Characterisation of an intron-split Solanales microRNA

Zahara Medina Calzada

A thesis submitted for the degree of
Doctor of Philosophy

University of East Anglia
School of Biological Sciences

September 2017

“Doing science is very romantic… when it works!”

(the friend of a friend of mine)

Abstract

MicroRNAs (miRNAs) are a distinct class of short endogenous RNAs with central roles in post-transcriptional regulation of gene expression that make them essential for the development and normal physiology of several groups of eukaryotes, including plants. In the last 15 years, hundreds of miRNA species have been identified in plants and great advances have been achieved in the understanding of plant miRNA biogenesis and mode of action. However, many miRNAs, generally those with less conventional features, still remain to be discovered. Likewise, further layers that regulate the pathway from miRNA biogenesis to function and turnover are starting to be revealed.

In the present work we have studied the tomato miRNA “top14”, a miRNA with a non-canonical pri-miRNA structure in which an intron is in between miRNA and miRNA*. We have found that this miRNA is conserved within the economically important Solanaceae family and among other members of the Solanales order also agriculturally relevant, 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*. In these three genera, two different pri-miRNA variants were detected; one spliced and the other one retaining the intron. After testing the mature miRNA production from the wild type tomato *MIRtop14*, from a version without intron and from another version without splicing capability, it was found that the intron influenced the accumulation of mature miRNA. Finally, a mRNA cleaved by this miRNA was identified; the mRNA coding for LOW PHOSPHATE ROOT (LPR), a protein which in Arabidopsis is involved in the arrest of root growth under phosphate starvation conditions. 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 may be regulated through intron retention. In case of being confirmed, it would add to the few recently reported examples of post-transcriptional regulation of a miRNA and should encourage the research of less known layers of miRNA regulation. Finally, the study of this miRNA sheds light to the crosstalk between miRNA biogenesis and splicing and, in a broader context, to the complex interactions between the different RNA regulatory networks operating in plants.

Table of Contents

List of Figures & Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Acknowledgements

I would like to start by thanking all those that made possible that I took a PhD in the first place. Thank you to Dr. Franscisco Javier Galego Rodriguez, Dr. Fred Tata, Dr. Nicole Royle, Dr. Yan Huang, Dr. Cesar Llave, Dr. Virginia Ruiz Ferrer and Ignacio Hamada for their encouragement, teachings and references.

I am profoundly grateful to my supervisor, Prof. Tamas Dalmay. Thank you for always considering my criteria 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.

I am also deeply thankful to my secondary supervisor, Dr. Ping Xu. Thank you for all the conversations about science and all your teachings on experimental procedures, especially on how to find a way out when everything seemed not to work.

I would also like to thank Dr. Ben Miller for all his help with the Golden Gate cloning, from the material provided to all his patient teachings.

I am also heartily thankful to Dr. Archana for her help with the degradome analysis and to Dr. Aurore Coince and Dr-to-be Martina Billmeier for all their help in the lab. Besides all your help at work, you three really lightened up my PhD days.

Thank you as well to Prof. Ronald Koes and Elizabeth C. Verbree for kindly sending *Petunia axillaris* S26 seeds.

Besides, I would like to thank all those that I have worked with throughout these years and who have contributed to helping me grow, sometimes as a scientist, sometimes as a person. I have been very lucky to get to work with so many different people from so many different places. A special thank you to Rocky Payet for correcting these acknowledgments.

Finally, I would not like to forget all those outside the science world who have nevertheless helped me to complete this PhD.

I will always be indebted to Prof. Enrique Baca Baldomero for all the generous help, and who has helped me to overcome the side of science that I find more difficult. I doubt this PhD would have been possible without your help. Thank you.

I am very grateful to my family for their loving care, despite none of them being especially interested in science. Thanks to my mother and her husband for their support, in particular economic, that has made possible for me to take such an uncertain career path. Thanks to my father, for reminding me of the important things in life, and for the candles. Thanks to my brother for being always there for anything I needed. Thanks to the rest of my family, especially from my father's side, for all the conversations and encouragement I received.

I would like to thank as well my family in-law, for having accepted me like another family member and having understood the decisions their son/ brother and I have taken, even if they would not have taken the same.

Thank you to my friends. To the ones in Norwich, for the evenings at the pub, the trips and the climbs. To the ones at home, for our WhatsApp conversations, their visits and always finding time to meet me whenever I decided to go to Spain. 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 Sara, for taking such good care of me when she was here in Norwich, and for passing on to me her strength every time we talk.

Last, but not least, I would like to thank Jose, my life journey partner. I could not have dreamt of a better partner for this adventure. Thanks for coming with me to an unknown city, for understanding each time 30 extra minutes in the lab ended up being 3 extra hours, for listening all time I talked about science despite not understanding it, for helping me survive these last months of writing, for enjoying with me the good moments and helping me out of the bad ones. Definitely, this PhD would not have been possible without you. Thank you.

Abbreviations

A	Adenine
AGO	Argonaute
ALMT1	ALUMINUM-ACTIVATED MALATE TRANSPORTER1
AMP1	ALTERED MERISTEM PROGRAM1
AS	Alternative splicing
Ata	*Arabidopsis thaliana*
ATP	Adenosine triphosphate
AtUBI10	Arabidopsis ubiquitin 10 promoter
BAC	bacterial artificial chromosome
BLAST	Basic Local Alignment Search Tool
Bp	Base pairs
C	Cytosine
CaMV p35S, t35S	CaMV 35S promoter, CaMV 35 terminator
CaMV	Cauliflower mosaic virus
CBC	Cap-binding complex
CDC5	Cell division cycle 5
cDNA	Complementary DNA
CDS	Coding DNA sequence
circRNAs	circular RNAs
Col0	Columbia 0 ecotype
CPL1	C-TERMINAL DOMAIN PHOSPHATASE-LIKE 1
CTR4	CTR1-like protein kinase
Cv	Cultivar
D-body	Dicing body
DCL1	Dicer-like 1 RNAse III endonuclease
DDL	DAWDLE
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNAse	Deoxyribonuclease
dNTPs	Deoxyribonucleotides
DRB2	DOUBLE-STRANDED RNA BINDING 2
DsRed	Discosoma sp. red fluorescent protein
DTT	Dithiothreitol

EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride
ER	Endoplasmic reticulum
eRNA	Enhancer RNA
EST	Expressed sequence tags
EtBr	Ethidium bromide
Fw	Forward
G	Guanine
gDNA	Genomic DNA
GRP7	GLYCINE-RICH RNA-BINDING PROTEIN 7
GUS	Beta-Glucuronidase
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
Ini	*Ipomoea nil*
IPTG	Isopropyl -D-1-thiogalactopyranoside
Kan	Kanamycin
KanR	Kanamycin resistance protein
Kb	Kilobase
KTN1	KATANIN1
LA	LANCEOLATE
LB	Luria-Bertani
LECA	last eukaryotes common ancestor
LIKE AMP1	LAM1
lmiRNAs	long miRNAs
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, miR	MicroRNA
MITE	Miniature inverted-repeat transposable element
MOS2	MODIFIER OF SNC1, 2
mRNA	Messenger RNA
Mut	Mutant
nat-miRNA	natural antisense microRNAs
Nbe	*Nicotiana benthamiana*
NCBI	National Center for Biotechnology Information
ncRNA	Non-coding RNA
NGS	Next generation sequencing
NOS	Nopaline synthase promoter
NOT2	Negative on TATA less 2
nt	Nucleotides
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PAR	promoter-associated RNA
PARE	Parallel analysis of RNA ends
Pax	*Petunia axillaris*
PAZ	Piwi Argonaute Zwille
P-bodies	mRNA-processing bodies
PCR	Polymerase chain reaction
PDR2	PHOSPHATE DEFICIENCY RESPONSE 2
phasiRNA	Phased, 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	Posttranscriptional gene silencing
RACE	Rapid amplification of cDNA ends
RACK1	RECEPTOR FOR ACTIVATED C KINASE 1
RCF3	REGULATOR OF CBF GENE EXPRESSION 3
RISC	RNA-induced silencing complex

RLM-RACE	RNA ligase-mediated RACE
RNA	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 sulfate
SE	SERRATE
Sec	Seconds
SGN	SOL Genomics Network
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
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
tasiRNA	trans-acting siRNAs
TBE	Tris-borate-EDTA
TCP4	TCP family transcription factor 4
T-DNA	Transfer DNA
TE	Transposable element
TEMED	Tetramethylethylenediamine

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	EXORIBONUCLEASE4

Chapter 1

Introduction

1.1. RNA biology overview

The first three species of RNA discovered were ribosomal RNA(rRNA)[1], transfer RNA (tRNA)[2] and messenger RNA(mRNA)[3]. Since all these RNAs are involved in protein synthesis, this led to the idea at the time that the role of RNA was basically confined to assist in protein synthesis.

Over time, the picture of RNA biology became more complex. Other kinds of RNAs, called small nuclear RNAs (snRNAs)[4], were discovered, along with RNA splicing[5,6] or RNA catalytic properties[7,8]. It started to become apparent that RNA was playing other important functions in the cell apart from those previously known.

Through the 80s and 90s, scattered reports of novel RNA species and activities added further insight to the field. These include the telomerase RNA component (TERC)[9], the X inactive specific (XIST) RNA[10,11], the first microRNA (miRNA) to be identified[12] or the small interfering RNAs (siRNAs)[13], along with the discovery of RNA-directed DNA methylation[14] and RNA interference[15,16].

When the 2000s came and with them the “omic” era, the high-throughput analysis methods showed the magnitude of what was already being suspected: most of the eukaryotic genome is transcribed into many kinds of different RNAs, of which protein-coding ones are just a minor fraction in complex organisms[17]

The discovery of long non-coding RNAs (lncRNAs)[18] as an abundant new class of transcriptional regulators was probably one of the biggest surprises, but many other species of RNAs have been described in the last 15 years, from Piwi-associated RNAs (piRNAs)[19–22] in animals to trans-acting siRNAs (tasiRNAs)[23,24] in plants, as well as the more recently reported promoter-associated RNAs[25] (PARs), enhancer RNAs (eRNAs)[26,27] or circular RNAs (circRNAs)[28–30], among others.

This diversity of RNAs carry out a broad range of functions which could be summarised into: 1) protein synthesis[31]; 2) regulation of gene expression, at different levels (chromatin, DNA, transcription, RNA splicing, editing, translation and turnover[32,33]); 3) genome stability, including telomere and centromere integrity and silencing of transposons[34]; 4) defence against viruses[35] 5) communication between cells[36,37].

Of course, all these RNA species and functions are not separate, they overlap and interact in multiple ways[38], creating an intricate network that underlies countless biological processes and that has been proposed to lay behind biological phenomena as important as the evolution of eukaryotic complexity[39].

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 lncRNAs and the rest of shorter non-coding RNAs, generally called small RNAs (sRNAs)[25].

Within the sRNA group, miRNA class is probably one of the best characterised since it was early discovered. The first report of a miRNA appeared in 1993 (although at the time it was called a short temporal RNA), when lin-4 was found in *Caenorhabditis elegans*[12]. It was observed that this small RNA regulated *C. elegans* developmental timing by annealing with lin-14 mRNA through antisense complementarity, which repressed lin-14 translation[12,40].

Seven years later, let-7, a second small RNA also repressing the translation of a mRNA, was again reported in *C. elegans*[41]. That same year, let-7 was found to be conserved in bilaterian animals, including humans[42]

It only took one more year for the explosion of discoveries on the miRNA field, with three different publications claiming the existence of multiple miRNAs conserved across animals[43–45]. One year later, the presence and conservation of miRNAs in the plant kingdom was shown in multiple reports[46–49] and virus-encoded miRNAs were added to the list shortly after[50].

From then until now, great advances have been achieved in the understanding of miRNAs, their biogenesis, functions and mode of action, while other facets of miRNAs are still being discovered or require further research.

1.2.2. What is a miRNA?

Conventionally, 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)[51,52].

As commented above, they have been found in animals, plants and viruses. However, as they present substantial differences, we are going to focus on the knowledge gathered about plant miRNAs in the following sections.

1.2.3. miRNA life cycle

1.2.3.1. Transcription

Most plant pri-miRNAs are synthesised from independent transcriptional units which give rise to a single mature miRNA each[53], although cases of intronic or poly-cistronic pri-miRNAs have also been reported[54–59].

Similarly to protein-coding transcripts, pri-miRNAs are usually transcribed by RNA polymerase II (Pol II) and therefore capped, polyadenylated and sometimes spliced[60–62]. Again similarly to other Pol II transcripts, the process is regulated by the Mediator complex, which recruits Pol II to miRNA gene (*MIR*) promoters[63], and by post-transcriptional modifications of the C-terminal domain of Pol II, which in plants have been reported to influence the capping of pri-miRNAs[64].

Transcription factors are also known to influence *MIR* expression. *MIR* promoters have been shown to contain the TATA box sequence[62] as well as other cis-regulatory motifs[65,66]. Specific transcription factors are reported to positively or negatively regulate the expression of individual *MIRs* in response to certain stresses or developmental stages[67–71]. In some cases, they create feedback loops in which a transcription factor regulated by a miRNA controls in turn the expression of its regulatory miRNA[72]. Besides, some transcription factors influence the expression of *MIRs* in general. This is the case of Negative on TATA less 2 (NOT2)[73] and Cell division cycle 5 (CDC5)[74], which interact with Pol II and enhance the transcription of pri-miRNAs broadly.

Finally, at least two more mechanisms of pri-miRNA transcription regulation should be added to the ones discussed above. One of them is the recently described existence of small peptides encoded by pri-miRNAs, which increase the transcription of their own encoding transcripts[75]. The second one is epigenetic regulation, with the examples of a histone acetyl transferase and a chromatin remodelling complex being required for the activation of some *MIR* promoters in *Arabidopsis*[76,77]

1.2.3.2. Processing

1.2.3.2.1. *Processing machinery*

pri-miRNAs fold creating one or more local stem-loop structures with the miRNA sequence located in one of the arms of the stem[78]. In plants, these hairpins are processed at a specific type of nuclear bodies called dicing bodies (D-bodies), which contain the complex of proteins that orchestrates the cleavage of the pri-miRNA into a miRNA/ miRNA* duplex, the processing complex[79–81].

The core protein of the processing complex and the one performing the cleavage is in most cases the Dicer-like 1 (DCL1) RNAse III endonuclease[47,49]. However, there are a few miRNAs known to be processed by DCL4[54] or by DCL3[82,83].

Other well stablished components of the complex are the proteins HYPONASTIC LEAVES1 (HYL1)[84,85] and SERRATE (SE)[86,87], which are necessary for the efficient and accurate cleavage of pri-miRNAs by DCL1[88,89]. Associated with them, with DCL1 and with the pri-miRNA, is TOUGH (TGH), another RNA binding protein also necessary for the efficient cleavage of pri-miRNAs, although not for the accuracy of the cleavage[90].

NOT2 and CDC5, two proteins already named because of their involvement in pri-miRNA transcription, are as well interacting with DCL1 and SE (but not HYL1) during pri-miRNA processing, promoting pri-miRNA processing possibly by recruiting these two processing factors to the nascent pri-miRNAs[73,74].

Likewise, proteins involved in pri-miRNA stabilization are also known to form part of the processing complex. This is the case of the nuclear cap-binding complex (CBC), which binds to the 5' cap structure of pri-miRNAs and to SE and NOT2[73,91–94], and of the proteins PLEIOTROPIC REGULATORY LOCUS 1 (PRL1)[95] and DAWDLE (DDL)[96], which are reported to interact with DCL1.

Interestingly, there have been recently discovered several processing factors which seem to affect miRNA accumulation, at least partially, by regulating pri-miRNA splicing. These include REGULATOR OF CBF GENE EXPRESSION 3 (RCF3, also known as HOS5)[97], RS40/RS41[97], STABILIZED 1 (STA1)[98] and GLYCINE-RICH RNA-BINDING PROTEIN 7 (GRP7)[99].

Other components of the DCL complex include: C-TERMINAL DOMAIN PHOSPHATASE-LIKE 1 (CPL1)[100], which regulates HYL1; MODIFIER OF SNC1, 2 (MOS2)[101] and THO/TREX complex core protein THO2[102], which seem to be involved

in the recruitment of pri-miRNAs to D-bodies; RECEPTOR FOR ACTIVATED C KINASE 1 (RACK1)[103] and SICKLE (SIC)[104], which also localize to D-bodies and influence pri-miRNA processing.

Finally, it has been recently reported that the protein DOUBLE-STRANDED RNA BINDING 2 (DRB2), a member of the DRB family as HYL1 (DRB1), is in some cases the one partnering with DCL1 instead of HYL1 during pri-miRNA processing[105]. This seems to define the later mode of action of the miRNA, which would direct cleavage of its target when processed by DCL1-HYL1 and translational repression when processed by DCL1-DRB2[105].

1.2.3.2.2. Processing patterns

Once the processing complex is assembled, the cleavage of the pri-miRNAs stem-loops takes place. This can happen following at least four different patterns of cleavage[106].

1) Short base-to-loop pattern, in which a first cleavage occurs ~15 nt up the stem from an internal bulge, and a second cleavage ~21 nt up the stem from the first cut, releasing the miRNA duplex[107–109].
2) Long base-to-loop pattern, which differs from the short base-to-loop pathway in the fact that after the second cleavage, a third cut, also ~21 nt up the stem, is necessary to release the miRNA duplex[106].
3) Short loop to base pattern, in which the first cut is determined by a final ~42 nt segment above the miRNA duplex that is cleaved out, followed by a second cut ~21 nt down the stem which releases the miRNA duplex[106].
4) Long loop-to-base pattern, with four sequential cleavages of the stem starting from the loop, until releasing the miRNA duplex[110,111].

It may be worth mentioning that bidirectional processing has been reported in miRNAs with a multi-branched loop[112]. However, only the short base-to-loop pattern was producing miRNA duplexes, with a non-canonical loop-to-base mechanism leading to an abortive pri-miRNA processing. See figure 1.1 for different processing patterns.

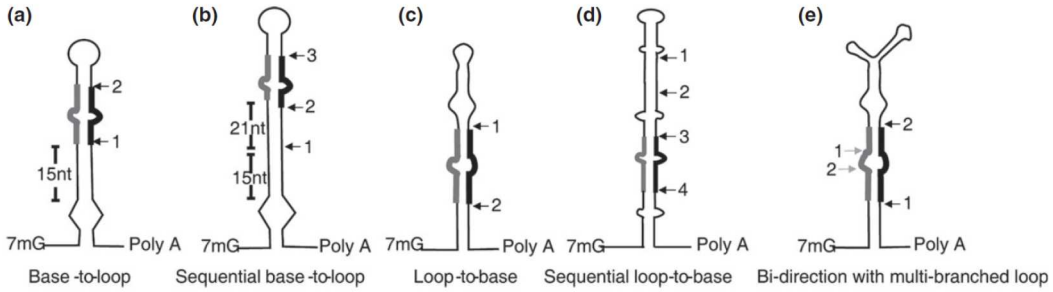


Figure 1.1. Processing patterns of pri-miRNAs depending on their structure. **A)** Short base-to loop. One cut 15 nt up the stem before releasing the miRNA duplex. **B)** Long base-to loop. Two cleavages, at 15 and 21 nt up the stem, before releasing the miRNA duplex. **C)** Short loop-to-base. First, ~42 nt loop is cut out before a second cut releases the miRNA duplex. **D)** Long loop-to-base. Four sequential cleavages of the stem starting from the loop, until releasing the miRNA duplex. **E)** Bidirectional processing, seen in some pri-miRNAs with multibranched loops, although some processing patterns are not productive. Figure from[390].

1.2.3.3. Methylation and nuclear export

The final product of the pri-miRNA processing is a miRNA/ miRNA* duplex, a small double stranded RNA with two nucleotide 3' overhangs formed of the miRNA (guide strand) and its complementary miRNA* (passenger strand)[47].

Once this duplex is released, the 3' terminal nucleotides of both strands are 2'-O-methylated by the protein HUA ENHANCER1 (HEN1)[113,114], which interacts with DCL1 and HYL1, possibly displacing SE[115]. However, HEN1 localizes to both nucleus and cytoplasm[79], so it is not clear where methylation occurs. Nevertheless, it has been demonstrated that this modification is essential to protect the duplex 3' termini from uridylation[113] and subsequent associated degradation[116,117].

In any case, the mature miRNA/miRNA* duplex is supposed to be transported to the cytoplasm, where the assembly of the RISC complex happens most probably, although the possibility of RISC assembling in the nucleus cannot be ruled out[118]. This nuclear export has been proposed to be carried out by the protein HASTY (HST), since it is the homologous to exportin 5, the protein in charge of the transport of pre-miRNAs out of the nucleus in animals[119]. However, although it has been shown that *hst* mutants are deficient in miRNA accumulation, no relative increase of miRNAs in

the nucleus was detected, which indicates that other or additional mechanisms of export still uncharacterised must exist in plants[119].

In an alternative model, it has been proposed that the RNA-induced silencing complex (RISC) is assembled in the nucleus, and the mature miRNA containing complex is then transported to the cytoplasm[120]. See figure 1.2 for plant miRNA life cycle.

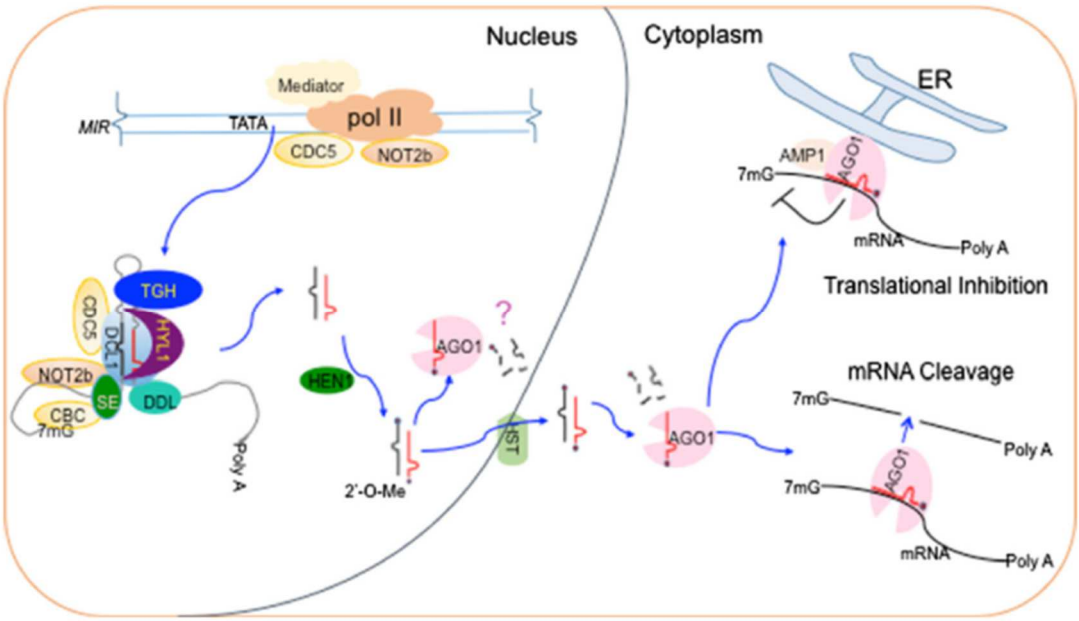


Figure 1.2. Plant miRNAs life cycle. After transcription of the pri-miRNA by Pol II, many protein factors are recruited to the transcript to form the processor complex. The core protein of the processor complex, DCL1, cleaves the pri-miRNA until releasing the miRNA duplex. miRNA/ miRNA* are then methylated by HEN1 and transported from the nucleus to the cytoplasm. miRNAs are finally loaded into an AGO protein, although it is not clear if this takes places in the nucleus or the cytoplasm. AGO is the main component of the riboprotein RISC complex, which directs either mRNA cleavage or translational repression, the last mechanism possibly occurring in the ER. Figure from[438].

1.2.3.4. RISC assembly

RISC is a ribonucleoprotein (RNP) complex which has an Argonaute (AGO) protein bounded to an sRNA[121] 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[122–124]. AGO proteins harbour four characteristic domains: N-terminal (N)[125], Piwi Argonaute Zwille (PAZ)[126,127] and middle (MID)[128,129] domains, which bind the 3' and the 5' end of sRNAs respectively, and C-terminal PIWI domain[125,130,131], 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[132]. However, examples in which the miRNA* is retained and leading RISC action have been reported[133,134].

The selection of the strand to be retained in the complex is determined by its lower 5' end thermostability[120,135,136]. 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 RISC[120]. The HYL1 regulator CPL1, has been shown to regulate this process as well[100].

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[137–139]. 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)[137–139]. Plant miRNAs usually have a 5' U, which is in accordance with AGO1 being the main AGO involved in the plant miRNA pathway[140]

1.2.3.5. Mechanism of action

In plants, RISC recognizes its mRNA target (or targets) through base pairing between miRNA and a sequence on the target mRNA, and directs posttranscriptional gene silencing (PTGS) by two mechanisms: mRNA cleavage and mRNA translational inhibition[141]**.** In most cases, complementarity between miRNA and target mRNA is nearly perfect[142,143]. This high degree of complementarity is a requirement for cleavage[144], while repression of translation can occur at a lower degree of complementarity as well[145], although high complementarity between miRNA and target is generally a requirement in plants whatever the mechanism of action[146].

In both mechanism, AGO1 seems to be the main enzyme performing the action[123,145]. As already explained above, whether a miRNA performs one or the other mechanism is determined by the protein partnering with DCL1 during its processing: if it is HYL1, it will carry out cleavage; if it is DRB2, it will direct translational repression[105]. This explains the observation that the levels of a particular target can be regulated by both mechanisms at the same time, both carried out by the same miRNA[145,147]

In the cleavage pathway, AGO1, through the slicer activity of its PIWI domain[123,125], performs a cut in the mRNA between the nucleotides complementary to the 10th and 11th nucleotide of the miRNA[148–150]. The 3' fragment with an unprotected 5' P end is subsequently 5' to 3' degraded by the EXORIBONUCLEASE4 (XRN4)[149,151], while the 5' fragment, which harbours an unprotected 3' OH end, is proposed to be degraded 3' to 5' by the exosome[151], a process facilitated by the 3' uridylation of the fragment[152] by the AGO1 interacting enzyme HEN1 suppressor 1 (HESO1)[153]. Interestingly, besides undergoing degradation, some fragments generated from miRNA cleavage of coding and non-coding transcripts act as precursors of secondary, RDR-dependent sRNAs called phasiRNAs (including tasiRNAs), which seem to be involved in reducing target transcript levels like miRNAs do (Reviewed in[154]).

While endonucleolytic cleavage has been widely studied, the translational repression pathway is less well known. Besides the involvement of DRB2[105], other proteins such as VARICOSE (VCS)[145], KATANIN1 (KTN1)[145], SUO[155], ALTERED MERISTEM PROGRAM1 (AMP1)[156] and AMP1 homologue, LIKE AMP1 (LAM1)[156] have been reported to take part specifically in this second pathway, but their exact role is still not clear.

The involvement of AMP1, an integral endoplasmic reticulum (ER) membrane protein, and its ability to prevent the association of miRNA target transcripts with the membrane-bound polysomes, suggest that miRNAs may act by impeding the initiation of translation of their targets in the ER[156]. On the other hand, another study showing the association of AGO1 and a portion of mature miRNAs with polysomes may indicate that repression of translation can occur also at a post-initiation stage[157]. In fact, a later study shows that RISC can block the recruitment of ribosomes or stop their movement once they have bound the 5' untranslated region (UTR), being able to inhibit translation at both the initiation or elongation steps[158]. miRNA directed mRNA translational inhibition has been suggested to take place at the mRNA-processing bodies (P-bodies) as well as in the ER, since both VCS and SUO localize to these cytoplasmic foci[145,155,159]. Finally, the discovery of KTN1, a microtubule-severing enzyme, in the miRNA translational inhibition pathway suggests that the microtubule network plays a role in this mechanism as well[145], which is consistent with the previous observations, given that microtubules are associated to both ER and P-bodies[160].

Although miRNA mediated repression of protein synthesis can lead to subsequent mRNA degradation in animals[161], in plants it seems that miRNA directed mRNA degradation is widely triggered by endonucleolytic cleavage: although the involvement of VCS, a decapping activator, in plant miRNA translational inhibition raised the possibility that plant miRNA targets could subsequently undergo decay through deadenylation and decapping like is usual in animals[145], further research showed no evidence of degradation of miRNA targets triggered by translational repression or any other slicer independent mechanism[158,162].

Initially, target mRNA cleavage was thought to be the hugely predominant mode of plant miRNAs action, an idea probably supported by the wrong supposition that high complementary between miRNA and target had to lead to cleavage[142], and by the existence of only two examples on mRNA translation inhibition[163–165] compared with the many reports on cleavage (see[166] for a compilation), a bias possibly promoted by the easier way to study cleavage just by sequencing the 5' cleaved fragments[150]. However, in light of later publications, repression of translation may be less rare than previously considered, although the exact contribution of each pathway to the regulation of gene expression is still difficult to estimate[141]. Nevertheless, the importance of the cleavage pathway is clear from the fact that slicer-defective AGO1 is not able to complement the *ago1* mutant[167], while mutations in the repression of translation pathway seem to cause less ubiquitous effects[155]. A recent study suggests that each mechanism may contribute to the regulation of plant homeostasis in a different way: miRNA directed cleavage by regulating a wide range of processes like development, metabolism or environment adaptation and miRNA directed repression of protein synthesis by more specifically regulating the response to environmental factors[168]. In another study, it has been proposed that translational repression could have evolutionary preceded cleavage in plant miRNAs[105]. In any case, further research in this area would be needed to clarify the part played by each mechanism.

Besides these two mechanisms of posttranscriptional gene regulation, a few studies showing gene regulation at the transcriptional level by miRNA directed DNA methylation in a moss, rice and *Arabidopsis* have been published[83,169–171]. In two of these studies, it was shown how DCL3 processed, AGO4 loaded, 24 nt long miRNAs (lmiRNAs) were directing cytosine methylation, both at the *MIR* and miRNA target loci[83,171]. Interestingly, a previous study already suggesting DNA methylation of their target loci by miRNAs 165 and 166 matches these later discoveries[169], since *ago1* mutants didn't show altered methylation[169] and *MIR165* and *MIR166* were later found

to produce 24nt variants[82]. However, the knowledge in this particular pathway in plants is still at an early stage and the mechanism of action are still widely unknown.

1.2.3.6. Turnover

To maintain a correct function, it is inferred that the accumulation of miRNAs must be controlled not only at their biogenesis, as we have already seen, but also by a mechanism of degradation when their action is no longer desirable as a result of internal or external changes[172]. However, the mechanisms of miRNA turnover are only starting to be deciphered.

It is known that the methylation of the miRNA duplex 3' ends by HEN1[114,173] is necessary to avoid their uridylation and subsequent triggering of miRNA degradation[174]. The nucleotidyl transferase HESO1 has been reported to be the main enzyme adding 3'-oligouridylate tails to unmethylated miRNAs[116,117], with another nucleotidyl transferase, UTP:RNA URIDYLYLTRANSFERASE (URT1)[175–177], supposed to be collaborating in the uridylation of miRNAs as well. Nevertheless, the identity of the enzyme carrying out the degradation of uridylated miRNAs remains unknown, although ribosomal RNA-processing protein 6 (RRP6), an exosome component that degrades uridylated small RNAs in the green algae Chlamydomonas[178], has been proposed as a plausible candidate[118,179].

By studying *hen1* maize, rice and *Arabidopsis* mutants, it has been discovered that miRNAs 3' trimming is common before uridylation, and that both truncation and tailing of miRNAs are AGO1 dependent[180]. This observation, together with the fact that HESO1 and URT1 interact with AGO1[153,176], suggest that trimming and uridylation occur on AGO loaded miRNAs[180]. As already explained earlier, it has been proposed that HESO1 is recruited to RISC to uridylate cleaved mRNAs, and that miRNAs are protected from this uridylation thanks to their 2'-*O*-methyl 3' ends[153]. However, how this mechanism also ensures the elimination of impaired or no longer needed miRNAs is not clear.

A family of 3'–5' exoribonuclease called SMALL RNA DEGRADING NUCLEASES (SDN) has been reported to be involved in miRNA degradation as well, since simultaneous knock down of SDN1, 2 and 3 in *Arabidopsis* resulted in elevated miRNAs levels and developmental defects[181]. Moreover, SDN1 was found to specifically process small ssRNA molecules, even if they were 2'-*O*-methylated[181]. Although SDN1 inability to degrade uridylated substrates rules out the possibility of it

being the enzyme carrying out uridylation triggered miRNA degradation[181], it is possible that SDN1 works in a previous step to HESO1 action, truncating the 3' end of the miRNA and leaving an unmethylated end free to be urydilated[118,179]. However, this is only a possible scenario that would still need to be confirmed.

1.2.4. miRNA genes, origin and evolution

1.2.4.1. miRNA genes (*MIR*) and genomic organization

As mentioned previously, plant miRNA genes are usually independent intergenic transcriptional units[53,54]. However, a small percentage of plant *MIR* locate within transcription units of other protein-coding or non-coding genes, generally in intronic regions but at UTRs or exons as well[53,54,182]. Nevertheless, intronic miRNAs generated in a way similar to animal mirtrons, that is, by a spliced intron giving raise directly to the small pre-miRNA hairpin[183,184], have only three validated examples in plants[54,185,186]. Finally, some miRNAs deriving from plant transposable elements have also been identified[187].

Most times, plant *MIRs* appear scattered in the genome and produce a single mature miRNA each[53]. However, a portion of *MIR* are grouped in genomic clusters, which are generally composed of tandem miRNA paralogues[58,188,189]. In some cases, there are pri-miRNA transcripts which produce more than one miRNA each (poly-cistronic *MIR*)[190,191], and although these poly-cistronic *MIR* usually comprise miRNAs which belong to the same family[58,188,189], there are also examples of poly-cystronic *MIR* harbouring non-homologous miRNAs[57,59].

Unlike their animal counterparts, plant *MIRs* show a great variability in length and structure, with sizes from slightly over 300bp to almost 5kb reported[58,192,193]. Besides, they usually contain introns, with 67% of independently transcribed *MIRs* estimated to carry at least one intron in *Arabidopsis*[193], and intron lengths varying from less than 100bp to over 3000bp[58,192,193]. In the majority of cases of independent intron-containing *MIR*, miRNA stem-loops locate to the first exon[193,194], although there are examples of hairpins locating to other exons[193] as well as to regions that can be either intronic or exonic depending on alternative splicing[56,59]. Furthermore, there has been discovered a whole class of miRNAs in rice, called natural antisense miRNAs (nat-miRNAs) because they are transcribed from the antisense strand of their target protein-coding genes, which have an intron in between their miRNA* and miRNA[195].

Alternative transcription start sites, alternative polyadenylation sites and alternative splicing have been observed in many independent plant *MIR*s[59,62,80,192,196–200] as well as in protein coding genes which host a miRNA hairpin in an intron[56,201]. As also referred to earlier, the TATA box sequence characteristic of Pol II transcribed genes and several cis-regulatory elements have been found at *MIR* promoters[62,65,66]. *trans*-acting factors interacting with these motifs have been found to regulate *MIR* expression in response to different conditions or developmental states[67–71], determining different miRNA expression patterns even between *MIR* family members[67,69,71].

Finally, *MIR* can be present in a genome as single-copy genes or they may have one to several homologues, forming part of a *MIR* multigene family which would produce slight variations of the same mature miRNA[189], as will be further seen in the following sections. Besides, a single *MIR* locus can also produce multiple miRNA variants with small differences in their size or sequence, the so-called isomiRs[202–204]. isomiRs are produced by differential Dicer cleavage, trimming or addition of nucleotides at the miRNAs ends, which gives place to 5' or 3' isoforms, or, in rare cases, by RNA editing of internal nucleotides, which originate polymorphic isoforms[204]. In plants, a recent study shows how different isomiRs can regulate their targets distinctly[205], another report which supports the already suggested biological importance of isomiRs[204,206,207].

1.2.4.2. *MIR* origin

New *MIR* genes have been proposed to originate from three different sources:

1) Inverted duplication of their target loci

A first model to explain *MIR* emergence in the genome came from the observation that some young *MIR* show complementarity with their target genes not only at the miRNA sequence, but through all the hairpin arm. Thus, a plausible scenario is that such *MIR* would have arisen from the inverted duplication of such loci[208].

According to this model, the inverted duplication of a loci would be able to create a near perfect hairpin, which thereby would be processed by DCL enzymes other than DCL1, which is known to favour imperfect hairpins. Such a way of processing would produce a heterogeneous population of siRNAs which, if the parental locus was a coding gene, would target its mRNA with different levels of

complementarity. If any of these interactions happened to be advantageous and specific, the beneficial siRNA sequence would be evolutionarily selected unlike the rest of the hairpin, which would accumulate random mutations. This process would eventually lead to the production of a specific miRNA, thanks to miRNA and target site evolutionary selection and to a shift to processing by DCL1 due to a now imperfect stem-loop. Over time, old *MIR* would appear unrelated to their originating locus besides their miRNA/ miRNA* sequence[208,209].

Supporting this model, several examples of *MIR* showing similarity to other loci in the genome have been reported since the model was first proposed[54,209–212]. Besides, non-conserved miRNAs have been seen to generally show higher hairpin complementarity[213] and to be more frequently processed by DCL2, 3 or 4 over DCL1 compared with conserved miRNAs[54,82]. Furthermore, the fact that the expression levels of conserved miRNAs is generally higher than that of young miRNAs[54,209] agrees with this model as well, as that could limit initial off target effects of the siRNA population[210,214]. Likewise, the fact that plant miRNAs and their targets align through their whole length would also be in agreement[215].

Variations of this model have been observed as well, with cases in which both arms of the pri-miRNA hairpin show complementarity with the originating locus, suggesting that a intralocus inverted duplication event preceded the subsequent duplication which would originate the *MIR*[54,210]. In addition, it has been noted that many miRNA genes show homology with loci which are not predicted to target, some showing targets different than the parental locus and others being proposed to lack targets altogether[209,210].

2) Random genomic sequences

Although the origination of *MIR* from inverted duplications of other loci is the model which has accumulated the most supporting evidence, this model cannot explain the origin of all *MIR*, since a portion of young *MIR* show no homology with other genomic loci[216].

Plant genomes are predicted to harbour a high abundance of fold-back sequences[143], which in some cases could gain transcriptional regulatory elements and produce RNAs with stem-loops that could be recognised by the miRNA machinery, giving rise to a new miRNA gene[216]. According to this model, a subset of *MIR* would arise from spontaneous evolution of genomic regions which are either highly degenerated inverted duplications or just self-complementary by chance[216].

This model would imply that new *MIR* would initially lack targets and thus would be quickly lost through mutational drift in most of the cases, especially considering the low probability of finding a target by chance in plants, where whole length complementarity between miRNA and target is necessary for interaction[215]. This difference in the requirement of miRNA-target complementarity between plants and animals has been argued to explain why this mechanism of *MIR* creation seems to be so common in animals while little relevant in plants[217]. In any case, a small percentage of these spontaneously evolved miRNAs could happen to complement an existing transcript, and if the interaction had any advantage for the plant such miRNA could be fixed and become eventually conserved[216].

In support of this model, *MIRs* which target genes other than their identified parental sequence have been reported[209], which would demonstrate that the random acquisition of a target is a plausible scenario[215]. Besides, this model would fit the observation that, when comparing *A. thaliana* and *A. lyrata*, several miRNA genes present in one of the *Arabidopsis* species have only non-miRNA homologous loci in the other one[210,216,218]. This indicates sequence change in one linage, either new miRNA gain or old miRNA degeneration, and since most of these genes are not found in related species[210,216] the first option seems to be more probable[213,216].

3) Transposable elements

Growing evidence points towards transposable elements (TEs), and in particular miniature inverted-repeat transposable elements (MITEs), as a main source of new *MIR* formation in both plants and animals[219].

MITEs are non-autonomous DNA-type TEs which derivate from full length ones: like them they are flanked by inverted repeats, but instead of a long transposase open reading frame (ORF) they just contain a short DNA sequence[220,221]. When transcribed, they have the ability to fold into hairpins and according to this model some of these hairpins can be recognized by the miRNA machinery and processed into miRNAs[187,222]. In a similar way than the spontaneous evolution model, a few of these miRNAs could find a complementary mRNA by chance, and consequently be either positively or negatively selected depending on whether the resulting regulation was advantageous or detrimental for the plant[223].

Since the first observations of MITE derived miRNAs in plants[187,222], several more examples have been reported[223–229]. However, this model presents the challenge that miRNA and siRNA pathways seem to intertwine in the processing of MITE; both cis-acting TE-repressing siRNAs and miRNAs tend to be produced,

making it difficult to differentiate bona fide miRNAs[230]. While some scientists consider this model controversial for this reason, the authors propose that full-length TE generating siRNAs could have transition through evolution to miRNA producing MITE, and thus the existence of MITE giving birth to both siRNAs and miRNAs would represent an intermediary state which supports the siRNA to miRNA evolutionary transition[187]. See figure 1.3 for plant *MIR* possible origins.

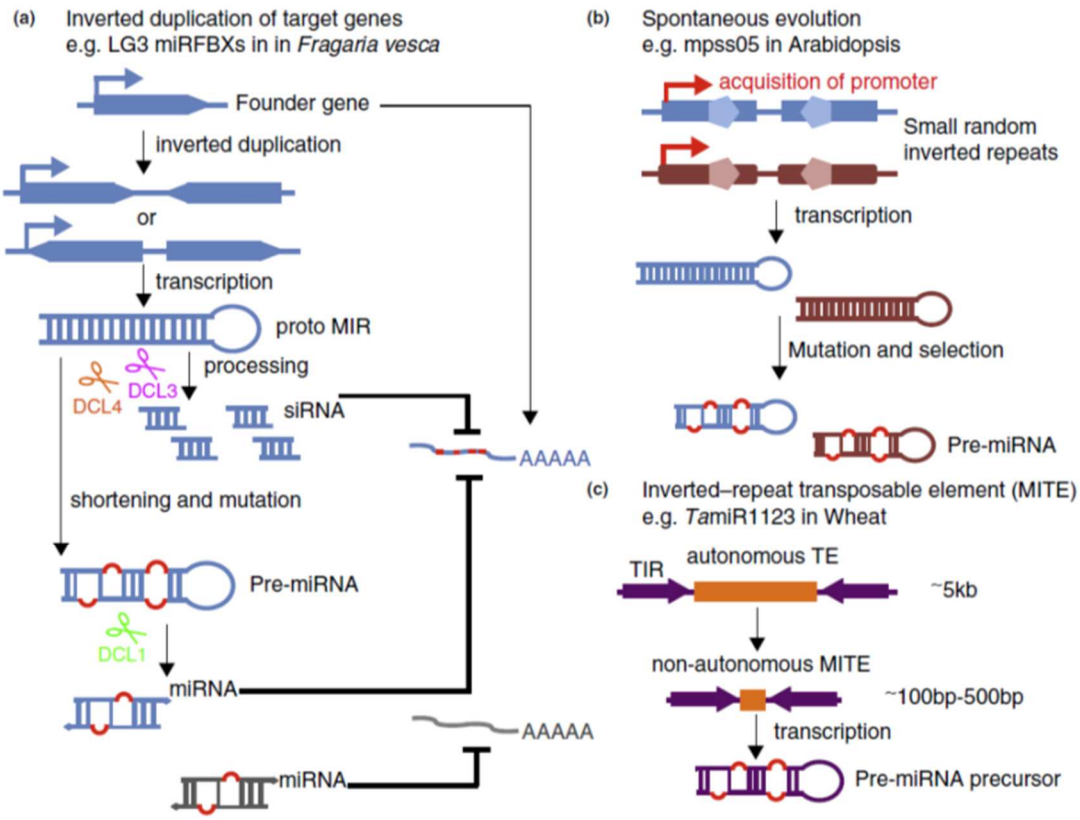


Figure 1.3. Three possible origins of new plant *MIR* genes. **A)** From inverted duplications of genes that will become the miRNA target. First, perfect hairpins are formed that will be processed by DCL different than DCL1 into siRNAs (proto *MIRs*). Then, mutations (red dots) accumulate in the hairpin and the hairpins start being processing by DCL1 into miRNAs. Several examples of miRNAs targeting their founder genes have been observed, although there are also cases in which a *MIR* originated from one gene target a different one. **B)** From random hairpin-forming transcribed genomic sequences. **C)** From MITE transposable elements that get recognised by the miRNA machinery. Figure from[250].

1.2.4.3. *MIR* evolution

Once *MIRs* arise *de novo*, local tandem duplications, chromosome segment duplications and whole genome duplications are responsible for the emergence of *MIR* gene paralogues which create multigene families[231]. These gene copies can keep on producing the same mature miRNA and thus increase the miRNA dosage, as seems to be the case in many plant gene clusters[57,225] or they can diverge and acquire new targets[232], different expression patterns[231] or just lose function[233].

MIR genes derived from a common ancestor form a family, a set of homologs which produce very similar mature miRNA sequences and generally show high similarity through their whole gene sequence as well[189]. At one extreme, there are *MIRs* belonging to multigene families conserved across all land plants[230,234], and a recent study suggests that some of them even between land plants and green algae as well[235]. On the other extreme, some *MIRs* appear as single copy species-specific genes[230,234]. All degrees of conservation appear in between, from *MIR* families present in all vascular plants, to those conserved in angiosperms or within a specific family only[230].

Most *MIRs* seem to be non-conserved, which implies that they are evolutionarily young *MIRs* and rapidly born and lost[54,209,234,236]. It has been suggested that, unlike highly conserved miRNAs that play fundamental roles and are not likely to be lost because they are positively selected, most of these young *MIRs* lack any relevant function; they are evolving neutrally and that is why they tend to be quickly lost by genetic drift[217,230,237]. This agrees with the observation that young *MIRs* tend to present low levels of expression[54], imprecise processing[218] and that the accumulation of their targets seems to be mostly unchanged in miRNA biogenesis mutants[209] or many just do not have identifiable targets at all[54,209,210]. Nonetheless, there are other possibilities rather than an absence of miRNA function that could explain the difficulty to predict and validate targets of non-conserved miRNAs. Such possibilities are that young miRNAs are interacting with their targets through different pairing rules than the ones used for the prediction[218,237], that they interact in a restricted spatiotemporal manner, like in a specific tissue or in response to certain stress[230], or that they are regulating they targets through translational inhibition[230].

Despite of the general support, the hypothesis of most *MIRs* being not conserved has been challenged based on two alleged pitfalls in the analyses leading to such affirmation. The first problem pointed out is that species-specific *MIRs* show a high percentage of annotated members which are not bona fide. The second

objection made is that *MIR* phylogenetics studies have been mainly performed using relatively distant plant species, so although the supposedly species-specific miRNA may not be conserved across all the evolutionary distance between the species analysed, it may be conserved among some other closer species[238].

In any case, once a miRNA has a target and this regulation is beneficial for the plant, selection will work to conserve this miRNA and to retain the regulatory interaction with the mRNA while the advantage lasts. Indeed, it has been confirmed that conserved miRNAs are under strong purifying selection, as well as their target sites and to a lesser extent their star miRNAs, with all of them showing lower evolutionary rates than their flanking regions[217,239–241]. Consistent with the hypothesis that most non-conserved miRNAs may lack targets and are consequently evolving neutrally while conserved miRNAs are playing important roles and as a result are under high selective constraint[217,230,237], conserved miRNAs show slower nucleotide variation rates than non-conserved ones[242,243].

Besides miRNA conservation, miRNAs and their targets have to coevolve to maintain their interaction, and in accordance to this there are several studies reporting how *MIR* evolutionary patterns are associated with the evolutionary patterns of their targets[243–246]. Nevertheless, conserved miRNA-target pairs can also be lost occasionally throughout evolution, as have been seen in certain phylogenetic linages[236].

Finally, the origin of the miRNA pathway in eukaryotes is still a matter of debate. Since no miRNAs conserved between animals and plants have been discovered and pathways in one and the other kingdom differ in their characteristics, the most supported hypothesis is that the miRNA system raised independently and converged in both linages[230,247–250]. Nonetheless, some authors argue otherwise based on the high *MIR* turnover observed in plants, that may have erased the tracks of common *MIR* sequences[251]. Further evidences would be needed thus to draw a conclusion.

1.3. Splicing in plants

1.3.1. Introduction

It was in 1977 that interrupted genes and splicing of mRNA were first discovered in adenovirus[5,6]. The phenomenon was confirmed in eukaryotes that very same year[252–254] and only one year later, the possibility that alternative splicing of introns and exons could create different mRNA variants of the same gene was proposed[255], to be observed experimentally shortly after[256,257].

Over the following years, great effort was made to discover the mechanisms and apparatus governing mRNA splicing. The chemical reactions necessary for splicing to occur were elucidated[258], as well as the machinery involved in the process[259–261].

When the study of alternative splicing at genomic and transcriptomic level became possible, the prevalence of the phenomenon in plants[262,263] and its key role in development and environmental response[264,265] was uncovered.

The differences between mRNA variants created by alternative splicing not only generate protein isoforms that may have different characteristics, but also can influence the transport, localization, stability and translation of the transcript[265–267]. Furthermore, the association of the splicing mechanism with other RNA processes such as nonsense-mediated decay and miRNA regulatory pathways are as well contributory to the whole regulation of gene expression that can be carried out through alternative splicing[265–267].

The importance of interrupted genes and splicing for eukaryotic evolution is becoming clear. It makes possible exon shuffling, a mechanism by which new genes are created by the rearrangement of exons[268,269]. It can also lead to alternative splicing, a phenomenon that expands proteome diversity and gene regulation, increasing coding capacity and evolutionary flexibility of genomes[267,270]. In agreement with this, it has been observed that alternative splicing contributes to organogenesis and speciation by creating different patterns of gene expression[271,272]. Furthermore, it has been speculated that alternative splicing may be a key contributor in the evolution of complexity[266,270,273], although publications claiming otherwise do also exist[274,275].

1.3.2. What is splicing?

Primary RNA transcripts undergo processing to become mature functional transcripts. Such processing includes 5' capping and 3' polyadenylation, and in intron-containing transcripts, also splicing. Splicing is the process by which introns are removed and exons are ligated together[266].

1.3.3. Types of Introns

There are four main types of introns: 1) tRNA or archaeal introns, which appear in archaea and tRNA genes in eukaryotes. They are removed by a cut and ligate enzymatic mechanism that makes them differ from all the other types of introns[276–278]. 2) Group I introns, which are present in bacteria and in eukaryotic organelles, as well as in nuclear rRNA genes in fungi and protists. They catalyse their own splicing through two transesterification reactions[278,279]. 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[278,280,281]. 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[278,282]. This type of introns can be further subdivided into U2-type and U12-type introns[283]. U2-type introns harbour the canonical GU and AG dinucleotides at the 5' and 3' end splicing sites respectively[284] and account for around 90% of introns in *Arabidopsis*[285] while U12-type introns present non-canonical splicing sites[286–288]. From now on we will focus on spliceosomal introns splicing.

1.3.4. The splicing process

1.3.4.1. Splicing machinery

Splicing is catalysed by the spliceosome[259–261], a RNP complex composed of five snRNAs that assemble with several proteins to create small nuclear ribonucleoproteins (snRNPs) plus up to hundreds of non-snRNP assisting proteins[289,290].

There are two types of spliceosomes: the so-called major or U2-dependant spliceosome, present in all eukaryotes, and the minor or U12-dependant

spliceosome, which coexist with the first one only in higher eukaryotes (including plants). Introns harbouring the canonical GT-AG splice sites (U2-type introns) are processed by the major spliceosome, with constituent RNPs U1, U2, U5 and U4/U6. In addition, the small subset of introns with non-canonical splicing sites (U12-type introns) are processed by the minor spliceosome, composed by the snRNPs U11, U12, U5 and U4atac/U6atac[289–291].

1.3.4.2. Splicing mechanism

The mechanism of splicing has been elucidated by studies in yeast and animals mainly, although later studies suggest that the process is very similar in plants, since core proteins and snRNAs involved are conserved across kingdoms[292], as also are conserved splicing consensus sequences[293,294].

The positions at which introns are excised and exons joined together are marked by consensus sequences called splice sites. The canonical 5' splice site (donor site) has a conserved GU while the canonical 3' splice site (acceptor site) is defined by an AG. Besides, introns present a polypyrimidine tract towards the 3' end and the so-called branch point 17-40 nt upstream of the acceptor site, the five nucleotide conserved sequence CURAY (R=purine and Y=pyrimidine)[295].

Splicing happens in two steps: In the first step, the 5' exon-intron junction is cleaved by the nucleophilic attack of the 2' OH group of a key adenosine at the intron branch site. A RNA intermediate called intron lariat is formed as a result. Subsequently, another nucleophilic attack takes place, this time from the newly created free 3' OH group at the upstream exon to the 3' intron-exon junction, which excise the lariat intron and ligate both exons[296]. See figure 1.4 for splicing mechanism.

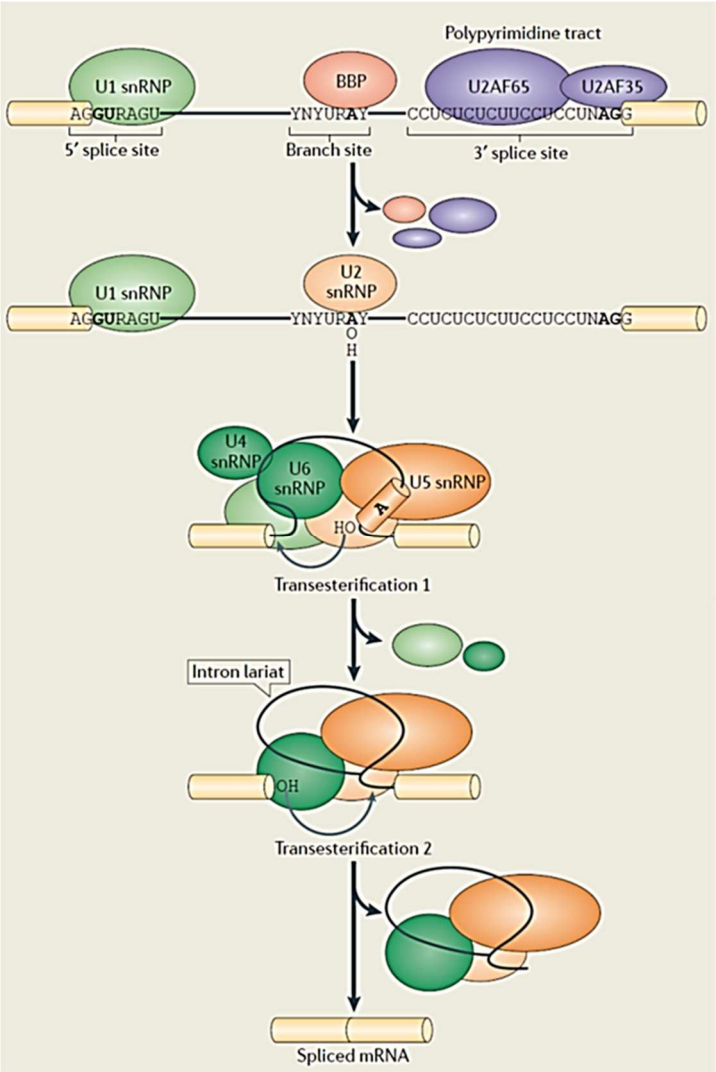


Figure 1.4. Splicing mechanism. The positions at which introns are excised and exons joined together are marked by consensus sequences called splicing sites (SS). The canonical 5' SS is GU and the canonical 3' SS is AG. Besides, introns present a polypyrimidine tract towards the 3' end and a conserved CURAY (plants)/ YNYURAY (mammals) sequence called branch point 17-40 nt upstream of the acceptor site. These sites are recognized in the pre-mRNA by the snRNPs U1, U2, U4, U5 and U6 and auxiliary factors (U2AF65, U2AF35 and BBP). Together, these factors form the spliceosome RNP complex, which catalyse splicing through two consecutive transesterifications. In the first one, the 5' exon-intron junction is cleaved by the nucleophilic attack of the 2' OH group of the key adenosine at the intron branch site. The so-called intron lariat, an RNA intermediate, is formed as a result. In the second transesterification, the newly created free 3' OH group at the upstream exon attacks the 3' intron-exon junction, excising the intron lariat and ligating both exons into the spliced mRNA form. Image from[266].

1.3.4.3. Splicing models

Splicing can be constitutive, in which case a single transcript variant is produced from a multi-exonic gene by always using the same set of splicing sites in each transcription[267].

In contrast, when splicing sites are sometimes not used or different ones are selected in different occasions, these leads to alternative splicing, where different transcripts are produced from a single gene[267].

Alternative splicing can be further divided into four different models: 1) Intron retention, in which splicing sites are sometimes not used and as a consequence an intron becomes an exon in a different transcript variant. This is the most common form of alternative splicing in plants, accounting for 40% of alternative splicing events in *Arabidopsis*. 2) Exon skipping, in which an exon may or may not be included in the mature transcript depending on which splicing sites are used. 3) Alternative 5' splice site, in which two 5' splice sites coexist and one or the other may be selected, creating a shorter or longer version of an exon. 4) Alternative 3' splice site, in which two alternative 3' splice sites are present and one or the other may be chosen in the same way that for the alternative 5' splice site model[267,285,297]. See figure 1.5 for different forms of alternative splicing and their frequency.

Frequently, splicing of a transcript involves the combination of several of these alternative splicing models, sometimes in the same splicing event. In fact, it is estimated that around 42% of splicing events are complex in *Arabidopsis*[267,285,297].

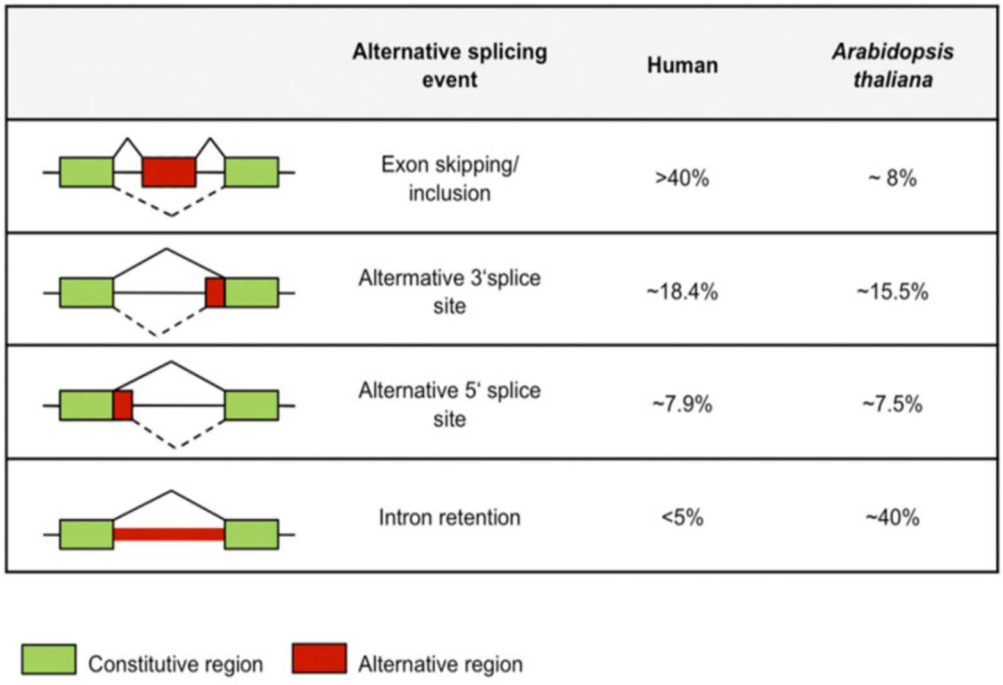


Figure 1.5. Types of alternative splicing and their frequency in humans and *Arabidopsis*. Exon skipping, in which an exon may or may not be included in the mature transcript depending on which SS are used. This is the most common form of alternative splicing in humans. Alternative 3' SS and alternative 5' SS, in which two 3' SS or 5' SS coexist and one or the other may be used, creating a shorter or longer version of an exon. Intron retention, in which SS are sometimes not used and as a consequence an intron becomes an exon in a different transcript variant. This is the most common form of alternative splicing in Arabidopsis. Figure from[267].

1.3.4.4. Splicing regulation and consequences

Despite initially thought to occur after transcription, there is now ample evidence showing that splicing takes place mainly co-transcriptionally, certainly in the case of constitutive splicing and usually for alternative splicing as well[298–300], although the removal of some alternatively spliced introns has been seen to happen post-transcriptionally[301,302]. This fact agrees with the current observations of splicing, transcription and chromatin structure influencing each other. Transcription is proposed to affect splicing through its involvement in the recruitment of splicing factors and through kinetic coupling, where the pace of elongation influences the splicing outcome[303]. In addition, chromatin can also influence splicing since certain histone modifications are associated with the recruitment of certain splicing factors and the positioning of nucleosomes seem to influence transcription kinetics too[304,305]

Apart from the influence of transcription and chromatin marks, the choice of whether a splicing site is used or not is mainly regulated by the interplay between cis-regulatory sequences and trans-acting splicing factors. There are four classes of cis-regulatory sequences: exon splicing enhancers, exon splicing silencers, intron splicing enhancers and intron splicing silencers. The trans-acting splicing factors are RNA binding proteins that join to these cis-regulatory elements and activate or inhibit the use of splicing sites by, for example, helping to recruit or to stabilise the spliceosome machinery[267,306,307]. The main families of splicing factors are the heterogeneous nuclear ribonucleoproteins (hnRNPs)[308] and Ser/Arg-rich (SR) proteins[309], that are involved in both constitutive and alternative splicing regulation; hnRNPs generally repressing and SR proteins generally activating splicing, although sometimes they may act otherwise depending on the context[306,307].

The finding of SR proteins being expressed differentially during development was probably one of the first lines of evidence of tissue-specific regulation of splicing in plants[310] while subsequent studies showing alternative splicing of SR transcripts themselves in response to development, hormones or abiotic stresses confirmed the correlation of splicing factors with biotic and abiotic cues[311–313].

Further studies have revealed that alternatively spliced transcripts can be related to cell type, developmental stage, circadian rhythm, pathogen infections or external conditions such as temperature, light or soil composition[264,265].

Besides, it is known that splicing interferes with post-transcriptional pathways. Alternative splicing has been reported to introduce premature termination codons in some transcripts, which usually determine their degradation through the nonsense-

mediated decay pathway and this cross-talk is thereby a way of regulating the levels functional transcripts[314,315]. Alternative splicing can also have an influence in the levels of mature miRNA, not only for intronic miRNAs[196,316], but as well for other intron-containing pri-miRNAs harbouring the miRNA sequence in an exon[197,198], as has been reported several times and we will further explore in section 4 of this chapter.

All these different facets of splicing form a picture of this process far more complex than originally thought. Splicing does not only produce different protein variants but also influences the abundance and fate of the transcripts by interacting with other cellular mechanisms and signals, and all of this under a strict regulation directed by both biotic and abiotic factors[264,265,267].

1.3.5. Origin and evolution of introns and splicing

The origin of introns remains a mystery to this day. However, several conflicting theories about their origin have been proposed: 1) “Introns-early” theory, which postulates that introns were already present in the ancestor of prokaryotes and eukaryotes and played a key role in the creation of the first genes by exon shuffling[255,268,317]. 2) “Introns-late” theory, which holds that spliceosomal introns emerged in eukaryotes through insertion into already functional genes of group II self-splicing introns from proto-mitochondria[318,319]. 3) “Introns-first” theory, which is a version of the “introns early” theory and sustains that introns are a vestige from the RNA world, in which exons would have raised from non-coding regions present between functional RNA genes[320,321]. 4) Finally, there are also publications suggesting a mixed model between the first two theories, supporting the existence of old and new introns that could have been originated through each one of the two mechanisms[322–324].

Although the primary origin of introns remains controversial, what is clear now is that at least some introns are recently acquired, since the recent gain of an intron has been reported in several occasions in both plants and animals[325–330]. In fact, it is probable that the mechanisms by which new introns proliferate are different from the ones by which they were first integrated in the eukaryotic genomes[331,332].

Among the proposed mechanisms for new intron gain are: 1) Group II intron insertion from an organelle into the nuclear genome, and subsequent processing through the spliceosome and thereby conversion into a spliceosomal intron[333,334]. As we have already seen, this mechanism is also proposed to explain the primary origin

of introns in eukaryotic genomes, but has also been suggested to account for recent gain of introns[335]. 2) Spliceosomal intron transfer between paralogous through homologous recombination, possibly using a reverse-transcribed non-spliced mRNA as substrate[327]. 3) Spliceosomal intron transposition, in which a spliced out intron is reverse spliced into the same or another transcript, and this transcript is subsequently reverse-transcribed into a cDNA that could recombine with the genome inserting the intron[336]. 4) Transposon insertion into a gene, which would be later removed by the spliceosome creating a new spliceosomal intron[337,338] 5) Tandem duplication of a exonic region harbouring an AGGT sequence that, after being duplicated, could be identified by the spliceosome as the 5' and 3' splicing sites of an intron, while the coding region would remain intact[335]. 6) Intronization, or the creation of an intron in a previously exonic region due to the generation of newly recognised splicing sites through point mutations[339].

Intron loss does contributes to the evolution of eukaryotic genes through mechanisms such as 1) Homologous recombination of the cDNA from an intronless transcript with its genomic locus[340] 2) genomic deletion, that can be of individual introns[341].

Besides, since it became apparent that homologous genes do not always harbour their introns in homologous position, the hypothesis of intron sliding was proposed, by which an intron would be able to relocate over short distances[342]. This hypothesis gained support as the phenomenon of alternative splicing became better understood, since the process could be driven by the shifting between alternative splicing sites in a given sequence, given that the emergence of new splicing sites by point mutations seems to be fairly common[343]. See figure 1.6 for examples of intron gain, loss and sliding.

The evolution of splicing is as controversial as that of introns. Besides group I and II self-splicing introns, the introns present in eukaryotic nuclear genomes need a spliceosome to be processed. Both spliceosomal components and introns have been identified in ancient eukaryotic linages, suggesting that at least a proto-spliceosome was present in the last eukaryotic common ancestor (LECA)[344–346]. According to the "introns-late" theory, the invasion of group II introns into the eukaryotic genome created the pressure to evolve a machinery to remove these introns, the spliceosome complex, at the same time that these group II introns gradually lost their capacity to self-splice[346,347]. However, the "introns-first" theory sustains that the first spliceosome-like machinery long predated the origin of eukaryotes and was an integral part of the

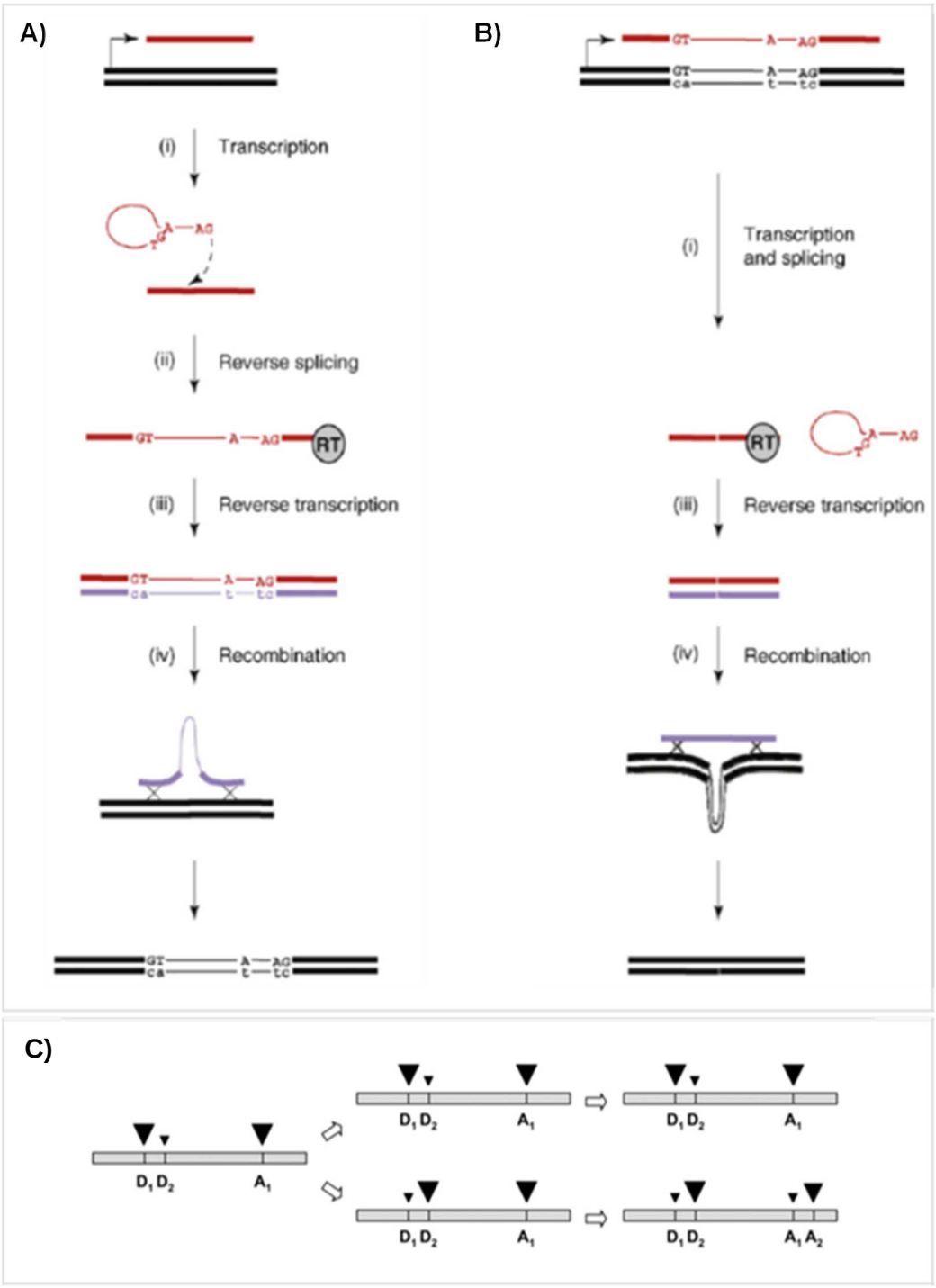


Figure 1.6. Models of intron gain, loss and sliding. **A)** Intron gain through spliceosomal intron transposition. An intronic RNA lariat is reverse spliced into a novel site of the same or another transcript. This transcript is then reverse transcribed into a cDNA that undergoes recombination with the genome, generating an intron gain allele. **B)** Intron loss through homologous recombination of the cDNA from a spliced transcript with its genomic locus, yielding an intron loss allele. **C)** Intron sliding from one position to a nearby one via AS. Divergence of two genes from a common ancestor after duplication, speciation, etc. D=donor site (5'SS). A=acceptor site (3'SS). Relative strength of the alternative SS is indicated by arrowheads of different sizes. While upper *MIR* remains unchanged, in the lower *MIR* point mutations can lead to the change of a SS from weak to strong (see D1 and D2) or to the emergence of a new SS (see A2), drifting the intron from D1-A1 to A2-D2 position. Figure A and B from[439]. Figure C from[343].

ribo-organisms living in the RNA word, to which the modern spliceosome is a remnant[320,321]. Great advance in the knowledge about the first stages of life would be needed to reach a conclusion.

Finally, how and when alternative splicing evolved is not straightforward either. Originally, it was suggested that it could have evolved from constitutive splicing through mutations at splicing sites that weakened them and/or through the evolution of splicing regulatory factors[348]. However, later research showing that the splicing sites of first eukaryotes were degenerate rather conserved has made more plausible an scenario in which alternative splicing was present, together with splicing itself, at least from the beginning of eukaryote evolution[349–351]. What seems to be clear is that alternative splicing has been favoured in complex multicellular organisms, which show much higher rates of alternative splicing than lower multicellular organisms, and of course than unicellular organisms, where alternative splicing is almost negligible[266,267].

1.4. Aims and objectives

The discovery of a miRNA with an unconventional intron-split pri-miRNA structure in tomato opened the possibility to a broader study. Its special feature made this miRNA a perfect candidate to be studied, since any findings could shed light in some of the less known aspects of miRNA biology, such as the regulation of miRNA biogenesis and of the levels of mature miRNA. With this aim, we have conducted the present multi-perspective study of this miRNA, called miRNAtop14.

First, we tried to determine the phylogenetic distribution of miRNAtop14. In this analysis we assessed not only the presence of the mature miRNA sequence in a species, but also the conservation of its pri-miRNA intron-exon structure and of its resulting secondary structure. Besides delimiting in which species this miRNA is present, the conservation of the intron and its position could be indicative of its importance in regulating the miRNA.

Subsequently, we tried to find out how the presence of the intron in between miRNA and miRNA* could affect the levels of mature miRNAtop14 and thereby be a potential step of post transcriptional regulation. With this aim we tested different tomato pri-miRNAtop14 splicing variants to evaluate their different outcomes.

Finally, we aimed to find out the targets of miRNAtop14. Such discovery would, in first place, indicate the function of miRNAtop14 and its importance for the plants containing it, which could be relevant from an applied point of view. Besides, this information could be joined with the data extracted from the previous investigations in order to design a study to determine whether the intron is ultimately influencing the biological role of miRNAtop14.

In conclusion, this study is an effort to broaden the knowledge in the field of miRNA biology and more specifically in the areas of miRNA evolution and biogenesis. More broadly, the current thesis tries to be a modest contribution towards the understanding of cell RNA biology itself, specifically in the overlapping area between miRNAs and splicing in plants.

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. Ailsa Craig, *Solanum lycopersicum* cv. MicroTom, *Nicotiana benthamiana*, *Petunia axillaris* line S26, *Ipomoea nil* cv. Kikyo-zaki, *Arabidopsis thaliana* ecotype Columbia (Col-0) and *Oryza sativa* ssp. Japonica cv. Nipponbare.

2.1.2. Total RNA extraction

RNA extraction was performed with Tri-reagent following the manufacturer's protocol (Ambion). The only differences were that after phenol-chloroform separation, 3 volumes of 100% ethanol instead of 1 volume of isopropanol were added to the recovered aqueous phase, followed by 10 minutes centrifugation and two washes with 75% ethanol instead of only one. In brief, tissue is frozen and grinded in liquid nitrogen and 1-2 mL of Tri-reagent are added to each 100 mg of tissue and incubated to dissociate nucleoprotein complexes. Then, RNA is isolated from DNA and proteins through the addition of chloroform and subsequent phase separation. Finally, the upper aqueous phase containing the RNA is recovered, and the RNA is precipitated through addition of 100% ethanol, washed two times with 75% ethanol and air dried before being dissolved in distilled water. The concentration and quality was measured using a NanoDrop spectrophotometer (Thermo Fisher scientific) at an absorbance ratio of A260/280 and A260/230 nm.

2.1.3. pGEM-t easy cloning method

After electrophoresis, bands to be cloned were cut from the gel and DNA was recovered using the Zymoclean Gel DNA Recovery Kit, following manufacturer's protocol (Zymo research). The concentration of the DNA was afterwards measured using a NanoDrop spectrophotometer (Thermo Fisher scientific).

Subsequently, if PCR amplification was carried out using a proofreading polymerase like Phusion (New England Biolabs), that do not leave overhangs, A tails were added to the amplicon molecules by incubating the PCR product at 70°C for 30

min. with 1µl GoTaq Flexi DNA polymerase (5 U/ µL, Promega), 1µl 2mM dATP, 1µl 25mM $MgCl_2$, 2µl 5X GoTaq Flexi buffer and water up to a total volume of 10µl.

Then, the tailed PCR amplicons were ligated into pGEM-T Easy vectors according to the ligation protocol provided by the pGEM-T Easy Vector Systems manufacturer.

Afterwards, *Escherichia coli* DH5α super-competent bacteria were transformed with this plasmids through a heat-shock method which consisted in mixing 20 µL of competent bacteria and ~1 µL of plasmid (20-200ng/µL), incubating them for 20 min. on ice, heat-shocking the cells for 45-50 seconds at 42°C and returning them to ice again for another 2 min. before adding 500µL of room-temperature LB medium.

The LB suspension was then incubated for 1-1.5 hours at 37°C with shaking and subsequently 100-200 µL of each transformation culture was plated onto duplicate LB/ampicillin/IPTG/X-Gal Petri dishes. Plates were incubated at 37°C for 16-24 hours. After this time, colonies were visible and those carrying an insert could be identified through blue/white screening.

Finally, to test whether a white colony was carrying the right insert, colony PCR, sequencing or both procedures were carried out. Colony PCR was performed as explained below (2.1.4) and sequencing was performed by Eurofins MWG Operon (Germany). For sequencing, plasmid DNA was isolated through miniprep using the QIAprep Spin Miniprep Kit. Apart from the first centrifugation step, in which overnight cell culture was pelleted by centrifuging at 4000g for 5 min. at 4°C, manufacturer's instructions (Qiagen) were followed.

2.1.4. Colony PCR

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

Each reaction made a total volume of 10 µL and contained 5.95 µL H2O, 2 µL 5X Green GoTaq flexi reaction buffer, 0.8 µL $MgCl_2$, 0.2 µL dNTPs (10 mM each), 0.5 µL of each primer (10 µM) and 0.2 µL GoTaq G2 DNA Polymerase (5 U/ µl, Promega). The PCR program consisted in a first cycle of 95°C for 2 min, followed by 30 cycles of 95°C for 30 sec, 50°C for 30 sec and 72°C for 1 min/ kb, followed by a final cycle of 72°C for 5 min.

2.1.5. 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 heating to mix, and then adding 2.75 ml of 40% acrilamide/ bis solution 19:1 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 was mixed with equal volume of gel loading buffer II (Ambion) and heated 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 Healthcare 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). Gel and membrane were stacked in between 0.5X TBE soaked Whatman filter paper. Transfer time was usually 35 min at 190 mA for one ~60 cm^2 gel.

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

The membrane was then hybridized with a ^{32}P end labelled DNA probe. For probe labelling, 2 µl of the probe (10µM, without 5'P) was mixed with 10 µl of water, 4µl of 5X forward buffer, 3µl of [γ-^{32}P]ATP (≥ 3,000 Ci/mmol) and 1µl T4 polynucleotide kinase (Life Technologies) and incubated at 37°C for 1 hour. After that, 30 µl of water were added to the mixture and all resulting 50 µl solution were 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 sent back to the hybridization oven at 37°C and with rotation, to let the hybridization proceed overnight.

The day after, the hybridization solution was washed away with 0.2x SSC, 0.1% SDS water solution. After a first rinse, the tube was washed twice for 20 min at

37°C in the hybridization oven and then with a final quick wash again. The membrane was then wrapped in Saran film and placed in a cassette facing a phosphor screen. After 6 or more hours of exposure, the film 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.

2.2. Material and methods chapter 3

2.2.1. PCR analysis of *MIRtop14* genomic locus (fig. 3.2.)

Genomic DNA was extracted from 6 leaf-discs (~60mg tissue) of *Solanum lycopersicum* cv. MicroTom, *Nicotiana benthamiana*, *Petunia axillaris* and *Ipomoea nil* grown in the conditions specified before (2.1.1). gDNA was extracted following Amani et al. protocol[352], after freezing and grinding the tissue in liquid nitrogen first. The concentration and quality of the DNA was measured using a NanoDrop spectrophotometer (Thermo Fisher scientific) at an absorbance ratio of A260/280 nm and A260/230 nm.

Each 20 µL PCR reaction contained 12.4 µL H_2O, 4 µL 5X Phusion GC buffer, 0.4 µL dNTPs (10 mM each), 1 µL of each primer (10 µM), 1 µL cDNA (10 ng/ µL), 0.2 µL Phusion DNA Polymerase (2.5 U/ µL, New England Biolabs). The PCR profile was as follows: 98°C for 30 sec, followed by 35 cycles of 98°C for 10 sec, 61°C for 20 sec and 72°C for 30 sec, followed by a final cycle of 72°C for 5 min.

PCR products were run in a 1.5% Agarose 0.5X TBE gel and a picture was taken with Typhoon FLA 9500 scanner (GE Healthcare Life Sciences) after EtBr staining. Bands were cut and cloned into pGEM-t easy vector according to protocol described in section 2.3.1. Finally, positive clones were sent for sequencing to Eurofins MWG Operon (Germany) and identity of the sequences was confirmed through BLAST alignment.

MIRtop14 sequences and primers are shown in appendix.

2.2.2. Northern blot detection of miRNAtop14 in different species (fig. 3.5.A)

For Northern blot analysis of the four Solanales species and *Arabidopsis thaliana* (fig. 3.5.A), seeds of *Solanum lycopersicum* cv. Ailsa Craig, *Nicotiana benthamiana*, *Petunia axillaris*, *Ipomoea nil* and *Arabidopsis thaliana* were sown in soil and grown at the temperature and light conditions indicated earlier (2.1.1). *S. lycopersicum* plants were grown for 24 days, N. *benthamiana* for 32 days, *P. axillaris* for 28 days, *I. nil* for 19 days and A. thaliana for 20 days. The aerial parts of three plantlets per species were used for RNA extraction.

Total RNA was extracted from each sample following the protocol in section 2.1.2 and subsequently, 10µg were analysed by sRNA Northern blot (section 2.1.5). The probe used was complementary to miRNAtop14 last 20 nucleotides, which are predicted to be the same in all four Solanales species analysed (only the first miRNAtop14 5' end nucleotide changes between species). The film was scanned using the Typhoon FLA 9500 scanner (GE Healthcare Life Sciences)

2.2.3. RT-PCR analysis of pri-miRNAtop14 (fig. 3.5.B)

The same total RNA samples used for *S. lycopersicum*, N. *benthamiana* and *P. axillaris* miRNAtop14 northern blot detection (2.2.2) were used for reverse-transcription and PCR analysis of pri-miRNAtop14.

For N. *benthamiana*, mRNA isolation from total RNA was carried out previous to reverse transcription using the Dynabeads mRNA purification kit, following the manufacturer's instructions (Life Technologies). Then, mRNA was reverse transcribed using SuperScript II reverse transcriptase following manufacturer's protocol (Invitrogen), with a few differences: the total volume after mixing resuspended RNA, oligo dT and dNTPs was adjusted to 13µL instead of to 12 µL with water, and only 1 µL of DTT instead of 2 µL was later added to the reaction. Besides, incubation at 42°C was carried out for 60 min. 1 µL of Generacer oligo dT primer (50 µM, 5'-GCTGTCAACGATACGCTACGTAACGGCATGACAGTGT_{24}-3') was used for the reverse transcription.

After reverse transcription, 1 µL RNase H (2 U/ µL, Invitrogen) was added to each reaction and the mixture was incubated at 37°C for 20 min.

cDNA was then PCR amplified in a 50 µL reaction containing 32 µL H_2O, 10 µL 5X Phusion GC buffer, 1 µL dNTPs (10 mM each), 2.5 µL of each primer (10 µM), 1.5 µL cDNA and 0.5 µL Phusion DNA Polymerase (2.5 U/ µL, New England Biolabs). The PCR program consisted of: 98°C for 30 sec, followed by 30 cycles of 98°C for 10 sec, 66°C for 10 sec and 72°C for 10 sec, followed by a final cycle of 72°C for 5 min.

PCR products were run in a 2% Agarose 0.5X TBE gel and a picture was taken with Typhoon FLA 9500 scanner (GE Healthcare Life Sciences) after EtBr staining. Bands were cut and cloned into pGEM-t easy vector according to protocol described in section 2.3.1. Finally, positive clones were send for sequencing to Eurofins MWG Operon (Germany) and identity of the sequences was confirmed through BLAST alignment.

For *S. lycopersicum* and *P. axillaris*, total RNA was used directly for PCR, but it was previously subjected to DNase treatment using the TURBO DNA-free Kit and following the manufacturers' instructions (Ambion). 6 µg of total RNA were DNase treated in a 10 µL reaction, after which DNAse was inactivated and RNA recovered as explained in the manufacturers' protocol.

Reverse transcription was then performed using SuperScript II reverse transcriptase following manufacturer's protocol (Invitrogen), with the only difference that the incubation at 42°C was carried out for 60 min. 2 µl of DNase treated total RNA (~ 1 µg) were used in the reverse transcription reaction in both species. For *S. lycopersicum*, 1 µl of 2 µM *Sly-MIRtop14* reverse primer (same used for gDNA and following cDNA PCR) was used for the reverse transcription. For *P. axillaris*, 1µl of dT_{15} primer (487µg/ µL) was used for the reverse transcription. Similar reactions but without reverse transcriptase added were also performed for both species, to have a control for any DNA contamination after PCR.

PCR reaction was subsequently carried out using both the reaction with and the reaction without reverse transcriptase in the two species. PCR conditions for each species were as follows.

S. lycopersicum 15 µL PCR reaction contained 9.3 µL H_2O, 3 µL 5X Phusion GC buffer, 0.3 µL dNTPs (10 mM each), 0.75 µL of each primer (10 µM), 0.75 µL cDNA and 0.15 µL Phusion DNA Polymerase (2.5 U/ µL, New England Biolabs). The PCR profile was as follows: 98°C for 30 sec, followed by 30 cycles of 98°C for 10 sec, 60°C for 20 sec and 72°C for 30 sec, followed by a final cycle of 72°C for 5 min.

P. axillaris 20 µL PCR reaction contained 12.4 µL H_2O, 4 µL 5X Phusion GC buffer, 0.4 µL dNTPs (10 mM each), 1 µL of each primer (10 µM), 0.6 µL DMSO, 0.4 µL cDNA and 0.2 µL Phusion DNA Polymerase (2.5 U/ µL, New England Biolabs). The PCR program was as follows: 98°C for 30 sec, followed by 35 cycles of 98°C for 10 sec, 62°C for 30 sec and 72°C for 30 sec, followed by a final cycle of 72°C for 5 min.

PCR products were run in a 1.5% Agarose gel and a picture was taken with Typhoon FLA 9500 scanner (GE Healthcare Life Sciences) after EtBr staining. Bands were cut and cloned into pGEM-t easy vector according to protocol described in section 2.3.1. Finally, positive clones were send for sequencing to Eurofins MWG Operon (Germany) and identity of the sequences was confirmed through BLAST alignment.

MIRtop14 sequences and primers are shown in appendix.

2.2.4. Northern blot detection of miRNAtop14 in different tissues (fig. 3.6)

For Northern blot detection of miRNAtop14 levels in tomato root, stem, leaves and leaflets (fig. 3.6), *S. lycopersicum* cv. Microtom was used. Root, stem, leaves and leaflets were collected from plants grown on soil for ~5 weeks.

Total RNA was then extracted from each tissue following the protocol in section 2.1.2.

2µg of total RNA from each sample were analysed by sRNA Northern blot (section 2.1.5). The probe used was the full 21 nucleotides complementary to Sly-miRNAtop14 and, after stripping, an oligonucleotide complementary to U6 RNA (5'-AGGGGCCATGCTAATCTTCTC-3'). The film was scanned using the Typhoon FLA 9500 scanner (GE Healthcare Life Sciences).

2.3. Material and methods chapter 4

2.3.1. Cloning of Sly-pri-miRNAtop14 sequences

Solanum lycopersicum L. cv. MicroTom seeds were sowed in soil and grown 9 weeks. The three youngest but completely open leaves were collected for RNA extraction, each one from a different tomato plant. Total RNA was extracted from each sample following the protocol in section 2.1.2 and subsequently, mRNA isolation was carried out using the Dynabeads mRNA purification kit, following the manufacturer's instructions (Life Technologies).

Then, mRNA was reverse transcribed using SuperScript II reverse transcriptase following manufacturer's protocol (Invitrogen), with a few differences: the total volume after mixing resuspended RNA, oligo dT and dNTPs was adjusted to 13µL instead of to 12 µL with water, and only 1 µL of DTT instead of 2 µL was later added to the reaction. Besides, incubation at 42°C was carried out for 60 min. 1µl of dT20 primer (50 µM) was used for the reverse transcription.

After reverse transcription, PCR was carried out directly to amplify Sly-*MIRtop14* two transcript variants. The 50 µL PCR reaction contained 32 µL H2O, 10 µL 5X Phusion HF buffer, 1 µL dNTPs (10 mM each), 2.5 µL of each primer (10 µM), 1.5 µL cDNA and 0.5 µL Phusion DNA Polymerase (2.5 U/ µL, New England Biolabs). The PCR program was as follows: 98°C for 1 min, followed by 40 cycles of 98°C for 10 sec, 64°C for 15 sec and 72°C for 20 sec, followed by a final cycle of 72°C for 5 min.

PCR products were run in a 1.5% Agarose 0.5X TBE gel and a picture was taken with Typhoon FLA 9500 scanner (GE Healthcare Life Sciences) after EtBr staining. Bands were cut and cloned into pGEM-t easy vector according to protocol described in section 2.3.1. Finally, positive clones were send for sequencing to Eurofins MWG Operon (Germany) and identity of the sequences was confirmed through BLAST alignment.

MIRtop14 sequences and primers are shown in appendix.

For golden gate cloning, both pri-miRNAtop14 sequences (with and without intron) were amplified from pGEM-t easy carrying plasmid. Miniprep was performed using the QIAprep Spin Miniprep Kit and 1 µL of 10 ng/ µL isolated plasmid were used as PCR template. Apart from that, the reaction mixture was the same as in the previous PCR which product was cloned in pGEM-t easy. The primers used had the

same *MIRtop14* matching sequence as the ones used in the previous PCR and shown in the appendix, but these ones had tails for golden gate cloning. The PCR program used this time was as follows: 98°C for 2 min, followed by 35 cycles of 98°C for 10 sec, 63°C for 30 sec and 72°C for 30 sec, followed by a final cycle of 72°C for 10 min.

2.3.2. Cloning of *Osa-MIR528* gDNA sequence

Oryza sativa was germinated in petri dishes over Whatman filter paper soaked in distilled water. After 12 days, sprouts were used for genomic DNA extraction according to Amani et al. protocol[352], after freezing and grinding the tissue in liquid nitrogen first. The concentration and quality of the DNA was measured using a NanoDrop spectrophotometer (Thermo Fisher scientific) at an absorbance ratio of A260/280 nm and A260/230 nm.

Subsequent PCR amplification of *Osa-MIR528* genomic sequence was carried out in a 20 µL PCR reaction containing 12.3 µL H2O, 4 µL 5X Phusion GC buffer, 0.4 µL dNTPs (10 mM each), 1 µL of each primer (10 µM), 0.5 µL gDNA (100ng/ µl), 0.6 µL DMSO and 0.2 µL Phusion DNA Polymerase (2.5 U/ µL, New England Biolabs). The PCR program was as follows: 98°C for 30 sec, followed by 35 cycles of 98°C for 10 sec, 56-65°C for 15 sec and 72°C for 90 sec, followed by a final cycle of 72°C for 5 min.

PCR products were run in a 1.5% Agarose 0.5X TBE gel and a picture was taken with Typhoon FLA 9500 scanner (GE Healthcare Life Sciences) after EtBr staining. Bands were cut from the gel and DNA was recovered using the Zymoclean Gel DNA Recovery Kit, following manufacturer's protocol (Zymo research). Subsequently, the recovered PCR product were assembled into a level 0 Golden Gate cloning vector and cloned according to the protocol described in next section (2.3.3). Positive clones were send for sequencing to Eurofins MWG Operon (Germany) and identity of the sequences was confirmed through BLAST alignment.

The sequences of the primers used for the amplification can be seen hereafter, with the nucleotides annealing to *O. sativa* gDNA in capital letters and the tails needed for golden gate cloning in lower case letters.

Forward primer: 5'-tgaagacggaatgCCAGTGCACCATGGCCGG-3'

Reverse primer: 5'-tgaagacggaagcTTGTCGTTGACAATACTACTCTTCT-3'

2.3.3. Golden gate cloning

The golden gate cloning method followed in this study was based in the golden gate system published by Weber et al.[353]. This system allows the assembly of multiple DNA fragments into a single insert piece thanks to several hierarchical assemblies facilitated by the cleavage by type II restriction enzymes, which cleave outside their recognition sites. This way, recognition sites can be cleaved out of the constructs after each assembly reaction.

Besides, the enzymatic restriction will produce specific four nucleotide overhangs that will pair only with other complementary overhangs, allowing the assembly of multiple fragments simultaneously in a fixed order.

In brief, simple modules (such as promoter, terminator, coding sequences, etc.) have to be first amplified with primers carrying a tail with an enzyme II recognition site followed by the distinct nucleotide sequence that will give place to the different overhangs in each case. This way, through restriction-ligation assembly reaction described hereunder, simple modules are cloned into level 0 vectors. Several of these simple modules harboured in level 0 vectors are then cut and assembled into a level 1 vector to create a full transcriptional unit. Finally, in the same way, several transcriptional units contained in a level 1 vector each can be cut and assembled together in a multigene piece into a level 2 vector.

Each assembly mixture contained 1 µL of vector backbone (100 ng), 1 µL of each additional assembly plasmid (100ng), 1.5 µL of 10x T4 ligase buffer, 0.15 µL 100X BSA, 1 µL T4 Ligase (400,000 cohesive end units/ ml, New England Biolabs), 1 µL type II restriction enzyme (Bpi I for level 0 and level 2 assembly, Bsa I for level 1 assembly) and water up to 15 µL.

The reaction consisted in 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.

The products of each reaction were then cloned into *E. coli* DH5α by the same heat-shock method used for pGEM-t easy cloning (described in section 2.1.3). For level 0 and level 1 constructs, positive colonies were recognized through blue-white screening. For level 2, red-white screening was used.

Finally, to test whether a white colony was carrying the right insert, colony PCR, sequencing or both procedures were carried out to test level 0 and level 1 constructs. Besides, level 2 constructs were checked by restriction digest additionally to specific modules sequencing, in order to confirm that the whole assembly was

correct given the impossibility to amplify the large multi-gene fragment. Colony PCR was performed as previously explained (2.1.4) and sequencing was performed by Eurofins MWG Operon (Germany). For sequencing and restriction digest, plasmid DNA was isolated through miniprep using the QIAprep Spin Miniprep Kit. Apart from the first centrifugation step, in which overnight cell culture was pelleted by centrifuging at 4000g for 5 min. at 4°C, manufacturer's instructions (Qiagen) were followed.

2.3.4. Site-directed mutagenesis

To create a non-spliceable intron containing *MIRtop14* construct, the 5'SS of the gene was mutated from G/GT to C/CC through site directed mutagenesis from the level 0 golden gate construct harbouring the WT *MIRtop14*.

First, primers were design to amplify the whole level 0 construct, but incorporating a mutation at the *MIRtop14* 5'SS:

Forward primer: 5'-GTTTAATTTATTAC**CCC**ATGTTATTTGTC-3' (it comprehends the 5' splicing site sequence, but harbours a CCC sequence at this position instead of the original GGT)

Reverse primer: 5'-AAACAATATTGATAAGCACTCTTT-3' (it is adjacent to the forward primer, immediately upstream, but oriented towards the opposite direction)

These primers had to be phosphorylated to allow the subsequent self-ligation of the amplicon.

PCR 25 µL reaction was as follows: 13 µL H_2O, 5 µL 5X Phusion HF buffer, 0.5 µL dNTPs (10 mM each), 1.25 µL of each primer (10 µM), 0.75 µL DMSO, 0.5 µL plasmid DNA (10 ng/ µL) and 0.25 µL Phusion DNA Polymerase (2.5 U/ µL, New England Biolabs). The PCR program consisted on: 98°C for 30 sec, followed by 25 cycles of 98°C for 10 sec, 60°C for 30 sec and 72°C for 90 sec, followed by a final cycle of 72°C for 5 min.

3 µL PCR product was used for self-ligation in a mixture with 14 µL H2O, 2 µL 10X T4 buffer and 1 µL T4 DNA ligase (New England Biolabs). After incubation at RT for 4 hours, 1.5 µL of ligation reaction were used to transform 20 µL *E. coli* DH5α following the heat sock protocol previously described.

Since all resulted colonies were white as expected, and the plasmids were all the same size, miniprep and sequencing of three colonies was carried out as specified

before. After sequence analysis it was found out that one of the colonies carried the CCC mutated 5' splicing site.

2.3.5. *Arabidopsis* transformation ("floral dip" method)

The *Arabidopsis thaliana* floral dip transformation method was carried out following the protocol by Andrew Bent[354] with small modifications.

Arabidopsis thaliana Col0 wild type seeds were sowed in soil, stratified in the dark at 4°C for 48-72h and then placed in a growth room at the conditions already stated.

Besides, the three level 2 constructs carried by *E. coli* DH5α bacteria were transferred into Agrobacterium tumefaciens GV3101 (pMP90) by electroporation using a MicroPulser Electroporation apparatus (Bio-Rad), following the instructions given by the manufacturer for the electroporation of *A. tumefaciens*.

Once *Arabidopsis* plants started showing many young unopened flower buds but before they produced mature flowers, around 5 weeks after planting, the floral dipping process was performed.

First, a single colony of each *Agrobacterium* strain carrying one level 2 vector of interest was inoculated in liquid LB media with the corresponding antibiotics to select for both the plasmid (Kan) and the bacteria strain (rifampicin and gentamicin), and grown at 28°C with shacking until reaching at least midlogarithmic or even approaching stationary phase. Then, the cells were pelleted by centrifugation and resuspend in an equal volume of 5% sucrose solution, and the suspension was subsequently diluted to an OD600=0.8.

Above-ground parts of *Arabidopsis* were then dipped in the *Agrobacterium* solution for 2 to 3 seconds, with gentle agitation. Several plants were dipped per construct. Dipped plants were placed under a dome for ~24 hours to maintain high humidity and then returned to their previous growing conditions. Once plants were dry, seeds were collected (T1 transformed seeds).

Arabidopsis thaliana T1 transformed seeds were surface sterilised through exposure for 6h to chlorine fumes produced by mixing 100 ml bleach (NaClO 6.15%) with 3 ml 12 M HCl into a desiccator jar. Sterile seeds were sowed into sterile petri plates with 50 mg/L kanamycin, 0.5X MS and 0.8% agar. Plates were stratified in the dark at 4°C for 48-72h and then placed in a growth room. After ~15 days, T1 seeds

which were growing healthy in media with Kan and showed dsRED fluorescence were transplanted to soil. We kept them growing under the same 22°C temperature and 16h light conditions until senescence. Once plants were dry, T2 seeds were collected independently from each plant, and each plant offspring was considered a transgenic line.

Arabidopsis thaliana T2 transformed and WT Col0 seeds were surface sterilised as described above and then sowed in petri plates with 0.5X MS, 0.8% agar and with or without 50 mg/ L Kan, respectively. Several transgenic lines were sowed per construct. Plates were stratified in the dark at 4°C for 48-72h and then placed in a growth room. Seedlings were collected after 15 days for RNA extraction.

2.3.6. Northern blot detection of miRNAtop14 in the three constructs (fig 4.4.A)

For Northern blot detection of miRNAtop14 levels in the different constructs, 20 healthy growing seedlings per transgenic line/ WT were pooled together for RNA extraction. Three transgenic lines per each one of the three constructs were analysed, together with the WT, making a total of ten samples. RNA extraction was performed as described before in section 2.1.2. 10µg of total RNA from each sample were analysed by sRNA Northern blot (section 2.1.5). The probes used were the full 21 nucleotides complementary to Sly-miRNAtop14, an oligonucleotide complementary to Osa-miRNA528 (5'-CTCCTCTGCATGCCCCTTCCA-3') and an oligonucleotide complementary to U6 RNA (5'-AGGGGCCATGCTAATCTTCTC-3'). The film was scanned using the Typhoon FLA 9500 scanner (GE Healthcare Life Sciences).

2.3.7. RT-PCR analysis of pri-miRNAtop14 in the three constructs (fig 4.4.B)

The same total RNA samples used for northern blot (2.3.6) were used for reverse-transcription and PCR analysis of pri-miRNAtop14.

Before reverse transcription, the RNA was DNase treated using the TURBO DNA-free Kit and following the manufacturers' instructions (Ambion). 4 µg of total RNA were DNase treated in a 10 µL reaction, after which DNAse was inactivated and RNA recovered as explained in the manufacturers' protocol.

Reverse transcription was then performed using SuperScript II reverse transcriptase following manufacturer's protocol (Invitrogen), with the only difference that the incubation at 42°C was carried out for 60 min. 1 µl of DNase treated total RNA (~ 0.6 µg) and 1µl of dT_{15} primer (487µg/ µL) were used for the reverse transcription reaction.

20 µL PCR reaction per transgenic line/ WT was subsequently carried out. The mixture consisted in 10.9 µL H_2O, 4 µL 5X Green GoTaq reaction buffer, 0.4 µL dNTPs (10 mM each), 1 µL of each primer (10 µM), 1 µL cDNA, 1.6 µL $MgCl_2$ and 0.2 µL GoTaq G2 DNA Polymerase (5 U/ µl, Promega). The PCR program was as follows: 95°C for 2 min, followed by 30 cycles of 95°C for 30 sec, 50°C for 30 sec and 72°C for 1 min, followed by a final cycle of 72°C for 5 min.

PCR products were run in a 1.5% Agarose 0.5X TBE gel and a picture was taken with Typhoon FLA 9500 scanner (GE Healthcare Life Sciences) after EtBr staining.

The sequences of the primers used for this PCR were:

Forward primer: 5'-TGGTGACTTTGATCTCAAAAGAGTGC-3'

Reverse primer: 5'-GAGAATTCTGGCTCCGTCGCTGT-3'

2.4. Material and methods chapter 5

2.4.1. RNA ligase-mediated rapid amplification of cDNA ends

The same total RNA samples used for *S. lycopersicum*, *N. benthamiana* and *P. axillaris* miRNAtop14 northern blot detection (2.2.2) and RT-PCR analysis of pri-miRNAtop14 (2.2.3) were used for RNA ligase-mediated rapid amplification of cDNA ends (RLM-RACE). This way, it was confirmed that the mature miRNAtop14 was present in the sample, and an absence of cleaved target could not be attributed, a priori, to lack of miRNA.

First, mRNA was isolated from total RNA using the Dynabeads mRNA purification kit, following the manufacturer's instructions (Life Technologies).

Subsequently, ligation to GeneRacer RNA Oligo (see table 2.1 for sequence) was performed. 0.5 µg mRNA were mixed with 0.5 µg GeneRacer RNA Oligo in a total volume of 14µl. After 5 min incubation at 65°C to relax RNA secondary structures and a quick chill on ice, 2 µl RNaseOut (40 U/ µl, Invitrogen), 2 µl 10X Ligase Buffer and 2 µl T4 RNA ligase (5 U/ µl, Ambion) were added. The mixture was then incubated at 37°C for 1 hour.

After that, mRNA was cleaned using the Dynabeads mRNA purification kit again, according to the manufacturer's instructions (Life Technologies). However, the 20 µl ligation reaction was first adjusted to a total of 100 µl volume with water before performing the cleaning. Afterwards, the RNA with the beads was resuspended in 12 µl of the Tris-HCl buffer provided by the kit, and 11 µl of clean ligated mRNA were recovered.

Subsequently, to the 11 µl suspension of ligated mRNA, 1 µl dNTPs (10 mM each) and 1 µl of 50 mM Generacer oligo dT (see table 2.1 for sequence) were added in the first step of the reverse transcription reaction. The rest of the reverse transcription was performed according to the SuperScript II reverse transcriptase manufacturer's protocol (Invitrogen), with minor changes. In brief, the ligated mRNA, dNTPs and Generacer oligo dT mixture was incubated at 65°C for 5 min and then 4 µl 5X First-Strand Buffer, 1 µl 0.1 M DTT and 1 µL RNaseOUT (40 units/ µL) were added. After a short 2 min incubation at 42°C, 1 µL (200 units) of SuperScript II RT (Invitrogen) was incorporated to the reaction, which was then incubated for another 60 min at 42°C before inactivating the reaction by heating at 70°C for 15 min.

Next, the cDNA was subjected to 7 cycles of PCR amplification. The full length of all ligated transcripts was amplified from the 5' ligated oligo to the tail included in the poly dT primer. The 50 µL PCR reaction included the whole 20 µL of first strand cDNA reaction plus 12.25 µL H_2O, 10 µL 5X Phusion GC buffer, 1.25 µL dNTPs (10 mM each), 2 µL of Generacer 5' primer (10µM), 2 µL of Generacer 3' primer (10µM) and 1 µL Phusion DNA Polymerase (2.5 U/ µL, New England Biolabs). The PCR program was as follows: 98°C for 1 min, followed by 7 cycles of 98°C for 30 sec, 68°C for 30 sec and 72°C for 5 min, followed by a final cycle of 72°C for 7 min.

Afterwards, the 7 cycle PCR products were cleaned using the Agencourt AMPure XP system (Beckman Coulter). 90 µl AMPure beads were added to the 50 µl PCR mixture in an Eppendorf tube. After 5 min incubation at RT, the tube was place in a magnetic stand. Once all beads had migrated to the wall, supernatant was discarded and two washes with 200 µl of 70% ethanol were performed. Without moving the tube from the stand, the pellet was dried for 5-7 min. After this, the tube was finally removed from the stand and the beads were resuspended in 16 µl H_2O. The bead suspension was then returned to the stand to separate beads from clean PCR product in solution, which was recovered.

For *Petunia axillaris* RLM-RACE results showed in figure 5.8, this last cleaning step was skipped. However, in other *P. axillaris* RLM-RACE reactions that showed the same negative outcome than the one shown in figure 5.8, this cleaning step had been performed (data not shown).

The 7 cycle PCR product was subsequently used to perform a first PCR and a consecutive nested PCR. These PCRs were specific to each transcript putatively cleaved by miRNAtop14 or to the control transcripts, known to be cleaved by a miRNA.

After PCR and electrophoresis, the bands of interest were cut from the gel, cloned and analysed following the pGEM-t easy cloning method described above (1.1.3)

2.4.2. *Solanum lycopersicum* RLM-RACE PCR (fig. 5.4)

Solanum lycopersicum first RLM-RACE PCR reaction contained 4 µL 5X Phusion GC buffer, 0.6 µL dNTPs (10 mM each), 0.6 µL DMSO, 0.4 µL cDNA, 1.8 µL of Generacer 5' primer (10 µM), 0.6 µL of gene-specific reverse primer (10 µM) and 0.2 µL Phusion DNA Polymerase (2.5 U/ µL, New England Biolabs) and water up to

20 µL. The first PCR program was a touchdown PCR as follows: 98°C for 30 sec, followed by 5 cycles of 98°C for 15 sec, 72°C for 60 sec, followed by another 5 cycles of 98°C for 15 sec, 70°C for 30 sec and 72°C for 30 sec, followed by 30 cycles of 98°C for 15 sec, 68°C for 30 sec and 72°C for 30 sec and a final 10 min at 72°C.

The PCR reaction to check for the presence of the non-cleaved transcripts was the same as this first PCR reaction, but instead of 1.8 µL of Generacer 5' primer (10 µM), 0.6 µL of gene specific forward primer (10 µM) were used.

After the PCR, 10 µL of the reaction were run in an agarose gel and the remaining 10 µL were cleaned by the Agencourt AMPure XP system (Beckman Coulter) as explained above. The clean PCR product was resuspended in 10 µL of water again and then used as template for nested PCR.

The nested RLM-RACE PCR reaction had a total volume of 15 µL in which 0.3 µL of the primary PCR product were added together with 3 µL 5X Phusion GC buffer, 0.45 µL dNTPs (10 mM each), 0.45 µL DMSO, 0.45 µL of Generacer Nested 5' primer (10 µM), 0.45 µL of gene-specific nested reverse primer (10 µM) and 0.15 µL Phusion DNA Polymerase (2.5 U/ µL, New England Biolabs). The PCR profile consisted in 98°C for 2 min, followed by 30 cycles of 98°C for 15 sec, 60-72°C for 30 sec and 72°C for 30 sec and a final 10 min at 72°C. The annealing temperatures specific for each primer were 65°C for target 2 and 6, 66°C for target 7, 67°C for target 3, 68°C for target 1, 8 and 9, 69°C for target 4 and 71°C for target 5 (*LPR*) and both controls (see table 5.1 for details of the targets). Primer sequences are compiled in table 2.1.

2.4.3. *Nicotiana benthamiana* RLM-RACE PCR (fig 5.6)

Nicotiana benthamiana RLM-RACE first PCR reaction and program were the same as for *Solanum lycopersicum*, but components were adjusted to perform a 50 µL reaction instead of a 20 µL reaction.

The PCR reaction to check for the presence of the non-cleaved transcripts was as well similar to the first PCR.

Nicotiana benthamiana nested PCR was exactly the same mixture and program as the one performed in tomato. It was performed using first PCR product either directly or cleaned by the Agencourt AMPure XP system (Beckman Coulter), obtaining the same results in both cases. The annealing temperatures, specific for

each primer, were 71°C for target 1, target 2 and two control transcripts and 60-64-68°C were used with similar results for *LPR1* and *LPR2* (see table 5.1 for details of the targets). Primer sequences are gathered in table 2.1.

2.4.4. *Petunia axillaris* RLM-RACE PCR (fig. 5.8)

Several different PCR mixtures and programs were used for *Petunia axillaris* RLM-RACE trying to optimize the conditions to detect any miRNAtop14 directed cleavage of *LPR* mRNA without success. The conditions used to produce figure 5.8 were the following ones.

The first PCR reaction was similar to the one carried out in *Solanum lycopersicum* and *Nicotiana benthamiana*, but with the components adjusted to make a total of 30 µL reaction. However, four slightly different reactions were performed changing the forward primer and its amount. This way, there were two reactions using the Generacer 5' primer (10 µM), one containing 2.7 µL of this primer and the other one containing 0.9 µL, and another two reactions using instead the Generacer Nested 5' primer (10 µM), again one containing 2.7 µL of this primer and the other one containing 0.9 µL. The program used consisted in 98°C for 2 min, followed by 35 cycles of 98°C for 15 sec and 72°C for 30 sec, followed by another 5 min at 72°C.

The PCR reaction to check for the presence of the non-cleaved transcripts was the same to the first PCR reaction.

Petunia axillaris nested PCR was also the identical mixture and program as the one performed in tomato and *Nicotiana*, but with the volumes adjusted to a 20 µL reaction. It was performed using first PCR product directly and the annealing temperatures of each primer were in all cases 62°C, and the primer sequences can be seen in table 2.1.

Primer	Sequence 5'- 3'
Generacer RNA oligo	CGACUGGAGCACGAGGACACUGACAUGGACUGAAGGAGUAGAAA
Generacer oligo dT	GCTGTCAACGATACGCTACGTAACGGCATGACAGTG(T)$_{24}$
Generacer 5' primer	CGACTGGAGCACGAGGACACTGA
Generacer 5' nested	GGACACTGACATGGACTGAAGGAGTA
Generacer 3' primer	GCTGTCAACGATACGCTACGTAACG
Sly T1 Rv	ATCCACAAGCGCAATCTCCACACAT
Sly T2 Rv	CGCAGTTAGCAGCAACAGGAGCAAG
Sly T3 Rv	GAATATTCAAGAAGCCCCGGCAACA
Sly T4 Rv	GGCCACAATATCCACGGCTTCCTTA
Sly T5 (LPR) Rv	ACTGTTGCTGTGTGCCGCGATGTAC
Sly T6 Rv	GCCCGCTTGACTGAGCTACCTGACT
Sly T8-9 Rv	GGACTGGAGGACACGGATGCTTCGA
Sly LA Rv	GCATTCAAGTAGAGCATATCCCTGTCAGGG
Sly SCW RV	GGCCCATTGCCCGCCATAACCGAT
Sly T1 Fw	ACACAAGTTCAGTTCCACCAGCAAGC
Sly T2 Fw	TGAGGGGTTGATGAATTTGGTCTCAAG
Sly T3 Fw	CGGGCTGCCGATAGTGCATTGTCTG
Sly T4 Fw	GTTTCCTTGGTGATGCAGGGCCAGG
Sly T5 (LPR) Fw	CCCCCACCCCACCCCTTTCTTCTTA
Sly T6 Fw	CTGATCCTGCAGCTCAGCTTGTGGC
Sly T8-9 Fw	CTCTTGCAAGCTGAACAGGCTGCCA
Sly T1 Rv nested	CTCCACACATCCAATGGCGA
Sly T2 Rv nested	GGAGCAAGTCCAGTAGTAAGTCC
Sly T3 Rv nested	GGCAACAATTTCGTGCGACT
Sly T4 Rv nested	ATTAGAAAGTACATCCCAAACCCCA
Sly T5 (LPR) Rv nested	TGGCATGTCTTGGAGTTCATCA
Sly T6 Rv nested	GAGATCAGTGCTGCCAGGAG
Sly T7 Rv nested	GCCGCAATTGACTGTACGAA
Sly T8-9 Rv nested	CCTTCGAAGCCCACTGGAAA
Sly C1 Rv	GCATTCAAGTAGAGCATATCCCTGTCAGGG
Sly C2 Rv	GGCCCATTGCCCGCCATAACCGAT
Nb T1 Rv	CGCCACCAATCGACGACCCT
Nb T2 Rv	GTGCAGTTGAAAGGCACTCAGCT
Nb LPR 1&2 Rv	TGGTACATGCCAATCTTGAGTGAC
Nb TCP4 Rv	TTCTGCATTACGTCGGTCCACTC
Nb SCW-6 Rv	TGCCTGCCATAGCCGATATCAAA
Nb T1 Fw	ACGAGACTGTGATTTCGCCGA
Nb T2 Fw	TTGGTGATGCAGGGCCAGGATAT
Nb LPR1 Fw	AGAGTGTTGGTGACTTTGTTCCT
Nb LPR2 Fw	AAGGGTATTGGTGACTTTGGTCC
Nb TCP4 Fw	GAATGGGAATGTTGCCAGTTCAA
Nb Scarecrow-6 Fw	CTCCAGCTGCTTCCCCATTTT
Nb T1 Rv nested	GGTTCCATTCTTCCCATCCCTCC
Nb T2 Rv nested	TCAGTCTCCAAGCTCTTGTCGCA
Nb LPR1 Rv nested	GTTTATCTTCTGCCATTGAAGCCAT
Nb LPR2 Rv nested	TGGGCATGTCTGGAAGTTTGT
Nb C1 Rv nested	CGTGGGTCAAAGAGCAGAAAATG
Nb C2 Rv nested	TGGCAAATTGTGCAACTGGTG
Pa LPR Rv	GGCCTCGATTGTAGGTCCAGGGACTGT
Pa TCP4 Rv	TGCAGAAGGGAAGTTGCATTGGC
Pa SCW Rv	AGCCCATTGCCCGCCATAACCGA
Pa LPR Fw	AGGGTGTTTGTGACTTTGTCCTGCA
Pa TCP4 Fw	ACTCAGAAAGCAAAAGCCAAGCCCA
Pa SCW Fw	CAGAGCTGGTCCAGACGGGGAAT
Pa LPR Rv nested	CGAGTGACTTAGGGACATGAGCACCA
Pa C1 Rv nested	CCTGCCCCTCATCTGCACCTTCA
Pa C2 Rv nested	ACTGTGCAACTGGTGAGATCTCAGAGA

Table 2.1. Oligo and primers sequences used for RLM-RACE analyses.

Chapter 3

miRNA “top14” characterization

3.1. Introduction

3.1.1. Identification and characterization 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[355]. Development of miRNA prediction bioinformatics algorithms was also crucial to allow the recognition of candidate miRNAs among the huge amount of sRNA reads in a library. 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[356–358].

In addition to de-novo discovery of miRNAs, the bioinformatics tools have to recognise the miRNAs that are homologous to already discovered ones either in the same or in other species. This has been traditionally achieved through the search of different nucleotide databases with previously identified miRNAs[188,359,360].

As a result of all these efforts to identify miRNAs, currently there are thousands of miRNAs annotated in over hundred different species[361]. However, concerns about the authenticity of some of these miRNAs have been raised[362]. In particular, it remains difficult to differentiate between miRNAs and endogenous siRNAs, given that the main difference between them is their biogenesis method[52,214]. 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[213].

An additional proof of miRNA authenticity and a first step in studying its function is the identification of targets[213]. Fortunately, computational miRNA target prediction in plants can be done relatively easily thanks to their high complementarity to the target site[142,143]. Furthermore, given that their mode of action is frequently cleavage[215], experimental validation of a target can be achieved directly by detecting the cleaved transcript through techniques such as 5' RACE[150] or its high throughput version, parallel analysis of RNA ends (PARE) libraries, also called degradome libraries[148,363].

Besides the chance of false positives, there is also a risk of false negatives in the bioinformatic search for miRNAs[213]. As an example, it has been claimed that non-conserved miRNAs are often not identified with some bioinformatics tools that rely on phylogenetic conservation[356,364]. In addition to this, different NGS technologies are

known to be biased towards or against certain sequences, artificially increasing or decreasing the abundance of some miRNAs, or even not detecting them at all[365]. These technical constraints may be delaying the discovery of certain miRNAs, especially if their features diverge from the most canonical ones.

3.1.2. Intron-split miRNAs

Plant *MIRs* usually contain introns, with a 67% of *Arabidopsis* pri-miRNAs estimated to have at least one intron[193]. 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[58,192,193]. miRNA hairpins are located in the first exon in most cases[193,194], although they may appear in other exons as well as in alternatively spliced regions that may be either intronic or exonic[56,59].

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[195,366]. This exon-intron 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[366]. In a later study, also in rice, it was found that most members of the miRNA444 family had a characteristic 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)[195]. Interestingly, all miRNA444 family members were found to have an intron in between their miRNA and miRNA*, even the two of them (miRNA444e and miRNA444f) which were not transcribed from the antisense strand of their target gene[195].

This peculiar exon-intron arrangement was subsequently reported also in miRNA444 members of maize and sorghum[58,367,368]. 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[368]. In a recent review paper they have been called intron-split miRNAs[369]. However, to our knowledge, up to date no intron-split miRNAs have been found outside the miRNA444 family.

3.1.3. Study of miRNAtop14

In 2008, Moxon et al.[370] 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 RNAfold. 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 were more distant from one another than what was usually observed in other miRNA/ miRNA* pairs. In fact, they were almost 700 nt apart although 98% of plant miRNA hairpins have a length of less than 336 nucleotides[371]. 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 came out; one of the expected genomic size between the primers but 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. With this information, I proceeded to further characterise this miRNA.

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.[372] 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.

3.2. Results

3.2.1. Identification of miRNAtop14 in Solanales

Both mature miRNAtop14 and genomic locus sequences have been identified in *Solanum lycopersicum*. With this information, a BLAST search was performed in different databases in order to determine whether this miRNA was also present in other related species.

First, *Solanum* species were examined, and after the miRNA had been detected in several species within the *Solanum* genus, the whole Solanaceae family was included in the search. Newly identified *MIRtop14* sequences were in turn used for BLAST searches against the next related species according to the Solanaceae phylogeny, which facilitated the identification of *MIRtop14* in up to four genera of the Solanaceae family, the four genera with higher amount of genomic sequences available (*Solanum*, *Capsicum*, *Nicotiana* and *Petunia*). We subsequently expanded the search to look for the miRNA within the whole Solanales order. Apart from Solanaceae, the only two species with sequenced genomes belonged to the Convolvulaceae family (*Ipomoea trifida* and *Ipomoea nil*), and miRNAtop14 was identified in both of them as well as in sweet potato (*Ipomoea batatas*). Finally, we tried to search for *MIRtop14* within the three closer orders to Solanales: Gentianales, Lamiales and Boraginales[373]. However, we couldn't identify the miRNA in either of these orders, despite of the fact that *Coffea canephora* and *Mimulus guttatus*, which belong to the Gentianales and Lamiales order, respectively, both have high quality draft genomes available[374,375].

All species in which miRNAtop14 has been identified and their phylogenetic relationships are shown in figure 3.1, while in table 3.1 the mature miRNAtop14 and miRNAtop14* sequence in each species are compiled.

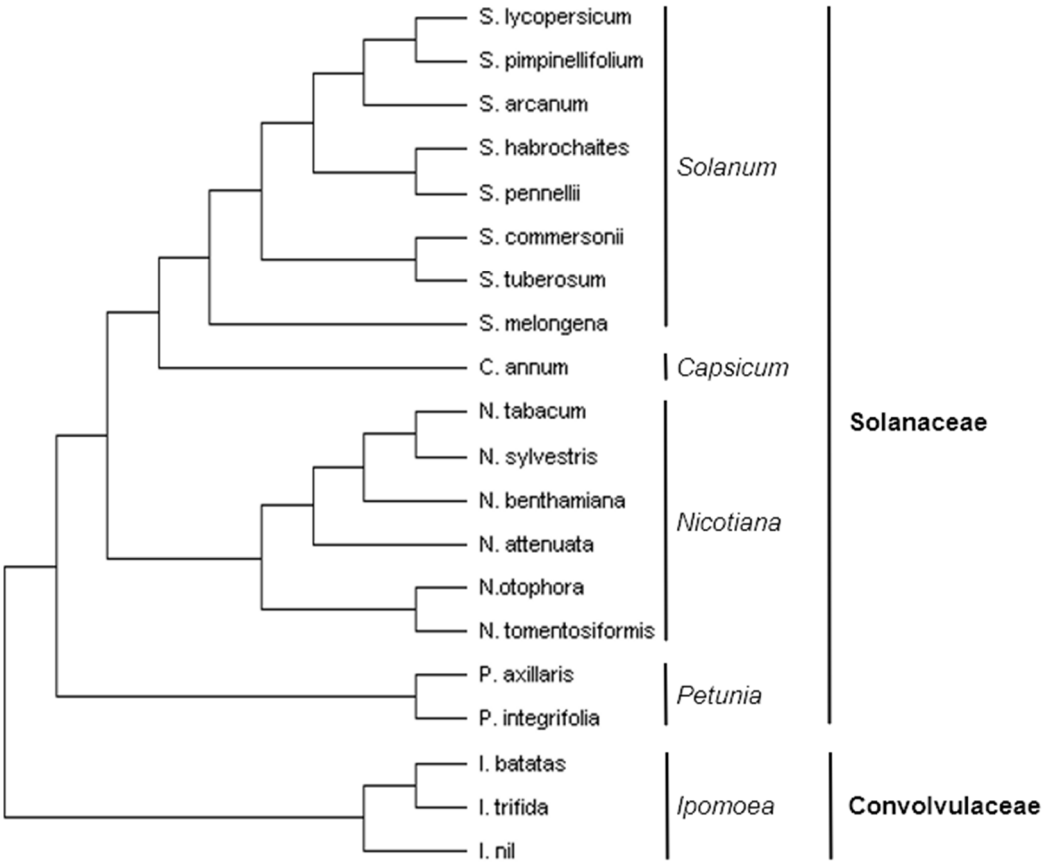


Figure 3.1. Cladogram depicting the evolutionary relationship among all species in which *MIRtop14* has been identified. All of them belong to the Solanales order. Data extracted from several studies of the phylogeny of these species[440–447].

Species	miRNA sequence (3p)	miRNA* sequence (5p)
Solanum lycopersicum	CUUGGGACCAAAGUCACCAAC	UGGUGACUUUGAUCUCAAAAG
Solanum pimpinellifolium	CUUGGGACCAAAGUCACCAAC	UGGUGACUUUGAUCUCAAAAG
Solanum arcanum	CUUGGGACCAAAGUCACCAAC	UGGUGAUUUUGGUCUCAAAAG
Solanum habrochaites	CUUGGGACCAAAGUCACCAAC	UGGUGUCUUUGGUCUCAAAAG
Solanum pennellii	CUUGGGACCAAAGUCACCAAC	UGGUGACUUUGGUCUCAAAAG
Solanum commersonii	CUUGGGACCAAAGUCACCAAC	UGGUGACUUUGGUCUCAAAAG
Solanum tuberosum	CUUGGGACCAAAGUCACCAAC	UGGUGACUUUGGUCUCAAAAG
Solanum melongena	UUUGGGACCAAAGUCACCAAC	UGGUGCCUUUGGUCUCUAAAG
Capsicum annuum	CUUGGGACCAAAGUCACCAAC	UGGUGAUUUUGGUCUCAAAAG
Nicotiana tabacum	UUUGGGACCAAAGUCACCAAC	UGGUGACUUUGGUCUCGAAAG
Nicotiana sylvestris	UUUGGGACCAAAGUCACCAAC	UGGUGACUUUGGUCUCGAAAG
Nicotiana benthamiana	UUUGGGACCAAAGUCACCAAC	UGGUGACUUUGGUCCCGAAAG
Nicotiana attenuata	UUUGGGACCAAAGUCACCAAC	UGGUGACUUAUGGUCUCGAAAG
Nicotiana otophora	UUUGGAACCAAAGUCACCAAC	UGGUGACUUUGGUCUCGAAAG
Nicotiana tomentosiformis	UUUGGAACCAAAGUCACCAAC	UGGUGACUUUGGUCUCAAAAG
Petunia axillaris	UUUGGGACCAAAGUCACCAAC	UGGUGACUUUGGUCUCGAAAG
Petunia integrifolia	UUUGGGACCAAAGUCACCAAC	UGGUGACUUUGGUCUCGAAAG
Ipomoea batatas	UUUGGGACCAAAGUCACCAAC	UGGUGACUUUGUACCCAAAGG
Ipomoea trifida	UUUGGGACCAAAGUCACCAAC	UGGUGACUUUGUACCCAAAAC
Ipomoea nil	UUUGGGACCAAAGUCACCAAC	UGGUGACUUUGUAUCCAAAAC

Table 3.1. miRNAtop14 and miRNAtop14* sequences in all the species in which the miRNA has been identified. miRNAtop14 is located on the 3' arm (3p) and miRNAtop14* is on the 5' arm (5p) of the miRNA hairpin. In red, nucleotides mismatching with *Solanum lycopersicum* sequence. National Center for Biotechnology Information (NCBI), SGN and Sweetpotato GARDEN databases were used to obtain these sequences[422,448,449].

3.2.2. *MIRtop14* genomic sequence and predicted pri-miRNA

MIRtop14 genomic sequences could be gathered in all species in which the mature miRNA sequence was identified. Presence of the putative pri-miRNAtop14 sequence in the genome was verified by PCR in four Solanales species, each one belonging to a different genus: *Solanum lycopersicum, Nicotiana benthamiana, Petunia axillaris* and *Ipomoea nil*. *Capsicum* was the only genus known to harbour miRNAtop14 which was not analysed. Subsequent cloning and sequencing confirmed the identity of these sequences, with the exception of *I. nil*, in which sequencing was not performed. The products of these amplifications can be seen in figure 3.2.

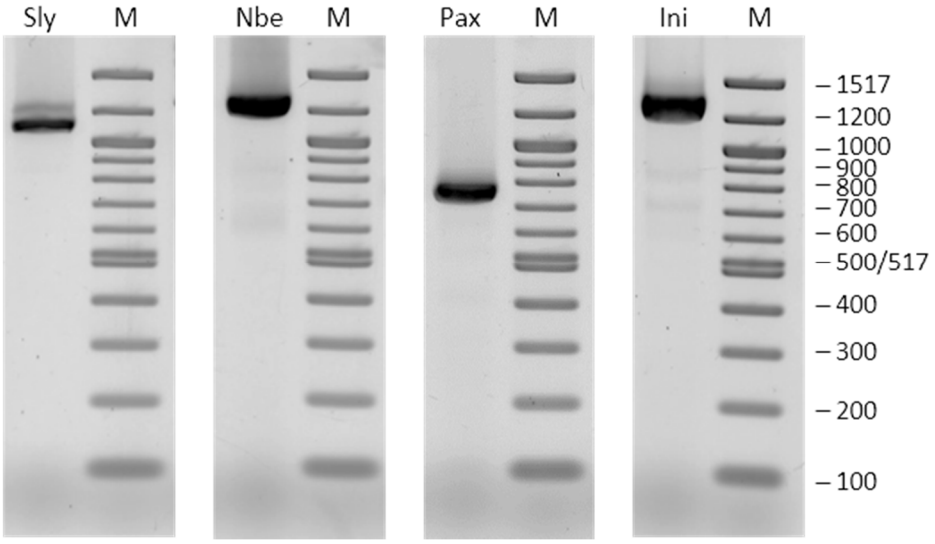


Figure 3.2. PCR from genomic DNA amplifying putative pri-miRNAtop14 sequence in the species *Solanum lycopersicum* (Sly), *Nicotiana benthamiana* (Nbe), *Petunia axillaris* (Pax) and *Ipomoea nil* (Ini). Expected product length from each species in nucleotides: Sly=1071, Nbe=1237, Pax=760, Ini=1247. M, 100bp marker. *MIRtop14* sequences and primers can be seen in appendix.

A ncRNA is predicted to be transcribed from this genomic locus in *Solanum lycopersicum*, several *Nicotiana* species and *Ipomoea nil* according to the NCBI database based on expressed sequence tags (EST) evidence. Besides, if looking directly at the EST or transcriptome shotgun assembly (TSA) collections also at NCBI, two more *Solanum* members as well as *Petunia axillaris, Petunia integrifolia* and *Ipomoea batatas* are predicted to transcribe this locus (see table 3.2).

Interestingly, in *Solanum* and *Nicotiana* species an intron is predicted to be excised from the region between miRNA and miRNA*, while in *Petunia axillaris* there is no intron predicted (see figure 3.3). In fact, when checking the genomic sequence

of *Petunia axillaris* and *Petunia inflata*, both have a very short stretch of DNA between miRNA and miRNA* compared with the long distance among them in the rest of the Solanaceae species studied (see table 3.2). However, even more surprisingly, in *Ipomoea nil*, one of the members of the Convolvulaceae family and therefore a species that has diverged earlier from all the rest of the Solanaceae family, the predicted ncRNA harbours an intron again, but this time the intron includes the miRNAtop14 sequence while miRNAtop14* lies within the upstream exon (see figure 3.3).

Besides, it may be interesting to note that among the EST reads curated at NCBI, there is evidence for transcripts both with and without intron in some species such as *Solanum lycopersicum* and *Nicotiana tabacum*, which would be an indicative of alternative splicing rather than constitutive splicing.

Finally, it is worth mentioning that in *Solanum lycopersicum, Solanum tuberosum* and *Solanum pimpinellifolium* a protein is predicted to be encoded from a 160-170 nucleotide sequence which expands across the mature miRNAtop14 site, according to ENSEMBL and SGN databases. In SGN, the protein predicted for *Solanum lycopersicum* is cycling DOF factor 2. Although such predictions cannot be considered very reliable because they are only based on in silico data, they cannot be discarded either.

Species	A) Distance miRNA-miRNA*	B) *MIR* transcript predicted?
Solanum lycopersicum	652 nt	yes, ncRNA
Solanum pimpinellifolium	660 nt	no
Solanum arcanum	619 nt	no
Solanum habrochaites	640 nt	no
Solanum pennellii	660 nt	no
Solanum commersonii	669 nt	No
Solanum tuberosum	668 nt	yes, EST
Solanum melongena	872 nt	yes, EST
Capsicum annuum	835 nt	No
Nicotiana tabacum	932 nt	yes, ncRNA
Nicotiana sylvestris	933 nt	yes, ncRNA
Nicotiana benthamiana	899 nt	no
Nicotiana attenuata	884 nt	yes, ncRNA
Nicotiana otophora	1159 nt	yes, ncRNA
Nicotiana tomentosiformis	1179 nt	no
Petunia axillaris	48 nt	yes, TSA
Petunia integrifolia	62 nt	yes, TSA
Ipomoea batatas	491 nt	yes, TSA
Ipomoea trifida	431 nt	no
Ipomoea nil	441 nt	yes, ncRNA

Table 3.2. miRNAtop14-miRNAtop14* distance and *MIRtop14* transcript prediction. **A)** Distance in nucleotides between miRNAtop14 and miRNAtop14* in the genome of the species predicted to harbour *MIRtop14* (miRNA and miRNA* sequences not included). **B)** Species in which *MIRtop14* has been predicted to be transcribed according to NCBI database. ncRNA=non-coding RNA predicted, nucleotide collection. EST=transcript only detected in the EST collection. TSA=transcript only detected in the TSA collection

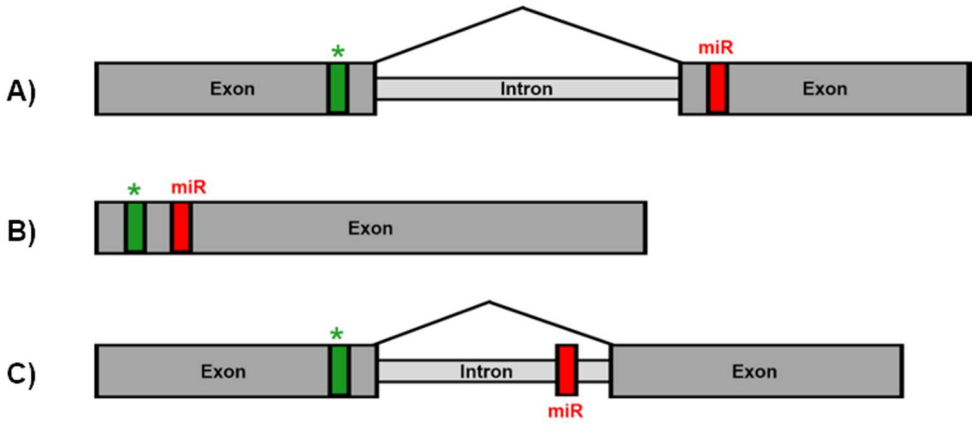


Figure 3.3. Exon-intron structure predicted for the different genera found harbouring miRNAtop14. **A)** Structure predicted for *Solanum*, *Capsicum* and *Nicotiana*. **B)** Structure predicted for *Petunia*. **C)** Structure predicted for *Ipomoea*.

3.2.3. miRNAtop14 primary transcript secondary structures

The same putative pri-miRNAtop14 sequences that were amplified from genomic DNA in four species of four different genus (*S. lycopersicum, N. benthamiana, P. axillaris and I nil*, see figure 3.2), were used for prediction of pri-miRNA secondary structures.

Once again, *Capsicum annum* was left out of the analysis because its pri-miRNAtop14 3' and 5' ends and intron boundaries were completely unknown. However, based on its genomic sequence, which shows an 840 nucleotide stretch between miRNA and miRNA*, and its phylogenetic position, which lays between *Solanum* and *Nicotiana* genera, *Capsicum annum* is expected to have roughly the same kind of structure as *Solanum* and *Nicotiana* do, with the intron in between the two miRNA stem arms, as seen in figures 3.4.A and 3.4.B.

Solanum lycopersicum and *Ipomoea nil* sequences chosen for the analysis of secondary structure were the ones corresponding to the NCBI predicted ncRNAs.

Nicotiana benthamiana pri-miRNA 5' and 3' ends were unknown, as were the splice sites of its intron. However, by alignment to *Nicotiana tabacum* NCBI predicted ncRNA it was possible to infer the intron beginning and end, which were later confirmed by RT-PCR (see next section). The alignment had a 91% identity from the very first nucleotide predicted to be transcribed in *Nicotiana tabacum* until position 1273 of the transcript, while from nucleotide 1273 to the 3' end of the transcript at position 1477 no homology was observed. Therefore, for secondary structure analysis the *Nicotiana benthamiana* sequence homologous to the first 1273

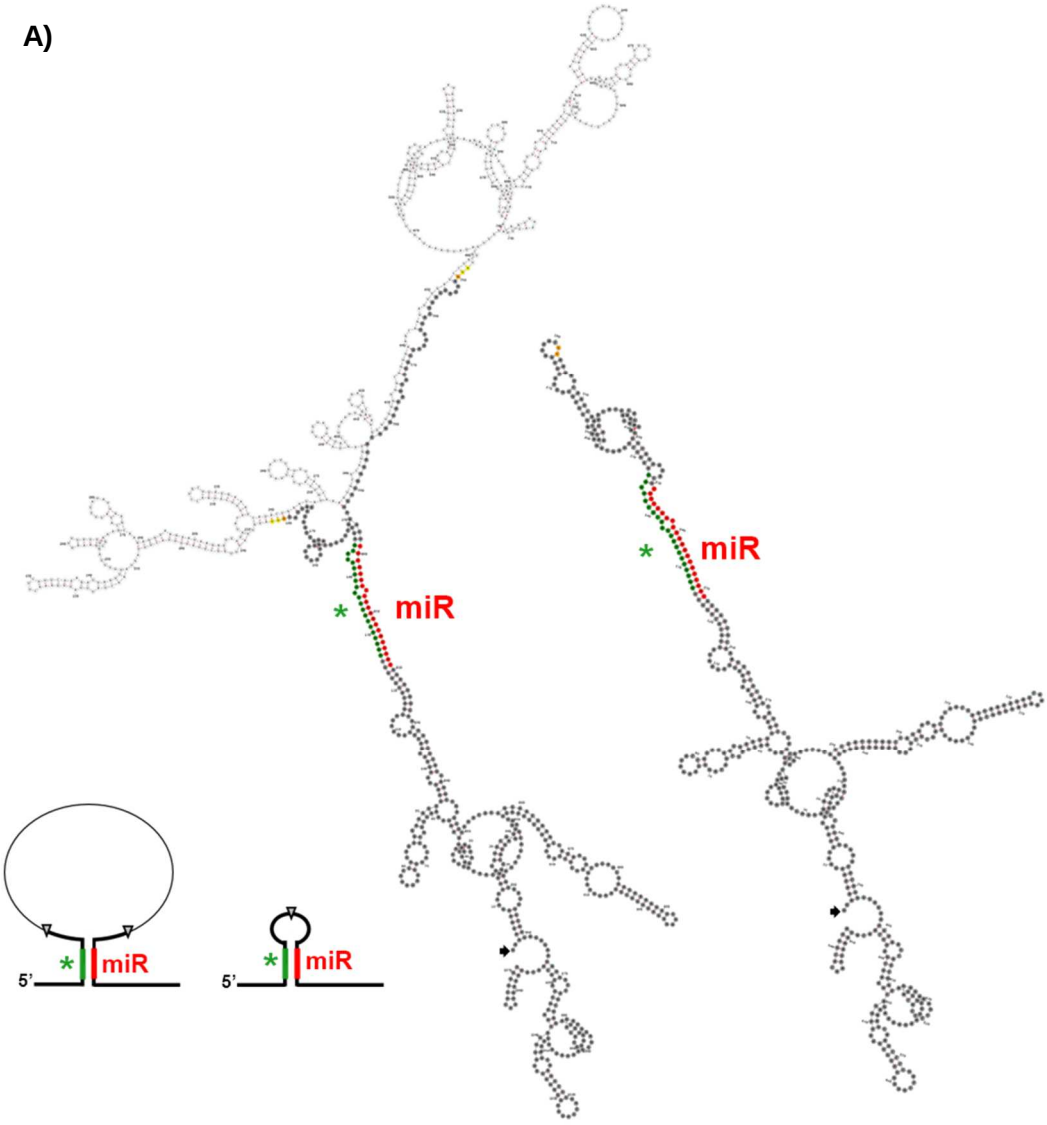
nucleotides of *N. tabacum* ncRNA was selected, in which both miRNAtop14 and miRNAtop14* sequences were included.

Finally, the *Petunia axillaris* pri-miRNAtop14 sequence chosen for secondary structure analysis was based on the transcript from this locus according to NCBI "transcriptome shotgun assembly". When aligning the NCBI predicted transcript to the *Petunia axillaris* genomic locus, which sequence had been validated by PCR and sequencing (see figure 3.2.), there were 26 mismatching nucleotides at the 5' end of the transcript sequence. We thus wondered whether the NCBI prediction of the *Petunia axilaris* transcript 5' end was correct, so decided to compare its genomic locus with *Solanum lycopersicum* and *Nicotiana benthamiana* putative transcripts. Both alignments expanded 49 nucleotides at the 5' end from the initially predicted transcript start site (once removed the 26 mismatching nucleotides). We therefore decided to include these 49 nucleotides as part of the *Petunia axilaris* pri-miRNAtop14 and used it to predict its secondary structure (the results of the RT-PCR analysis, shown in the next section 3.2.4., confirmed the presence of this 49 nucleotides at the beginning of the transcript).

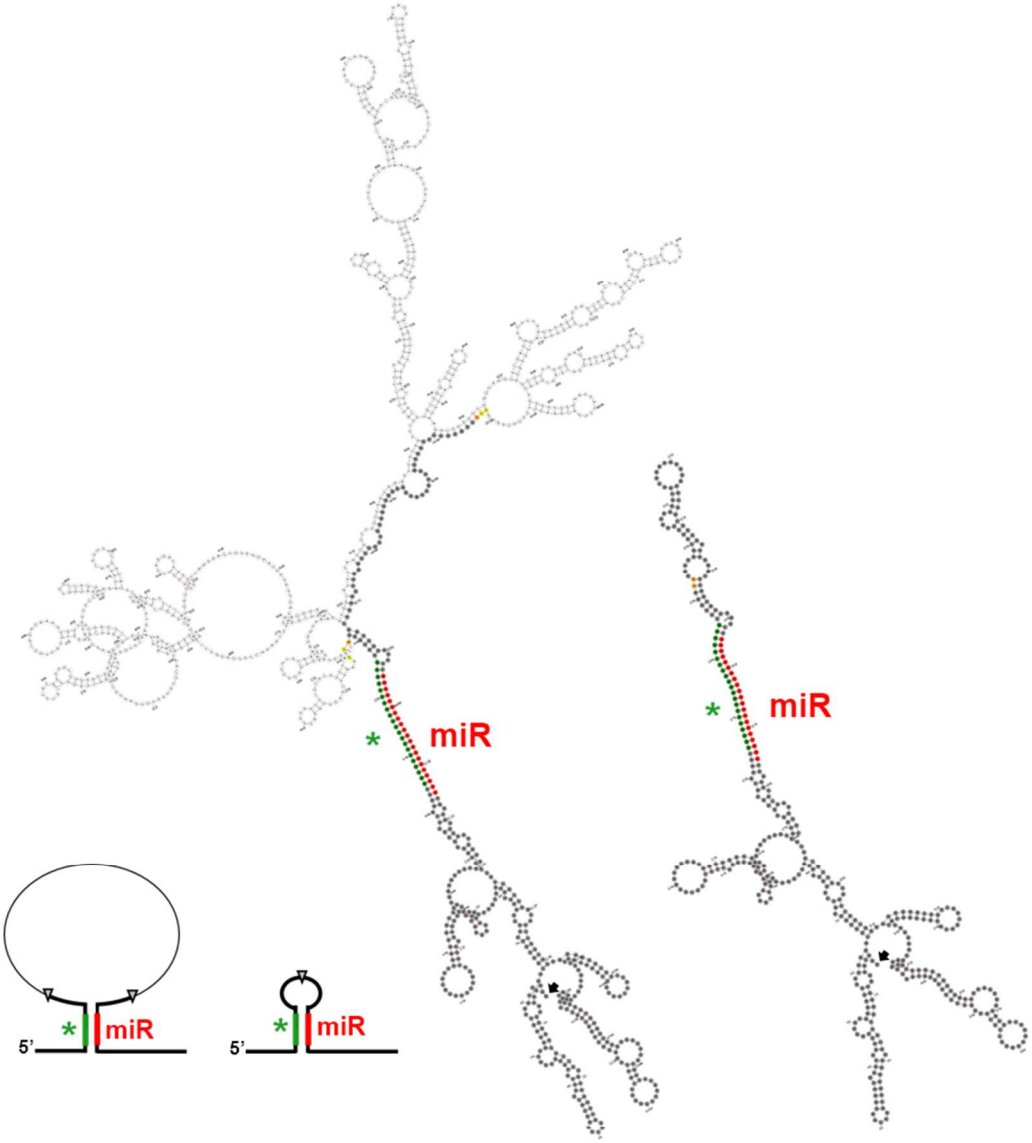
Optimal secondary structure of the four sequences described above was calculated as the one with minimum free energy using RNAfold[376]. For the species harbouring an intron, which are all of them apart from *Petunia axillaris*, both versions of the sequence, with and without intron, were analysed.

The predicted secondary structures are shown in figure 3.4. As can be observed, in all of them the miRNA and miRNA* pair together creating a hairpin, which is in accordance with the structure necessary for a pri-miRNA to be processed into a mature miRNA[51,52]. Interestingly, in both *Solanum lycopersicum* and *Nicotiana benthamiana* a hairpin with the same stem arms is created independently of whether the pri-miRNAs are folded with the intron included or excluded; the only difference is the length of the upper loop. Of course, in *Ipomea nil* spliced primary transcript the miRNA-like stem-loop is not created, since in this species the miRNA sequence lays within the intron. However, a hairpin with miRNAtop14/miRNAtop14* duplex is formed when the intron-containing transcript is folded.

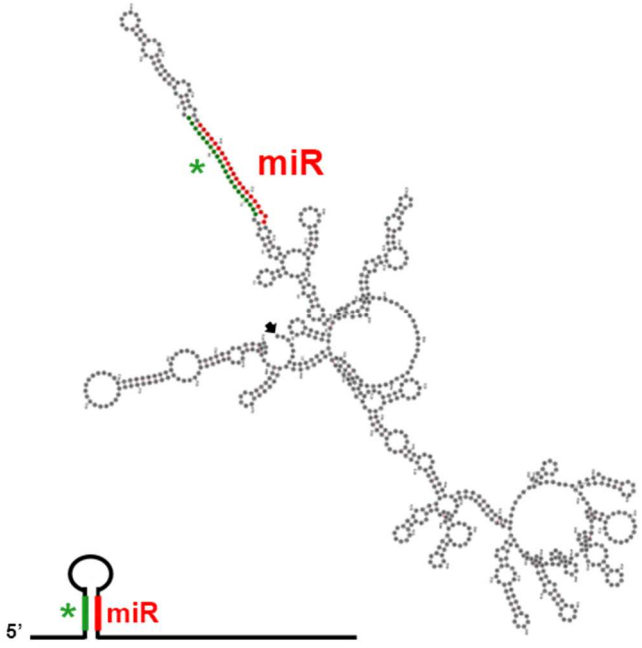
A)



B)



C)



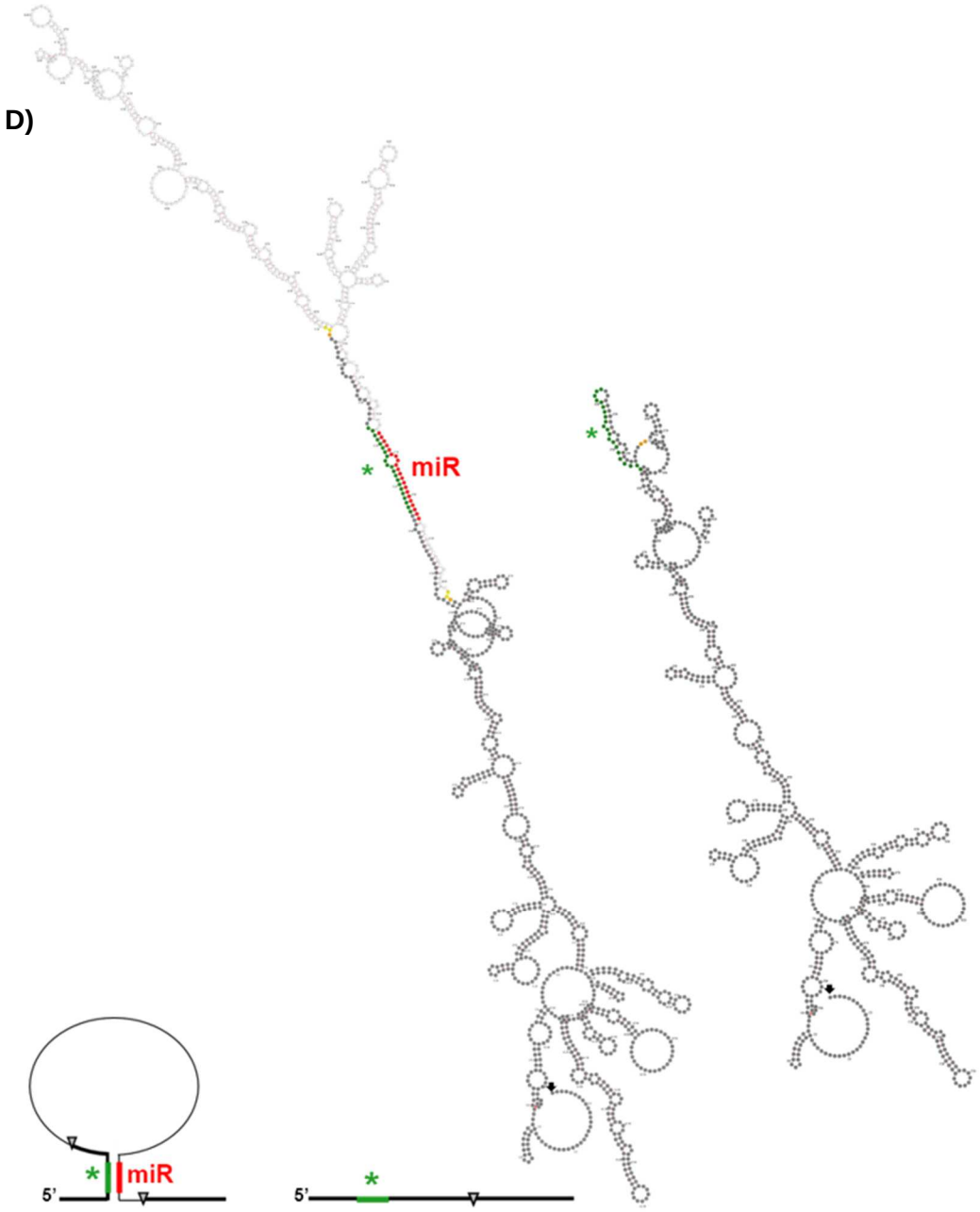


Figure 3.4. pri-miRNAtop14 secondary structure and schematic representation of the resulting miRNA hairpin, spliced (right) and non-spliced (left) variants. Species studied: **A)** *Solanum lycopersicum* **B)** *Nicotiana benthamiana* **C)** *Petunia axillaris* **D)** *Ipomoea nil*. pri-miRNAtop14 secondary structure was predicted by RNAfold[376] as the one with minimum free energy from the putative pri-miRNA sequence and visualizations were created using *forna* tool. miRNAtop14 and miRNAtop14* are indicated in red and green, respectively, in both the schemes and the secondary structure representation. In the schemes, exons are represented as bold lines, introns as thin lines and SS and exon-exon junctions as grey triangles. In the secondary structure representations, nucleotides belonging to an exon are dark grey while the ones within the intron are light grey. GU-AC dinucleotides at the SS are coloured yellow. First exon 3' end nucleotide and last exon 5' end nucleotide are coloured orange to mark the exon-exon junction after splicing. The first 5' residue of each transcript is indicated with a black arrow and the structures are always oriented with the miRNA stem-loop at the top.

3.2.4. miRNAtop14 mature miRNA and primary transcript expression detection

The four species previously studied; *S. lycopersicum, N. benthamiana, P. axillaris* and *I. nil*, were subsequently analysed by sRNA Northern blot to confirm mature miRNAtop14 expression in vivo. Along with them, a sample from *A. thaliana* was also tested as negative control. The samples consisted of total RNA extracted from the aerial part of three plantlets per species, all around one month old. The same probe was used for the detection of mature miRNAtop14 in the five samples: the complementary of the mature miRNAtop14 sequence without the first 5' nucleotide, which is the only nucleotide that varies between these species (see table 3.1). The results are shown in figure 3.5.A., where it can be appreciated that miRNAtop14 is expressed in all four Solanales species examined but not in *Arabidopsis thaliana*.

Total RNA from *S. lycopersicum* and *P. axillaris* and total mRNA from *N. benthamiana* was reverse transcribed and tested for pri-miRNAtop14 expression by PCR. The products of the amplification were run in an agarose gel and can be seen in figure 3.5.B.

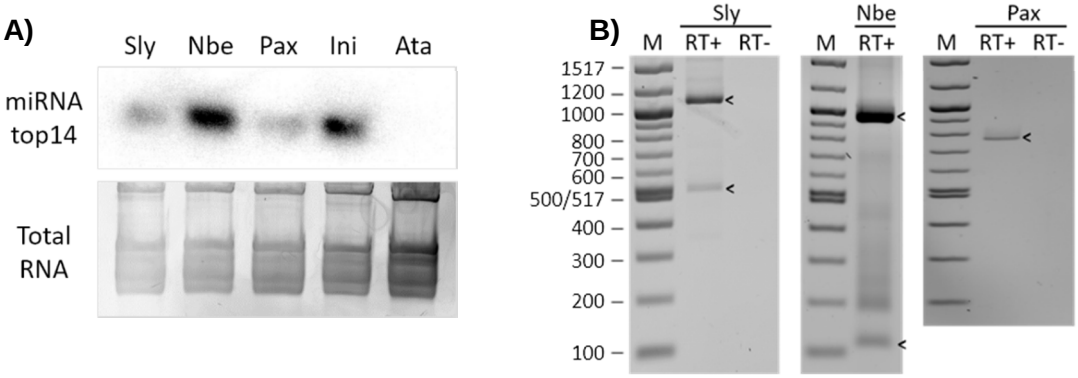


Figure 3.5. Detection of miRNAtop14 and pri-miRNAtop14 in different plant species. **A)** Detection of mature miRNAtop14 in *Arabidopsis thaliana* (Ata) and four species of Solanales: *Solanum lycopersicum* (Sly), *Nicotiana benthamiana* (Nbe), *Petunia axillaris* (Pax) and *Ipomoea nil* (Ini) by Northern blot. Ethidium bromide stained total RNA is included as loading control. **B)** Detection of pri-miRNAtop14 after total RNA reverse transcription in Sly and Pax, and mRNA reverse transcription in Nbe (RT+) by PCR. Controls without reverse transcriptase enzyme were included to rule out genomic DNA contamination where indicated (RT-). Bands showing the expected length are marked by an arrowhead. Expected amplicons length from each species in nucleotides: Sly=1071 & 503, Nbe=941 & 111, Pax=760. M, 100bp marker. *MIRtop14* sequences and primers can be seen in appendix.

Both spliced and non-spliced *S. lycopersicum* and *N. benthamiana MIRtop14* transcripts, as well as a single transcript in *P. axillaris*, were amplified and sequenced. Although the full length of these transcripts cannot be determined in the absence of RACE analysis, the expression of a transcript which spans at least from miRNAtop14* to miRNAtop14 sequence, encompassing both, was confirmed in all three species, as well as the position of an intron in between miRNAtop14 and miRNAtop14* in *S. lycopersicum* and *N. benthamiana*.

Finally, in *Solanum lycopersicum*, different levels of miRNAtop14 have been detected between different parts of the plant (leaf, root and stem), between different leaves and even between different leaflets of the same leaf (see figure 3.6.). Furthermore, when trying to replicate these experiments it has been found that these levels are highly fluctuating (see figure 3.6.). 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.

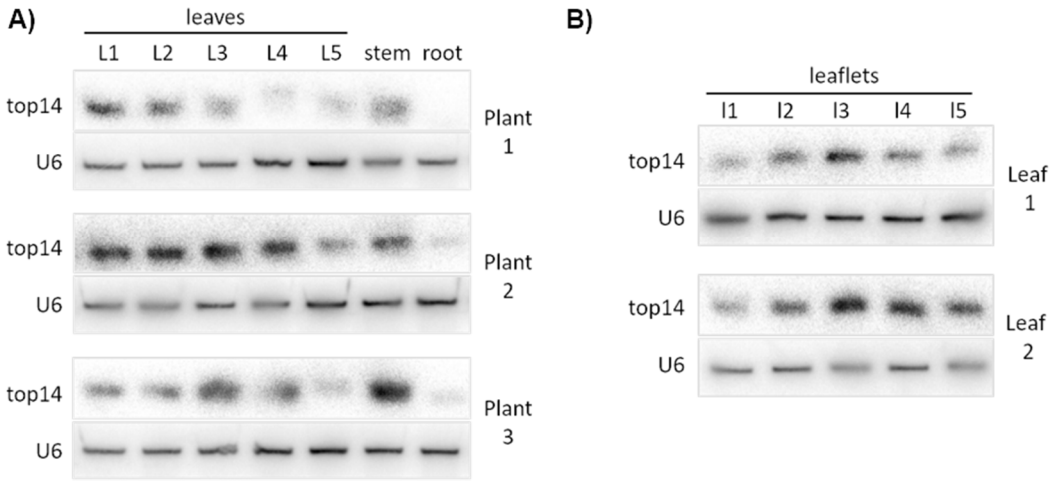


Figure 3.6. Detection of mature miRNAtop14 levels in *Solanum lycopersicum* root, stem, leaves and leaflets. **A)** Detection of mature miRNAtop14 levels in *Solanum lycopersicum* root, stem and leaves by Northern blot. Different leaves from the same plant were named L1 to L5, from the oldest to the youngest. **B)** Detection of mature miRNAtop14 levels in *Solanum lycopersicum* leaflets by Northern blot. Different leaflets from the same leaf were name l1 to l5, from the bottom left to the bottom right of the leaf, in clockwise order. U6 detection is included as loading control.

3.3. Discussion

3.3.1. Discovery of miRNAtop14

miRNAtop14 has been identified in a couple of independent studies, the first one in *S. lycopersicum* and a second one in *N. benthamiana*[370,372]. However, the connection between the two orthologous miRNAs was not noted, neither was published the existence of an intron in the middle of both miRNA hairpins. In this study, these findings have been connected and the phylogenetic conservation of miRNA top14 within the Solanales order has been determined.

It is nevertheless surprising that a miRNA present in a plant family as studied as Solanaceae, and which has been so easily detected by Northern blot in all four species tested (see figure 3.5), had not been identified earlier. In fact, although levels of mature miRNAtop14 were highly variable, miRNAtop14 presence was detected in all Northern blot experiments performed (data not shown), which indicates that constitutive level of miRNAtop14 is high enough to be routinely detected.

Two main reasons can probably explain why miRNAtop14 has not been characterised earlier. First, the sequence bias inherent to small RNA NGS platforms[365]. Indeed, when miRNAtop14 was first identified in the Dalmay lab the sRNA libraries were sequenced by 454 Life Sciences using pyrosequencing technology[370]. Subsequent tomato sRNA libraries prepared in the same laboratory, but sequenced by Genome Analyser II (Illumina) yielded very small number of reads from this miRNA[377]. In agreement with this, Baksa et al.[372] sRNA libraries were sequenced by HiSeq 2000 (Illumina), a third different sRNA NGS system that may show differences in miRNA detection compared with Genome Analyser II. The second reason is the presence of the intron in between miRNA and miRNA*, which makes the identification of this miRNA impossible by the existent bioinformatic tools when the reads are mapped to the genome, since the miRNA and miRNA* reads map too far from each other. Since there is no genome sequence available for *N. benthamiana,* Baksa et al.[372] used EST sequences to map sequence reads and this allowed them to identify miRNAtop14, as the intron sequence was absent from the ESTs analysed. This difficulty has already been observed for nat-miRNA, which also have these characteristic exon-intron structure, and therefore are not detected by usual miRNA prediction tools[368]. Therefore, future work involving high-throughput miRNA identification could benefit from including analysis tools that enable of the search for miRNAs with less conventional exon-intron structures.

3.3.2. *MIRtop14* phylogenetic study

MIRtop14 has been identified in two main Solanales families: Solanaceae and Convolvulaceae. However, this order includes three more smaller families: Montiniaceae, Hydroleaceae, and Sphenocleaceae[373]. These three families are considered to form a sister clade to the main clade formed by Solanaceae and Convolvulaceae, two large families which have a well stablished close relationship[373]. According to our analysis, *MIRtop14* is confined to the Solanales order, since we could not identify it in any of the orders closer to it: Gentianales, Lamiales and Boraginales[373]. It must therefore have emerged at some point at the beginning of Solanales divergence, either at the origin of Solanales (82-86 millon years ago[378]) or just after the separation between the two clades (Solanaceae-Convolvulaceae and Montiniaceae-Hydroleaceae-Sphenocleaceae). It is therefore a linage-specific, relatively young miRNA, and as such it appears in a single copy in each genome, unlike the highly conserved miRNAs that tend to form multicopy gene families[234].

From the phylogenetic analysis of *MIRtop14*, the high heterogeneity of its primary transcript between different genera stands out. While in *Solanum* and *Nicotiana* (and putatively in *Capsicum*), there is an intron in the middle of the miRNA stem loop, the closest studied genus to have diverged from these three genera, *Petunia*, doesn't have any intron in its whole pri-miRNAtop14 transcript. Surprisingly, *Ipomoea*, the first of all of these genera to have diverged from the rest, is predicted to have an intron again, although in a different position relative to the miRNA hairpin. The most probable explanation for this phylogenetic pattern is that an intron was already present in the common ancestor of all *MIRtop14* studied, and that it was probably lost in *Petunia* genus a posteriori. In fact, in the recent evolution of plants, intron losses have outnumbered intron gains[379]. Besides, changes in intron position between homologous genes, like the one seen here between *I. nil* and *N. benthamiana/ S. lycopersicum MIRtop14*, are a commonly observed phenomenon which has been explained by the so-called process of intron sliding[343].

Despite their heterogeneity, all the pri-miRNAs from the different species tested produced mature miRNAtop14 at highly detectable levels (see figure 3.5). Furthermore, the mature miRNA sequence is very conserved among all species harbouring this *MIR*, with only one nucleotide difference between *Capsicum* and most *Solanum* species when compared to *Nicotiana, Petunia* and *Ipomoea* species (see table 3.1). The only two exceptions to this are *N. otophora* and *N. tomentosiformis*, where there is a second nucleotide change between their miRNAtop14 and the one

in the rest of *Nicotiana* species. This low variation among miRNAtop14 sequences suggest that they are under purifying selection, an evolutionary mechanism that eliminates deleterious variations and which is the predominant force maintaining the sequence of those miRNAs that have a function[217,240]. However, to confirm this hypothesis it would be necessary to do a bioinformatic analysis of all *MIR* sequences comparing the rate of evolutionary change between miRNA and flanking regions[210,240]. This would indirectly suggest that *MIRtop14* has a biological role that makes it being conserved.

Besides, it is interesting to point out that the nucleotide change between miRNAtop14 in different species is a substitution at the 5' end position from a U to a C, since it has been reported that sRNAs starting with U preferentially bind AGO1, while the ones starting with C predominantly bind AGO5[137,380]. However, a later study looking specifically at miRNAs shows that, although to a lesser extent, AGO1 also accepts miRNAs with a 5'C and AGO5 accepts miRNAs with a 5'U[381]. In this same study, it was found that miRNAs with a C in the 5' position will be sorted to AGO1 if they harbour a G or a U at position 9 and to AGO5 if they have a C or an A at position 9 instead[381]. According to this report, miRNAtop14 would be bound to an AGO5 complex in all studied species, since it starts either by C or by U, but always harbour a C at the 9th position. Thereby, it could be speculated that the change from 5'U to 5'C would just better fit the sorting into AGO5. However, this hypothesis would contradict the finding of AGO5 expression being restricted to somatic cells flanking megaspore cells and mother cells in *A. thaliana*[382], if this was confirmed to be also the AGO5 expression pattern in tomato, as it is likely to be the case[383]. Regardless, an experimental approach such as AGO complex immunoprecipitation[124,137] would be necessary to determine the AGO protein into which miRNAtop14 is loaded.

Finding the origin of *MIRtop14* was not among the main aims of our study, so we did not carry out any in depth analysis on this topic. However, a preliminary analysis using NCBI nucleotide BLAST did not identify any high homology between *MIRtop14* sequence and a protein-coding gene, including its known target (see chapter 5), neither to a transposable element. By ruling out that *MIR14* originated from an inverted gene duplication[208] or from a transposon[219], the spontaneous evolution model would be the one left to explain *MIRtop14* origin[216], and according to it *MIRtop14* could have arisen from a randomly transcribed genomic sequence with a hairpin[216]. Nevertheless, a more systematic analysis would be necessary to reach any conclusion on the origin of *MIRtop14* with some confidence.

It is striking from the pri-miRNA secondary structure analysis that the miRNA and miRNA* sequences are predicted to pair in all secondary structures of all species studied, with the putative exception of spliced *Ipomoea* transcript, where the miRNA arm is predicted to locate within the intron. The conservation of this pairing is likely another indicative of miRNAtop14 playing a biological role. While low divergence in the miRNA sequence probably reflects purifying selection to maintain base pairing with its targets, low divergence in the miRNA* region suggests purifying selection to keep the complementarity between miRNA and miRNA*[230].

Finally, the secondary structure of *S. lycopersicum, N. benthamiana* (both transcripts) and *I. nil* (non-spliced transcript) pri-miRNA agrees with a short base-to-loop pattern of processing: they all show long unstructured loops incompatible with a loop-to-base mechanism of processing and a stem at the base of about 15 nucleotides in length, which matches the short base-to-loop processing pathway but would be too short for the long base-to-loop mechanism to take place[106]. Intriguingly, the secondary structure of pri-miRNA from *P. axillaris* would better agree with a short loop-to-base processing pattern, since it has a very structured loop of less than 50 nucleotides in total while its lower stem is only about 8 nucleotides long, which would be usually considered too short for a base-to-loop mode of processing[106]. Nevertheless, to experimentally confirm the way of processing of each transcript it would be necessary to detect their processing intermediates, which could be carried out through methods such as specific parallel amplification of RNA ends (SPARE)[106,384].

3.3.3. pri-miRNA and miRNAtop14 expression

From the pri-miRNA expression analysis we could establish the presence and position of an intron in *N. benthamiana* and *S. lycopersicum* and the lack of introns in *P. axillaris*. Likewise, the transcription of the whole DNA stretch including both miRNAtop14 and star was as well validated (see appendix for *MIRtop14* sequences and primers). However, although there are predictions based on EST reads, pri-miRNA transcription start sites and polyA sites could not be confirmed in the absence of 5' and 3' RACE analysis, respectively. Nevertheless, it could be noted that when performing reverse transcription PCR of *S. lycopersicum* and *P. axillaris* pri-miRNAs, levels of product were much lower when using the most 5' primer than when using an overlapping (*Solanum*) or contiguous (*Petunia*) primer immediately downstream. On the other hand, the levels of amplification of these immediately downstream primers

were relatively similar to other further downstream primers (data not shown). This observation could be due to a difference in efficiency between primer pairs, but it could as well be an indication that the 5' transcription start site lays within the sequence of the most upstream primer in these two species, or that there are alternative transcription start sites. RACE analysis of pri-miRNAtop14 would be appropriate to answer this question.

Pri-miRNA transcription analysis also showed that *MIRtop14* is subjected to alternative splicing in *Solanum* and *Nicotiana*, since both transcript variants, with and without intron, could be detected after reverse transcription from mRNA or DNAse treated total RNA amplified with a polyT primer. Otherwise, if *MIRtop14* was constitutively spliced, only the intronless transcript would have been amplified. This form of alternative splicing, called intron retention, in which introns are retained in a fully processed, mature mRNA, is indeed the most common mode of alternative splicing in plants[285,385,386]. These results agree with the finding that among the ESTs curated at NCBI, there are some with and some without intron in *Solanum* and *Nicotiana* species. Although intron retention prediction through EST data has been widely used[297,387], it has been argued that it may not always be reliable, since some ESTs may come from partially processed mRNAs or genomic DNA contamination[285,297]. In any case, RT-PCR is usually considered a confirmation of alternative splicing taking place[267,285,386].

miRNA expression analysis in different species indicated that mature miRNAtop14 is produced in all four Solanales species studied, three belonging to the Solanaceae family and one to the Convolvulaceae family, but not in *Arabidopsis*. To better define the species to which miRNAtop14 expression is confined to and at which point did this miRNA appeared, additional Northern analyses could be carried out including samples from other Convolvulaceae members as well as from members of other families within Solanales. Besides, members from other orders close to Solanales, such as Gentianales and Lamiales, could also be tested. Despite that *MIRtop14* genomic sequence has not been identified in these families, if it was present in them it is probable it could be detected with the same probe used in the already performed Northerns, given the high conservation shown by miRNAtop14 sequence among species.

Finally, miRNAtop14 Northern analysis from different parts of tomato plant was not suggestive of any specific spatial pattern of expression. It could be noted, however, that levels seemed to be lowest in root and highest in shoot. Besides, when

looking at different leaflets within a leaf, miRNAtop14 accumulation appeared to be growing from the base to the tip. However, an overall pattern of expression could not be elucidated and the highly changing miRNA levels observed between leaves led us to think that some kind of stimulus-induced expression on top of constitutive or developmentally regulated expression is most likely.

The transformation of tomato, *Nicotiana* or another miRNAtop14 producing plant with a construct harbouring a reporter gene (e.g. GUS) under the miRNAtop14 promoter would help to unravel the expression pattern of miRNAtop14. Additionally, either the transformed or the wild type plants could be subjected to different stimulus in order to assess changes in the levels of miRNA expression as response, either by an assay to detect the change in the marker in transgenic plants or directly by miRNA detection through sRNA Northern blot in wild type plants. Alternatively, in situ detection of miRNAtop14 could be carried out with the same aim[388].

Chapter 4

Intron influence in miRNAtop14 biogenesis

4.1. Introduction

4.1.1. pri-miRNAs processing and splicing crosstalk

Most plant pri-miRNAs are independent transcriptional units generally producing a single mature miRNA[53,54]. Similar to protein-coding transcripts, they are capped, polyadenylated and the majority contain introns[60,62,193].

There are several studies of independent plant *MIRs* with alternative transcription start sites, alternative polyadenylation sites and/or going through alternative splicing[59,62,80,192,196–200]. Likewise, protein coding genes which host a miRNA hairpin in an intron has also been found to have alternative splicing or alternative polyadenylation sites[56,201].

Among miRNA stem-loops contained within an intron of a protein coding gene, there are examples in which splicing of the intron increases the level of mature miRNA[56] and examples of the opposite, in which miRNA is upregulated under splicing inhibition and selection of an alternative polyA site within the intron[201]. Interestingly, in both cases alternative splicing and associated miRNA accumulation correlate with a specific environmental condition, which in these examples is heat stress[56,201].

There are also examples where the miRNA stem-loop includes an exonic and an intronic part, and thereby can only be processed into mature miRNA from the unspliced pri-miRNA variant[199,200]

Another interesting example is the group of nat-miRNAs, where miRNA and miRNA* are located in two different exons separated by an intron and splicing is required for the formation of the hairpin[195].

A case has also been reported where a polycistronic *Arabidopsis MIR* produces two different miRNAs, miRNA842 and miRNA846, from three different alternatively spliced transcript variants (see figure 4.1). miRNA846 is located in an exon and expressed in splicing isoform 1, whereas it is not produced in the other two isoforms because its hairpin is truncated, with the miRNA laying in an intron and the miRNA* in an exon. On the contrary, miRNA842 hairpin is completely exonic in isoform 2 and completely intronic in the other two isoforms, potentially being produced in all three variants[59]. The alternative splicing event is mediated by abscisic acid (ABA), which indicates that it must be related to a biological function[59].

In most cases, however, the miRNA stem-loop is located in a single exon of the pri-miRNA[193,194], so splicing is not necessary a priori for the formation of the

miRNA hairpin and subsequent miRNA processing. Nevertheless, in 2013, two publications demonstrated the influence of adjacent pri-miRNA introns on the levels of mature miRNAs located in an exon[197,198].

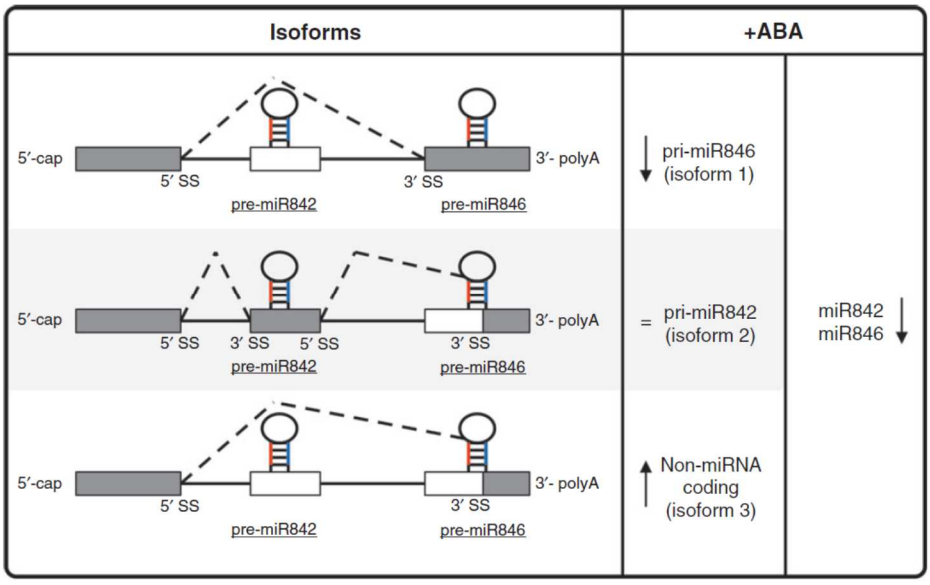


Figure 4.1. AS variants from *Arabidopsis MIR842/846* and associated regulation of pri-miRNA and mature miRNA levels upon ABA application. miRNA846 is located in an exon and expressed in splicing isoform 1, but it is not produced in the other two isoforms because its hairpin is truncated, with the miRNA laying in an intron and the miRNA* in an exon. In *Arabidopsis*, miR842 and miRNA846 arise from different alternatively spliced products. miRNA842 is located in an exon and expressed in splicing isoform 2, whereas it is located in an intron but potentially expressed too in the other two isoforms. Application of ABA to *Arabidopsis* seedling increase the production of isoform 3 while reducing the production of isoform 1, reducing the expression of both miRNA842 and miR846 (indicating that miRNA842 is probably expressed at higher level from isoform 1 than from isoform 3). Grey boxes represent constitutive exons and white boxes alternatively spliced exons in each particular pri-miRNA isoform. Horizontal black lines represent introns. Broken lines denote alternative splicing events. Pre-miRNA842 and pre-miRNA846 stem loops are indicated and their miRNA/miRNA* duplexes are depicted with the miRNA in red and the miRNA* in blue. Broken black lines indicate AS events. Changes in the amount of pri-miRNA isoforms and mature miRNAs 842 and 846 after ABA application are indicated with arrows and equal sign. Image from[193].

4.1.2. Influence of pri-miRNA introns in mature miRNA accumulation

Bielewicz et al.[198] studied the influence of the single intron contained in Ath-*MIR161* and *Ath-MIR163* on the levels of their mature miRNAs. To do so, they generated several transgenes of the *MIR* genes: one WT, one intron-less and three non-sliceable versions (5'SS mutated, 3'SS mutated; 3'SS + 5'SS mutated). They subsequently tested them either by transforming *MIR163* knock-down *Arabidopsis* with the *MIR163* constructs or by performing transient expression *in N. benthamiana* with the *MIR161* constructs. In both cases, mature miRNA levels were strongly reduced in the intron-less *MIRs* compared to the WT.

Besides, a similar mature miRNA reduction occurred when 5'SS or 5'SS + 3'SS mutants harbouring the intron were tested, but not in an equally non-spliceable 3'SS mutant. This peculiarity was interpreted as the 5'SS being the main motif involved in enhancing mature miRNA levels, rather than the intron excision event itself. Interestingly, it was also found that a proximal polyA site in 5'SS mutant, which in WT *MIR* is used around 40% of the times, is used in 80% of the occasions and in 5'SS + 3'SS mutant in over 95%, which was interpreted as a possible shield effect of this proximal polyA site by the functional 5'SS[198].

In addition, it was noted that the amounts of different pri-miRNAs constructs did not correlate with their corresponding mature miRNA levels, indicating that the intron was influencing processing, rather than gene expression[198].

Finally, to probe the biological significance of the intron, they measured the levels of miRNA163 and its mRNA target in WT and 5'SS + 3'SS mutated plants under a pathogen infection, known to induce miRNA163 accumulation. While in WT target levels decreased, in the mutant target levels were shown to rise instead[198].

With all these results, they hypothesised that the enhancement of miRNA accumulation was determined by the interaction between the miRNA biogenesis machinery and the spliceosome, possibly involving the spliceosomal component U1 snRNP, which binds 5'SS[198].

In another publication, Schwab et al.[197] also researched the influence of pri-miRNA introns on mature miRNA accumulation, this time of *Ath-MIR163* and *Ath-MIR172*. They created several *MIR163* transgene variants with and without intron, and consistently found higher levels of miRNA from the intron containing variants, both in *N. benthamiana* leaves transient expression and in stably transformed *A. thaliana MIR163* knock-down seedlings. The same observation was made for

MIR172a, which harbours two introns downstream of the miRNA stem-loop. Constructs without any intron yielded lower amounts of miRNA than constructs with one or both introns.

Besides, to test whether this enhancing effect was caused by features particular to miRNA introns, introns of protein-coding transcripts were placed downstream of the intron-less *MIR163* transgene, obtaining a similar enhancement of miRNA accumulation. However, when placing these same introns upstream of the intron-less MIR163, the enhancing effect was much weaker. Furthermore, when testing MIR172b, which has an intron 5' and another 3' of the miRNA hairpin, constructs with the 5' intron showed the same reduced accumulation of mature miRNA172b as the intron-less constructs[197].

Subsequently, to assess whether the excising of the intron was itself necessary for the positive effect in mature miRNA accumulation, they generated a non-spliceable transgene for *MIR163* and another one for *MIR172a* by mutating the 5'SS. miRNA mature levels from non-spliceable constructs were slightly reduced compared with WT in *N. benthamiana* transient transformation but rather similar in *Arabidopsis* stable transformation, indicating that splicing itself is not required for intron enhancement of miRNA accumulation[197].

Next, the group decided to investigate if increased miRNA levels from intron-containing *MIR* could be a consequence of increased pri-miRNA levels, since a known effect of introns is the intron-mediated enhancement of gene expression. With this aim, an ectopic copy of *MIR163* either with or without intron was introduced into *dcl1 Arabidopsis* mutants, in order to measure pri-miRNA levels without the interference of processing. Although there was slightly higher amount of pri-miRNA from the intron-containing *MIR*, the difference in levels of miRNA primary transcript between intron-less and intron-containing construct did not explain the great difference previously observed in mature miRNA levels, and therefore the intron-mediated enhancement of gene expression was discarded as cause of the intron effect in miRNA levels[197].

Finally, they tested whether pri-miRNA processing could in turn affect the splicing of introns as well. For this experiment, they measured the amounts of pri-miRNA with excised vs. retained intron by transforming WT and *dcl1* mutant with *MIR163*, *MIR172a* and *MIR172b* constructs. It was found that splicing was more efficient in a *dcl-1* mutant than in WT background for those constructs with a 3' intron (*MIR163* and *MIR172a*), while unchanged when the intron was upstream to the

miRNA stem-loop (*MIR172b*). This result could indeed suggest that dicing has a negative effect in splicing of downstream introns[197].

After all these analyses, they concluded that the intron induced increase of miRNA accumulation had to be due either to enhanced pri-miRNA processing or to reduced mature miRNA turnover[197].

There are a few differences in the results between one and the other study, the most obvious is probably the different levels of mature miRNA obtained when mutating *MIR163* 5'SS (strongly reduced in Bielewicz et al.[198] study while almost unchanged in the study by Schwab et al.[197]). Both groups attribute this discrepancy either to the different promoters used in each study (*MIR163*[198] vs. CaMV35S[197]) or to the different 5'SS point mutations, that in the case of Schwab et al. experiment may not have been enough to abolish the recruitment of binding factors (such as U1 snRNP, according to Bielewicz et al.[198]). In any case, it is clear in both investigations that, at least in the *MIRs* studied, introns downstream of the miRNA hairpin have a positive effect in the accumulation of the mature miRNA[197,198]. Furthermore, the evidences shown from both works point towards the existence of a cross-talk between pri-miRNA splicing and processing that regulates each other[197,198,389] (see fig. 4.2). However, further research is needed to elucidate the specific mechanisms by which this happens.

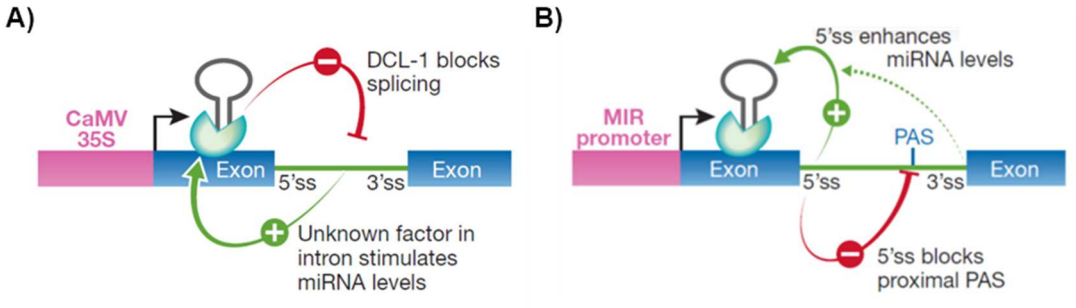


Figure 4.2. Enhancement of miRNA processing by plant introns. **A)** Schwab et al[197] created *MIR* transgenes with an exonic miRNA hairpin and a downstream intron (green line) under the CaMV 35S promoter (pink box). From their observations, they propose that unknown factors within the intron positively influence miRNA processing while miRNA processing negatively affects splicing, suggesting an interaction between processing and splicing machinery. **B)** Bielewicz et al[198] created *MIR* transgenes with an exonic miRNA hairpin and a downstream intron (green line) under the native *MIR* promoter (pink box). From their results, they propose that a functional 5'SS is key in the intron mediated enhancement of miRNA processing, while it also plays a role in supressing a proximal polyA site within the intron. They suggest that splicing factors binding the 5'SS interact with the miRNA processing machinery. PAS, polyA site. Image adapted from[389].

4.1.3. Study of miRNAtop14 intron influence

Until recently, the pri-miRNA region outside the stem loop had been largely ignored[389]. However, recent studies suggest that pri-miRNA introns and their splicing and/ or alternative splicing influence mature miRNA accumulation and may be regulating miRNA levels spatially and temporally in response to specific conditions[194,390].

pri-miRNAtop14 has an uncommon structure in which miRNA and miRNA* are in two different exons separated by an intron, which is spliced out or retained in the two alternatively spliced transcript variants (see chapter 3).

This characteristic prompt us to hypothesise that the non-spliced variant, despite of creating miRNA-miRNA* pairing, would probably show reduced levels of mature miRNAtop14 given its complex miRNA fold-back structure with a large loop, which does not match most predicted miRNA hairpin structures and associated modes of processing, although some other examples of relatively long loops have also been observed in plants[106].

On the other hand, given the results obtained by Bielewicz et al.[198] and Schwab et al.[197] showing the positive influence of introns in mature miRNA accumulation, we speculated that pri-miRNAtop14 intron may enhance miRNA biogenesis. Supporting this hypothesis, the same intron enhancing effect has been observed in intron-split hairpin RNAs (hpRNAs) engineered to induce PTGS in plants[391,392]. It was found that including a spliceable intron in between the two arms of the hairpin enhanced efficiency of PTGS to almost 100%, much higher than in other construct with non-spliceable spacers in between the arms[391,392].

In sight of this evidence, we decided to create a system to check whether *MIRtop14* intron is influencing mature miRNA levels, and whether it is doing so in two antagonistic ways (enhancement and repression) determined by alternative splicing.

4.2. Results

4.2.1. A system to assess intron influence in mature miRNAtop14 levels

Solanum lycopersicum MIRtop14 was chosen to analyse whether there were any differences in the amount of mature miRNAtop14 coming from each of the two alternatively spliced transcript variants, one retaining the intron and the other one with the intron spliced, as well as from a third variant coming from the intron-retaining transcript mutated to impede splicing.

Each of the three *MIRtop14* variants was assembled with three more transcriptional units (two controls of transformation and one control of gene expression) into independent constructs which were used to transform *Arabidopsis thaliana* inflorescences. miRNAtop14 levels were then analysed in T2 generation of plants transformed with the three different constructs.

The process, in more detail, consisted in reverse transcription of total RNA and PCR amplification of both pri-miRNAtop14 alternative splicing variants. The sequences chosen for amplification were almost the same as the ones amplified in the analysis of pri-miRNAtop14 expression, the only difference was that in this case the 5' primer was 17 nucleotides upstream compared to the previous analysis (see figure 3.5 for Sly-pri-miRNAtop14 expression analysis and appendix for *MIRtop14* sequences and primers).

Subsequently, both pri-miRNAtop14 variants were cloned first into the pGEMT-easy vector and then transferred to a level 0 golden gate vector. At this point, site-directed mutagenesis was used to mutate the 5'SS sequence of the pri-miRNA harbouring the intron, which was modified from the G/GT exon/intron splice site consensus sequence to C/CC. The inability of the resulting transcript to undergo spicing was assured by reverse transcription PCR.

Every one of the three *Sly-MIRtop14* constructs was subsequently assembled into a full transcriptional unit under the CaMV 35S promoter and terminator hosted in a level 1 golden gate construct.

In parallel, three more level 1 golden gate plasmids harbouring full transcriptional units for Kanamycin resistance protein (KanR), Discosoma sp. red fluorescent protein (DsRed) and *Osa-MIR528* were either gathered or assembled de novo. *Osa-MIR528* was placed under the CaMV 35S promoter and terminator as *MIRtop14*, KanR was under the Nopaline synthase promoter (pNOS) and terminator

(tNOS) and dsRED under the *Arabidopsis* ubiquitin 10 promoter (AtUBI10) and the tNOS terminator.

Both KanR and dsRED genes were chosen as selectable markers of plant transformation. Besides, the monocot specific *Osa-MIR528* gene was included to have a control of miRNA expression between different transgenic lines from the T-DNA insert.

Finally, all three control transcriptional units were assembled together into three independent golden gate level 2 binary constructs with a different *MIRtop14* variant each (see fig. 4.3). These three multigene constructs were introduced into *Agrobacterium tumefaciens* GV3101 (pMP90) to mediate *Arabidopsis* female gametophyte cell lineages transformation through the method known as “floral dip”. Transformed individuals among the offspring plants were identified as those growing in Kanamycin containing media and showing dsRED fluorescence. These individuals (T1) were grown until producing seed, and each gave place to an independent transgenic line.

T2 seeds were again sowed on Kanamycin containing media, and 20 healthy growing seedlings per transgenic line were pooled together to analyse miRNAtop14 levels. Three independent transgenic lines per construct were chosen for the analysis.

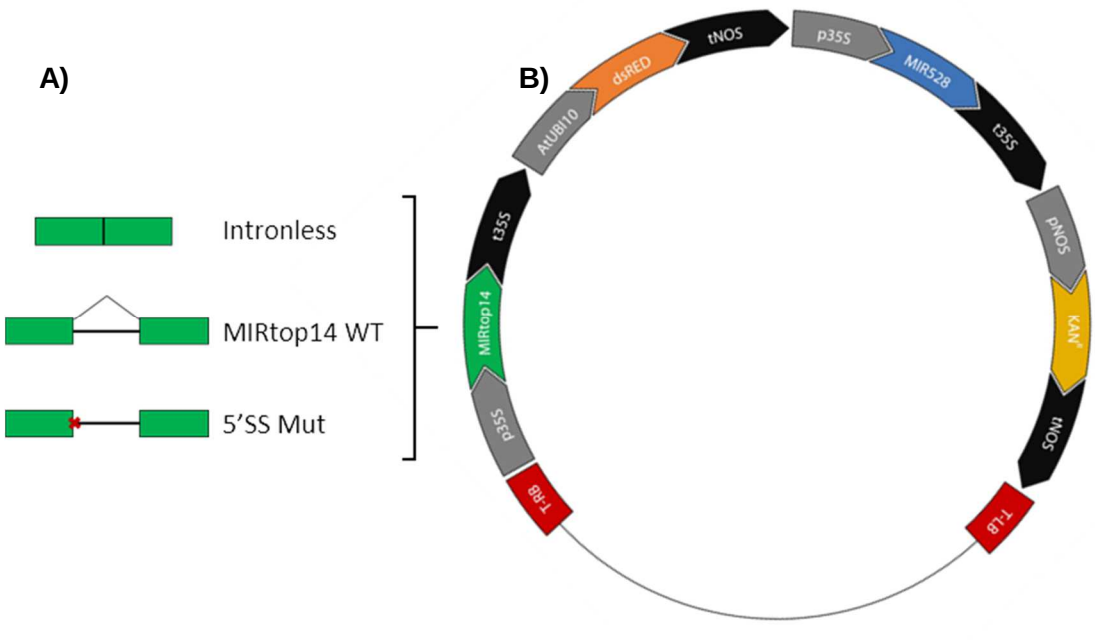


Figure 4.3. Scheme of the constructs used for *Arabidopsis* transformation. **A)** Three different constructs only differing in *MIRtop14* sequence were created: one harbouring the wild type *MIRtop14* sequence with an spliceable intron ("*MIRtop14 WT*", two green boxes separated by a black line), another one consisting in the two *MIRtop14* exons without intron ("Intronless", two consecutive green boxes) and a final one similar to wild type *MIRtop14*, but with the 5'SS mutated from G/GT to C/CC making the intron non-spliceable ("5'SS Mut", two green boxes separated by a black line and a red cross indicating the mutation at the 5'SS). **B)** The rest of the plasmid apart from *MIRtop14* sequence was the same in all three constructs, consisting in four full transcriptional units in between the T-DNA right and left borders ("T-RB" and "T-LB" red squares in the image): *MIRtop14* sequence (green box) under CaMV 35S promoter and terminator (p35S and t35S grey boxes); dsRED sequence (orange box) under AtUBI10 promoter and NOS terminator (AtUBI10 and tNOS grey boxes); *MIR528* sequence (blue box) under CaMV 35S promoter and terminator (p35S and t35S grey boxes); KanR sequence (yellow box) under NOS promoter and terminator (pNOS and tNOS grey boxes).

4.2.2. Intron influence in *S. lycopersicum* miRNAtop14 levels

Ten total RNA samples were analysed by Northern blot for small RNA detection. The samples analysed corresponded to three biological replicates (three independent transgenic lines) per each *MIRtop14* construct (with intron, without intron and with non-spliceable intron, see figure 4.3). A sample from WT Col-0 *Arabidopsis* was as well included in the analysis as negative control for pri-miRNAtop14 gene expression.

Each total RNA sample was extracted from a total of 20 seedlings per transgenic line (or WT) grown for 16 days in media with Kanamycin (except for WT, where Kanamycin was not added). To verify that each sample harboured the expected *MIRtop14* construct (or none), PCR was carried out to amplify all alternatively spliced pri-miRNAtop14 sequences (see fig. 4.4.B).

RNA samples were separated by electrophoresis in a denaturing polyacrylamide-urea gel and subsequently transferred and fixed to a nylon membrane. miRNAtop14 was detected by hybridization with a ^{32}P-radiolabeled DNA probe complementary to its full 21 nucleotides sequence. The results of this Northern blot analysis are shown in fig. 4.4.A.

Although there are variations in miRNAtop14 levels between different transgenic lines, there is a trend that the intron-containing WT *MIRtop14* produces the highest amount of mature miRNA while *MIRtop14* without intron produces slightly lower levels and the 5'SS mutated version of the intron-containing *MIRtop14*, which cannot undergo splicing, produces the lowest levels of all. These results suggest that the intron of pri-miRNAtop14 could have a positive effect on mature miRNA accumulation, but at the same time restricts miRNA processing when it is not spliced. Besides confirming the enhancement of miRNA production by *MIR* introns, these results suggest that an alternative splicing event might be regulating the levels of mature miRNAtop14 depending on the alternate retention or splicing of its pri-miRNA intron.

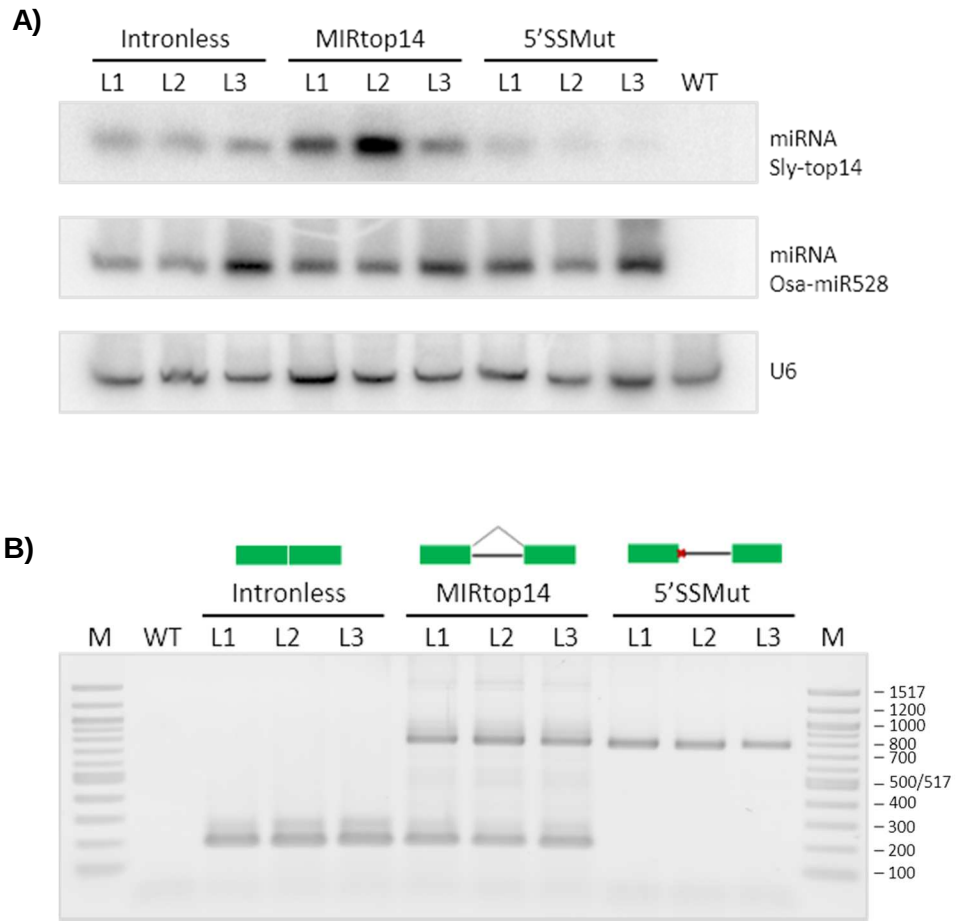


Figure 4.4. Effect of the intron in mature miRNAtop14 accumulation. **A)** Northern blot detection of mature miRNAtop14 in *A. thaliana* wild type (WT) and *A. thaliana* transformed with the three *MIRtop14* constructs (Intronless, *MIRtop14* and 5'SS MUT; see fig. 4.3). Three independent transgenic lines (L1, L2 and L3) were analysed per construct. Osa-miRNA528 detection was included as internal control of gene expression. U6 detection was included as loading control. **B)** Detection of pri-miRNAtop14 in each sample after total RNA reverse transcription. Bands of 752 nucleotides correspond to the amplification of pri-miRNAtop14 with intron and bands of 184 nucleotides correspond to the amplification of pri-miRNAtop14 without intron. M, 100bp marker.

4.3. Discussion

The unusual *MIRtop14* exon-intron structure prompted us to further study the possible functional implications of this structure in miRNAtop14 regulation. With this aim we created a system to measure mature miRNAtop14 coming from different variations of this exon-intron structure: we assessed the WT *MIR*, an intronless variant and a 5'SS mutated variant in which splicing was inhibited. The results show an enhancement of mature miRNA levels by the intron, as long as it can be spliced. Otherwise, it negatively influences miRNAtop14 accumulation (see figure 4.4).

Similar results have been obtained in three independent transgenic lines from each one of the three *MIRtop14* variants, each line comprising a pool of 20 different seedlings. Besides, an internal control of gene expression was included in this analysis to reach a confident conclusion from these results. Such a control is required because we cannot discard the possibility of some of the transgenic lines harbouring more than one T-DNA insert, which can lead either to increased expression or to silencing, mainly through PTGS[393]. Likewise, we neither can ignore the influence of the location in the chromosome in which an insertion has taken place in its level of expression, a phenomenon called position effect[394,395]. Therefore, in our constructs we have included as control *Osa-MIR528* within the same T-DNA as *MIRtop14*, so they both will be in the same copy number and in the same chromosomal locus. This allows the normalisation of the levels of mature miRNAtop14 based on the levels of mature miRNA528 detected through Northern analysis, making it possible to confirm that any differences in miRNAtop14 accumulation between *MIRtop14* three variants will be a consequence of their exon-intron structures and not of other factors. Consequently, it can be concluded that *MIRtop14* exon-intron structure is influencing mature miRNA levels.

In previous studies, the positive effect of intron on miRNA accumulation has been reported to occur when introns were downstream of the miRNA hairpin, while introns upstream of the miRNA hairpin showed no effect on mature miRNA levels[197]. This has been interpreted as indicative that whatever the factors mediating the observed effect of intron on miRNA accumulation are, they must be acting either before or during spliceosome recruitment[197]. Interestingly, *MIRtop14* has a structure different from the other miRNAs studied, with an intron in between miRNA and miRNA*, rather than upstream or downstream of their hairpin. Therefore, it is revealing that an intron in the middle of the miRNA stem-loop has a similar positive effect on miRNA accumulation to an intron located downstream to the stem-loop,

because such observation supports that the enhancement of miRNA accumulation is determined simultaneously to spliceosome recruitment, rather than before.

Although our results reinforce the observation of introns having a positive effect in miRNA processing, further experiments would be needed to determine at which level this enhancement is taking place: increased pri-miRNA expression, increased pri-miRNA processing or reduced miRNA turnover. Increased processing is the option favoured by previous studies, mainly because enhanced *MIR* expression was discarded by measuring pri-miRNA levels and comparing them with mature miRNA levels[197,198]. Likewise, we could perform quantitative RT-PCR with the samples analysed by Northern (see figure 4.4) to measure the levels of pri-miRNAtop14 transcripts and correlate them with miRNAtop14 levels. Besides, we could also replicate the experiment of *Arabidopsis* transformation with the three *MIRtop14* containing constructs using *dcl1* mutants instead of WT plants, as done by Schwab et al.[197]; this way we could properly assess whether the intron is having any effect in *MIRtop14* expression putting aside the interference in pri-miRNA levels caused by processing.

Apart from a positive influence of the intron on miRNAtop14 accumulation, we could verify that the large loop in between miRNAtop14 and miRNAtop14* hinders miRNA processing, since the non-spliceable *MIRtop14* shows a very low amount of mature miRNA. This was an expected result given that, if the miRNA hairpin is processed from base-to-loop as predicted, the excised long pre-miRNA molecule would be more prone to create secondary structures or interactions that could interfere with Dicer cleaving. However, it is interesting to note that, although very reduced, miRNA biogenesis can still take place despite of the intron not being spliced. Based on this observation, the affirmation that nat-miRNAs need to be spliced in order to be processed[195] may need to be examined.

If *MIRtop14* is indeed alternatively spliced as our results suggest both in *Solanum* and *Nicotiana* (see chapter 3), that would imply that the production of miRNAtop14 can be post-transcriptionally regulated: the spliced transcript variant will produce high levels of miRNA (to which the presence of the intron in the pre-mRNA will contribute positively) while the transcript variant with the intron retained will produce very low levels of miRNA, if any at all.

The affirmation that miRNAtop14 levels are regulated post-transcriptionally through alternative splicing is controversial, as it is difficult to explain why *MIRtop14* would have two alternatively spliced variants instead of being constitutively spliced,

when only the spliced variant seems to have a function. It could be argued that the intron-retaining pri-miRNAtop14 transcript is just a result of mis-splicing and that it does not respond to any biologically relevant function. This is a claim that was often made for intron retention events, especially in animals, but that current evidence is dismissing[396].

Alternatively, it could be alleged that the low levels of miRNAtop14 produced by the non-spliced variant are indeed playing a role and are the response to a requirement for a low amount of miRNAtop14 in some specific cells/tissues, developmental points or environmental conditions.

It has been also proposed that mRNA levels can be modulated by changing the ratio between functional transcripts and nonsense or intron-retaining transcripts[397], which could be the role for the alternative splicing event of *MIRtop14*.

Another interesting possibility comes from a study in the plant *Marsilea vestita*, where intron retention controls translation in a developmentally regulated manner; certain mRNAs are produced and kept with introns until reaching specific developmental stages, at which point they go through splicing and induce a quick protein production[302]. Similarly, intron-containing pri-miRNAtop14 may be initially produced and stored to allow a rapid miRNAtop14 production at some point, when some specific cue would trigger its splicing and subsequent miRNA biogenesis.

In order to assess whether *MIRtop14* alternative splicing is indeed playing a role in the regulation of the production of mature miRNAtop14 levels, we could perform quantitative RT-PCR to measure pri-miRNA levels, and more specifically to estimate the amount of each alternatively spliced variant. By testing the amount of each pri-miRNA transcript in samples with different levels of mature miRNAtop14, it would be possible to unravel any correlation between them and mature miRNA amount. Finally, if there is such correlation, finding a stimulus or spatio-temporal pattern associated with the change in the ratio between the two pri-miRNA variants would confirm that miRNAtop14 levels are being regulated through a biologically relevant alternative splicing event.

Chapter 5

miRNAtop14 mRNA target: *LPR*

5.1 Introduction

5.1.1. miRNAs function, mode of action and roles

The miRNAs' main function is to regulate the level of expression of genes[78]. As already seen in the introduction, they perform this regulation by targeting one or several specific mRNAs and reducing their translation[78]. They achieve this by assembling into a RNP complex, the so-called RISC, and leading it towards the target transcript through base-pairing[121].

Once the RISC-target interaction is established, one of two processes can follow; either mRNA degradation (in the case of plants, triggered by the direct cleavage of the mRNA by Ago), or inhibition of translation of the mRNA (in which case mRNA decay does not seem to follow in plants)[141]. Which of the two pathways is followed is determined by which protein (HYL1 or DRB2) associates with DCL1 during the miRNA biogenesis, according to a recent study[105]. However, it is still unclear whether there are biological implications that favour the occurrence of one process over the other.

In any case, both actions lead to a reduction in protein levels from the targeted genes, which together with other cellular regulation processes adjust the levels of proteins in a specific place and moment determined by different internal or external cues. This regulation may affect single proteins, but in many cases, miRNAs influence whole regulatory networks by targeting key genes, such as transcription factors or hormone receptors[166].

The regulatory influence of miRNAs in development is long known thanks to experiments with mutants in components of the miRNA biogenesis pathway[398–402]. Later it was discovered that miRNAs play an important role in the response to environmental stimuli. One study found that specific miRNAs were expressed when *Arabidopsis thaliana* plants were grown under drought, high salinity or low temperature conditions[403] and another study showed that miR395 was induced under low-sulphate stress[143]. Besides miRNA responses to the most common stresses such as drought, salinity, temperature or nutrient deficiency stresses, involvement of miRNAs in responses to stimuli as varied as hypoxia, UV B irradiation, oxidative stress or bacterial pathogenesis has been reported over the years[404]. An interesting aspect of miRNA function is that it can influence plant development based on the environmental conditions, therefore acting as a link between both[405]. For instance,

plants can vary their root architecture depending on the availability of nutrients in the soil through processes involving miRNAs activity[406].

5.1.2. miRNA-target evolution

When *MIRs* arise *de novo*, they may already have a target, usually if they come from the inverted duplication of their target gene [407], or they may have no target, as has been suggested for many young miRNAs[230,237,408]. This second possibility would occur mainly in miRNAs derived from transposons[219,223] or from hairpin folding RNAs that appear randomly in the genome[216]. In this second case, most miRNAs lacking a target would be lost by mutational drift[230,237,408]. However, over time a few of these miRNAs could eventually find a complementary transcript and, if the interaction gives an advantage, the miRNA-target pair would be fixed[216,223]. For instance, purifying selection has been shown to contribute to the conservation of miRNAs and target sites[217,408].

5.1.3. Multicopper oxidases and LPR

Multicopper oxidases (MCOs) are a family of enzymes widely distributed across life taxa[409,410]. They are characterised by containing four copper atoms arranged in two centres; a type 1 centre with one atom and a type 2/ type 3 centre with three atoms each[411]. Through electron transfer from one to the other centre, they perform substrate oxidation together with the reduction of dioxygen to water[411]. Structurally, MCOs are formed of multiple domains homologous to cupredoxin[412]. Depending on their substrate, they can be divided in several subclasses, such as laccases, ascorbate oxidases, ferroxidases or bilirubin oxidases[409,410].

Two proteins belonging to the MCO family of enzymes are the so-called LOW PHOSPHATE ROOT1 (LPR1) and its close paralogue LPR2, both first identified in Arabadipsis[413]. More specifically, they are ferroxidases, since both of them have been probed to display iron oxidation activity[414].

The modification of root architecture in response to media low in phosphate is a long-known phenomenon; proliferous roots allow plants to better explore the soil and to expand their surface to uptake this nutrient[415]. In *Arabidopsis*, low phosphate media determines a reduced growth of the primary root and an increased growth of the lateral roots[416].

LPR1 was initially mapped as an Arabidopsis quantitative trait locus (QTL) connected to primary root growth arrest under low phosphate in the soil[417]. Later on, the *LPR1* gene responsible for this trait was identified, as well as a paralogous gene, *LPR2*[413]. In this study, experiments with *lpr1* and *lpr2* loss of function mutants indicated that both LPR1 and LPR2 play a role in the arrest of primary root growth under low phosphate, with *lpr2* showing a lower effect than *lpr1* and *lpr1, lpr2* double mutant showing an additive effect between both genes[413] (see fig. 5.1.A). Besides, *LPR1* mRNA was detected in both *Arabidopsis* leaves and roots, but further experiments focused only in LPR1 localization within the root, since it is where its only known function was carried out[413]. This way, it was find out that, although with differences between alleles, LPR1 is mainly expressed in the root tip, within the meristematic region and the root cap[413]. Furthermore, it was demonstrated that it was necessary that the root tip was in contact with the low phosphate media, rather than any other part of the plant, to trigger the arrest of growth response[413].

In a subsequent publication, LPR1 was specifically localized to the ER in root cells[418]. Furthermore, it was found to interact with another protein also present in the ER of root tip cells; PHOSPHATE DEFICIENCY RESPONSE 2 (PDR2)[418]. While *pdr2* mutants have an oversensitive response to low phosphate and develop an extremely short root with a very reduced meristem under phosphate deficiency conditions[419], *pdr2, lpr1, lpr2* triple mutants show a phenotype similar to that of *lpr1, lpr2* double mutant[418] (see fig. 5.1.B). It was therefore proposed that PDR2 could act upstream in the process, regulating LPR1/LPR2 and all together adjusting the activity of the primary root meristem under low phosphate conditions[418].

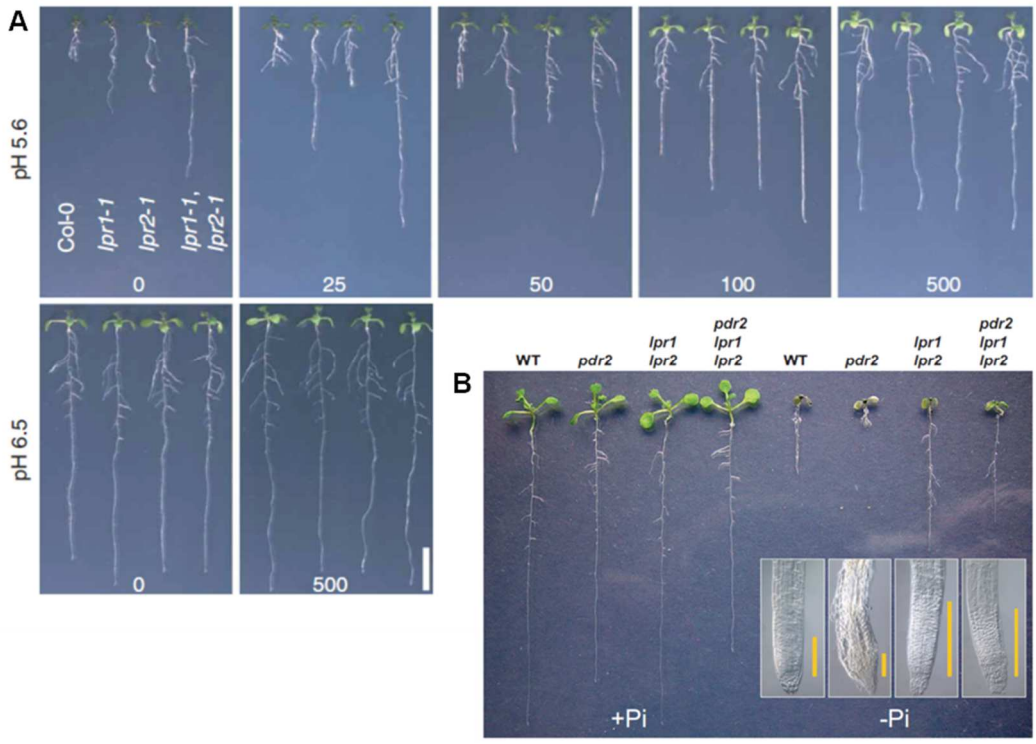


Figure 5.1. LPR function in *Arabidopsis* root growth arrest under low phosphate. **A)** LPR1 and LPR2 display a similar function in root growth arrest under low phosphate. The effect of LPR1 is stronger, but both effects are cumulative. Seeds of wild type Col0, *lpr1* and *lpr2* single mutants and *lpr1, lpr2* double mutant were grown in media with the concentrations of phosphate indicated at the bottom of each picture (µM) and the indicated pH and resultant phenotypes are shown. White scale bar, 1 cm. **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 phosphate and the resultant phenotypes are shown. Insets show enlarged root tips, with yellow bars indicating meristem length. Image adapted from[413,414].

A later publication shows the first evidence of a mechanism by which LPR1/ LPR2 and PDR2 could control root growth[414]. In this study, LPR1 expression in the root apical meristem is detected in the cell wall apart from the ER, and both LPR1 and LPR2 are found to have ferroxidase activity[414]. LPR1-dependent Fe oxidation in the cell wall, specifically in the apoplast, would initiate redox cycling and the creation of reactive oxygen species (ROS), which would trigger callose deposition that, in turn, would impair symplastic communication in the apical root meristem[414]. The inhibition of symplastic communication would lead to meristem reduction, determining the arrest in root growth observed in low phosphate conditions[414]. Besides, it is suggested that it is not only phosphate deficiency, but the antagonistic interaction between phosphate and iron availability, which triggers this response[414].

Finally, a very recent study divides this response into two processes: the rapid inhibition of cell elongation in the elongation zone of the apical root (just above the meristem) and the subsequent inhibition of the cell division in the meristem[420] (fig. 5.2).

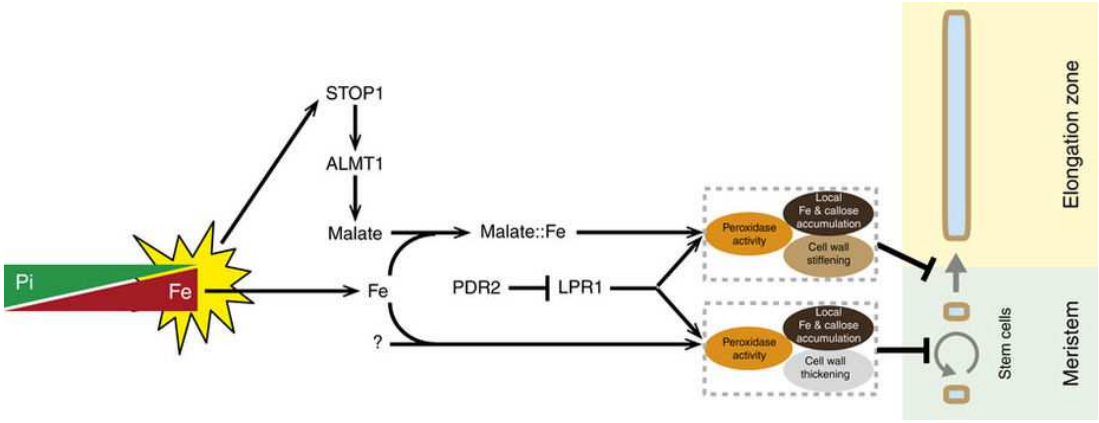


Figure 5.2. Scheme depicting the pathway that inhibits cell division and cell elongation processes in the root as response to a low Pi/ Fe ratio. A low Pi/ Fe ratio activates the transcription factor STOP1, which induces the expression of *ALMT1*, which in turn facilitates malate exudation into the apoplast. Concurrently, PDR2 stops inhibiting LPR1, which starts oxidising Fe^{2+} into Fe^{3+} also in the apoplast. As a result, Fe^{3+}-malate complexes are formed in the apoplast, giving place to iron accumulation, LPR1-ferroxidase dependent redox cycling, peroxidase activity, ROS creation, callose deposition and cell wall stiffening, which determines the cease of root cell elongation. In a second process independent of STOP1 or ALMT1 but also triggered by a low Pi/ Fe ratio, PDR2 and LPR1 are involved in an unknown pathway that equally leads to iron accumulation, redox cycling, peroxidase activity, ROS creation, callose deposition and, in this case, cell wall thickening[420]. The final result is the inhibition of cell division in the meristem. Image from[420].

In the first process, a low phosphate/ iron ratio post-transcriptionally activates the transcription factor SENSITIVE TO PROTON TOXICITY1 (STOP1) to induce the expression of its target, the malate channel encoding gene *ALUMINUM-ACTIVATED MALATE TRANSPORTER1* (*ALMT1*), which will facilitate malate exudation into the apoplast. At the same time, PDR2 stops inhibiting LPR1 which starts oxidising Fe^{2+} into Fe^{3+}, also in the apoplast. Consequently, Fe^{3+}-malate complexes will be formed in the apoplast, giving place to iron accumulation, LPR1-ferroxidase dependent redox cycling, ROS creation and callose deposition as already explained in the previous study[414,420]. Besides, peroxidase-dependent cell wall stiffening is also observed in this study, which determines the cease of root cell elongation[420].

The second process (in which cell division in the meristem is inhibited) does not require the activity of STOP1 or ALMT1, as have been observed in *stop1* and

almt1 mutants. However, it does involve PDR2 and LPR1 and the pathway of iron accumulation, redox cycling, ROS creation, callose deposition and peroxidase activity, that determines, in this case, cell wall thickening[420].

5.1.4. Study of miRNAtop14 target

When top14 was first discovered in tomato, Moxon et al.[370] predicted a target for this sRNA; the transcript variant 2 of CTR1-like protein kinase (CTR4). However, when they tried to validate this target performing RNA ligase-mediated 5' amplification of cDNA ends (RLM-RACE), no cleavage was detected[370].

Later on, when Baksa et al.[372] identified top14 in *Nicotiana benthamiana*, they also proposed a target based on their PARE libraries data. However, this data did not show an irrefutable miRNAtop14 directed cleavage of the target as can be appreciated in the T-plot (see fig. 5.3), and confirmation by RLM-RACE was not carried out.

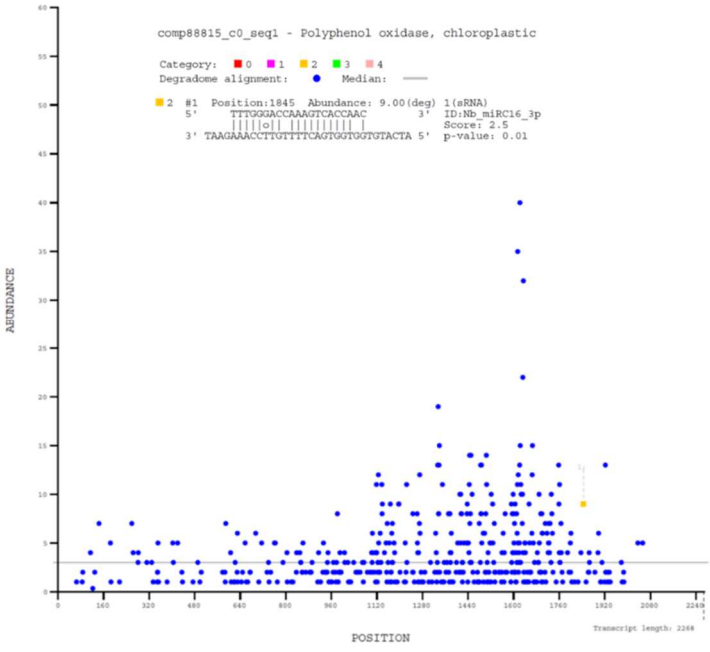


Figure 5.3. T-plot of *Nicotiana benthamiana* polyphenol oxidase mRNA comp88815_c0_seq1, from Baksa et al.[372] degradome data. Blue dots represent abundance vs position of degradome reads. The yellow dot indicates the reads coming from the position where miRNAtop14 would putatively cleave, which corresponds to a category 2 out of 4, 0 being the most reliable cleavage. Image from[372], supplementary material.

In this study, we have again attempted to find evidence of any transcript cleaved by miRNAtop14-RISC. With this aim, we first performed RLM-RACE analysis and degradome library search of putative targets in *Solanum lycopersicum*. Once a target was identified, we carried out RLM-RACE and degradome search to check miRNAtop14 directed cleavage of this target in *Nicotiana benthamiana*, and then RLM-RACE in *Petunia axillaris*. Besides, we have also identified this target in most of the species of Solanales where we had previously identified miRNAtop14, and we have made a prediction of the interaction in each case.

5.2. Results

5.2.1. Identification of miRNAtop14 target in *Solanum lycopersicum*

5.2.1.1. RLM-RACE

A list of putative *Solanum lycopersicum* miRNAtop14 (Sly-miRNAtop14) targets was obtained through psRNAtarget server, version 2011, which analyses target-site accessibility as well as complementarity between target transcript and miRNA based on the specific rules of plant miRNA-mRNA interactions[421]. A total of 9 possible targets were predicted by this tool when comparing miRNAtop14 sequence against *Solanum lycopersicum* transcript cDNA library version 2.4, SGN[422], using the default parameters (see table 5.1 caption for details). The list of predicted Sly-miRNAtop14 targets together with each complementarity score (expectation), target accessibility, miRNA-mRNA alignment and mode of target inhibition can be seen in table 5.1.

All nine putative Sly-miRNAtop14 targets were subsequently tested experimentally by RLM-RACE, despite some of them being predicted to undergo translational repression rather than cleavage according to psRNAtarget (see table 5.1). Total mRNA from three tomato plantlets aerial parts was used for the analysis. *GRAS24* and *LANCEOLATE* (*LA*) transcripts were included in the analysis as experimental controls. These two transcripts were chosen because miRNA-directed cleavage had been confirmed for both by RLM-RACE in tomato[423,424]. The presence of all transcripts in the sample was confirmed by RT-PCR amplification before the RLM-RACE assay. The results of this experiment are shown in fig. 5.4.A.

Target 5 (*Solyc05g008290.2.1*) was the only one showing a clear band of the expected amplicon size. This band was recovered from the gel and cloned into pGEM-T easy vector. 26 clones were sequenced and 25 of them were the amplification product of the transcript precisely cleaved between nucleotides complementary to the 10th and 11th nucleotide of miRNAtop14 (see fig. 5.4.B). This analysis clearly indicated that *Solyc05g008290.2.1* transcript is subjected to miRNAtop14 directed cleavage.

According to SGN, transcript Solyc05g008290.2.1 is predicted to encode a bilirubin oxidase enzyme, which is a type of multicopper oxidase protein407(Sakurai et al., 2007),408(Solomon et al., 1996). However, when performing BLAST search of the transcript sequence against NCBI nucleotide collection, the identical mRNA is

predicted to code for a multicopper oxidase, but for the ferroxidase LPR2 instead of a bilirubin oxidase.

As already seen in the introduction, LPR is a protein first discovered in *Arabidopsis thaliana* where there are two variants: LPR1 and LPR2. However, our search in both SGN and NCBI indicated that there is only one variant of *LPR* in *Solanum lycopersicum*. Surprisingly, although in NCBI the tomato *LPR* mRNA is called *LPR2*, when performing BLAST of this sequence against TAIR10 Transcripts[425], the best alignment is with *A. thaliana LPR*1 (although *LPR*2 also shows homology). Therefore, we concluded that miRNAtop14 target in tomato is indeed *LPR*, but decided not to specify the variant in the absence of more information.

No.	Target accession, description	Exp.	UPE	miRNA-target alignment	mode of action
T1	Solyc08g074640.1.1, Polyphenol oxidase	1	20.406	miRNA 20 AACCACUGAAACCAGGGUUC ::::::::::::::::::: Target 1723 GUGGUGACUUUGGUCCCAAG	Cleavage
T2	Solyc06g008810.2.1, Auxin F-box protein 5	2.5	13.888	miRNA 21 CAACCACUGAAACCAGGGUUC :::: ::: :::::::.:::: Target 973 GUUGAUGAAUUUGGUCUCAAG	Cleavage
T3	Solyc10g080180.1.1, U3 small nucleolar RNA-associated protein 18	2.5	14.166	miRNA 20 AACCACUGAAACCAGGGUUC ::: ::::: ::::.:::: Target 443 UUGAUGACUAUGGUUCCAAG	Translation
T4	Solyc07g066260.2.1, Protein phosphatase 2C	2	17.499	miRNA 20 AACCACUGAAACCAGGGUUC :::::::.::: :::.:::: Target 1225 UUGGUGAUUUUUGUCUCAAG	Translation
T5	Solyc05g008290.2.1, Cupredoxin / Bilirubin oxidase (LPR in NCBI)	3	18.982	miRNA 20 AACCACUGAAACCAGGGUUC :::::::::::::::. : Target 136 UUGGUGACUUUGGUCCUCAA	Cleavage
T6	Solyc03g115840.2.1, DnaJ homolog subfamily C member 10	3	20.11	miRNA 21 CAACCACUGAAACCAGGGUUC :::::: :. :::::::.:::: Target 1888 GUUGGUCAUCUUGGUCUCAAG	Cleavage
T7	Solyc03g118010.2.1, RISC, nuclease component Tudor-SN	3	19.192	miRNA 20 AACCACUGAAACCAGGGUUC :::::: :::: ::::::: Target 2537 UUGGUGCCUUUAAUCCCAAG	Translation
T8	Solyc08g062910.2.1, Elongation factor EF-2	3	19.407	miRNA 20 AACCACUGAAACCAGGGUUC :::::: ::::::::::.:: Target 2000 UUGGUG-CUUUGGUCCUGAG	Cleavage
T9	Solyc08g062920.2.1, Elongation factor EF-2	3	19.407	miRNA 20 AACCACUGAAACCAGGGUUC :::::: ::::::::::.:: Target 4242 UUGGUG-CUUUGGUCCUGAG	Cleavage

Table 5.1. *Solanum lycopersicum* miRNAtop14 predicted targets by psRNAtarget server. **Column 1)** Target number, given to identify each target in our RLM-RACE experiments. **Column 2)** Target accession and description, according to SGN transcript cDNA library version 2.4. **Column 3)** Expectation (Exp.), a score for miRNA-target complementarity, was set to a maximum of 3. **Column 4)** UPE, target accessibility as the maximum energy to unpair the target site, was set to a maximum of 25. **Column 5)** miRNA-target alignment gives the position of the first aligning nucleotide of the target and the last aligning nucleotide of the miRNA, considering the first position the 5' end in both cases. The sequence of the target is written from the 5' to the 3' end (from left to right) and the sequence of the miRNA in the opposite direction. **Column 6)** miRNA mode of action was predicted to be cleavage whenever there were not mismatches between target and miRNA nucleotides 9 and 11, and translational repression otherwise.

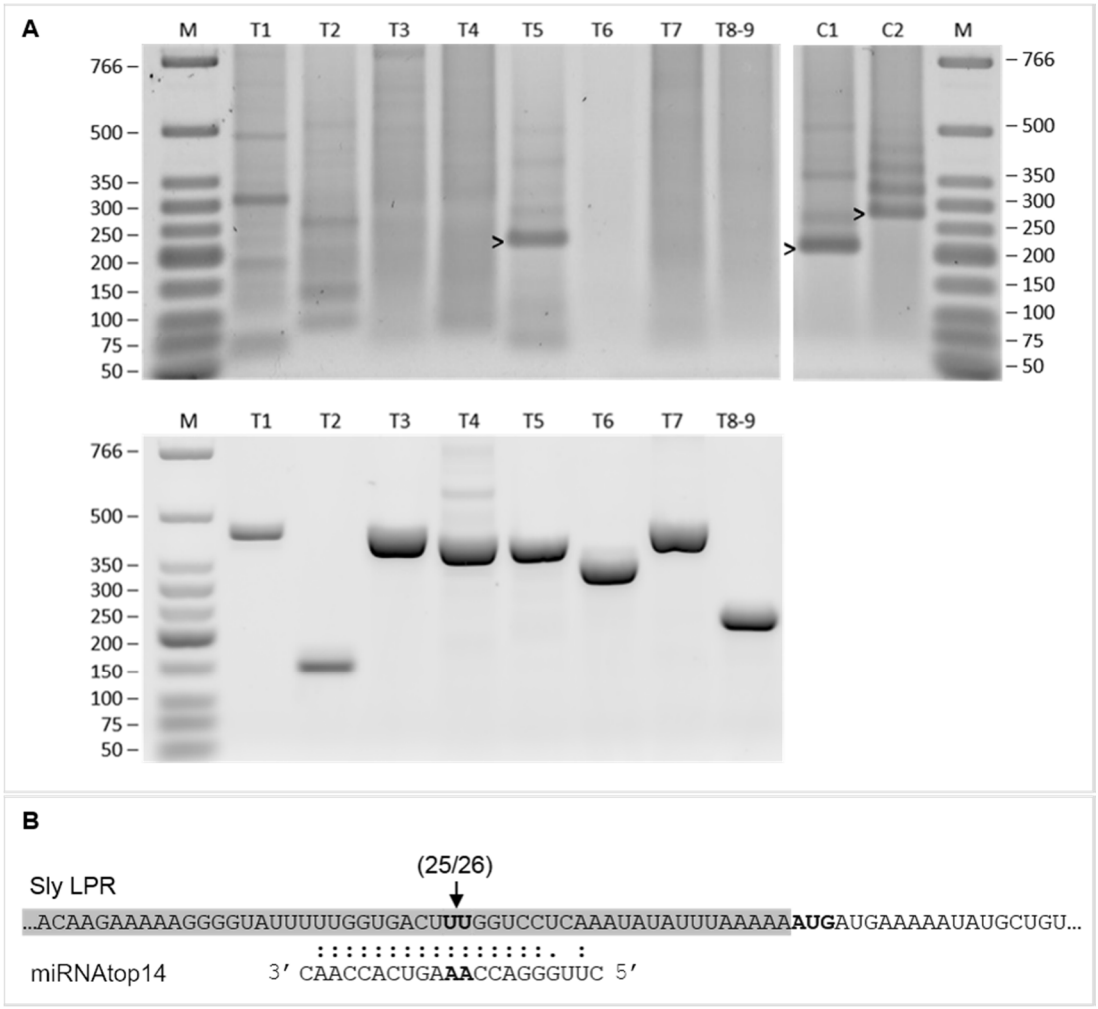


Figure 5.4. RLM-RACE analysis and miRNAtop14-*LPR* targeting in *Solanum lycopersicum.* **A)** RLM-RACE analysis of the 9 predicted targets predicted in tomato, compiled in table 5.1, each labelled with a target number (e. g. T1) according to the table 5.1. Target T8 and target T9 share the same sequence in the region of the putative miRNAtop14 cleavage, so they could not be independently analysed (they have the same primers) and therefore their common product is labelled as T8-9. **Top)** Nested PCR products from RLM-RACE run in an agarose gel. Two control targets are included: C1, LA cleaved by miRNA319 and C2, GRAS24 cleaved by miRNA171. Bands showing the expected length are marked by an arrowhead (T5, C1 and C2). Expected product length from each target in nucleotides: T1=77, T2=145, T3=114, T4=145, T5=197, T6=158, T7=203, T8-9=140, C1=209, C2=274. Note: T1 and T2 show bands that are proximate to the expected amplicon size, so these bands were recovered and cloned as well. T1 cloning failed while T2 band resulted to be the product of an unspecific amplification. **Bottom)** Control of target mRNA presence in the sample; PCR amplification across putative miRNAtop14 directed cleavage. All targets are present in the sample in non-cleaved form. Expected product length from each target in nucleotides: T1=437, T2=149, T3=434, T4=405, T5=378, T6=348, T7=449, T8-9=235. M, low molecular weight marker **B)** miRNA-T5 (*LPR*) target site interaction scheme and results of the cloning and sequencing of the RLM-RACE products. Shadowed in grey 5'UTR followed by the translation start codon in bold. 25 out of 26 clones showed the cleaved position indicated by the arrow, which correspond to the expected cleaved position between miRNA nucleotides 10th and 11th (in bold).

5.2.1.2. Degradome

In parallel to RLM-RACE analysis, a published *Solanum lycopersicum* degradome library[426] was searched for evidence of miRNAtop14 cleaved transcripts. This degradome library was composed of three libraries; one from fruit, one from flower and one from leaf tissue.

First, we looked for any transcripts complementary to miRNAtop14, allowing up to 4 mismatches and one gap, within the *Solanum lycopersicum* transcript cDNA library version 2.4, SGN[422]. Then, from the resulting list of transcripts, we discarded those with no degradome reads within 50 nucleotides upstream or downstream of miRNAtop14 complementary site, leaving only 26 transcripts left. Finally, we screened the 26 transcripts that remained and their associated degradome reads to identify anyone with reads in which the first nucleotide was complementary to miRNAtop14 10^{th} nucleotide, which would indicate cleavage directed by this miRNA.

There were six transcripts with reads that could come from a hypothetical miRNAtop14 directed cleavage. When combining the reads form the three libraries, in only 1 of the 6 transcripts these reads corresponded to a category 0 degradome peak, in other words, only one of these transcripts had its maximum number of reads coming from the miRNAtop14 cleaving site (see fig. 5.5). Suitably, this transcript was the same identified by RLM-RACE to be cleaved; *Solyc05g008290.2.1*, which codes for the protein LPR and is therefore confirmed to be a miRNAtop14 target.

Although all potential targets analysed by RLM-RACE were initially included in the degradome search due to its complementarity with miRNAtop14, apart from target 5 (LPR) only target 8/ target 9 had 2 reads from miRNAtop14 cleavage site in the tomato fruit library. These two transcripts share the same sequence for most of their length and miRNAtop14 complementary site is within an identical sequence region, so it was impossible to determine from which of the 2 transcripts were the 2 reads produced. In any case, there are other locations within these transcripts with a much higher number of reads than the two reads coming from the predicted miRNAtop14 cleaving site, which makes both unlikely miRNAtop14 targets.

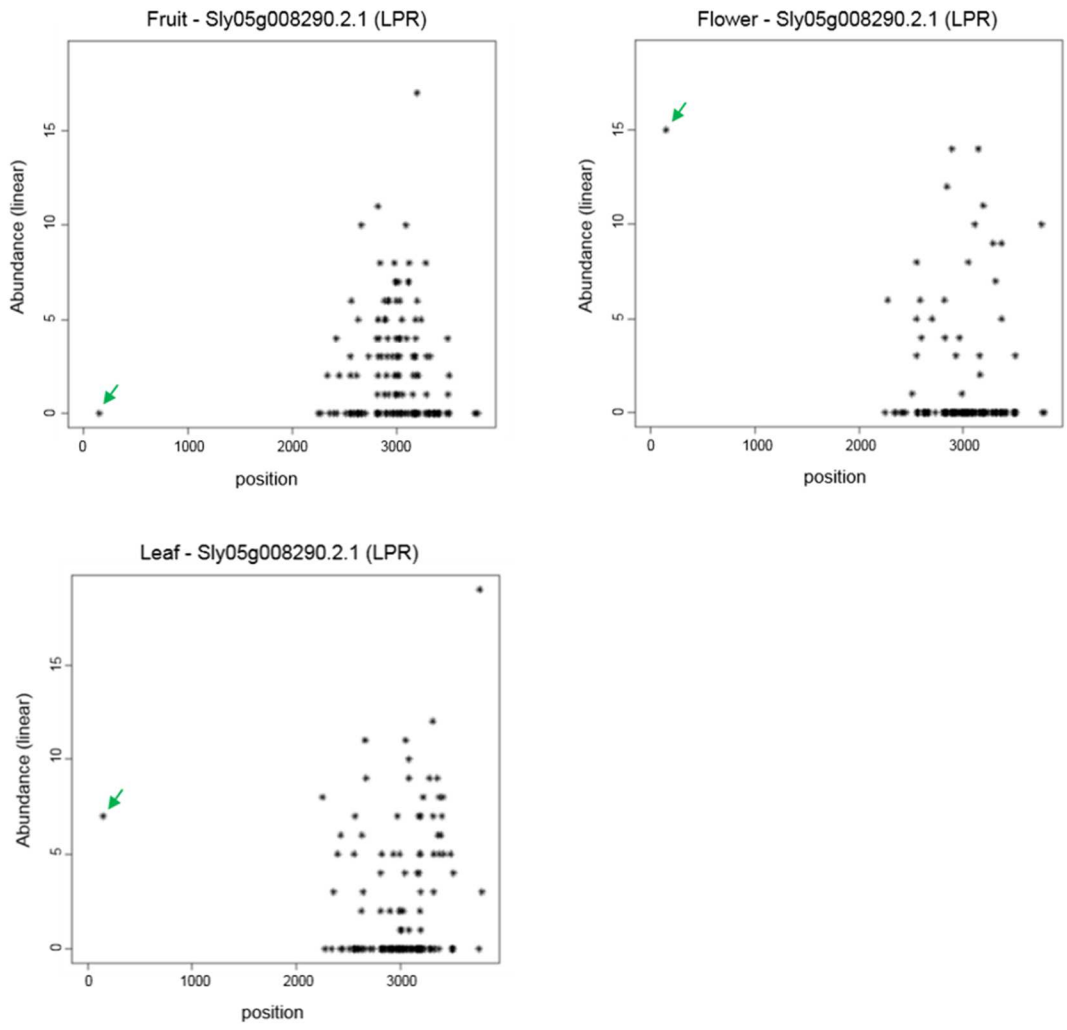


Figure 5.5. T-plots of *Solanum lycopersicum LPR* from Lopez-Gomollon et al.[426] degradome data, three tissues analysed. Reads coming from the putative miRNAtop14 cleavage are indicated by green arrows. As seen in the image, reads were detected from both flower (15) and leaf (7) tissues, but not from fruit. When the three libraries are combined miRNAtop14 cleavage position harbours the maximum number of reads, 22, from across the *LPR* mRNA, and becomes a category 0 degradome peak with the closest maximum at 19 reads.

5.2.2. Identification of miRNAtop14 target in *Nicotiana benthamiana*

5.2.2.1. RLM-RACE

Nicotiana benthamiana miRNAtop14 (nbe-miRNAtop14) targets were predicted through psRNAtarget server, version 2011[421]. This analysis yielded a total of 22 possible miRNAtop14 targets from *Nicotiana benthamiana* transcript library Niben101, SGN[422], using the same default parameters used previously for the similar analysis in tomato. Of these putative 22 nbe-miRNAtop14 targets, there are 12 with an expectation of 2.5 or lower (the default parameters consider a maximum expectation of 3). These 12 predicted targets are shown in table 5.2, together with expectation, target accessibility, miRNA-mRNA alignment and mode of target inhibition of each one.

Target accession, description	Exp.	UPE	miRNA-target alignment	mode of action
Niben101Scf02994g03003.1, Cupredoxin / copper oxidase (LPR2)	1	16.238	miRNA 20 AACCACUGAAACCAGGGUUU :::::::::::::::::..:: Target 187 UUGGUGACUUUGGUCCUGAA	Cleavage
Niben101Scf13776g01016.1, Cupredoxin / copper oxidase (LPR1)	2	22.836	miRNA 21 CAACCACUGAAACCAGGGUUU :::::::::::::: :::..:: Target 188 GUUGGUGACUUUGUUCCUGAA	Cleavage
Niben101Scf00180g08002.1, Polyphenol oxidase 1 (PPO1)	2.5	13.508	miRNA 21 CAACCACUGAAACCAGGGUUU : :::::::::: ::.::::: Target 1851 GGUGGUGACUUUUGUUCCAAA	Translation
Niben101Scf04384g02010.1, Polyphenol oxidase A1	2.5	19.048	miRNA 21 CAACCACUGAAACCAGGGUUU : :::::::::: ::.::::: Target 225 GAUGGUGACUUUUGUUCCAAA	Translation
Niben101Scf04384g02014.1, Polyphenol oxidase 3	2.5	17.71	miRNA 21 CAACCACUGAAACCAGGGUUU : :::::::::: ::.::::: Target 1840 GGUGGUGACUUUUGUUCCAAA	Translation
Niben101Scf03619g05002.1, Polyphenol oxidase 3	2.5	16.309	miRNA 20 AACCACUGAAACCAGGGUUU :::::::: ::::::::.:: Target 1752 GUGGUGACUCUGGUCCCGAA	Translation
Niben101Scf03733g00006.1, Protein phosphatase 2C	2.5	22.127	miRNA 20 AACCACUGAAACCAGGGUUU :::::::.::: :::.:::. Target 200 UUGGUGAUUUUUGUCUCAAG	Translation
Niben101Scf17372g00015.1, Protein phosphatase 2C	2.5	21.66	miRNA 20 AACCACUGAAACCAGGGUUU :::::::.::: :::.:::. Target 881 UUGGUGAUUUUUGUCUCAAG	Translation
Niben101Scf17372g00016.1, Protein phosphatase 2C	2.5	23.037	miRNA 20 AACCACUGAAACCAGGGUUU :::::::.::: :::.:::. Target 1234 UUGGUGAUUUUUGUCUCAAG	Translation
Niben101Scf05484g02007.1, Protein phosphatase 2C (PP2C)	2.5	22.092	miRNA 20 AACCACUGAAACCAGGGUUU :::::::.::: :::.:::. Target 1467 UUGGUGAUUUUUGUCUCAAG	Translation
Niben101Scf05241g02004.1, Protein phosphatase 2C	2.5	19.969	miRNA 20 AACCACUGAAACCAGGGUUU :::::::.::: :::.:::. Target 1589 UUGGUGAUUUUUGUCUCAAG	Translation
Niben101Scf05173g03012.1, Protein of unknown function	2.5	22.024	miRNA 20 AACCACUGAAACCAGGGUUU ::: :::.:: ::::::::: Target 2805 UUGCUGAUUUGGGUCCCAAA	Translation

Table 5.2. *Nicotiana benthamiana* miRNAtop14 predicted targets by psRNAtarget server. **Column 1)** Target accession and description, according to SGN transcript cDNA library Niben101. **Column 2)** Expectation (Exp.), a score for miRNA-target complementarity, was set to a maximum of 2.5. **Column 3)** UPE, target accessibility as the maximum energy to unpair the target site, was set to a maximum of 25. **Column 4)** miRNA-target alignment gives the position of the first aligning nucleotide of the target and the last aligning nucleotide of the miRNA, considering the first position the 5' end in both cases. The sequence of the target is written from the 5' to the 3' end (from left to right) and the sequence of the miRNA in the opposite direction. **Column 5)** miRNA mode of action was predicted to be cleavage whenever there were not mismatches between target and miRNA nucleotides 9 and 11, and translational repression otherwise. Transcripts later analysed by RLM-RACE are shadowed in grey.

As table 5.2 shows, there are two transcripts with higher expectation (complementarity with miRNAtop14) than the rest, which also are the only ones predicted to be cleaved. Both transcripts code for multicopper oxidases according to SGN, and when their sequences are used for "nucleotide collection" NCBI search, the closest homologous are LPR-like proteins, first from *Nicotiana* and then from other Solanales. In fact, when tomato *LPR* sequence is used for BLAST search against *N. benthamiana* transcript library Niben101, SGN, these two transcripts are the only ones aligning with it (~95% identity), which indicates that these are the only two *N. benthamiana LPR* genes. The alignment of each *N. benthamiana LPR* sequence with *Arabidopsis thaliana LPR1* and *LPR2* TAIR sequences shows that Niben101Scf02994g03003.1 is slightly more similar to *LPR2* (Score 74 *LPR2* and 68 *LPR1*) and Niben101Scf13776g01016.1 is slightly more similar to *LPR1* (Score 60 *LPR1* and 58 *LPR2*). Thereby, we decided to refer to *N. benthamiana* transcripts as *LPR1* or *LPR2* accordingly.

RLM-RACE was performed to test cleavage of *N. benthamiana LPR1* and *LPR2*. Together with these two putative targets, we included in the analysis Niben101Scf00180g08002.1. (Polyphenol oxidase 1, *PPO1*) and Niben101Scf05484g02007.1 (Protein phosphatase 2C, *PP2C*). Although these transcripts are only predicted to be translationally repressed in *N. benthamiana* (see table 5.2), they are homologous to *Solanum lycopersicum* target 1 and target 2, respectively, which are the only targets with a higher probability than *LPR* to be cleaved in tomato, according to psRNAtarget prediction (see table 5.1). Furthermore, *PPO1* transcript was identified by Baksa et al.[372] as a target of miRNAtop14 in *N. benthamiana* according to their degradome data. Given that the evidence for this cleavage was not very conclusive, since the T-plot in their publication indicates that this putative target corresponds to a category 2 (see fig. 5.3), we considered adequate to test it by RLM-RACE as well.

Total mRNA from the shoot system of three *Nicotiana benthamiana* plantlets was used for the analysis. *N. benthamiana* scarecrow-like protein 6 (SCL6) and TCP4 transcription factor transcripts were included in the RLM-RACE analysis as experimental controls. These two transcripts were chosen because they are homologous to tomato GRAS24 and LA, respectively, and both had been detected as category 0 targets in Baksa et al. *N. benthamiana* degradome[372]. Besides, it was checked that all four RLM-RACE target candidates were present in the total mRNA sample by RT-PCR amplification. RLM-RACE products of the four putative targets and controls were run in agarose gels and are shown in fig. 5.6.A

LPR1 and *LPR2* transcripts showed both a band of the correct amplicon length, while neither of the other two transcripts had visible amplification at the expected size. The bands from *LPR1* and *LPR2* RLM-RACE amplification were recovered from the gel and cloned each into pGEM-T easy vector. 20 clones from *LPR2* were sequenced and all of them corresponded to the precise cleavage of this transcript by miRNAtop14-RISC. However, when another 20 clones from the transformation with *LPR1* were sequenced, we realised that the reverse nested primer used for *LPR1* RLM-RACE was not specific for *LPR1*, but was able to amplify both *LPR1* and *LPR2* transcripts instead, which was not very surprising given the similarity between their sequences. As a result, we decided to analyse a higher number of clones in order to get at least 20 from *LPR1*, which finally made a total of 43 clones tested. Of these 43 clones, 20 corresponded to *LPR1* transcript, of which 19 showed the expected cleavage site. The remaining 23 clones came from *LPR2* and 22 of them indicated the expected miRNAtop14 directed cleavage. Overall, 19 clones out of 20 for *LPR1* and 42 clones out of 43 for *LPR2* were cleaved between the nucleotides complementary to the 10^{th} and 11^{th} nucleotide of miRNAtop14 (see fig. 5.6.B.). These results confirm that in *Nicotiana benthamiana*, similarly to *Solanum lycopersicum, LPR* transcripts are subjected to miRNAtop14 directed cleavage. However, the cleavage of *PPO1* by miRNAtop14 that was suggested by Baksa et al.[372] degradome data remains possible, although we were not able to verify it since *PPO1* mRNA was not present in our sample (see fig. 5.6.A).

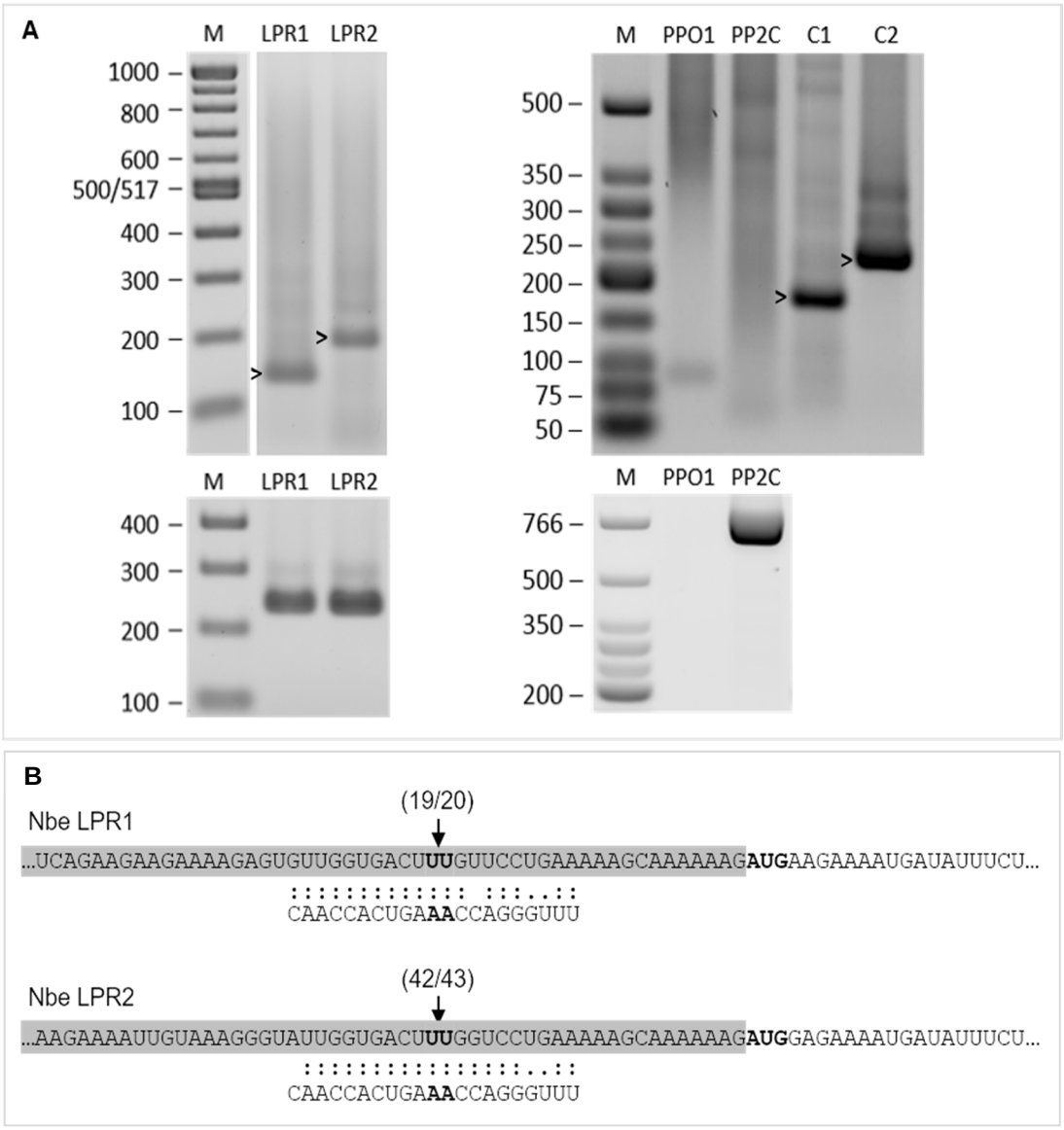


Figure 5.6. RLM-RACE analysis and miRNAtop14-*LPR* targeting in *Nicotiana benthamiana.* **A)** RLM-RACE analysis of cleavage by miRNAtop14 of its *LPR1*, *LPR2*, *PPO1* and *PP2C* predicted targets in *N. benthamiana*, shadowed in grey in table 5.2. **Top)** Nested PCR products from RLM-RACE run in an agarose gel. Two control targets are included: C1, TCP4 cleaved by miRNA319 and C2, SCL6 cleaved by miRNA171. Bands showing the expected length are marked by an arrowhead (LPR1, LPR2, C1 and C2). Expected product length from each target in nucleotides: LPR1=125, LPR2=179, PPO1=145, PP2C=236, C1=182, C2=237. **Bottom)** Control of target mRNA presence in the sample; PCR amplification across putative miRNAtop14 directed cleavage. All targets but PPO1 are present in the sample in non-cleaved form. Expected product length from each target in nucleotides: LPR1=232, LPR2=233, PPO1=563, PP2C=778. M, 100bp marker for LPR1 and LPR2 gels, low molecular weight marker for PPO1 and PP2C gels. **B)** miRNA-*LPR1* and *LPR2* target site interaction schemes and results of the cloning and sequencing of the RLM-RACE products. Shadowed in grey 5'UTR followed by the translation start codon in bold for both transcripts. 19 out of 20 clones in *LPR1* and 42 out of 43 clones in *LPR2* showed the cleaved position indicated by the arrows, which correspond to the expected cleaved position between miRNA nucleotides 10th and 11th (in bold).

5.2.2.2. Degradome

As already seen above, Baksa et al.[372] had produced a *Nicotiana benthamiana* degradome library in which they had identified the target of miRNAtop14 as the *PPO1* mRNA Niben101Scf00180g08002.1. However, the reads detected at this miRNAtop14 cleavage site constituted a category 2 degradome peak, indicating that there was a higher number of reads from several other positions in the transcript (fig. 5.3). We consequently tried to confirm this target by RLM-RACE, but failed to do so because *PPO1* transcript was not present in our sample (see fig. 5.6.A).

In contrast, we identified *LPR1* and *LPR2* as miRNAtop14 targets by RLM-RACE. Thereby, we decided to check the data from Baksa et al.[372] PARE library for these two transcripts to determine whether the cleavage we had detected by RLM-RACE could be confirmed analysing their degradome data.

Baksa et al.[372] degradome cDNA library is constituted of 10 separate libraries which include leaf, stem, root, flower and seedling tissue, two replicas of each. We decided to search for our 2 targets of interest keeping the data of each of these libraries independent. Surprisingly, we could only find a very low number of reads coming from miRNAtop14 cleaving position in one of the replicas of the root library, and only for *LPR1* transcript (see t-plot in fig 5.7). This may indicate a difference in the conditions in which the plants were grown for our RLM-RACE experiment compared with the conditions at which they were grown to make these PARE libraries. Otherwise, it could just indicate that the two experimental methods, although similar in their basis, may have enough differences to favour the recovery of some transcript fragments over others in each case, which would also explain the incongruous result.

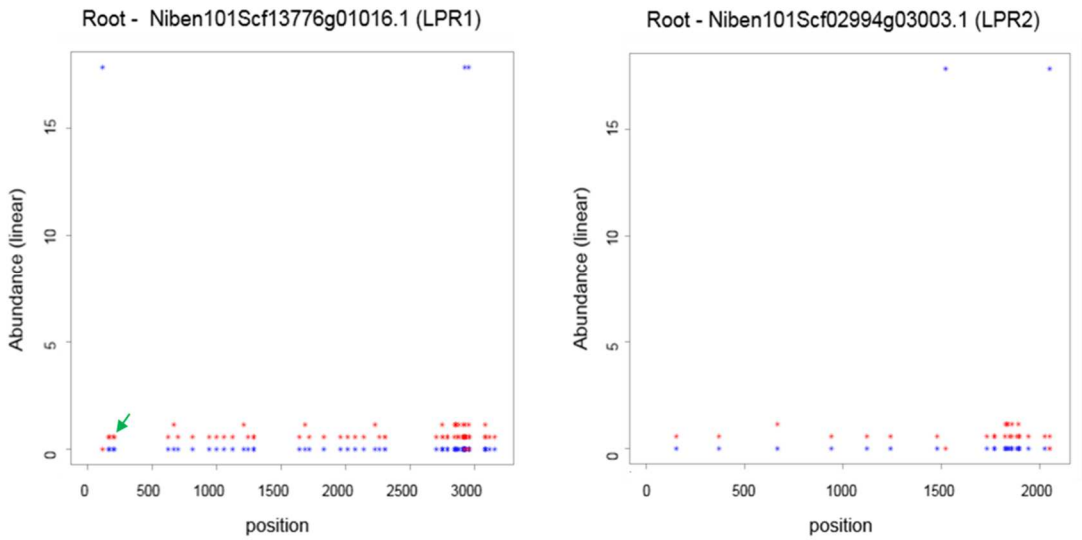


Figure 5.7. T-plots of *LPR1* and *LPR2* from *Nicotiana benthamiana* root degradome library data from Baksa et al.[372], two replicas. Root is the only tissue out of the five analysed in which reads from putative miRNAtop14 directed cleavage could be detected. Reads coming from each replica are indicated as red or blue stars. The peak of reads produced from the potential miRNAtop14-RISC cleavage is indicated by a green arrow. As seen in the image, there are only reads from *LPR1* transcript and only in one of the replicas, and the number is 0.57 after normalisation.

5.2.3. miRNAtop14 targeting of *LPR* in other Solanales species

5.2.3.1. miRNAtop14-*LPR* complementarity in Solanales

After the previous analysis in which *LPR* was detected to be a target of miRNAtop14 in both tomato and *Nicotiana benthamiana*, we decided to look for *LPR* transcripts in the rest of Solanales in which miRNAtop14 had already been identified (see table 3.1). We searched various NCBI and SGN sequencing collections and found at least one *LPR* variant in all Solanales species quested, sometimes two, with the exception of *Ipomoea batatas* in which no *LPR* sequence was found, which could possibly be explained by the limited amount of sequenced genome available. See results of this search in table 5.3.

Subsequently, we tested whether miRNAtop14 was predicted to target *LPR* (one or two variants) in each one of the species studied through psRNAtarget. The prediction is made based on miRNA-mRNA complementarity and target accessibility, and the mode of action of the miRNA (cleavage or translational repression) is

suggested based on the presence of any mismatches in the central positions of the miRNA-target alignments. The results of this analysis are compiled in table 5.3.

Species	LPR paralog	Exp.	UPE	miRNA-target alignment	mode of action
Solanum lycopersicum	LPR	3	18.982	miRNA 20 AACCACUGAAACCAGGGUUC :::::::::::::::::. : Target 136 UUGGUGACUUUGGUCCUCAA	Cleavage
Solanum pimpinellifolium	LPR, not 5' UTR	5' UTR not available, targeting prediction could not be performed			
Solanum arcanum	LPR	3	17.743	miRNA 20 AACCACUGAAACCAGGGUUC :::::::::::::::::. : Target 63 UUGGUGACUUUGGUCCUCAA	Cleavage
Solanum habrochaites	LPR	3	19.125	miRNA 20 AACCACUGAAACCAGGGUUC :::::::::::::::::. : Target 139 UUGGUGACUUUGGUCCUCAA	Cleavage
Solanum pennellii	LPR	3	17.958	miRNA 20 AACCACUGAAACCAGGGUUC :::::::::::::::::. : Target 53 UUGGUGACUUUGGUCCUCAA	Cleavage
Solanum commersonii	LPR	3	13.307	miRNA 20 AACCACUGAAACCAGGGUUC :::::::::::::::::. : Target 43 UUGGUGACUUUGGUCCUCAA	Cleavage
Solanum tuberosum	LPR	3	24.483	miRNA 20 AACCACUGAAACCAGGGUUC :::::::::::::::::. : Target 212 UUGGUGACUUUGGUCCUCAA	Cleavage
Solanum melongena	LPR	0.5	19.119	miRNA 20 AACCACUGAAACCAGGGUUU ::::::::::::::::::.:: Target 41 UUGGUGACUUUGGUCCUAAA	Cleavage
Capsicum annuum	LPR	3.5	14.038	miRNA 20 AACCACUGAAACCAGGGUUC :::::::::::::::.:. : Target 102 UUGGUGACUUUGGUUCUCAA	Cleavage
Nicotiana tabacum	LPR1	1	11.75	miRNA 21 CAACCACUGAAACCAGGGUUU ::::::::::::::::::..:: Target 29 GUUGGUGACUUUGGUCCUGAA	Cleavage
	LPR2	1	18.932	miRNA 21 CAACCACUGAAACCAGGGUUU ::::::::::::::::::..:: Target 116 GUUGGUGACUUUGGUCCUGAA	Cleavage
Nicotiana sylvestris	LPR	1	20.326	miRNA 21 CAACCACUGAAACCAGGGUUU ::::::::::::::::::..:: Target 910 GUUGGUGACUUUGGUCCUGAA	Cleavage
Nicotiana benthamiana	LPR1	2	22.836	miRNA 21 CAACCACUGAAACCAGGGUUU :::::::::::::: :::..:: Target 188 GUUGGUGACUUUGUUCCUGAA	Cleavage
	LPR2	1	16.238	miRNA 20 AACCACUGAAACCAGGGUUU ::::::::::::::::::..:: Target 187 UUGGUGACUUUGGUCCUGAA	Cleavage
Nicotiana attenuata	LPR	No targeting predicted			
Nicotiana otophora	LPR	3.5	18.719	miRNA 21 CAACCACUGAAACCAAGGUUU :::::::::::: :: :..:: Target 136 GUUGGUGACUUUAGUCCUGAA	Translation
Nicotiana tomentosiformis	LPR	2.5	18.932	miRNA 21 CAACCACUGAAACCAAGGUUU ::::::::::::::: :..:: Target 123 GUUGGUGACUUUGGUCCUGAA	Cleavage
Petunia axillaris	LPR	No targeting predicted			
Petunia integrifolia	LPR1	No targeting predicted			
	LPR2	No targeting predicted			
Ipomoea batatas	LPR not available	LPR sequence not available, targeting prediction could not be performed			
Ipomoea trifida	LPR1	No targeting predicted			
	LPR2	No targeting predicted			
Ipomoea nil	LPR1	No targeting predicted			
	LPR2	No targeting predicted			

Table 5.3. Analysis of miRNAtop14-*LPR* complementarity in Solanales in which miRNAtop14 has been detected using psRNAtarget. **Column 1)** Solanales species. **Column 2)** Number of *LPR* paralogous found in the species analysed. **Column 3)** Expectation (Exp.), a score for miRNA-target complementarity, was set to a maximum of 5, less possible restrictive value. **Column 4)** UPE, target accessibility as the maximum energy to unpair the target site, was set to a maximum of 100, less possible restrictive value. **Column 5)** miRNA-target alignment gives the position of the first aligning nucleotide of the target and the last aligning nucleotide of the miRNA, considering the first position the 5' end in both cases. The sequence of the target is written from the 5' to the 3' end (from left to right) and the sequence of the miRNA in the opposite direction. **Column 6)** miRNA mode of action was predicted to be cleavage whenever there were not mismatches between target and miRNA nucleotides 9 and 11, and translational repression otherwise. **Note,** in two species the analysis could not be carried out: in *Solanum pimpinellifolium*, in which *LPR* mRNA sequence was lacking the 5'UTR region, and in *Ipomoea batatas,* where not *LPR* sequence could be found.

As can be seen in table 5.3., in all *Solanum* species, in *Capsicum annum* and in all *Nicotiana* species but one (*Nicotiana attenuata*) miRNAtop14 is predicted to target *LPR*. Interestingly, the target site is located in the 5'UTR of the transcript in all of them, 14 nucleotides upstream of the transcription start site in *Solanum* (apart than in *Solanum melongena* where it is only 3 nucleotides upstream), 15 nucleotides upstream in *Capsicum annuum* and 13 nucleotides upstream in both *Petunia* and *Nicotiana* (with the exception of *Nicotiana otophora*, where it is only 11 nucleotides upstream of the transcription start site). Finally, the mode of action of the miRNA when targeting is predicted to be cleavage in all cases, with the only exception of *Nicotiana otophora* where it is translational repression.

On the other hand, in both *Petunia* species and the two *Ipomoea* species that could be analysed, miRNAtop14 is not predicted to target *LPR* due to the low complementarity between both. This complementarity is virtually non-existent for *Ipomoea*, where no putative target site could be identified at all. However, in *Petunia*, a site complementary to miRNAtop14 can be identified 12 nucleotides upstream of the transcription start site (see fig. 5.8.B), a location similar to where miRNAtop14 targets all the rest of Solanaceae transcripts. Although the pairing nucleotides are not enough to determine miRNA targeting in plants according to the canonical rules[142,144], alternative rules for some specific cases like 5'UTR targeting have been suggested[158], so we decided to test *Petunia axillaris LPR* for miRNAtop14 directed cleavage by RLM-RACE.

5.2.3.2. miRNAtop14 targeting of *LPR* by cleavage in *Petunia axillaris*

Petunia is, of the four Solanaceae genera studied, the first one to have diverged, and thereby the other three genera (*Solanum, Capsicum* and *Nicotiana*) are more closely related among them. In agreement with this, *Petunia* is also the only one of the four genera in which miRNAtop14 is not predicted to target *LPR*, apart from the exception of *Nicotiana attenuata* species(see table 5.3). In order to assess experimentally whether this prediction is true, we proceeded to check miRNAtop14 directed cleavage of *LPR* transcript in *Petunia* by RLM-RACE.

The analysis was performed using total mRNA from the aerial part of three *Petunia axillaris* plantlets. Different primers and PCR conditions were tried for both RLM-RACE PCR and nested PCR amplifications. The homologous of tomato GRAS24 and LA transcripts (GRAS family and TCP4 transcription factors, respectively) were included in the analysis as experimental controls, as had been done for tomato and *N. benthamiana*. Although the cleavage of these transcripts had not been tested in *Petunia* to our knowledge, we expected it to be conserved in this species since it was confirmed to happen in species much more distant to *Solanum* and *Nicotiana* than *Petunia*, as is the case of *Arabidopsis*[150,232]. Finally, the presence of the putative target mRNA in the sample was also confirmed by RT-PCR. The agarose gel with the results of the RLM-RACE analysis for the target and the controls, including some of the different PCR conditions tried, is shown in fig. 5.8.A.

As can be appreciated in the figure, there is no band of the expected size in any of *the P. axillaris* samples in which the cleavage of the *LPR* target was tested. However, the cleaved fragments of both control transcripts are detected in the very same *Petunia* samples, as well as the non-cleaved *LPR* and miRNAtop14. Consequently, we concluded that miRNAtop14 does not direct cleavage of *LPR* in *Petunia axillaris*, at least in the conditions at which the plants tested were grown. This result matches psRNAtarget analysis, in which no cleavage of *LPR* was predicted by *Petunia* miRNAtop14. However, the possibility of translation inhibition could not be discarded, especially given that the target site is located in the 5'UTR of the transcript[158].

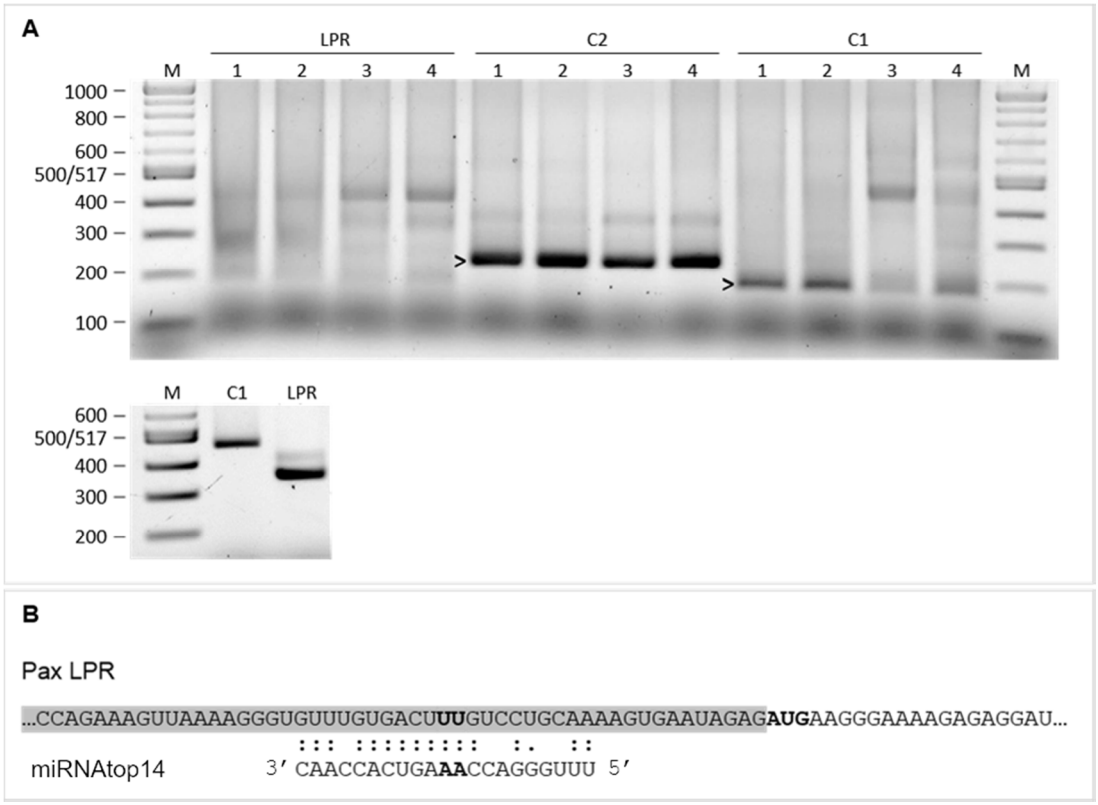


Figure 5.8. RLM-RACE analysis and miRNAtop14-*LPR* putative targeting in *Petunia axillaris.* **A)** RLM-RACE analysis of cleavage by miRNAtop14 of its *LPR* predicted target in *Petunia axillaris*. **Top)** Nested PCR products from RLM-RACE run in an agarose gel. Two control targets are included: C1, TCP4 cleaved by miRNA319 and C2, GRAS cleaved by miRNA171. Bands showing the expected length are marked by an arrowhead (C1 and C2). Expected product length from each target in nucleotides: LPR=244, C1=174, C2=223. Numbers 1, 2, 3, 4 on top of the wells correspond to different conditions in the first RLM-RACE PCR on the Generacer primer. In a 30µl reaction volume: 1=2.7µl Generacer primer, 2=0.9µl Generacer primer, 3=2.7µl Generacer nested primer, 4=0.9µl Generacer nested primer. **Bottom)** Control of target mRNA presence in the sample; PCR amplification across putative miRNAtop14 directed cleavage. Both *LPR* and C1 are present in the sample in non-cleaved form. Expected product length from each target in nucleotides: LPR=344, C1=451. M, 100bp marker. **B)** miRNA-*LPR* target site interaction scheme. Shadowed in grey 5'UTR followed by the translation start codon in bold. No miRNAtop14 cleavage was detected by RLM-RACE (see gel on the top), so not cloning and sequencing analysis was carried out.

5.3. Discussion

5.3.1. Identification of miRNAtop14 target

RLM-RACE[150] has been the method of preference to validate targets of miRNAs in plants given that, unlike in animals, it is thought that most plant miRNAs direct cleavage of their targets rather than supressing their translation[215]. PARE method is based on the same principles as RLM-RACE, the detection of cleaved transcripts through their characteristic 5' P, but in PARE it is done in a high-throughput scale[148,363].

In *S. lycopersicum*, both RLM-RACE and PARE analysis results nicely agree as expected, both showing clear *LPR* directed cleavage. However, it is striking that in *N. benthamiana* there is a difference between RLM-RACE and PARE results given that both methods are based on the same principle. While RLM-RACE shows a clear miRNAtop14 directed cleavage of both *LPR1* and *LPR2*, PARE only detected very low level of cleaved *LPR1* in one out of ten libraries.

This can be a result of slight differences in the methodologies. For instance, degradome libraries could be biased towards certain sequences as it has been observed in sRNA libraries[365]. Besides, PARE may be less sensitive towards transcripts that are in lower amount among the whole group of cleaved transcripts that is being searched by this method, while in RLM-RACE each transcript is tested individually, so the absence of competition with the whole population of cleaved transcripts can make this method more sensitive to detect less abundant cleaved mRNAs.

Another reason for such discordant results could be biological, that is, that the samples analysed by one and the other method had indeed very different levels of cleaved *LPR*. On one hand, this would be in agreement with the high variability observed in the levels of miRNAtop14 in tomato (see figure 3.6), which could suggest miRNAtop14 expression to be inducible; in this case the differences in cleaved *LPR* could be explained by differences in the growth conditions between the two samples. On the other hand, the fact that miRNAtop14 is ubiquitously detected in all plant samples that we have tested make us think that there must be some level of *MIRtop14* expression stimulus-independent, and thereby the very low level of *LPR* cleavage detected in the *N. benthamiana* PARE libraries could have other reasons.

The incongruity between the high levels of miRNAtop14 (see figure 3.5.A) and of non-cleaved *LPR* (see figure 5.4.A and 5.6.A) in the same *S. lycopersicum* and *N.*

benthamiana samples may be explained by different factors. For instance, miRNAtop14 could be repressing translation at the same time as directing cleavage of the *LPR* population, a phenomenon known to happen with other plant miRNAs[145,147]. Besides, it may be that *LPR* is not accessible to miRNAtop14 in some instances; they may be present in different cells or tissues, what is called mutual exclusion[214,427], or *LPR* target site may be protected by mRNA secondary structures or binding proteins[428]. Finally, miRNAtop14 could have more targets not identified in this study that could be competing with *LPR*.

In fact, we were surprised by the observation that, in tomato, *LPR* was detected to be cleaved by miRNAtop14 by both RLM-RACE and degradome while *PPO*, another miRNAtop14 predicted target with higher complementarity, was not detected by either. However, in *N. benthamiana* both *LPR1* and *LPR2* have higher complementarity to miRNAtop14 than *PPO*, but *PPO* is the one that has been proposed to be targeted by Baksa et al.[372] based on degradome data (although in our experiment both *LPRs* are detected by RLM-RACE and not *PPO*). This indicates that, as has already been reported, miRNA-transcript complementarity is not the only parameter that determines miRNA-RISC targeting, and other parameters such as the discussed above (target site accessibility, different patterns of expression, etc.) can influence miRNA-target interaction[428] and even possibly change it depending on the conditions.

5.3.2. miRNA-target coevolution and interaction

It was expected that *MIRtop14* was playing a biological role based on the high conservation of its mature miRNA sequence and the maintenance of the miRNA-miRNA* stem-loop across species, as well as from the high levels of mature miRNA being detected by the Northern analyses. However, proving that miRNAtop14 directs the cleavage of a protein-coding mRNA is the conclusive evidence.

The importance of miRNAtop14 directed LPR regulation is evident in both *Solanum* and *Nicotiana* genera, where the miRNA-*LPR* target site shows near-perfect complementarity in all seven and in five out of six species studied, respectively (see table 5.3). Besides, in *Capsicum* genus, where only *Capsicum annum* species was analysed, miRNAtop14 also showed enough complementarity to *LPR* target site for cleavage to be performed.

In addition to the conservation of miRNAtop14 target site in *LPRs* of different species, miRNAtop14 target site is also conserved among paralogous *LPR* variants present in the same species. Such is the case in *Nicotiana benthamiana* and in *Nicotiana tabacum*, which harbour two *LPR* paralogs each, both predicted to be cleaved by miRNAtop14 in both species.

These predicted cleavages were later confirmed by RLM-RACE in *Solanum lycopersicum* only *LPR* and in *Nicotiana benthamina LPR1* and *LPR2* variants.

In *Petunia* genus, however, *LPRs* were not predicted to be miRNAtop14 targets in any of the two species studied. Nevertheless, partial complementarity including the 10th and 11th miRNA nucleotides is observed between miRNAtop14 and *LPR* at the same position where the target site is present in *Nicotiana, Capsicum* and *Solanum* species. Still, as predicted by the in-silico analysis, when performing RLM-RACE in *Petunia axillaris* no miRNAtop14 directed cleavage was detected. Although it is difficult to reach a conclusive proof based only in negative data like in this case, all evidences seem to indicate that *Petunia LPRs* are indeed not cleaved by miRNAtop14.

It is nevertheless interesting to note that miRNAtop14 is targeting *LPR* 5'UTR rather than its coding region. While plant miRNAs require almost full complementarity with their targets for either cleavage or repression of translation to take place[158,234], in the specific case in which the target site is within the 5'UTR, partial complementarity between miRNA and target is sufficient to cause repression of translation[158]. This would open a possibility for *Petunia LPRs* to be regulated by miRNAtop14 through translational repression. Furthermore, since non-cleaved *LPR* transcripts have been detected in both *S. lycopersicum* and *N. benthamiana* RLM-RACE samples, it cannot be dismissed that miRNAtop14 may regulate *LPR* by both cleavage and translational repression, as it is known to occur in other plant miRNA-target pairs[145,157].

If we wanted to test whether translational inhibition is taking place, it would be necessary to measure protein levels of LPR and compare them with *LPR* mRNA levels and miRNAtop14 levels. Alternatively, we could take advantage of the fact that *LPR* target site is in its 5'UTR region to check the possibility of translational inhibition in *Petunia*. We would just need to create two constructs with two transcriptional units each: one construct with *Pax-MIRtop14* and *Pax-LPR* 5'UTR fused to a reporter gene such as GUS and another construct identical but with *Pax-LPR* 5'UTR mutated at the miRNAtop14 predicted target site. Subsequently, we could transform any plant

species lacking miRNAtop14 such as *Arabidopsis* with these constructs, and measure the differences in the reporter protein expression between both.

In *Ipomoea* genus, where *Ipomea nil* species was studied more in depth, not even a low level of complementarity could be found between miRNAtop14 and any of its two *LPR* variants. miRNAtop14 does not seem to be regulating *LPR* in this genus.

It, however, remains intriguing that, despite of the lack of any known targets in *Ipomoea* and possibly in *Petunia* genera, miRNAtop14 sequence is equally conserved in these species, so is the interaction between miRNA and miRNA*. One possible explanation is that miRNAtop14 has another target which is shared by all species, and that *LPR* is a later, additionally gained target. In support of this possibility, it has been observed in *Arabidopsis* as well as in animals that older miRNAs have a greater number of targets than young ones[218,408]. To explain this observation, it has been hypothesized that, as miRNA with targets tend to be conserved, they also tend to be expressed at higher levels or broader patterns and this would increase the chances of pairing with other mRNAs that could eventually be fixed as new targets if the interaction is beneficial for the plant[408].

On the other hand, since we could not find homology between *MIRtop14* and any other protein coding gene, it is also plausible that *MIRtop14* raised randomly from a hairpin forming transcript[216] and existed initially as a miRNA without target (as it is predicted for most young miRNAs[217,230,237]) until eventually captured *LPR* as a target. However, in this case, miRNAtop14 is expected to be evolving neutrally on those clades where it has no target, and therefore it would show divergence with the conserved miRNAtop14 sequence or would have already been lost, as tend to happen with young miRNAs without function[209,217,230,237].

It would be difficult to conclude at which point miRNAtop14 started regulating *LPR*. If *Petunia LPR* was found to be regulated by miRNAtop14, it would indicate that the interaction between miRNAtop14 and *LPR* started before the divergence of this genus. Otherwise, miRNAtop14-*LPR* interaction may have started after *Petunia* divergence, or their interaction may have been lost in this genus. Similarly, although *Ipomoea* and miRNAtop14 are not currently interacting, it cannot be dismissed that such interaction was lost in this genus too. In fact, it cannot even be discarded that miRNAtop14 and *LPR* started interacting at the point when this miRNA first arose.

It could be hypothesised that miRNAtop14-*LPR* initial interaction could have been facilitated by the position of the target site at the 5'UTR, which would have allowed miRNAtop14 to decrease *LPR* translation despite a relatively low

complementarity between miRNA and target[158]. Later on, once such regulation of *LPR* resulted beneficial, higher levels of complementarity between miRNAtop14 and *LPR* would have been achieved through miRNA-target site coevolution, leading to directed cleavage once the complementarity between both was sufficient. In support of this theory, it has been suggested that translational inhibition could be the preferred mode of action of non-conserved miRNAs, which would explain the general lack of targets from young miRNAs validated at the mRNA level[145].

5.3.3. miRNAtop14 function in Solanales

Based on the activity of LPR, a multi-copper ferroxidase[414], and on its known function in *Arabidopsis*[413,414,418,420], it could be predicted that miRNAtop14 plays a role in phosphate and/or iron homeostasis. However, it should be a function specific of Solanaceae, since despite of the ubiquitous presence of LPR in plants[429], miRNAtop14 only regulates LPR in this specific family.

Another important aspect to note is that LPR's known function in *Arabidopsis* takes place in the root [413], while according to our northern analysis, mature miRNAtop14 shows lower levels in root, compared to shoot and leaves (see figure 3.6). Nevertheless, LPR is expressed in *Arabidopsis* leaves[413], and an analogous function linking phosphate/iron homeostasis with plant development is plausible to take place in either shoot or leaves.

For instance, ferroxidase activity by multicopper oxidases has already been detected in vascular plants, where it has been suggested to be involved in tissue lignification given their high expression in lignifying vascular tissues[430]. Furthermore, LPR ferroxidases themselves have been suggested to play a role in lignin polymerization during phosphate starvation, since it was observed that phosphate deprivation led to increased lignification in WT *Arabidopsis* root tips, but not in *lpr1, lpr2* double mutants[431]. Therefore, LPR could be playing a role assisting in lignification. Such a role would fit the patterns of miRNAtop14 expression observed by the Northern analyses. First, lignification occurs in specific cells and tissues, mainly associated to the vascular system, such as xylem[432]. Accordingly, stem shows the highest levels of miRNAtop14 accumulation (see figure 3.6.A). Second, biosynthesis of lignin is developmentally activated as the vascular system is developed[432], which could explain the gradient in miRNAtop14 amount observed from tip to bottom of individual leaves (see figure 3.6.B). Finally, there is strong evidence demonstrating that lignin is produced and quickly deposited in cell walls in response to both biotic

and abiotic stresses, even in cells that do not normally accumulate lignin, where it acts as barrier to pathogens and/or by limiting further growth[433]. This could account for the high variations on the levels of miRNAtop14 found when replicating experiments, sometimes showing increased accumulation of the miRNA, suggesting a stimulus-induced way of expression (data not shown).

In any case, further research is needed to determine LPR function, as well as which advantage provides the regulation of *LPR* by miRNAtop14, in Solanaceae. Furthermore, it would also be interesting to investigate whether this miRNA has any additional function involving other targets, as could be suggested by its high conservation in species where it is not targeting *LPR*.

Finding out the pattern of expression of *MIRtop14* and possibly of LPR would be one of the first ways to obtain information about this miRNA function. To achieve this, *MIRtop14* harbouring plants could be transformed with constructs in which a reporter gene was placed under *MIRtop14* or *LPR* promoter. Besides, these transgenic plants could be used to test changes in *MIRtop14* or *LPR* expression under certain conditions or stresses. Based on our knowledge of LPR function in *Arabidopsis*, a first step could be assessing whether there are any changes in *MIRtop14* or *LPR* expression in response to a low phosphate/ iron ratio in the soil. Additionally, we could also test factors known to produce stress-induced lignification, such as drought[434], pathogen infection[433], or wounding, the last one known to occur in tomato[435].

To deepen the knowledge on how miRNAtop14 regulates LPR in vivo, the constructs with *LPR* promoter followed by a reporter gene could easily include miRNAtop14 target site within the promoter region, taking advantage of miRNAtop14 target site location at *LPR* 5'UTR. Subsequently, we could create constructs in which the miRNAtop14 target site had been mutated and compare both patterns of *LPR* expression in the plant, with and without miRNAtop14 regulation.

Creating *MIRtop14* and *LPR* knock-out and overexpression mutants could be another strategy to figure out their function. This approach, especially *MIRtop14* mutation, has the risk of obtaining a weak phenotype that would not be very informative of miRNAtop14 function, especially if miRNAtop14's main function responds to a specific stimulus which may not be present. However, given the broad expression of *MIRtop14*, it would be expected to see some phenotypic consequences after its mutation.

Finally, it would be interesting to research the possibility of miRNAtop14 regulating another target apart from *LPR*. In fact, PPO proteins, one of which is predicted to be *N. benthamiana* target according to Baksa et al. degradome[372] data, are also good candidates to be regulated by miRNAtop14 in tomato according to psRNAtarget prediction. One possibility to test the regulation of other putative targets would be to perform RLM-RACE in samples growth under different conditions to try to detect regulation directed under specific stimulus. Another possibility would be to test whether any of the remaining predicted targets are being regulated by translational repression rather than by cleavage. To address this, we would need to measure the protein levels of the target relating them with the levels of its mRNA and miRNAtop14 in the sample, for example, between wild-type and *MIRtop14* knock-out mutants.

Chapter 6

Summary and general discussion

After the discovery that miRNAs are ubiquitous molecules in animals and plants at the beginning of the century, the last 15 years have witnessed an explosion in both the finding of new miRNA species and the broadening of our knowledge in miRNA biology. While great numbers of “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 unveil 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[370,372].

Thanks to the growing number of plant reference genomes available, besides other sequence collections such as EST, 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.

Apart 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 findings indicate that miRNAtop14 is a Solanales specific miRNA, which makes this miRNA an interesting candidate to be further studied if only because the economic importance of this order, as such research may unveil some distinct characteristic of Solanales that could be utilised 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. Surprisingly, no cleavage has been detected in *Petunia axillaris*, despite of *Petunia LPR* mRNA and miRNAtop14 showing partial complementarity, and no complementarity at all has been found between *LPR* mRNA and miRNAtop14 in *Ipomoea*.

In *Arabidopsis*, LPR proteins are involved in the pathway that arrests root growth under low phosphate conditions[413,414]. However, despite its wide distribution in the plant kingdom[429], 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[413].

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.

From comparing the results of the study of miRNAtop14 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 not target site to be found at all in *Ipomoea* LPR.

Generally, miRNA and target sites are supposed to coevolve in order to maintain its interaction and thus the miRNA function[243–246]. 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[217,230,237].This raises the question of why does miRNAtop14 maintain 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 this 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 discarded the possibility of miRNAtop14 directing translational repression

of mRNA without cleavage, a kind of interaction which RLM-RACE would not reveal 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.

Finally, the study of *MIRtop14* peculiar intron-exon structure has been the last main focus of our study. The *MIRtop14* phylogenetic conservation analysis have shown that the presence of an intron in between miRNA and miRNA* is conserved across *Solanum*, *Capsicum* and *Nicotiana* species. However, the whole intron has been experimentally confirmed to be absent in *Petunia*, and in *Ipomoea* it is predicted to engulf miRNAtop14 sequence itself, although experimental corroboration remains to be carried out. Besides, the pri-miRNAtop14 search in EST libraries has revealed that *MIRtop14* produces two pri-miRNA variants; one retaining the intron and the other one with the intron spliced. This finding has been experimentally confirmed in *S. lycopersicum* and *N. benthamiana* by RT-PCR of polyadenylated transcripts.

pri-miRNAtop14 intron-split arrangement has only been previously observed in the *MIR444* family, in the so-called nat-miRNAs[195]. This publication assumes a connection between the characteristic intron-exon structure and the *cis*-antisense nature of the newly found miRNAs, even so two of the *MIR444* family members do not have an antisense target. Such association agrees with their hypothesis that the presence of the intron is important for the formation of these *MIRs* because it prevents the annealing between the pri-miRNA and its antisense target[195]. Accordingly, they suggest further searches for this kind of MIRs, since more miRNAs of this type may have been easily missed given their peculiar characteristics.

On one hand, our research confirms the need to look for other *MIR* that may have been missed because of the presence of an intron in their stem-loop. On the other hand, pri-miRNAtop14 transcript is not antisense to any known mRNA, which indicates that *MIRs* with this intron-exon structure but not targeting antisense transcripts may exists as well, in which case they would be better defined as intron-split miRNA than as nat-miRNA[369].

While the biological reason behind the apparent requirement of an intron to be present in between miRNA and miRNA* may be to restrict base-pairing between pri-miRNA and its mRNA targets in nat-miRNAs[195], the reason for the conservation of this structure in the rest of intron-split miRNAs remains to be unravelled since the above explanation would not apply. Furthermore, in the case of *MIRtop14*, the intron-

retaining pri-miRNA variant would not impede the annealing with any hypothetical anti-sense transcript anyway.

An alternative possibility, taking into account the finding of the two pri-miRNA variants being produced from *MIRtop14*, is that the intron may be enabling the differential production of miRNAtop14 under different circumstances. From what it is known about pri-miRNA processing, it is probable that the distinct secondary structures created by the two pri-miRNA variants would be diced with varying efficiencies, with the intron containing variant being expected to produce a lower quantity of miRNA or not at all. Consequently, the higher amount of one or the other pri-miRNA variant in the cell would influence the levels of mature miRNA being produced. Thereby, controlling the percentage of each pri-miRNA transcript could be used as an additional layer of miRNA regulation beyond the activation or repression of transcription. This would confer the miRNA with a broader regulatory system and hence with higher flexibility to respond to different stimulus.

In this regard, it may be interesting to note that the intron-split structure of *MIRtop14* is conserved only on those genera in which miRNAtop14 targets LPR and not in the rest. This may be a mere coincidence or it may, otherwise, be speculated that this structure could be relevant in some way for the interaction between miRNAtop14 and LPR. Since the need for this pri-miRNA to avoid base paring with an antisense transcript is discarded, the hypothesis of this intron-split structure enabling a wider miRNA regulation through the production of the two pri-miRNAtop14 transcripts could better fit this observation. For instance, the wider flexibility of response gained by miRNAtop14 with this additional way of regulation would increase the potential of the miRNA to regulate more than one target, and thus it would have allowed the acquisition of LPR as new target besides the hypothesised target shared by all species in which miRNAtop14 is conserved.

This hypothesis which contemplate that the intron may be important for the regulation of pri-miRNA processing and thus for the modulation of miRNA levels and associated function has not been discarded to occur in nat-miRNAs, since it has not been experimentally tested; in the only study published it was assumed that all transcripts were constitutively spliced and that non-spliced pri-miRNAs would not be able to undergo dicing[195].

To study this conjecture, we have first analysed if the amount of mature miRNA produced from the spliced and the intron containing Sly-pri-miRNA variant differ from each other. As previously described in chapter 4, we have done so by

creating an artificial system in which *A. thaliana* plants, which do not contain *MIRtop14*, were independently transformed with three *Sly-MIRtop14* variants; wild type, a variant without intron and a variant in which the intron is non-spliceable.

After comparing the level of miRNAtop14 between the transgenic plants carrying MIRtop14 without intron or with an unspliceable intron, it was verified that the intron retaining pri-miRNA variant yields a much lower amount of miRNA than the spliced one, although not null. This result confirms that this intron-exon structure, by allowing the creation of the two pri-miRNA variants, makes possible a posttranscriptional change in the levels miRNAtop14 and, therefore, could potentially have value to the plant if it was used as additional regulatory element of the miRNA.

Besides, by comparing the accumulation of mature miRNAtop14 from the wild type versus the intron-lacking *MIRtop14* transformed plants using the same system as above, it was observed that the presence of the intron on the *MIR* enhances the levels of mature miRNA (as long as it can be spliced). A similar positive effect on miRNA accumulation had been previously reported for some other *MIRs* harbouring introns in positions downstream of the miRNA stem-loop[197,198]. In these studies, it was speculated that this was due to an increase in pri-miRNA processing, which was suggested to be achieved by the positive influence of the crosstalk between splicing and processing factors. This scenario would perfectly fit our results, which together with the results from the two previous studies would further support the hypothesis that the presence of the intron is advantageous for the *MIR* functioning.

The first step towards elucidating whether miRNAtop14 is being regulated at the posttranscriptional level has been taken by finding that two pri-miRNAs variants associated with different levels of mature miRNA accumulation are being produced from *MIRtop14*. However, several questions remain to be answered before being able to confirm that such event is indeed actively regulating the levels of miRNAtop14, and before understanding the complete mechanism of regulation.

First, it would need to be examined whether the production of one over the other pri-miRNAtop14 variant is being used *in vivo* to achieve different levels of miRNA accumulation in the cell. If such was the case, the ratio between spliced/non-spliced transcripts should show variation associated with the levels of total mature miRNA. This would indicate that mature miRNA levels are, at least partially, regulated through splicing. Otherwise, if this ratio remained constant, it would indicate that no regulation is taking place at this level. Finally, it should be possible to associate the production of one or the other pri-miRNA variant to some stimulus that should be

relevant for miRNAtop14 function, which would indicate that the regulation of miRNAtop14 levels is biologically meaningful.

In order to explore this, it would be necessary to first identify two circumstances in which miRNAtop14 levels vary, whether it is in response to a stimulus such as an abiotic or biotic stress, or between different stages of development or different tissues of the plant. For example, scarce of phosphate in the soil would be a plausible factor to be determining downregulation of miRNAtop14 in the root if LPR had the same function in tomato than in *Arabidopsis*, since LPR is known to be upregulated in the root of *Arabidopsis* in low phosphate conditions.

Once two conditions which determine differential miRNAtop14 accumulation are known, the levels of intron-retaining and of spliced pri-miRNAtop14 should be measured and related to those of mature miRNA. As already discussed, if they are involved in miRNAtop14 regulation, the ratio between one and the other pri-miRNA variant should change among samples from one or the other condition.

In contrast, if no change in this ratio is detected, it probably indicates that no regulation is taking place at this point of miRNA biogenesis and would open the question of why two pri-miRNA variants are being produced. One possibility would be that this regulation is still taking place under different conditions, since this miRNA is likely to be modulating more than one target. It would still be possible that while miRNAtop14 responds to one stimulus in order to regulate a specific target in a specific tissue or developmental stage, it could also respond to a different stimulus in a different tissue or developmental stage to regulate a second target. In this scenario, while the production of one over the other pri-miRNA variant may not be influenced by one of the stimulus, it can still be influenced by the other. Another possibility would be that the retained intron is altering the localization of the pri-miRNA. This mechanism has been seen in neurons, in which introns retained in some transcripts facilitate their targeting to dendrites, where they are finally spliced[436]. Finally, it would also be possible that the production of these two pri-miRNA transcripts does not have any value for the *MIR*, but it is instead a case of non-functional alternative splicing[437].

Lastly, in the case it was confirmed that miRNAtop14 levels are indeed being regulated by the different ratio between its pri-miRNA variants, it would still have to be elucidated the mechanism by which this is happening. In this case, two different regulatory scenarios would be possible:

1) There are two alternative splicing variants: an intron containing variant which produces very low level of miRNA and a spliced variant which produces high level

of miRNA. The non-spliced variant is not spliced if more mature miRNA is eventually required; in the case more mature miRNA is needed, more spliced variant is transcribed. The ratio between both variants would determine the levels of mature miRNA. A similar mechanism, in which mRNA abundance would be modulated through changing ratios between functional and nonsense/intron-retaining alternatively spliced variants has been already proposed[397].

2) The non-spliced transcript could be a “storage form” that would undergo splicing just when high levels of mature miRNA are required. Removal of the retained intron would enable efficient pri-miRNA processing and facilitate the quick rise of mature miRNA levels in response to some specific condition. Such scenario has been already observed in the plant *Marsilea vestita*, where some mRNAs retain their introns until reaching specific developmental stages, at which point they go through splicing and the translation into proteins is triggered[302].

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

1. Palade, G. E. A small particulate component of the cytoplasm. *J. Cell Biol.* **1,** (1955).

2. Hoagland, M. B., Stepheson, M. L., Scott, J. F., Hecht, L. I. & Zamecnik, P. C. A soluble ribonucleic acid intermediate in protein synthesis. *J. Biol. Chem.* **231,** 241–57 (1958).

3. Brenner, S., Jacob, F. & Meselson, M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* **190,** 576–581 (1961).

4. Weinberg, R. A. & Penman, S. Small molecular weight monodisperse nuclear RNA. *J. Mol. Biol.* **38,** 289–304 (1968).

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

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

7. Kruger, K. *et al.* Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* **31,** 147–157 (1982).

8. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35,** 849–57 (1983).

9. Greider, C. W. & Blackburn, E. H. The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell* **51,** 887–98 (1987).

10. Brown, C. J. *et al.* The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71,** 527–42 (1992).

11. Brockdorff, N. *et al.* The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* **71,** 515–26 (1992).

12. Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75,** 843–54 (1993).

13. Hamilton, A. J. & Baulcombe, D. C. A Species of Small Antisense RNA in Posttranscriptional Gene Silencing in Plants. *Science (80-.).* **950,** 10–13 (1999).

14. Wassenegger, M., Heimes, S., Riedel, L. & Sänger, H. L. RNA-directed de novo methylation of genomic sequences in plants. *Cell* **76,** 567–76 (1994).

15. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391,** 806–811 (1998).

16. Waterhouse, P. M., Graham, M. W. & Wang, M. B. Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 13959–64 (1998).

17. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).

18. Okazaki, Y. *et al.* Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420,** 563–573 (2002).

19. Aravin, A. *et al.* A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442,** 203–7 (2006).

20. Grivna, S. T., Beyret, E., Wang, Z. & Lin, H. A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev.* **20,** 1709–1714 (2006).

21. Girard, A., Sachidanandam, R., Hannon, G. J. & Carmell, M. A. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442,** 199–202 (2006).

22. Watanabe, T. *et al.* Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev.* **20,** 1732–1743 (2006).

23. Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H. L. & Poethig, R. S. SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. *Genes Dev.* **18,** 2368–2379 (2004).

24. Vazquez, F. *et al.* Endogenous trans-Acting siRNAs Regulate the Accumulation of Arabidopsis mRNAs. *Mol. Cell* **16,** 69–79 (2004).

25. Kapranov, P. *et al.* RNA Maps Reveal New RNA Classes and a Possible Function for Pervasive Transcription. *Science (80-.).* **316,** 1484–1488 (2007).

26. Kim, T.-K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465,** 182–187 (2010).

27. De Santa, F. *et al.* A Large Fraction of Extragenic RNA Pol II Transcription Sites Overlap Enhancers. *PLoS Biol.* **8,** e1000384 (2010).

28. Salzman, J., Gawad, C., Wang, P. L., Lacayo, N. & Brown, P. O. Circular RNAs Are the Predominant Transcript Isoform from Hundreds of Human Genes in Diverse Cell Types. *PLoS One* **7,** e30733 (2012).

29. Jeck, W. R. *et al.* Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19,** 141–157 (2013).

30. Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495,** 333–338 (2013).

31. Harvey Lodish, Arnold Berk, S Lawrence Zipursky, Paul Matsudaira, David Baltimore, and J. D. *The Three Roles of RNA in Protein Synthesis.* (200AD).

32. Morris, K. V & Mattick, J. S. The rise of regulatory RNA. *Nat. Rev. Genet.* **15,** 423–37 (2014).

33. Cech, T. R. & Steitz, J. A. The noncoding RNA revolution - Trashing old rules to forge new ones. *Cell* **157,** 77–94 (2014).

34. Bierhoff, H., Postepska-Igielska, A. & Grummt, I. Noisy silence: Non-coding RNA and heterochromatin formation at repetitive elements. *Epigenetics* **9,** 53–61 (2014).

35. Ding, S.-W. & Voinnet, O. Antiviral Immunity Directed by Small RNAs. *Cell* **130,** 413–426 (2007).

36. Dinger, M. E., Mercer, T. R. & Mattick, J. S. RNAs as extracellular signaling molecules. *J. Mol. Endocrinol.* **40,** 151–9 (2008).

37. Colombo, M., Raposo, G. & Théry, C. Biogenesis, Secretion, and Intercellular Interactions of Exosomes and Other Extracellular Vesicles. *Annu. Rev. Cell Dev. Biol* **30,** 255–89 (2014).

38. Carninci, P. RNA Dust: Where are the Genes? *DNA Res.* **17,** 51–59 (2010).

39. Taft, R. J., Pheasant, M. & Mattick, J. S. The relationship between non-protein-coding DNA and eukaryotic complexity. *BioEssays* **29,** 288–299 (2007).

40. Wightman, B., Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75,** 855–62 (1993).

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

42. Pasquinelli, A. E. *et al.* Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408,** 86–89 (2000).

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

44. Lau, Nelson C.; Lim, Lee P.; Weinstein, Earl G.; Bartel, D. P. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science (80-.).* **294,** 858–862 (2001).

45. Lee, R. C. *et al.* An extensive class of small RNAs in Caenorhabditis elegans. *Science* **294,** 862–4 (2001).

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

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

48. Mette, M. F., van der Winden, J., Matzke, M. & Matzke, A. J. M. Short RNAs can identify new candidate transposable element families in Arabidopsis. *Plant Physiol.* **130,** 6–9 (2002).

49. Park, W., Li, J., Song, R., Messing, J. & Chen, X. CARPEL FACTORY, a Dicer Homolog, and HEN1, a Novel Protein, Act in microRNA Metabolism in Arabidopsis thaliana. *Curr. Biol.* **12,** 1484–1495 (2002).

50. Pfeffer, S. *et al.* Identification of virus-encoded microRNAs. *Science (80-.).* **304,** 734–6 (2004).

51. 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. A uniform system for microRNA annotation. *RNA* **9,** 277–279 (2004).

52. Meyers, B. C. *et al.* Criteria for Annotation of Plant MicroRNAs. *Plant Cell Online* **20,** 3186–3190 (2008).

53. Griffiths-Jones, S., Saini, H. K., Van Dongen, S. & Enright, A. J. miRBase: Tools for microRNA genomics. *Nucleic Acids Res.* **36,** D154–D158 (2008).

54. Rajagopalan, R., Vaucheret, H., Trejo, J. & Bartel, D. P. A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Genes Dev.* **20,** 3407–25 (2006).

55. Brown, J. W. S., Marshall, D. F. & Echeverria, M. Intronic noncoding RNAs and splicing. *Trends in Plant Science* **13,** 335–342 (2008).

56. Yan, K. *et al.* Stress-Induced Alternative Splicing Provides a Mechanism for the Regulation of MicroRNA Processing in Arabidopsis thaliana. *Mol. Cell* **48,** 521–531 (2012).

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

58. Zhang, L. *et al.* A Genome-Wide Characterization of MicroRNA Genes in Maize. *PLoS Genet.* **5,** e1000716 (2009).

59. Jia, F. & Rock, C. D. 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 (2013).

60. Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *Eur. Mol. Biol. Organ. J.* **23,** 4051–4060 (2004).

61. Cai, X., Hagedorn, C. H. & Cullen, B. R. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* **10,** 1957–66 (2004).

62. Xie, Z. *et al.* Expression of Arabidopsis MIRNA genes. *Plant Physiol.* **138,** 2145–2154 (2005).

63. Kim, Y. J. *et al.* The role of Mediator in small and long noncoding RNA production in Arabidopsis thaliana. *EMBO J.* **30,** 814–822 (2011).

64. Hajheidari, M., Farrona, S., Huettel, B., Koncz, Z. & Koncz, C. CDKF;1 and CDKD Protein Kinases Regulate Phosphorylation of Serine Residues in the C-Terminal Domain of Arabidopsis RNA Polymerase II. *Plant Cell* **24,** 1626–1642 (2012).

65. Megraw, M. *et al.* MicroRNA promoter element discovery in Arabidopsis. *RNA* **12,** 1612–9 (2006).

66. Zhao, X., Zhang, H. & Li, L. Identification and analysis of the proximal promoters of microRNA genes in Arabidopsis. *Genomics* **101,** 187–194 (2013).

67. Yamasaki, H., Hayashi, M., Fukazawa, M., Kobayashi, Y. & Shikanai, T. SQUAMOSA Promoter Binding Protein-Like7 Is a Central Regulator for Copper Homeostasis in Arabidopsis. *Plant Cell* **21,** 347–61 (2009).

68. Baek, D., Park, H. C., Kim, M. C. & Yun, D.-J. The role of Arabidopsis MYB2

in miR399f-mediated phosphate-starvation response. *Plant Signal. Behav.* **8,** e23488 (2013).

69. Grigorova, B. *et al.* LEUNIG and SEUSS co-repressors regulate miR172 expression in Arabidopsis flowers. *Development* **138,** 2451–2456 (2011).

70. Cho, H. J. *et al.* SHORT VEGETATIVE PHASE (SVP) protein negatively regulates miR172 transcription via direct binding to the pri-miR172a promoter in *Arabidopsis*. *FEBS Lett.* **586,** 2332–2337 (2012).

71. Yumul, R. E. *et al.* POWERDRESS and Diversified Expression of the MIR172 Gene Family Bolster the Floral Stem Cell Network. *PLoS Genet* **9,** (2013).

72. Wu, G. *et al.* The Sequential Action of miR156 and miR172 Regulates Developmental Timing in Arabidopsis. *Cell* **138,** 750–759 (2009).

73. Wang, L. *et al.* NOT2 Proteins Promote Polymerase II-Dependent Transcription and Interact with Multiple MicroRNA Biogenesis Factors in Arabidopsis. *Plant Cell* **25,** 715–727 (2013).

74. Zhang, S., Xie, M., Ren, G. & Yu, B. CDC5, a DNA binding protein, positively regulates posttranscriptional processing and/or transcription of primary microRNA transcripts. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 17588–93 (2013).

75. Lauressergues, D. *et al.* Primary transcripts of microRNAs encode regulatory peptides. *Nature* **520,** 90–93 (2015).

76. Kim, W. *et al.* Histone acetyltransferase GCN5 interferes with the miRNA pathway in Arabidopsis. *Cell Res.* **19,** 899–909 (2009).

77. Choi, K. *et al.* Regulation of microRNA-mediated developmental changes by the SWR1 chromatin remodeling complex in Arabidopsis thaliana. *Plant Physiol.* **171,** 1128–1143 (2016).

78. Bartel, D. P. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* (2004). doi:10.1016/S0092-8674(04)00045-5

79. Fang, Y. & Spector, D. L. Identification of Nuclear Dicing Bodies Containing Proteins for MicroRNA Biogenesis in Living Arabidopsis Plants. *Curr. Biol.* **17,** 818–823 (2007).

80. Song, L., Han, M.-H., Lesicka, J. & Fedoroff, N. Arabidopsis primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 5437–42 (2007).

81. Fujioka, Y., Utsumi, M., Ohba, Y. & Watanabe, Y. Location of a Possible miRNA Processing Site in SmD3/SmB Nuclear Bodies in Arabidopsis. *Plant Cell Physiol.* **48,** 1243–1253 (2007).

82. Vazquez, F., Blevins, T., Ailhas, J., Boller, T. & Meins, F. Evolution of Arabidopsis MIR genes generates novel microRNA classes. *Nucleic Acids Res.* **36,** 6429–6438 (2008).

83. Wu, L. *et al.* DNA Methylation Mediated by a MicroRNA Pathway. *Mol. Cell* **38,** 465–475 (2010).

84. Han, M.-H., Goud, S., Song, L. & Fedoroff, N. The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 1093–8 (2004).

85. Vazquez, F. *et al.* The nuclear dsRNA binding protein HYL1 is required for

microRNA accumulation and plant development, but not posttranscriptional transgene silencing. *Curr. Biol.* **14,** 346–51 (2004).

86. Lobbes, D., Rallapalli, G., Schmidt, D. D., Martin, C. & Clarke, J. SERRATE: a new player on the plant microRNA scene. *EMBO Rep.* **7,** 1052–1058 (2006).

87. Yang, L., Liu, Z., Lu, F., Dong, A. & Huang, H. SERRATE is a novel nuclear regulator in primary microRNA processing in Arabidopsis. *Plant J.* **47,** 841–850 (2006).

88. Kurihara, Y., Takashi, Y. & Watanabe, Y. The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *RNA* **12,** 206–12 (2006).

89. Dong, Z., Han, M.-H. & Fedoroff, N. The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 9970–5 (2008).

90. Ren, G. *et al.* Regulation of miRNA abundance by RNA binding protein TOUGH in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 12817–21 (2012).

91. Gregory, B. D. *et al.* A Link between RNA Metabolism and Silencing Affecting Arabidopsis Development. *Dev. Cell* (2008). doi:10.1016/j.devcel.2008.04.005

92. Laubinger, S. *et al.* Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 8795–800 (2008).

93. Kim, S. *et al.* Two Cap-Binding Proteins CBP20 and CBP80 are Involved in Processing Primary MicroRNAs. *Plant Cell Physiol.* **49,** 1634–1644 (2008).

94. Raczynska, K. D. *et al.* The SERRATE protein is involved in alternative splicing in Arabidopsis thaliana. *Nucleic Acids Res.* **42,** 1224–1244 (2014).

95. Zhang, S., Liu, Y., Yu, B., Gelvin, S. & Zhang, S. PRL1, an RNA-Binding Protein, Positively Regulates the Accumulation of miRNAs and siRNAs in Arabidopsis. *PLoS Genet.* **10,** e1004841 (2014).

96. Yu, B. *et al.* The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 10073–8 (2008).

97. Chen, T., Cui, P. & Xiong, L. The RNA-binding protein HOS5 and serine/arginine-rich proteins RS40 and RS41 participate in miRNA biogenesis in Arabidopsis. *Nucleic Acids Res.* **43,** 8283–8298 (2015).

98. Ben Chaabane, S. *et al.* STA1, an Arabidopsis pre-mRNA processing factor 6 homolog, is a new player involved in miRNA biogenesis. *Nucleic Acids Res.* **41,** 1984–1997 (2013).

99. Koster, T. *et al.* Regulation of pri-miRNA processing by the hnRNP-like protein AtGRP7 in Arabidopsis. *Nucleic Acids Res.* **42,** 9925–9936 (2014).

100. Manavella, P. A. *et al.* Fast-forward genetics identifies plant CPL phosphatases as regulators of miRNA processing factor HYL1. *Cell* **151,** 859–70 (2012).

101. Wu, X. *et al.* A role for the RNA-binding protein MOS2 in microRNA maturation in Arabidopsis. *Cell Res.* **23,** 645–657 (2013).

102. Francisco-Mangilet, A. G. *et al.* THO2, a core member of the THO/TREX complex, is required for microRNA production in Arabidopsis. *Plant J.* **82,** 1018–1029 (2015).

103. Speth, C., Willing, E.-M., Rausch, S., Schneeberger, K. & Laubinger, S. RACK1 scaffold proteins influence miRNA abundance in Arabidopsis. *Plant J.* **76,** 433–445 (2013).

104. Zhan, X. *et al.* Arabidopsis proline-rich protein important for development and abiotic stress tolerance is involved in microRNA biogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 18198–203 (2012).

105. Reis, R. S., Hart-smith, G., Eamens, A. L., Wilkins, M. R. & Waterhouse, P. M. Gene regulation by translational inhibition is determined by Dicer partnering proteins. *Nat. Plants* **1,** 1–6 (2015).

106. Bologna, N. G. *et al.* Multiple RNA recognition patterns during microRNA biogenesis in plants. *Genome Res.* **23,** 1675–1689 (2013).

107. Mateos, J. L., Bologna, N. G., Chorostecki, U. & Palatnik, J. F. Identification of MicroRNA Processing Determinants by Random Mutagenesis of Arabidopsis MIR172a Precursor. *Curr. Biol.* **20,** 49–54 (2010).

108. Song, L., Axtell, M. J. & Fedoroff, N. V. RNA Secondary Structural Determinants of miRNA Precursor Processing in Arabidopsis. *Curr. Biol.* **20,** 37–41 (2010).

109. Werner, S., Wollmann, H., Schneeberger, K. & Weigel, D. Structure Determinants for Accurate Processing of miR172a in Arabidopsis thaliana. *Curr. Biol.* **20,** 42–48 (2010).

110. Addo-Quaye, C. *et al.* Sliced microRNA targets and precise loop-first processing of MIR319 hairpins revealed by analysis of the Physcomitrella patens degradome. *RNA* **15,** 2112–21 (2009).

111. Bologna, N. G., Mateos, J. L., Bresso, E. G. & Palatnik, J. F. A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159. *EMBO J.* **28,** 3646–3656 (2009).

112. Zhu, H. *et al.* Bidirectional processing of pri-miRNAs with branched terminal loops by Arabidopsis Dicer-like1. *Nat. Struct. Mol. Biol.* **20,** 1106–1115 (2013).

113. Yu, B. *et al.* Methylation as a crucial step in plant microRNA biogenesis. *Science* **307,** 932–5 (2005).

114. Yang, Z., Ebright, Y. W., Yu, B. & Chen, X. HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2’ OH of the 3’ terminal nucleotide. *Nucleic Acids Res.* **34,** 667–675 (2006).

115. Baranauske, S. *et al.* Functional mapping of the plant small RNA methyltransferase: HEN1 physically interacts with HYL1 and DICER-LIKE 1 proteins. *Nucleic Acids Res.* **43,** 2802–2812 (2015).

116. Ren, G., Chen, X. & Yu, B. Uridylation of miRNAs by HEN1 SUPPRESSOR1 in arabidopsis. *Curr. Biol.* **22,** 695–700 (2012).

117. Zhao, Y. *et al.* The arabidopsis nucleotidyl transferase HESO1 uridylates unmethylated small RNAs to trigger their degradation. *Current Biology* **22,** 689–694 (2012).

118. Rogers, K. & Chen, X. Biogenesis, turnover, and mode of action of plant microRNAs. *Plant Cell* **25,** 2383–99 (2013).

119. Park, M. Y., Wu, G., Gonzalez-Sulser, A., Vaucheret, H. & Poethig, R. S. Nuclear processing and export of microRNAs in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 3691–3696 (2005).

120. Eamens, A. L., Smith, N. A., Curtin, S. J., Wang, M.-B. & Waterhouse, P. M. The Arabidopsis thaliana double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes. *RNA* **15,** 2219–35 (2009).

121. Meister, G. & Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. *Nature* **431,** 343–349 (2004).

122. Vaucheret, H., Vazquez, F., Crété, P. & Bartel, D. P. The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. *Genes Dev.* **18,** 1187–1197 (2004).

123. Baumberger, N. & Baulcombe, D. C. Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 11928–33 (2005).

124. Qi, Y. *et al.* Biochemical Specialization within Arabidopsis RNA Silencing Pathways. *Mol. Cell* **19,** 421–428 (2005).

125. Song, J.-J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal Structure of Argonaute and Its Implications for RISC Slicer Activity. *Science (80-.).* **305,** (2004).

126. Song, J.-J. *et al.* The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat. Struct. Biol.* **10,** 1026–1032 (2003).

127. Ma, J.-B., Ye, K. & Patel, D. J. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* **429,** 318–322 (2004).

128. Ma, J. B. *et al.* Structural basis for 5 ’-end-specific recognition of guide RNA by the A-fulgidus Piwi protein. *Nature* **434,** 356–372 (2005).

129. Parker, J. S., Roe, S. M. & Barford, D. Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* **434,** 663–6 (2005).

130. Liu, J. *et al.* Argonaute2 Is the Catalytic Engine of Mammalian RNAi. *Science (80-.).* **305,** (2004).

131. Rivas, F. V *et al.* Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat. Struct. Mol. Biol.* **12,** 340–349 (2005).

132. Iki, T. *et al.* In Vitro Assembly of Plant RNA-Induced Silencing Complexes Facilitated by Molecular Chaperone HSP90. *Mol. Cell* **39,** 282–291 (2010).

133. Zhang, X. *et al.* Arabidopsis Argonaute 2 Regulates Innate Immunity via miRNA393*-Mediated Silencing of a Golgi-Localized SNARE Gene, MEMB12. *Mol. Cell* **42,** 356–366 (2011).

134. Devers, E. A., Branscheid, A., May, P. & Krajinski, F. Stars and Symbiosis: MicroRNA- and MicroRNA*-Mediated Transcript Cleavage Involved in Arbuscular Mycorrhizal Symbiosis. *Plant Physiol.* **156,** 1990–2010 (2011).

135. Khvorova, A. *et al.* Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115,** 209–16 (2003).

136. Schwarz, D. S. *et al.* Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115,** 199–208 (2003).

137. Mi, S. *et al.* Sorting of Small RNAs into Arabidopsis Argonaute Complexes Is Directed by the 5′ Terminal Nucleotide. *Cell* **133,** 116–127 (2008).

138. Montgomery, T. A. *et al.* Specificity of ARGONAUTE7-miR390 Interaction and Dual Functionality in TAS3 Trans-Acting siRNA Formation. *Cell* (2008). doi:10.1016/j.cell.2008.02.033

139. Takeda, A., Iwasaki, S., Watanabe, T., Utsumi, M. & Watanabe, Y. The mechanism selecting the guide strand from small RNA duplexes is different among Argonaute proteins. *Plant Cell Physiol.* **49,** 493–500 (2008).

140. Vaucheret, H. Plant ARGONAUTES. *Trends Plant Sci.* **13,** 350–358 (2008).

141. Iwakawa, H. & Tomari, Y. The Functions of MicroRNAs: mRNA Decay and Translational Repression. *Trends Cell Biol.* **25,** 651–665 (2015).

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

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

144. Mallory, A. C. *et al.* MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5′ region. *EMBO J.* **23,** (2004).

145. Brodersen, P. *et al.* Widespread translational inhibition by plant miRNAs and siRNAs. *Science* **320,** 1185–90 (2008).

146. Liu, Q., Wang, F. & Axtell, M. J. Analysis of complementarity requirements for plant MicroRNA targeting using a Nicotiana benthamiana quantitative transient assay. *Plant Cell* **26,** 741–53 (2014).

147. Beauclair, L., Yu, A. & Bouché, N. microRNA-directed cleavage and translational repression of the copper chaperone for superoxide dismutase mRNA in Arabidopsis. *Plant J.* **62,** 454–462 (2010).

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

149. German, M. a *et al.* Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nat. Biotechnol.* **26,** 941–946 (2008).

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

151. Souret, F. F., Kastenmayer, J. P. & Green, P. J. AtXRN4 degrades mRNA in Arabidopsis and its substrates include selected miRNA targets. *Mol. Cell* **15,** 173–183 (2004).

152. Shen, B. & Goodman, H. M. Uridine Addition After MicroRNA-Directed Cleavage. *Science (80-.).* **306,** (2004).

153. Ren, G. *et al.* Methylation protects microRNAs from an AGO1-associated

activity that uridylates 5’ RNA fragments generated by AGO1 cleavage. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 6365–70 (2014).

154. Fei, Q., Xia, R. & Meyers, B. C. Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. *Plant Cell* **25,** 2400–15 (2013).

155. Yang, L., Wu, G. & Poethig, R. S. Mutations in the GW-repeat protein SUO reveal a developmental function for microRNA-mediated translational repression in Arabidopsis. *Proc. Natl. Acad. Sci.* **109,** 315–320 (2012).

156. Li, S. *et al.* MicroRNAs Inhibit the Translation of Target mRNAs on the Endoplasmic Reticulum in Arabidopsis. *Cell* **153,** 562–574 (2013).

157. Lanet, E. *et al.* Biochemical Evidence for Translational Repression by Arabidopsis MicroRNAs. *Plant Cell* **21,** 1762–8 (2009).

158. Iwakawa, H. & Tomari, Y. *Molecular Insights into microRNA-Mediated Translational Repression in Plants*. *Molecular Cell* **52,** (2013).

159. Xu, J., Yang, J.-Y., Niu, Q.-W. & Chua, N.-H. Arabidopsis DCP2, DCP1, and VARICOSE form a decapping complex required for postembryonic development. *Plant Cell* **18,** 3386–98 (2006).

160. Hamada, T. *et al.* RNA Processing Bodies, Peroxisomes, Golgi Bodies, Mitochondria, and Endoplasmic Reticulum Tubule Junctions Frequently Pause at Cortical Microtubules. *Plant Cell Physiol.* **53,** 699–708 (2012).

161. Jonas, S. & Izaurralde, E. Towards a molecular understanding of microRNA-mediated gene silencing. *Nat. Rev. Genet.* **16,** 421–433 (2015).

162. Arribas-Hernández, L., Kielpinski, L. J. & Brodersen, P. mRNA Decay of Most Arabidopsis miRNA Targets Requires Slicer Activity of AGO1. *Plant Physiol.* **171,** 2620–32 (2016).

163. Aukerman, M. J. & Sakai, H. Regulation of Flowering Time and Floral Organ Identity by a MicroRNA and Its APETALA2-Like Target Genes. *PLANT CELL ONLINE* **15,** 2730–2741 (2003).

164. Chen, X. A MicroRNA as a Translational Repressor of APETALA2 in Arabidopsis Flower Development HHS Public Access. *Sci. March* **26,** 2022–2025 (2004).

165. Gandikota, M. *et al.* The miRNA156/157 recognition element in the 3′ UTR of the Arabidopsis SBP box gene SPL3 prevents early flowering by translational inhibition in seedlings. *Plant J.* **49,** 683–693 (2007).

166. Jones-Rhoades, M. W., Bartel, D. P. & Bartel, B. MicroRNAS and their regulatory roles in plants. *Annu. Rev. Plant Biol.* **57,** 19–53 (2006).

167. Carbonell, A. *et al.* Functional Analysis of Three Arabidopsis ARGONAUTES Using Slicer-Defective Mutants. *Plant Cell* **24,** 3613–3629 (2012).

168. Reis, R. S., Hart-Smith, G., Eamens, A. L., Wilkins, M. R. & Waterhouse, P. M. MicroRNA regulatory mechanisms play different roles in arabidopsis. *J. Proteome Res.* **14,** 4743–4751 (2015).

169. Bao, N., Lye, K.-W. & Barton, M. K. MicroRNA Binding Sites in Arabidopsis Class III HD-ZIP mRNAs Are Required for Methylation of the Template Chromosome. *Dev. Cell* **7,** 653–662 (2004).

170. Khraiwesh, B. *et al.* Transcriptional Control of Gene Expression by

MicroRNAs. *Cell* **140,** 111–122 (2010).

171. Chellappan, P. *et al.* siRNAs from miRNA sites mediate DNA methylation of target genes. *Nucleic Acids Res.* **38,** 6883–6894 (2010).

172. Sanei, M. & Chen, X. Mechanisms of microRNA turnover. *Curr. Opin. Plant Biol.* **27,** 199–206 (2015).

173. Yu, B. *et al.* Methylation as a crucial step in plant microRNA biogenesis. *Science (80-.).* **307,** 932–935 (2005).

174. Li, J., Yang, Z., Yu, B., Liu, J. & Chen, X. Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. *Curr. Biol.* **15,** 1501–1507 (2005).

175. Sement, F. M. *et al.* Uridylation prevents 3′ trimming of oligoadenylated mRNAs. *Nucleic Acids Res.* **41,** 7115–7127 (2013).

176. Wang, X. *et al.* Synergistic and Independent Actions of Multiple Terminal Nucleotidyl Transferases in the 3' Tailing of Small RNAs in Arabidopsis. *PLoS Genet* **11,** (2015).

177. Tu, B. *et al.* Distinct and Cooperative Activities of HESO1 and URT1 Nucleotidyl Transferases in MicroRNA Turnover in Arabidopsis. *PLoS Genet.* **11,** (2015).

178. Ibrahim, F. *et al.* Uridylation of mature miRNAs and siRNAs by the MUT68 nucleotidyltransferase promotes their degradation in Chlamydomonas. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 3906–11 (2010).

179. Rogers, K. & Chen, X. MicroRNA biogenesis and turnover in plants. *Cold Spring Harb. Symp. Quant. Biol.* **77,** 183–194 (2012).

180. Zhai, J. *et al.* Plant MicroRNAs Display Differential 3' Truncation and Tailing Modifications That Are ARGONAUTE1 Dependent and Conserved Across Species. *Plant Cell* **25,** 2417–2428 (2013).

181. Ramachandran, V. & Chen, X. Degradation of microRNAs by a family of exoribonucleases in Arabidopsis. *Science* **321,** 1490–2 (2008).

182. Yang, G. D. *et al.* Genomewide analysis of intronic microRNAs in rice and Arabidopsis. *J. Genet* **91,** 313–324 (2012).

183. Ruby, J. G., Jan, C. H. & Bartel, D. P. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448,** 83–6 (2007).

184. Okamura, K., Hagen, J. W., Duan, H., Tyler, D. M. & Lai, E. C. The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. *Cell* **130,** 89–100 (2007).

185. Zhu, Q. H. *et al.* A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. *Genome Res.* **18,** 1456–1465 (2008).

186. Joshi, P. K. *et al.* Identification of mirtrons in rice using MirtronPred: A tool for predicting plant mirtrons. *Genomics* (2012). doi:10.1016/j.ygeno.2012.04.002

187. Piriyapongsa, J. & Jordan, I. K. Dual coding of siRNAs and miRNAs by plant transposable elements. *Rna* **14,** 814–821 (2008).

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

(2008).

189. Li, A. & Mao, L. Evolution of plant microRNA gene families. *Cell Res.* | *Cell Res.* **17,** 212–218 (2007).

190. Chuck, G., Cigan, A. M., Saeteurn, K. & Hake, S. The heterochronic maize mutant Corngrass1 results from overexpression of a tandem microRNA. *Nat. Genet.* **39,** 544–549 (2007).

191. Wang, S. *et al.* Molecular evolution and selection of a gene encoding two tandem microRNAs in rice. *FEBS Lett.* **581,** 4789–4793 (2007).

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

193. Stepien, A. *et al.* Posttranscriptional coordination of splicing and miRNA biogenesis in plants. *Wiley Interdiscip. Rev. RNA* (2016). doi:10.1002/wrna.1403

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

195. Lu, C. *et al.* Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). *Proc. Natl. Acad. Sci. U. S. A.* **105,** 4951–6 (2008).

196. Kruszka, K. *et al.* Developmentally regulated expression and complex processing of barley pri-microRNAs. *BMC Genomics* **14,** 34 (2013).

197. Schwab, R., Speth, C., Laubinger, S. & Voinnet, O. Enhanced microRNA accumulation through stemloop-adjacent introns. *EMBO Rep.* **14,** 615–21 (2013).

198. Bielewicz, D. *et al.* Introns of plant pri-miRNAs enhance miRNA biogenesis. *EMBO Rep.* **14,** 622–8 (2013).

199. Barciszewska-Pacak, Maria; Knop, K., Jarmolowski, A. & Szweykowska-kulinska, Z. Arabidopsis thaliana microRNA162 level is posttranscriptionally regulated via splicing and polyadenylation site selection. *Acta Biochim. Pol.* **63,** 811–816 (2016).

200. Hirsch, J. *et al.* Characterization of 43 non-protein-coding mRNA genes in Arabidopsis, including the MIR162a-derived transcripts. *Plant Physiol.* **140,** 1192–204 (2006).

201. Knop, K. *et al.* Active 5′ splice sites regulate the biogenesis efficiency of Arabidopsis microRNAs derived from intron-containing genes. *Nucleic Acids Res.* **323,** gkw895 (2016).

202. Morin, R. D. *et al.* Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res.* **18,** 610–21 (2008).

203. Morin, R. D. *et al.* Comparative analysis of the small RNA transcriptomes of Pinus contorta and Oryza sativa. *Genome Res.* **18,** 571–84 (2008).

204. Neilsen, C. T., Goodall, G. J. & Bracken, C. P. IsomiRs – the overlooked repertoire in the dynamic microRNAome. *Trends Genet.* **28,** 544–549 (2012).

205. Jeong, D.-H. *et al.* Comprehensive investigation of microRNAs enhanced by

analysis of sequence variants, expression patterns, ARGONAUTE loading, and target cleavage. *Plant Physiol.* **162,** 1225–45 (2013).

206. Ameres, S. L. & Zamore, P. D. Diversifying microRNA sequence and function. *Nat. Rev. Mol. Cell Biol.* **14,** 475–88 (2013).

207. Guo, L. & Chen, F. A challenge for miRNA: multiple isomiRs in miRNAomics. *Gene* **544,** 1–7 (2014).

208. Allen, E. *et al.* Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana. *Nat. Genet.* **36,** 1282–1290 (2004).

209. Fahlgren, N. *et al.* High-throughput sequencing of Arabidopsis microRNAs: Evidence for frequent birth and death of MIRNA genes. *PLoS One* **2,** (2007).

210. Fahlgren, N. *et al.* MicroRNA gene evolution in Arabidopsis lyrata and Arabidopsis thaliana. *Plant Cell* **22,** 1074–89 (2010).

211. He, H., Liang, G., Li, Y., Wang, F. & Yu, D. Two young MicroRNAs originating from target duplication mediate nitrogen starvation adaptation via regulation of glucosinolate synthesis in Arabidopsis thaliana. *Plant Physiol.* **164,** 853–65 (2014).

212. Xia, R., Ye, S., Liu, Z., Meyers, B. C. & Liu, Z. Novel and Recently Evolved MicroRNA Clusters Regulate Expansive F-BOX Gene Networks through Phased Small Interfering RNAs in Wild Diploid Strawberry 1[OPEN]. *Plant Physiol.* (2015). doi:10.1104/pp.15.00253

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

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

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

216. Felippes, F. F. de, Schneeberger, K., Dezulian, T., Huson, D. H. & Weigel, D. Evolution of Arabidopsis thaliana microRNAs from random sequences. *RNA* **14,** 2455–9 (2008).

217. Nozawa, M., Miura, S. & Nei, M. Origins and evolution of microRNA genes in plant species. *Genome Biol. Evol.* **4,** 230–239 (2012).

218. Ma, Z., Coruh, C. & Axtell, M. J. Arabidopsis lyrata small RNAs: transient MIRNA and small interfering RNA loci within the Arabidopsis genus. *Plant Cell* **22,** 1090–103 (2010).

219. Roberts, J. T., Cardin, S. E. & Borchert, G. M. Burgeoning evidence indicates that microRNAs were initially formed from transposable element sequences. *Mob. Genet. Elements* **4,** e29255 (2014).

220. Bureau, T. E. & Wessler, S. R. Tourist: a large family of small inverted repeat elements frequently associated with maize genes. *Plant Cell* **4,** 1283–94 (1992).

221. Feschotte, C. & Mouchès, C. Evidence that a family of miniature inverted-repeat transposable elements (MITEs) from the Arabidopsis thaliana genome has arisen from a pogo-like DNA transposon. *Mol. Biol. Evol.* **17,** 730–7 (2000).

222. Yao, C. *et al.* Cloning of novel repeat-associated small RNAs derived from hairpin precursors in Oryza sativa. *Acta Biochim. Biophys. Sin. (Shanghai).* **39,** 829–834 (2007).

223. Li, Y., Li, C., Xia, J., Jin, Y. & Tacker, M. Domestication of Transposable Elements into MicroRNA Genes in Plants. *PLoS One* **6,** e19212 (2011).

224. Zhang, Y., Jiang, W.-K. & Gao, L.-Z. Evolution of MicroRNA Genes in Oryza sativa and Arabidopsis thaliana: An Update of the Inverted Duplication Model. (2011). doi:10.1371/journal.pone.0028073

225. Sun, J., Zhou, M., Mao, Z. & Li, C. Characterization and Evolution of microRNA Genes Derived from Repetitive Elements and Duplication Events in Plants. *PLoS One* **7,** (2012).

226. Ou-Yang, F. *et al.* Transposable element-associated microRNA hairpins produce 21-nt sRNAs integrated into typical microRNA pathways in rice. *Funct. Integr. Genomics* **13,** 207–16 (2013).

227. Kurtoglu, K. Y., Kantar, M. & Budak, H. New wheat microRNA using whole-genome sequence. *Funct. Integr. Genomics* **14,** 363–379 (2014).

228. Yu, M., Carver, B. F. & Yan, L. TamiR1123 originated from a family of miniature inverted-repeat transposable elements (MITE) including one inserted in the Vrn-A1a promoter in wheat. *Plant Sci.* **215,** 117–123 (2014).

229. Dai, S. *et al.* Widespread and evolutionary analysis of a MITE family Monkey King in Brassicaceae. *BMC Plant Biol.* **15,** 149 (2015).

230. Cuperus, J. T., Fahlgren, N. & Carrington, J. C. Evolution and Functional Diversification of *MIRNA* Genes. *Plant Cell* **23,** 431–442 (2011).

231. Maher, C., Stein, L. & Ware, D. Evolution of Arabidopsis microRNA families through duplication events. *Genome Res.* **16,** 510–519 (2006).

232. Palatnik, J. F. *et al.* Sequence and Expression Differences Underlie Functional Specialization of Arabidopsis MicroRNAs miR159 and miR319. *Dev. Cell* **13,** 115–125 (2007).

233. Morea, E. G. O. *et al.* Functional and evolutionary analyses of the miR156 and miR529 families in land plants. *BMC Plant Biol.* **16,** 40 (2016).

234. Axtell, M. J. & Bowman, J. L. Evolution of plant microRNAs and their targets. *Trends in Plant Science* **13,** 343–349 (2008).

235. Alaba, S. *et al.* The liverwort Pellia endiviifolia shares microtranscriptomic traits that are common to green algae and land plants. *New Phytol.* **206,** 352–67 (2015).

236. Chávez Montes, R. a *et al.* Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs. *Nat. Commun.* **5,** 3722 (2014).

237. Axtell, M. J. Evolution of microRNAs and their targets: Are all microRNAs biologically relevant? *BBA - Gene Regul. Mech.* **1779,** 725–734 (2008).

238. Taylor, R. S., Tarver, J. E., Hiscock, S. J. & Donoghue, P. C. J. Evolutionary history of plant microRNAs. *Trends Plant Sci.* **19,** 175–182 (2014).

239. Guo, X. *et al.* Selection and mutation on microRNA target sequences during rice evolution. *BMC Genomics* **9,** 454 (2008).

240. Ehrenreich, I. M. & Purugganan, M. D. Sequence Variation of MicroRNAs and Their Binding Sites in Arabidopsis. *PLANT Physiol.* **146,** 1974–1982 (2008).

241. Takuno, S. & Innan, H. Selection fine-tunes the expression of MicroRNA target genes in arabidopsis thaliana. *Mol. Biol. Evol.* **28,** 2429–2434 (2011).

242. Abrouk, M. *et al.* Grass microRNA gene paleohistory unveils new insights into gene dosage balance in subgenome partitioning after whole-genome duplication. *Plant Cell* **24,** 1776–92 (2012).

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

244. Zhao, M., Meyers, B. C., Cai, C., Xu, W. & Ma, J. Evolutionary Patterns and Coevolutionary Consequences of MIRNA Genes and MicroRNA Targets Triggered by Multiple Mechanisms of Genomic Duplications in Soybean. *Plant Cell* **27,** 546–562 (2015).

245. Liu, T. *et al.* Global investigation of the co-evolution of MIRNA genes and microRNA targets during soybean domestication. *Plant J.* **85,** 396–409 (2016).

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

247. Millar, A. A. & Waterhouse, P. M. Plant and animal microRNAs: Similarities and differences. *Functional and Integrative Genomics* (2005). doi:10.1007/s10142-005-0145-2

248. Cerutti, H. & Casas-Mollano, J. A. On the origin and functions of RNA-mediated silencing: From protists to man. *Current Genetics* **50,** 81–99 (2006).

249. Ghildiyal, M. & Zamore, P. D. Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* **10,** 94–108 (2009).

250. Cui, J., You, C. & Chen, X. The evolution of microRNAs in plants. *Curr. Opin. Plant Biol.* **35,** 61–67 (2017).

251. Moran, Y., Agron, M., Praher, D. & Technau, U. The evolutionary origin of plant and animal microRNAs. *Nat. Ecol. Evol.* **1,** 0027 (2017).

252. Brack, C. & Tonegawa, S. 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 (1977).

253. Jeffreys, A. J. *et al.* The rabbit beta-globin gene contains a large large insert in the coding sequence. *Cell* **12,** 1097–108 (1977).

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

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

256. Alt, F. W. *et al.* Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3' ends. *Cell* **20,** 293–301 (1980).

257. Early, P. *et al.* Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. *Cell* **20,** 313–319 (1980).

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

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

260. Grabowski, P. J., Seller, S. R. & Sharp, P. A. A Multicomponent Complex Is Involved in the Splicing of Messenger RNA Precursors. *Cell* **42,** 345–353 (1985).

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

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

263. Genomics, B. *et al.* Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. *BMC Genomics* **7,** (2006).

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

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

266. Kornblihtt, A. R. *et al.* Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* **14,** 153–165 (2013).

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

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

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

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

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

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

273. Yang, X. *et al.* Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. *Cell* **164,** 805–817 (2016).

274. Brett, D., Pospisil, H., Valcárcel, J., Reich, J. & Bork, P. Alternative splicing and genome complexity. *Nat. Genet.* **30,** 29–30 (2002).

275. Tress, M. L., Abascal, F. & Valencia, A. Alternative Splicing May Not Be the Key to Proteome Complexity. *Trends in Biochemical Sciences* **42,** 98–110 (2017).

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

277. Abelson, J., Trotta, C. R. & Li, H. tRNA splicing. *J. Biol. Chem.* **273,** 12685–8 (1998).

278. Rodríguez-Trelles, F., Tarrío, R. & Ayala, F. J. Origins and Evolution of Spliceosomal Introns. *Annu. Rev. Genet* **40,** 47–76 (2006).

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

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

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

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

283. Sharp, P. A. & Burge, C. B. Classification of introns: U2-type or U12-type. *Cell* **91,** 875–879 (1997).

284. Brown, J. W. S., Smith, P. & Simpson, C. G. Arabidopsis consensus intron sequences. *Plant Mol. Biol.* **32,** 531–535 (1996).

285. Marquez, Y., Brown, J. W. S., Simpson, C., Barta, A. & Kalyna, M. Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. *Genome Res.* **22,** 1184–1195 (2012).

286. Jackson, L. J. A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res.* **19,** 3795–3798 (1991).

287. Hall, S. L. & Padgett, R. A. Conserved Sequences in a Class of Rare Eukaryotic Nuclear Introns with Non-consensus Splice Sites. *J. Mol. Biol.* **239,** 357–365 (1994).

288. Lewandowska, D. *et al.* Determinants of plant U12-dependent intron splicing efficiency. *Plant Cell* **16,** 1340–52 (2004).

289. Wahl, M. C., Will, C. L. & Lührmann, R. The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell* **136,** 701–718 (2009).

290. Will, C. L. & Lührmann, R. Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3,** 1–2 (2011).

291. Hoskins, A. A. & Moore, M. J. The spliceosome: A flexible, reversible macromolecular machine. *Trends in Biochemical Sciences* **37,** 179–188 (2012).

292. Wang, B.-B. & Brendel, V. The ASRG database: identification and survey of Arabidopsis thaliana genes involved in pre-mRNA splicing. *Genome Biol.* **5,** R102 (2004).

293. Simpson, C. G. *et al.* Mutational analysis of a plant branchpoint and polypyrimidine tract required for constitutive splicing of a mini-exon. *RNA* **8,** 47–56 (2002).

294. Iwata, H. & Gotoh, O. Comparative analysis of information contents relevant to recognition of introns in many species. *BMC Genomics* **12,** 45 (2011).

295. Reddy, A. S. N. Alternative Splicing of Pre-Messenger RNAs in Plants in the Genomic Era. *Annu. Rev. Plant Biol* **58,** 267–94 (2007).

296. Luehrsen, K. R., Taha, S. & Walbot, V. Nuclear Pre-mRna Processing in Higher Plants. *Prog. Nucleic Acid Res. Mol. Biol.* **47,** 149–193 (1994).

297. Barbazuk, W. B., Fu, Y. & McGinnis, K. M. Genome-wide analyses of alternative splicing in plants: Opportunities and challenges. *Genome Research* **18,** 1381–1392 (2008).

298. Beyer, A. L. & Osheim, Y. N. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* **2,** 754–765 (1988).

299. Khodor, Y. L. *et al.* Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. *Genes Dev.* **25,** 2502–2512 (2011).

300. Ameur, A. *et al.* Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat. Struct. Mol. Biol.* **18,** 1435–1440 (2011).

301. Bhatt, D. M. *et al.* Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150,** 279–290 (2012).

302. Boothby, T. C., Zipper, R. S., Van der Weele, C. M. & Wolniak, S. M. Removal of Retained Introns Regulates Translation in the Rapidly Developing Gametophyte of Marsilea vestita. *Dev. Cell* **24,** 517–529 (2013).

303. Bentley, D. L. Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* **15,** 163–175 (2014).

304. Alló, M. *et al.* Chromatin and alternative splicing. *Cold Spring Harb. Symp. Quant. Biol.* **75,** 103–11 (2010).

305. Luco, R. F. *et al.* Epigenetics in alternative pre-mRNA splicing. *Cell* **144,** 16–26 (2011).

306. Syed, N. H., Kalyna, M., Marquez, Y., Barta, A. & Brown, J. W. S. Alternative splicing in plants – coming of age. *Trends Plant Sci.* **17,** 616–623 (2012).

307. Meyer, K., Koester, T. & Staiger, D. Pre-mRNA Splicing in Plants: In Vivo Functions of RNA-Binding Proteins Implicated in the Splicing Process. *Biomolecules* **5,** 1717–1740 (2015).

308. Weighardt, F., Biamonti, G. & Riva, S. The roles of heterogeneous nuclear ribonucleoproteins (hnRNP) in RNA metabolism. *BioEssays* **18,** 747–756 (1996).

309. Schaal, T. D. & Maniatis, T. Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol. Cell. Biol.* **19,** 1705–19 (1999).

310. Lopato, S., Waigmann, E. & Barta, A. Characterization of a novel arginine/serine-rich splicing factor in Arabidopsis. *Plant Cell* **8,** 2255–2264 (1996).

311. Lopato, S. *et al.* atSRp30, one of two SF2/ASF-like proteins from Arabidopsis thaliana, regulates splicing of specific plant genes. *Genes Dev.* **13,** 987–1001 (1999).

312. Lazar, G. & Goodman, H. M. The Arabidopsis splicing factor SR1 is regulated by alternative splicing. *Plant Mol. Biol.* **42,** 571–81 (2000).

313. Palusa, S. G., Ali, G. S. & Reddy, A. S. N. Alternative splicing of pre-mRNAs of Arabidopsis serine/arginine-rich proteins: regulation by hormones and stresses. *Plant J.* **49,** 1091–1107 (2007).

314. Palusa, S. G. & Reddy, A. S. N. Extensive coupling of alternative splicing of pre-mRNAs of serine/arginine (SR) genes with nonsense-mediated decay. *New Phytol.* **185,** 83–89 (2010).

315. Kalyna, M. *et al.* Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. *Nucleic Acids Res.* **40,** 2454–69 (2012).

316. Kruszka, K. *et al.* Transcriptionally and post-transcriptionally regulated microRNAs in heat stress response in barley. *J. Exp. Bot.* **65,** 6123–6135 (2014).

317. Doolittle, W. F. Genes in pieces: were they ever together? *Nature* **272,** 581–582 (1978).

318. Cavalier-Smith, T. Selfish DNA and the origin of introns. *Nature* **315,** 283–4 (1984).

319. Cavalier-Smith, T. Intron phylogeny: a new hypothesis. *Trends Genet.* (1991). doi:10.1016/0168-9525(91)90377-3

320. Jeffares, D. C., Poole, A. M. & Penny, D. Relics from the RNA world. *J. Mol. Evol.* **46,** 18–36 (1998).

321. Poole, A. M., Jeffares, D. C. & Penny, D. The path from the RNA world. *J. Mol. Evol.* **46,** 1–17 (1998).

322. Rogozin, I. B., Wolf, Y. I., Sorokin, A. V, Mirkin, B. G. & Koonin, E. V. Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. *Curr. Biol.* **13,** 1512–1517 (2003).

323. Roy, S. W. Recent evidence for the exon theory of genes. *Genetica* **118,** 251–66 (2003).

324. Koonin, E. V. The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? *Biol. Direct* **1,** 22 (2006).

325. Giroux, M. J. *et al.* De novo synthesis of an intron by the maize transposable element Dissociation. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 12150–4 (1994).

326. Logsdon, J. M. *et al.* Seven newly discovered intron positions in the triose-phosphate isomerase gene: evidence for the introns-late theory. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 8507–11 (1995).

327. Hankeln, T., Friedl, H., Ebersberger, I., Martin, J. & Schmidt, E. R. A variable intron distribution in globin genes of Chironomus: evidence for recent intron gain. *Gene* **205,** 151–160 (1997).

328. Frugoli, J. A., McPeek, M. A., Thomas, T. L. & McClung, C. R. Intron loss and gain during evolution of the catalase gene family in angiosperms. *Genetics* **149,** 355–65 (1998).

329. Tarrio, R., Rodríguez-Trelles, F. & Ayala, F. J. A new Drosophila spliceosomal intron position is common in plants. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 6580–3 (2003).

330. Coghlan, A. & Wolfe, K. H. Origins of recently gained introns in Caenorhabditis. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 11362–7 (2004).

331. Roy, S. W. & Gilbert, W. Rates of intron loss and gain: Implications for early eukaryotic evolution. *Proc. Natl. Acad. Sci.* **102,** 5773–5778 (2005).

332. William Roy, S. & Gilbert, W. The evolution of spliceosomal introns: patterns, puzzles and progress. *Nat. Rev. Genet.* **7,** 211–221 (2006).

333. Cavalier-Smith, T. Intron phylogeny: a new hypothesis. *Trends Genet.* **7,** 145–148 (1991).

334. Stoltzfus, A. On the Possibility of Constructive Neutral Evolution. *J. Mol. Evol.* **49,** 169–181 (1999).

335. Rogers, J. H. How were introns inserted into nuclear genes? *Trends Genet.* **5,** 213–216 (1989).

336. Sharp, P. A. On the Origin of RNA Splicing and lntrons. *Cell* **42,** 397–400 (1985).

337. Crick, F. Split genes and RNA splicing. *Science (80-.).* **204,** (1979).

338. Iwamoto, M., Maekawa, M., Saito, A., Higo, H. & Higo, K. Evolutionary relationship of plant catalase genes inferred from exon-intron structures: isozyme divergence after the separation of monocots and dicots. *TAG Theor. Appl. Genet.* **97,** 9–19 (1998).

339. Irimia, M. *et al.* Origin of introns by 'intronization' of exonic sequences. *Trends Genet.* **24,** 378–81 (2008).

340. Fink, G. R. Pseudogenes in yeast? *Cell* **49,** 5–6 (1987).

341. Kent, W. J. & Zahler, A. M. Conservation, regulation, synteny, and introns in a large-scale C. briggsae-C. elegans genomic alignment. *Genome Res.* **10,** 1115–25 (2000).

342. Rogozin, I. B., Lyons-Weiler, J. & Koonin, E. V. Intron sliding in conserved gene families. *Trends Genet.* **16,** 430–432 (2000).

343. Tarrío, R., Ayala, F. J. & Rodríguez-Trelles, F. Alternative splicing: a missing piece in the puzzle of intron gain. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 7223–8 (2008).

344. Simpson, A. G. B., MacQuarrie, E. K. & Roger, A. J. Eukaryotic evolution: Early origin of canonical introns. *Nature* **419,** 270–270 (2002).

345. Collins, L. & Penny, D. Complex spliceosomal organization ancestral to extant eukaryotes. *Mol. Biol. Evol.* **22,** 1053–1066 (2005).

346. Irimia, M. & Roy, S. W. Origin of spliceosomal introns and alternative splicing. *Cold Spring Harb. Perspect. Biol.* **6,** (2014).

347. Rogozin, I. B., Carmel, L., Csuros, M. & Koonin, E. V. Origin and evolution of spliceosomal introns. *Biol. Direct* **7,** 11 (2012).

348. Ast, G. How did alternative splicing evolve? *Nat. Rev. Genet.* **5,** 773–782 (2004).

349. Irimia, M., Rukov, J., Penny, D. & Roy, S. Functional and evolutionary analysis of alternatively spliced genes is consistent with an early eukaryotic origin of alternative splicing. *BMC Evol. Biol.* **7,** 188 (2007).

350. Schwartz, S. H. *et al.* Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. *Genome Res.* **18,** 88–103 (2008).

351. Keren, H., Lev-Maor, G. & Ast, G. Alternative splicing and evolution: diversification, exon definition and function. *Nat. Rev. Genet.* **11,** 345–355 (2010).

352. Amani, J., Kazemi, R., Abbasi, A. R. & Salmanian, A. H. Short Communication A simple and rapid leaf genomic DNA extraction method for polymerase chain reaction analysis. *Iran. J. Biotechnol.* **9,** 69–71 (2011).

353. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLoS One* **6,** e16765 (2011).

354. Bent, A. Arabidopsis thaliana floral dip transformation method. *Methods Mol. Biol.* **343,** 87–103 (2006).

355. Simon, S. A. *et al.* Short-read sequencing technologies for transcriptional analyses. *Annu Rev Plant Biol* **60,** 305–333 (2009).

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

357. Gomes, C. P. C. *et al.* A Review of Computational Tools in microRNA Discovery. *Front. Genet.* **4,** 81 (2013).

358. Akhtar, M. M., Micolucci, L., Islam, M. S., Olivieri, F. & Procopio, A. D. Bioinformatic tools for microRNA dissection. *Nucleic Acids Res.* gkv1221-(2015). doi:10.1093/nar/gkv1221

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

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

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

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

363. German, M. A., Luo, S., Schroth, G., Meyers, B. C. & Green, P. J. 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 (2009).

364. Kleftogiannis, D. *et al.* Where we stand, where we are moving: Surveying computational techniques for identifying miRNA genes and uncovering their

regulatory role. *J. Biomed. Inform.* **46,** 563–573 (2013).

365. Raabe, C. A., Tang, T.-H., Brosius, J. & Rozhdestvensky, T. S. Biases in small RNA deep sequencing data. *Nucleic Acids Res.* **42,** 1414–26 (2014).

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

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

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

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

370. Moxon, S. *et al.* Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. *Genome Res.* **18,** 1602–9 (2008).

371. Thakur, V. *et al.* Characterization of statistical features for plant microRNA prediction. *BMC Genomics* **12,** 108 (2011).

372. Baksa, I. *et al.* Identification of Nicotiana benthamiana microRNAs and their targets using high throughput sequencing and degradome analysis. *BMC Genomics* **16,** 1025 (2015).

373. Refulio-Rodriguez, N. F. & Olmstead, R. G. Phylogeny of Lamiidae. *Am. J. Bot.* **101,** 287–299 (2014).

374. Denoeud, F. *et al.* The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science (80-.).* **345,** 1181–1184 (2014).

375. Hellsten, U. *et al.* Fine-scale variation in meiotic recombination in Mimulus inferred from population shotgun sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 19478–82 (2013).

376. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6,** 26 (2011).

377. Mohorianu, I. *et al.* Profiling of short RNAs during fleshy fruit development reveals stage-specific sRNAome expression patterns. *Plant J.* **67,** 232–46 (2011).

378. Wikstrom, N., Savolainen, V. & Chase, M. W. Evolution of the angiosperms: calibrating the family tree. *Proc. R. Soc. B Biol. Sci.* **268,** 2211–2220 (2001).

379. Roy, S. W. & Penny, D. Patterns of Intron Loss and Gain in Plants: Intron Loss-Dominated Evolution and Genome-Wide Comparison of O. sativa and A. thaliana. *Mol. Biol. Evol.* **24,** 171–181 (2006).

380. Takeda, A., Iwasaki, S., Watanabe, T., Utsumi, M. & Watanabe, Y. The Mechanism Selecting the Guide Strand from Small RNA Duplexes is Different Among Argonaute Proteins. *Plant Cell Physiol.* **49,** 493–500 (2008).

381. Thieme, C. J., Schudoma, C., May, P., Walther, D. & Schneeberger, K. Give it AGO: the search for miRNA-Argonaute sorting signals in Arabidopsis thaliana indicates a relevance of sequence positions other than the 5 - position alone. (2012). doi:10.3389/fpls.2012.00272

382. Tucker, M. R. *et al.* Somatic small RNA pathways promote the mitotic events

of megagametogenesis during female reproductive development in Arabidopsis. *Development* **139,** 1399–1404 (2012).

383. Bai, M. *et al.* Genome-wide identification of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families and their expression analyses in response to viral infection and abiotic stresses in Solanum lycopersicum. *Gene* **501,** 52–62 (2012).

384. Schapire, A. L. *et al.* Construction of Specific Parallel Amplification of RNA Ends (SPARE) libraries for the systematic identification of plant microRNA processing intermediates. *Methods* **67,** 36–44 (2013).

385. Iida, K. *et al.* Genome-wide analysis of alternative pre-mRNA splicing in Arabidopsis thaliana based on full-length cDNA sequences. *Nucleic Acids Res.* **32,** 5096–5103 (2004).

386. Ner-Gaon, H. *et al.* Intron retention is a major phenomenon in alternative splicing in Arabidopsis. *Plant J.* **39,** 877–885 (2004).

387. Haas, B. J. *et al.* Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. doi:10.1093/nar/gkg770

388. Wheeler, G., Valoczi, A., Havelda, Z. & Dalmay, T. *In Situ* Detection of Animal and Plant MicroRNAs. *DNA Cell Biol.* **26,** 251–255 (2007).

389. Dhir, A. & Proudfoot, N. J. Feed backwards model for microRNA processing and splicing in plants. *Nat. Publ. Gr.* **14,** (2013).

390. Zhang, S., Liu, Y. & Yu, B. New insights into pri-miRNA processing and accumulation in plants. *Wiley Interdisciplinary Reviews: RNA* **6,** 533–545 (2015).

391. Smith, N. A. *et al.* Total silencing by intron-spliced hairpin RNAs. *Nature* **407,** 319–20 (2000).

392. Varsha Wesley, S. *et al.* Construct design for efficient, effective and high-throughput gene silencing in plants. *Plant J.* **27,** 581–590 (2001).

393. Lechtenberg, B., Schubert, D., Forsbach, A., Gils, M. & Schmidt, R. Neither inverted repeat T-DNA configurations nor arrangements of tandemly repeated transgenes are sufficient to trigger transgene silencing. *Plant J.* **34,** 507–17 (2003).

394. Weiler, K. S. & Wakimoto, B. T. Heterochromatin and Gene Expression in Drosophila. *Annu. Rev. Genet.* **29,** 577–605 (1995).

395. Matzke, a J. & Matzke, M. a. Position effects and epigenetic silencing of plant transgenes. *Curr. Opin. Plant Biol.* **1,** 142–148 (1998).

396. Wong, J. J.-L., Au, A. Y. M., Ritchie, W. & Rasko, J. E. J. Intron retention in mRNA: No longer nonsense: Known and putative roles of intron retention in normal and disease biology. *BioEssays* **38,** 41–49 (2016).

397. Filichkin, S. A. *et al.* Environmental Stresses Modulate Abundance and Timing of Alternatively Spliced Circadian Transcripts in Arabidopsis. *Mol. Plant* **8,** 207–227 (2015).

398. Bohmert, K. *et al.* AGO1 defines a novel locus of Arabidopsis controlling leaf development. *EMBO J.* **17,** 170–80 (1998).

399. Telfer, A. & Poethig, R. S. HASTY: a gene that regulates the timing of shoot

maturation in Arabidopsis thaliana. *Development* **125,** 1889–98 (1998).

400. Jacobsen, S. E., Running, M. P. & Meyerowitz, E. M. Disruption of an RNA helicase/RNAse III gene in Arabidopsis causes unregulated cell division in floral meristems. *Development* **126,** 5231–43 (1999).

401. Lu, C. & Fedoroff, N. A mutation in the Arabidopsis HYL1 gene encoding a dsRNA binding protein affects responses to abscisic acid, auxin, and cytokinin. *Plant Cell* **12,** 2351–2366 (2000).

402. Chen, X., Liu, J., Cheng, Y. & Jia, D. HEN1 functions pleiotropically in Arabidopsis development and acts in C function in the flower. *Development* **129,** 1085–94 (2002).

403. Sunkar, R. & Zhu, J.-K. Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. *Plant Cell* **16,** 2001–19 (2004).

404. Sunkar, R., Li, Y. F. & Jagadeeswaran, G. Functions of microRNAs in plant stress responses. *Trends in Plant Science* **17,** 196–203 (2012).

405. Li, C. & Zhang, B. MicroRNAs in Control of Plant Development. *J. Cell. Physiol.* **231,** 303–313 (2016).

406. Couzigou, J.-M. & Combier, J.-P. Plant microRNAs: key regulators of root architecture and biotic interactions. *New Phytol.* **212,** 22–35 (2016).

407. Allen, E. *et al.* Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana. *Nat. Genet.* **36,** 1282–1290 (2004).

408. Nozawa, M. *et al.* Evolutionary Transitions of MicroRNA-Target Pairs. *Genome Biol. Evol.* **8,** 1621–1633 (2016).

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

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

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

412. Murphy, M. E. P., Lindley, P. E. & Adman, E. T. Structural comparison of cupredoxin domains: Domain recycling to construct proteins with novel functions. *Protein Sci.* **6,** 761–770 (1997).

413. Svistoonoff, S. *et al.* Root tip contact with low-phosphate media reprograms plant root architecture. *Nat. Genet.* **39,** 792–796 (2007).

414. Müller, J. *et al.* Iron-Dependent Callose Deposition Adjusts Root Meristem Maintenance to Phosphate Availability. *Dev. Cell* **33,** 216–230 (2015).

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

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

417. REYMOND, M., SVISTOONOFF, S., LOUDET, O., NUSSAUME, L. & DESNOS, T. Identification of QTL controlling root growth response to

phosphate starvation in Arabidopsis thaliana. *Plant, Cell Environ.* **29,** 115–125 (2006).

418. Ticconi, C. A. *et al.* ER-resident proteins PDR2 and LPR1 mediate the developmental response of root meristems to phosphate availability. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 14174–9 (2009).

419. Ticconi, C. A., Delatorre, C. A., Lahner, B., Salt, D. E. & Abel, S. Arabidopsis pdr2 reveals a phosphate-sensitive checkpoint in root development. *Plant J.* **37,** 801–814 (2004).

420. Balzergue, C. *et al.* Low phosphate activates STOP1-ALMT1 to rapidly inhibit root cell elongation. *Nat. Commun.* (2017). doi:10.1038/ncomms15300

421. Dai, X. & Zhao, P. X. psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.* **39,** W155-9 (2011).

422. Fernandez-Pozo, N. *et al.* The Sol Genomics Network (SGN)--from genotype to phenotype to breeding. *Nucleic Acids Res.* **43,** D1036–D1041 (2015).

423. Huang, W., Xian, Z., Kang, X., Tang, N. & Li, Z. Genome-wide identification, phylogeny and expression analysis of GRAS gene family in tomato. *BMC Plant Biol.* **15,** 209 (2015).

424. Ori, N. *et al.* Regulation of LANCEOLATE by miR319 is required for compound-leaf development in tomato. *Nat. Genet.* **39,** 319 (2007).

425. Berardini, T. Z. *et al.* The arabidopsis information resource: Making and mining the ?gold standard? annotated reference plant genome. *genesis* **53,** 474–485 (2015).

426. Lopez-Gomollon, S., Mohorianu, I., Szittya, G., Moulton, V. & Dalmay, T. Diverse correlation patterns between microRNAs and their targets during tomato fruit development indicates different modes of microRNA actions. *Planta* **236,** 1875–1887 (2012).

427. Kawashima, C. G. *et al.* Sulphur starvation induces the expression of microRNA-395 and one of its target genes but in different cell types. *Plant J.* **57,** 313–321 (2009).

428. Brodersen, P. & Voinnet, O. Revisiting the principles of microRNA target recognition and mode of action. *Nat. Rev. Mol. Cell Biol.* **10,** 141–148 (2009).

429. Ming, R. *et al.* Genome of the long-living sacred lotus (Nelumbo nucifera Gaertn.). *Genome Biol.* **14,** R41 (2013).

430. Hoopes, J. T. & Dean, J. F. D. Ferroxidase activity in a laccase-like multicopper oxidase from Liriodendron tulipifera. *Plant Physiol. Biochem.* **42,** 27–33 (2004).

431. Ziegler, J. *et al.* Non-targeted profiling of semi-polar metabolites in Arabidopsis root exudates uncovers a role for coumarin secretion and lignification during the local response to phosphate limitation. *J. Exp. Bot.* **67,** 1421–1432 (2016).

432. Ros Barceló, A. Lignification in plant cell walls. *Int. Rev. Cytol.* **176,** 87–132 (1997).

433. Sattler, S. E. & Funnell-Harris, D. L. Modifying lignin to improve bioenergy feedstocks: strengthening the barrier against pathogens?†. *Front. Plant Sci.* **4,** 70 (2013).

434. Gall, H. *et al.* Cell Wall Metabolism in Response to Abiotic Stress. *Plants* **4,** 112–166 (2015).

435. Lagrimini, L. M., Vaughn, J., Erb, W. A. & Miller, S. A. Peroxidase Overproduction in Tomato: Wound-induced Polyphenol Deposition and Disease Resistance. *Biotechnol. HORTSCIENCE* **28,** 218–221 (1993).

436. Buckley, P. T. *et al.* Cytoplasmic Intron Sequence-Retaining Transcripts Can Be Dendritically Targeted via ID Element Retrotransposons. *Neuron* **69,** 877–884 (2011).

437. Bush, S. J., Chen, L., Tovar-Corona, J. M. & Urrutia, A. O. Alternative splicing and the evolution of phenotypic novelty. *Philos. Trans. R. Soc. B Biol. Sci.* **372,** 20150474 (2017).

438. Xie, M., Zhang, S. & Yu, B. microRNA biogenesis, degradation and activity in plants. *Cellular and Molecular Life Sciences* **72,** 87–99 (2015).

439. Roy, S. W. & Irimia, M. Mystery of intron gain: new data and new models. *Trends Genet.* **25,** 67–73 (2009).

440. Aoki, S. & Ito, M. Molecular Phylogeny of Nicotiana (Solanaceae) Based on the Nucleotide Sequence of the matK Gene. *Plant Biol.* **2,** 316–324 (2000).

441. Doganlar, S., Frary, A., Daunay, M. C., Lester, R. N. & Tanksley, S. D. A comparative genetic linkage map of eggplant (Solanum melongena) and its implications for genome evolution in the Solanaceae. *Genetics* **161,** 1697–1711 (2002).

442. Clarkson, J. J. *et al.* Phylogenetic relationships in Nicotiana (Solanaceae) inferred from multiple plastid DNA regions. *Mol. Phylogenet. Evol.* **33,** 75–90 (2004).

443. Olmstead, R. G. *et al.* A molecular phylogeny of the Solanaceae Richard. *Taxon* **57,** 1159–1181 (2008).

444. Eserman, L. A., Tiley, G. P., Jarret, R. L., Leebens-Mack, J. H. & Miller, R. E. Phylogenetics and diversification of morning glories (tribe Ipomoeeae, Convolvulaceae) based on whole plastome sequences. *Am. J. Bot.* **101,** 92–103 (2014).

445. Aversano, R. *et al.* The Solanum commersonii Genome Sequence Provides Insights into Adaptation to Stress Conditions and Genome Evolution of Wild Potato Relatives. *Plant Cell* **27,** 954–68 (2015).

446. Pease, J. B., Haak, D. C., Hahn, M. W. & Moyle, L. C. Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation. *PLOS Biol.* **14,** e1002379 (2016).

447. Bombarely, A. *et al.* Insight into the evolution of the Solanaceae from the parental genomes of Petunia hybrida. *Nat. Plants* **2,** 16074 (2016).

448. NCBI Resource Coordinators. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **45,** D12–D17 (2017).

449. Hirakawa, H. *et al.* Survey of genome sequences in a wild sweet potato, Ipomoea trifida (H. B. K.) G. Don. *DNA Res.* **22,** 171–179 (2015).

Appendix

MIRtop14 sequences and primers used in the different analyses performed.

Exon sequences are shaded in grey while introns are not shaded. Intron GT-AG consensus splicing site sequences are highlighted in yellow. miRNAtop14 is highlighted in red and miRNAtop14* is highlighted in green.

Forward primers sequences are indicated in orange characters. Reverse primers sequences correspond to the reverse complement of the sequences indicated in blue characters. Sequences of the primers used for genomic DNA analysis (fig. 3.2) are in bold and sequences of the primers used for pri-miRNA analysis (fig. 3.5.B) are underlined. Note that for *S. lycopersicum* and *P. axillaris*, the same primers were used for gDNA and pri-miRNA analysis. Note that for *I. nil*, pri-miRNA analysis was not performed.

Finally, in S. lycopersicum it is shown the whole sequence included in the constructs created to assess the influence of the intron in mature miRNAtop accumulation (see scheme of the three constructs in fig 4.3). The three different *MIRtop14* sequences included in the constructs correspond to the full sequence shown: 1) with intron, 2) without intron, 3) with intron but with the 5'SS site mutated from G/GT to C/CC. The reverse primer used is the same as the one used for gDNA and pri-miRNA analyses, while the forward primer is 17 nt upstream of the forward primer used previously for gDNA and pri-miRNA analyses. The sequences of both primers used are indicated in italics.

Solanum lycopersicum MIRtop14

5'*TAGGCCACTTATATTTG****TGAAAGAATAA*****ATATAATGTACTGCATC**ATATCAATAATCTTATCTATAA
AAGGAGTAAAATTTGCTACATTGAATATACCATAGAGGAAATTGAATTAATGAAGAAGTAGAAGATGA
CACTTTGTTGGTGACTTTGATCTCAAAAGAGTGCTTATCAATATTGTTTGTTTAATTTATTACGGTAT
GTTATTTGTCTTATTTTACTTTAGTAAACTATTTATTGAAACTTCTTTCAAAGATTAGTTTTTTTAGT
CGAAAGTTTTTTGAAAGCATATTTTATATTGAGCAAGAGGTAAGAATAAGATTTATATACATTTCATG
ACCTGCTTATGAAATCATACTGAATATGTTATTATTATTATTTATTTTAGCAAAATGCTCATGTAGTA
TATTGTATTGTACCTTAATTGAATGTTTGATGCATGTTTGTGGGGCGGTGAAGCAATATTTGATCCTA
TATATGAAATAGTTTTTTAGTTGAAGATCGTTCAGCTGATCGTCCTTGAATTTATATACCTTTTATAA
TATAAAAGTGTCAAATTTGTAATTTAAAAAAAATAATCACCATTCTTATATACTATTTGTAAGAATGA
TCCTTGTTATTTTTACTTGAACTTAGAAATGATTTAGATGATTTACTTTAAAATTTTAAACATATTGT
AGTCTTGCTTTTTTGCATTACATGCCTTTTTGCTCTAACCTTTGTTCAAGAAATCGTGTACAATGGAA
TTGTGAAACAATATTTGTAGGAATATCAAGCATTATTATCATCACTATATGGACACATGGCATTATAC
TCTTGGGACCAAAGTCACCAACAGAGTGTCAAATTTCTCATCATTCCGAGCTCAAACACAGCGACGGA
GCCAGAATTCTCATCGAAGAAAATATCAAAATATTTTCTTAATTAATGATCTCTATTGTTTTGTTGTT
TTTTAGTAGTGCCTTGTTATTTTCTTTGGAGAATGTAACCAGCTAGATGTAATCCCCACATATATGTA
ATATAATAAATCAATCCCCTACCTTTTGGGGGTATGTTTACTA***AGCTATGCAACATAAAGTAACACTT
TA***

Nicotiana benthamiana MIRtop14

5′GGACAAGATTTTACTACTTGTAGTACAATCAGTGTACTTGCTAGTATTGTTGTAAAAAATTGTACG
GTTGTGACATGGTAGAACTAATCTATAAAAGTAGTCGAATTTGCCCCACTCGATATGCTAAGGGGAAA
TTGAAGAAGAAGCAGAAGATGATACGTTGTTGGTGACTTTGGTCCCGAAAGAATGTTGGTTAGTATTA
ATTGTTTGTATACACGCACTTACGCCTAAAATATAGATTTTTCACTATATGTTTTCAATTTACATAAA
CAAAACCTTTTTTCCCCAAAAATGTGGTCCAACTCTTTTCTTAAACTCTTATATATACAAGAACACCC
ACGTAAGCTTTTGTATATAGCGGAAACATAACGCACCGTCAGATTAATTGTCTTGTTTCCTCATACAT
TTTTGTGTGGTCTATGGCATCAACCACTAATTTTCTCATTACCAAAATGGCTACCATTTTTACATATC
TTTTTCTGTGTTCCATAAACCGCAATTTTTATATTTTATTTCAATATACCTTAAACCAAATTAATAAA
TTATTCTTATGTAAAGTGTCGATAAGATTGTCCCTTTTTCTTTACAACAACTCAGTTTCCTTATTTGA
AAATGTTAAGATCAGTCAACTTCATCATTTCATGTGTTTTCTCCTTTCCGCTAATTACAAGTGCATTT
ACCAAGTTTTTTGTATTACATTCCCTCTGTGCTAAATTGATCAAATAGGTTATTTTGACGGACAACAT
TATTAAATTAGATTAACAAATTATGCGCATGAAAACCCAACTTTTTCAAAAAGGCAAACTTGCAGTCT
TGTATACGATGAAACTGTAGGACTTTAGCTTTTATTTACTGATAATACAATTAAATATTTACACTCCA
TTGCAATTTAACCTAGTGCATAAGGTTGCAGCTTGTGTTATTGTTTAGGTTAGTAGTTCAATTGTTTT
ATGTAGAGTTACCTGCAATTACATTTCAAGCGAGCAGATAATGTTGAAAATTTACATTATCAATGTAT
ATAGTTAAGGTCATTGTTGAATGCTACGAAAGAAGAAGCATCAGTGTATGGACACAGCATTAAACTTT
TGGGACCAAAGTCACCAACAGAGTCTCAAATTTCTCATCTCCAACATCCAACTGAGGGGAACAGCCAA
CAACTTTGTTTCTCAATGATTTCTATTGTTATGTTTTTGTATATTAGTGACGTTTTGTTCCACTTGGA
GAATGTAACATATAG

Petunia axillaris MIRtop14

5′GTCAAATTTGCTACACTTGAAATGCCAAGTGCAAACAGGAAGAAATCGAAGAAGATGACACTTTGC
TGGTGACTTTGGTCTCGAAAGAGTTTTCGTGCTAATTATTATTGTTATGACACGGCATGACATAACAC
TTTTTGGGACCAAAGTCACCAACATAGTATCAAATTTCTCATCCATCCGAGGAAACTCTACCTATGCAA
TTAGCTCGACTCCAGGTATATATAGATAGCTCTTAATATAAGCGAAATCAGGAGTTGAAGCTTATAAA
TTCAGGATATATTTAATAGTACAACATAATTAGAGATTGAAACATATATTTACATGTCAACTAGAAGG
TGCAAAAATAGTTAGGGAGTGAATTTTTTTTGCCAGCCGGGCATTTTAGAGATGAGGCTCAATTTCTC
TCAGTGATACACCCTGCAATCTCAGTATCCCTGCTTACTTGCCAACTTAGGTTCCAAGCAACAAGTTA
TGATGCTTCATAGAGTATGTCGAGAAATTGCTCTCCATGTGCATGTTAAGAGGCCAAAATGTTAGTCA
GGCATTAGAGCAGCTCAATTCATTCCCTTAGTCCTATAAACGATGTAATGAATACAAGGAAGCCGAAC
CCCAAAGTTAGTTAATTGTCTGCAATGTTAGTTAAGTGTCTGATGTAGCGGAGCAAGTGATCAGCACT
TGCCTGGTTCTGTAAAATGTAAATATTTTTTTGGTTGAATGAAGTTACTTTACCAAAAAAAGAAAAAG
ACGTTTGAGAACTC

Ipomoea nil MIRtop14

5′TCCTCCAACCTCAAAGCAAAATAATACATAGTACTAAATTAGTACTTCGGAGAGGCTTAGATAACA
ATGGAGAACAACAACTAGGCTTGGTAACAGTACGTACTACAATTGTTGGGAGAAGAAGAAGTAGCAGT
ACTACAATTGTGTACTAAATTAAAAGGAAAATATAATTTAAATATGTAACCCCAAGAAACTTTGTATA
AAAACCTCAGGGCTATTGTGTCTGGGAGATCACAGAAAAGCAAGAATGAATTGAGCATTTTATATATT
TAGAGGGCAAAAGTGAAGAAGATGGCAGTTTGTTGGTGACTTTGTATCCAAAACTGTGGGTCATGGTT
TCATCTCAGGTTTGAGTTTAAGATAGGGAATGTTTGGTTCATATTGCTTAGGACTTCTTTTTATATAC
TTTTTGGAATGGGAGGGGCAATGTTGTCATTTTTACTTCCAATATGATTTTGCTATAGGTATTAAAGT
ATTAGATTAGATTACTTTAGGGAAGAAAAAAAAAAAATCAATACTTTTATGTACACAACATTTTAACC
TTTTATTTTCAGTTAGAGCATTATGGACATTGTAGACTTTTTTTTTATTCTTTTATTTGAATGACCAG
CAAAATCTGTTGTCACTACTTAAGAGTGTGCATTAGGTAAATCGGCTAGAATTTTCTTATAAATTAGA
TATATATGAGGTGGAGTAAGATTTTAATTTGGTTTTCAATTTATGAAGTTGTGGGTTTGTAGAGTTGC
ATCCATTTCCAACACTATTTTGGGACCAAAGTCACCAACACACTGTCATCATTCTCAGCTTCAGAAGA
TCTCAATTAAGATACTGTACAATAAGACACGACCAGCCATTTCACAAACAACTCAGGGAACATTGAGA
TTATTTATGCTTAGTTTTTGCTTCATTCTTAATCATCATATTTTTATGCTTCTACTTAAGGTACATGT
TTTGTATTGTGATCCAGGCCGAATTGAGGGTTTCTTCTGCTTTAATTTGCTCAAGTGCTTGATTAGTT
CAGCTGGAGACTAGCTATTGAAGACGACAATATCAAGAAAATTTCAATTGTAAAACAAAGATGTTTAT
GTTTTGTTTGTCTTACTTTTTGGAACTCAAACAATATTGGACTTGTTATTGGGTCATGGTCCTCCCAA
GAGTTTAGACTGGTTTGTAGGTTTGGGTTTGTGGACTACCCATTATCTACATGGGTCCAATGTATATC
CAGATACATATCTACATATCTACCA