Golden gate cloning technology's capability of multiple insert cloning

1.3.1 Modular Cloning (Moclo) System

A modular cloning system (Moclo) was created based on the Golden Gate cloning strategy. The Moclo system used in this work is comprised of two kits: the Golden Gate MoClo Plant Tool Kit and the Golden Gate MoClo Plant Parts Kit. All the plasmids that are part of the two kits, also possess genes that confer resistance to antibiotics (kanamycin, carbenicillin, spectinomycin) and reporter genes (lacZa, cRED). The Golden Gate MoClo Plant Tool Kit contains vector backbones to clone new standard parts (e.g., promoters, coding sequences, terminators) (level 0), assemble transcriptional units (level 1) or multigene vector/constructs (level 2, M, P), as shown in Picture 4. The Golden Gate MoClo Plant Parts Kit contains standardized parts (modules), such as reporter genes, promoters and terminators, which enable Golden Gate construction (Engler et al., 2014; Werner et al., 2012).

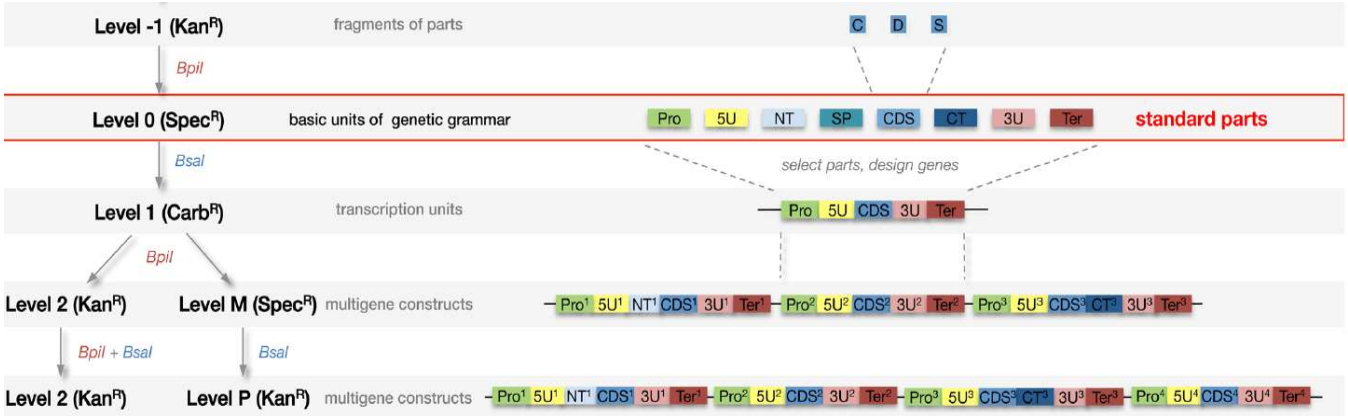

Figure 6 Golden Gate Moclo Assembly Standard. Standard parts (level 0) are assembled into level 1 transcriptional units with type IIS BsaI restriction enzyme. Level 1 transcriptional units are flanked by two type IIS BbsI recognition sites for downstream assembly into level 2 multigene vectors (Engler et al., 2014).

1.4 *Agrobacterium rhizogenes*-mediated hairy root transformation

Hairy root transformation is a genetic engineering technique used to introduce transgenes into plants through the use of *Agrobacterium rhizogenes*. *A. rhizogenes* is a soil bacterium that contains the Ri (Root inducing) plasmid. The Ri plasmid carries GALLS genes that can transfer a portion of its DNA, known as transfer DNA (T-DNA).

After an agrobacterium infection, T-DNA from the Ri plasmid is transferred to the host cell and integrated into the plant genome. The expression of genes located in the T-DNA causes the induction of “hairy roots”. By using T-DNA binary system consisting of a T-DNA vector binary plasmid that contains the transgene and a Ri helper plasmid, that contains the genes required for transfer of the T-DNA into the plant. This system is similar to *A. tumefaciens* transformation system, but it requires shorter transformation periods while maintaining high transformation efficiency. As a result, this transformation system is particularly suited for applications where rapid and efficient generation of transgenic is required, including the production of different compounds, as well as for studying plant-microbe interactions or plant gene function. Recently it has being used as a tool to study gene function through CRISPR/Cas9- mediated gene editing, exploiting the shorter regeneration process to verify the gene-editing efficiency of various CRISPR/Cas9 vectors (Jedličková et al., 2022; Kiryushkin et al., 2022; Nguyen et al., 2022).

Aim of the thesis

In previous works it was revealed that the symbiotic interactions formed between the model organism *L. japonicus* and beneficial microorganisms, like the rhizobium *M. loti*, are regulated by various mechanisms in the plant. On one hand, triterpene biosynthesis pathway encoded by AMY2 (OSC8) seems to regulate nodulation; on the other hand, Shaggy-like kinase 1 (LSK1) regulates nodulation. During the present thesis the aim is to understand how these two different mechanisms regulate nodulation and whether there is a connection between symbiosis and triterpene production. To that end, a series of CRISPR/Cas9-mediated stable silenced plant lines were generated for the different genes that take part in these mechanisms. A Golden Gate-based cloning strategy was adopted to build the CRISPR/Cas9 genome editing system. With the genome editing system ready at hand, the mutation efficiency of the different constructs via *A. rhizogenes*-mediated hairy root transformation was evaluated.

2. Materials and Methods

2.1 Golden Gate-based plasmid construct assembly

In order to create all the necessary constructs, we used the Moclo Plant Parts kit and the Moclo Plant Tool kit that are based on the Golden Gate cloning scheme (Engler et al., 2014b). The Moclo Plant Tool kit has a collection of vectors and each one of them produces different sticky ends when cleaved by the IIS enzymes (BsaI, BbsI). Depending on the sequence of those sticky ends, the contents of the level 1 vectors can be inserted into the positions 1, 2, 3, 4 or 5 in the level 2 vector in either forward or reverse orientation. This is a crucial step because the position and orientation of each transcriptional unit in the l2 construct affects its function. For example, if two transcriptional units were to have the same orientation, it could cause transcriptional interference between the two promoters. For the cloning reactions we employed a 2:1 molar ratio for quantity(ng) of insert to vector. To calculate the quantity of the insert we used the following formula**:**

insert quantity(ng)= 2*vector quantity(ng)*insert size(kb)/ vector size(kb)

Many of the constructs and vectors used for this thesis were already constructed beforehand or were part of the Moclo Plant Tool kit. In general, the assembly of a new construct required the following steps: a digestion-ligation cloning reaction to create the construct, a transformation protocol to insert the plasmid into competent cells, the selection of the colonies that received the plasmid, the isolation of the amplified construct and the verification of its contents. The digestion-ligation reaction made use of the IIS restriction enzymes BsaI/BbsI and the T4 ligase to assemble the new constructs. For the construction of the level 1 transcriptional units a short digestion-ligation protocol (40 cycles of digestion-ligation) had been used, while a longer one (70 cycles of digestion-ligation) was employed for the level 2 constructs(Table 1).

Cycling Conditions			
37°C	5 min	Repeat 69 times	BsaI/BbsI digestion
16°C	5 min	Repeat 69 times	T4 Ligation
80°C	20 min	Once	Enzyme inactivation
16°C	-	Wait	-

Table 1. Cycling conditions employed for Golden Gate cloning. Depending on the assembly of either level or level 2 constructs the number of repeats was 39 and 69 respectively.

2.1.1 Heatshock Transformation

The cloning reaction was used to transform *E. coli* DH5α chemi-competent cells via a heat-shock protocol. The *E. coli* DH5α chemi-competent cells stored at -80°C were transferred on ice. In this protocol we added 5 μl for l0 and l1 constructs or 10 μl for l2 constructs of the cloning reaction to the competent cells. The cells were left on ice for 20 min to incubate, then moved to 42°C for 55sec as a heatshock and back on ice for 5 min to recover. 1 ml of LB medium was added to the competent cells, which were left to grow at 37°C with either shaking at 160 rpm or without shaking for 45 min (l0 and l1) and 90 min (l2). Following, the cells were precipitated by centrifugation at 10,000g for 30 sec. The supernatant was partially discarded, and the precipitate was redissolved in the remaining liquid. The cells were then transferred onto plates containing an antibiotic selecting for the transformed bacteria due to the occurrence of the corresponding resistance gene in the vector backbone. They were left to grow at 37°C overnight.

2.1.2 Colony Selection

The antibiotic resistance selected for bacterial cells that have received a vector, but it could not distinguish if the vector/constructs carried the DNA inserts. So, in order to select the colonies with the recombinant vector, further screening was required. The blue-white screening was used for the l1 constructs, by adding 50μl of X-gal (40mg/ml) and 10μl of IPTG (0,5M) in the plates, while the replacement of an artificial operon Cred allowed for a red-white screening without the addition of additional compounds in the l2 constructs.

2.1.3 Plasmid Isolation and identification

A sample of the selected colonies was added into LB liquid cultures with the corresponding antibiotic and were incubated at 37°C and 210 rpm overnight. The Macherey-Nagel plasmid purification kit was used to extract the construct carrying plasmids from the bacterial colonies following the manufacturer's instructions. After the plasmid purification, the concentration of the plasmid DNA was calculated using a nanodrop spectrophotometer. The verification process required a diagnostic digestion with restriction enzymes, usually XmnI, EcoRI and HindIII.

For the digestion, two reactions (Table 2) were prepared for every sample with and without the restriction enzyme that resulted in a cut and an uncut plasmid. When the reactions were prepared, they were incubated at 37^0C for 90min.

Component	Volume
Plasmid DNA	1μl
Cutsmart (NEB) Buffer	1μl
Enzyme (NEB)	0,2-0,5μl
ddH_2O	Bring volume to 10μl

Table 2. Components for a generic diagnostic digestion for plasmids.

The identification of the constructs was achieved through a gel electrophoresis using an 1% agarose gel and in a later stage with Sanger sequencing.

2.2 Level 0 module collection

Some of the level 0 modules were already part of the Moclo Plant Parts kit, while others, like the LjUB500i promoter, the NOS terminator, the hCas9, AtCas9, the SV40 NLS and Bar CDS1 module, were previously constructed.

2.3 Level 1 transcriptional units

Most of the level 1 transcriptional units had been constructed beforehand for other experiments.

2.3.1 Level 1 reporter cassette

Level 0 modules *Lotus japonicus* Ubiquitin 9LjUb500i) promoter (Pro+5U), turbo Red Fluorescent Protein (tRFP) coding sequence and terminator of actin 2 gene were subcloned into level 1 acceptor(1F) pICH41332 of 1st position and forward orientation through a BsaI/T4 digestion-ligase reaction. The components and the steps of the reaction are shown in (Table 1 &Table 3).

Level 1 reporter cassette 1F	BsaI/T4 digestion-ligase
pICH47732 1F vector (4968)	50 ng
LjUb500i Pro+5U (1127bp)	22 ng

TurboRFP (696bp)	14 ng
Act2 Ter (485bp)	10 ng
BsaI (NEB) 20U/µl	0,5µl
T4 DNA ligase 400U/µl	0,25µl
Cutsmart 10x	1µl
T4 ligase buffer 10x	1µl
ddH_2O	Till 10µl

Table 3. Level 1 reporter cassette digestion-ligation reaction mixture

2.3.2 gRNA cassettes

gRNAs targeting the genomic and coding sequences of *AMY2*, *CYP71D353*, *LSK1* and *OSC3* were already designed. The key consideration when designing the gRNAs were: the selection of unique sequences that had the PAM motif, an analysis to minimize the risk of off-target effects and the GC content. A more detailed description of on the selection criteria for the gRNAs was done in previous works. Each gRNA was synthesized by annealing two oligonucleotides (Table 5), which resulted in the creation of a dsDNA sequence with 5' and 3' overhangs.

The hybridization reactions are shown in Table 4. The steps required for the process were an incubation at 95◦C for 5', followed by another incubation at RT for 20'. This process was repeated three times and then the samples were placed on ice.

gRNA annealing	Volume(µl)
Forward primer 10pmol/µl	9
Reverse primer 10pmol/µl	9
T4 ligase Buffer 10x	2

Table 4. Components for the gRNA annealing hybridization reaction .

Primer name	Sequence 5'→3'
AMY2-gRNA1 _F	TTC G TACATATGCATGCGAATACT
AMY2-gRNA1_R	AAC AGTATTCGCATGCATATGTA C
AMY2-gRNA3 _F	TTC G CGACACACATAGCTTCTTGG
AMY2-gRNA3 _R	AAC CCAAGAAGCTATGTGTGTCG C
AMY2-gRNA4 _F	TTC **G** CACTGCATAGAAAAGCCATA
AMY2-gRNA4_R	AACTATGGCTTTTCTATGCAGTG **C**
AMY2-gRNA5 _F	TTC **G**TGACATCTCGAGTTTCACAT
AMY2-gRNA5_R	AACATGTGAAACTCGAGATGTCA **C**
OSC3-gRNA1_ F	TTCGTTCAAGCCCATGATGGCCAC
OSC3-gRNA1_ R	AACGTGGCCATCATGGGCTTGAAC
OSC3-gRNA2_ F	TTCGTATTAATCATGTACGTTACG
OSC3-gRNA2_ R	AACCGTAACGTACATGATTAATAC
CYP71D353-gRNA1_F	TTCGTCTTGGATAACTCTCTCAGG
CYP71D353-gRNA1_R	AACCCTGAGAGAGTTATCCAAGAC
CYP71D353-gRNA2_F	TTCGTAATCACAAGTAGCCAGCGT
CYP71D353-gRNA2_R	AACACGCTGGCTACTTGTGATTAC
LSK1_gRNA1_F	TTCGTCGAGACGGACAGCCGAAGC
LSK1_gRNA1_R	AACGCTTCGGCTGTCCGTCTCGAC
LSK1_gRNA2_F	TTCGTGGTTGGGACTGGATCCTT
LSK1_gRNA2_R	AACAAGGATCCAGTCCCAACCAC
LSK1_gRNA3_F	TTCGGTCAACTGTGCGCATAACC
LSK1_gRNA3_R	AACGGTTATGCGCACAGTTGACC

LSK1_gRNA4_F	TTCGCTCAGATTAAAGCGCACCCG
LSK1_gRNA4_R	AACCGGGTGCGCTTTAATCTGAGC

Table 5. gRNA oligonucleotides: Nucleotides colored with orange represent 3-nt 5' overhangs that are complimentary to gRNA acceptor fusion sites after digestion with SapI. The size of the gRNA sequence was 20-nt, with 3-nt NGG added into 5' as the PAM sequence recognized by Cas9 endonuclease of Streptococcus pyogenes For each gRNA sequence that does not start with guanine, it was added at the beginning, after the 5' overhang to enhance stability during RNA-DNA or RNA-RNA interactions. Hence some oligonucleotides(colored with light blue) have 24 bp instead of 23bp.

Each ds gRNA with overhangs was subcloned into the appropriate gRNA acceptor. gRNAs were mixed with gRNA acceptor 3F or gRNA acceptor 4F in a SapI/T4 ligase reaction protocol (Table 1,Table 6).

gRNA cassette		**SapI/T4 digestion-ligation**
gRNA acceptor 3F	**gRNA acceptor 4F**	50ng
ds gRNA (23-24 bp)		1μl
SapI 10U/μl		1μl
T4 ligase 400u/μl		0,25μl
Cutsmart 10x		1μl
T4 ligase buffer 10x		1μl
ddH_2O		Till 10μl

Table 6. gRNA cassette digestion-ligation reaction mixture

5μl of digestion-ligation reaction were transformed into DH5a chemo-competent cells, incubated in plates and liquid cultures with carbenicillin 100μg/ml. Plasmid DNA was isolated using the Macherey Nagel NucleoSpin Plasmid- Miniprep kit and purified plasmids were sent for Sanger sequencing.

gRNA	**gRNA acceptor**
AMY2-gRNA1 F/R	3F acceptor
AMY2-gRNA3 F/R	4F acceptor
AMY2-gRNA4 F/R	3F acceptor
AMY2-gRNA5 F/R	4F acceptor
LSK1-gRNA1 F/R	3F acceptor
LSK1-gRNA2 F/R	3F acceptor
LSK1-gRNA3 F/R	4F acceptor
LSK1-gRNA4 F/R	4F acceptor
CYP71D353-gRNA1 F/R	3F acceptor
CYP71D353-gRNA2 F/R	4F acceptor
OSC3-gRNA1 F/R	3F acceptor
OSC3-gRNA2 F/R	4F acceptor

Table 7. List of gRNA cassettes. Those colored light blue were already prepared.

2.4 Level 2 constructs

The level 1 transcriptional units for the reporter gene, plant selection marker gene, guide RNAs(gRNAs) and Cas9 cassettes were assembled into the level 2 vector pAGM4723 in a digestion-ligation reaction using BbsI/T4 ligase (Table 8). The appropriate end-linker plasmids were used to connect the selection marker to the right overhang of the vector. A negative control was also assembled by switching the gRNA cassette with the corresponding gRNA acceptor (Table 9). In order to optimize the cloning protocol, a modified version was used in

which, the amount of plasmid DNA for both the vector and inserts was increased, while the addition of Bovine Serum Albumin (BSA) helped to stabilize the enzymes in the reaction (Table 9). The steps of the reaction are shown in Table 1.

Level 2 constructs with two gRNA cassettes	BbsI/T4 digestion ligation	Level 2 constructs with one gRNA cassette	BbsI/T4 digestion ligation
Vector_pAGM4723 (12919bp)	50 ng	Vector_pAGM4723 (12919bp)	50 ng
1F_LjUb_tRFP_act2 (2351bp)	50 ng	1F_LjUb_tRFP_act2 (2351bp)	50 ng
2R_2x35sΩ_atCas9_nos (5440)	120 ng	2R_2x35sΩ_atCas9_nos (5440)	120 ng
3F_gRNA_cassete (987)	20 ng	3F_gRNA_cassete (987)	20 ng
4F_gRNA_cassete (987)	20 ng	-	-
5R_Bar_selection_marker (1908)	40 ng	4R_Bar_selection_marker (1908)	40 ng
End-linker five	1 ng	End-linker four	1 ng
BbsI (NEB) 10U/µl	0,5µl	BbsI (NEB) 10U/µl	0,5µl
T4 DNA ligase 400U/µl	0,25µl	T4 DNA ligase 400U/µl	0,25µl
Cutsmart 10X	1µl	Cutsmart 10X	1µl
T4 ligase buffer 10X	1µl	T4 ligase buffer 10x	1µl
ddH_2O	Till 10µl	ddH_2O	Till 10µl

Table 8. Level 2 multigene construct digestion-ligation reaction mixtures

Level 2 construct Cas9 only (negative control)	BbsI/T4 digestion ligation original	BbsI/T4 digestion ligation modified.
Vector_pAGM4723 (12919bp)	50 ng	100 ng
1F_LjUb_tRFP_act2 (2351bp)	50 ng	100 ng
2R_2x35sΩ_atCas9_nos (5440)	120 ng	240 ng
3F_gRNA_acceptor (961)	20 ng	40 ng
4F_gRNA_acceptor (961)	20 ng	40 ng
5R_Bar_selection_marker (1908)	40 ng	80 ng
End-linker 5	1 ng	1 ng
BbsI (NEB) 10U/µl	0,5µl	0,5µl
T4 DNA ligase 400U/µl	0,2µl	0,5µl
Cutsmart 10x	1µl	1,5µl
T4 ligase buffer 10x	1µl	1,5µl
BSA	-	1,5µl
ddH_2O	Till 10µl	Till 15µl

Table 9. Level 2 multigene construct modified and original digestion-ligation reaction for Cas9 only construct.

10µl of each reaction were used to transform DH5a chemo-competent cells, transformed cells were placed onto LB plates containing kanamycin 50µg/ml. After an overnight incubation, white colonies were picked and transferred into liquid LB cultures with kanamycin (50µg/ml) and incubated overnight. After the incubation the process was the same as with the level 1 transcriptional unit constructs, with the only difference being the use of the XmnI restriction enzyme in the diagnostic digestion as shown in (Table 2).

Target gene region	L2 CRISPR/Cas9 constructs
AMY2: intron 1 and exon 4	AMY2:gRNA1-gRNA3
AMY2 : intron 9 and intron 10	AMY2:gRNA4-gRNA5
LSK1: exon 3 and exon 5	LSK1:gRNA1-gRNA3
LSK1: exon 4 and exon 10	LSK1:gRNA2-gRNA4
CYP71D353: exon1	CYP71D353:gRNA1-gRNA2
OSC3:	OSC371D353:gRNA1-gRNA2
-	Cas9-only (negative control)

Table 10: List of assembled level 2 binary vectors(constructs). All constructs possess the transcriptional units for the reporter gene (tRFP), selection marker gene (BASTA) and the Cas9 cassette (2x35S_Ω_AtCas9_tnos). The names were based on the gRNA cassettes of each construct.

2.5 *Agrobacterium rhizogenes*-mediated hairy root transformation

2.5.1 *Lotus japonicus* seed scarification and germination

Seeds of the *Lotus japonicus* cultivars MG20 and Gifu were used in this experiment. Seeds were submerged in sulfuric acid H_2SO_4 at room temperature while being gently shaken. After the seeds started to show signs of discoloration, the acid was carefully removed and then cold sterilized dH_2O was used to wash the seeds in five repeated steps. After completing the washing, seeds were sterilized using a solution containing 20% bleach (NaOCl) and 0,1% tween20 (Polysorbate 20) while shaking in room temperature for 12 to 15 minutes. At this point, the seeds started getting swollen and turning white, showing that the seed coat ruptured, thus promoting germination. After that, the seeds were washed 8 times with cold sterilized ddH_2O to remove any traces of bleach. Finally, the seeds were submerged in ddH_2O, covered with tinfoil, and kept at 4°C for 48hrs to allow imbibement. After two days the seeds were transferred to 1/2MS (Murashige-Skoog) plates and covered with tin foil. The plates were placed vertically at 22 °C. The tinfoil was removed three days later, and the percentage of germination was evaluated. The germinated seeds were then maintained vertically at 22 °C with a 16h light/ 8h dark photo period for 4 days.

2.5.*2 Agrobacterium rhizogenes* transformation and culture preparation

The level 2 constructs were transformed into the electo-competent *Agrobacterium rhizogenes* strain LBA1334 via electroporation. 1µl of each construct (10ng/µl) was added to 50µl of *A. rhizogenes* electro-competent cells. After gently mixing, the mixture was transferred into a sterilized cold electroporation cuvette. The cuvette was then inserted into the chamber of the micropulser electroporetor (Biorad). The appropriate conditions were chosen for the competent cells and then an electric pulse was delivered. 1ml of LB was then added into the cuvette and the cells were gently resuspended. The mixture was then transferred into an Eppendorf tube and incubated for 2hrs at 28°C. After the recovery period the cells were centrifuged at 10.000 g for 30sec, with the excess supernatant being discarded and cells resuspended in the remaining supernatant. The mixture was then streaked onto LB plates containing kanamycin 50µg/ml and rifampicin 25µg/ml and placed into an incubator at 28°C for 3 days.

Three days before the hairy root transformation, liquid cultures, containing kanamycin 50µg/ml and rifampicin 25µg/ml, were prepared from single colonies and incubated at 28°C for 2 days. The day before the hairy root transformation was to take place, 1ml of each liquid culture was transferred to LB plates containing kanamycin 50µg/ml and rifampicin 25µg/ml. and incubated at 28°C overnight. After 24 hours, each plate was covered with a thick layer of bacterial cells.

2.5.3 Hairy Root transformation

Each bacterial layer was first diluted by adding 3ml of sterilized ddH_2O to the plate and then resuspended by scrapping the surface of the plate with a pipette's tip. Part of the resuspended culture was then used to soak a sterilized filter paper placed in an empty plate. The seedlings were placed on the soaked filter paper and the primary root of each seedling was cut right under the hypocotyl using a sterilized scalpel. Then, a small amount of the remaining culture was injected to the wounded site(Figure 7). The seedlings were then transferred to ½ MS plates (10-15 seedlings per plate). The process was repeated for every construct and then the plates were covered with tin foil and placed vertically at 22 °C. The tin foil was removed after 2 days and the plates were maintained vertically to maintain root orientation, at 22 °C with a 16h light/ 8h dark photoperiod for 2 days. Five days after the co-cultivation with *A. rhizogenes*, the explants were transferred to ½ MS plates (10-15 per plate) with ceftazidime 300μg/ml ,in order to stop the co-cultivation with the bacteria, and the plates once more maintained vertically at 22 °C with a 16h light/ 8h dark photoperiod.

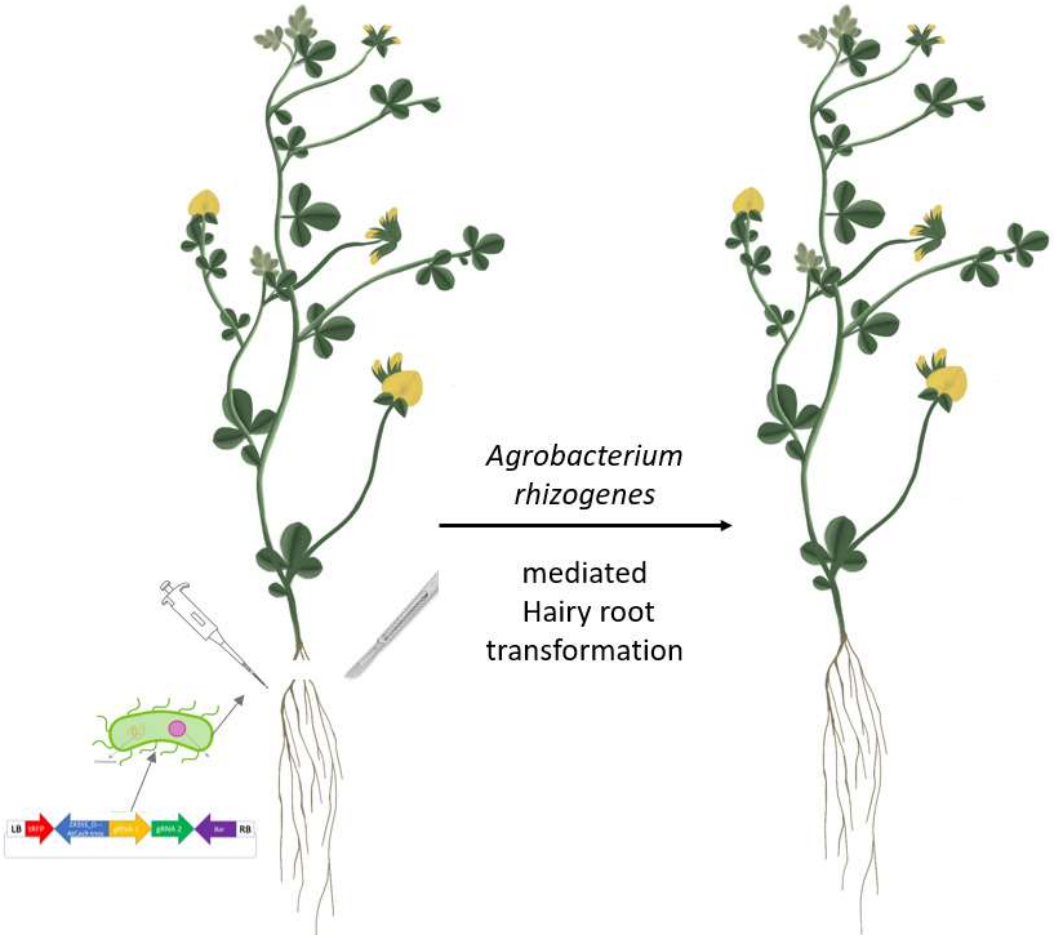

Figure 7 Visualization of the workflow for A. rhizogenes mediated Hairy root transformation. On the left the process of infecting the L. japonicus plants with the transformed Agrobacteria by wounding-cutting off the roots. On the right the plants after regeneration of "hairy" roots.

2.6. Sampling-*Mesorhizobium loti* inoculation

2.6.1 Root screening-first sampling-transferring to pots

3 weeks post hairy root transformation, regenerated roots were observed under a stereomicroscope for Red (orange) fluorescent protein TurboRFP (tRFP). The excitation filter of the stereoscope allowed for the transmission of wavelength ranging from 560-580nm, while the emission filter allowed for observation of wavelengths ranging from 600-660 nm

allowing for the excitation and observation of tRFP. Some transformed roots were saved for preliminary mutation screening by cutting them off, freezing them in liquid nitrogen and storing them at -80°C until analysis. The non-transformed roots were discarded and a single transformed root remained on each plant. Plants were then transferred to pots containing a sterilized mixture of two kinds of sand and vermiculite in a 2:1:1 ratio (MIT2000:Yellow Silica: Vermiculite). 30ml of Hoagland nutrient solution without nitrogen mixed with dH_2O in 1:1 ratio was added to each pot before planting. Plants were watered every two to three days alternately with dH_2O and Hoagland solution (with nitrogen for control plants and without nitrogen for plants inoculated with rhizobium) until sampling.

2.6.2 *Mesorhizobium loti culture preparation*

Six days before inoculation a *Mesorhizobium loti* stain R7A culture was transferred to a YMB plate and incubated at 28°C for three days. A single colony was then transferred into a liquid YMB medium and incubated at 28°C, 160rpm for three days.

2.6.3 Mesorhizobium loti inoculation and second sampling

1 week after planting, half the number of plants was inoculated with 1ml of liquid culture containing *Mesorhizobium loti*. 4 weeks post rhizobia infection, roots were again checked for fluorescence and transformed ones were sampled for mutation analysis.

2.7 DNA extraction

DNA was extracted from the sampled roots using the CTAB protocol. 100mg of frozen grinded tissue were incubated with 300µl of preheated CTAB extraction buffer at 65°C for 20 minutes, while occasionally mixing. After the incubation, an equal amount of chloroform: isoamyl alcohol was added. After vortexing and centrifuging the samples at 11.000g for 5min, the supernatant was transferred to a new tube. Next, cold isopropanol was added in a 0.7 volume ratio into the aqueous phase. The mixture was incubated at -20°C for 10min and then centrifuged at 11.000g for 15min. The supernatant was discarded, and the pellet was washed with 500µl of cold 70% ethanol. The samples were then centrifuged for 5min at 11.000 g to secure the pellet. Supernatant was discarded and the tubes were left to dry on the bench. After drying, DNA pellet was resuspended in ddH_2O with RNAse (49/1 ratio).

2.8 Mutation analysis Q5-T7

The regions of *Lotus japoninus* genome targeted by gRNAs were amplified using the Q5 High-Fidelity DNA polymerase in a 15µl reaction to verify the possible mutagenesis due to the CRISPR/Cas9 sytem . The same process was repeated in regions that might be affected as off-targets of the CRISPR/Cas9 system, to confirm the specificity of the gRNAs . The cycling steps, as well as the components of the Q5 PCR reaction. can be seen at Table 11 & Table 12. After the reaction, the samples were analyzed by agarose gel 1% electrophoresis.

Component	Volume µl	Volume µl
Q5 High-Fidelity 2X Master Mix	7,5	25
10 µM Forward Primer	0,75	2,5
10 µM Reverse Primer	0,75	2,5
DNA sample	2	2
ddH_2O	Till 15	Till 50

Table 11. Components of Q5 PCR reaction. The 15µl reaction was used for the T7 assay while 50l reaction was used for the Gel extraction.

Temperature	Time	-
98°C	30sec	Once
98 °C	10sec	Repeat 35
Ta	30sec	Repeat 35
72 °C	1min	Repeat 35
72 °C	2min	Once

Table 12. Steps of Q5 PCR reaction

Primers	Sequence 5'→3'	Tm	Ta
mut_AMY2-intron9_F	GCACAGCACAAAAAAAGAATT	63°C	66°C
mut_AMY2-intron10_R	TCAAGAATTCTCAGGTCTGCTT	63°C	66°C
mut_AMY2-exon4_F	CGGATGGAATTGTTATCCAAG	58°C	61°C
mut_AMY2-exon4_R	CTGGATGCATAGGAAGAAATG	58°C	61°C
mut_AMY2-intron1_F	ATTCAAACCATACTTCGCAC	58°C	61°C
mut_AMY2-intron1_R	CTTGGCGTATTCTCTCTTCG	58°C	61°C
mut_LSK1-exon 3_F	GACCAATAAACTTGGTGTG	56°C	59°C
mut_LSK1-exon 5_R	ACCACAACAAATTAAAATGTGA	56°C	59°C
LSK1_gRNA1_off F	AACCGGTACCAACGACACG	58°C	61°C
LSK1_gRNA1_off R	CCCGTTTCATTTCAGGGAGGG	58°C	61°C
LSK1_gRNA3_off F	ACGATGACAAACTCCAATTACACG	60°C	64°C
LSK1_gRNA3_off R	CAGACAATCTCATACATGGCGG	60°C	64°C

Table 13. List of primers used in Q5. The temperature during the annealing (Ta) was determined as 3 degrees above the Tm of each primer.

The Q5 amplified samples were used in the T7 assay reaction where every sample was separated in two reactions. The T7 endonuclease 1 (T7E1) is a structure-selective enzyme that recognizes and cleaves structural deformities in heteroduplexed DNA like mismatched or nicked double-stranded DNA. The T7E1 is widely used on mismatch detection assays as a method for evaluating the activity of site-specific nucleases, like the CRISPR/Cas9 system mutations (Sentmanat et al., 2018). The components are shown in Table 14 with the cycling steps of the denaturation-hybridization reaction being shown at Table 15. After those steps 0,7µl of T7 were added into one of the two reactions for every sample and then all samples were incubated at 37 ^{0}C for 30 min. After the incubation, 1µl of EDTA 0,5M was added to deactivate the T7 enzyme. Reactions were analyzed by electrophoresis in 2% agarose gel. The T7E1 assay validated CRISPR reagents by comparing the banding patterns of the cut products between control and experimental samples to determine the presence and frequency of. Representative T7-positive samples were prepared for Sanger sequencing to verify the mutation efficacy of each gRNA and identify the type of indels.

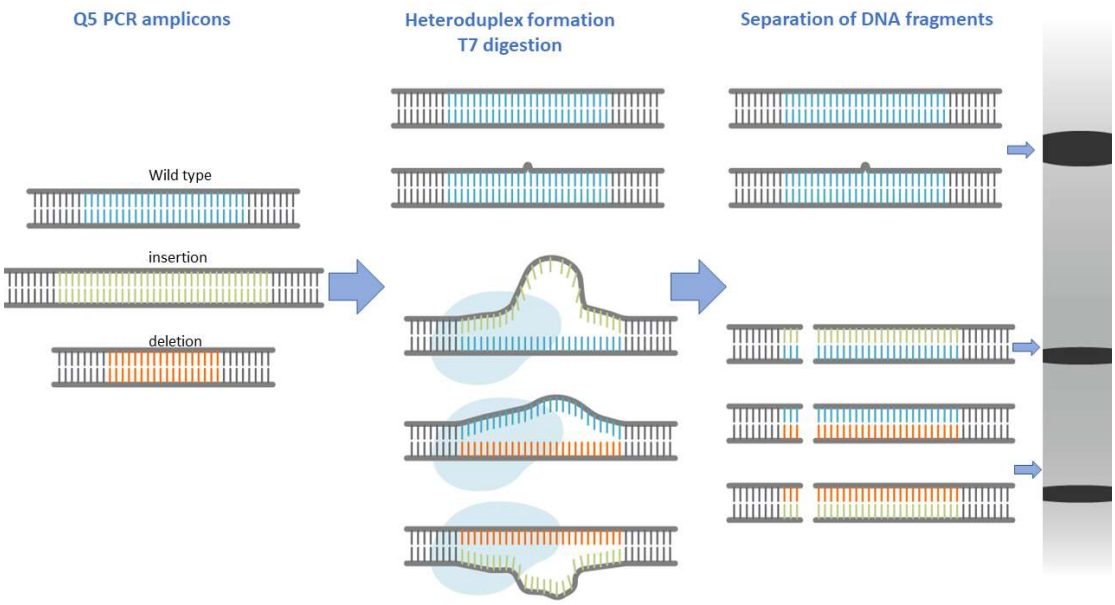

Figure 8 Schematic of T7 assay. Amplification of control and CRISPR/Cas9 edited sequences(Left). Formation of heteroduplexed DNA though denaturation-hybridization(Middle) and cleavage of mismatched DNA though treatment with T7EI. Analysis of fragments run on a gel(Right).

Component	Volume µl
T7 Buffer 2X	2
Q5 amplicon	5
ddH_2O	12

Table 14. Components of T7 denaturation-hybridization reaction mixture

Temperature	Time	heating rate
95°C	5min	3 °C/sec
85°C	1sec	2 °C/sec
75°C	1sec	0,1 °C/sec
65°C	1sec	0,1 °C/sec
55°C	1sec	0,1 °C/sec
45°C	1sec	0,1 °C/sec
35°C	1sec	0,1 °C/sec
25°C	1sec	0,1 oC/sec

Table 15. Cycling conditions for the T7 assay

DNA samples were once more amplified using the Q5 High-Fidelity DNA polymerase in a 50µl reaction in order to procure a greater concentration of the amplified product. The cycling steps, as well as the components of the Q5 PCR reaction, can be seen at Table 11 & Table 12. The samples were then analyzed by agarose gel 1% electrophoresis and expected bands were cut for downstream purification and cloning.

2.9 Gel Extraction of amplified target regions-Cloning

Amplicons were then purified using the Macherey-Nagel GEL AND PCR CLEAN-UP kit following the manufacturer's instructions. The concentration of each amplicon was measured using a nanodrop spectrophotometer. In the case of LSK1 target region, purified amplicons were also sent for sequencing to .

2.9.1 TA cloning

Amplicons were prepared for TA cloning by adding one adenine nucleotide to their 3' blunt end (process known as A-tailing). The components of the A-tailing reaction are shown in Table 16. Reactions were incubated at 70°C for 30'.

Component	Volume µl
ddH_2O	0,8
Buffer A	1
dATPs	0,2
Kapa Taq	1
Purified products (amplicons)	7

Table 16. Components of A-tailing enzymatic reaction mixture

The amplicons were then ready to be cloned into the pGEM®-T Easy Vector. The components of the cloning reaction are shown in Table 17.

Component	Volume µl
ddH_2O	4,5
T4 ligase buffer	1
T4 ligase	1
the pGEM®-T Easy Vector	0,5
Amplicons	3

Table 17. Components of the TA cloning reaction mixture

2.9.2 CloneJET PCR Cloning

For the analysis of mutations in AMY2 target regions, as well as LSK1, blunt end amplicons were also ligated to the cloning vector pJET, followed by a 5' incubation at room temperature. Cloning conditions are shown in Table 18.

Component	Volume µl
2x Reaction Buffer	10
Purified PCR product	1
pJET1.2/ blunt cloning vector (50ng/µl)	1
T4 ligase	1
ddH_2O	Up to 20

Table 18. Components of the CloneJET PCR Cloning reaction mixture.

All cloning reactions were used to transform *E. coli* DH5a competent cells. Plasmid DNA was isolated with same method as mentioned above and sent for Sanger sequencing.

Primers	5→3 Sequence	Tm
T7 (pGEM)	TAATACGACTCACTATAGG	63°C
Sp6 (pGEM)	TCTATAGTGTCACCTAAAT	63°C
PJET F	CGACTCACTATAGGGAGAGCGGC	62°C

PJET R	AAGAACATCGATTTTCCATGGCAG	58°C
LSK1_mut1F (amplicon)	GACCAATAAACTTGGTGTG	56°C
LSK1_mut1R (amplicon)	ACCACAACAAATTAAAATGTGA	56°C

Table 19. Primers used for sequencing the pGEM and pJET plasmid vectors.

2.10 Analysis of sequencing results

After receiving the sequencing results, we proceeded with the analysis. We compared the sequence derived from the chromatograms with the wild-type sequence. During these comparisons, we verified the presence of mutations in the samples. We created separate files containing the alternate sequence for each observed mutation, using SnapGene software. After collecting the sequence of all the observable mutations, we proceeded with aligning those sequences using the wild-type sequence as a reference. The alignment was done using the software Geneious Prime 2023.1.2. After performing the alignment, we analyzed the mutations for their possible effect in the translation of the encoded gene.

2.11 Media and solutions

In various stages of the experiment the following media and solutions were used:

- LB (liquid): add 10gr of meat peptone, 10gr of NaCl and 5gr yeast extract per litre of dH_2O.
- LB (agar): add 10gr of meat peptone, 10gr of NaCl, 5gr yeast extract and 15gr of bacteriological Agar per litre of dH_2O.
- YMB (liquid): add 10 gr mannitol, 0,5 gr K_2HPO_4, 0,2 gr $MgSO_4$ $7H_2O$, 0,1 gr NaCl, 1 gr Yeast Extract per litre of dH_2O.
- YMB (agar): add 10 gr mannitol, 0,5 gr K_2HPO_4, 0,2 gr $MgSO_4$ $7H_2O$, 0,1 gr NaCl, 1 gr Yeast Extract and 15 gr agar per litre of dH_2O.
- Water Agar: add 8gr of agar to 1lt of dH_2O.
- ½ MS (agar): 2,2gr of Murashige and Skoog salt, 20gr of sucrose and 8gr of bacteriological Agar per litre of dH_2O.
- CTAB: 8ml of 100mM Tris HCl (pH=8), 3,2ml of 20mM EDTA (pH=8,), 0,8gr PVP, 28ml of1,4M NaCl, 1,6gr cetyltrimethylammonium bromide (CTAB) and ddH_2O till volume=80ml.
- Chloroform-isoamyl: add 48ml of chloroform to 2ml of isoamyl alcohol.
- Hoagland nutrient solution without nitrogen: add 2 ml $MgSO_4$ 1M, 1 ml KH_2PO_4 1M, 1 ml Fe EDTA 0,1M, 1 ml Microelements (2,86 gr/L H_3BO_3, 1,81 gr/L $MnCl_2$ $4H_2O$, 0,11 gr/L $ZnCl_2$, 0,05 gr/L $CuCl_2$ H_2O, 0,02 gr/L H_2MoO_4), 10 ml $CaCl_2$ 0,5 M, 5 ml KCL 1M into 1lt of dH_2O.

Note: All media and solutions apart from the chloroform-isoamyl mixture were autoclaved at 121°C for 20'.

3. Results

3.1 Level 1 Cloning

Apart from the 1F_Lj500_tRFP_act2 (Figure 9), the reporter gene and some gRNA expression cassettes, all the other level 1 transcriptional units were already available in the lab. Following the procedure mentioned in section 2.1, we prepared the cloning reaction, transferred the resulting construct into competent cells and isolated the plasmid DNA. We performed diagnostic digestion on 1F_Lj500_tRFP_act2 clones with the EcoRI enzyme and analyzed the reactions with gel electrophoresis. The sizes of the expected bands were 881bp and 5804bp. A clone with the correct digestion pattern was sent for Sanger sequencing. In the case of gRNA cassettes (Figure 10), samples were sent directly for Sanger sequencing without prior diagnostic digestion. After receiving the sequencing results, we made use of the prepared maps to compare them, thus confirming their nucleotide sequence.

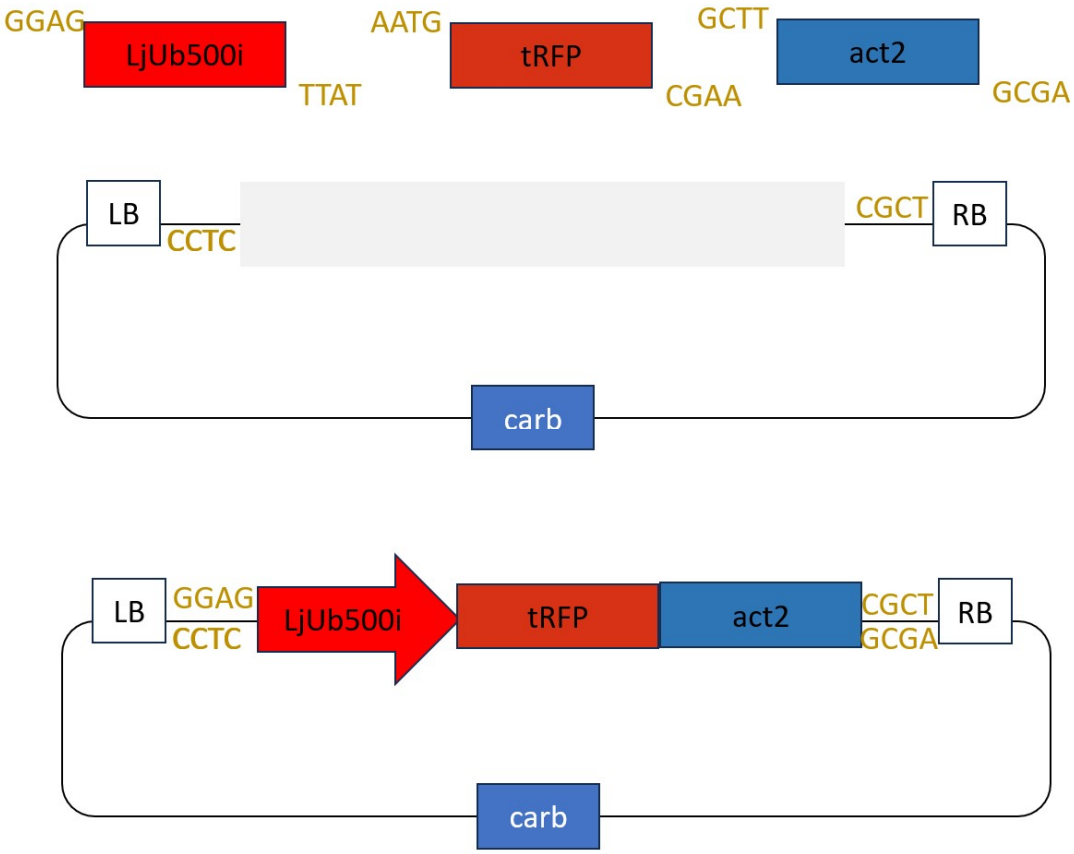

Figure 9 Visualization of the assembly of the level 1 reporter gene transcriptional unit though a digestion-ligation reaction of level 0 standard parts and an acceptor vector. The digestion was achieved with the use of the type IIS RE BsaI and the ligation with the T4 ligase.

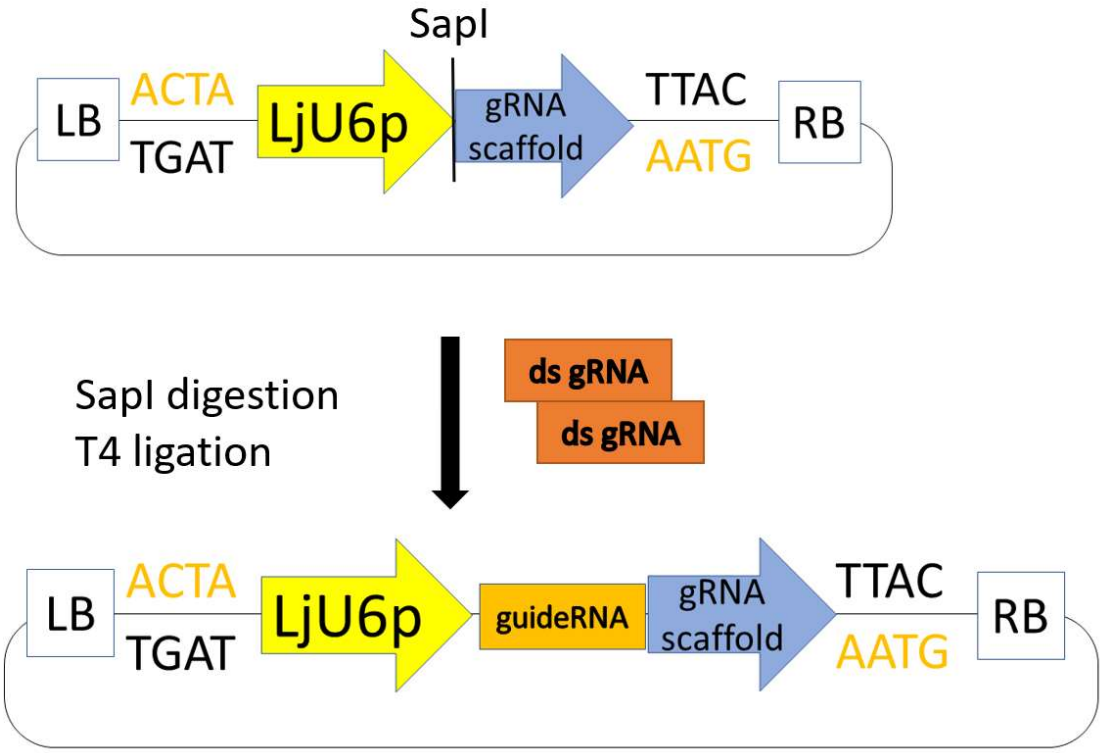

Figure 10 Visualization of the assembly of the gRNA's cassettes through a SapI / T4 digestion ligation reaction between a dsDNA containing the gRNA's spacer sequence and a level 1 TU the contains the gRNA acceptor.

3.2 Assembly of Level 2 constructs

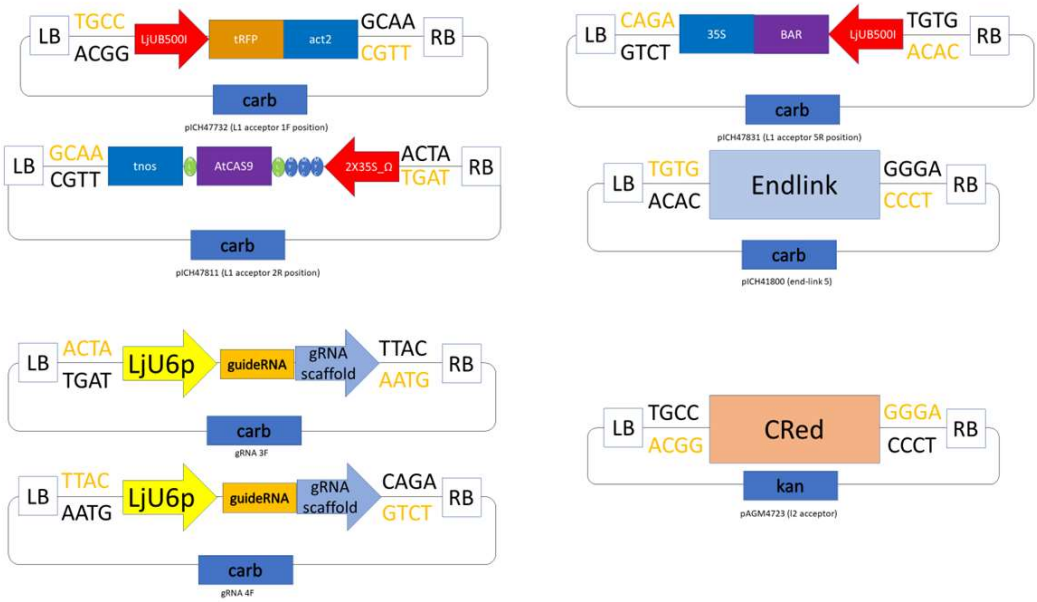

Figure 11 Level 2 library. Visualization of l1 transcriptional units(gRNA and cas9 cassettes , selection and reporter gene, endlinker) and destination vector taking part in the assembly of the l2 constructs.

By following the procedure described in section 2.1, we prepared the cloning reaction with all the required parts (Figure 11), transformed the competent cells, and isolated the plasmid DNA. A full list of the assembled constructs can be seen in table 10 and a visualization of the constructs in Figure 12. All of the level 2 constructs were first created by following the reaction shown in Table 8, with an acceptable level of success.

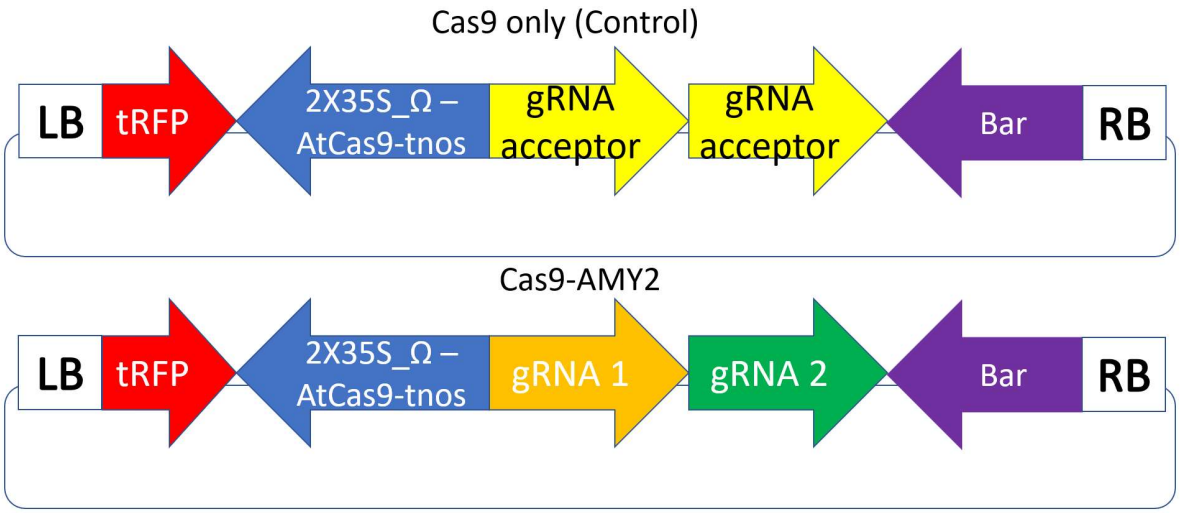

Figure 12 Visualization of an assembled level 2 multigene construct showing the position and orientation of the transcriptional units within the construct. From left to right we have the reporter gene (tRFP) in forward orientation, the Cas9 cassette in reverse orientation, the two gRNA cassettes in forward and the selection gene (Bar) in reverse.

After plasmid isolation, we chose 4-6 colonies per construct for digestion with the XmnI enzyme for validating our constructs based on expected banding patterns as described here. Since all constructs have the same sequence except for the few nucleotides that correspond to the gRNAs, the expected digestion band motif was the same for nearly all constructs: 11.153, 3525, 961, 430, 427bp (Figure 13). The colonies with the correct band pattern were sent for Sanger sequencing. After receiving the sequencing results, we compared them with the previously designed maps to validate the inserts.

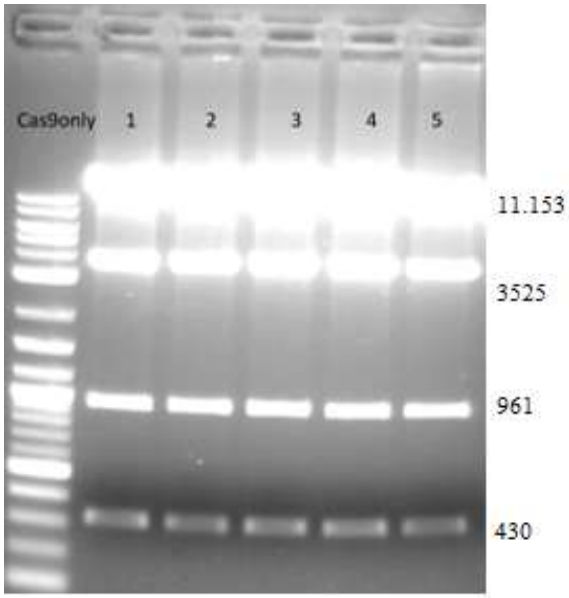

Figure 13 Representative image of a diagnostic digestion of a level 2 construct (Cas9only) run on gel for verification of successful cloning. DNA Ladder (0.1-10.0 kb)

As previously mentioned, there were some problems while assembling the level 2 constructs. The efficiency of the transformation was low for some constructs and in many cases, even white colonies gave an incorrect digestion pattern. After several trials, we decided to use a modified protocol for the cloning reaction, shown at Table 9. This increased the efficiency of the cloning reactions, allowing us to successfully assemble the rest of the constructs.

3.3 *Agrobacterium rhizogenes*-mediated hairy root transformation

The level 2 constructs were introduced into *A. rhizogenes* LBA 1334 strain by electroporation. Even though ten constructs were generated, including the Cas9-only and the pICH47732 (1F vector) that both function as a control, not all of them were utilized in the hairy root transformations of this master thesis (Table 18). All constructs were transformed into Gifu ecotype plants and some of them were transformed into MG20 ecotype plants as well.

***Agrobacterium rhizogenes* constructs**
pICH47732 (1F vector) (negative control for tRFP fluorescence)
Cas9 only (negative control for mutation analysis)
AMY2:gRNA1-gRNA3
AMY2:gRNA4-gRNA5
LSK1:gRNA1-gRNA3
LSK1:gRNA2-gRNA4
CYP71D353:gRNA1-gRNA2
OSC3:gRNA1-gRNA2

Table 20. List of constructs transformed into Agrobacterium rhizogenes.

We analyzed the plants transformed with AMY2:gRNA4-gRNA5 and LSK1:gRNA1-gRNA3. When observing the phenotypic characteristics of the hairy roots (length, number of vertical roots), there was no notable difference between the controls and the CRISPR/Cas9 roots. The verification of transformation success was performed by root observation under a stereoscope for tRFP fluorescence (Figure 14).

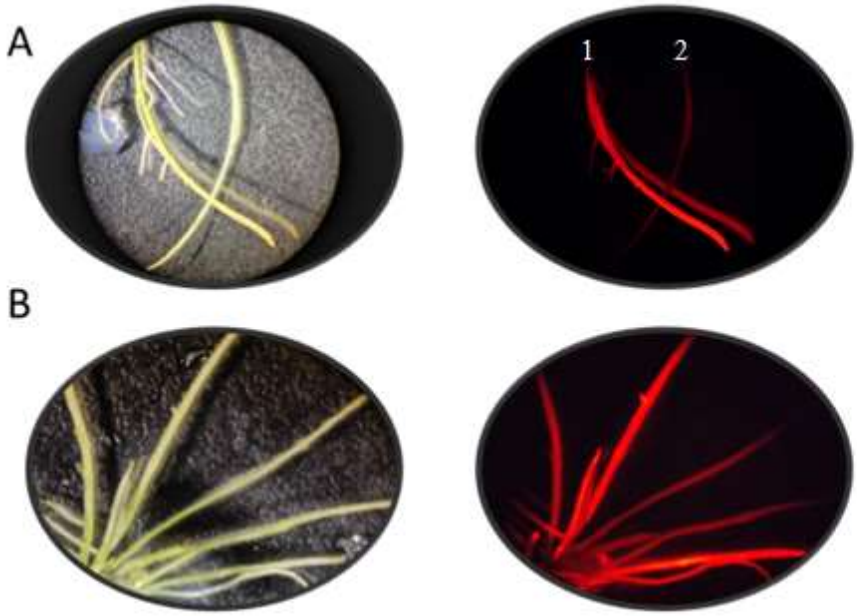

Figure 14tRFP positive hairy roots under stereoscope (in white light in the left and under laser in the right). A) Two roots are seen, the left one (1) was transformed, the one in the right (2) was not. B) More examples of tRFP positive roots (same plant). Excitation filter (560-580nm), Emission filter (600-660 nm).

3.4 Mutation analysis Q5-T7

3.4.1 Q5 PCR

To analyze the mutations in the transgenic hairy roots, regions of the genome including one or two target sites of gRNAs were PCR amplified using Q5® High-Fidelity DNA polymerase. A non-template control (NTC) was used as a negative control for the Q5 reaction to ensure that there were not any contaminations. DNA samples derived from transgenic roots transformed by the constructs AMY2:gRNA4-gRNA5 and LSK1:gRNA1-gRNA3 were used to search for mutations, while those transformed by Cas9only and pICH47732 (1F) served as controls .The target sites amplified by Q5 are shown at Table 21 . In Figure 15, samples that were amplified for the AMY2 intron 9- intron 10 region are shown. The expected band at 612bp is present in most samples. A second band at 300 bp, indicating a deletion in that region is detected in sample g.

Genomic region	Amplicon size bp
AMY2: intron9 – intron10	612
LSK1 : exon 3 -exon5	892

Table 21. List of amplicon sizes generated by Q5 polymerase for each target region.

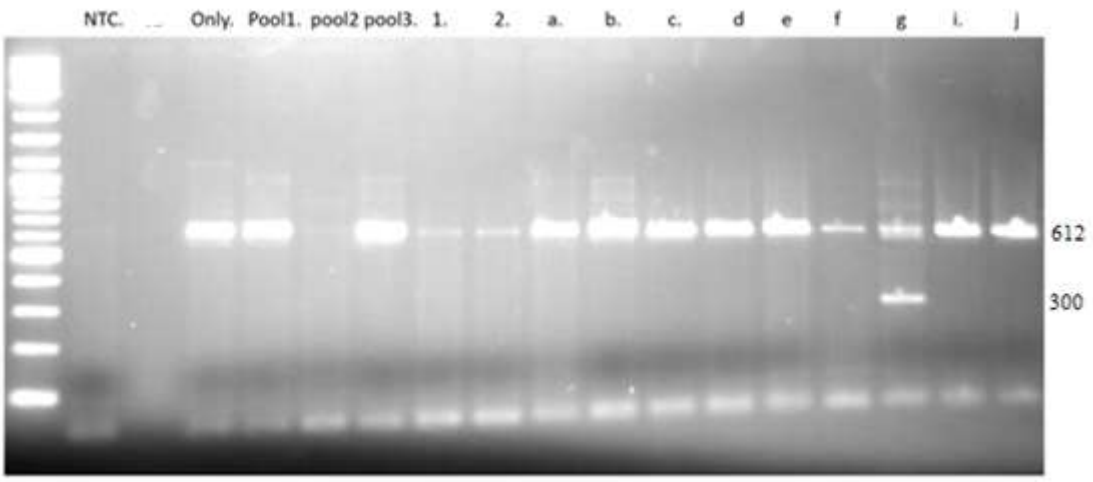

Figure 15. Q5 polymerase PCR reaction at the amplified AMY2 intron 9 and intron 10 target regions run on gel. The NTC is the negative control with the rest being samples derived from transgenic roots transformed by the Cas9only and AMY2:gRNA4-gRNA5 constructs. The Only is a root sample transformed by the Cas9only construct, the pool(1-3) are samples derived from multiple roots while the a-j derived from single roots transformed by the AMY2:gRNA4-gRNA5 construct. DNA Ladder (0.1-10.0 kb)

3.4.2 T7 endonuclease assay for the LSK1 exon 3- exon 5 region.

The T7 assay was used to detect mutations on the samples amplified by the Q5 PCR. Some mutations were detected in samples of transgenic roots 3 weeks post transformation with the LSK1:gRNA1-gRNA3 construct. The results are shown in Figure 16. In the case of absence of a mutation, the only expected band would be at 892bp, as exemplified in sample 3. In the rest of the samples, secondary bands were observed, indicating the detection of possible mutations. We made an inference on the site of the mutation based on the motif of the bands. In sample 1 the mutations likely happened in the region targeted by LSK1:gRNA1, since there is a faint band around 800bp. A second band should have been seen at around 90bp but due to the size and the likely faint signal we were not able to observe it. Similarly, samples 2 and 4 show similar motifs with bands around 800bp, 600bp and 250bp. From those motifs it likely that mutations occurred on the region of both gRNAs but in different cells.

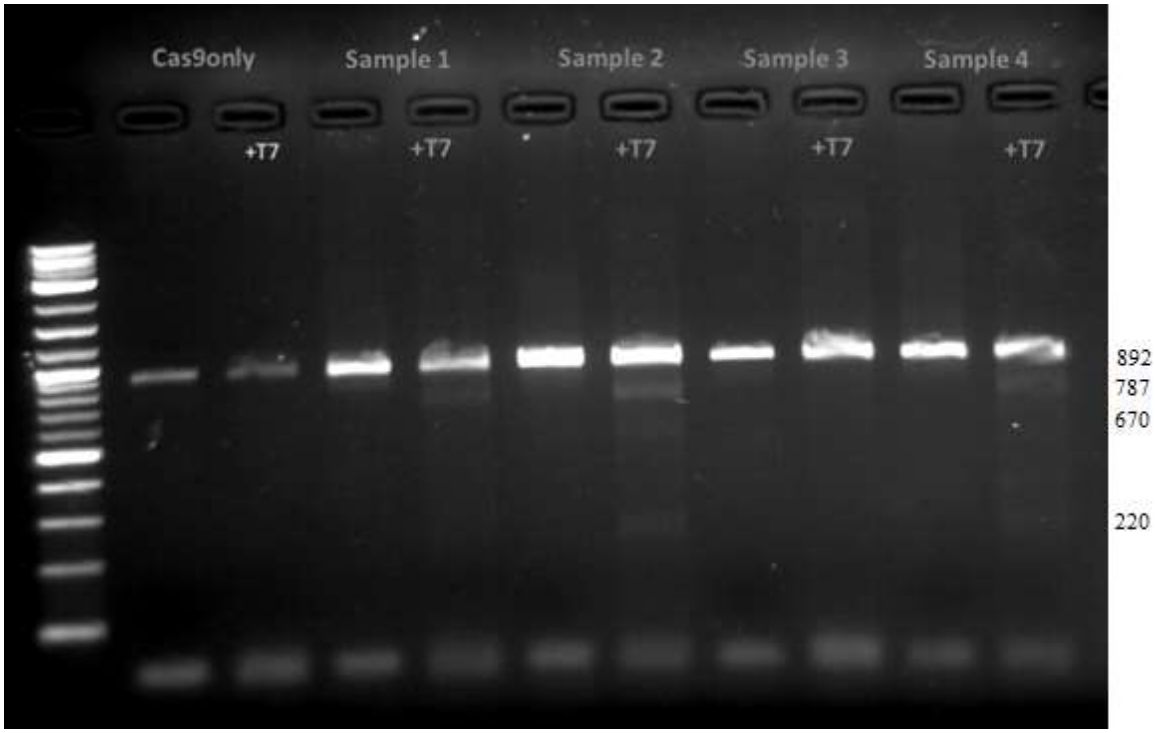

Figure 16 T7 endonuclease assay at the amplified LSK1 exon3- exon5 region run on gel. The Cas9 only serves as a negative control with the rest of the samples being from transgenic roots transformed by the LSK1:gRNA1-gRNA3 construct. Mutations detected in samples 1,2 and 4. DNA Ladder (0.1-10.0 kb)

After verifying the presence of mutations in the target site, we also checked for possible off-target effects of Cas9, by amplifying genomic regions where the off-target might have taken place. For the LSK1:gRNA1-gRNA3 samples we amplified with two sets of primers one for each gRNA. Using the T7 assay we did not detect any off-target activity of Cas9 (Figure 17& Figure 18).

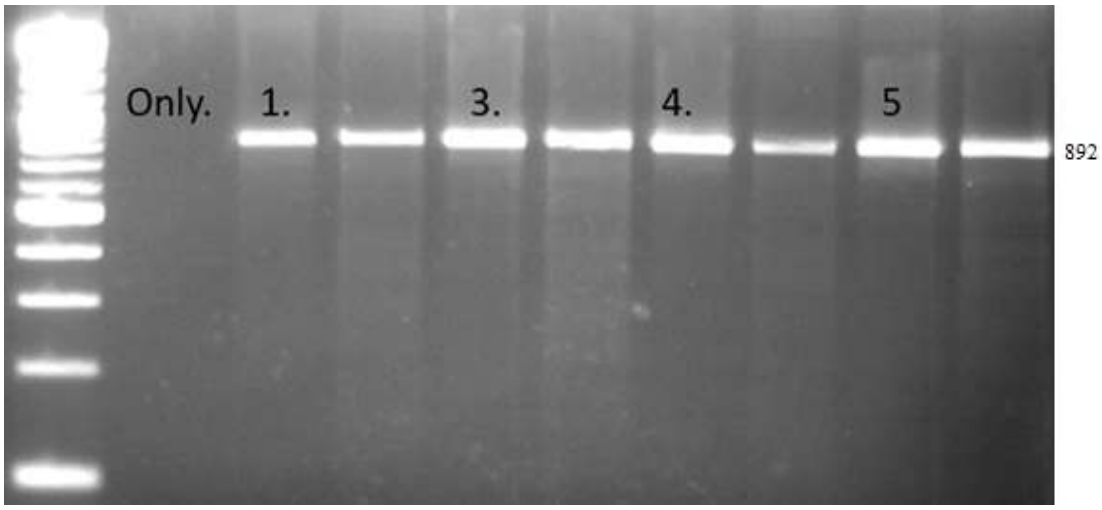

Figure 17 T7 endonuclease assay for the amplified LSK1:gRNA1 possible off-target sequence run on gel. No off-target effects detected in either control(Only) or genome edited(1,3,4,5) samples. DNA Ladder (0.1-10.0 kb)

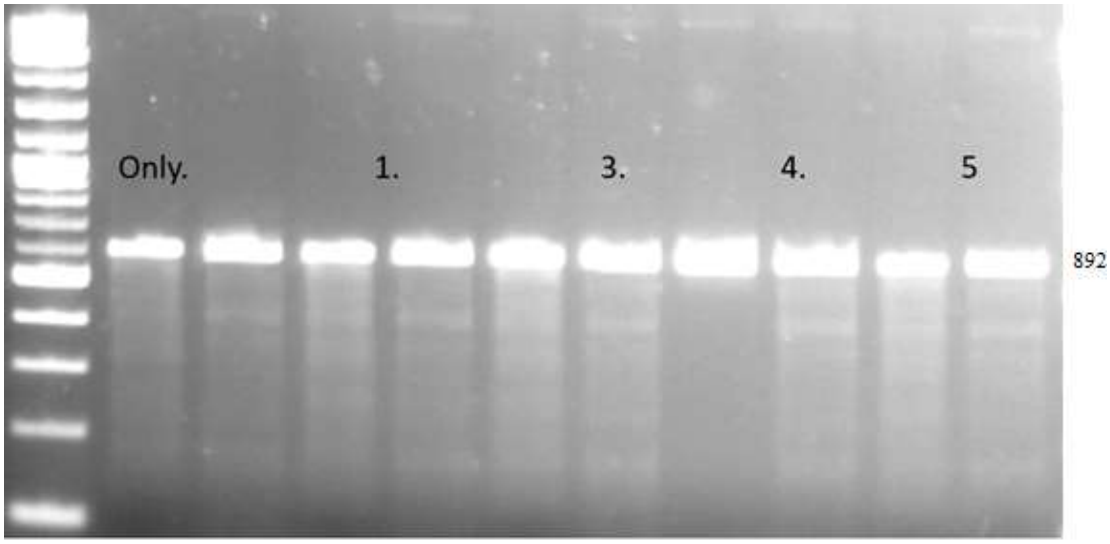

Figure 18 T7 endonuclease assay for the amplified LSK1:gRNA3 possible off-target sequence run on gel. No off-target effects detected in either control(Only) or genome edited(1,3,4,5) samples. DNA Ladder (0.1-10.0 kb)

3.4.3 T7 endonuclease assay for the AMY2 intron 9- intron 10 regions.

The T7 assay was used once again to detect mutations on the samples amplified by the Q5 PCR. Some mutations were detected in samples of transgenic roots 3 weeks post hairy root transformation (Figure 19) with the AMY2:gRNA4-gRNA5 construct. Further mutations were detected in samples collected 4 weeks after either a mock inoculation (Figure 20) or an inoculation with R7A (Figure 21) (in total 8 weeks post hairy root transformation). In the absence of mutations, the only expected band would be at 612bp. In the samples where secondary bands were observed, indicating the detection of possible mutations. The bands detected at around 500 bp indicated that the mutation occurred on the AMY2:gRNA4. In this case, the second band expected at 100bp was missing, likely due to low resolution of the staining of the agarose gel, as explained before. The bands observed at around 400 bp and 200bp indicated that the mutation occurred on the AMY2:gRNA5 but the second band was also missing in some samples. In some samples, we observed mutations on both sites, as depicted by a combination of the previous band patterns.

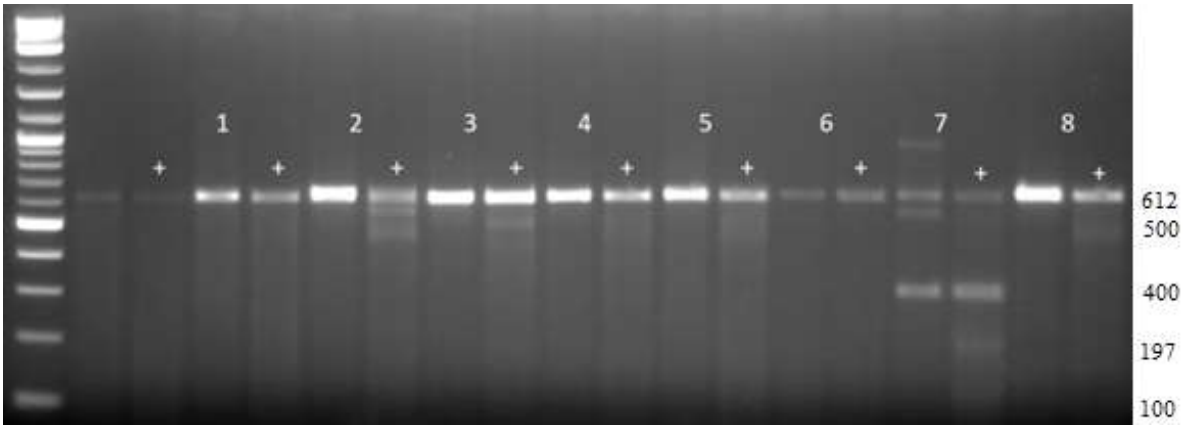

Figure 19 T7 endonuclease assay for the amplified AMY2 intron 9 intron 10 region run on gel. The Cas9 only (non-numbered sample) serves as a negative control with the rest of the samples being from transgenic roots 3 weeks post transformation with the AMY2:gRNA4-gRNA5 construct. Mutations detected in samples 2,3, 7 and 8. DNA Ladder (0.1-10.0 kb)

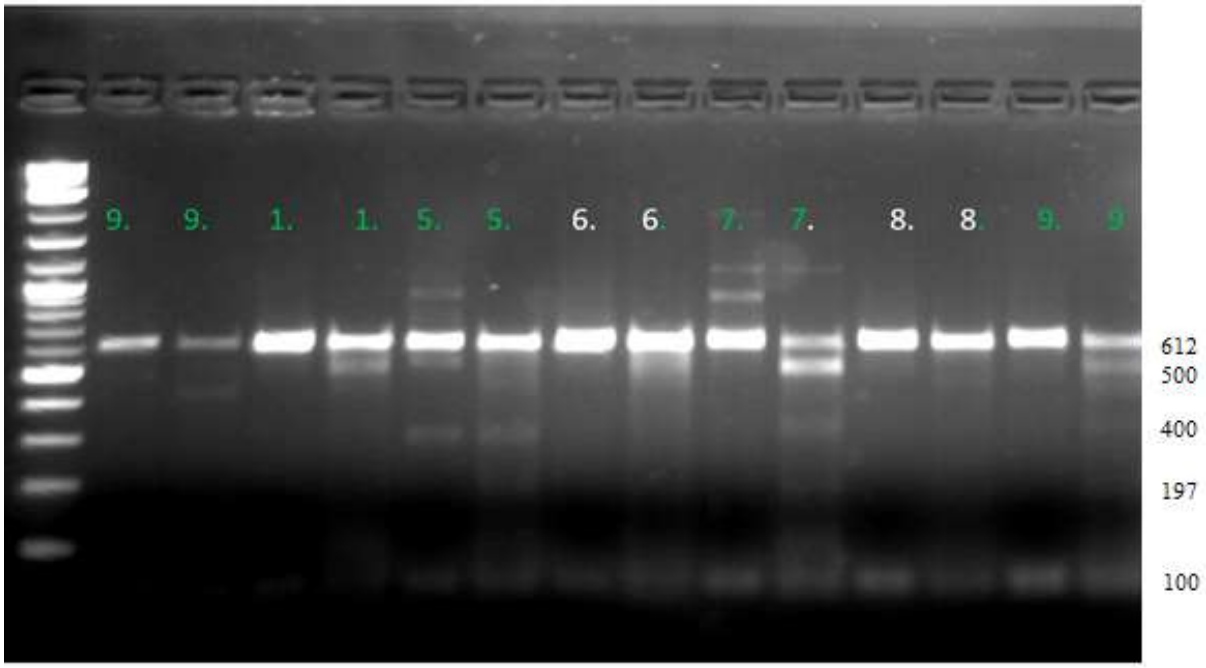

Figure 20. T7 endonuclease assay for the amplified AMY2 intron 9- intron 10 region run on gel. Samples are from transgenic roots transformed with the AMY2:gRNA4-gRNA5 construct 4 weeks post mock inoculation (R7A-). Samples with putative mutations detected are numbered in green. DNA Ladder (0.1-10.0 kb)

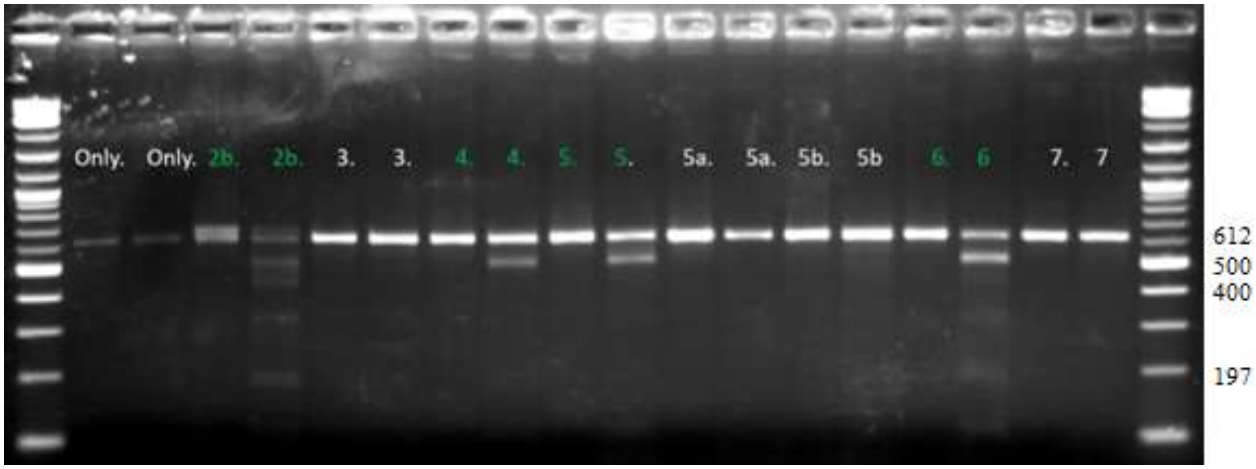

Figure 21. T7 endonuclease assay for the amplified AMY2 intron 9- intron 10 region run on gel. The Cas9 only sample serves as a negative control with the rest of the samples being from transgenic roots transformed with the AMY2:gRNA4-gRNA5 construct 4 weeks post mock inoculation (R7A+). Samples with putative mutations detected are numbered in green. DNA Ladder (0.1-10.0 kb)

3.5 Analysis of Sanger sequencing results

After verifying the presence of mutations in the target region the next step was to prepare the samples for Sanger sequencing and have the results analyzed in order to confirm and identify the type of INDELS on the *LSK1* and *AMY2* genes. The samples were compared to the wild type sequence in order to pinpoint the differences in the sequence. For the LSK1 gene, we analyzed twelve clones but only two caried a modified allele The clones were derived from different roots transformed by the LSK1 gRNA1-gRNA2 construct. One was a 1bp deletion and the other was a 4bp deletion (Figure 22).

Since the LSK1:gRNA3 targets the exon 5 of *LSK1*, we checked if those mutations result in changes in the translation of the gene. Both mutations disrupt the reading frame, creating a premature termination codon. Regarding the LSK1:gRNA1, the analysis of the Sanger chromatograms did not reveal any mutations.

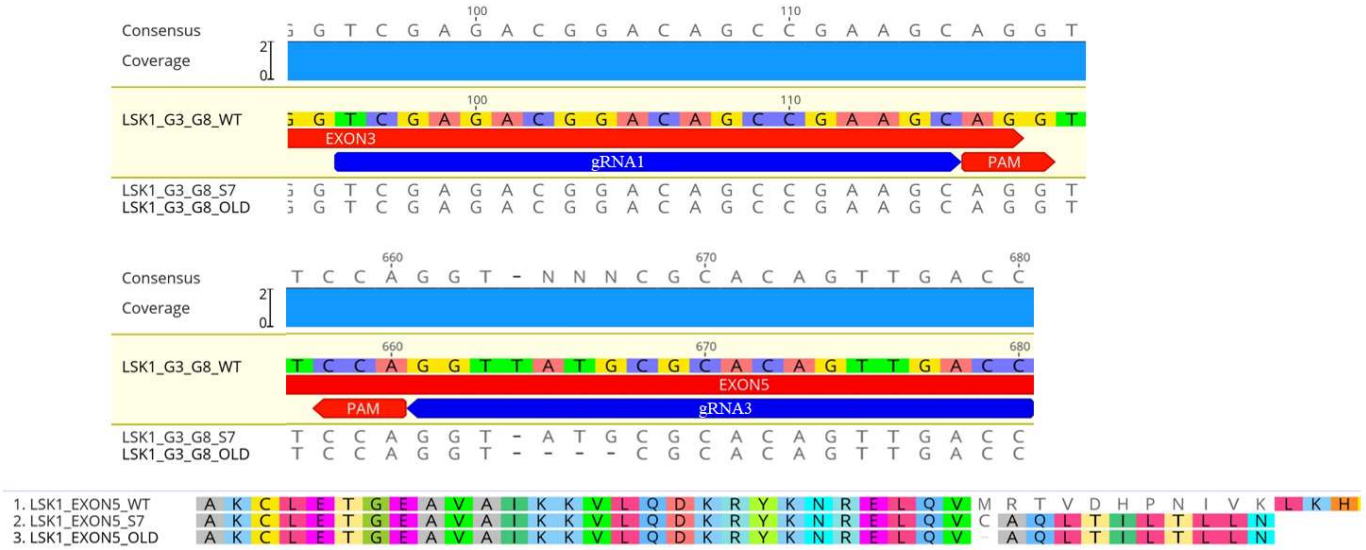

Figure 22. Analysis of the LSK1 gene, exon 3 to exon 5 region. Short deletions in the LSK1:gRNA3 targeted sequence. Analysis revealed that the deletion caused a frameshift that creates a premature stop codon.

For the *AMY2* gene, multiple mutations were detected across various samples. Based on the type of mutations that were uncovered, we grouped them together for the analysis. In the first group (Group A) we placed the samples in which we identified short indel mutations in the regions targeted by the gRNAs. In general, for the region targeted by the AMY2:gRNA5 we observed the deletion of four specific base pairs and one point insertion, while in the region targeted by the AMY2:gRNA4 we observed deletions of diverse sizes and a point insertion that appeared in many samples (Figure 23).

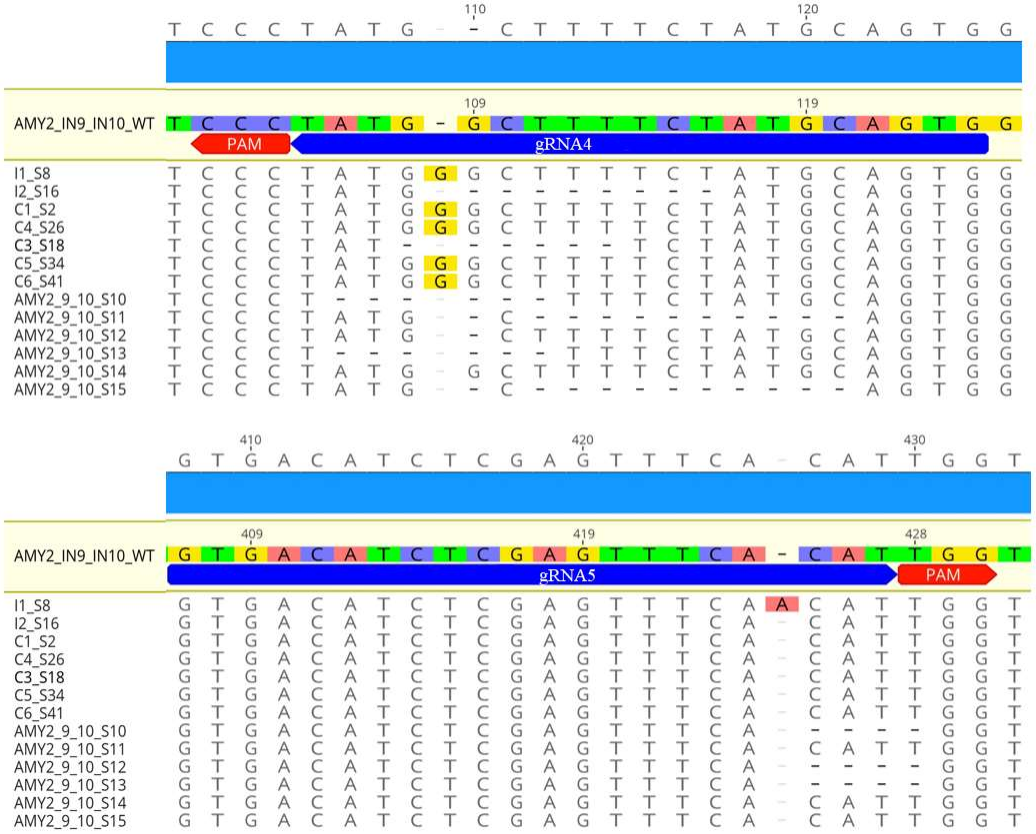

Figure 23. Analysis of AMY2 gene intron 9 to intron 10 region. Group A, short indels in the sequence targeted by gRNAs.

In the second group (Group B) are the samples in which we observed the deletion of the DNA sequence that contained whole of exon10. In that case no frameshift would occur, but the deletion of exon 10 would impact the function of the AMY2 protein considering that the active site is located in that particular exon (Figure 24).

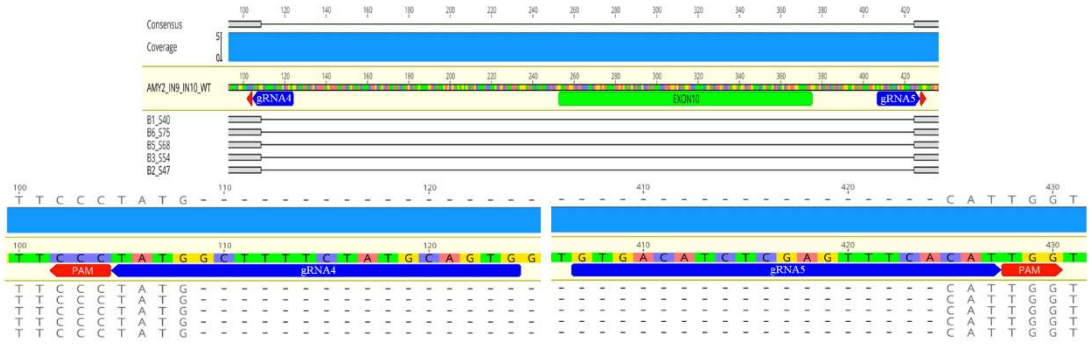

Figure 24. Analysis of AMY2 gene, intron 9 to intron 10 region. Group B: exon 10 knockout.

In the third group (Group C), we placed the samples where we detected an insertion, which occurred within the region corresponding to AMY2:gRNA4 (Figure 25). When annotated, the sequence of that insertion was revealed to correspond to a region of Ri plasmid of *Agrobacterium rhizogenes*. Since the insertion took place within an intronic region it should not affect the transcription or translation of AMY2.

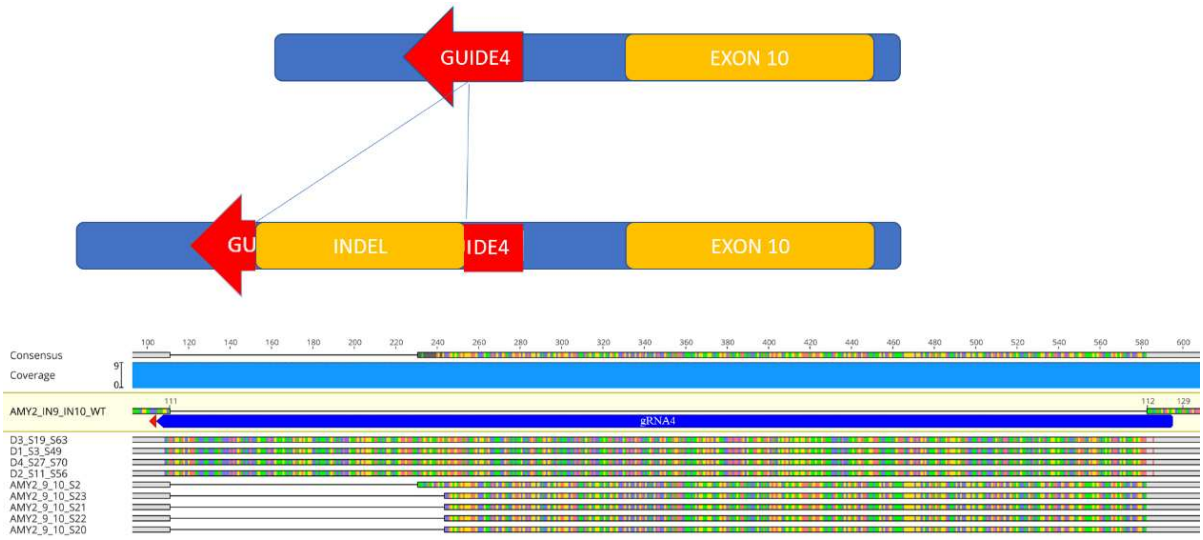

Figure 25. Analysis of AMY2 gene, intron 9 to intron 10 region. Group C, insertion on the AMY2:gRNA4 target region.

4. Discussion-Upcoming Experiments

The Legume (Fabaceae) family encompasses a diverse range of plants that are valued for their high nutritional content in many cultures around the world. Under nitrogen-poor conditions, they are capable of forming symbiotic relationships with bacteria of the genus *Rhizobium*, thus covering their own needs in nitrogen and helping replenish the nitrogen levels in the soil. As a result their presence has a positive impact on both natural and agricultural ecosystems (Márquez & Stougaard, 2005). In addition, legumes, like most plants, are known for producing a vast number of specialized metabolites that confer advantageous characteristics to the plants during their natural development or in adverse conditions (Open et al., 2020). One of the keys to understand and improve the natural resistance to stress factors lies in understanding the plant physiology as well as the pathways relating to the secondary metabolism and symbiotic relationships. The model plant *L. japonicus* is one of the species employed to uncover the details behind such intricate regulations (Küster, 2013; Stougaard, 2013).

The study of nodulation in *L. japonicus* revealed an interwoven association with the triterpene metabolism, part of the vast secondary metabolism, and with a family of *GSK3b/SHAGGY-like* kinases. In particular, the *AMY2* gene seems to regulate aspects of nodulation such as nodule organogenesis and progression of infection (Krokida et al., 2013), while *LSK1* seems to regulate the expression of AON (AutO regulation of Nodulation) components (Garagounis et al., 2019). At the same time, *in vitro* experiments provided evidence about lupeol's binding to LSK1, thus possibly linking all three pathways (unpublished data). We, thus, aimed to investigate the putative involvement of the triterpene pathways and *LSK1* in rhizobial symbiosis. The first step to uncover these interactions is the functional analysis of mutant plant lines without active components in these pathways. In our lab, a CRISPR/Cas9 system was optimized for use in *L. japonicus* by testing various *Cas9* genes. During this study, we assembled this CRISPR/Cas9 system as a Golden Gate-based binary vector and employed it for *A. rhizogenes*-mediated hairy root transformation. This method was used to assess the functionality of various designed gRNAS for targeting *AMY2* and *LSK1*.

A substantial number of gRNAs were designed to modify and silence the target genes. Each construct was designed to target a specific gene. The construct that contained our system is comprised of a reporter gene (tRFP), a selection marker gene (BAR) that will be of use in later stages of the experiments for the selection of transgenic tissues, two gRNAs that would target the same gene and the Cas9 cassette.

During the assembly of the l2 constructs, there seemed to be a problem with the cloning reaction procedure that resulted in either low-level or unsuccessful assembly of some of the constructs. To overcome that hurdle, we modified the protocol by increasing the vector-insert concentration without changing the ratio and by adding BSA in the cloning reaction, as suggested by the literature (Farell & Alexandre, 2012). The modified protocol allowed for the successful assembly of the remaining constructs. In the end, we managed to assemble six constructs that would target the various target genes and a seventh one that would function as a negative control. Due to time constraints, only three of the constructs targeting the *AMY2* and *LSK1* genes and the control constructs were used within the context of this work.

To differentiate between transformed and non-transformed roots, we used the tRFP fluorescence. This screening was used for two reasons. Firstly, because the presence of fluorescence allows the visual selection of the transgenic roots. In addition, the reporter gene

was placed near the Left Border (LB) of the T-DNA and since the RB is usually the initiation site for T-DNA integration, with the LB being the last to be integrated into the genome and more prone to imperfect transfer, the expression of the fluorescent protein indicates for a successful integration of T-DNA (Gelvin, 2021; Sahab & Taylor, 2022).

The tRFP screening technique was used to determine the efficiency of the *A. rhizogenes* to transform the roots and to differentiate between the transformed and non-transformed roots even in the same plant when observed under the stereoscope. The verification of mutations on the tissues that exhibited fluorescence in the next steps of the established procedure, indicates that our screening method is effective. When screening for mutations using the T7 assay, we also attempted to determine the efficiency of the CRISPR/Cas9 system. For LSK1 we were able to find indications that there were mutations in the region targeted by LSK1:gRNA3 and possibly in the LSK1:gRNA1 region as well. However, when analyzing the sequencing results, we only identified mutations only on the LSK1:gRNA3 sequence. That could mean that the efficiency of the LSK1:gRNA3 was relatively low and that of LSK1:gRNA1 might be zero, although a greater number of clones needs to be analyzed by sequencing to confirm this. Although only a few mutations were generated, these resulted also to premature termination codons besides frameshifts, with the stop codons being the desired types of mutation for the generation of silenced plant lines. Based on the sequence of the gRNAs targeting the LSK1 gene, we also checked for possible off-target effects of Cas9. The off-target sites are usually dependent on the gRNA, since Cas9 can tolerate up to 3 mismatches between the gRNA and genomic sequence (Guo et al., 2023). With the use of the T7 assay no mutations were detected in the off-target sites, indicating that the designed gRNAs are not only efficient but also specific.

In the case of AMY2 a greater number of mutations were indicated from the T7 assay. The sequencing results also confirmed this, as mutations were identified in the regions corresponding to both guides (AMY2:gRNA4 & AMY2:gRNA5), showing that both of them were efficient in causing mutagenesis. As previously mentioned, we grouped the mutations based on the differences observed using the wild type as a reference. The most abundant mutations were small indels in the regions of the two gRNAs, which showed the efficiency of the system but would be useless in the following experiments since they occurred in intronic regions leaving the coding sequence unaffected. The excision of AMY2 exon 10 was also detected, with exon 10 being flanked by the introns targeted by the gRNAs. Exon 10 contains the active site of AMY2, so it was a desired target for the generation of a silenced line. The excision of the exon due to targeting of flanking sequences was also shown in other works (Mou et al., 2017). These results were promising not only for the efficiency of the system but most importantly about the prospect of generating the silenced lines. Apart from those, we were also able to detect interesting results. During the Q5 PCR we detected bands of a size greater than expected when amplifying a region of the AMY2 gene. Sequencing revealed that those extra nucleotides were the result of an insertion ranging from around 300-400bp that occurred in the region of AMY2:gRNA4. When attempting to annotate the sequence of the insertion, the results showed that in some samples it corresponded to the Ri plasmid of *A. rhizogenes* or in synthetic binary vectors. There is a possibility that a DSB occurred in the sequence targeted by the gRNA and that part of the Ri plasmid or the binary construct were integrated at the site, due to the functions of the GALLS proteins encoded by the Ri plasmid (Gelvin, 2021; Magori & Citovsky, 2011).

In conclusion, during this master thesis we assembled a series of binary vectors containing a CRISPR/Cas9 system optimized for *L. japoninus* and targeting various genes. These constructs were introduced to *A. rhizogenes*, and we assessed the efficiency of the gRNAs of inducing mutations in transgenic hairy roots. We evaluated three sets of gRNAs targeting the LSK1 and AMY2 genes and uncovered gene editing effect in both cases. However, due to the sparse number of analyzed samples, we were unable to detect all the putative mutations that might have taken place as suggested by the T7 assay. For the same reason it was also difficult to compute the actual mutation efficacy of the gRNAs.

In the upcoming months, more samples will be analyzed to answer these questions. That being said, since the functionality of the gRNAs was confirmed, we will proceed with the *A. tumefaciens*-mediated stable transformation to acquire the *amy2* and *lsk1* silenced mutant lines. In the future, the rest of the binary vectors will be assessed for the efficiency of their gRNAs targeting *AMY2, OSC3, LSK1* and *CYP71D353*. After confirming their functionally, we shall also proceed with the generation of stable mutant lines for those genes as well. Our aim is to utilize the CRISPR/Cas9-mediated gene edited seeds for functional analysis to better understand the interconnection between triterpene metabolism, symbioses and *LSK1*.

5. References

1. DE FARIA, S. M., LEWIS, G. P., SPRENT, J. I., & SUTHERLAND, J. M. (1989). Occurrence of nodulation in the Leguminosae. *New Phytologist*, *111*(4), 607–619. https://doi.org/10.1111/j.1469-8137.1989.tb02354.x
2. Delis, C., Krokida, A., Georgiou, S., Peña-Rodríguez, L. M., Kavroulakis, N., Ioannou, E., Roussis, V., Osbourn, A. E., & Papadopoulou, K. K. (2011). Role of lupeol synthase in Lotus japonicus nodule formation. *New Phytologist*, *189*(1), 335–346. https://doi.org/10.1111/j.1469-8137.2010.03463.x
3. Engler, C., Gruetzner, R., Kandzia, R., & Marillonnet, S. (2009). Golden gate shuffling: A one-pot DNA shuffling method based on type ils restriction enzymes. *PLoS ONE*, *4*(5). https://doi.org/10.1371/journal.pone.0005553
4. Engler, C., Kandzia, R., & Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE*, *3*(11). https://doi.org/10.1371/journal.pone.0003647
5. Engler, C., Youles, M., Gruetzner, R., Ehnert, T. M., Werner, S., Jones, J. D. G., Patron, N. J., & Marillonnet, S. (2014a). A Golden Gate modular cloning toolbox for plants. *ACS Synthetic Biology*, *3*(11), 839–843. https://doi.org/10.1021/sb4001504
6. Engler, C., Youles, M., Gruetzner, R., Ehnert, T. M., Werner, S., Jones, J. D. G., Patron, N. J., & Marillonnet, S. (2014b). A Golden Gate modular cloning toolbox for plants. *ACS Synthetic Biology*, *3*(11), 839–843. https://doi.org/10.1021/sb4001504
7. Farell, E. M., & Alexandre, G. (2012). Bovine serum albumin further enhances the effects of organic solvents on increased yield of polymerase chain reaction of GC-rich templates. *BMC Research Notes*, *5*(1), 1. https://doi.org/10.1186/1756-0500-5-257
8. Ferguson, B. J., Mens, C., Hastwell, A. H., Zhang, M., Su, H., Jones, C. H., Chu, X., & Gresshoff, P. M. (2019). Legume nodulation: The host controls the party. *Plant Cell and Environment*, *42*(1), 41–51. https://doi.org/10.1111/pce.13348
9. Garagounis, C., Tsikou, D., Plitsi, P. K., Psarrakou, I. S., Avramidou, M., Stedel, C., Anagnostou, M., Georgopoulou, M. E., & Papadopoulou, K. K. (2019a). Lotus SHAGGY-like kinase 1 is required to suppress nodulation in Lotus japonicus. *Plant Journal*, *98*(2), 228–242. https://doi.org/10.1111/tpj.14207
10. Garagounis, C., Tsikou, D., Plitsi, P. K., Psarrakou, I. S., Avramidou, M., Stedel, C., Anagnostou, M., Georgopoulou, M. E., & Papadopoulou, K. K. (2019b). Lotus SHAGGY-like kinase 1 is required to suppress nodulation in Lotus japonicus. *Plant Journal*, *98*(2), 228–242. https://doi.org/10.1111/tpj.14207
11. Gelvin, S. B. (2021). Plant dna repair and agrobacterium t−dna integration. *International Journal of Molecular Sciences*, *22*(16). https://doi.org/10.3390/ijms22168458
12. Guo, C., Ma, X., Gao, F., & Guo, Y. (2023). Off-target effects in CRISPR/Cas9 gene editing. *Frontiers in Bioengineering and Biotechnology*, *11*(March), 1–11. https://doi.org/10.3389/fbioe.2023.1143157
13. Iturbe-Ormaetxe, I., Haralampidis, K., Papadopoulou, K., & Osbourn, A. E. (2003). Molecular cloning and characterization of triterpene synthases from Medicago truncatula and Lotus japonicus. *Plant Molecular Biology*, *51*(5), 731–743. https://doi.org/10.1023/A:1022519709298
14. Jedličková, V., Mácová, K., Štefková, M., Butula, J., Staveníková, J., Sedláček, M., & Robert, H. S. (2022). Hairy root transformation system as a tool for CRISPR/Cas9-directed genome editing in oilseed rape (Brassica napus). *Frontiers in Plant Science*, *13*(August), 1–15. https://doi.org/10.3389/fpls.2022.919290
15. Khalil, A. M. (2020). The genome editing revolution: review. *Journal of Genetic*

Engineering and Biotechnology, *18*(1). https://doi.org/10.1186/s43141-020-00078-y
16. Kiryushkin, A. S., Ilina, E. L., Guseva, E. D., Pawlowski, K., & Demchenko, K. N. (2022). Hairy CRISPR: Genome editing in plants using hairy root transformation. In *Plants* (Vol. 11, Issue 1). MDPI. https://doi.org/10.3390/plants11010051
17. Koonin, E., & Makarova, K. (2019). Origins and evolution of CRISPR-Cas systems. *Royal Society Publising*, *2019*. https://doi.org/http://dx.doi.org/10.1098/rstb.2018.0087
18. Krokida, A., Delis, C., Geisler, K., Garagounis, C., Tsikou, D., Peña-Rodríguez, L. M., Katsarou, D., Field, B., Osbourn, A. E., & Papadopoulou, K. K. (2013). A metabolic gene cluster in Lotus japonicus discloses novel enzyme functions and products in triterpene biosynthesis. *New Phytologist*, *200*(3), 675–690. https://doi.org/10.1111/nph.12414
19. Küster, H. (2013). Medicago truncatula. In *Brenner's Encyclopedia of Genetics: Second Edition* (Vol. 4). Elsevier Inc. https://doi.org/10.1016/B978-0-12-374984-0.00915-3
20. Lee, S., Lee, C. Y., Lee, J., Seo, H.-H., & Hwang, K.-C. (2015). Cut and paste the genome: Genome editing for research and therapy. *Journal of Cellular Biotechnology*, *1*(1), 95–106. https://doi.org/10.3233/jcb-15009
21. Magori, S., & Citovsky, V. (2011). Epigenetic control of Agrobacterium T-DNA integration. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1809*(8), 388–394. https://doi.org/10.1016/j.bbagrm.2011.01.007
22. Márquez, A. J. (Antonio J. ., & Stougaard, J. (2005). *Lotus japonicus handbook*. Springer.
23. Mou, H., Smith, J. L., Peng, L., Yin, H., Moore, J., Zhang, X. O., Song, C. Q., Sheel, A., Wu, Q., Ozata, D. M., Li, Y., Anderson, D. G., Emerson, C. P., Sontheimer, E. J., Moore, M. J., Weng, Z., & Xue, W. (2017). CRISPR/Cas9-mediated genome editing induces exon skipping by alternative splicing or exon deletion. *Genome Biology*, *18*(1), 4–11. https://doi.org/10.1186/s13059-017-1237-8
24. Nguyen, D. Van, Hoang, T. T. H., Le, N. T., Tran, H. T., Nguyen, C. X., Moon, Y. H., Chu, H. H., & Do, P. T. (2022). An Efficient Hairy Root System for Validation of Plant Transformation Vector and CRISPR/Cas Construct Activities in Cucumber (Cucumis sativus L.). *Frontiers in Plant Science*, *12*(February). https://doi.org/10.3389/fpls.2021.770062
25. Open, F. T., Erb, M., & Kliebenstein, D. J. (2020). *Plant Secondary Metabolites as Defenses , Regulators , and Primary Metabolites : The Blurred*. *184*(September), 39–52. https://doi.org/10.1104/pp.20.00433
26. Phillips, D. R., Rasbery, J. M., Bartel, B., & Matsuda, S. P. (2006). Biosynthetic diversity in plant triterpene cyclization. *Current Opinion in Plant Biology*, *9*(3), 305–314. https://doi.org/10.1016/j.pbi.2006.03.004
27. Sahab, S., & Taylor, N. (2022). Studies on Pure Mlb® (Multiple Left Border) Technology and Its Impact on Vector Backbone Integration in Transgenic Cassava. *Frontiers in Plant Science*, *13*(February). https://doi.org/10.3389/fpls.2022.816323
28. Sawai, S., Shindo, T., Sato, S., Kaneko, T., Tabata, S., Ayabe, S. I., & Aoki, T. (2006). Functional and structural analysis of genes encoding oxidosqualene cyclases of Lotus japonicus. *Plant Science*, *170*(2), 247–257. https://doi.org/10.1016/j.plantsci.2005.08.027
29. Sawai, S., Shindo, T., Sato, S., Kaneko, T., Tabata, S., Ayabe, S. I., Aoki, T., Thimmappa, R., Geisler, K., Louveau, T., O'Maille, P., & Osbourn, A. (2014). Triterpene biosynthesis in plants. *Plant Science*, *170*(2), 247–257. https://doi.org/10.1146/annurev-arplant-050312-120229
30. Sentmanat, M. F., Peters, S. T., Florian, C. P., Connelly, J. P., & Pruett-Miller, S. M.

(2018). A Survey of Validation Strategies for CRISPR-Cas9 Editing. *Scientific Reports*, *8*(1), 1–8. https://doi.org/10.1038/s41598-018-19441-8

31. Stougaard, J. (2013). Lotus japonicus. In *Brenner's Encyclopedia of Genetics*. https://doi.org/10.1016/b978-0-12-374984-0.00883-4
32. Thimmappa, R., Geisler, K., Louveau, T., O'Maille, P., & Osbourn, A. (2014). Triterpene biosynthesis in plants. *Annual Review of Plant Biology*, *65*, 225–257. https://doi.org/10.1146/annurev-arplant-050312-120229
33. Wan, L., Wang, Z., Tang, M., Hong, D., Sun, Y., Ren, J., Zhang, N., & Zeng, H. (2021). Crispr-cas9 gene editing for fruit and vegetable crops: Strategies and prospects. *Horticulturae*, *7*(7), 1–19. https://doi.org/10.3390/horticulturae7070193
34. Wang, Q., Liu, J., & Zhu, H. (2018). Genetic and molecular mechanisms underlying symbiotic specificity in legume-rhizobium interactions. *Frontiers in Plant Science*, *9*(March), 1–8. https://doi.org/10.3389/fpls.2018.00313
35. Wendy, C., & Pederson, D. (2017). Mechanisms and conseqeunces of double-strand break formation in chromatin. *Physiology & Behavior*, *176*(5), 139–148. https://doi.org/10.1002/jcp.25048.Mechanisms
36. Werner, S., Engler, C., Weber, E., Gruetzner, R., & Marillonnet, S. (2012). Fast track assembly of multigene constructs using golden gate cloning and the MoClo system. *Bioengineered Bugs*, *3*(1), 38–43. https://doi.org/10.4161/bbug.3.1.18223
37. Xu, Y., & Li, Z. (2020). CRISPR-Cas systems: Overview, innovations and applications in human disease research and gene therapy. *Computational and Structural Biotechnology Journal*, *18*, 2401–2415. https://doi.org/10.1016/j.csbj.2020.08.031