Function of a microRNA gene containing an intron in *Solanum lycopersicum*

Afrah Mohammed Bawazeer

A thesis submitted for the degree of

Doctor of Philosophy

University of East Anglia

School of Biological Sciences

August 2022

“Science is like magic **BUT REAL”**

-Arthur C. Clarke

Abstract

MicroRNAs (miRNAs) are short single stranded RNA molecules. They are ~21 nucleotide long that are derived from long hairpin structures that result in post transcriptional gene silencing. MiRNA molecules are incorporated into a protein complex called RISC (RNA-induced silencing complex) and take this complex to target mRNA for degradation or inhibition of translation. It has been shown that miRNA molecules and other short RNA molecules intrinsically linked to the regulation of many genes that control the development, growth, and differentiation of plants.

In this study, we have characterised a novel tomato miRNAtop14. This miRNA is unusual as its primary transcript (pri-miRNA) contains a ~700nt intron, which is spliced out. MiRNA top 14 has been found to be conserved within the economically important *Solanaceae* family and among other agriculturally relevant members of the *Solanales* order, like in sweet potato, while its peculiar intron-split pri-miRNA structure is exclusively kept in the more closely related genera *Solanum, Capsicum and Nicotiana.* A mRNA cleaved by this miRNA was identified; the mRNA coding for LOW PHOSPHATE ROOT (LPR), a protein, which is involved in the arrest of root growth under phosphate starvation conditions in Arabidopsis. Interestingly, although LPR is widely conserved in plants, included in all the ones harbouring miRNAtop14, LPR cleavage was found to occur only in the three genera where the intron-split pri-miRNA structure is conserved.

The current study indicates that MIRs encoded by less canonical loci should be included in future miRNA searches, since they may be producing mature miRNAs with a function, as seen in this investigation. Furthermore, our results suggest that this

miRNA (top14) seems to be involved in plant growth and development as our experiments have indicated that the deletion of MIRtop14 seems to affect the root development. Moreover, as it is mentioned above, during the working on this project the experiments have found a target for top14 so called LOW PHOSPHATE ROOT1 (LPR1) in roots and the analyses by Northern blot have shown a high expression level of this miRNA in roots. Finally, the discovery of this miRNA and the study of it so far have opened up numerous avenues of possible investigation. Further study would indicate how this miRNA functions in the plant.

Table of Contents

List of Figures

Acknowledgements

First and foremost, I am extremely/ deeply grateful to my supervisor, Prof. Tamas Dalmay. Thank you for your always insightful comments and suggestions. Thank you for always considering my circumstances and otherwise discussing it with me in such a gentle way. Thank you for always being approachable and for being so understanding to all kind of troubles I have faced throughout these years. Your immense knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life.

I would like to thank the Saudi Arabian Embassy in London and Saudi Ministry of Higher Education for the studentship that allowed me to conduct this thesis,

Thank you as well to my secondary supervisor Dr. Simon Moxon for all the conversation and suggestions about the project.

I would like to offer my special thanks to Dr. Ben Miller for his invaluable advice, continuous support, and patience during my PhD study. Thank you for teaching on experimental procedures, especially on how to find a way out when everything seemed not to work.

Additionally, I would like to express my sincere gratitude to Dr. Martina Billmeier for her treasured support which was really influential in shaping my experiment methods and critiquing my results.

I would also like to extend my sincere thanks to Mrs. Maria-Elina Mannarelli for her technical support on my study.

I would like to thank all the members in Prof.Tamas Dalmay's lab; especially Dr Zahara Calzada and Dr. Rocky Payet. It is their kind help and support that have made my study and life in the UK a wonderful time.

I am very heartily thankful and grateful to my mother and father for their loving care, despite none of them being especially interested in science. Thank to my parents for their support, in particular economic, that has made possible for me to take such an uncertain career path. Without their tremendous understanding and encouragement in the past few years, it would be impossible for me to complete my study. I am deeply thankful to thank my sisters and brothers, whom without this would have not been possible.

Thank you, my friends, in Norwich for the evenings and trips that helped to lower the PhD stress. To the ones at home -Saudi Arabia - for our WhatsApp conversations and video calls especially during the Covid-19 pandemic time. You have really helped me to stay sane when things got rough. I really appreciate your friendship. I would like to give a special thanks here to my friend Noha, for taking such good care of me, while she is in Saudi, and for passing on to me her strength every time we talk.

Finally, I would like to express my gratitude and apologies to my daughters (Lara and Sidra) for being away from them even after the work hours to do the lab work and these last months of writing. Definitely, this PhD would not have been possible without their warm smiles. Thank you.

List of Abbreviations

A	Adenine
AGO	Argonaut
AS	Alternative splicing
APS	Ammonium Persulfate
Ata	*Arabidopsis thaliana*
ATP	Adenosine triphosphate
BAC	Bacterial artificial chromosome
BLAST	Basic Local Alignment Search Tool
BP	Base pairs
CaMV p35S, T35S	CaMV 35S promoter, CaMV 35 terminator
CBC	Cap-binding complex
CDC5	Cell division cycle 5
cDNA	Complementary DNA
CDS	Coding DNA sequence
circRNAs	Circular RNAs
CPL1	C-TERMINAL DOMAIN PHOSPHA T ASE-LIKE 1
CTR4	CTR1-like protein kinase
DCL-1	Dicer-like 1 RNAse III endonuclease
DDL	DAWDLE
DMSO	Dimethyl sulfoxide
DNA	Deoxy ribonucleic acid
DNAse	Deoxy ribonuclease
dNTPs	Deoxy ribonucleotides
DRB2	DOUBLE-STRANDED RNA BINDING 2
CRISPR- Cas9	Clustered Regularly Interspaced Short Palindromic Repeats
DSB	Double-strand break
DsRed	Discosoma sp. Red fluorescent protein
DTT	Dithiothreitol
E. coli	Escherichia coli
EtBr	Ethidium bromide

FW	Forward
G	Guanine
gRNA	Guide RNA
GRP7	GLYCINE-RICH RNA-BINDING PROTEIN 7
GUS	Beta-Glucuronidase
HCL	Hydrochloric Acid
HDR	High-fidelity homology directed repair
HEN1	HUA ENHANCER1
HESO1	HEN1 suppressor 1
hnRNP	Heterogeneous nuclear ribonucleoprotein particle
hnRNPs	Heterogeneous nuclear ribonucleoproteins
hpRNAs	Hairpin RNAs
HSP90	HEAT SHOCK PROTEIN 90
HST	HASTY
HYL1	HYPONASTIC LEAVES1
IIS	Type-II restriction endonucleases enzymes
Ini	*Ipomoea nil*
IPTG	Isopropyl -D-1-thiogalactopyranoside
Kan	Kanamycin
Kb	Kilobase
LA	LANCEOLATE
LacZ	*lacZ* β-galactosidase
LB	Luria-Bertani
LECA	last eukaryotes common ancestor
LIKE AMP1	LAM1
lncRNA	Long non-coding RNA
LPR	Low Phosphate Root
MCO	Multicopper oxidase
MID	Middle
Min	Minutes
MIR	miRNA gene
miRISC	miRNA guided RNA-induced silencing complex

miRNA	MicroRNA
MITE	Miniature inverted-repeat transposable element
MoClo	Modular Cloning
MOS2	MODIFIER OF SNC1, 2
mRNA	Messenger RNA
MS	Murashige - Skoog
Mut	Mutant
nat-miRNA	Natural antisense microRNAs
Nbe	*Nicotiana benthamiana*
NCBI	National Centre for Biotechnology Information
ncRNA	Non-coding RNA
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
NOS	Nopaline synthase promoter
NOT2	Negative on TATA less 2
nt	Nucleotides
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PAM	Protospacer Adjacent Motif
PAR	promoter-associated RNA
PARE	Parallel analysis of RNA ends
PAZ	Piwi Argonaute Zwille
P-bodies	mRNA-processing bodies
PCR	Polymerase chain reaction
PDR2	PHOSPHATE DEFICIENCY RESPONSE 2 Phased,
phasiRNA	Secondary, small interfering RNA
piRNA	Piwi-interacting RNA
pNOS, tNOS	NOS promoter, NOS terminator
Pol II	RNA polymerase II
PP2C	Protein phosphatase 2C
PPO1	Polyphenol oxidase 1
pre-miRNA	Precursor miRNA
pri-miRNA	Primary miRNA

PRL1	PLEIOTROPIC REGULATORY LOCUS 1
PTGS	Post-transcriptional gene silencing
RACE	Rapid amplification of cDNA ends
RCF3	Regulator of CBF gene expression 3
RIF	Rifampicin
RISC	RNA-induced silencing complex
RLM-RACE	RNA ligase-mediated
RNA	RACE Ribonucleic acid
RNAi	RNA interference
RNase	Ribonuclease
RNP	Ribonucleoprotein
ROS	Reactive oxygen species
rRNA	Ribosomal RNA
RRP6	Ribosomal RNA-Processing protein 6
RT	Room temperature
RT-PCR	Reverse transcription PCR
RV	Reverse
SCL6	Scarecrow-like protein 6
SDN	Small RNA degrading nucleases
SDS	Sodium dodecyl sulphate
SE	Serrate
Sec	Seconds
SGN	SOL Genomics Network
sgRNAs	Single guide RNAs
SIC	Sickle
siRNA	Small interfering RNA
Sly	*Solanum lycopersicum*
snRNA	Small nuclear RNA
snRNP	Small nuclear ribonucleoprotein particle
SPARE	Specific parallel amplification of RNA ends
SpCas9	*S. pyogenes* Cas9
Spp.	Subspecies
SR protein	Ser/Arg-rich protein

sRNA	Small RNA
SS	Splicing site
SSC	Saline-sodium citrate
STA1	STABILIZED 1
STOP1	SENSITIVE TO PROTON TOXICITY1
T	Thymine
TALENs	Transcription activator- like nucleases
tasiRNA	trans-acting siRNAs
TBE	Tris-borate-EDTA
TCP4	TCP family transcription factor 4
T-DNA	Transfer DNA
TE	Transposable element
TERC	Telomerase RNA component
TGH	TOUGH
tRNA	Transfer RNA
TSA	Transcriptome shotgun assembly
U	Uracil
URT1	UTP:RNA URIDYLYLTRANSFERASE
UTR	Untranslated region
UV	Ultraviolet light
VCS	Varicose
WT	Wild type
XIST	X inactive specific transcript
XRN4	Exribonuclease4
ZFNs	zinc-finger nucleases

Chapter 1

Introduction

1.1. Introduction

Micro RNAs (miRNA) are a class of non-coding, endogenous small (short) RNAs which act as gene expression regulators. miRNA abundance at the intracellular level is generally regulated under multiple control levels like transcriptional regulation, pri-miRNA and pre-miRNA processing and finally RNA-induced silencing complex assembly (RISC) (Wang et al. 2019). The first discovery of miRNA (lin-4) was reported in *Caenorhabditis elegans,* when it was observed that miRNA lin-4 was negatively regulating the lin-14 protein coding mRNA (Lee & Rosalind et al.1993).

miRNAs play an important role in many biological processes, such as development, differentiation, proliferation, growth, apoptosis, tumour formation, metabolism, stress response and disease development (Zhao et al. 2019). They function by targeting mRNAs and repressing their translation or by degrading them (Li et al. 2019).

1.2 miRNAs in Plants

1.2.1 Introduction

The high degree of overlap in the characteristics of the multiple non-coding RNA species has made their classification challenging. However, an arbitrary threshold of 200 nucleotides in length has been traditionally used to perform a first division between long noncoding RNA (lncRNAs) and the rest of shorter non-coding RNAs, which are generally called small RNAs (sRNAs) (Kapranov et al., 2007)

Within the sRNA group, the miRNA class is probably one of the best characterised since it was discovered before the others. The first report of a miRNA appeared in 1993 (although at that time it was called a short temporal RNA), when lin-

4 was reported, which is a small RNA that was discovered in *Caenorhabditis elegans* (Lee et al.,1993). It was observed that this small RNA regulated *C. elegans* developmental timing by annealing with lin-14 mRNA through antisense complementarity, which inhibited lin-14 translation (Lee et al.,1993, Wightman & Ruvkun,1993).

Seven years later, let-7, which is the second published miRNA was shown to repress the translation of a mRNA in *C. elegans* (Revkun, et al,.2020). That same year, let-7 was found to be conserved in bilaterian animals, including humans (Pasquinelli et al., 2000). Just one year later different publications demonstrated the existence of multiple miRNAs conserved across animals (Lagos-Quintana et al., 2001; Lee et al., 2001). Another year later, the presence and conservation of miRNAs in the plant kingdom was also confirmed in multiple reports (Llave, et al., 2002; Park, et al., 2002) and virus-encoded miRNAs were shortly added to the list (Pfeffer, et al., 2004)

From that point until now, great advances have been accomplished to understanding of miRNAs, its biogenesis, functions, and mode of action, while different aspects are still being discovered or require further research.

1.2.2. What is a miRNA?

miRNAs are defined as small RNAs of 20-25 nucleotides in length which are produced through the precise cleavage by RNase III enzymes of hairpin structures folded within longer, single-stranded transcripts called primary miRNAs (pri-miRNA) (Ambros, et al.,2004; Meyers, et al., 2008).

1.2.3. The Distinction Between siRNA and miRNA

microRNA (miRNAs) and short interfering RNA (siRNAs), are both molecules which regulate gene expression by suppressing the target gene (Brodersen and Voinnet, 2006). Since both siRNA and miRNA molecules have the same characteristic features of Dicer products (20–25 nucleotides length, 5′-phosphate, and 3′-hydroxyl), thus, it is difficult to distinguish between them. Therefore, miRNAs molecules can be distinguished from siRNAs molecules based on distinctive aspects of their biogenesis. (Ambros, et. al., 2003). The main difference between them is that microRNAs molecules are derived from long single-stranded endogenous transcripts that folds into a hairpin like secondary structure and is cleaved by DCL1, whereas siRNAs molecules are produced from long endogenous or exogenous dsRNA molecules (bimolecular duplexes or long hairpins) (Ambros, et. al., 2003) that are cleaved by DCL3 or DCL4 (Brodersen and Voinnet, 2006). siRNAs and microRNAs have similar function (Ambros et al., 2003) as a result, it is not easy to differentiate between the two unless the biogenesis pathway is revealed.

1.2.4. Biogenesis of miRNAs

1.2.4.1 Introduction

To understand the relation between the genomic organization of miRNAs to biogenesis and function, the following is a basic overview of what is known regarding these two processes. miRNA molecules are generated through a series of posttranscriptional biogenesis steps. It is now established that there are three forms of microRNA; primary-miRNA (pri-miRNA), precursor-miRNA (pre-miRNA) and mature microRNA (Lee et al., 2002). Pri-miRNA is around (~100-200), pre-miRNA is

around (~70nt) and the mature microRNA is around (20nt-25nt) long. Pri-miRNA is localised in nucleus, whereas pre-miRNA can be found in both nucleus and cytoplasm and the mature miRNA is found in the cytoplasm (Lee et al., 2002).

MIRNA genes are transcribed by RNA polymerases II enzyme complex (RNA Pol II) to generate the primary transcripts, which is referred to as (pri-miRNAs) (Lee et al., 2004a). The pri- miRNA is characterized by a hairpin RNA structure which is recognized and in animals it is cleaved by a second enzyme complex called Drosha, and its cofactor DGCR8, to generate a shorter hairpin of about 70nt length called precursor miRNAs (pre-miRNAs). Pre-miRNAs are exported from the nucleus to the cytoplasm via Exportin-5 to be finally cleaved by RNase III endonuclease Dicer to generate dsRNA miRNAs of 21-25 bp (Seitz and Zamore, 2006). One of the strands, the so-called miRNA∗ strand, is degraded while the other (mature miRNA) is incorporated into a protein complex called RISC (RNA- induced silencing complex). RISC without a miRNA is not functional but when it is loaded with a miRNA, the miRNA guides this complex to target mRNA that contain sequences complementary to the miRNA. Regulation takes place either by inhibition of translation of the mRNAs when complementarity between miRNA and mRNA is not so high or by mRNA degradation when there is a near perfect match between miRNA and its target mRNA (Guo et al., 2010). (Figure 1.)

1.2.4.2. *Pri-miRNA* Processing

MicroRNAs are produced by cleaving a long transcript called a pri-miRNA into a shorter transcript called a pre-miRNA, which is then cleaved once again to create the mature miRNA. In animals the enzyme complex Drosha performs the first cleavage step to form pre-miRNA, however plant genomes do not have a Drosha homolog

(Waititu et al. 2020). Instead, it has been shown that an enzyme complex known as Dicer-like performs both cleavage steps (Quévillon Huberdeau & Simard, 2019). There are four Dicer-like genes in A. thaliana; DCL1 which is involved in miRNA production, DCL2 which has a role in virus resistance, DCL3 has an important role in transcriptional silencing and DCL4 which produces 21 nt long RDR6 dependent tasi RNAS (Wang et al. 2019; Lee et al., 2004b; Dunoyer et al., 2005; Henderson et al., 2006). DCL1 has been shown to be vital in the development and growth of the plant as well as in the production of miRNAs since DCL1 mutants are embryonic lethal (Kurihara and Watanabe, 2004). It is also regulated in a negative feedback manner, by a miRNA, MİRI62 Okazaki et al. 2018). DCL1 forms a complex with HYL1 (which is a dsRNA binding protein) within the nucleus and in mutants with a reduced function HYL1, an increased accumulation of pri-miRNA and a decrease in the levels of pre-miRNAs and mature miRNAs are observed (Wang et al. 2019; Kurihara et al. 2006; Song et al., 2007). Therefore, it is thought that HYL1 is involved in the positioning of the cleavage site in the DCL1 pri-miRNA and pre-miRNA processing (Kurihara et al., 2006). Serrate was also shown to be involved in the regulation of pri-miRNA processing. It is found in the nucleus and is acts along with HYL1 and DCL1 as a general regulator of pri- to mature miRNA (Lobbes et al., 2006) As discussed earlier, miRNA genes are transcribed by RNA pol II and, these transcripts are capped therefore it was hypothesised that the capping proteins CBP20 and CBP80 could bind to these capped ends and play a role in the processing of miRNAs. However, it seems that CBP20 and CBP80 may have overlapping functions with Serrate and HYL1 (Kim et al., 2018)

1.2.4.3. *pre-miRNA* Processing

Animal and plant pri-miRNAs are processed into pre-miRNAs which contain an imperfect stem loop (or hairpin) structure. Many animal miRNAs are located in introns (Li & Yu, 2021) however, exons (either translated or untranslated) also contain pre-miRAs (Li & Yu, 2021). The mature miRNA is located in one of the arms of this secondary structure. The processing machinery recognizes the hairpin structure and staggered cleavages are produced on the dsRNA. Through the cuts the mature miRNA and miRNA star (miRNA*) are released. The miRNA* is the miRNA resulting from the other side of the hairpin, usually assumed to be non-active. The mature miRNA is loaded to RISC while the miRNA* is degraded. (Reinhart et al. , 2002) showed that in Arabidopsis DICER-LIKE1 (DCL1) which is a kind of RNAse type III enzyme produces all the miRNA precursor cuts.

A detailed mutagenesis study reveals that plant precursor processing starts with a next to loop cleavage processing. The precursors are then cut three more times in the intervals of 20 to 22 nucleotides in a direction following ‘loop to base’ (Hajieghrari & Farrokhi, 2021). This process continues until the mature miRNA is released.

1.2.4.4. Methylation and Nuclear Export

Once the miRNA transcript has been processed into mature miRNA by DCL-1 it is then methylated at the 3’ end and then exported to the cytoplasm. The miRNA complex is methylated by Hen1, which acts both on miRNA duplexes and siRNA duplexes (Wang et al. 2019; Yu et al., 2005). It methylates the ribose of the last nucleotide in the miRNA, this prevents uridylation, which is the addition of Uridine residues to the last nucleotide (Li et al., 2005). Uridylation would result in the degradation of the miRNA. There are many supposed benefits of methylation, this

modification could protect miRNAs from exonucleases attacking the 3’ end, or it could aid the recognition of plant Ago proteins or stop miRNAs acting as primers, especially as plant miRNAs are very highly complementary to their targets (Wang et al. 2019). Nevertheless, methylation seems to be a key modification of miRNAs to protect them from degradation by RNases (Okazaki et al. 2018).

In plants the mature miRNAs are significantly more abundant in the cytoplasm than in the nucleus, and it is likely that they function in the cytoplasm, therefore a transport mechanism is needed to transport mature miRNAs from the nucleus where they are produced, to the cytoplasm where they can cause cleavage of mRNAs. HASTY is a member of the importin ß family of nuclear cytoplasmic transport receptor (Figure 1.) (Lu et al., 2022). Exportin 5 is important in stabilising pre-miRNAs and transporting pre-miRNAs, tRNA, out of the nucleus and mutations of Exportin 5 result in increased pre-miRNA degradation. However, there is no conclusive evidence so far that HST is needed for miRNA production, and exp5 and HST seems to be functionally separate (Park et al., 2005). However, the hst1 mutant exhibited a reduced growth and an abnormally organised SAM (Shoot Apical Meristem) which are both factors that can be controlled by miRNAs (Bollman et al., 2003).

1.2.4.5. RISC Loading

Once the mature miRNA has been produced, processed and exported out of the nucleus, it is then loaded onto RISC. DRB1 seems to direct this loading step (Wang et al. 2019). The loaded RISC complex enables the miRNA of direct cleavage of mRNAs causing postranscriptional silencing. The loaded miRNA guides RISC cleavage of mRNAs at an exact position, between the nucleotides that are paired with the 10th and 11th nucleotides of the miRNA (Elbashir et al., 2001; Bartel, 2004).

It was unknown for a long period where the RISC assembled. Recent research (Bologan et al., 2018) revealed that EXPO1 exports RISC to the cytoplasm after it is assembled in the nucleus. However, the information available at this time does not rule out the potential that some miRNAs are exported in their duplex forms and assembled in the cytosol (Figure 1) (Wang et al. 2019). Among the double strands one strand of miRNA/miRNA duplex is assembled with Argonaute (AGO) protein. The other strand (passenger strand) is degraded as well as ejected (Wang et al. 2019). The gene silencing procedure is mediated by miRNAs using translation inhibition or target cleavage. RISC is guided by miRNAs to target genes by the help of base pairing (Quévillon Huberdeau & Simard, 2019).

The RISC assembly requires HSP90 (molecular chaperone) for activity. The process is facilitated through Cyclophilin 40/Squint, and this process is inhibited using Protein phosphatase 5 (PP5) (Okazaki et al. 2018). CYP40 has different domains. The TPR domain of CYP40 deals with variants of HSP 90 which contains the MEEVD sequence (C-terminal). The entire process of RISC loading is positively as well as negatively regulated by two Importin-Beta family of proteins. The two regulatory proteins are Transporting 1 (TRN 1) and Enhanced miRNA Activity 1 (EMA1) (Wang et al. 2019).

RISC is a ribonucleoprotein (RNP) complex which has an Argonaute (AGO) protein bound to an sRNA (Wang et al. 2019) as the core component. In the plant miRNA pathway, the sRNA is a mature miRNA and the AGO protein is AGO1 in most of the cases (Quévillon Huberdeau & Simard, 2019). AGO proteins harbour four characteristic domains: N-terminal (N) (Okazaki et al. 2018), Piwi Argonaute Zwille (PAZ) and middle (MID) (Wang et al. 2019), domains, which bind the 3’ and the 5’ end

of sRNAs respectively, and C-terminal PIWI domain (Quévillon Huberdeau & Simard, 2019), which has the slicer endonuclease activity.

In the most broadly accepted model of RISC assembly, the whole miRNA/miRNA* duplex is loaded onto the AGO protein with the help of HEAT SHOCK PROTEIN 90 (HSP90), to subsequently be dissociated and the guide strand (miRNA) retained while the passenger strand (miRNA*) discarded (Okazaki et al. 2018). However, examples in which the miRNA* is retained and leading RISC action have been reported (Quévillon Huberdeau & Simard, 2019).

The selection of the strand to be retained in the complex is determined by its lower 5' end thermostability (Quévillon Huberdeau & Simard, 2019). This selection is assisted, at least in AGO1- RISC, by the protein HYL1, which has been shown to dictate which strand is incorporated onto the complex, allegedly by loading directionally the miRNA/miRNA* duplex onto RISC120. The HYL1 regulator CPL1, has been shown to regulate this process as well (Wang et al. 2019).

Additionally, the identity of the nucleotide at the 5' end of the sRNA strand strongly influences its sorting into one or another member of the AGO family (Wang et al. 2019). AGO1 preferentially recruits sRNAs starting with a uridine (U) while AGO2 and AGO4 bind sRNAs with adenine (A) at the 5' end and AGO5 sRNAs with a 5' terminal cytosine (C) (Quévillon Huberdeau & Simard, 2019). Plant miRNAs usually have a 5' U, which is in accordance with AGO1 being the main AGO involved in the plant miRNA pathway (Quévillon Huberdeau & Simard, 2019).

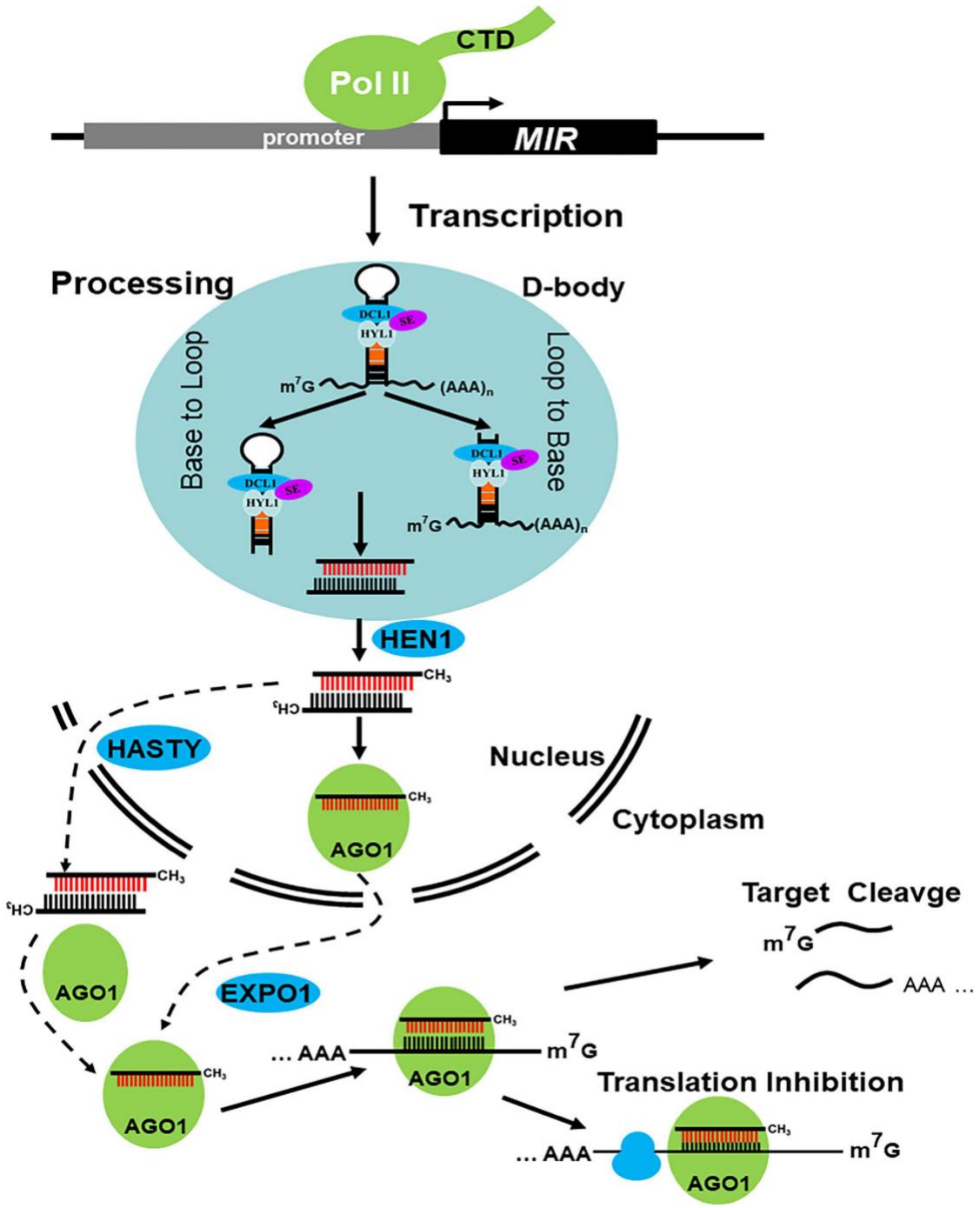


Figure 1. A diagram of miRNA biogenesis in plants.

MIR genes are initially transcribed by Pol II into pri-miRNAs that fold back to form hairpin structure. Splicing and further processing in nuclear involve the interactive functions of HYL1, DCL1 and SE. miRNAs and pre-miRNAs are sequentially processed by DCL1 to yield one or several phased miRNA/miRNA* duplexes, which are methylated by HEN1, which happens in nucleus. And then, the miRNA load into AGO1 protein and export via EXPO1 or the methylated duplex translocated from nucleus to cytoplasm via HASTY. The miRNA is selected and incorporated into dedicated AGO1-containing RISC that directs translation inhibition or cleavage of the target mRNA transcript. (Adopted from Lu et al., 2022)

1.2.4.6. RNA Silencing Pathways in plant

In eukaryotes, RNA silencing is a process that has evolved over time. In plants, the RNA silencing pathway has diversified to meet various functional needs (Samarfard et al. 2020). RNA silencing in plants can be divided into four functionally separate pathways based on the source of dsRNA and the functional target of sRNAs: microRNA (miRNA) pathway, trans-acting small interfering RNA (tasiRNA) pathway, RNA-directed DNA methylation pathway, and exogenic RNA silencing pathway. Plants have evolved multiple RNA silencing factors in tandem with the evolution of RNA silencing pathways. The model plant Arabidopsis, for example, has four DCLs, six RNA-dependent RNA polymerases (RDRs), and ten AGOs, along with several other factors (Guoet et al.2019).

1.2.4.7. RNA Silencing Pathway in Animal

The miRNA pathway is one of the most understood animal RNA silencing pathways (Reichholf et al. 2019). In the process of miRNA biogenesis, the enzyme complex Drosha performs the first cleavage step to form pre-miRNA, while RNase III dicer helps in cleaving pre-miRNAs and 21-23 nucleotide long mature miRNA is loaded to RISCs. In mammalian animals four AGO proteins are present which include AGO1-4. AGO proteins bind with siRNAs and miRNAs. AGO2 is capable of inducing endonucleolytic cleavage through perfect base pairing with the targets (Pong & Gullerova, 2018).

The collected studies from Drosophila, C. elegans and mammalian cells (Kim, et al. 2009) indicate a conserved animal mechanism that is analogous to, but distinct from, plant miRNA biogenesis. Most animal miRNAs are transcribed by RNA polymerase II, although a subset of animal miRNAs are products of RNA polymerase

III (Borchert et al. 2006). The major difference compared with plants is the segregated cleavage of miRNA precursors by nuclear and cytoplasmic RNase III enzymes. All animals use the Drosha RNase III enzyme, which partners with the double-stranded RNA-binding domain protein DGCR8 (known as Pasha in invertebrates), to liberate pre-miRNA hairpins from pri-miRNA transcripts. The lengths of pre-miRNAs are more consistent in animals than in plants, with most in the 55- to 70-nucleotide range, however, select Drosophila pre-miRNAs can approach 200 nucleotides (Ruby et al. 2007). Following nuclear export of pre-miRNAs by Exportin-5, they are cleaved into miRNA/miRNA* duplexes by cytoplasmic Dicer (a single enzyme in C. elegans and vertebrates, and Dicer-1 in Drosophila). These are loaded into miRNA effector Argonautes (Drosophila dAGO1, C. elegans ALG1/2, and vertebrate Ago1 to Ago4). Of the mammalian Ago proteins, only Ago2 has Slicer activity (Liu et al. 2004& Meister et al. 2004), Drosophila dAGO1 has Slicer activity, but appears to have poorer turnover than its paralog dAGO2, the major carrier of endogenous siRNAs (endo-siRNAs) (Okamura et al. 2004 & Forstemann et al.2007). Curiously, while plant miRNAs are universally methylated at their 3' ends by HEN1, most products of animal miRNA genes are not. An exception regards Drosophila miRNA* strands, which are preferentially loaded into dAGO2 All dAGO2 cargoes, including miRNA* strands, endo-siRNAs and exogenous siRNAs from viruses or artificial dsRNA, are methylated by HEN1 as single-stranded species (Czech et al. 2009 & Horwich et al. 2007). In addition to this core machinery, several accessory factors influence the biogenesis efficiency, fidelity and sorting of animal miRNAs (Kim et al. 2010). Notably, a growing number of these factors act in cell-specific or state-specific manners to regulate miRNA production or activity, indicating that neither process is constitutive.

1.2.5. Genomic Location of miRNAs

Knowing genomic location of miRNAs have helped in learning information about the biogenesis and function of specific miRNAs and miRNAs in general (Lu et al., 2005; Glazov et al., 2008). MiRNA genes can be located in several different kind of genomic regions: intergenic, intronic and exonic (Olena and Patton, 2010). It was reported that the majority of miRNAs are found between coding genes or within introns (Brodersen and Voinnet, 2006). Intergenic miRNAs are found between genes and can be a single microRNA or a cluster of microRNAs, they are separate from known transcription regions and have their own distinct promotors (Olena and Patton, 2010). Intronic miRNAs are located in the introns of annotated genes and share the promoter of their host gene. They can be either clustered or have just a single miRNA in the intron. Exonic miRNAs are the rarest of the three types, they often overlap with an exon and intron, and shares the promoter of the host gene (Olena and Patton, 2010). (Figure 2.)

The majority of plant miRNAs tend to be in an intergenic position, in contrast, animal miRNAs tend to be intronic (Merchan et al., 2009). Most plant microRNAs are singular, however, the majority of animal miRNAs are clustered. It has been reported that only 20% of plant miRNAs are clustered. The majority of reported plant miRNA clusters tend to be homologues copies, whereas in animals miRNAs in a cluster are usually unrelated (Merchan et al., 2009). (Figure. 2.)

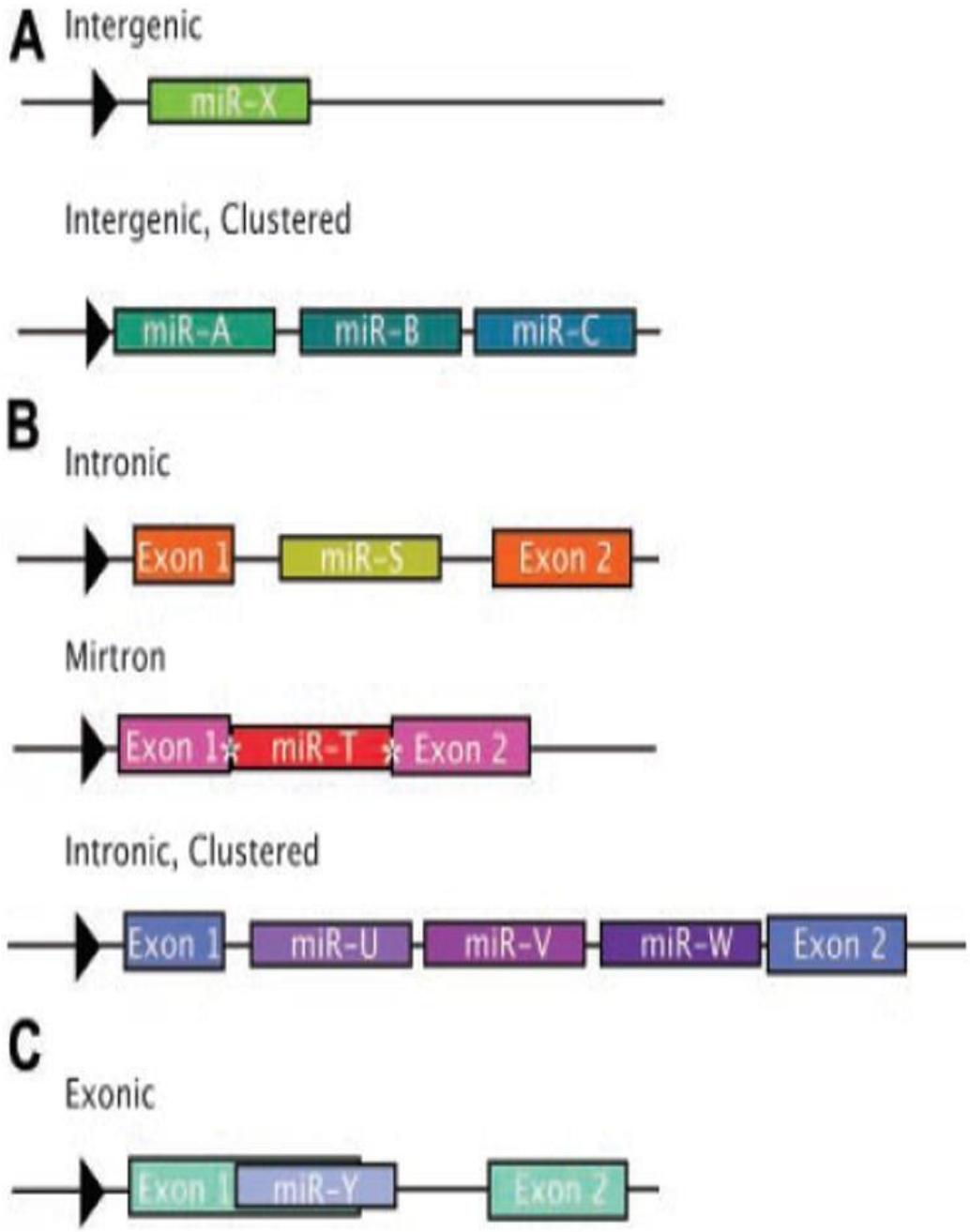


Figure 2. Genomic location of microRNAs (miRNA).
(A) Intergenic miRNAs are found in genomic regions distinct from known transcription units. These miRNAs can be monocistronic (top part) with their own promoters (black arrowhead) or polycistronic, where several miRNAs are transcribed as cluster of primary transcripts (bottom part); (B) intronic miRNAs are found in the introns of annotated genes, both protein coding and noncoding RNA; (C) exonic miRNAs are far more rare than either of the types above and often overlap an exon and an intron of a noncoding gene. (Taken from Olena and Patton, 2010)

1.2.6. MicroRNA Targets Identification

Because of several complete genome sequences are available, it would not be difficult to find targets of miRNAs if they were perfectly complementary to each other (Dalmay, 2008). The identification of genes targeted by a miRNA can be achieved by two main approaches: bioinformatics prediction or experimental identification. First of all, once the sequence of the miRNA is known and if the target organisms' genome is sequenced and annotated then as the target sequence will most probably be highly complementary to the miRNA computational programmes can identify possible targets. This is more likely to be successful in plants than in animals as plant miRNAs tend to have far greater homology to their targets than animals (Dalmay, 2008). Second, experimental approaches include over-expression or down-regulation of miRNAs, these will lead to a decrease or increase the observable effect of the miRNA and therefore prediction can be made as the possible location and role of the target gene. Identification and validation more targets will help in improving the prediction and clear understanding the biological role of miRNA molecules (Dalmay, 2008).

1.2.7. Role of Plant miRNAs in Development of Plants

Many studies have shown that microRNAs and other short RNAs intrinsically linked in the regulation of many genes that control the development, growth and differentiation of plants. The following are a few examples of miRNAs regulating developmentally important genes. In 2008, the discovery of 350 small RNAs in tomato fruit and 700 in tomato leaves has been reported, and with the exception of the conserved miRNAs, 90% of the small RNAs were found to be unique to tomato (Itaya et al., 2008). Therefore, the role of microRNAs in regulation of plant development is likely to be highly diverse and species specific.

The meristem is a group of cells where development of plants originates from, this group of undifferentiated cells will divide and differentiate to become the cells of the organs that make up the plant. During germination the root and shoot apical meristem (SAM) are formed very early, later additional meristems will form such as floral meristems and additional shoot meristems (Chuck et al., 2009). NAC stands for three genes which were discovered as they contain a particular domain, (the NAC domain): NAM (no apical meristem), ATAF1&2, and CUC2 (cup-shaped cotyledon) (Souer et al., 1996; Aida et al., 1997). The NAC genes considered as one of the largest families of plant-specific genes and have been found in a wide range of species. Using several complete plant genomic sequences helped to identify 117 NAC genes in Arabidopsis, 151 in rice, 79 in grape, 26 in citrus, 163 in poplar, and 152 each in soybean and tobacco (Rushton et al., 2008; Hu et al., 2010; Nuruzzaman et al., 2010, 2012a; Le et al., 2011)

It has been reported that the boundary domains around the organs at Arabidopsis floral and shoot apical meristems are regulated by miR164, its target genes are NAC domain family of transcription factors including CUC1(Cup Shaped Coteyledon 1) and CUC2 (Laufs et al., 2004; Nikovics et al., 2006).

There are three miR164 genes: miR164A, miR164B, and miR1640C. MiR164A is important in serration of the leaf margin, miR164C is controlling the production of petal number, and it is not clear what the specific role of miR164B is, as the mutant plant lacking miR164B does not show any modifications in aerial organs (Mallory et al, 2004a). Many abnormalities have been seen in the development of Arabidopsis, such as extra petals, misshapen leaves and missing sepals due to the expression of a miR164 resistant cuc1(Mallory et al., 2004a).

In addition, it was shown that miR127 is controlling the flowering time and the correct formation of floral organs in Arabidopsis. The target sequences of mi172 are a small subfamily called APETALA2. Part of this family is TOE1 and TOE2 which are floral repressors. The pri-miRNA is produced from EAT (Early Activation Tagged), as a result when EAT is over expressed, the plant shows an early flowering phenotype (Aukerman and Sakai, 2003).

MiRNA molecules are also involved in controlling the development of leaves. In Arabidopsis, there are two HD-ZIPIll transcription factor genes PHABULOSA (PHB) and PHAVOLUTA (PHV), both have complementary regions to miR165 and miR166. When the binding sites of miRNAs are mutated, the result/impact is disordered leaf polarity including abaxial (outer) to adaxial (inner) transformation of leaf cells (Mallory et al., 2004b).

The position of abaxial cells is found in the outer part of the leaf primodium and adaxial, the inner part, these go on to become the upper side and the lower side of the leaf, respectively (Mallory et al. 2004b). The expression of HD-ZIPIll genes is the reason behind this axial polarity (Juarez et al., 2004). In maize plant when the miRNA166 binding site is mutated, the mutant Rolled leaf1- Original (Rld1-O) has a disrupted miR166 complementary binding site. This causes in an upward curled leaf due to an abaxial/ adaxial partial reversal of leaf polarity (Juarez et al., 2004).

As a result of these studies, the post transcriptional regulation of genes by microRNAs is vital for the correct development of all plant organs.

1.3. Splicing in Plant

1.3.1 Introduction

Splicing of mRNA and interrupted genes were first discovered in adenovirus in 1977 (Berget, et al., 1977; Chow, et al., 1977) and at the same year, the phenomenon was confirmed in eukaryotes (Brack, et al., 1977; Breathnach, et al., 1977). One year later, the possibility that alternative splicing of introns and exons could create different mRNA variants of the same gene was suggested (Gilbert, 1978), which was soon confirmed experimentally (Alt, et al., 1980; Early, et al., 1980).

Over the next years, a lot of effort was made to discover the mechanisms and the machinery of mRNA splicing. The machinery involved in the process for splicing (Brody & Abelson, 1985; Frendewey & Keller, 1985) as well as the chemical reactions necessary were clarified (Cech, 1983).

When the study of alternative splicing at genomic level became possible, the prevalence of the phenomenon (Wang & Brendel, 2006; Genomics et al., 2006) and its key role in plant development and environmental response (Staiger & Brown, 2013; Filichin et al., 2015) was discovered.

The differences between mRNA variants created by alternative splicing do not only produce protein isoforms that may have different characteristics, but also can influence the transport, localization, stability and translation of the transcript. In addition, the association of the splicing mechanism with other RNA processes such as miRNA regulatory pathways is also contributory to the whole regulation of gene expression that can be carried out through alternative splicing (Filichin et al., 2015; Reddy et al., 2013).

The importance of splicing process for eukaryotic evolution and interrupted genes is becoming clearer. It makes exon shuffling possible, which is a mechanism

by which new genes are created by the rearrangement of exons (Gilbert, 1987; Patty, 1999). Furthermore, it can lead to alternative splicing, which is a phenomenon that expands proteome diversity (Reddy et al., 2013; Nilsen & Graveley, 2010). In fact, it has been observed that alternative splicing plays a role in organogenesis and speciation by creating different patterns of gene expression (Barbosa-Morais et al., 2012; Merkin et al., 2012).

1.3.2. What is Splicing?

RNA splicing is a modification of the precursor messenger RNA (pre-mRNA) transcript into a mature messenger RNA (mRNA) prior the translation process. Splicing is a process by which introns are removed from a pre-mRNA and exons are joined to produce a mature messenger RNA that contains only exons (Kornblihtt, et. al., 2013).

1.3.3. Types of Intron

At least four main types of introns have been identified: Transfer RNA introns, Group I introns, Group II Introns, and Spliceosomal introns.

1) tRNA introns, which appear in archaea and tRNA genes in eukaryotes. They are removed by a cut and ligate enzymatic mechanism that makes them different from all the other types of introns (Greer et al, 1983; Rodriguez-Trelles et al, 2006).

2) Group I introns, which are found in bacteria and eukaryotic organelles, as well as in nuclear rRNA genes in fungi and protists. They catalyse their own splicing through two transesterification reactions (Rodriguez-Trelles et al, 2006; Cech, 1990).

3) Group II introns, which have been found in bacteria and in the organelles of plants, fungi and protists. They self-splice through two transesterifications as group I introns,

but the mechanism is different between both (Rodriguez-Trelles et al, 2006; Michel, et al, 1982; Michel, et al, 1989).

4) Spliceosomal introns, which are the introns present in eukaryotes nuclear genomes. Unlike all the other types of introns, they are spliced by the spliceosome complex, but the splicing mechanism is the same as for group II introns (Rodriguez-Trelles et al, 2006; Padgetr et al, 1986). The focus in the following parts is on spliceosomal introns splicing.

1.3.4. Intron-Split miRNAs Splicing in plants

Plant MIRs often contain introns, it was reported that 67% of Arabidopsis pri-miRNAs have at least one intron (Stepien, et al., 2016). Both plant pri-miRNAs and their introns show a large variability in length, with sizes from 300bp to 5000bp for pri-miRNAs, and from 100bp to 3000bp for their introns (Stepien, et al., 2016; Zhang, et al., 2009). miRNA hairpins are situated in the first exon in most cases (Stepien, A. et al. 2016; Szweykowska-Kulińska et al., 2013), despite the fact that they may be in other exons and additionally in alternatively spliced regions that might be either intronic or exonic (Yan, et. al, 2012; Jia and Rock, 2013).

Another fascinating probability is the one in which an intron is in between the miRNA and the miRNA* sequences, dividing the miRNA stem-loop (Sunkar, et. al, 2005; Lu, et.al, 2008). This exon-intron-exon structure was first seen in a member of the miRNA444 family in rice (Sunkar, et. al, 2005). In a later report, also in rice, it was discovered that most members of the miRNA444 family had a feature that was not seen before in any other miRNAs. They were transcribed from the antisense strand of their own target; therefore, they were called natural antisense miRNAs (nat-miRNAs) (Lu, et.al, 2008). Besides, it may be interesting to note that, all miRNA444 family members were found to have an intron between their miRNA and miRNA*, even those

two (miRNA444e and miRNA444f), which were not transcribed from the antisense strand of their target gene (Lu, et.al, 2008).

Subsequently, this special exon-intron-exon arrangement was additionally revealed in miRNA444 members of maize and sorghum (Zhang, et al., 2009; Paterson, et al., 2009; Thieme, et al., 2011). Moreover, SplamiR, which is a bioinformatic tool, was created to recognize this specific sort of pri-miRNAs with an intron amidst the miRNA hairpin, as they were not predicted by existing bioinformatic tools, unless spliced (Thieme, et al., 2011). These pri-miRNAs have been called intron-split miRNAs (Pek, et al., 2015).

1.4. CRISPR- Cas9 technology (Clustered Regularly Interspaced Short Palindromic Repeats)

1.4.1. Introduction

Genome editing is a technique that changes the DNA of an organism. Using these technologies genetic materials can be added, altered, or removed in particular genome locations. One of the popular technologies of genome editing is CRISPR-Cas9 (Zhan et al. 2019). In case of infection with a virus, generally small pieces of DNA are captured by the bacteria. This complex is inserted into the own DNA of bacteria following a specific pattern. These segments are called CRISPR arrays. The presence of a CRISPR array helps the bacteria to remember viruses later (Medline, 2022).

Bacterial cells use the CRISPR-Cas system for protecting themselves from any kind of new infection. As already indicated, the CRISPR-Cas systems enables the bacteria cells to memorize the infection that has happened before (Dimitriu et al.

2022). As soon as the phage attaches with a bacterial cell and injects the DNA within the cell, a small sequence is inserted with the CRISPR sequence of the bacterial DNA. (Figure 3.). The variable CRISPR sequence sections thus prepare a library of all the encountered pathogens of bacteria (Tyumentseva et al. 2021). The CRISPR sequence library is transferable, and the bacterium can pass the pathogenic memory to its offspring as well. More than half of the known bacteria have CRISPR-Cas as a defence system (Klompe and Sternberg, 2018). The two popular CRISPR-Cas classes are Class 1 containing protein complexes and class 2 system deals with one cutting protein (Wangn and Cui, 2020).

In case the same viral infection happens to the bacteria then RNA segments are produced from the CRISP arrays which help in attaching and recognising specific regions of the viral DNA. As soon as the identification process is done, Cas9 or any similar enzyme cuts the viral DNA apart (Najafi et al. 2022). With the cleavages in the DNA, the virus is disabled. This bacterial system can be recreated in eukaryotes to generate targeted mutations in the genome. A small RNA piece fused with a guide sequence is expressed in cells. These attach to the target sequence of the DNA.. These segments are similar to the RNA segments produced by the bacteria from the CRISPR array (Harrington et al. 2020). The target sequence is cleaved by the help of Cas9 enzyme. Other than Cas9, Cpf1 enzyme is also used. As soon as the DNA is cleaved, the DNA repair machinery of the cell joins the cleaved ends but during this process it adds and/or deletes a few nucleotides around the cleavage site (Hussain et al. 2019).

In the CRISPR dependent gene silencing process the number of off targets are less than in RNAi (Salanga & Salanga, 2021). The risk of associated immune

responses due to CRISPR is low. The spacer sequences of CRISPR create RNA sequences of short length. These sequences also work as the guide for matching the DNA sequences (Javed et al. 2018). As already discussed in the above sections, the CRISPR system produces different types of enzymes and Cas9 is one of those enzymes that bind to the targeted DNA sequence and cuts it.

1.4.2. What is CRISPR system?

CRISPR arrays are regions that are found in the genomes of bacterial "immune system" which help defend against invading viruses. These regions are composed of repeated DNA sequences and spacers (Figure 3.). Often bacteria can be infected by viruses therefore bacteria are using CRISPR immune system to protect against these attacks by degrading the genome of the invading virus (Bolotin et al., 2005, Jinek et al., 2012).

1.4.3. How does CRISPR-Cas9 System Work?

The mechanism of action of the CRISPR–Cas system, to protect bacteria from repeated viral attack, works through three main stages: Adaptation stage, Expression/Maturation stage, and Targeting stage (Barrangou, R. and Marraffini, L. 2014).

In the first stage which is known as adaptation or 'spacer acquisition', sequences from the viral genome are inserted into the CRISPR array of the bacterial genome (Barrangous et al., 2007). It is this spacer sequence that is integrated into the CRISPR array to form the immunological memory (Jiang et al.,2017). During the second stage, where immunity is executed, which is called expression and maturation

stage, during CRISPR RNAs (crRNA) biogenesis phase, the CRISPR array is transcribed into a precursor RNA transcript that is further processed to generate smaller units of RNA known as CRISPR RNAs (crRNAs) or guide RNAs (gRNAs), each one of these crRNAs containing a single spacer flanked by a part of the repeat sequence. These crRNAs are combined with one or more Cas proteins to form the active Cas–crRNA complex (Amitai, et al., 2016, Marraffini, 2015, Jiang et al.,2017). Third, in the targeting stage the Cas–crRNA complex searches the cell for nucleic acid targets that are recognized by base-pairing with complementary crRNA sequences. The spacer sequences in these RNAs are used as guides to direct the cleavage of the viral genome by the Cas endonucleases. Successful recognition leads to the cleavage and degradation of the target DNA (Garneau et al., 2010; Gasiunas et al., 2012; Jinek et al., 2012). (Figure 4.)

1.4.4. Advantages of CRISPR/Cas9 over ZFN and TALENs

ZFN and TALENs are other methods to introduce mutations to specific positions in the genome (reference). However, the CRISPR/Cas9 system has several advantages over ZFN and TALENs in terms of its simplicity, flexibility, and affordability. There are several differences between CRISPR/Cas9, ZFN and TALENs. The most important difference is that the CRISPR system relies on RNA-DNA recognition, rather than on protein-DNA Binding mechanism (Jansen, et al., 2002, Jinek, et al., 2012, B. Liu, et al., 2015). Thus, it is more workable and easier to construct a customized CRISPR/Cas9 complex by only changing the gRNA sequence instead of engineering a new protein. The target sequence needs to be immediately upstream of a PAM sequence (5′-NGG-3′) (Jinek, et al., 2012) because the latter is essential for target recognition by Cas9.

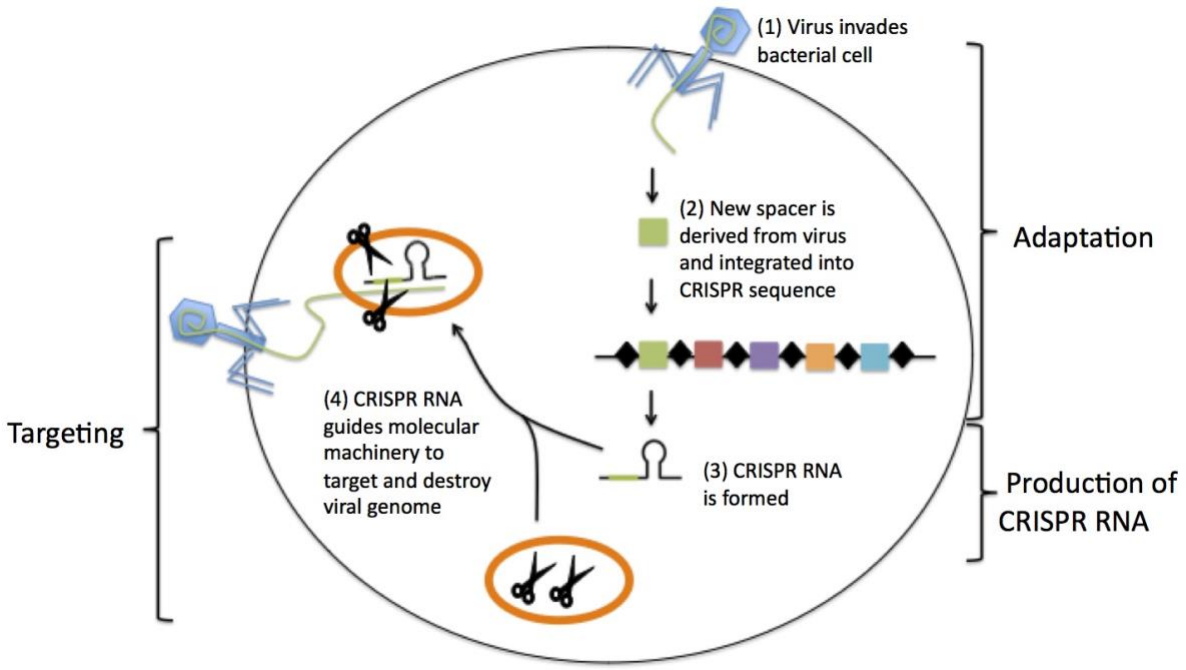


Figure 3. How does CRISPR work? The Three stages of CRISPR-immunity system: Adaptation, Production, and Targeting. CRISPRs are regions in the bacterial genome. These regions contain clusters repeats of DNA (black diamonds) and spacers (colored boxes). During adaption stage, new spacers (DNA fragments) obtain from the invading virus genome are inserted into the CRISPR region by an unknown mechanism. In production of crRNA, the CRISPR sequence is transcribed by Cas to generate small crRNAs. During targeting process, crRNA guides Cas9 (molecular machinery) to DNA sequences of the invading virus. The match between the crRNA and target sequences determine the nucleolytic cleavage of invading virus DNA as a result degrade it. Figure adapted from (Barrangou and Marraffini, L. 2014).

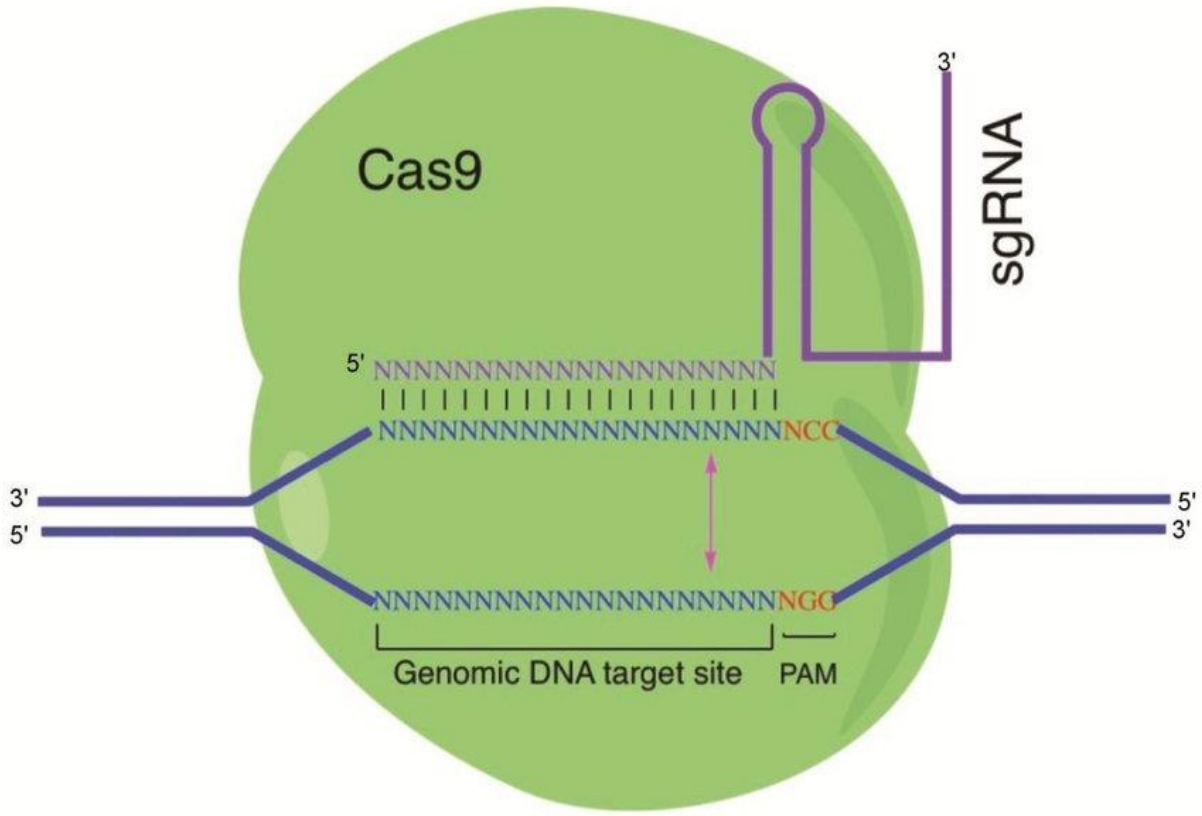


Figure 4. The structure of CRISPR/Cas9 system lustration of the CRISPR/Cas9 system. The Cas9 protein interacts with the single- guide RNA (sgRNA) to direct endonuclease activity proximal to the protospacer adjacent motif (PAM)sequence. Custom-designed sgRNAs recognize their target sequence and allow Cas9 endonuclease to cleave the sense strand 3 base pairs (bp) and antisense strand 3 bp upstream of the PAM sequence (NGG). Binding of sgRNAs to the target sites induces Cas9 endonuclease to create a double-strand break on the genomic target. Figure adapted from (Chuan et. al., 2018).

1.4.5. Editing or silencing Genes with CRISP-Cas9

Instead of relying on bacteria to provide crRNAs or gRNAs, researchers design and produce small RNA molecules that match the DNA region of interest, to direct the Cas9 enzyme to snip the targeted DNA (Barrangous et al., 2007).

The CRISPR-Cas9 technology consists of two main key molecules, which make a mutation into the DNA molecules. These key molecules are an enzyme called Cas9 and a piece of RNA called guide RNA (gRNA). The Cas9 enzyme is working as molecular scissors, which has the ability to cut the double strands of DNA at a specific position in the genome. As a result of this cut, that location of the DNA is edited, or a piece of DNA is removed if two gRNAs are deployed. In terms of the gRNA, it consists of an approximately 20nt long pre- designed RNA sequence, which is located in a longer RNA scaffold. The scaffold part attaches to DNA sequence and the gRNA guides the Cas9 enzyme to the specific region of the genome. As a result, the Cas9 enzyme cuts at the right location in the genome. Since the guide RNA (gRNA) is designed to bind to a specific sequence in the DNA region of interest, it has RNA sequence that is complementary to the target DNA sequence in the genome which means that the gRNA would only bind to the target sequence and no other sequence of the genome. The Cas9 enzyme follows the gRNA to the target location in the DNA sequence and causes a cut across both strands of the DNA. However, the editing of the specific sequences depends on the type of repair strategy being used by a cell, such as nonhomologous end joining (NHEJ) or homologous directed repair (HDR), as presented below in detail and summarized in figure 5. (Hussain et al., 2018)

1.4.5.1 Nonhomologous End Joining (NHEJ)

In order to achieve site-specific genomic engineering, the CRISPR-associated enzyme Cas9 causes a double-strand break (DSB) at the position designated by the guide RNA (Xi-Dian, et al., 2019). Cells repair the DSB using the NHEJ or HDR pathway. The low fidelity of NHEJ, which is prone to errors, may result in a base deletion or insertion (indel) after repair, resulting a mutation (Bernheim et al., 2017).

The NHEJ pathways are further classified into traditional and alternative NHEJ pathways. But because NHEJ is prone to errors, the pathway's final product frequently has extra or missing DNA sequences, which could lead to a non-functional coding sequence (Pannunzio et al., 2018). NHEJ is the predominant DSB repair pathway and is responsible for most DSB repairs throughout the cell cycle (Arnoult et al., 2017). To attach to DNA termini, NHEJ needs Ku. This increases the variety of NHEJ's enzymatic components, which include a nuclease, a ligase, and two polymerases (Mateos-Gomez et al., 2017). Intriguingly, each of these enzymatic components is unique for its capability in working on a broad range of incompatible DNA ends coupled with flexibility in loading order, leading to several possible junctional consequences from one DSB. Direct ligation of DNA end configurations is possible. If these ends cannot be joined, they may be processed further until ligatable configurations are obtained, which are typically stabilized by little more than 4 bp of terminal micro-homology. The processing of DNA ends results in the addition or deletion of nucleotides, accounting for the fact that original DNA sequences are rarely restored after NHEJ repair of DSBs (Pannunzio et al., 2018).

The following factors are required for NHEJ repair regardless of end structure and dictate the major events of the pathway. Broken ends are recognized by loading

of the Ku70/Ku80 heterodimer. Ku then acts as a scaffold for recruitment of a kinase (DNA-PKcs) and a two subunit DNA ligase (XRCC4-ligase IV), together with some other factors (PAXX, XLF), this complex holds a pair of DNA ends together, forming a paired end complex.The paired end complex then ligates compatible DNA ends together, thus repairing the break (Arnoult et al., 2017).

This is a simplified, streamlined version of this pathway and does not consider the missing or damaged nucleotides that are common to biological sources of DSBs, and which need to be processed. Processing occurs prior to ligation as incompatible DNA ends interfere with that step. Accordingly, NHEJ has a vast toolbox of processing factors, including polymerases (Pol μ and Pol λ), nucleases (Artemis), and structure-specific end cleaning enzymes (Aprataxin, Tdp2) that function to make ends better substrates for ligation (Xi-dian, et al., 2019).

The method used to fix a DSB is not chosen at random, this has an impact on the outcome of genome editing. Compared to NHEJ, HDR is a less prevalent type of DSB repair, but when used properly, it can significantly affect targeted genome editing. However, in undivided cells, which is the majority of cells in vivo, the HDR pathway is limited. Thus, HDR-mediated genome editing methods are limited to *in vivo* applications (Xi-dian, et al., 2019).

1.4.5.2. Homology-Directed Repair (HDR)

The HDR pathway, which has a high fidelity but a low incidence, is the second DSB repair process. The targeted nuclease cleaves the DNA under the guidance of an external repair template. This can increase the probability of homologous recombination (HR) by about 1,000-fold. Notably, HDR is able to precisely modify the genome using a variety of methods, including conditional gene knockout, gene knock-in, gene substitution, and point mutations (Arnoult et al., 2017). Accurate insertions, base substitutions between DSB sites or two DSBs, and other modifications are produced by the HDR pathway using homologous donor DNA sequences from sister chromatids or foreign DNA. To get the intended result, genomic engineering requires this kind of precise modification. Sister chromatids are only available in the S and G2 phase, therefore, HDR is limited to these phases of the cell cycle (Salsman et al., 2017).

1.4.5.2.1. Homology directed repair pathways

HDR mechanisms can be divided into two main camps: non-conservative and conservative. The conservative methods are characterized by use of a homologous donor (e.g., sister chromatid, plasmid, etc.), and consists of three pathways: classical double-strand break repair (DSBR), synthesis-dependent strand-annealing (SDSA), and break-induced repair (BIR). The non-conservative method is single strand annealing (SSA), an error prone mechanism that notably does not require a donor template (Salsman et al., 2017).

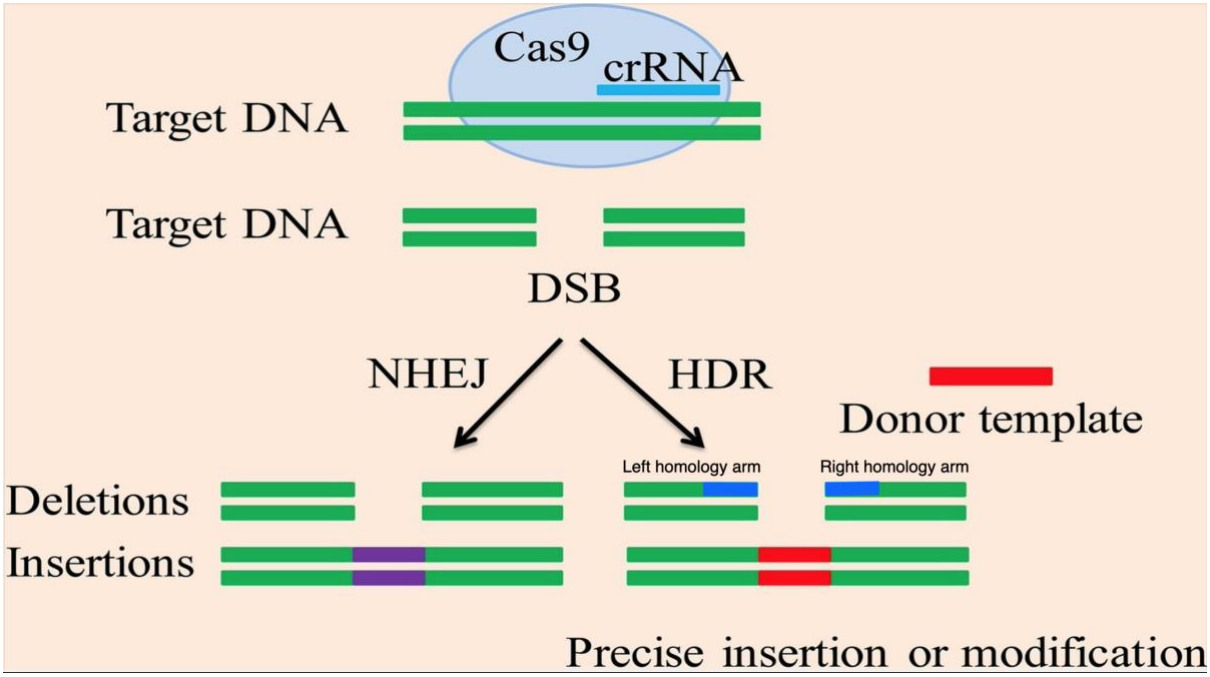


Figure 5. CRISPR/Cas9-mediated DSB repair mechanism. The CRISPR-associated enzyme Cas9 breaks down the target DNA to create a DSB, the two repeated sequences are further used as templates to produce short crRNAs. Methods for DSB repair include the NHEJ and HDR pathway. The NHEJ pathway creates accurate deletions and insertions. The HDR pathway uses homologous donor DNA sequences from sister chromatids or foreign DNA to create accurate insertions, base substitutions between DSB sites or two DSBs, and other modifications. Figure adapted from. (Xi-dian, et al., 2019)

1.4.6. CRISPR/Cas9 Mediated Gene Editing in Tomato

Tomato is an effective genomic model to check the feasibility of different genetic mechanisms. Tomato plant has 500 genotype rich genomic sequences, mutant populations and transcriptome data. All of these resources help in testing the process of CRISPR/Cas9 editing in the tomato species. The designing of the constructs for editing essential genes in the tomato is completely based on the vector system (Reem & Eck, 2019). Thus, the CRISPR/Cas9 Formation in the tomato plant is enhanced by the application of the golden gate cloning system. The coning system contains the NPTII (neomycin phosphotransferase II) cassetteas a marker gene that leads to resistance to Kanamycin (Reem & Eck, 2019). In case of Arabidopsis, U6 promoter driven guide RNA and Cas9 are the two important factors for gene editing.

In order to edit a genome successfully, selection and design of guide RNA are some crucial steps. The factor that acts as the major constraint in this process is PAM (protospacer adjacent motif). It is placed in such a site that can be denoted as 5’-NGG-3’ (Reem & Eck, 2019). In order to modify the expression level of the gene, the gene promoter region has to be targeted. The targeting procedure can be executed with the help of a number of gRNAs. Using the CRISPR-Direct and the CRISPR-P web tools gRNAs can be easily designed (Reem & Eck, 2019).

Chapter 2

Material and Methods

2.1. General material and methods

2.1.1. Plant materials and growth conditions

All plants were grown at 22°C and 16h light / 8h dark in a growth room. The species and cultivars / ecotypes used in this study were: *Solanum lycopersicum* cv. MicroTom and *Nicotiana benthamiana* (*Nbe*).

2.1.1.1. Growth of tomato seedlings on plates

Tomato seeds were sterilised by vortex for 2 minutes in 70 % ethanol. After removal of ethanol the seeds were rinsed once with sterile distilled water. It was followed by shaking with sodium hypochlorite solution for half hour. The seeds were rinsed 3 times with sterile distilled water. Seeds were placed on Murashige - Skoog (MS) plates and left to grow for 1 week under long day conditions (16-hours light and 8-hours dark).

2.1.1.2. Growth of *Nicotiana benthamiana* on plates

Nbe seeds were sterilised by sterilisation solution, which was 1 ml of NaOCI, 5 µl of Triton 100 times and 4ml of sterile distilled water. Seeds were put in an Eppendorf tube and 1 ml of the sterilisation solution was added. After 10 minutes they were centrifuged for 10 seconds, and the liquid was taken off. The liquid was replaced with sterile distilled water. These steps were repeated five times. The seeds were left in water and then distribute.

Seeds were placed on MS plates and left to grow for 1 week under long day conditions (16-hours light and 8-hours dark).

2.2. Total RNA extraction

RNA extraction was performed with Tri - reagent following the manufacturers protocol (Ambion). In brief, tissue was frozen, grinded in liquid nitrogen with 1 - 2 ml of Tri – reagent solution is added to each 100 mg of tissue and vortexed. The ground tissue was incubated at room temperature for 5 minutes. 200 µl chloroform was added to each sample, mixed by vortex and incubated at room temperature for 8 minutes. Next the samples were centrifuged in a cold room at 14000 rcf for 15 minutes. Finally, the upper aqueous phase containing the RNA was moved to another tube, 500 µl of Isopropanol was added and vortexed. Samples were left over night in -80 °C to precipitate. Next day, the samples were left to thaw and centrifuged in the cold room at 14,000 rcf for 8 minutes. White pellet was formed at the bottom of the tube which contained the RNA. Samples were washed two times with 950 µl 80% ethanol and air dried before dissolved in distilled water. The concentration and quality were measured using a NanoDrop spectrophotometer (Thermo Fisher scientific) at an absorbance ratio of A260 / 280 and A260 / 230 nm.

2.3. DNA Extraction

Total DNA was extracted from *Solanum lycopersicon* cv. (MicroTom) and *Nbe* plants. After the harvesting, the plant tissue was immediately frozen in liquid nitrogen. The frozen tissue was either immediately ground up to a fine powder by a pestle under liquid nitrogen or stored at -80 °C. Ten % volume of Extraction buffer (200 mM Tris-HCl (pH 7.5) 250 mM NaCl, 25 mM EDT and 0.5% SDS) was added to ~30 mg of plant tissue and mix for 5 second. The samples were incubated at 60°C for 30 min. 500 µl of chloroform:iso-amyl alcohol (24:1) was added, mixed and centrifuged at 15,000 g for 5 min at 4°C. The supernatant was transferred to a new Eppendorf tube and added an equal volume of Isopropanol. The samples were mixed and incubated

at -20°C 30 min. The samples were centrifuged at 15,000 × g for 5 minutes at RT. The supernatant was discarded, and the pellet was left to air dry then dissolved in 100 µl TE buffer (10 mM Tris, bring to pH 8.0 with HCl and 1 mM EDTA. Finally, for precipitation of starch and other insoluble polysaccharides samples were incubated on ice for 5 min. and centrifuged at 15000 g for 2 min. At the end supernatant was transferred containing the DNA into new tubes. The DNA extraction was performed as described in the protocol (Genomic DNA extraction from leaves) (Amani et al., 2011)

2.4. DNA and RNA quantification

DNA and RNA were quantified using a NanoDrop Spectrophotometer (Thermo Scientific). The ratio absorbance between 260: 280 of 1.5 µl solution was used to confirm whether the DNA and RNA were acceptable quality.

2.5. Polymerase Chain Reaction (PCR)

Two different polymerases were used depending on what the PCR product would be used for. The basic amplification just for visualising on agarose gel, the Go-Taq was used, and a PCR reaction was used for Restriction-ligations to assemble fragments by Golden Gate cloning method.

The PCR reactions using Go-Taq Flexi DNA Polymerase was run with the following conditions: an initial 95°C for 2 minutes, a cycle of 95°C for 1 minute, 42°C for 1 minute, and 72 °C for 1 minute was repeated 30 times, followed by 72 °C for 5 minutes and held at 4°C.

For Golden Gate method, the PCR reaction was run using restriction and ligations with the following conditions: 25 cycles of 3 min at 37°C (restriction) followed by 4 min at 16°C (ligation) and a final cycle of 5 min at 50°C and 5 min at 80°C.

2.6. Agarose Gel electrophoresis

The samples were loaded to a 1% or 1.5% agarose gel in TBE containing 5-10 µl of 1% of ethidium bromide and gels were run in 0.5% TBE buffer. The gels were visualised by a Typhoon FLA 9500 scanner.

2.7. Colony PCR

To perform colony PCR, a sterile tip was used to pick individual colonies and dip into each PCR reaction tube.

Each reaction made a total volume of 10 µl and contained 6.5 µl H_2O, 2 µl 5X Green GoTaq flexi reaction buffer, 0.8 µl, $MgCl_2$ 0.2 µl dNTPs, 0.2 µl of each primer and 0.1 µl GoTaq G2 DNA Polymerase (5 µl, Promega). The PCR program consisted of a first cycle of 94°C for 120 sec, followed by 29 cycles of 94°C for 30 sec, 58°C for 30 sec and 72°C for 90 sec / kb, followed by a final cycle of 72°C for 10 min and hold at 4°C.

2.8. GUS Staining with vacuum

Nbe plants were taken and rolled into a 15 ml falcon tube and covered with Staining solution (GUS buffer containing 01 mM EDTA, 50 mM $NaPO_4$, 0.1% Triton X 100 and water + 14 µl of 0.5M X-GIcA stock solution). The tubes were placed into a vacuum chamber and extracted air with a pump. Plants were vacuum infiltrated for 5-10 minutes or longer if the liquid has not penetrated into the tissue. The tubes were incubated at 37°C for overnight and the staining solution was removed. The Plants were transferred to petri dishes and washed with several changes of 100% ethanol and left shaking in 100% EtOH overnight or until tissue cleared. During the ethanol

washes the chlorophyll is extracted into the ethanol making it easier to see the blue staining of the plant tissue.

2.9. Golden Gate Cloning Method

2.9.1. Preparation of competent DH10B / DH5α / Top10 *Escherichia coli*

5 ml culture of *E. coli* DH5α, DH10B and TOP10 was grown overnight at 37 °C in Luria -Bertani medium (LB). From the overnight culture,1 ml was added to 100 mL of pre - warmed LB and cultured at 37 °C with vigorous aeration at 250 - 350 rpm (BR3. 11, Jouan centrifuges, UK) until the OD600 was equal to 0. 48. The cells were chilled on ice and centrifuged at 2500 rpm for 10 min. The pellet was dissolved in 30 ml of ice-cold filter sterilised TB1 pH 5. 8 (30 mM KOAc, 50 mM $MnCl_2$, 100 mM RECI, 10 mM $CaCl_2$, 15 % (w / v) glycerol), and kept on ice for 2 hrs. The cells were centrifuged at 2500 rpm for 5 min and the pellet was gently taken up into 4 ml of ice-cold filter sterilised TB2 (10 mM MOPS (pH 7), 75 mM $CaCl_2$, 10 mM RbCI, 15 % (w / v) glycerol). The cells were aliquoted on dry ice into 100 ml aliquots and stored at - 80 °C prior to use for transformation.

2.9.2. Cloning into *Escherichia coli* (*E. coli*) DH5α, DH10B and Top10

2.9.2.1. Heat shock transformation

For level 1- construct, *E. coli*, DH5α competent cells was used and DH10B and Top10 for level 2, 1 µl of DNA was added to *E. coli*, DH5α, DH10B and Top10 competent cells (100 µl), mixed gently, and kept on ice for up to 30 minutes. The cells were heat shocked at 42°C for 30 seconds and returned to ice for 1-2 minutes. This

was followed by the addition of 500 ul of LB and incubated and shacked at 200 RPM for one hour at 37°C.

2.9.2.2. Plating transformed *E. coli* DH-5a, DH-10B, and Top10

20 µl – 100 µl of transformed *E. coli* DH10B, DH5α and Top10 cells were plated onto the LB media in a sterile environment using an ethanol. Plates were cultured at 37°C overnight to enable colony formation. Single colonies were picked using a sterile toothpick or pipette tip to master plate and colony PCR was carried out.

2.9.2.3. Selection of transformants (Level1 and Level2)

For level 1, standard blue-white selection was used to identify colonies with a recombinant plasmid. Transformed cells were plated onto LB Agar containing 100 µg/ml ampicillin (AMP), 100 µg /ml IPTG and 40 µg /ml X - gal. Although success was generally high, 6-12 transformed colonies were routinely evaluated by colony PCR to identify truly positive colonies for continued verification prior to plasmid isolation and sequence verification.

For level 2, the kanamycin (KAN) resistance and the CRed color (red- white) selection was used to identify colonies with a recombinant plasmid. Transformed cells were plated onto LB Agar containing 100 µg/ml KAN. Colonies were routinely evaluated by colony PCR to identify truly positive colonies for continued verification prior to plasmid isolation and sequence verification.

2.9.2.4. Plasmid isolation from bacteria cultures and DNA sequencing

Plasmid DNA was isolated/purified from overnight *E*. coli culture in 5 ml LB using the QIAprep Spin Miniprep Kit (Qiagen). Overnight cell culture was pelleted by

centrifuging at 4000 rpm for 3 min at 4 °C as in the manufacturer's instructions and the DNA was eluted from the spin columns using 50 µl dH_2O. Sequencing was performed by Eurofins MWG Operon (Germany) and identity of the sequences was confirmed by BLAST alignment.

2.9.3. Preparing of competent *Agrobacterium* cells (EHA-105)

5 ml of LB with 50 µg/ml rifampicin (RIF) was inoculated with *Agrobacterium* (EHA-105 strain) and grown overnight at 28°C with 180rpm. The 5 ml culture was added to 500 ml of LB with 50 µg/ml RIF, this was grown overnight at 28°C with 180rpm, and once the OD600 reached 0.6-1.0 the cultures were cooled on ice for 15 minutes then centrifuged at 5000 rpm for 10 minutes at 4°C. The medium was discarded, the cell pellet was re-suspended with ice-cold water, and centrifuged again at 5000rpm for 10 minutes at 4°C. This was repeated twice and for the final time the cells were washed with 10% sterile ice-cold glycerol. Once the 10% glycerol has been removed the pellet was suspended in 1-2 ml of 10% glycerol and aliquoted into 400 µl.

2.9.4. *Agrobacterium* Transformation- Electro-transformation (EHA-105)

40 ng of plasmid DNA was diluted to 8 ng/µl and added to 200 µl of *Agrobacterium* competent cells on ice and stirred gently with a pipette tip. Then this was added to a cuvette, electroporated, and 1 ml of LB was added immediately to the *Agrobacterium* cells. The cells were incubated at 28°C for 1 hour at 180 rpm, then spread onto LB plates with 50 µg/ml RIF and KAN which were incubated for 2-3 days at 28°C to obtain single colonies. Once colonies were visible, colony PCRs, as described in section 2.1.7.**,** were run to ensure that the correct construct has been taken up.

2.10. Tomato cotyledon fragment transformation

The tomato cotyledon fragment transformation was carried out following the next steps:

1. Seed sterilization and germination

Wild type Microtom seeds were washed 2-3 times by sterile H_2O. The seeds were sterilised by dipping in liquid 50% bleach solution or 5-10% NaOCl until the seed coat was completely white (approximately 20-30 minutes). The seeds were washed by sterilized water for 5-6 times. They were transferred onto growth medium, which contained: MS + 3% glucose + 0,8 % agar, pH=6.0. The seeds were placed in the growth chamber under fluorescent light (lighting time is 12/24h)

2. Cotyledon fragment preparation – Cut and pre-culture

Cotyledons of 7-8 day old plants were cut into 0.1-0.3 cm segments. The segments were placed onto plates containing MS + 3% glucose + putrescine 1 mM + 1.0 mg/l zeatin + 0.1 mg/l IAA + 0,8 % agar, pH = 5.8. The plates were kept in dark at 26° C for 2 days and used for transformation.

3*. Agrobacterium* preparation (EHA-105)

Agrobacterium cells (EHA-105) from glycerol stock were streaked on to agar plates containing 50mg/l KAN and RIF, and the plates were incubated at 28° C for 1-3 days. To make a primary culture, a single colony was picked and inoculated into a test tube containing 10ml LB medium+ antibiotics (kAN and RIF). The culture was Incubated in a shaker at 28° C for overnight. To make secondary culture, when OD600 of the primary culture reached 0.8-1.0, 5 ml of primary culture was inoculated into 50ml fresh

LB + antibiotics (KAN and RIF). The culture was incubated in a shaker at 28°C for 3-4h until OD600 reached 0.5-0.8. Acetosyringone (200 µM) was added and incubated for additional 1h. *Agrobacterium* cells were collected by centrifuging at 6000 rpm for 10 minutes. The pellets were re-suspended by 50 ml fresh LB and used directly for transformation.

4. Transformation

The tomato cotyledon segments from step 2 were mixed in agro-bacterium suspension and kept for 30-60 minutes in a rocking equipment or gently inverted after each 5 minutes for 1 minute. The plant samples were taken out and dried on autoclaved Whatman paper. The plant samples were transferred onto plates containing co-culture medium: MS + 3% glucose putrescine 1 mM + 1.0 mg/l zeatin + 0.1 mg/l IAA + AS 200 µM + 0,8 % agar, pH = 5.8. The plant samples were kept in co- culture at 25° C for 2 days in dark.

5. *Agrobacterium* killing and plant selection

The plant samples were transferred onto bacteria killing and plant selection medium: MS + 3% glucose + 1 mM putrescine + 0.5 mg/l zeatin + 0.05 mg/l IAA + 250 mg/l cefotaxime (CTX) + 50 mg/l KAN + 0,8 % agar, pH = 5.8. The plates were kept in dark at 26° C. After 3-4 days, the plates were placed under lighting conditions. The samples were kept on the medium continuously until shoots appeared. The samples were transferred onto fresh medium after each 10-15 days.

6. Shoot elongation

When real shoots appeared, the shoot containing callus was transferred into growth chambers with medium: MS/2 + 3% glucose + 1 mM putrescine + 0.5 mg/l BAP + 250 mg/l CTX + 25 mg/l KAN + 0,8 % agar, pH = 5.8. The samples were kept on the

medium continuously until the shoots reached the length of 2-3 cm, when they were transferred onto fresh medium after each 10-15 days.

7. Rooting

The shoot was cut and placed onto medium: MS/2 + 3 % glucose + 1 mM putrescine + 1.0 mg/l IBA+ 50 mg/l KAN + 0,8 % agar, pH = 5.8.

8. Hardening

In a growth chamber, temperature is 26 °C and high humidity, the caps were opened, and the test tubes were covered by para-film. A small hole was made on the para-film on the first day, then 2 holes more were made on each following day. The para-film was removed on the fourth day. On the fifth day the plants were transferred into vermiculite pots and watered and covered by plastic bag. After 3-4 days, the plastic bag was partly opened and after 1 week it was completely opened. The plants should be transferred into soil when they reached the height of 5-10 cm but, unfortunately, they were died when they reached the height of 2-4 cm.

2.11. sRNA Northern blot

Total RNA samples were run in a 15% polyacrylamide-urea gel prepared by mixing 3.15 g urea, 1.875 ml water and 0.75 ml 5X TBE, briefly heated to mix, and 2.75 ml of 40% acrilamide/ bis solution 19:1 was added followed by 3.75 µl of TEMED and 75 µl of 10% Ammonium Persulfate (APS) just before pouring the mixture into a gel cast (BIorad).

Each sample were mixed with equal volume of gel loading buffer II (Ambion) and heated for 2 min at 70°C to denature any secondary structures before loading. After running (~100V, 2 hours approx.), the gel was stained with Ethidium bromide

and a picture was taken with Typhoon FLA 9500 scanner (GE Heatthcare Life Sciences) to record the quality and equal loading of the RNA samples.

Subsequently, RNA was transferred to a same size Hybond-NX nylon membrane through semi-dry blotting in a V20-SDB semi-dry blotter apparatus (Thermo Fisher scientific). The membrane was placed on top of the gel, facing the anode side, and both were stacked in between 0.5X TBE soaked Whatman filter paper. Transfer time was usually 35 min at 190 mA for one ~60 2 cm gel.

Once transfer was finished, RNA was attached to the membrane through chemical cross-linking. The membrane was placed on top of a Whatman filter paper with the RNA side facing up and soaked with 5 ml of cross-linking solution, wrapped with Saran cling film and baked at 60°C for 1.5-2 hours. The crosslinking solution consisted of 61.25 µl of 12.5 M 1-methylimidazote. 5 µl 12M hydrochloric acid (HCI) and 0.186 g of 1-ethyl-3-(3dimethylaminopropyl) carbodiimide hydrochloride (EDC) dissolved in water to make a final volume of 6 ml.

The membrane was hybridized with a ^{32}P end labelled DNA probe. For probe labelling, 2 µl of the oligonucleotide (10 µM, without 5'P) was mixed with 10 µl of water, 4 µl of 5X forward buffer, 2 µl of (γ- ^{32}P) ATP (3,000 Ci/mmol) and 1 µl T4 polynudeotide kinase (Life Technologies) and incubated at 37°C for 1 hour. After incubation, 30 µl of water was added to the mixture and all resulting 50 µl solution was transferred to the tube containing 5 ml of PerfectHyb Plus Hybridization Buffer (Sigma) where the membrane had been pre-hybridizing at 37°C with rotation for 1-2 hours. Subsequently, the hybridization tube was put back to the oven at 37°C and with rotation, to let the hybridization proceed overnight.

The following day, the hybridization solution was replaced with wash solution (0.2xSSC, 0.1% SDS). After a quick rinse, the membrane was washed twice for 20

min at 37°C in the rotating oven and with a final quick wash again. The membrane was wrapped in Saran film and placed in a cassette facing a phosphor screen. After 3-6 or more hours of exposure, the screen was scanned using the Typhoon FLA 9500 scanner (GE Healthcare Life Sciences).

If the same membrane was used for a subsequent hybridization with another probe, the first probe was stripped from the membrane by incubating it in 0.1% SDS at 80-90°C with rotation for ~1h or until no signal of radiation was detected with a Geiger counter. The probe used was the full 21 nucleotides complementary to Sly-miRNAtop14 and, after stripping, an oligonucleotide complementary to U6 RNA (AGGGGCCATGCTAATCTTCTC)

Chapter 3

miRNA “top14” Characterization and Expression

3.1. Introduction

3.1.1. Identification of miRNAs

In recent years, there has been a substantial progress in the identification of new miRNAs thanks to the development and improvement of next generation sequencing (NGS) technologies. Development of MIRNA predicting bioinformatic algorithms were also crucial to allow the identification of candidate miRNAs among the huge amount of sRNA reads in sequencing libraries. Prediction of miRNAs is based on parameters such as their length, phylogenetic conservation, pre-miRNA secondary structure and distribution of reads on the predicted pre-miRNAs (Mendes, *et al*., 2009, Akhtar, *et al*.,2015).

As a result of all these efforts to identify miRNAs, currently there are thousands of miRNAs annotated in over a hundred different species (Kozomara & Griffiths-jones, 2014). However, concerns about the authenticity of some of these miRNAs have been raised (Meng *et al*., 2012). In particular, it remains difficult to differentiate between miRNAs and endogenous siRNAs, given that the main difference between them is their biogenesis (Meyers, *et al*., 2008, Voinnet, 2009). However, sRNA NGS has the advantage of showing the distribution of reads across the genomic locus, allowing the distinction between single precisely diced miRNA- miRNA* pairs and the populations of randomly distributed reads characteristic of sRNAs (Jones-Rhoades, 2012).

In terms of expression, it has been shown that several miRNAs are differentially expressed during stress conditions. miR171 expression changes during cold, drought and salt stresses in Arabidopsis. In addition, miR395 and miR398 expression is strongly induced by sulphur and phosphate starvation, respectively. (Çakır et al., 2021).

An additional proof of miRNA authenticity and a first step in studying its function is the identification of target mRNAs (Jones-Rhoades, 2012). Fortunately, computational miRNA target prediction in plants can be done relatively easily due to their high complementarity to the target site (Rhoades, *et al*., 2002, Jones-Rhoades & Bartel, 2004). Consequently, their mode of action is frequently cleavage of target mRNAs (Axtell, *et al*.,2011). Experimental validation of a target can be achieved directly by detecting the cleaved transcript through techniques such as 5' RACE (Llave, 2002) or its high throughput version, parallel analysis of RNA ends (PARE) libraries, also called degradome sequencing (Addo-Quaye, *et al*., 2009, German, *et al*., 2009). This has been done previously for this study and in my work the focus has been the expression of top14miRNA by detecting the mature miRNA by Northern blot technique and detecting primary transcript by GUS staining technique.

3.1.2. Intron-split miRNAs

Plant *MIR*s usually contain introns, with a 67% of *Arabidopsis* pri-miRNAs estimated to have at least one intron (Stepien, *et al*., 2016). Both plant pri-miRNAs and their introns show a great variability in length, with sizes reported from 300bp to 5000bp for pri-miRNAs and from 100bp to 3000bp for their introns (Zhang, *et al*., 2009, Stepien, *et al*., 2016, Szweykowska-Kulinska *et al*., 2013). miRNA hairpins are usually located in the first exon (Stepien, *et al*., 2016, Szweykowska-Kulinska *et al*., 2013) although they may appear in other exons, as well as in alternatively spliced regions that may be either intronic or exonic (Yan, *et al*., 2012, Jia. *et al*., 2013)

Another interesting possibility reported is the one in which an intron appears in between the miRNA and the miRNA* sequence, dividing the miRNA stem- loop (Lu, *et al*., 2008, Sunkar, *et al*., 2005). This exon-intron-exon structure was first observed in a member of the miRNA444 family in rice, where the pri-miRNA fold-back structure

could only be predicted from the processed transcript but not from the genomic locus, suggesting the presence of an intron (Sunkar, *et al*., 2005). In a later study, also in rice, it was found that most members of the miRNA444 family had a characteristic feature not observed previously in any miRNA: they were transcribed from the antisense strand of their own target, for which they were called natural antisense miRNAs (nat-miRNAs) (Lu, *et al*., 2008). Interestingly, all MIRNA444 family members were found to have an intron in between their miRNA and miRNA*, even the two exceptions (miRNA444e and miRNA444f), which were not transcribed from the antisense strand of their target gene (Lu, *et al*., 2008).

This peculiar exon-intron-exon arrangement was subsequently reported also in miRNA444 members of maize and sorghum (Zhang, *et al*., 2009, Paterson, *et al*., 2009, Thieme, *et al*., 2011). Furthermore, a bioinformatic tool (SplamiR) was developed to identify this particular kind of pri-miRNAs with an intron in the middle of the miRNA hairpin, since they were not predicted by existing bioinformatic tools, unless spliced (Thieme, *et al*., 2011). In a recent review paper, they have been called intron-split miRNAs (Pek, *et al*., 2015). However, to our knowledge, up to date no intron-split miRNAs have been found outside the miRNA444 family.

3.1.3. Identification of miRNAtop14

In 2008, Moxon et al performed high-throughput sequencing of sRNAs from tomato leaf, bud and different stages of fruit development. After filtering out matches to tRNA and rRNA sequences, the remaining 18 to 30 nucleotide reads were mapped to SOL Genomics Network (SGN) tomato "bacterial artificial chromosome" (BAC) sequences, since the whole tomato genome sequence was not available at that time. Those reads aligning to the BAC sequences were subsequently analysed by checking whether there was also a plausible miRNA* sequence among

the BAC sequences, and by testing whether the predicted genomic MIR could fold into a miRNA-like hairpin using RNA fold. As a result, several known and putative new miRNAs were identified. One of the predicted miRNAs was the 14th most abundant read from the combination of the libraries and was therefore called "top14" in this publication.

Once the whole tomato genome was sequenced, it was noticed that miRNAtop14 and its putative miRNA* sequence was more distant from one another than was usually observed in other miRNA/ miRNA* pairs. (Figure 6.). In fact, they were almost 700 nt apart although 98% of plant miRNA hairpins have a length of less than 336 nucleotides (Thakur, *et al*.,2011). Therefore, a reverse transcription PCR (RT-PCR) was carried out from tomato total RNA with the aim of gathering more information about miRNAtop14 primary transcript. Surprisingly, when the PCR product was resolved in an agarose gel, two bands were visible; one of the expected genomic size between the primers but also another one around half the size of the upper band. When the two bands were sequenced, it was discovered that both amplified transcripts were indeed produced from *MIRtop14*, but the shorter one had an excised stretch in the middle with the GT- AG canonical intron splicing sites at the 5'-3' end, respectively, indicating that the pri-miRNA contained an intron in between miRNA and miRNA* sequences.

Besides other analyses, miRNAtop14 presence was predicted in several species within the Solanales order, including *Nicotiana benthamiana*. While this research was in progress, Baksa *et al*. 2015 published a study where they had sequenced sRNAs libraries from several *Nicotiana benthamiana* tissues, detecting mature miRNAtop14 in all of them, which they named Nb_miRC16_3p.

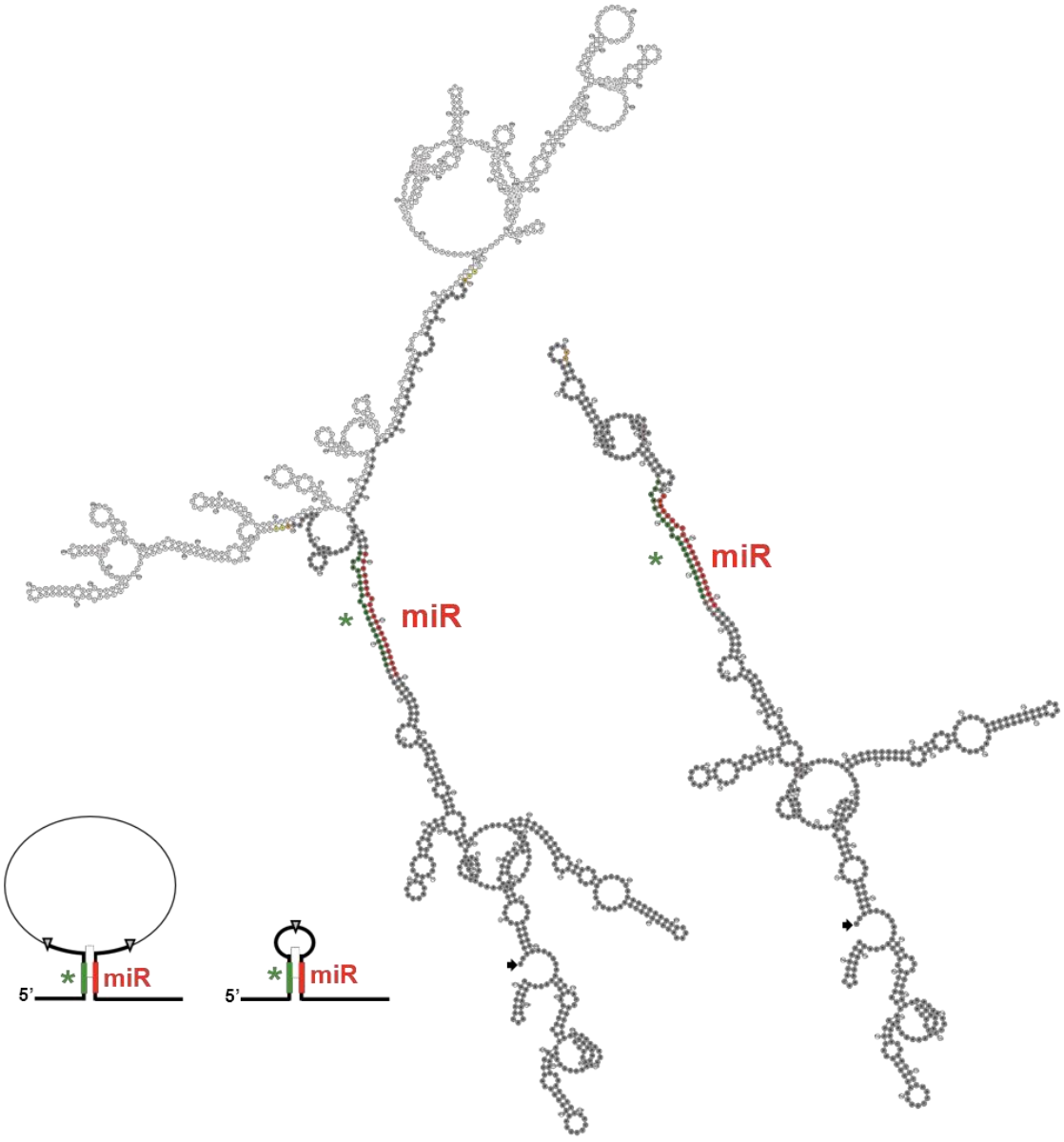


Figure 6. pri-miRNAtop14 secondary structure. pri-miRNAtop14 secondary structure.and schematic representation of the resulting miRNA hairpin of *Solanum lycopersicum*, spliced (right) and non-spliced (left) variants. Figure adapted from Zahara Calzada's thesis (not published).

3.1.4. Target of miRNAtop14

Target mRNAs were predicted for miRtop14 by Simon Moxon and then experimentally validated by Zahara Calzada, a former PhD student in the Dalmay laboratory. She was able to validate one predicted target LPR1 that belongs to the multicopper oxidase family.

Multicopper oxidases, MCOs, are a family of enzymes which could be founded widely in many species. These enzymes are containing four copper atoms in two centres; a type 1 centre with one atom and a type 2/ type 3 centre with three atoms each (Solomon et al., 1996). Multicopper oxidases can be divided in several subclasses, depending on their substrate, such as laccases, ascorbate oxidases, ferroxidases or bilirubin oxidases (Hoegger, et al., 2006; Sakurai, et al., 2007).

LOW PHOSPHATE ROOT1 (LPR1) and its close paralogue LOW PHOSPHATE ROOT2 (LPR2) are two proteins which belong to the MCO family of enzymes, they also both have been identified in Arabidopsis (Svistoonoff, et al., 2007). More particularly, they are ferroxidases, as both of them have been probed to display iron oxidation activity (Muller, et al., 2015).

Root architecture modification, responding to a media of low phosphate is a long- known phenomenon; proliferous roots help plants to improve exploring the soil and to extend their surface to uptake this nutrient (Raghothama, 1999). In Arabidopsis, low phosphate media reducing the growth of the primary root and increasing the growth of the lateral roots (Williamson, et al., 2001). (Figure 7.)

It was found that, in Arabidopsis there is a connection between LPR1 gene and the trait of primary root growth arrest under low phosphate in the soil (Reymond, et al.,

2006). Later on, a study has discovered the LPR1 gene responsible for this trait, as well as a paralogous gene, LPR2 (Svistoonoff, et al., 2007). Furthermore, at the same study, loss of function experiments with lpr1 and lpr2 identified that, both LPR1 and LPR2 play a role in the arrest of primary root growth under low phosphate. lpr2 showed a lower effect than lpr1, and lpr1/lpr2 double mutant showed an additive effect between two genes (Svistoonoff, et al., 2007) (Figure 7.). In addition, although LPR1 mRNA was identified in leaves and roots of Arabidopsis, further studies focused only on root, as it is the place that its function was discovered. Therefore, it was demonstrated that LPR1 is expressed in the root tip, between the meristematic region and the root cap (Svistoonoff, et al., 2007). Also, it was reported that the trait of arrest of growth response happened, when the root tip was connected to the low phosphate media more than any other part of the plant (Svistoonoff, et al., 2007).

Later on, LPR1 was, in particular, found in the endoplasmic reticulum in root cells (Ticconi, et al., 2009). Moreover, at the same study it was claimed that there is an interaction between LPR1 and a protein called PHOSPHATE DEFICIENCY RESPONSE 2 (PDR2), which is also localized to the endoplasmic reticulum of root tip cells (Ticconi, et al., 2009). Furthermore, it was found that under low phosphate conditions media, pdr2 mutant has a strong response to that media and develop an extremely short root with a much-reduced meristem (Ticconi, et al., 2004), while the triple mutants of pdr2, lpr1, lpr2 show a phenotype similar to double mutant of lpr1, lpr2 (Ticconi, et al., 2009) (See fig.7.). Therefore, it was concluded that, under low phosphate conditions, it is possible for PDR2 to act upstream in the process, also regulating LPR1/LPR2 and all together adjusting the activity of the primary root meristem (Ticconi, et al., 2009).

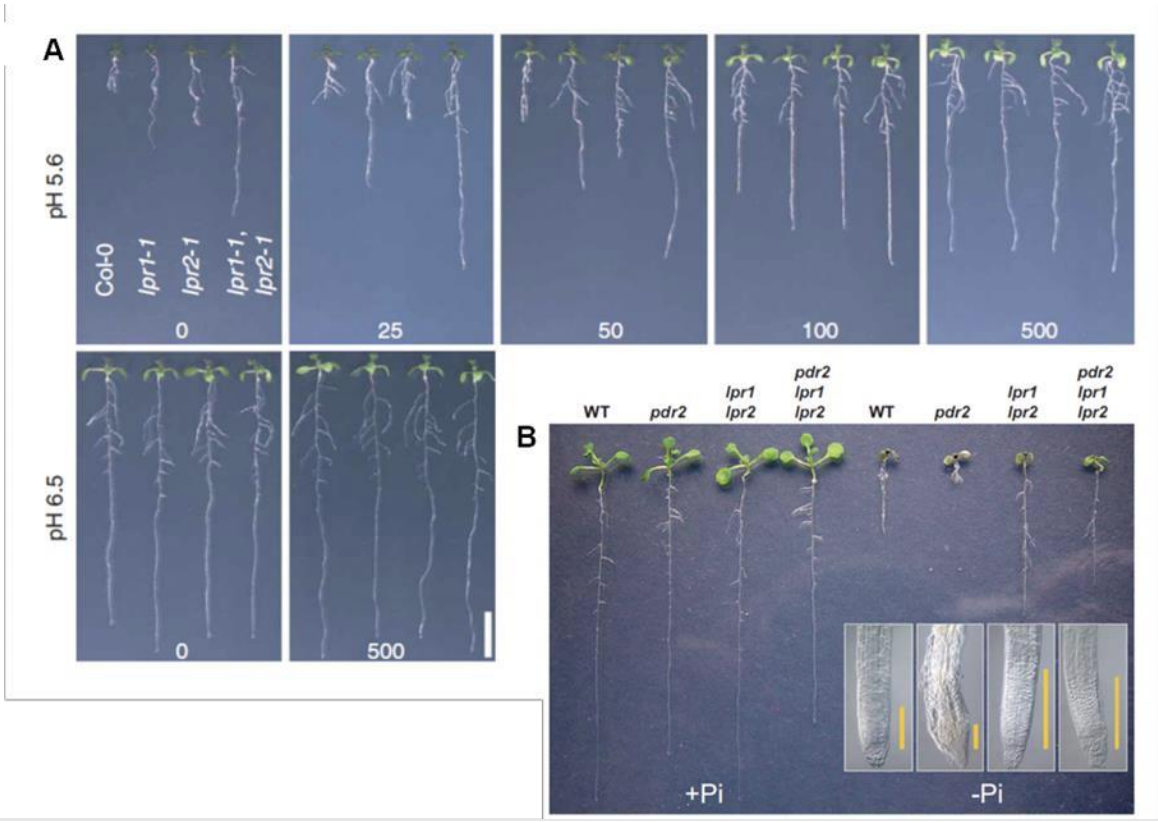


Figure 7. The relationship between LPR1 and LPR2 and root growth under low pi.

A) LPR1 and LPR2 have a similar function in root growth arrest under low Pi. The effect of LPR1 is stronger. Seeds of wild type Col0, lpr1 and lpr2 single mutants and lpr1, lpr2 double mutant were grown in media with the concentrations of Pi indicated at the bottom of each picture (µM) and the indicated pH and resultant phenotypes are shown.

B) LPR proteins interact with PDR2, acting downstream of it. Seeds of wild type Col0, pdr2, lpr1, lpr2 double mutant and pdr2, lpr1, lpr2 triple mutant were grown in media with sufficient or scarce Pi and the resultant phenotypes are shown. Insets show enlarged root tips, with yellow bars indicating meristem length. Figure adapted from (Muller, et al., 2015; and Svistoonoff, et al., 2007).

3.2. Result

3.2.1. miRNAtop14 Mature miRNA Expression Detection by Northern blot in Different Species and Tissues

Nicotiana benthamiana was analysed by sRNA Northern blot to confirm mature miRNAtop14 expression in vivo. miRNAtop14 has been detected in different parts of the plant, leaves, and root.

Total RNA was extracted from different parts of *Nicotiana benthamiana* leaf and root of three plants, following the protocol in section 2.1.2. and subsequently, 10ug was analysed by sRNA Northern blot (section 2.1.13). The same probe was used for the detection of mature miRNAtop14 in the three samples: the complementary of the mature miRNAtop14 sequence. The membrane was exposed to a phosphor-imager plate and the plate was scanned using the Typhoon FLA 9500 scanner (GE Healthcare Life Sciences). The results are shown in (Figure 8) where it can be seen that miRNAtop14 seems to be expressed in roots at higher level than in leaves and there is a slight gradient across leaves where the expression level is lowest in oldest leaves and highest in youngest leaves. However, the following bar chart graph (Figure 9) is for the northern blot data for N. benthamiana which represent statistical analysis -with ImageQuant- of the findings. It shows that there is no significant difference between the expression levels in the tissues ($p > 0.05$). Further research would be needed to determine the precise pattern of miRNAtop14 expression.

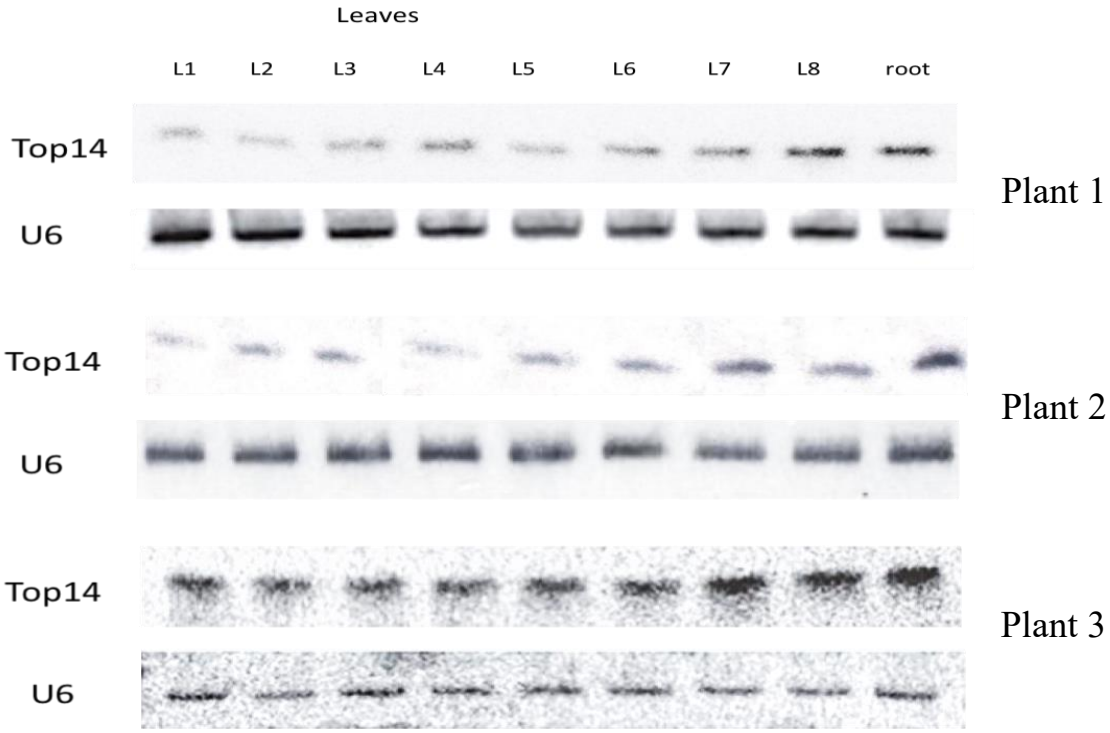


Figure 8. Detection of mature miRNAtop14 in N. benthamiana. Detection of mature miRNAtop14 levels in wild type N. benthamiana leaves and root by Northern blot. Different leaves from the same plant were named L1 to L8 from the oldest to the youngest. The bands showing the expression level of top14. U6 detection is included as loading control.

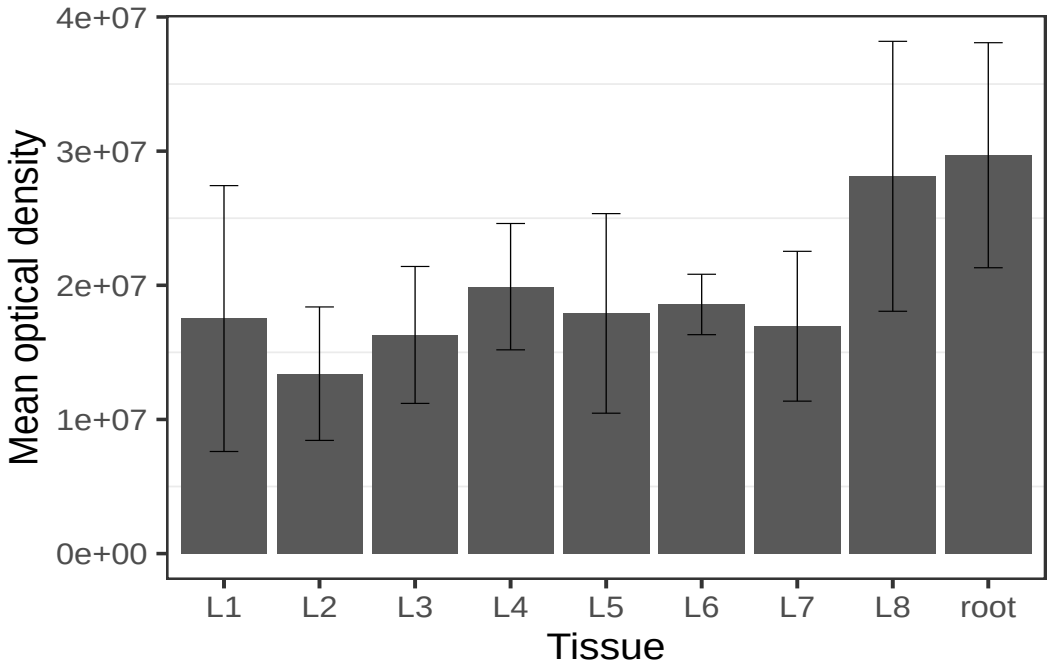


Figure 9. Statistical analysis of mature miRNAtop14 in N. benthamiana. Statistical analysis of mature miRNAtop14 levels in wild type N. benthamiana by ImageQuant. The tissue is different leaves from three plants were named L1 to L8 from the oldest to the youngest and roots.

In addition, different levels of miRNAtop14 have been detected in different parts of *Solanum lycopersicum* (cv. Microtom), plant (leaves, root and stem) (see figure 10.)

Root, stem, and leaves were collected from *S. lycopersicum* cv. Microtom plants grown in soil for ~5 weeks. Total RNA was then extracted from each tissue following the protocol in section 2.1.2.

2ug of total RNA from each sample was analysed by sRNA Northern blot (Section 2.1.13). The probe used was the full 21 nucleotides complementary to Sly-miRNAtop14 and, after stripping, an oligonucleotide complementary to U6 RNA. The membrane was exposed to a phosphor-imager plate and the plate was scanned using the Typhoon FLA 9500 scanner (GE Healthcare Life Sciences).

Furthermore, in the attempt to replicate these experiments, it has been found that these levels are highly fluctuate (Figure 10.). This may be an indicative of induced expression of miRNAtop14, possibly under specific stimulus. However, further research would be needed to determine the precise pattern of miRNAtop14 expression.

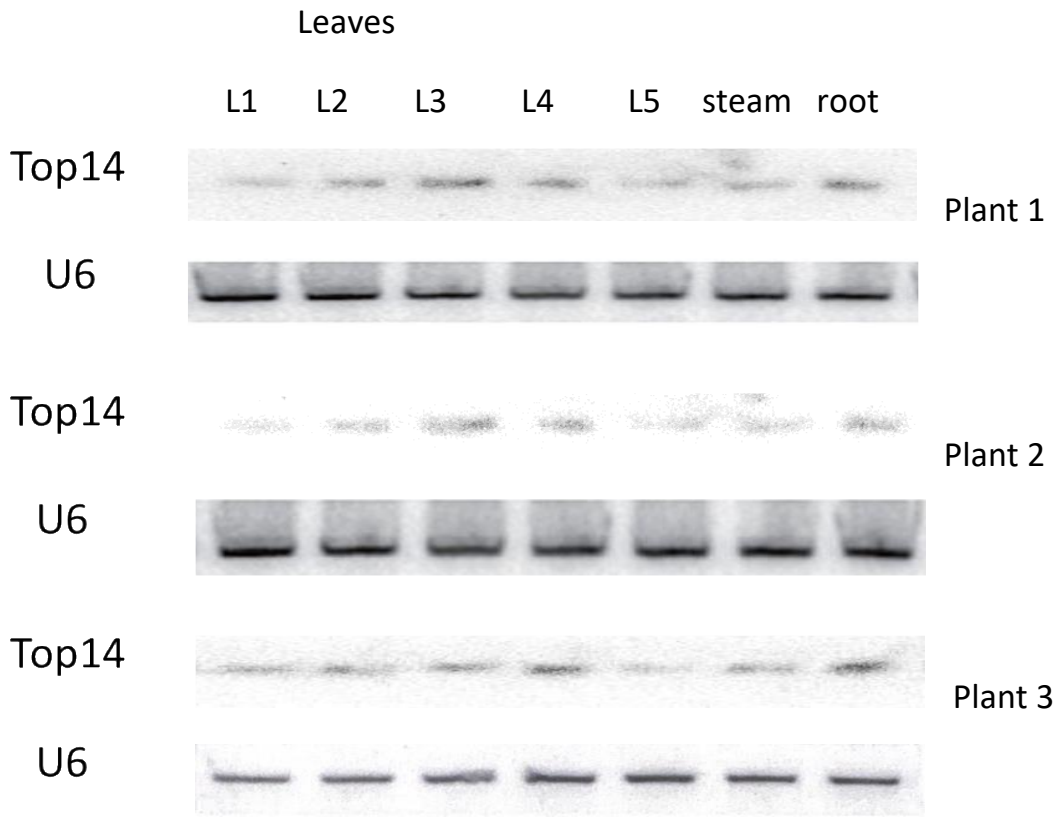


Figure 10. Detection of mature miRNAtop14 in Microtom. Detection of mature miRNAtop14 levels in wild type Solanum lycopersicum (cv. Microtom) in root, stem, and leaves by Northern blot. Different leaves from the same plant were named L1 to L5, from the oldest to the youngest. The bands showing the expression level of top14. U6 detection is included as loading control.

3.2.2. miRNAtop14 Primary Transcript Expression Detection by GUS Staining

Seeds of transgenic *Nicotiana benthamiana* seeds with a construct harbouring the reporter gene GUS under the top14 promotor (generated by a previous PhD student Zahara Calzada) were sown and the seedlings were analysed by GUS staining to confirm primary miRNAtop14 expression in vivo. GUS activity has been detected in different parts of the plant (leaf, root, and stem). Seeds of transgenic *Nicotiana benthamiana* were sown on medium and grown at the temperature and light condition indicated in (2.1.1) for two, three, and four weeks. Then, the whole plants were subjected to GUS staining as described in 2.1.8.

The results are shown in figure 11. and 12. where it can be seen that GUS (and therefore) milRNAtop14 is expressed everywhere and the expression level seems to be the highest in roots. (Figure 11. and 12.)

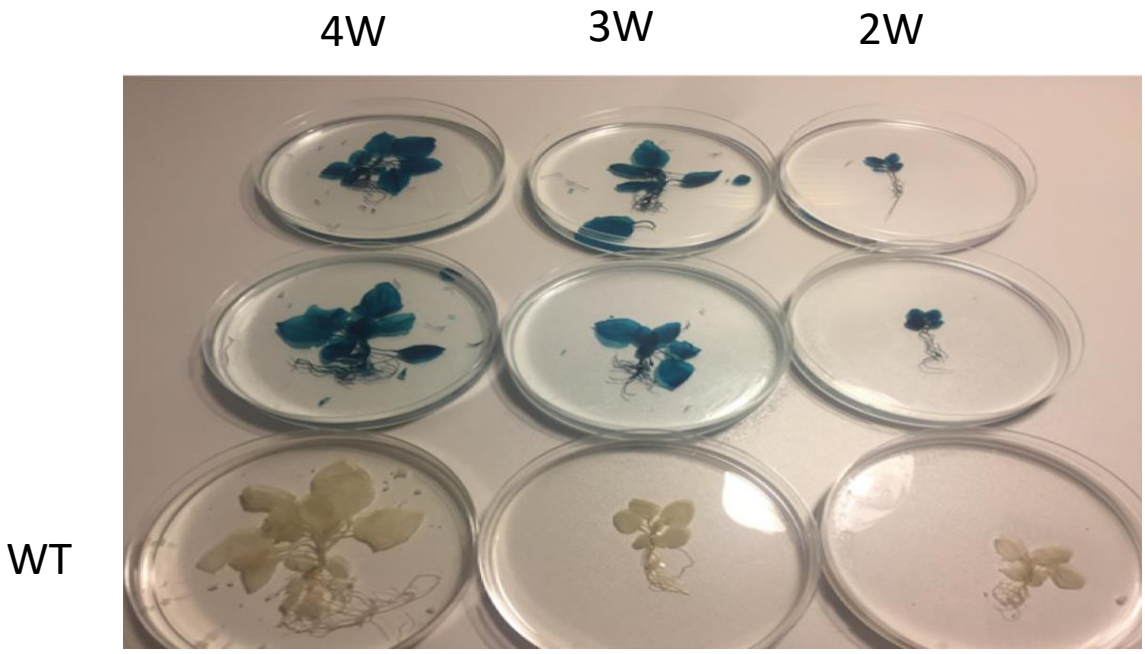


Figure 11. Detecting of GUS activity as a proxy for primary miRNAtop14 levels in transgenic Nicotiana benthamiana root, stem and leaves by GUS Staining. Three plants were used in different age (two, three, and four weeks). Wild type is included as loading control.

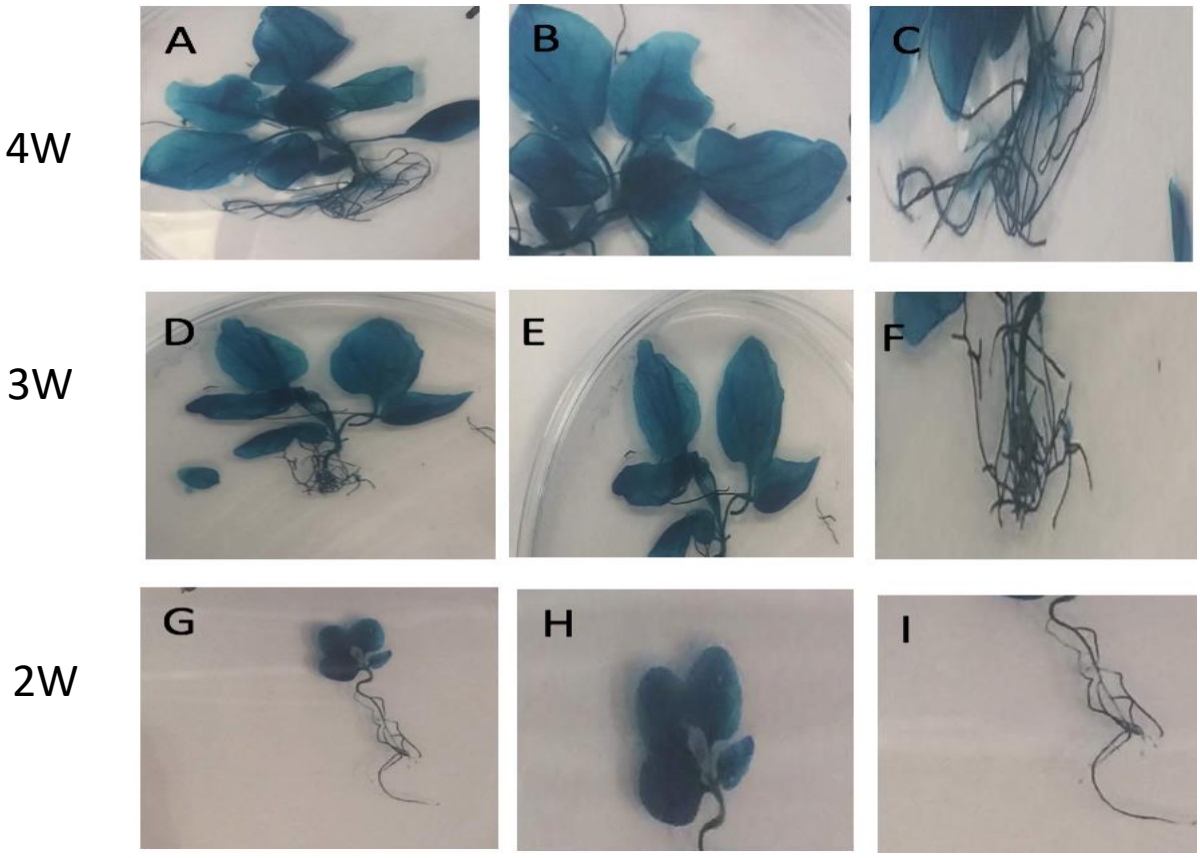


Figure 12. Detecting of GUS activity as a proxy for primary miRNAtop14 levels in transgenic Nicotiana benthamiana roots, stems and leaves by GUS Staining. Three plants were used in different age (two, three, and four weeks). Images of the whole plants (A,D,G) and their higher-magnification images are presented as follows: B and C from A, E and F from D, and H-I from G

3.3. Discussion

miRNAtop14 has been identified in a couple of independent studies, the first one in S.lycopersicum and N. benthamiana (Moxon et al., 2008 & Baksa et al., 2015). However, the expression pattern was not described.

Many miRNAs are expressed ubiquitously across all tissues, however, there are examples for miRNAs that are expressed in a tissue specific manner. Tissue specific expression pattern can give a clue about the function of the miRNA. Probably the most striking tissue specific expression is shown by miR395 (Kawashima et al 2009). miR395 is very strongly induced during sulphur starvation but only in the phloem companion cells of the vascular system. Interestingly, one of its targets (SULTR2;1) is also induced by sulphur starvation but it is primarily expressed in xylem parenchyma cells. This indicates that the function of miR395 is to keep SULTR2;1 expression restricted to xylem cells during sulphur starvation. This example shows that miRNAs can have very specific expression pattern and it can contribute to our understanding to its function. In this study, we have attempted to establish expression pattern of miRNAtop14 in S. lycopersicum (cv. Microtom) and N. benthamiana, that would help to understand the functionality role of top14.

3.3.1. Mature miRNAtop14 Expression

Northern blot analysis of wild type *Solanum lycopersicum* (cv. Microtom) and *Nicotiana benthamiana* was carried out to detect mature miRNAtop14. The analysis of Microtom included root, stem and leaves, while the *Nicotiana benthamiana* analysis was restricted to leaves and root. Though the Northern blots showed that expression level seems to be highest in root and lowest in shoot of N. benthamiana, the statistical analysis showed no significant difference between the expression levels in the tissues.

Also, no consistent differential expression was detected in tomato. One potential reason for the lack of statistically significant difference across the leaves, despite the potential gradient from oldest to youngest leaf, is that the biological replicates were too different from each other. This increased the error bars on the graphs and therefore the seemingly different tissues (root and old leaf and young and old leaves) were not statistically different from each other. This could be addressed by either using many more replicates or trying to find a way to make the replicates more similar to each other including a synchronisation of leaf age. If there is a real difference in mature miRNAtop14 level in old and young leaves (and a gradient in between) then it is very important that the leaves appear at the same time on the different plants used for the experiment.

Analysis of miRNAtop14 expression in wild type Microtom by qPCR in order to assess changes in the levels of miRNAtop14 expression and compare the data/results with Northern blot analysis for the same plant would help to unravel the expression level and expression pattern of miRNAtop14. In addition, analysing miRNAtop14 level in individual leaflet of the compound leaves of tomato may reveal differential expression in leaflets at different positions.

3.3.2. pri-miRNAtop14 Expression

Whole transgenic *Nicotiana benthamiana* plants expressing GUS from the miRNAtop14 promoter were analysed by GUS staining to investigate pri-miRNAtop14 expression. GUS expression seemed to be very similar across all leaves and higher in root than in leaves.

Future experiments may try to quantify the pri-miRNAtop14 expression level including the both the spliced and un-spliced forms. First of all, the GUS staining may be saturated, and subtle differences could have been missed. The seedlings could be

stained again but the reaction could be stopped earlier when the tissues are not dark blue, yet. In addition, the GUS activity could be measured to obtain a more accurate level of GUS expression, rather than just looking at the intensity of the colour.

Analysis of wild type *Nicotiana benthamiana* by qRT-PCR, would help to unravel the level of splicing of pri-miRNAtop14 in different leaves. Primers could be designed at each side of the intron and using a short extension time during the PCR cycle, only the spliced variant would be quantified. In a separate reaction a forward primer could be used that anneals to an intron region and a reverse primer that anneals to the second exon. This reaction would only quantify the un-sliced form. Both spliced and un-spliced forms could be analysed in the different age leaves, as well as shoot and root tissues.

Unfortunately, we did not find a strong tissue specific expression for neither the pri- nor the mature miRNAtop14 but as discussed above, further, more careful studies may still establish a gradient for mature miRNAtop14 across leaves. In that case, if pri-miRNAtop14 expression is constant across the different leaves and tissues, the potential difference would be most likely due to differential splicing in the different age leaves and tissues. This could be tested by the qRT-PCR experiments suggested above.

Chapter 4

“top14” biological function: Generating a CRISPR-Cas9 construct by Golden Gate cloning method

4.1. Introduction

The most important requirement of biological research and biotechnology is the availability of effective DNA assembly methods. Therefore, various methods of cloning DNA fragments of interest have been developed. Conventional methods generally require several cloning steps to generate the construct of interest. A single DNA fragment is transferred from the donor plasmid or PCR product to the recipient vector. Various methods have been developed in recent years to facilitate and speed up this process. One such method is Golden Gate cloning (Engler et al., 2009).

Golden Gate cloning or Golden Gate assembly (Engler, et al., 2008) is a molecular cloning technique that permits researchers to use type-II 'restriction endonucleases enzymes (IIS) and T4 DNA ligase to simultaneously and purposefully convert multiple DNA fragments into one fragment *in vitro* (Engler, et al., 2009). The most typically used Type IIS enzymes include BsaI, BsmBI, and BbsI. Compared with standard type II restriction enzymes such as EcoRI and BamHI, these enzymes cut DNA outside of their recognition sites. Thus, a non-palindromic overhang can be formed (Weber, et al., 2011). Since there are 256 potential overhang sequences (assuming there are four unpaired nucleotides), multiple fragments of DNA could be assembled by using combinations of overhang sequences (Weber, et al., 2011). Additionally, because the final product does not have a Type II restriction enzyme recognition site, the properly ligated product cannot be cut again by the restriction enzyme, meaning that the reaction is irreversible (Weber, et al., 2011). A common thermal cycler protocol cycles between 37 °C (best for restriction enzymes) and 16 °C (optimal for ligases) several times (Engler, et al., 2009). While this technique

can be used to build a single insert, researchers have used it to assemble many pieces of DNA at the same time. (Engler, et al., 2014).

Using the Golden Gate cloning process, up to ten pieces can be assembled at once in a recipient plasmid (Engler, et al., 2009). Cloning is accomplished by pipetting all plasmid donors, the recipient vector, a type IIS restriction endonuclease enzyme, and ligase into a single tube and incubating the mixture in a thermal cycler. (Engler, et al., 2008). Due to the simplicity of the cloning procedure, the majority of clones obtained after transformation contain the expected construct. Golden Gate cloning, on the other hand, required the use of carefully designed donor and recipient plasmids. (Engler, et al., 2008).

Golden Gate cloning method is based on type IIS restriction enzymes' unique ability to cleave outside of their recognition site (Engler, et al., 2008). These recognition sites are removed in the cleavage process when they are put to the far 5' and 3' ends of any DNA fragment in inverse orientation, allowing two DNA pieces flanked by compatible sequence overhangs to be ligated seamlessly. (Figure 13.) The directed assembly of several fragments is possible because type IIS restriction sites can be engineered to form various overhangs, which are referred to as fusion sites (Lebedenko, et al., 1991).

All DNA segments are delivered as uncut plasmids for assembly, and the target vector, T4 DNA ligase, and type IIS restriction enzyme are combined in a single reaction mix. The application of restriction-ligation allows for exceptionally efficient assembly of numerous fragments. Golden Gate is one of the advanced cloning techniques that can be used for genome editing constructs. It is applicable for the overall assembling of multiple DNA fragments to create effector nucleases called TAL (Zhang et al. 2021). The CRISPR technology has also adopted Golden Gate Cloning

for the insertion of oligonucleotides that specify the target sequence of the gRNA (Zhang et al. 2021).

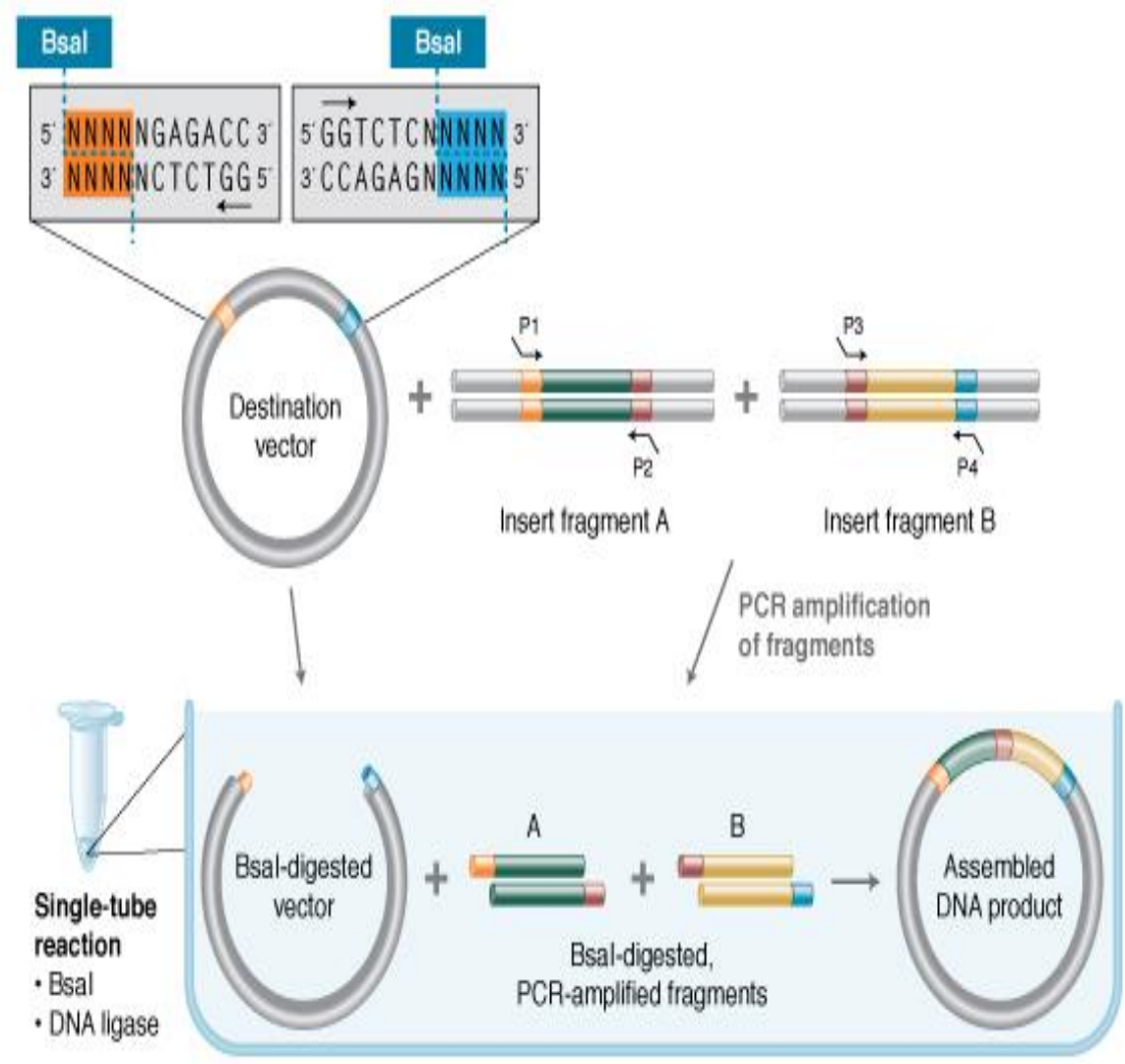


Figure 13. Design requirement of Golden Gate Assembly. Constructs are assembled by mixing in one tube all module plasmids (or PCR fragments) and a destination vector together with the appropriate type IIS enzyme and ligase. Image adapted from New England Biolabs Inc.

4.2. Cloning Standard

Cloning standards are employed in DNA assembly to prevent changes in cloning efficiency and plasmid function caused by compatibility issues between the restriction sites on the insert and those on the vector (Engler, et al., 2014). The cloning standards at Golden Gate Assembly are divided into two categories. The single-gene construct is created by adding genetic components such as promoters, open reading frames, and terminators to the first-tier Golden Gate assembly. Then, in second-tier Golden Gate assembly, a multigene construct is created by combining many constructs created in first-tier assembly. The modular cloning (MoClo) methodology and the GoldenBraid2.0 standard is performed to enhance second-tier assembly (Casini, et al., 2015).

4.3. MoClo System

MoClo uses a parallel approach, with restriction sites for BpiI on both sides of the inserts in all tier-one (level 0) constructs. The vector, also known as the "destination vector," has an outward-facing BsaI restriction site with a drop-out screening cassette, where genes will be inserted (Casini, et al., 2015). LacZ is a popular screening cassette that is replaced in the destination vector by the multigene construct (Casini, et al., 2015). The layout of the ultimate multigene construct is determined by the overhangs on each tier-one build and the vector, which are different but complementary to the overhang of the next section (Casini, et al., 2015).

Cloning Golden Gate normally begins with level 0 modules (Engler, et al., 2014). If the level 0 module is too large, cloning will begin with level -1 fragments, which should be sequenced before the large construct can be cloned (Engler, et al., 2014). The level 0 modules do not need to be sequenced again if starting from level -

1 fragments, but they must be sequenced if starting from level 0 modules (Engler, et al., 2014).

4.3.1. Level 0 Modules

Level 0 modules provide the foundation of the MoClo system and include genetic elements such as a promoter, a 5' untranslated region (UTR), a coding sequence, and a terminator. In Golden Gate Cloning technique, the core sequences of level 0 modules should not contain type IIS restriction enzymes sites for BsaI, BpiI, or Esp3I while flanked by two BsaI restriction sites inverted orientation. During the cloning procedure, level 0 modules without type IIS restriction sites bordering them can add the BsaI sites (Engler, et al., 2014).

If any undesirable restriction sites exist in the level 0 modules, they can be edited *in silico* by removing one nucleotide from the type IIS restriction site. During this step, it needs to be ensured that the introduced mutation has no effect on the genetic function encoded by the sequence of interest. A silent mutation in the coding sequence is desirable since it has no effect on the protein sequence or function of the gene of interest (Engler, et al., 2014).

4.3.2. Level -1 Fragments

Large level 0 modules are cloned with the help of level -1 pieces. Blunt-end cloning with restriction ligation can be performed to clone level -1 fragments. The vector used to clone level -1 fragments should not contain the type IIS restriction site BpiI, which is required for the assembly step that follows. Furthermore, in the following assembly stage, the vector should have a different selection marker than the destination vector; for example, if spectinomycin resistance is applied in level 0

modules, level-1 fragments should have another antibiotic resistance, such as ampicillin (Engler, et al., 2014).

4.3.3. Level 1 Constructs

The position and orientation of each gene in the final construct are determined by the level 1 destination vector (Marillonnet, et al., 2015). There are fourteen different level 1 vectors available, each with a different outside fusion site sequence but identical inside fusion sites. As a result, all vectors can include the identical level 0 pieces (Marillonnet, et al., 2015). Because all level 1 vectors are binary plasmids, they are employed for transient expression in plants via *Agrobacterium* (Marillonnet, et al., 2015).

4.3.4 Level 2 Constructs

Because of the insertion of level 1 modules, level 2 vectors have two inverted BpiI sites. The upstream fusion site is compatible with level 1 vector-cloned genes, whereas the downstream fusion site has a universal sequence. Each clone allows for the insertion of 2-6 genes into the same vector (Marillonnet, et al., 2015).

It is not recommended to include more genes in one cloning phase, as this would result in erroneous constructions (Marillonnet and colleagues, 2015). On the one hand, this may result in more restriction sites in the construct, as this open construct allows for the inclusion of other genes. On the other hand, this can eliminate restriction regions, where the closed construct prevents new genes from being added.

Thus, End-linkers with BsaI or BsmBI internal restriction sites, as well as blue or purple markers, are required for constructs comprising more than six genes.

The restriction site and the marker should be alternated in each cloning stage. In addition, two restriction enzymes are required: BpiI for releasing level 1 modules

from level 1 constructs and BsaI/BsmBI for digesting and opening the recipient level 2 plasmid.

The right colonies should alternate from blue to purple at each cloning phase while screening, but if a "closed" end-linker is used, the colonies will be white (Marillonnet, et al., 2015).

4.4. Result

4.4.1. Selecting Genomic Targets and Designing gRNAs

CRISPR-Cas9 is a unique technology that helps researchers to edit certain genome locations specifically. Currently, CRISPR-Cas9 is the most precise and versatile method of genetic manipulation as removal, alteration and addition of nucleotides can be achieved with high specificity (Soyars et al. 2018). The CRISPR-Cas9 consists of two main key molecules, which are an enzyme called Cas9 and a piece of RNA called guide RNA (gRNA). In terms of the gRNA, it consists of an approximately 20nt long pre- designed RNA sequence, which is located in a longer RNA scaffold. The scaffold part attaches to DNA sequence and the gRNA guides the Cas9 enzyme to the specific region of the genome which is called a genomic target (Cooper et al. 2022). One of the important steps in this method is selecting the target site in the genome. The genomic target is ~20 nucleotide DNA sequence, that is immediately upstream to a Protospacer Adjacent Motif (PAM), which is a three-nucleotide motif specific to each Cas enzyme (Brooks et al, 2014). The PAM serves as a binding signal for Cas9 enzyme, but the exact sequence depends on which Cas protein is used. In our study to select the best two gRNA sequences for MIRNAtop14 gene, the targets were selected by using online tools such as CRISPR-P 2.0 and E-CRISP. We chose target regions that were at each side of the mature MIRNAtop14

sequence and were just upstream of a 5’ NGG motif, that is the PAM for Cas9. (Figure 14.). In this study the selected gRNA target sequences were separated by forty base pairs and an approximately sixty base pair deletion was expected (underlined). (Figure 14.). As the target base pairing is achieved by the gRNA to the target site, a double-strand break is mediated by Cas 9 a few nucleotides upstream the PAM.

In addition, there are a few other factors that are important for gRNA design. The first is the GC content of the guiding sequence that should be 40-80% (Rönspies et al. 2022). The other important aspect is the length of the gRNA and the average length of the gRNA is around 20bp. The GC content and length together influence the specificity of the gRNA because they both affect the strength of annealing between the gRNA and the target site. If the GC content is too low and/or the gRNA is too short, the Cas9 may be guided to unintended sites, which is called off-targeting. For example, it has been observed that the guide sequences which are shorter than 17 bp can target multiple loci and therefore have high off-target activity (Coelho et al. 2020). The CRISPR/Cas9 construct was designed to target two target sites at each side of MIRNAtop14 with the intention to create an approximately 60bp deletion to delete the entire mature MIRNAtop14 sequence (Belhaj et al., 2013). The two sgRNAs were expressed from the Arabidopsis U6 promoter that requires a G at the first position, therefore the first nucleotide of the guide sequence was designed to be a “G”. Both of gRNAs constructs were designed 48 nt long including the guide sequences, which match the target sequences, followed by the fusion sites (ATTG, GTTT,) and finally the restriction site for Esp3I enzyme (Brooks et al, 2014) (Figure 15.).

5′*TAGGCCACTTATATTTGTGAAAGAATA*ATATAATGTACTGCATCATATCAATAATCTTA
TCTATAAAAGGAGTAAAATTTGCTACATTGAATATACCATAGAGGAAATTGAATTAATGAA
GAAGTAGAAGATGACACTTTGTTGGTGACTTTGATCTCAAAAGAGTGCTTATCAATATTGT
TTGTTTAATTTATTACGGTATGTTATTTGTCTTATTTTACTTTAGTAAACTATTTATTGAA
ACTTCTTTCAAAGATTAGTTTTTTTAGTCGAAAGTTTTTTGAAAGCATATTTTATATTGAG
CAAGAGGTAAGAATAAGATTTATATACATTTCATGACCTGCTTATGAAATCATACTGAATA
TGTTATTATTATTATTTATTTTAGCAAAATGCTCATGTAGTATATTGTATTGTACCTTAAT
TGAATGTTTGATGCATGTTTGTGGGGCGGTGAAGCAATATTTGATCCTATATATGAAATAG
TTTTTTAGTTGAAGATCGTTCAGCTGATCGTCCTTGAATTTATATACCTTTTATAATATAA
AAGTGTCAAATTTGTAATTTAAAAAAAATAATCACCATTCTTATATACTATTTGTAAGAAT
GATCCTTGTTATTTTTACTTGAACTTAGAAATGATTTAGATGATTTACTTTAAAATTTTAA
ACATATTGTAGTCTTGCTTTTTTGCATTACATGCCTTTTTGCTCTAACCTTTGTTCAAGAA
ATCGTGTACAATGGAATTGTGAAACAATATTTGTAGGAATATCAAGCATTATTATCATCAC
TATATGGACACATGGCATTATACTCTTGGGACCAAAGTCACCAACAGAGTGTCAAATTTCT
CATCATTCCGAGCTCAAACACAGCGACGGAGCCAGAATTCTCATCGAAGAAAATATCAAAA
TATTTTCTTAATTAATGATCTCTATTGTTTTGTTGTTTTTTAGTAGTGCCTTGTTATTTTC
TTTGGAGAATGTAACCAGCTAGATGTAATCCCCACATATATGTAATATAATAAATCAATCC
CCTACCTTTTGGGGGTATGTTTACTA*AGCTATGCAACATAAAGTAACACTTTA*

Figure 14. Solanum lycopersicum MIRtop14 sequence. Exon sequences are shaded in grey while introns are not shaded. Intron GT- AG consensus SS sequences are highlighted in yellow miRNAtop14 is highlighted in red and miRNAtop14* is highlighted in green. Blue and green nucleotides are the targets sequences. choosing target regions of miRNAtop14 the sequence that contained the requisite binding region, or protospacer adjacent motif (PAM), for Cas9 cleavage, which is a 5'-NGG located immediately after the 20 bp target DNA (in purple). The selected gRNA targets sequences were separated by forty base pairs and sixty base pairs deletion was expected (underlined).

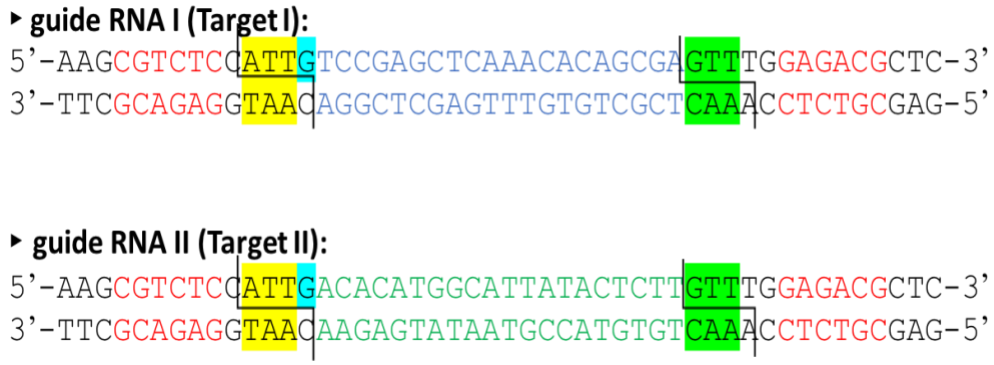


Figure 15. Structure of guide RNAI and guide RNAII construct. The gRNA is 21 nt long. The blue and green bp are the guide sequences. The fusion sites are shaded in green and yellow. The red bp are the restriction site for Esp3I enzyme. The two sgRNAs were expressed by using a small RNA promotor, "Arabidopsis U6 Promotor". First nucleotide in the guide sequence should be a "G" (shaded in blue) if U6 promotor is used. Guide sequence should match the target except for the first nucleotide (5' G) that does not have to match.

4.4.2. Generating a CRISPR Cas9-sgRNAs Construct

To knock out miRNAtop14, a Cas9-sgRNAs construct was designed by using Golden Gate cloning approach, including two sgRNAs alongside the Cas9 endonuclease gene with the intention to create a large, defined deletion (Belhaj et al., 2013). The CRISPR Cas9-sgRNAs construct was designed in two levels. In Level 1 constructs, two plasmids were designed. First, a vector carrying the two sgRNAs fragments placed under the Arabidopsis U6 promoter which were digested by the BsaI restriction enzyme (Figure 16.). Second, a plasmid carrying the *S. pyogenes* Cas9 (SpCas9) gene, which was expressed from the p35S promoter and t35S terminator and the EsP3I restriction enzyme (IIS) was used for digesting (Figure 17.). Both level 1 vectors used Ampicillin as a selection marker. For level 2 construct, Cas9-sgRNAs were assembled to have the final construct. In the final construct, two restriction enzymes were needed, where BpiI is used for releasing level 1 modules from level 1 constructs and BsaI for digesting and opening the recipient level 2 plasmid. In addition, Kanamycin was used in level 2 as a selection marker (Figure 18.).

Level 1 constructs pL1M-R2-p35S-AtCas9-t35S, pL1M-F4-pAtU6-26-miTOP14_1sgRNA1, pL1M-F5-pAtU6-26-miTOP14_2 sgRNA2 and the linker pL1M-ELB-3-49277 were assembled into the level 2 vectors pL2B-KAN-CRISPR-, pL2B-KAN-CRISPR-miTOP14 and their linker pL1M-ELE-5-41800.

Finally, colony PCR and sequencing procedures were carried out to test, whether white colonies were carrying the right insert (Cas9 and sgRNAs). Colony PCR was performed as previously explained in section (2.1.7) by using Cas9 gene forward and reverse primers and single guide RNAs (sgRNAs) forward and reverse primers.

FOR-AtCas9	AAGTACTTGAAAGCAGCTG
REV-AtCas9	GAACCGCTCTTATCAAGAAG

FOR-sgRNA	CCCGCCAATATATCCTGTC
REV-sgRNA	GCGGACGTTTTTAATGTACTG

In addition, sequencing was performed by Eurofins MWG Operon (Germany). The results are shown in (Figure 18.) where it can be seen that a final construct that contained two single guide RNAs (sgRNAs) alongside the Cas9 endonuclease gene with the intention to create large, defined deletions to knock out miRNAtop14.

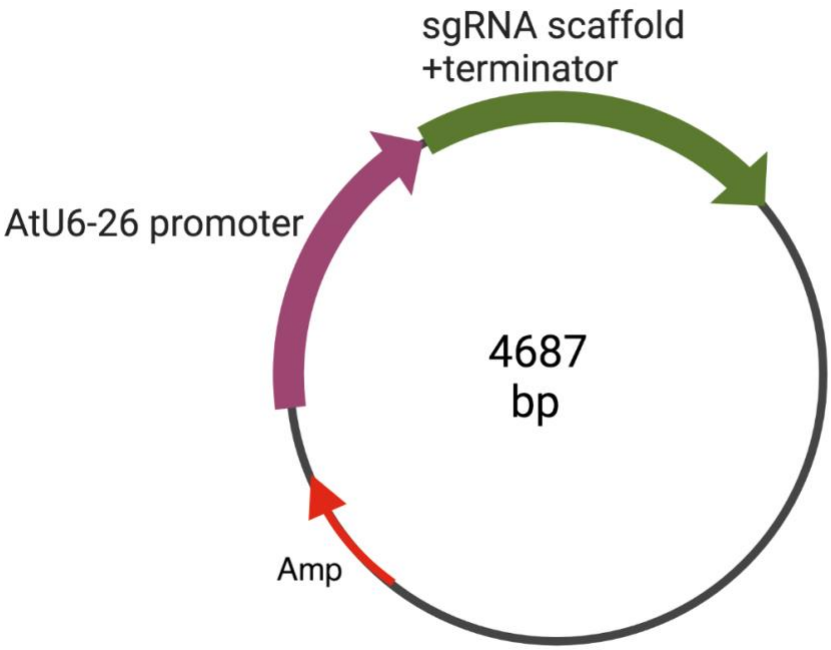


Figure 16. Level 1 construct carrying sgRNA. Level 1 construct carrying sgRNA fragments placed under the Arabidopsis U6 promoter which were digesting by using Bsal restriction enzyme**.**

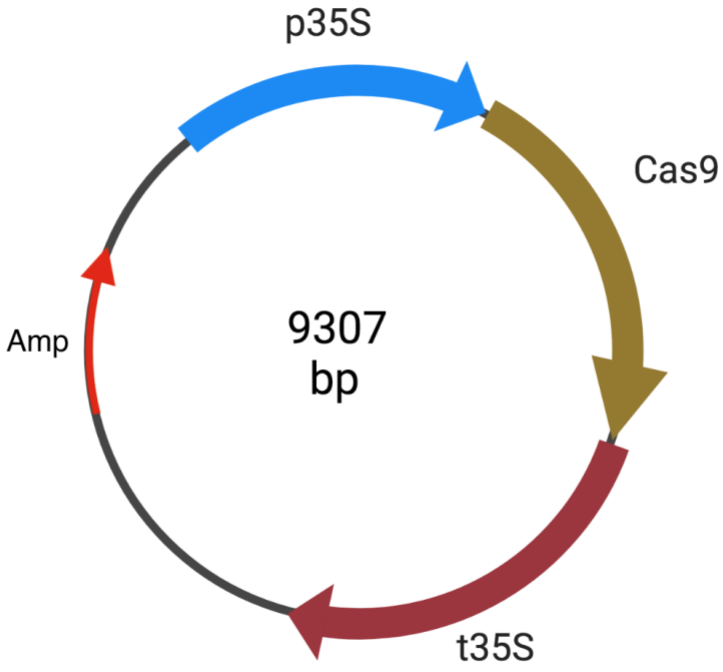


Figure 17. Level 1 construct carrying Cas9 gene. The S. pyogenes Cas9 (SpCas9) was used in this study, which was expressed by using p35S promoter and t35S terminator and EsP3l restriction enzyme (IIS) was used for digesting.

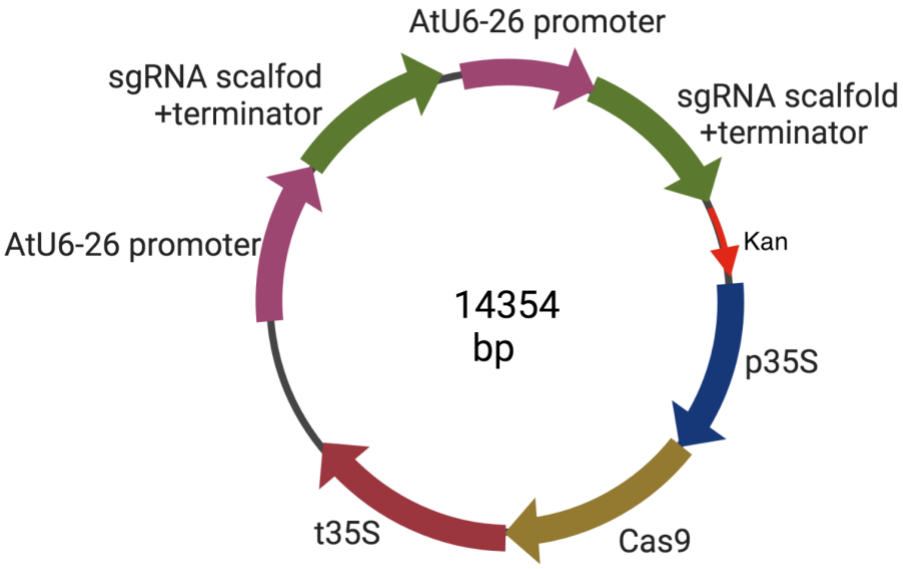


Figure 18. Final construct Cas9-sgRNAs plasmid**.** The final construct is carrying the Cas9-gene and two sgRNAs, presented here show that all elements required for the design of a completely automated knocking out system are now in place. Two restriction enzymes were needed, BpiI is used for releasing level 1 modules from level 1 constructs and BsaI is for digesting and opening the recipient level 2 plasmid.

4.5. Discussion

Golden gate cloning relies on the Type IIS reaction enzymes. In order to execute the cloning by using the golden gate cloning method, a cassette is designed with the gene of interest flanked by compatible ends to sticky ends produced by type IIS restriction enzymes' such as Bbsl and Bsal (Chiasson et al. 2019). These cleavage sites are located upstream and downstream of the recognition sites of the restriction sites, therefore after successful ligation of the gene of interest, the recognition sites get eliminated. Any DNA fragment which has an unpaired TGGA sequence at the 5' end and an TCCG overhang at the 3' end can ligate into a specific vector. (Hinz et al. 2021). The overhang of the entry-level DNA can be presented in the original plasmid or this can be added through the addition of the PCR-based amplification.

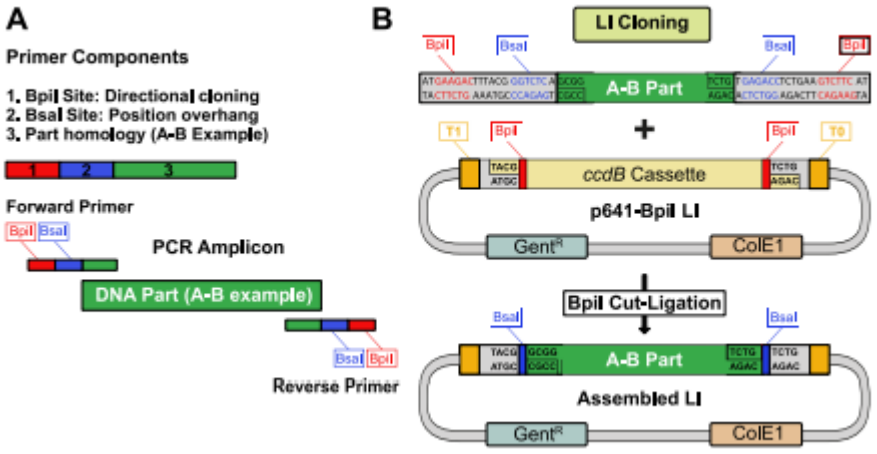


Figure 19. **Demonstration of Golden gate cloning**

(Source: Chiasson et al. 2019)

Figure 19. shows an overview of the primer components that are required for the cloning of a new DNA part (A). Other than the homology of the targeted parts, the BpiI restriction site is also contained within the primer for the directional level 1 cloning process (Chiasson et al. 2019). Similarly, for the level 2 cloning process, a BsaI site is

also required. In order to perform a cut-litigation of BpiI with the amplified part and in order to generate an entry-level clone of the MK system, p641-BpiI backbone support is required (Chiasson et al. 2019). T0 and T1 terminator sequences are contained within the p641-Bpi plasmid where the clones are inserted within the flanking region.

The CRISPR system consists of a CRISPR-associated protein 9 (*Sp*Cas9) endonuclease fused with a nuclear localization signal and a single-stranded guide RNA (gRNA) that directs the *Sp*Cas9 to the target locus. A gRNA is a piece of RNA that functions as a guide for the enzyme protein (Cas). It occurs naturally in prokaryotes, however, can also be designed to be used for targeted editing in eukaryotes if a Cas enzyme is also provided. (Kim et al. 2021).

The aim of this study was to knock out top14, in order to deduce the function by analysing the phenotype of the plants in the absence of top14 miRNA. There are two approaches to reach our goal of disruption top14: either using one or two gRNAs. We could have designed one gRNA to the middle of the mature top14 sequence, which would result in the expression a mutant top14. Although, that would not target the original top14 targets but potentially may target other mRNAs. Therefore, we chose to design two gRNAs to target two flanking sites of the mature top14 sequence which would result in two cuts, leading to the deletion of an approximately 60bp region, including the mature top14 miRNA sequence, rather than expressing a mutant top14. Oh et al. 2020 have mentioned that by expressing of several gRNAs with *Sp*Cas9 proteins in a single cell can edit multiple genes or induce a large deletion of a specific chromosome. In addition, there is always a possibility that a gRNA doesn't bind to its target. If only one gRNA is used and it does not bind to its target, Cas does not cleave and there is no mutation. If two gRNAs are used and one of them does not

bind, at least the other mediates cleavage and a mutation is introduced, although not the desired deletion. Considering that the gene editing constructs usually have to be delivered into the plants using tissue culture technique that is very time consumable, the increase in efficiency by using two gRNAs is an important factor (Oh et al. 2020). Moreover, according to Xie et al. their two-gRNA expression system in rice are more effective than just one gRNA expression system: two gRNAs are connected by a tRNA precursor sequence and processed into individual gRNAs after transcription under the control of U6 promoter (Xie et al. 2020). To find out if this method is generally applicable to gene editing in other plants, more investigation is required. Off-target analysis for the selected gRNAs is shown in appendix.

Chapter 5

Transforming Tomato Plants with CRISPR-Cas9 Construct

5.1. Introduction

5.1.1. Genome Editing Methods Overview

During the past several years, there has been fast development of gene editing methods that make it possible to directly target regions of chromosomes in a sequence-specific way. Two of these approaches, zinc-finger nucleases (ZFNs) (Urnov et al., 2010) and transcription activator- like nucleases (TALENs) (Joung and Sander, 2013), are based on protein-DNA interactions, whereas a third method, (CRISPR)/Cas9 endonuclease (Liu and Fan, 2014), is an RNA-guided DNA endonuclease system. Each technology has advantages and disadvantages in terms of cost, ease of construction, efficiency of targeting, and specificity (Liang et al., 2014). Of the three genome- editing technologies, CRISPR/Cas9 has been applied during the past eight years with applications to bacterial, animal, human, and most recently, plant systems (Pennisi, 2013). The reason behind that is due to a greater number of advantages of CRISPR/Cas9, including the straightforward construct design and assembly compared to ZFNs and TALENs.

5.1.2. CRISPR/Cas9 Editing in Plants

In 2013 the first CRISPR/Cas9 editing reports were published in plants with success in both transient expression and stable transgenic lines. There have been reports on three crop species, rice (Zhang et al.,2014), sorghum and wheat (Jiang et al.,2013), as well as evidence of efficient application in model species, such as *Arabidopsis thaliana* (Li et al.,2013) and *N. benthamiana* (Nekrasov et al., 2013), (Wang et al., 2014). For stable transgenic lines, changes in the genes of primary transformants in Arabidopsis and rice have been shown to continue in the next

generation (Feng et al., 2014; Zhang et al., 2014). As a result, CRISPR/Cas9 is rapidly becoming the ideal method for gene editing in plants, despite the need for further tests to determine whether efficacy is universal.

Brooks et al. (2014) and others (Feng et al., 2014; Gao and Zhao, 2014; Zhang et al., 2014) suggest that CRISPR/Cas9 is ready to rapidly become the technology that will allow genetic knockouts to be generated. As an example, in 2014 Brooks and others introduced mutations by gene editing to Solyc11g064850 (AGO7) that controls leaf development. After successful editing of AGO7, they have developed two constructs to target the three homologs of Solyc11g064850, to test the gene editing efficiency of CRISPR/Cas9 beyond *AGO7* and explore its potential for reverse genetics. In the tomato genome, there are three homologs of Solyc11g064850, two of which have very similar nucleotide sequence, suggesting possible redundancy (Consortium, 2012). Two sgRNAs were incorporated, one designed to target the Solyc08g041770 and one designed to simultaneously target a conserved region in the second to last exon of Solyc07g021170 and Solyc12g044760. Eight T0 plants have been regenerated for each construct (Brooks et al., 2014). By doing PCR genotyping they identified that six of eight transgenic lines (75%) recovered from transformation with the CRISPR/Cas9 construct targeting Solyc08g041770 carried mutations, and most plants were likely chimeric (Brooks et al., 2014). PCR genotyping showed that two plants had the expected deletion (Brooks et. al., 2014).

In 2014, the transient use of CRISPR/Cas9 was reported in tomato roots (Ron et al., 2014). At specified loci in so-called “hairy root” structures, mutations were introduced which were induced by *Agrobaterium rhizogenes* carrying a CRISPR/Cas9 transgene. However, there was no regeneration of transgenic plants (Ron et al., 2014).

To our knowledge, for the first time, in 2014 Brooks and others have demonstrated CRISPR/Cas9- induced mutations and their heritability by subsequent generations in stable transgenic tomato lines.

5.1.3. Generating Knockout Using CRISPR-Cas9 System

The reason behind choosing CRISPR-Cas9 technology in our study is that among dicot crops, tomato, *Solanum lycopersicum*, is an ideal crop for using CRISPR/Cas9 gene editing and that because of diploidy of the genome, the genome sequence is a high quality and the transformational methods are efficient and available (Consortium, 2012 & Van Eck et al., 2006)).

CRISPR system can be used to generate knockout cells by co-expressing an endonuclease like Cas9 (also known as Cpf1) and a gRNA specific to the targeted gene. The genomic target is ~20 nucleotide DNA sequence, that is present immediately adjacent to a Protospacer Adjacent Motif (PAM) (Brooks et al, 2014). In other words, the location of sequence targets should be upstream of a PAM sequence (5′-NGG-3′) (Belhaj et al., 2015). The PAM serves as a binding signal for Cas9 enzyme, but the exact sequence depends on which Cas protein is used. Once expressed, the Cas9 protein and the gRNA form a ribonucleoprotein complex through interactions between the gRNA scaffold and surface-exposed positive charged grooves on Cas9. In gRNA binding, Cas9 undergoes a conformational change which shifts the molecule from a non-DNA binding inactive conformation into an active DNA binding conformation. Importantly, the gRNA is mainly free to interact with the target DNA in the spacer region (Brooks et al, 2014 & Belhaj et al., 2015). Only if the gRNA spacer sequence has sufficient homology with the target DNA, will Cas9 cleave in a given locus.

When the Cas9-gRNA complex binds with putative DNA target, the seed sequence (8-10 bases at the 3′ end of the gRNA targeting sequence) is annealed to the target DNA. If the seed and target DNA sequences match, the gRNA will still anneal to the target DNA in a 3′ to 5′ direction. Therefore, mismatches between the target sequence within the 3′ seed sequence completely abolish target cleavage, while mismatches toward the 5′ end distal to the PAM often still allows target cleavage to be achieved. (Brooks et al, 2014). A second conformational change is made to Cas9 after target binding that positions the nuclease domains, called RuvC and HNH, in order to cleave opposite strands of the target DNA. A double-strand break (DSB) within the target DNA (~3-4 nucleotides upstream of the PAM sequence) is the result for Cas9-mediated DNA cleavage (Belhaj et al., 2015). Then the result of DSB is repaired by one of two general repair pathways. The first one is the efficient but error-prone non-homologous end joining (NHEJ) pathway. Second, the less efficient but high-fidelity homology directed repair (HDR) pathway. The NHEJ repair pathway is the most active repair mechanism, and it frequently causes small nucleotide insertions or deletions at the DSB site. The reliability of the NHEJ-mediated DSB repair has significant practical impacts, because there will be a variety of mutations in the cell population of Cas9 and gRNA. In most cases, NHEJ creates small indels in the target DNA that result in amino acid deletion, insertion or frameshift mutations leading to premature stop codons in the ORF. A loss of function mutation in the target gene is the ideal end result. The strength of the knockout phenotype should be validated experimentally for a particular mutant cell (Brooks et al, 2014).

5.2. Result

5.2.1. Deliver the CRISPR/Cas9-sgRNA Construct to the Plants

The *Agrobacterium-mediated* transformation (AMT) method specifically relies on the capability of *Agrobacterium tumefaciens* (a bacterial pathogen) to transfer foreign genes into several host plants. In recent years, AMT has become one of the most effective and commonly used methods to generate different transgenic plants. Gene transfer to plants from bacteria, however, is a complex process that includes a number of steps. In this particular context, the AMT has been found to be one of the most effective ways of gene transformation (Jin et al., 2022). *Agrobacterium tumefaciens* is a common soil phytopathogen, which directly infects wound sites of plants. Correspondingly, this bacteria becomes responsible for crown gall disease in plants by delivering transfer (T)-DNA from their cells into the specific host plant cells through a bacterial type IV secretion system (T4SS). This characteristic of the pathogen has been used in the gene transfer process. It has also been noted that AMT is a highly evolved and complex mechanism that involves genetic determinants of the bacteria, as well as the host plant cell.

It has also been found that, in general, five steps are included in the overall AMT process, in the context of gene transfer from bacteria to plants. These AMT steps include colonisation of specific species of bacteria, bacterial virulence system's introduction, T-DNA transfer complex, transfer of T-DNA and finally T-DNA's integration into the host plant genome (Jin et al., 2022). Through these five steps, these bacterial strains transfer a single-copy of T-DNA along with Virulence (Vir) effector proteins to the host plant cells (Roushan et al., 2022). As a consequence, this

protocol enables the effective insertion of non-rearranged, stable and single-copy sequences into the plant genome. Accordingly, using this method, it becomes possible to transfer comparatively large segments of DNA with minimal rearrangement. It has also been proven that through the use of this method it is possible to create transgenic plants with high fertility and quality as well. Taking all of the characteristics and advantages of this gene transfer protocol into consideration, it has been used to deliver the final construct Cas9-sgRNAs.

Following the method, as explained by Stanton Gelvin (2018), the AMT of tomato cultivar M82 has been performed as discussed in 2.9.4 section. The cotyledon segments, from 6 to 8-day-old seedlings have been pre-cultured for one day followed by inoculation with *A*. *tumefaciens* strain EHA-105 containing the Cas9- sgRNAs construct. In the following step, co-cultivation took place for two days and then the cotyledon segments were transferred to a selective regeneration medium containing 75mg/l kanamycin. When shoots were approximately 1.5-2 cm tall, they were transferred to a selective rooting medium that also contains 75 mg/l kanamycin. (Figure 20.).

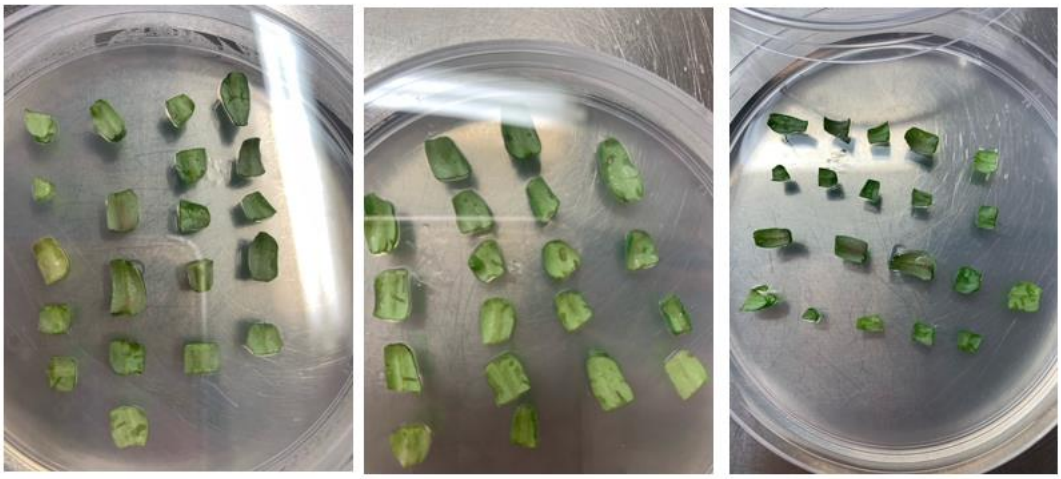
5.2.2. Regenerate and Screen Transgenic Plant for Gene Knockout Events

Gene knockout refers to a process that enables perturbing of the genomic DNA of a model organism of a cell. Thus, that specific gene's expression is stopped permanently. In this particular context, the targeted genes are damaged to make them non-functional (Prihatna et al., 2018). In this process, the plants are transferred into rooting medium after becoming 1.5-2 cm in height (Figure 20.(c)), with those transgenic plants screened for gene knockout events. Contrastingly, these plants died when they showed 2-4 cm in height. (Figure 20. (d)). Although, we have attempted

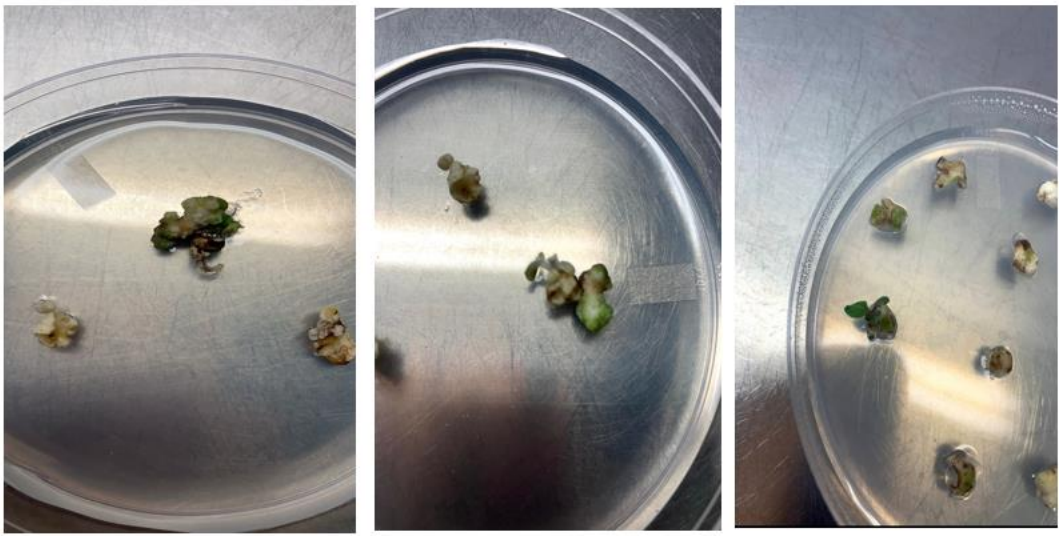
many times to regenerate transgenics plants with CRISPR- Cas9 deletion of top14, the shoots could not survive, even with many attempts, when transferred to the rooting medium. This became a serious issue in our efforts to transform tomato plants and eventually we did not manage to root any of the shoots.

(a)

(b)



(C)



(d)

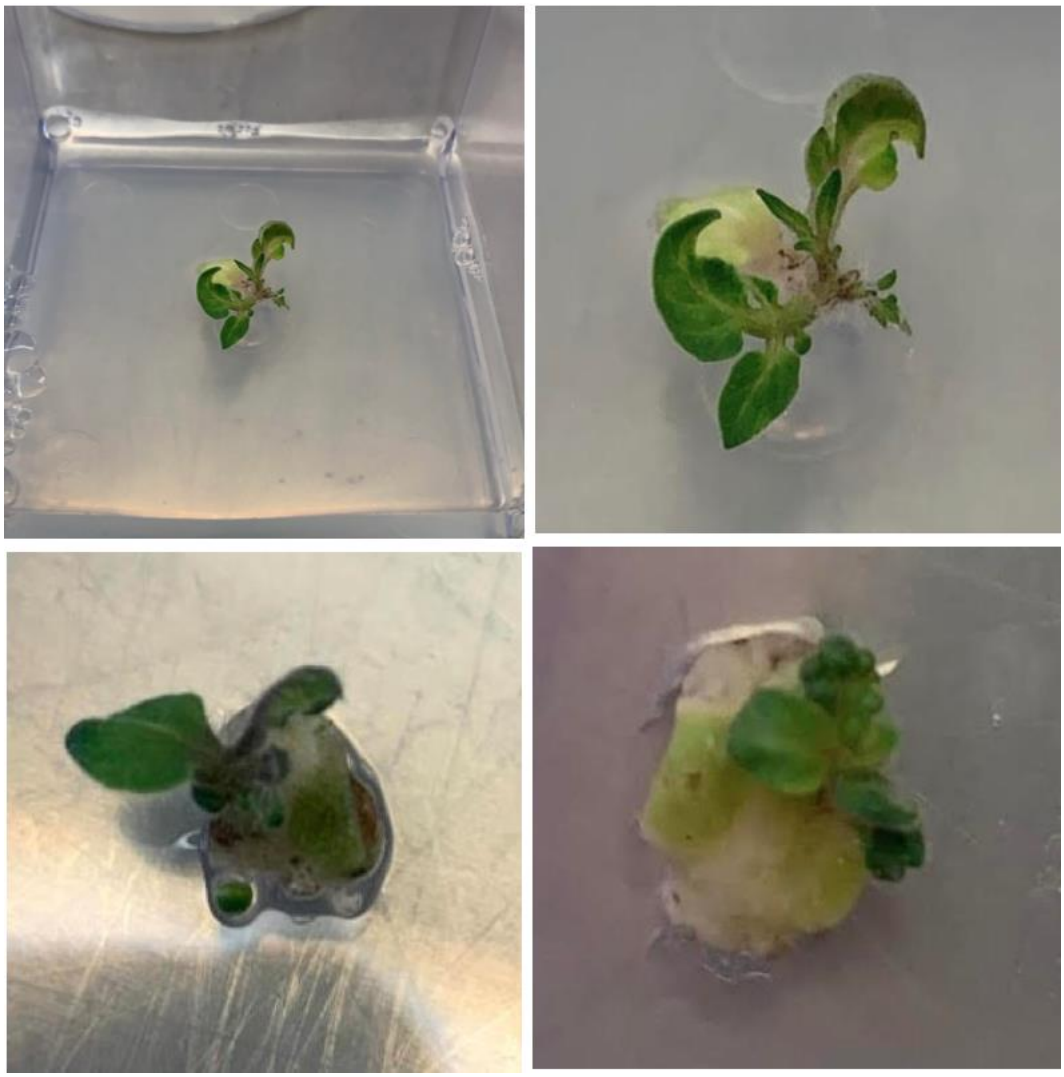


Figure 25. Transformation of Micro-Tom. (**a**) Seedlings of 8 days after imbibition. (**b**) Explants on co-culture medium. (**c**) Explants on callus induction medium (**d**) shoot regeneration from callus on shoot induction medium. Shots reach 3-5 cm

5.3. Discussion

5.3.1. Potential pitfalls of CRISPR/Cas9

To identify the biological function of MIRNAtop14 gene, CRISPR-Cas9 approach was applied in this study by following the steps described by Brooks et al, 2014. Despite the fact that CRISPR/Cas9 has a broad range of use in research, there are several aspects that affect its efficiency and specificity. Although, we have attempted many times to regenerate transgenic plants with CRISPR- Cas9 deletion of top14, the shoots could not survive when transferred to the rooting medium. Several potential reasons could be behind this such as off-target effects, Cas9 activity, target site selection and sgRNA design, and that top14 may be essential for growth and development. The following sections are to discuss these reasons.

5.3.1.1. Off-target effects

Off-target effects can be defined as unintended cleavages causing mutations at untargeted genomic locations with a similar but not identical sequence to the target site. It is not exactly known why the Cas9 protein cleaves some off-target sites and not others (Modrzejewski et al., 2019). Electroporation showed low numbers of off-target mutations in plant protoplasts when the sgRNA and Cas9 delivered as RNP complexes (Subburaj et al., 2016). Moreover, delivering RNP complexes by liposome-mediated transfection revealed the reduction of off-target mutations compared to plasmid DNA transfection (Liang et al., 2015).

The CRISPR-Cas9 system is well studied in archaea and bacterial genomes. These shows that not all the positions of the gRNA are required to match the sequence of the target DNA; thus, off-target sites are cleaved (Wu et al., 2014). The cutting or

binding may not lead to any functional consequences when the off-target sites lay outside of regulatory elements or targeted genes. Off-target effects could affect the plant's development and growth as the issue of off-targeting in plants still not well understood (Feng et al., 2014; Gao and Zhao, 2014; Zhang et al., 2014). In addition, Ron et al, (2014) have used CRISPR-Cas9 in order to make mutations in tomato roots specifically in loci so called "hairy root", however, there was no regeneration of transgenic plants (Ron et al., 2014). Moreover, in 2014 for the first time, to our knowledge, Brooks and others have used CRISPR/Cas9 to induce mutations in stable transgenic lines of tomato, as well as the heritability of these mutations through subsequent generations. However, they recommended that future genomic sequence be required to assess if, and to what extent, CRISPR/Cas9 causes off-targeting in tomato because in the animal field CRISPR/Cas9 system was reported to have a high off-target rate (Carroll, 2013).

Brooks et. al., 2014 suggested that future genome sequencing is needed to evaluate of whether, and to what extent, CRISPR/Cas9 causes off- targeting in tomatoes (Brooks et. al., 2014). In the animal field, the CRISPR/Cas9 system was reported to have a high off-target rate (Carroll, 2013). Several recent studies have, however, demonstrated the high CRISPR/Cas9 specificity in plants (Feng et al., 2014; Gao and Zhao, 2014; Zhang et al., 2014). The problem of off-targeting in plants still needs to be addressed systemically, but a range of approaches can reduce the potential impacts by using a newly developed algorithm to select the least predicted CRISPR / Cas9 sgRNAs (Xie et al., 2014). Off-target analysis for the selected gRNAs is shown in appendix.

5.3.1.2. Cas9 activity

As a genome editing tool, the CRISPR/Cas9 system cleaves particular nucleotides based on sequence complementarity with the sgRNA. The Cas9 protein goes through a significant conformational shift with the interaction of sgRNA and target DNA. As a result, Cas9's catalytic nuclease lobe spins by ~100°, producing nucleic acid-cleaving activity (Jinek et al., 2014 & Eslami-Mossallam et al., 2022).

In general, sgRNAs that bind to 20 nucleotide target sequences can direct Cas9 to detect genomic locations. However, Hsuet et al. observed that sgRNAs with +85 nucleotide tracrRNA tails boosted Cas9 activity and generated more mutations. They also noticed that both concatenated and interspaced two base mismatches, in the proximal region of PAM, significantly decreased Cas9 activity. This impact was increased to encompass three concatenated mismatches (Hsu et al., 2013), and in most genes, three or more inter-spaced and five concatenated mismatches were found to inhibit Cas9 cleavage activity (Hsu et al., 2013 & Eslami-Mossallam et al., 2022). Furthermore, the outcomes of additional research revealed that severely shortened guide RNA would also cause Cas9 to lose cleavage function (Kiani et al., 2015). Therefore, sgRNA design improvement and careful target site selection are required to deliver greater Cas9 cleavage efficiency. Additionally, it was demonstrated that extension of the sgRNA sequence at the 3′ end improved DNA cleavage activity (Jinek et al., 2013). Theoretically, additional number of sgRNA /Cas9 complexes could encourage greater editing effectiveness. The unavoidable complementarity of nonspecific regions in the genome, however, could result in off-target effects if sgRNA/Cas9 complexes are produced in excess (FU et al., 2013 & Eslami-Mossallam et al., 2022). Thus, the quantities of sgRNA and Cas9 should be taken into account to increase on-target mutation rates as well as the activity of Cas9.

5.3.1.3. Target site selection and sgRNA design

Among the potential reasons of CRISPR/Cas9 system's pitfalls, sgRNA design is of utmost importance. Cas9/sgRNA complexes can be used for genome editing or catalytically inactive Cas9 (dCas9)/sgRNA complexes can be used for gene regulation because CRISPR/Cas9 systems are extremely programmable. These applications necessitate the development of effective, precise sgRNAs. Rational sgRNA design still poses a significant difficulty, as it has to take into account a variety of factors (Eslami-Mossallam et al., 2022). Previously, it was thought that the presence of PAM plus an adjacent complementary target sequence would allow Cas9/sgRNA complexes to cleave double-stranded DNA. Numerous studies, however, revealed that some sgRNAs were ineffective or even inert (Wang et al., 2014, Doench et al., 2014 & Mehravar et al., 2019). A number of sequence characteristics in and around the target sequences that predict sgRNA efficiency have been discovered as a result of the growing body of experimental data on the application of CRISPR/Cas9 systems for genome engineering (Peng et al., 2019).

The 5' end of sgRNAs that append a G (guanine) (for example, GX19NGG) is required for expression from a U6 promoter (Wang et al., 2014). Aside from that, G is preferred in the first or second position nearest to PAM, which could help Cas9 load (Wang et al., 2014), but C (cytosine) is significantly unfavourable in the same places. Third, the presence of T (thymine) at the four nucleotide sites next to PAM is undesired because numerous U (uracil) in sgRNA results in reduced gRNA expression (Xu et al., 2015). G is preferable in the PAM-distal region, and A(adenine) is preferred in the centre of sgRNA (Wang et al., 2014). Overall, sgRNAs that are G-rich and A-depleted are more stable and effective (Moreno et al., 2015). Additionally, the SpCas9 PAM has been revealed to have novel characteristics that affects gRNA activity in a

reproduceable manner. For instance, there is a bias for C and against T in the variable nucleotide of NGG (Doench et al., 2014 & Hussain et al., 2019). SpCas9 works best when used with a long PAM sequence of CGGH to produce DSBs. On the other hand, TGGG exhibits the least activity (Doench et al., 2014 & Ishino et al., 2018). Purines are also favoured at the majority of sites in sgRNAs (Xu et al., 2015). The majority of the nucleotides in the spacer region in the CRISPR/dCas9 system collectively contribute to sgRNA efficiency, in contrast to the "seed sequences" that primarily influence sequence preferences in the CRISPR/Cas9 system (Xu et al., 2015). With so many criteria being put forth, the creation of sgRNAs is now made easier by an increasing number of computer tools. The majority of the sgRNA design tools available today either support the SpCas9 system or a number of other orthogonalCas9 systems from various bacterial species (Ishino et al., 2018).

5.3.1.4. Function of top14

Another possible reason for our inability to generate transgenic tomato plants could be that top14 is essential for growth and development processes.

Shoot regeneration happens, as shoots and small leaves developed (see figure 20. (d)), therefore the problem is with root formation. A previous -PhD student- Zahara Medina Calzada, identified LOW PHOSPHATE ROOT1 (LPR1) as a target for miRNA top14 (not published) in roots. It has also been proposed that in *Arabidopsis thaliana* the direct contact between low phosphate (Pi) medium and primary root tip led to inhibition of root growth while loss of function mutations in (LPR1) reduce this inhibition (Svistoonoff et. al. 2007). In wild type tomato, miRNAtop14 is expressed at high level in root, which downregulates LPR1 but if the CRISPR/Cas9 approach deleted miRNA top14, there would be no top14 expression. As a result, LPR1 would be present at a higher level in root which could lead to inhibition of roots formation. Although, in

principle, this would only be an issue in low-Pi medium, since LPR1 level is even higher than in wild type plants, it could interfere with root growth.

In order to assess the effect of the level of (Pi) on formation and growth of roots, we could increase the level of phosphate (Pi) in the regeneration medium. Wild type plants, with normal level of LPR1, are only sensitive to low Pi medium, but higher level of LPR1 may lead to sensitivity to even higher level of Pi, since low level of LPR1 – in lpr1 plants – can tolerate low level of (Pi).

Furthermore, there are several examples that miRNAs can affect organ development, such as miR164 (Laufs et al., 2004; Nikovics et al., 2006). It has been reported that the boundary domains around the organs at Arabidopsis floral and shoot apical meristems are regulated by miR164 through targeting the NAC domain family of transcription factors (Laufs et al., 2004; Nikovics et al., 2006).There are three miR164 genes: miR164A, miR164B, and miR1640C. MiR164A is important in serration of the leaf margin, miR164C is controlling the production of petal number, and it is not clear what specific role miR164B has, as the mutant plant lacking miR164B does not show any modifications in aerial organs (Mallory et al, 2004). Many abnormalities have been seen in the development of Arabidopsis, such as extra petals, misshapen leaves and missing sepals due to the expression of a miR164 resistant cuc1(Mallory et al., 2004).

In order to assess whether mirtop14 is indeed playing a role in the development, growth and differentiation of plants, we could design a control construct that expresses Cas9 and a pair of guides RNAs targeting a gene that is not involved in growth and development (e.g., a resistance gene). If transgenic gene edited plants can be regenerated with that construct but not with the top14 targeting construct, then it could

be concluded that top14 plays an important role in growth and development. If no transgenic plant would be regenerated with the control construct, then there is a technical issue with the protocol.

5.3.2. Assessing the Activity of the CRISPR Constructs

Considering that off-target effect is one of the possible reason that we were not able to regenerate a whole plant after transforming the Cas9/gRNA construct, it is important to reflect on how we would do it differently in the future. We generated only one construct, but it would be beneficial to generate several different constructs, all aiming at deleting miRNAtop14 but targeting different target sites. After building different Cas9-sgRNAs constructs, it could save time and effort to assess their efficiency. CRISPR constructs can be transiently expressed, before going through stable transformation, which would then cause DNA cleavage and mutations. Transient expression assays provide a practical and convenient tool for basic research in plant biology (Reed et al., 2018). Transient expression assays have been increasingly more effective in recent years for characterizing unknown gene function. They have been developed for gene function studies (Hellens et al.,2005; Lee and Yang., 2006) and have also proved helpful for assessing the activity of gene constructs before undertaking stable transformation (Sparkes et al., 2006). The dominant technology used for transient expression is *Agrobacterium* infiltration (Fei et al., 2021). After infiltrating the Agrobacterium strains harbouring the different CRISPR constructs, the infiltrated tissue is analysed to determine if the intended edits were made and if there are unwanted edits elsewhere (off-targets). The genomic sequence at the target site can be analysed using various methods such as PCR, Southern blotting, Sanger sequencing and Next generation sequencing (NGS) (Pribylova et al., 2022).

To characterize the gene-edited tissue, the first step is extracting the genomic DNA from the infiltrated tissue. This can be done using various methods, but a common method is to use a commercial kit designed for this purpose. Once the DNA is extracted, it can be analysed using PCR. It is a commonly used method of plant genetics to amplify the region of DNA containing the target sequence, which can then be sequenced to confirm that the desired edits have been made. To do this, the DNA is first amplified using primers that flank the target sequence. The amplified DNA is then sequenced by Sanger sequencing (Veillet et al., 2019).

If the intention is deletion, the first step towards confirmation of CRISPR success is to use PCR, to verify that the region was deleted. Primers should be used that can amplify a region that includes the potentially deleted sequence and should extend at least 100 base pairs on either side of the deletion (Demeke & Dobnik, 2018). The amplicon (PCR product) should be large enough that you can easily see it on an agarose gel (Coulther et al., 2019). The next step of the verification process is to use sanger sequencing to verify that the intended region has been deleted (Manghwar et al., 2019).

Also, next generation sequencing (NGS) is one of the recommended methods for investigation of CRISPR edits (Pribylova et al., 2022). This process can screen many sgRNAs (CRISPR constructs) to find the most efficient one and it can also detect off-target modifications. Once the most efficient sgRNA (CRISPR construct) is found, it can be used for stable transformation. The reason that the regenerated shoots were not tested for Cas9- induced mutations by PCR or Sanger sequencing is that we did not want to lose them as their number were very limited. The regenerated shoots were very short (2-4 cm in height) and the size of the leaves were very small also the number of them were limited. Every time they were kept to be bigger so we can have at least

one leaf to be tested without losing the whole plant but unfortunately they did not survive.

Chapter 6

Summary and general discussion

Following the discovery that miRNAs are ubiquitous molecules in animals and plants at the beginning of the century, the last 15 years have seen an explosion in both the identification of new miRNA species and the widening of our understanding of miRNA biology. While many “traditional” miRNAs have been identified in many plant species and the general principles of miRNA processing and mode of action have already been elucidated, current research is now focusing on the characterization of less conventional miRNAs, which can also help to uncover some less well understood aspects of miRNA biology, such as the regulation of biogenesis. In the current work, we have focused on studying one of those less conventional miRNAs, miRNAtop14, a miRNA with an intron in between its miRNA and miRNA* sequences in both tomato and *Nicotiana benthamiana*, the plant species where it was first discovered (Moxon et al., 2008 and Baksa et al., 2015).

A large number of plant reference genomes, as well as other sequence collections such as ESTs, are now available. We have been able to determine that the MIRtop14 gene is present across members of the *Solanaceae* family, as well as at least in one genus of the *Convolvulaceae* family, the closest family to *Solanaceae* within the *Solanales* order. However, despite of searching–in all the orders close to *Solanales*, it has not been identified beyond this linage.

Aside from the in-silico prediction, the presence of MIRtop14 locus in the genome, as well as the production of mature miRNAtop14 have been both experimentally confirmed in the species *Solanum lycopersicum, Nicotiana benthamiana, Petunia axillaris* and *Ipomoea nil*, the first three belonging to *Solanaceae* and the last one to *Convolvulaceae* family.

These data suggest that miRNAtop14 is a *Solanales* specific miRNA, making it a promising target for further research, because the economic importance of this order. Additional research could unveil some unique characteristic of *Solanales* that could be useful in agricultural development.

The biological role of miRNAtop14 has also been studied. RLM-RACE experiments have revealed that miRNAtop14 directs the cleavage of LPR protein-coding mRNA in both *S. lycopersicum* and *N. benthamiana*, including the two LPR paralogues present in the last species.

In *Arabidopsis*, LPR proteins are involved in the pathway that arrests root growth under low phosphate conditions (Svistoonoff et al. 2007 and Müller et al*.,* 2015). However, despite its wide distribution in the plant kingdom (Ming, R. et al., 2013) there are currently no reports about the role of LPR in other plants, neither whether its function is confined to root or LPR may also be playing another task in other tissues, since this protein is known to be expressed in other parts of the plant apart from root (Svistoonoff et al., 2007).

Based on the known function of its target in Arabidopsis, and supposing such function would be conserved in *Solanales*, it can be hypothesised that miRNAtop14 expression would change in response to phosphate stress conditions in order to fine tune the required levels of its target. It would be interesting to test this postulate, which could be done by growing plants in normal vs. low phosphate conditions and measuring the levels of both miRNAtop14 and LPR by, for example, Northern and Western blot, respectively. Besides, measurement and comparison of root length between plants grown in both phosphate conditions would indicate if the mechanism of root growth arrest under scarce phosphate observed in Arabidopsis also applies to *Solanales*.

Our experiments have indicated that the deletion of MIRtop14 seems to affect the root development. As mentioned above, during the working on this project the experiments have found a target for top14 so called LOW PHOSPHATE ROOT1 (LPR1) in roots (not published) and the analyses by GUS staining seems to show a high expression level of miRNAtop14 in roots as explained in a chapter 3. Svistoonoff and others (2007) have proposed that in *Arabidopsis thaliana* there is a link between low phosphate (Pi) medium and inhibition of root growth while loss of function mutations in (LPR1) reduce this inhibition. In wild type plants miRNAtop14 is expressed at high level in root, and downregulates LPR1. However, because the deletion of MIRtop14 by CRISP-Cas9, there would be no miRNAtop14, therefore, LPR1 would be present at a higher level in root which may lead to inhibition of roots formation. In wild type plants, this would only be an issue in low-Pi medium, but LPR1 level is even higher than in wild type plants, thus it could interfere with root growth.

From comparing the results of the study of miRNAtop14 (not published) conservation and those from studying its interaction with LPR a surprising yet interesting observation is inferred; while miRNAtop14 sequence is very conserved among all Solanales, with only the 5' end nucleotide changing in some cases, this is not the case for miRNAtop14 target site in LPR, up to the point where there is no target site to be found at all in Ipomoea LPR (Zahara Medina Calzada, 2017; not published).

Generally, miRNA and target sites are supposed to coevolve in order to maintain its interaction and thus the miRNA function (Xie et al, 2017; Zhang et al., 2016). In fact, when a miRNA does not have any interaction with a mRNA target it does not achieve conservation because it tends to be quickly lost by genetic drift (Xie et al, 2017; Park and Beal, 2019). This raises the question of why miRNAtop14 maintains such a high conservation in all harbouring species, including those ones

where it does not interact with LPR, where it would be expected to be lost or at least degenerated.

The most obvious explanation would probably be that it may have another target across all these species, and that LPR targeting has been newly acquired in the last diverging branch of Solanaceae which comprises Solanum, Capsicum and Nicotiana genera. One way to try to solve this question would be to carry out RLM-RACE experiments including additional putative targets not already tested. Besides, it cannot be ruled out that miRNAtop14 causes translational repression of mRNA without cleavage. This kind of interaction would not be revealed by RLM-RACE and should be examined by an alternative method such as measuring target protein levels in different conditions and comparing them with target mRNA and miRNAtop14 levels. In any case, the door remains open to research the role of miRNAtop14 beyond LPR regulation, as an additional target is likely to exist.

The primary aim of this study was to investigate the biological function of miRNAtop14. CRISPR-Cac9 technology was applied to knock out MIRtop14 to identify the biological role. Although, there was transgenics tomato plantlets after transformation with the CRISPR-Cas9 construct, the shoots died when they reached 2-4 cm. Despite the many advantages of this system, there are some challenges to the current Cas9-based tools, such as off-target effects, Cas9 activity, target site selection and sgRNA design, and that top14 is essential for growth and development

The CRISPR/Cas9 system has so far demonstrated that it is a reliable and adaptable tool for gene regulation and genome editing. However, it became clear through further study of CRISPR that this technique was not as simple as first believed. Recent studies have examined various elements that may have an impact on the CRISPR/Cas9 system. We may make greater use of this technique, as well as

increase its specificity and efficiency, by addressing its potential pitfalls. (Stovicek et al., 2017; Modrzejewski et al., 2019; Waldrip et al., 2020).

An important pitfall in all CRISPR genome editing applications with Cas9 proteins is the potential for introducing unexpected mutations, frequently at off-target Cas9 binding sites. A number of studies have demonstrated that this system can induce a substantial amount of off-target mutagenesis. (Pattanayak et al., 2013; Kuscu et al., 2014; Polstein et al., 2015; Ma et al., 2016; Park and Beal, 2019). One significant source of these undesirable results lies in the predisposition of Cas9/sgRNA complexes to attach to off-target locations in the genome that include mismatches, which may cause unexpected DNA editing events and mutagenesis (Pattanayak et al., 2013; Kuscu et al., 2014; Polstein et al., 2015; Ma et al., 2016; Park and Beal, 2019).

Additionally, the Cas9 expression level can affect the Cas9 (or dCas9) off-targeting binding kinetics, as genome-wide studies have indicated that higher expression of Cas9 is linked to increased off-target binding than lower Cas9 expression levels (Waldrip et al., 2020). Also, imperfect base pairing at off-target binding sites probably locks Cas9 into a structural conformation that is not capable of cleavage, which explains why Cas9 binds more off-target sites than it cleaves (Kuscu et al., 2014; Wu et al., 2014; Tsai et al., 2015).

To Reduce off-target Binding and Cleavage by Cas proteins, it is crucial to precisely design sgRNAs to minimize mutations that are not desired. Consequently, before creating and optimizing sgRNA expression vectors, it is recommended that using web-based tools to identify suspected sgRNA target sites and determine whether there might be any potentially problematic off-targeting events. Numerous in silico tools have been designed and are thoroughly assessed such as Off-Spotter,

Cas-offinder and CRISPR-ERA. (Lee et al., 2016; Stovicek et al., 2017; Raschmanová et al., 2018; Deaner and Alper, 2019; Liu et al., 2019; Ding et al., 2020; Manghwar et al., 2020; Naeem et al., 2020; Sledzinski et al., 2020).

Lastly, it has been reported that off-targeting can affect plant development and growth and the issue of off-targeting in plants still needs to be addressed systematically (Feng et al., 2014; Gao and Zhao, 2014; Zhang et al., 2014). Genomic sequencing was recommended to assess if, and to what extent, CRISPR/Cas9 causes off-targeting in tomato because in the animal field a relatively high off-target rate was reported (Carroll, 2013).

Alternatively, the importance of top14 in development and growth could explain the premature death of the transgenic lines. Many studies have shown that microRNAs intrinsically linked to the regulation of many genes that control the development, growth and differentiation of plants. Many abnormalities have been seen in the development of Arabidopsis, such as extra petals, misshapen leaves and missing sepals due to the expression of a miR164 resistant cuc1 (Mallory et al., 2004). For future work, to prove the essentiality of miRNAtop14, it is important to use a control CRISPR-Cas9 construct to target a gene that is not involved in development and growth.

Although still partially incomplete, the first steps have been taken towards the full characterization of MIRtop14, including its biological role and possible post-transcriptional regulation. The completion of this study would increase the current knowledge about miRNA regulation, a field of study just starting to emerge, and would shed light to the increasingly evident crosstalk between different RNA pathways, as is the case of splicing and miRNA biogenesis. In addition, the identification of a MIR with

such an unusual exon-intron structure could open the door to search for more miRNA produced by unconventional pri-miRNAs, including other intron-split miRNAs.

References

Addo-quaye, C., Eshoo, T. W., Bartel, D. P. & Axtell, M. J. (2009). Endogeneous siRNA and microRNA targets identified by sequencing of the Arabidopsis degradome. *Curr. Biol*. 18, 758–762

Aida, M., Ishida, T., Fukaki, H., Fujisawa, H., and Tasaka, M. (1997). Genes involved in organ separation in Arabidopsis: an analysis of the cup-shaped cotyledon mutant. *Plant Cell*.June,9(6) 841-57.

Akhtar, M. M., Micolucci, L., Islam, M. S., Olivieri, F. & Procopio, A. D. (2015). Bioinformatic tools for microRNA dissection. *Nucleic Acids Res.* gkv1221-

Alt, F. W., Bothwell, A.L., Knapp,M., Siden, E., Mather, E., Koshland, M., Baltimore, D. (1980). Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3' ends. *Cell* 20(2)**,** 293– 301

Ambros, V., Bartel, B., Bartel, D. P., Burge, C. B., Carrington, J. C., Chen, X., ... Tuschl, T. (2003). A uniform system for microRNA annotation. *RNA*, *9*(3), 277–279.

Ambros, Victor; Bartel, Bonnie; Bartel, David P.; Burge, Christopher B.; Carrington, James C.; Chen, Xuemei; Dreyfuss, Gideon; Eddy, Sean R.; Griffiths-Jones, Sam; Marshall, Mhairi; Matzke, Marjori; Ruvkun, Gary; Tuschl, T. (2004). A uniform system for microRNA annotation. *RNA* 9, 277–279

Amitai G. and Sorek R. (2016). CRISPR -Cas adaptation: insights into the mechanism of action. Nat. Rev. *Microbiol.14,* 67-76

Arnoult, N., Correia, A., Ma, J., Merlo, A., Garcia-Gomez, S., Maric, M., et al. (2017). Regulation of DNA repair pathway choice in S and G2 phases by the NHEJ inhibitor CYREN. Nature 549, 548–552. doi: 10.1038/nature24023

Aphasizheva, I., Alfonzo, J., Carnes, J., Cestari, I., Cruz-Reyes, J., Göringer, H. U., ...& Aphasizhev, R. (2020). Lexis and grammar of mitochondrial RNA processing in trypanosomes. *Trends in parasitology*, *36*(4), 337-

Aukerman, M. and Sakai, H. (2003).Regulation of flowering time and floral organ identity by microRNA and its APETALA2-like target genes.plant cell online 15,2730.

Axtell, M. J., Westholm, J. O. & Lai, E. C. (2011). Vive la diff rence: biogenesis and evolution of microRNAs in plants and animals. Genome Biol. 12, 221

Baksa, I. Nagy, T., Barta, E., Havelda, Z., Varallyay, E., Silhavy, D., Burgyan, J., and Szittya, G. (2015). Identification of Nicotiana benthamiana microRNAs and their targets using high throughput sequencing and degradome analysis. *BMC Genomics* 16**,** 1025.

Baldi, L., Hacker, D. L., Adam, M., & Wurm, F. M. (2007). Recombinant protein production by large-scale transient gene expression in mammalian cells: state of the art and future perspectives. *Biotechnology letters*, *29*(5), 677-684.

Barbosa-Morais, N. L. et al. (2012). The Evolutionary Landscape of Alternative Splicing in Vertebrate Species. *Science* (80). 338

Barrangou, R. & Marraffini, L. (2014). CRISPR-Cas Systems: Prokaryotes Upgrade to Adaptive Immunity. *Molecular Cell* 54, 234-244.

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

Bartel, D.**P**. (2009). MicroRNA Target Recognition and Regulatory Functions.*Cell* 136, 215–233.

Belhaj, K., **Chaparro-Garcia, A**., **Kamoun, S**., **Patron, N**.**J**., **and Nekrasov, V**. **(**2015).Editing plant genomes with CRISPR/Cas9, In Current Opinion in *Biotechnology*.76-84

Bernheim, A., Calvo-Villamanan, A., Basier, C., Cui, L., Rocha, E. P. C., Touchon, M., et al. (2017). Inhibition of NHEJ repair by type II-A CRISPR- Cas systems in bacteria. Nat. Commun. 8:2094. doi: 10.1038/s41467-017-02350-1

Berget, S. M., Moore, C. & Sharp, P. A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. U. S. A.* 74, 3171–5

Bologna, N. G., Iselin, R., Abriata, L. A., Sarazin, A., Pumplin, N., Jay, F., et al. (2018). Nucleo-cytosolic Shuttling of ARGONAUTE1 prompts a revised model of the plant MicroRNA Pathway. Mol. Cell 69, 709–719.

Bolotin, A., Quinquis, B., Sorokin, A. and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 2551–2561

Brack, C. & Tonegawa, S. (1977). Variable and constant parts of the immunoglobulin light chain gene of a mouse myeloma cell are 1250 nontranslated bases apart. *Proc. Natl. Acad. Sci. U. S. A.* 74, 5652–6

Breathnach, R., Mandel, J. L. & Chambon, P. (1977). Ovalbumin gene is split in c hicken DNA. *Nature* 270, 314–319

Brodersen, P., and Voinnet, O. (2006). The diversity of RNA silencing pathway in plants.Trends in Genetics 22, 268-280.

Brody, E. & Abelson, J. (1985). The quot; spliceosome& quot; yeast pre- messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science (80).* 228**,**

Brooks, C., Nekrasov, V., Lippman1, Z., and Van, J., (2014). Efficient gene editing in tomato in the first generation using the CRISPR/Cas9 system.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and Van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defence in prokaryotes. *Science* 321, 960–964

Çakır, Ö., Arıkan, B., Karpuz, B. & Turgut-Kara, N., (2021). Expression analysis of miRNAs and their targets related to salt stress in Solanum lycopersicum H-2274. Biotechnology & Biotechnological Equipment, 35(1), pp.283-

Campbell, M., Haas, B., Hamilton, J., Mount, S., Buell, C.(2006). Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis.BMC Genomics.3277

Carroll, D. (2013) Staying on target with CRISPR-Cas. *Nature biotechnology* **31:** 807-809

Casini, A, Storch, M., Baldwin, Geoffrey S., Ellis, T. (2015) Bricks and blueprints: methods and standards for DNA assembly.

Cech, T. (1990). Self-Splicing of Group I Introns. *Annu. Rev. Biochem.* 59, 543–568.

Cech, T. R. (1983). RNA Splicing: Three Themes with Variations. *Cell* 34**,** 713–716.

Charoonnart, P., Tuant, H., Webb, C., Purton, S., Robinson, C., & Saksmerprome, V. (2022). Enhanced production of antiviral dsRNA in the Chlamydomonas reinhardtii chloroplast via a novel convergent promoter expression system.

Chen, X., Liu, J., Janssen, J. M., **Goncalves, M.** (2017). The chromatin structure differentially impacts high-specificity CRISPR-Cas9 nuclease strategies. *Mol. There. Nucleic Acids*. 8, 558-563

Chiasson, D., Giménez-Oya, V., Bircheneder, M., Bachmaier, S., Studtrucker, T., Ryan, J., ...& Parniske, M. (2019). A unified multi-kingdom Golden Gate cloning platform. Scientific reports, 9(1), 1-12.

Chockalingam, K., Peng, Z., Vuong, C. N., Berghman, L. R., & Chen, Z. (2020). Golden Gate assembly with a bi-directional promoter (GBid): A simple, scalable method for phage display Fab library creation. Scientific reports, 10(1), 1-14.

Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. (1977) An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. Cell 12, 1–8

C**huan- Chen, Y., Sheng, J., trang, P., Liu, F.** (2018). Potential Application of the CRISPR/Cas9 Systemagainst Herpesvirus Infections. *Viruses* 10, 1-10

Chuck, G., Candela, H. and Hake, S. (2009). Big impacts by small RNAs in plant development, In Current Opinion. *Plant Biology* 12, 81-86

Clement, K., Rees, H., Canver, M. C., Gehrke, J. M., Farouni, R., Hsu, J. Y., ...& Pinello, L. (2018). Analysis and comparison of genome editing using CRISPResso2. bioRxiv, 392217.

Coelho, M. A., De Braekeleer, E., Firth, M., Bista, M., Lukasiak, S., Cuomo, M. E., & Taylor, B. J. (2020). CRISPR GUARD protects off-target sites from Cas9 nuclease activity using short guide RNAs. Nature communications, 11(1), 1-12.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339 (6121): 819–823.

Consortium T., (2012). The tomato genome sequence provides insights into fleshy f ruit evolution. *Nature* 485: 635-641

Cooper, S., Wadsworth, E. S., Schnaufer, A., & Savill, N. J. (2022). Organisation of minicircle cassettes and guide RNA genes of Trypanosoma brucei. RNA, rna-079022

Coulther, T. A., Stern, H. R., & Beuning, P. J. (2019). Engineering polymerases for new functions. Trends in biotechnology, 37(10), 1091-

Crozet, P., Navarro, F. J., Willmund, F., Mehrshahi, P., Bakowski, K., Lauersen, K. J., ... & Lemaire, S. D. (2018). Birth of a photosynthetic chassis: a MoClo toolkit enabling synthetic biology in the microalga Chlamydomonas reinhardtii. ACS synthetic biology, 7(9), 2074-2086. 10.1021/acssynbio.8b00251\

Dalmadi, Á., Gyula, P., Bálint, J., Szittya, G., & Havelda, Z. (2019). AGO-unbound cytosolic pool of mature miRNAs in plant cells reveals a novel regulatory step at AGO1 loading. Nucleic acids research, 47(18), 9803-9817

Dalmay, T. (2008). Identification of genes targeted by microRNAs. *Biochemical Society Transactions,* 1194-1196.

Dang, Y., Jia, G., Choi, J., Ma, H., Anaya, E., Ye, C., Shankar, P., W. Haoquan. (2015). Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. Genome Biol., 16 (2015), p. 280

Deaner, M., and Alper, H. S. (2019). Enhanced Scale and Scope of Genome Engineering and Regulation Using CRISPR/Cas in Saccharomyces cerevisiae. FEMS Yeast Res. 19, foz076. doi:10.1093/femsyr/foz076

Delidow, B. C., Lynch, J. P., Peluso, J. J., & White, B. A. (1993). Polymerase chain reaction : basic protocols. Methods in molecular biology (Clifton, N.J.), 15, 1–29.

Demeke, T., & Dobnik, D. (2018) Critical assessment of digital PCR for the detection and quantification of genetically modified organisms. Analytical and Bioanalytical Chemistry, 410(17), 4039-4050.

Devaux, Y., Stammet, P., Friberg, H., Hassager, C., Kuiper. M.A., Wise, M.P., and Nielsen, N. (2015). Biomarker subcommittee of TTM trial (Target Temperature Management After Cardiac Arrest, NCT01020916). *Crit Care*

Di Felice, F., Micheli, G. & Camilloni, G., (2019).Restriction enzymes and their use in molecular biology: An overview. Journal of biosciences, 44(2), pp.1-8.10.1007/s12038-019-9856-8

Dimitriu, T., Kurilovich, E., Łapińska, U., Severinov, K., Pagliara, S., Szczelkun, M. D., & Westra, E. R. (2022). Bacteriostatic antibiotics promote CRISPR-Cas adaptive immunity by enabling increased spacer acquisition. Cell host & microbe, 30(1), 31-40.

Ding, W., Zhang, Y., and Shi, S. (2020). Development and Application of CRISPR/Cas in Microbial Biotechnology. Front. Bioeng. Biotechnol. 8, 711.

Doench JG, Hartenian E, Graham DB, Tothova Z,Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ& Root DE. (2014) Rational design of highly actives gRNAs for CRISPR-Cas9-mediated gene inactivation.Nat Biotechnol32, 1262–1267.

Dörnbrack, K., Beck, J., & Nassal, M. (2022). Relaxing the restricted structural dynamics in the human hepatitis B virus RNA encapsidation signal enables replication initiation in vitro. PLoS pathogens, 18(3), e1010362.

Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R., Hood, L. (1980). Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. *Cell* **20,** 313–319

Engler, C., Carola; Marillonnet, Sylvestre (2014). Golden Gate cloning. Method in Molecular Biology.

Engler C, Gruetzner R, Kandzia R, Marillonnet S. (2009). Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzmes*. PLoS One. 4(5):e5553*

Engler C., Kandzia R., and Marillonnet, S. (2008). A One Pot, One Step, Precision Cloning Method with High Throughput Capability". 11-05

Eslami-Mossallam, B., Klein, M., Smagt, C. V., Sanden, K. V., Jones, S. K., Hawkins, J. A., ... & Depken, M. (2022). A kinetic model predicts SpCas9 activity, improves off-target classification, and reveals the physical basis of targeting fidelity. Nature communications, 13(1), 1-10.

Fei Y, Nyik o T, Molnar A. 2021. Non-perfectly matching small RNAs can induce stable and heritable epigenetic modifications and can be used as molecular markers to trace the origin and fate of silencing RNAs. Nucleic Acids Research 49: 1900–1913.

Feng Z, Mao Y, Xu N, Zhang B, Wei P, Yang DL, Wang Z, Zhang Z, Zheng R, Yang L, Zeng L, Liu X, Zhu JK (2014). Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proc Natl Acad Sci U S A*

Filichkin, S., Priest, H. D., Megraw, M. & Mockler, T. C. (2015). Alternative splicing in plants: Directing traffic at the crossroads of adaptation and environmental stress. Current Opinion. *Plant Biology* 24, 125–135

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

Frendewey, D. & Keller, W. (1985). Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. *Cell* **42,** 355–67

Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D,Joung JK & Sander JD. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleasesin human cells.Nat Biotechnol31, 822–826

Gantner, J., Ordon, J., Ilse, T., Kretschmer, C., Gruetzner, R., Löfke, C., ...& Stuttmann, J. (2018). Peripheral infrastructure vectors and an extended set of plant parts for the Modular Cloning system. PloS one, 13(5), e0197185.

Gao Y, Zhao Y (2014) Specific and heritable gene editing in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* 111: 4357-4358

Gao, Y., Liu, H., Zhang, C., Su, S., Chen, Y., Chen, X., ...& Gan, J. (2021). Structural basis for guide RNA trimming by RNase D ribonuclease in Trypanosoma brucei. Nucleic Acids Research, 49(1), 568-583.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magad n, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA 67- 71.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria E2579-E2586.

Gelvin B. S. (2003) *Agrobacterium*-Mediated Plant Transformation: the Biology behind the “Gene- Jockeying” Tool. *Microbiology and Molecular Biology* Reviews 67(1): 16–37

German, M. A., Luo, S., Schroth, G., Meyers, B. C. & Green, P. J. (2009). Construction of Parallel Analysis of RNA Ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. Nat. Protoc. 4, 356–362

Ghadessy, F. J., Ramsay, N., Boudsocq, F., Loakes, D., Brown, A., Iwai, S., ...& Holliger, P. (2004). Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. Nature biotechnology, 22(6), 755-

Gilbert, W. (1978). Why genes in pieces? Nature 271, 501–501.

Gilbert, W. (1987). The exon theory of genes. *Cold Spring Harb. Symp. Quant. Biol.* 52, 901–5

Glazov, E.A., Cottee, P.A., Barris, W.C., Moore, R.J., Dalrymple, B.P., Tizard, M.L. (2008). A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach. Genome Res. 18, 957–964.

Golden Gate Assembly, New England Biolabs Inc. accessed 23/8/2015 "https://www.neb.com/applications/cloning-and-synthetic-biology/dna-assembly-and- cloning/golden-gate-assembly

Greer, C. L., Peebles, C. L., Gegenheimer, P. & Abelson, J. (1983). Mechanism of Action of a Yeast RNA Ligase in tRNA Splicing. *Cell* **32,** 537–546

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature.* 35–40.

Guo, Q., Liu, Q., Smith, N. A., Liang, G., & Wang, M. B. (2019). RNA Silencing in Plants: Mechanisms, Technologies and Applications in Horticultural Crops. Current genomics, 17(6), 476–489.

Hajieghrari, B., & Farrokhi, N. (2021). Plant RNA-mediated gene regulatory network. Genomics.https://pdf.sciencedirectassets.com/272501/1-s2.0-S0888754321X00079/1-s2.0-S0888754321004481/main.pdf?X-Amz-Security-

Harrington, L.B., Ma, E., Chen, J.S., Witte, I.P., Gertz, D., Paez-Espino, D., Al-Shayeb, B., Kyrpides, N.C., Burstein, D., Banfield, J.F. & Doudna, J.A., (2020). A scoutRNA is required for some type V CRISPR-Cas systems. Molecular cell, 79(3), pp.416-424.

Hellens, R.P., Allan, A.C., Friel, E.N., Bolitho, K., Grafton, K., Templeton, M.D., Karunairetnam, S., Gleave, A.P. and Laing, W.A. (2005) Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. Plant Methods, 1, 13.

Heller, R. C., Chung, S., Crissy, K., Dumas, K., Schuster, D., & Schoenfeld, T. W. (2019). Engineering of a thermostable viral polymerase using metagenome-

derived diversity for highly sensitive and specific RT-PCR. Nucleic acids research, 47(7), 3619-3630.

Henderson. IR., Zhang, X., Lu, C., Johnson, L., Meyers, B.C., Green, P.J., and Jacobsen, S.E. (2006). Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. Nat Genet 38: 721–725

Hernanz-Koers, M., Gandía, M., Garrigues, S., Manzanares, P., Yenush, L., Orzaez, D., & Marcos, J. F. (2018). FungalBraid: A GoldenBraid-based modular cloning platform for the assembly and exchange of DNA elements tailored to fungal synthetic biology. Fungal Genetics and Biology, 116, 51- 61.

Hernanz-Koers, M., Gandía, M., Garrigues, S., Manzanares, P., Yenush, L., Orzaez, D., & Marcos, J. F. (2018). FungalBraid: A GoldenBraid-based modular cloning platform for the assembly and exchange of DNA elements tailored to fungal synthetic biology. Fungal Genetics and Biology, 116, 51-61.

Hinz, A. J., Stenzler, B., & Poulain, A. J. (2021).Golden gate assembly of aerobic and anaerobic microbial bioreporters. Applied and environmental microbiology, 88(1), e01485-21.

Hoegger, P. J., Kilaru, S., James, T. Y., Thacker, J. R. and K es, U. (2006). Phylogenetic comparison and classification of laccase and related multicopper oxidase protein sequences. *FEBS J.* 273**,** 2308–26

Hong ZHANG, B., Ping PAN, X., Lian WANG, Q., Cobb, G. P. & Anderson, T. A. (2005). Identification and characterization of new plant microRNAs using EST analysis. *Cell Res.* 15, 336–360

Hu, R., Qi, G., Kong, Y., Kong, D., Gao, Q., and Zhou, G. (2010). Comprehensive analysis of NAC domain transcription factor gene family in Populus trichocarpa. *BMC Plant Biol.*

Hussain, A., Imran, Q.M. & Yun, B.W., (2019).CRISPR/Cas9-Mediated Gene Editing in Grain Crops.In Recent Advances in Grain Crops Research (p. 9).London, UK: IntechOpen

Hussain, W., Mahmood, T., Hussain, J., Ali, N., Shah, T., Qayyum, S., et al. (2018). CRISPR/Cas system: a game changing genome editing technology, to treat human genetic diseases. Gene 685, 70–75. doi: 10.1016/j. gene.2018.10.072

Hsu PD, Scott DA, Weinstein JA, Ran FA,Konermann S, Agarwala V, Li Y, Fine EJ, Wu X,Shalem Oet al.(2013) DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol31,827–832.28

Ibal, J. C., Pham, H. Q., Park, C. E., & Shin, J. H. (2019). Information about variations in multiple copies of bacterial 16S rRNA genes may aid in species identification. PLoS One, 14(2), e0212090.

Indriani, S. (2020). Cloning Techniques and Its Application in Food Industry: A Short Review. Genetic Engineering, 15, 12

Ishino, Y., Krupovic, M., & Forterre, P. (2018). History of CRISPR-Cas from encounter with a mysterious repeated sequence to genome editing technology. Journal of bacteriology, 200(7), e00580-

Itaya, A. Bundschuh, R. Archual, A., Joung, J., Fei, Z., Dai, X., Zhao, P., Tang, Y., Nelson, R., and Ding, B. (2008). Small RNAs in tomato fruit and leaf development. BBA - *Gene Regulatory Mechanisms* Volume 1779, Pages 99-107

Jansen, R., Embden, J., Gaastra, W., Schouls, L. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol*., 43, pp. 1565-1575

Javed, M. R., Sadaf, M., Ahmed, T., Jamil, A., Nawaz, M., Abbas, H., & Ijaz, A. (2018).CRISPR-Cas system: history and prospects as a genome editing tool in microorganisms. Current microbiology, 75(12), 1675-

Jia, F. & Rock, C. D. (2013).MIR846 and MIR842 comprise a cistronic MIRNA pair that is regulated by abscisic acid by alternative splicing in roots of Arabidopsis. Plant Mol. Biol. 81, 447–460.

Jiang, F., and Doudna. J.A. (2017). CRISPR–Cas9 structures and mechanisms Annu. Rev. *Biophys.,* 46, pp. 505-529

Jiang, W. Bikard, D., Cox, D., Zhang, F., Marraffini, L. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems Nat. *Biotechnol*., 31, pp. 233-239

Jin, K., Tian, N., da Silva Ferreira, J. F., Sandhu, D., Xiao, L., Gu, M., ...& Liu, S. (2022). Comparative Transcriptome Analysis of Agrobacterium tumefaciens Reveals the Molecular Basis for the Recalcitrant Genetic Transformation of Camellia sinensis L. Biomolecules, 12(5), 688.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J., Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity Science, 337. pp. 816-821

Jinek M, East A, Cheng A, Lin S, Ma E & Doudna J. (2013) RNA-programmed

genome editing in humancells. eLife2, e00471.

Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E,Ma E, Anders C, Hauer M, Zhou K, Lin Set al.(2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation.Science343,1247997

Juarez, M., Kui, J., Thomas, J., Heller, B., and Timmermans, M. (2004). microRNA-mediated repression of rolled leaf specifies maize leaf polarity. *Nature* 428, 84-88

Jones-Rhoades, M. W. & Bartel, D. P. (2004). Computational Identification of Plant MicroRNAs and Their Targets, Including a Stress-Induced miRNA. Mol. *Cell* 14, 787–799

Jones-Rhoades, M. W. (2012) Conservation and divergence in plant microRNAs. *Plant Mol. Biol*. 80, 3–16.

Joung, J.K., and Sander, J.D. (2013). TALENs: a widely applicable technology for targeted genome editing. *NatRev Mol Cell Biol* 14: 49-55

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337 (6096): 816–821.

Jia, F. & Rock, C. D. (2013). MIR846 and MIR842 comprise a cistronic MIRNA pair that is regulated by abscisic acid by alternative splicing in roots of Arabidopsis. *Plant Mol. Biol*. 81, 447–460.

Jones D.H., Doherty A, and Wu H. (2005)**.** Review of methodologies and a protocol for the *Agrobacterium*-mediated transformation of wheat. Plant Methods 1: 5

Kapranov, P. Cheng, J., Dike, S., Nix, D., Duttagupta, R., Willingham, A., Stadler, P., Hertel, J., Hackermuller, J., Hofacker, I., Bell, I., Cheung, E., Drenkow, J., Dumais, E., Patel, S., Helt, G., Ganesh, M., Ghosh, S., Piccolboni, A., Sementchenko, V., Tammana, H., Gingeras, T. (2007).RNA Maps Reveal New RNA Classes and a Possible Function for Pervasive Transcription. *Science* (80). 316, 1484–1488

Kiani S, Chavez A, Tuttle M, Hall RN, Chari R, Ter-Ovanesyan D, Qian J, Pruitt BW, Beal J, Vora Set al. (2015) Cas9 gRNA engineering for genome editing, activation and repression. Nat Methods12,1051–1054.

Kim, W. N., Kim, H. J., Chung, Y. S., & Kim, H. U. (2021). Construction of Multiple Guide RNAs in CRISPR/Cas9 Vector Using Stepwise or Simultaneous Golden

Gate Cloning: Case Study for Targeting the FAD2 and FATB Multigene in Soybean. Plants, 10(11), 2542.https://doi.org/10.3390/plants10112542

Kingston, E. R., & Bartel, D. P. (2021). Ago2 protects Drosophila siRNAs and microRNAs from target-directed degradation, even in the absence of 2′-O-methylation. RNA, 27(6), 710-724.

Klompe, S.E. and Sternberg, S.H., 2018. Harnessing "A Billion Years of Experimentation": the ongoing exploration and exploitation of CRISPR–Cas immune systems. The CRISPR journal, 1(2), pp.141-158.

Kloosterman, W.P. and Plasterk, R.H. (2006**).** The diverse functions of microRNAs in animal development and disease. Dev. Cell 11, 441–450

Kluesner, M. G., Nedveck, D. A., Lahr, W. S., Garbe, J. R., Abrahante, J. E., Webber, B. R., & Moriarity, B. S. (2018). EditR: a method to quantify base editing from Sanger sequencing. The CRISPR journal, 1(3), 239-250.10.1089/crispr.2018.0014

Kornblihtt, A. R., Schor, I. E., Allo, M., Dujardin, G., Petrillo, E., Munoz, M. J. (2013)**.** Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* 14, 153–165.

Kozomara, A. & Griffiths-Jones, S. (2014) MiRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*. 42, D68-73.

Kuscu, C., Arslan, S., Singh, R., Thorpe, J., and Adli, M. (2014). Genome-wide Analysis Reveals Characteristics of Off-Target Sites Bound by the Cas9 Endonuclease. Nat. Biotechnol. 32, 677–683. doi:10.1038/nbt.2916

Lagos-Quintana, M. et al. (2001).Identification of novel genes coding for small expressed RNAs. Science 294, 853–8

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001)**.**Identification of Novel Genes Coding for Small Expressed RNAs.Science 294, 853–858

Laufs, P., Peaucelle, A., Morin, H., and Traas, J. (2004). MicroRNA regulation of the CUC gene is required for boundary size control in Aribidopsis meristems. *Development* 131, 4311.

Lander, E.S. (2016). The heroes of CRISPR Cell, 164 pp. 18-28

Liang X., Potter J., Kumar S., Zou Y., Quintanilla R., Sridharan M., Carte J., Chen W., Roark N., Ranganathan S. (2015). Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J. Biotechnol.* 208:44–53.

Li JF, Norville JE, Aach J, McCormack M, Zhang D, Bush J, Church GM, Sheen J (2013). Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. *Nat Biotechnol* 31**:** 688-691

Liu, B., Xu, H., Miao, J., Zhang, A., Xiaojin, K., Li, W., Zhou, L., Xie, H., Sirois, P., Li, K. (2015), CRISPR/Cas: a faster and more efficient gene editing system. J. Nanosci. *Nanotechnol*., 15, pp. 1946-1959

Liu, L., and Fan, X.D. (2014). CRISPR-Cas system: a powerful tool for genome engineering. Plant Mol Biol 209-218

Lebedenko E., Birikh K., Plutalov O., Berlin Y. (1991 Method of artificial DNA splicing by directed ligation (SDL). *Nucleic Acids Res*. Dec 25; 19(24):6757-61.

Lee, C. M., Cradick, T. J., Fine, E. J., and Bao, G. (2016). Nuclease Target Site

Selection for Maximizing On-Target Activity and Minimizing Off-Target Effects in Genome Editing. Mol. Ther. 24, 475–487. doi:10.1038/mt.2016.1

Lee, M.W. and Yang, Y. (2006) Transient expression assay by agroinfiltration of leaves. In Arabidopsis Protocols (Salinas, J. and Sanchez-Serrano, J.J., eds), pp. 225–229. New York: Humana Press.

Lee, R. C. and Ambros, V. (2001). An extensive class of small RNAs in Caenorhabditis elegans. *Science* 294, 862–4.

Lee, Y., Jeon, K., Lee, J.-T., Kim, S., and Kim, V. N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. *The EMBO Journal*, *21*(17), 4663– 4670.

Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S.H., and Kim, V.N. (2004a). MicroRNA genes are transcribed by RNA polymerase II. 4051–4060.

Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. (2004). Distinct Roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA Silencing Pathways. *Cell,* Volume 117, Issue 1, 69 – 81

Li JF, Norville JE, Aach J, McCormack M, Zhang D, Bush J, Church GM, Sheen J (2013). Multiplex and homologous recombination-mediated genome editing

in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9.Nat Biotechnol 31: 688-691

Li, C., Wong, A. Y., Wang, S., Jia, Q., Chuang, W. P., Bendena, W. G., ...& Hui, J. H. (2018). miRNA-mediated interactions in and between plants and insects. International Journal of Molecular Sciences, 19(10), 3239.

Li, J., Manghwar, H., Sun, L., Wang, P., Wang, G., Sheng, H., ...& Zhang, X. (2019). Whole genome sequencing reveals rare off-target mutations and considerable inherent genetic or/and somaclonal variations in CRISPR/Cas9-edited cotton plants. Plant biotechnology journal, 17(5), 858-868.

Li, L., Veksler-Lublinsky, I., & Zinovyeva, A. (2019). HRPK-1, a conserved KH-domain protein, modulates microRNA activity during Caenorhabditis elegans development. PLoS genetics, 15(10), e1008067.

Li, M., & Yu, B. (2021). Recent advances in the regulation of plant miRNA biogenesis. RNA biology, 18(12), 2087-2096.

Liang X., Potter J., Kumar S., Zou Y., Quintanilla R., Sridharan M., Carte J., Chen W., Roark N., Ranganathan S. (2015). Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. J. Biotechnol. 208:44–53.

Liang, Z., Zhang, K., Chen, K., Gao, C., (2021). Targeted Mutagenesis in Zea mays Using TALENs and the CRISPR/Cas System,

Liao, C., Ttofali, F., Slotkowski, R. A., Denny, S. R., Cecil, T. D., Leenay, R. T., ... & Beisel, C. L. (2018). One-step assembly of large CRISPR arrays enables multi-functional targeting and reveals constraints on array design. bioRxiv, 312421.

Lin, J. (2018). Agrobacterium-Mediated Transformation and Gene Editing of Phaeoacremonium minimum.University of California, Davis.

Lin S, Staahl BT, Alla RK & Doudna JA. (2014) Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9delivery.eLife3, e04766

Liu, L., and Fan, X.D. (2014).CRISPR-Cas system: a powerful tool for genome engineering. Plant Mol Biol 209-218

Liu, B., Xu, H., Miao, J., Zhang, A., Xiaojin, K., Li, W., Zhou, L., Xie, H., Sirois, P., Li, K. (2015), CRISPR/Cas: a faster and more efficient gene editing system. J. Nanosci. Nanotechnol., 15, pp. 1946-1959

Liu, G., Zhang, Y., and Zhang, T. (2019). Computational Approaches for Effective

CRISPR Guide RNA Design and Evaluation. Comput. Struct. Biotechnol. J. 18, 35–44. doi:10.1016/j.csbj.2019.11.006

Le, D.T., Nishiyama, R., Watanabe, Y., Mochida, K., Yamaguchi-Shinozaki, K., Shinozaki, K., and Tran, L.S. (2011). Genome-wide survey and expression analysis of the plant-specific NAC transcription factor family in soybean during development and dehydration stress. *DNA Res.* 63-76.

Lu C, Tej SS, Luo S, Haudenschild CD, Meyers B, Green P. (2005). Elucidation of the small RNA component of the transcriptome. *Science*. 309, 1567–1569.

Lu, C., Jeong, D., Kulkarni, K., Pillay, M., Nobuta, K., German, R., Thatcher, S., Maher, C., Zhang, L., Ware, D., Liu, B., Cao, X., Meyers, B., and Green, P. (2008). Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). *Proc. Natl. Acad. Sci. U. S. A.* 105, 4951–6.

Lu, Z., Yu, X., Shengbo, C., Min, S., Xianda, J., Zhuoli, H., Shuai, G., (2022). Mechanisms of MicroRNA Biogenesis and Stability Control in Plants. Frontiers in Plant Science

Llave, C. (2002). Cleavage of Scarecrow-like mRNA Targets Directed by a Class of Arabidopsis miRNA. *Science* (80). 297, 2053–2056.

Llave, C., Kasschau, K. D., Rector, M. A. & Carrington, J. C. (2002). Endogenous and silencing associated small RNAs in plants. *Plant Cell* 14, 1605–19

Ma, H., Tu, L.-C., Naseri, A., Huisman, M., Zhang, S., Grunwald, D., et al. (2016). CRISPR-Cas9 Nuclear Dynamics and Target Recognition in Living Cells. J. Cell. Biol.214, 529–537. doi:10.1083/jcb.201604115

Ma, Z. & Zhang, X., (2018).Actions of plant Argonautes: predictable or unpredictable?. Current opinion in plant biology, 45, pp.59-67

Mallory, A. C., Reinhart, B. J., Jones-Rhoades, M. W., Tang, G., Zamore, P. D., Barton, M. K., and Bartel, D. P. (2004). MicroRNA control of *PHABULOSA* in leaf development: importance of pairing to the microRNA 5′ region. *The EMBO Journal*, *23*(16), 3356–3364.

Mao, Y., Yang, X., Zhou, Y., Zhang, Z., Botella, J.R. & Zhu, J.K., (2018).Manipulating plant RNA-silencing pathways to improve the gene editing efficiency of CRISPR/Cas9 systems. Genome biology, 19(1), pp.1-15.

Manghwar, H., Li, B., Ding, X., Hussain, A., Lindsey, K., Zhang, X., et al. (2020).

CRISPR/Cas Systems in Genome Editing: Methodologies and Tools for sgRNA Design, Off-Target Evaluation, and Strategies to Mitigate Off-Target Effects. Adv. Sci.7, 1902312. doi:10.1002/advs.201902312

Mateos-Gomez, P. A., Kent, T., Deng, S. K., McDevitt, S., Kashkina, E., Hoang, T. M., et al. (2017). e helicase domain of Poltheta counteracts RPA to promote alt-NHEJ. Nat. Struct. Mol. Biol. 24, 1116–1123. doi: 10.1038/nsmb.3494

Marillonnet, S., & Grützner, R. (2020). Synthetic DNA assembly using golden gate cloning and the hierarchical modular cloning pipeline. Current protocols in molecular biology, 130(1), e115.

Marillonnet, Sylvestre; Werner, Stefan. (2015) “Assembly of Multigene Constructs Using Golden Gate Cloning" *Method in Molecular Biology*.

Marr, E. J., Wright, H. W., Sargison, N. D., Nisbet, A. J., & Burgess, S. T. (2018). Gene silencing by RNA interference in the ectoparasitic mite, Psoroptes ovis. *Veterinary research*, *49*(1), 1-

Marraffini LA. (2015)**.** CRISPR-Cas immunity in prokaryotes. *Nature* 526(7571):55-61

Martínez-García, E., Goñi-Moreno, A., Bartley, B., McLaughlin, J., Sánchez-Sampedro, L., Pascual del Pozo, H., ...& de Lorenzo, V. (2020). SEVA 3.0: an update of the Standard European Vector Architecture for enabling portability of genetic constructs among diverse bacterial hosts. Nucleic acids research, 48(D1), D1164-D1170.

Medline (2022). What are genome editing and CRISPR-Cas9?

Medžiūnė, J., Kapustina, Ž., Žeimytė, S., Jakubovska, J., Sindikevičienė, R., Čikotienė, I., & Lubys, A. (2022). Advanced preparation of fragment libraries enabled by oligonucleotide-modified 2′, 3′-dideoxynucleotides. Communications Chemistry, 5(1), 1-8.

Mehravar M, Shirazi A, Mehrazar MM, Nazari M. (2019) *In Vitro* Pre-validation of Gene Editing by CRISPR/Cas9 Ribonucleoprotein. Avicenna J Med Biotechnol. 11(3):259-263. PMID: 31380000; PMCID: PMC6626505.

Mekler, V., Minakhin, L., Semenova, E., Kuznedelov, K., & Severinov, K. (2016). Kinetics of the CRISPR-Cas9 effector complex assembly and the role of 3′-terminal segment of guide RNA. Nucleic acids research, 44(6), 2837-2845.

Mendes, N. D., Freitas, A. T. & Sagot, M. F. (2009). Survey and Summary: Current tools for the identification of miRNA genes and their targets. *Nucleic Acids Research* 37**,** 2419–2433

Meng, Y., Shao, C., Wang, H. & Chen, M. (2012). Are all the miRBase-registered microRNAs true? A structure- and expression-based re-examination in plants. *RNA Biol.* 9, 249–53

Merchan, F., Boualem, A., Crespi, M., and Frugier, F. (2009). Plant polycistronic precursors containing non-homologous microRNAs target transcripts encoding functionally related proteins. *Genome Biology* 10, R136

Merkin, J., Russell, C., Chen, P. & Burge, C. B. (2012). Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* **338,** 1593–9

Meyers, B. C. Axtell, M. J., Bartel, B., Bartel, D., Baulcombe, D., Bowman, J. L., Cao, X., Carrington, J. C., cChen, X., Green, P. J., Jones, S. G., Jacobsen, S. E., Mallory, A.C., Martienssen, R. A., Poethig, R. S., Qi, Y., Vaucheret, H., Voinnet, O., Watanabe, Y., Weigel, D., Kang Zhu, J.(2008). Criteria for Annotation of Plant MicroRNAs. *Plant Cell* Online 20, 3186–3190

Michel, F., Jacquier, A. & Dujon, B. (1982). Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. *Biochimie* 64, 867–881

Michel, F., Umesono, K. & Ozeki, H. (1989). Comparative and functional anatomy of group II catalytic introns -a review *. *Gene* 82, 5–30

Millar, A. A., Lohe, A., & Wong, G. (2019). Biology and function of miR159 in plants. *Plants*, *8*(8), 255.

Modrzejewski, D., Hartung, F., Sprink, T., Krause, D., Kohl, C., and Wilhelm, R. (2019). What is the available evidence for the range of applications of genome-editing as a new tool for plant trait modification and the potential occurrence of associated off-target effects: a systematic map? *Environ Evid*. 8:11.

Moreno-Mateos MA, Vejnar CE, Beaudoin JD,Fernandez JP, Mis EK, Khokha MK & Giraldez AJ. (2015) CRISPR scan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo.Nat Methods12,982–988

Moro, A., Driscoll, T. P., Boraas, L. C., Armero, W., Kasper, D. M., Baeyens, N., ... & Nicoli, S. (2019). MicroRNA-dependent regulation of biomechanical genes establishes tissue stiffness homeostasis. *Nature cell biology*, *21*(3), 348-358.

Moxon, S., Jing, R., Szittya, G., Schwach, F., Rusholme Pilcher, R. L., Moulton, V., & Dalmay, T. (2008). Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. *Genome Research*, *18*(10), 1602–1609.

Müller, J. Toev, T., Hesters, M., Teller, J., Moore, K. L., Hause, G., Dinesh, D. C., Burstenbinder, K., Abel, S. (2015). Iron-Dependent Callose Deposition Adjusts Root Meristem Maintenance to Phosphate Availability. Dev. Cell 33, 216–230. Function of a microRNA gene containing an intron in tomato

Najafi, S., Tan, S.C., Aghamiri, S., Raee, P., Ebrahimi, Z., Jahromi, Z.K., Rahmati, Y., Nahand, J.S., Piroozmand, A., Jajarmi, V. & Mirzaei, H., (2022). Therapeutic potentials of CRISPR-Cas genome editing technology in human viral infections. Biomedicine & Pharmacotherapy, 148,

Mollory, A., Dugas, D., Bartel, D., and Bartel, B. (2004). MicroRNA Regulation of NAC-Domain Targets Is Required for Proper Formation and Separation of Adjacent Embryonic, Vegetative, and Floral Organs. *Current Biology* 14, 1035-1046.

Moreira, D., Pereira, A. M., Lopes, A. L., & Coimbra, S. (2020). The best CRISPR/Cas9 versus RNA interference approaches for Arabinogalactan proteins' study. *Molecular Biology Reports*, *47*(3), 2315-2325.

Marillonnet, Sylvestre; Werner, Stefan. (2015) "Assembly of Multigene Constructs Using Golden Gate Cloning" *Method in Molecular Biology*.

Muller, J. Toev, T., Hesters, M., Teller, J., Moore, K. L., Hause, G., Dinesh, D. C., Burstenbinder, K., Abel, S. (2015). Iron-Dependent Callose Deposition Adjusts Root Meristem Maintenance to Phosphate Availability. *Dev. Cell* **33,** 216–230. Function of a microRNA gene containing an intron in tomato

Naeem, M., Majeed, S., Hoque, M. Z., and Ahmad, I. (2020). Latest Developed

Strategies to Minimize the Off-Target Effects in CRISPR-Cas-Mediated Genome Editing. Cells 9, 1608. doi:10.3390/cells9071608

Najafi, S., Tan, S.C., Aghamiri, S., Raee, P., Ebrahimi, Z., Jahromi, Z.K., Rahmati, Y., Nahand, J.S., Piroozmand, A., Jajarmi, V. & Mirzaei, H., (2022). Therapeutic potentials of CRISPR-Cas genome editing technology in human viral infections. Biomedicine & Pharmacotherapy, 148, p.112743

Nekrasov V, Staskawicz B, Weigel D, Jones JD, Kamoun S. (2013). Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. *Nat Biotechnol* 31**:** 691-693

Nikovics, K., Blein, T., Peaucelle, A., Ishida, T., Morin, H., Aida, M., and Laufs, P. (2006). The balance between the MIR164 and CUC2 genes controls leaf margin serration in Arabidopsis. *Plant cell Online* 18,2929.

Nilsen, T. W. & Graveley, B. R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463**,** 457–463.

Nuruzzaman, M., Manimekalai, R., Sharoni, A.M., Satoh, K., Kondoh, H., Ooka, H., and Kikuchi, S. (2010). Genome-wide analysis of NAC transcription factor family in rice Gene. 30-44.

Nuruzzaman, M., Sharoni, A. M., Satoh, K., Kondoh, H., Hosaka, A., and Kikuchi, S. (2012). A genome-wide survey of the NAC transcription factor family in monocots and eudicots, in Introduction to Genetics – DNA Methylation, Histone Modification and Gene Regulation. *Hong Kong: iConcept Press*

Okazaki, K., Kato, H., Iida, T., Shinmyozu, K., Nakayama, J. I., Murakami, Y., & Urano, T. (2018).RNAi-dependent heterochromatin assembly in fission yeast Schizosaccharomyces pombe requires heat-shock molecular chaperones Hsp90 and Mas5. Epigenetics & chromatin, 11(1), 1-15.

Olena, A. F., and Patton, J. G. (2010). Genomic Organization of microRNAs. *Journal of Cellular Physiology*, 540–545.

Padgetr, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. & Sharp, P. A. (1986). SPLICING OF MESSENGER RNA PRECURSORS. *Ann. Rev. Biochem* 55, 1119–150

Park, W., Li, J., Song, R., Messing, J. & Chen, X. (2002). CARPEL FACTORY, a Dicer Homolog, and HEN1, a Novel Protein, Act in microRNA Metabolism in Arabidopsis thaliana. *Curr. Biol*. 1484–1495

Parkes, I.A., Runions, J., Kearns, A. and Hawes, C. (2006) Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. Nature Protoc. 1, 2019–2025.

Patel, A. A., and Steitz, J. A. (2003). Splicing double: Insights from the second spliceosome. *Nature* 4, 960–970

Paterson, A. H. *et al.* (2009). The Sorghum bicolor genome and the diversification of grasses. *Nature* 457, 551–556.

Pattanayak, V., Lin, S., Guilinger, J. P., Ma, E., Doudna, J. A., and Liu, D. R. (2013). High-throughput Profiling of Off-Target DNA Cleavage Reveals RNA-Programmed Cas9 Nuclease Specificity. Nat. Biotechnol. 31, 839–843. doi:10.1038/nbt.2673

Patthy, L. (1999). Genome evolution and the evolution of exon-shuffling — a review k. *Gene* 238, 103–114.

Pek, J. W. and Okamura, K. (2015). Regulatory RNAs discovered in unexpected places. *Wiley Interdiscip. Rev. RNA* 6**,** 671–686.

Pennisi, E. (2013). The CRISPR craze. *Science* 833-836

Peng R, Lin G, and Li J. (2019) Potential pitfalls of CRISPR/Cas9-mediated genome editing. The Febs Journal.

Pannunzio, N. R., Watanabe, G., and Lieber, M. R. (2018). Nonhomologous DNA end-joining for repair of DNA double-strand breaks. J. Biol. Chem. 293, 10512–10523. doi: 10.1074/jbc.TM117.000374

Přibylová A, Fischer L, Pyott DE, Bassett A, Molnar A. (2022). DNA methylation can alter CRISPR/Cas9 editing frequency and DNA repair outcome in a target-specific manner. PMID: 35524464.

Park, S., and Beal, P. A. (2019). Off-Target Editing by CRISPR-Guided DNA Base Editors. Biochemistry 58, 3727–3734. doi:10.1021/acs.biochem.9b00573

Raghothama, K. G. (1999). PHOSPHATE ACQUISITION. *Annu. Rev. Plant Physiol. Plant Mol. Biol* 50**,** 665–93.

Plant Breeding and Genomics (2019). The Polymerase Chain Reaction (PCR).https://plant-breeding-genomics.extension.org/the-polymerase-chain-reaction-

Polstein, L. R., Perez-Pinera, P., Kocak, D. D., Vockley, C. M., Bledsoe, P., Song, L., et al. (2015). Genome-wide Specificity of DNA Binding, Gene Regulation, and Chromatin Remodeling by TALE- and CRISPR/Cas9-based Transcriptional Activators. Genome Res. 25, 1158–1169. doi:10.1101/gr.179044.114

Pong, S. K., & Gullerova, M. (2018). Noncanonical functions of micro RNA pathway enzymes–Drosha, DGCR 8, Dicer and Ago proteins. FEBS letters, 592(17), 2973-2986.

Prihatna, C., Barbetti, M. J., & Barker, S. J. (2018). A novel tomato fusarium wilt tolerance gene. Frontiers in microbiology, 9, 1226. DOI:

Pryor, J. M., Potapov, V., Kucera, R. B., Bilotti, K., Cantor, E. J., & Lohman, G. J. (2020). Enabling one-pot Golden Gate assemblies of unprecedented

complexity using data-optimized assembly design. PLoS One, 15(9), e0238592.

Qin D. Next-generation sequencing and its clinical application. Cancer Biol Med. 2019 (1):4-10. doi: 10.20892/j.issn.2095-3941.2018.0055. PMID: 31119042; PMCID: PMC6528456.

Qin, G., Wu, S., Zhang, L., Li, Y., Liu, C., Yu, J., ... & Zhang, Z. (2022). An Efficient Modular Gateway Recombinase-Based Gene Stacking System for Generating Multi-Trait Transgenic Plants. Plants 2022, 11, 488.

Quévillon Huberdeau, M., & Simard, M. J. (2019). A guide to micro RNA-mediated gene silencing. The FEBS journal, 286(4), 642-652.

Quillet, A., Saad, C., Ferry, G., Anouar, Y., Vergne, N., Lecroq, T., & Dubessy, C. (2020). Improving bioinformatics prediction of microRNA targets by ranks aggregation. Frontiers in Genetics, 10.

Raschmanová, H., Weninger, A., Glieder, A., Kovar, K., and Vogl, T. (2018). Implementing CRISPR-Cas Technologies in Conventional and Non-conventional Yeasts: Current State and Future Prospects. Biotechnol. Adv. 36, 641–665. doi:10.1016/j.biotechadv.2018.01.006

Reed, J., Osbourn, (2018). A. Engineering terpenoid production through transient expression in *Nicotiana benthamiana*. *Plant Cell Rep* **37,** 1431–1441

Reddy, A. S. N., Marquez, Y., Kalyna, M. & Barta, A. (2013). Complexity of the alternative splicing landscape in plants. *Plant Cell* 25**,** 3657–83

Reem, N. T., & Eck, J. V. (2019). Application of CRISPR/Cas9-mediated gene editing in tomato.In Plant Genome Editing with CRISPR Systems (pp. 171-182). Humana Press, New York, NY.

Reichholf, B., Herzog, V. A., Fasching, N., Manzenreither, R. A., Sowemimo, I., & Ameres, S. L. (2019). Time-resolved small RNA sequencing unravels the molecular principles of microRNA homeostasis. Molecular cell, 75(4), 756-768.

Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B. & Bartel, D. P. (2002).MicroRNAs in plants. Genes Dev. 16, 1616–26

Ren, B., Liu, L., Li, S., Kuang, Y., Wang, J., Zhang, D., ... & Zhou, H. (2019). Cas9-NG greatly expands the targeting scope of the genome-editing toolkit by recognizing NG and other atypical PAMs in rice. Molecular plant, 12(7), 1015-1026.

Reymond, M., Svistoonoff, S., Loudet, O., Nussaume, L. and Desnos, T. (2006). Identification of QTL controlling root growth response to phosphate starvation in Arabidopsis thaliana. *Plant, Cell Environ.* 29**,** 115–125

Rhoades, M. W. et al. Prediction of Plant MicroRNA Targets The major challenge in determining miRNA functions. *Cell* 110, 513–520 (2002).

Rojas, A. M., Drusin, S. I., Chorostecki, U., Mateos, J. L., Moro, B., Bologna, N. G., ... & Palatnik, J. F. (2020). Identification of key sequence features required for microRNA biogenesis in plants. Nature communications, 11(1), 1-11.

Rodr guez-Trelles, F., Tarr o, R. & Ayala, F. J. (2006). Origins and Evolution of Spliceosomal Introns. *Annu. Rev. Genet* 40**,** 47–76

Ron M, Kajala K, Pauluzzi G, Wang D, Reynoso MA, Zumstein K, Garcha J, Winte S, Masson H, Inagaki S, Federici F, Sinha NR, Deal R, Bailey-Serres J, Brady SM. (2014). Hairy root transformation using Agrobacterium rhizogenes as a tool for exploring cell type-specific gene expression and function using tomato as a model. *Plant physiology*

Rönspies, M., Schindele, P., Wetzel, R., & Puchta, H. (2022). CRISPR–Cas9-mediated chromosome engineering in Arabidopsis thaliana. Nature Protocols, 1-27.

Roushan, M. R., Shao, S., Poledri, I., Hooykaas, P. J. J., & van Heusden, G. P. H. (2022). Increased Agrobacterium-mediated transformation of Saccharomyces cerevisiae after deletion of the yeast ADA2 gene. Letters in Applied Microbiology, 74(2), 228-237.

Rosalind C. Lee, Rhonda L. Feinbaum, Victor Ambros, (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 75(5), 843-854

Rozov, S. M., Permyakova, N. V., & Deineko, E. V. (2019). The problem of the low rates of CRISPR/Cas9-mediated knock-ins in plants: approaches and solutions. International journal of molecular sciences, 20(13), 3371

Rushton, P.J., Bokowiec, M.T., Han, S., Zhang, H., Brannock, J.F., Chen, X., Laudeman, T.W., and Timko, M.P. (2008). Tobacco transcription factors: novel insights into transcriptional regulation in the Solanaceae. Plant Physiol. 280-95.

Ruvkun, G. et al. (2000). The 21-nucleotide let-7 RNA regulates developmental timing In Caenorhabditis elegans. *Nature* 403, 901–906.

Sakurai, T., K. (2007). Basic and applied features of multicopper oxidases, Cueo, bilirubin oxidase, and laccase. *Chem. Rec.* 7, 220–9

Salanga, C.M. & Salanga, M.C., (2021). Genotype to phenotype: CRISPR gene editing reveals genetic compensation as a mechanism for phenotypic disjunction of morphants and mutants. International Journal of Molecular Sciences, 22(7), p.3472.

Sainsbury F, Lomonossoff GP (2014). "Transient expressions of synthetic biology in plants". Current Opinion in Plant Biology. SI: Physiology and metabolism. 19: 1–7.

Salsman, J., Masson, J. Y., Orthwein, A., and Dellaire, G. (2017). CRISPR/Cas9 gene editing: from basic mechanisms to improved strategies for enhanced genome engineering in vivo. Curr. Geneer. 17, 263–274. doi: 10.217 4/156652321766617112209462 9

Samarfard, S., McTaggart, A. R., Sharman, M., Bejerman, N. E., & Dietzgen, R. G. (2020).Viromes of ten alfalfa plants in Australia reveal diverse known viruses and a novel RNA virus. Pathogens, 9(3)

Sledzinski, P., Nowaczyk, M., and Olejniczak, M. (2020). Computational Tools and Resources Supporting CRISPR-Cas Experiments. Cells 9, 1288. doi:10.3390/cells9051288

Souer, E., van Houwelingen A., Kloos D., Mol J., and Koes R. (1996). The no apical meristem gene of Petunia is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. *Cell* 85, 159–170

Shomron, Noam, and Carmit Levy. (2009). MicroRNA-biogenesis and pre-mRNA splicing crosstalk. *Journal of Biomedicine and Biotechnology* 594678.

Sunkar, R., Girke, T., Jain, P. K. & Zhu, J.-K. (2005). Cloning and characterization ofmicroRNAs from rice. Plant Cell 17, 1397–411

Seitz, H. and Zamore, P. D. (2006). “Rethinking the microprocessor,” *Cell*, vol. 125, no. 5, pp. 827– 829

Shapulatov, U., van Hoogdalem, M., Schreuder, M., Bouwmeester, H., Abdurakhmonov, I. Y., & van der Krol, A. R. (2018). Functional intron-derived miRNAs and host-gene expression in plants. Plant methods, 14(1), 1-13

Shoffner, G. M., Peng, Z., & Guo, F. (2020). Structures of microRNA-precursor apical junctions and loops reveal non-canonical base pairs important for processing. bioRxiv.

Shukla, G., Singh, J., and Barik, S. (2011). MicroRNAs: Processing, Maturation, Target Recognition and Regulatory Functions. Mol Cell Pharmacol.3(3): 83–92.

Solomon, E. I., Sundaram, U. M. and Machonkin, T. E. (1996). Multicopper Oxidases and Oxygenases. *Chem. Rev.* 96, 2563–2606

Soyars, C. L., Peterson, B. A., Burr, C. A., & Nimchuk, Z. L. (2018). Cutting edge genetics: CRISPR/Cas9 editing of plant genomes. Plant and Cell Physiology, 59(8), 1608-1620.

Staiger, D. & Brown, J. W. S. (2013). Alternative splicing at the intersection of biological timing, development, and stress responses. *Plant Cell* **25,** 3640–56

Stanton Gelvin, D. (2018). Theme 4-New Plant Breeding Techniques. In Vitro Cellular & Developmental Biology-Plant, 54(1), S84-S106. DOI:10.1007/s11627-018-9923-0

Stepien, A. Knop, K., Dolata, J., Taube, M., Bajczyk, M., Barcizewska-Pacak, M., Pacak, A., Jarmolowski, A., Szweykowska-kulinska, Z. (2016). Posttranscriptional coordination of splicing and miRNA biogenesis in plants. *Wiley Interdiscip. Rev. RNA*

Stovicek, V., Holkenbrink, C., and Borodina, I. (2017). CRISPR/Cas System for Yeast Genome Engineering: Advances and Applications. FEMS Yeast Res. 17, fox030. doi:10.1093/femsyr/fox030

Subburaj S., Chung S.J., Lee C., Ryu S.M., Kim D.H., Kim J.S., Bae S., Lee G.J. (2016) Site-directed mutagenesis in Petunia× hybrida protoplast system using direct delivery of purified recombinant Cas9 ribonucleoproteins. *Plant. Cell Rep.* 2016; 35:1535–1544.

Sunkar, R., Girke, T., Jain, P. K. & Zhu, J.-K. (2005). Cloning and characterization of microRNAs from rice. *Plant Cell* 17**,** 1397–411.

Sunkar, R. & Jagadeeswaran, G. (2008). In silico identification of conserved microRNAs in large number of diverse plant species. *BMC Plant Biol.* 8**,** 37.

Svistoonoff, S. Cerff, A., Reymond, M., Sigoillot-Claude, C., Ricaud, L., Blamchet, A., Nussaume, L., and Desnos, T. (2007). Root tip contact with

low- phosphate media reprograms plant root architecture. *Nat. Genet.* 39**,** 792–796

Szarzynska, B. et al. (2009). Gene structures and processing of Arabidopsis thaliana HYL1-dependent pri-miRNAs. *Nucleic Acids Res.* 37, 3083–93

Szweykowska-Kuli ska, Z., Jarmolowski, A. & Vazquez, F. (2013). The crosstalk between plant microRNA biogenesis factors and the spliceosome. *Plant Signal. Behav.* 8, e26955

Thakur, V. Wanchana, S., Xu. M., Bruskiewich, R., Paul Quick, W., Mosig, A., and Zju, X. (2011). Characterization of statistical features for plant microRNA prediction. *BMC Genomics* 12**,** 108.

Thieme, C. J., Gramzow, L., Lobbes, D. and Theißen, G. SplamiR. (2011). Prediction of spliced miRNAs in plants. *Bioinforma. Orig. Pap.* 13**,** 1215–122310

Thomson J.A. (2016). Genetic Engineering of Plants. Biotechnology. Encyclopedia of Life Support Systems.

Tsai, S. Q., Zheng, Z., Nguyen, N. T., Liebers, M., Topkar, V. V., Thapar, V., et al. (2015). GUIDE-seq Enables Genome-wide Profiling of Off-Target Cleavage by CRISPR-Cas Nucleases. Nat. Biotechnol. 33, 187–197. doi:10.1038/nbt.311

Tyumentseva, M., Mikhaylova, Y., Prelovskaya, A., Tyumentsev, A., Petrova, L., Fomina, V., ...& Akimkin, V. (2021). Genomic and phenotypic analysis of multidrug-resistant Acinetobacter baumannii clinical isolates carrying different types of CRISPR/Cas systems. Pathogens, 10(2), 205.

Ueki, S., Magori, S., Lacroix, B., & Citovsky, V. (2013). Transient gene expression in epidermal cells of plant leaves by biolistic DNA delivery. In Biolistic DNA Delivery (pp. 17-26).Humana Press, Totowa, NJ.

Uniyal, A. P., Mansotra, K., Yadav, S. K., & Kumar, V. (2019).An overview of designing and selection of sgRNAs for precise genome editing by the CRISPR-Cas9 system in plants. 3 Biotech, 9(6), 1-19.

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., and Gregory, P.D. (2010). Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* 11: 636-646

Van Eck J, Kirk DD, Walmsley AM. (2006). Tomato (Lycopersicum esculentum). *In* K Wang, ed, Methods in Molecular Biology, Agrobacterium Protocols, Vol 343. *Humana Press Inc.*, Totowa, NJ, pp 459-473

Veillet, F., Chauvin, L., Kermarrec, M. P., Sevestre, F., Merrer, M., Terret, Z., ... & Chauvin, J. E. (2019). The Solanum tuberosum GBSSI gene: a target for assessing gene and base editing in tetraploid potato. Plant cell reports, 38(9), 1065-1080.

Voinnet, O. (2009). Origin, Biogenesis, and Activity of Plant MicroRNAs. *Cell* 136, 669–687

Waititu, J. K., Zhang, C., Liu, J., & Wang, H. (2020).Plant non-coding RNAs: origin, biogenesis, mode of action and their roles in abiotic stress. International Journal of Molecular Sciences, 21(21), 8401

Waldrip, Z. J., Jenjaroenpun, P., DeYoung, O., Nookaew, I., Taverna, S. D., Raney, K. D., et al. (2020). Genome-wide Cas9 Binding Specificity in Saccharomyces cerevisiae. PeerJ 8, e9442. doi:10.7717/peerj.9442

Wang, B.-B. & Brendel, V. (2006). Genomewide comparative analysis of alternative splicing in plants. *Proc. Natl. Acad. Sci. U. S. A.* 103**,** 7175–80

Wang, J., Mei, J., & Ren, G. (2019).Plant microRNAs: biogenesis, homeostasis, and degradation. Frontiers in plant science, 10, 360

Wang T, Wei JJ, Sabatini DM & Lander ES. (2014) Genetic screens in human cells using the CRISPR-Cas9system.Science343,80–84

Wang, X., Wang, Y., Dou, Y., Chen, L., Wang, J., Jiang, N., ...& Ren, G. (2018). Degradation of unmethylated miRNA/miRNA* s by a DEDDy-type 3′ to 5′ exoribonuclease Atrimmer 2 in Arabidopsis. Proceedings of the National Academyof Sciences,

Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C, Qiu JL. (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat Biotechnol*

Wang, Z. & Cui, W., (2020) CRISPR-Cas system for biomedical diagnostic platforms. View, 1(3), p.20200008.

Wightman, B., Ha, I. & Ruvkun, G. (1993) Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. *Elegans. Cell 7*5, 855–62

Williamson, A., & Leiros, H. K. S. (2019). Structural intermediates of a DNA–ligase complex illuminate the role of the catalytic metal ion and mechanism of phosphodiester bond formation. Nucleic acids research, 47(14), 7147-7162

Williamson, L. C., Ribrioux, S. P., Fitter, A. H. and Leyser, H. M. (2001) Phosphate availability regulates root system architecture in Arabidopsis. *Plant Physiol.* 126**,** 875–82.

Weber E., Engler C., Gruetzner R., Werner S., and Marillonnet S. (2011**).** A Modular Cloning System for Standardized Assembly of Multigene Constructs. -02-18

Wong, N., Liu, W., Wang, X. (2015). WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome Biol.,* 16 (2015), p. 218

Wu, X., Scott, D. A., Kriz, A. J., Chiu, A. C., Hsu, P. D., Dadon, D. B., et al. (2014). Genome-wide Binding of the CRISPR Endonuclease Cas9 in Mammalian Cells. Nat. Biotechnol. 32, 670–676. doi:10.1038/nbt.2889

Xi-Dian, T., Fei, G., Ming-Jie, L., Qin-Lei, F., De-Kun, C., Wen-Tao, M. (2019). Methods for Enhancing Clustered Regularly Interspaced Short Palindromic Repeats/Cas9-Mediated Homology-Directed Repair Efficiency. Frontiers in Genetics.

Xie, J. et al. (2017). Adaptive evolution and functional innovation of Populus -specific recently evolved microRNAs. *New Phytol.* **213,** 206–219.

Xie K, Zhang J, Yang Y. (2014). Genome-wide prediction of highly specific guide RNA spacers for CRISPR-Cas9-mediated genome editing in model plants and major crops. *Molecular plant 7:* 923-926

Xu H, Xiao T, Chen CH, Li W, Meyer CA, Wu Q, WuD, Cong L, Zhang F, Liu JSet al. (2015) Sequence determinants of improved CRISPR sgRNA design. Genome Res25, 1147–1157.

Xu, T.F., Zhao, X.C., Jiao, Y.T., Wei, J.Y., Wang, L.and Xu, Y. (2014) A pathogenesis related protein, VpPR-10.1, from Vitis pseudoreticulata: an i nsight of its mode of antifungal activity. PLoS ONE, 9, e95102.

Xu, W., Shen, P., Li, R., Liu, B., & Yang, L. (2022). Development of an event-specific droplet digital PCR assay for quantification and evaluation of the transgene DNAs in trace samples of GM PRNP-knockout goat. Foods, 11(6), 868. DOI

Yamamoto, T., Hoshikawa, K., Ezura, K. et al. (2018). Improvement of the transient expression system for production of recombinant proteins in plants. *Sci Rep* **8,** 4755

Yan, K., Liu, P., Ai Wu, C., Yang, G., Xu, R., Guo, Q., Huang, J., Zheng, C. (2012). Stress-Induced Alternative Splicing Provides a Mechanism for the Regulation of MicroRNA Processing in Arabidopsis thaliana. *Mol. Cell* 48, 521–531

Zahn, K. E., Jensen, R. B., Wood, R. D., & Doublié, S. (2021). Human DNA polymerase θ harbors DNA end-trimming activity critical for DNA repair. Molecular cell, 81(7), 1534-1547.

Zhang H, Zhang J, Wei P, Zhang B, Gou F, Feng Z, Mao Y, Yang L, Zhang H, Xu N, Zhu JK (2014) The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnol J* 12: 797-807

Zhang, B., Pan, X., Cannon, C. H., Cobb, G. P. & Anderson, T. A. (2006). Conservation and divergence of plant microRNA genes. *Plant* J. 46, 243–59

Zhang, H., Demirer, G. S., Zhang, H., Ye, T., Goh, N. S., Aditham, A. J., ...& Landry, M. P. (2019). DNA nanostructures coordinate gene silencing in mature plants. Proceedings of the National Academy of Sciences, 116(15), 7543-7548

Zhang, L., Ming Chia, J., Kumari, S., Stein, J., Liu, Z., Narechania, A., Maher, C., Guill, K., McMullen, M., Ware, D. (2009). A Genome-Wide Characterization of MicroRNA Genes in Maize. *PLoS Genet. 5*, e1000716

Zhang, Y., Park, K. Y., Suazo, K. F., & Distefano, M. D. (2018). Recent progress in enzymatic protein labelling techniques and their applications. Chemical Society Reviews, 47(24), 9106-9136.

Zhang, Y., Ren, Q., Tang, X., Liu, S., Malzahn, A. A., Zhou, J., ... & Qi, Y. (2021). Expanding the scope of plant genome engineering with Cas12a orthologs and highly multiplexable editing systems. Nature communications, 12(1), 1-11.

Z**hang, Y., Xia, R., Kuang, H. & Meyers, B. C.** (2016). The Diversification of Plant NBS-LRR Defense Genes Directs the Evolution of MicroRNAs That Target Them. *Mol. Biol. Evol.* **33,** 2692–2705).

Zhao, Y., Wang, Z., Zhang, W., & Zhang, L. (2019). MicroRNAs play an essential role in autophagy regulation in various disease phenotypes. Biofactors, 45(6), 844-856.

Zhou, x., Ruan, J., Wang, G., and Zhang, W. (2007). Characterization and identification of microRNA core promoters in four model species.

Appendix

Off-target analysis for the selected gRNAs.

Seq-id	sgRNA-id	Score	Sequence	strand	pos	%GC
seq19306 1089 bp	Guide1	0.6952	TCCGAGCTCAAACACAGCGACGG	+	878	55%
Seq19306 1089 bp	Guide2	0.2061	ACACATGGCATTATACTCTTGGG	+	818	35%