UNIVERSITY OF TROMSØ UiT

FACULTY OF BIOSCIENCES, FISHERIES AND ECONOMICS
DEPARTMENT OF ARCTIC AND MARINE BIOLOGY

Functional Analysis of microRNA 840 in *Arabidopsis*

Zhijian Zhong

BIO-3910 Master`s thesis in Biology

May 2011

Table of contents

Acknowledgements

First of all, with great appreciation and gratitude, I'd acknowledge my thanks to my principal supervisor, Prof. Kirsten Krause**.** It is due to her patient instruction that I was able to complete the thesis in the end. In my master thesis project, from the research topic selection, plan of experiments establishment, result analysis, to the composition of the thesis, every detail has embodied the endeavor and wisdom of my supervisor A. Furthermore, I have benefitted a lot from her rigorous and realistic academic attitude and the working style of excellence.

My heartfelt thanks then must go to my co-supervisor Janina Fuss (Ph.D. Cand.). The huge amount of specific guidance from her has accompanied me all the way, from scientific knowledge, the experimental operation and analysis to the writing of the thesis. Her extensive knowledge and careful help is one of the most important factors for the accomplishment of the thesis.

I am also grateful to all the people in the Molecular Environments Research Group at the Department of Arctic and Marine Biology, UiT. Everyone in the group offers selfless help and support to each other in this energetic and warm family. Special thanks are given to Bernd Ketelsen (Ph.D. Cand.), Dr. Ullrich Hermann, Dr. Yin Lan, Alena Didriksen (MSc.), Stian Olsen and Leidulf Lund, for their numerous assistance and encouragement. Last, but not at all least, thanks for our former group member Stephanie Hainbuch who prepared the vectors and transgenic plants for this master project.

Also greatly indebted to my good friends Dr. Haowei Chen, Dr. Chunlin Shi, Dr. Xiaohui Zhang and Guangjin Ding whose advice and help are very useful for my courses learning, master project conducting and thesis writing.

I also wish to express my deep appreciation to all the experts and professors who review my thesis and attend the thesis defense for their time and consideration.

Finally, I would extend my sincere thanks to my family for their care and help. Their constant support behind makes me concentrate on my study and work, and also own enough courage to accept challenges and solve problems.

Abbreviation

A	Adenine
Amp	Ampicillin
ATMT	*Agrobacterium-tumefaciens*-mediated-transformation
ATP	Adenosine 5'-triphosphatase
Bp	Base pair
C	Cytosine
D	Daltons
EB	Ethidium bromide
EDTA	Ethylenediamine tetra acetic acid
dsDNA	double stranded Deoxyribonucleic Acid
g	Gram
G	Guanine
Gent	Gentamicin
h	Hour
IPTG	Isopropy-β-D-thigalatopyranoside
kb	Kilobases (pair)
Kan	kanamycin
L	Liter
LB	Luria-Bertani
M	Mole
mg	Milligram
min	Minute
ml	Milliliter
miRNAs	microRNAs
mRNA	Messenger ribonucleic acid
nat-siRNA	natural antisense transcript siRNA
nt	nucleotide
OD	Optical density
PCR	Polymerase chain reaction
piRNA	PIWI-interacting RNAs
PPR	pentatricopeptide repeat protein
RNAi	RNA Interference
rcf	relative centrifugal force
rpm	Revolutions per minute
RT	Room Temperature
siRNAs	short interfering RNAs
SsDNA	single stranded Deoxyribonucleic Acid
TAE	Tris Acetate EDTA
Ta-siRNA	Trans-acting siRNA
Tris	Tri-[hydroxymethyl]methyl-2-aminoethanesulfon
µg	Microgram
µL	Microliter
U	Uracil
UTR	untranslated region

Abstract

Three mechanisms have been implicated for plant microRNAs (miRNAs) to regulate gene expression, i.e. directing target RNA cleavage, transcriptional silencing and translational repression. A great number of target genes of plant miRNAs were predicted using different algorithms and verified through experimental methods.

In the present work, one miRNA from *Arabidopsis*, microRNA 840 (miR840) was initially analyzed through the high-through put sequencing and bioinformatic method (Rajagopalan *et al.*, 2006). MiR840 expresses from the complementary strand of its predicted target gene, *AtWhirly3*. *AtWhirly3* encodes a homolog of the potato transcriptional regulator p24. MiR840 is located between 2 genes, At2g02740 (Whirly 3, Why 3) and At2g02750 (Pentatricopeptide Repeat Protein, PPR). Whirly 3 protein belongs to the Whirly protein family which is one of the main interesting topics in our group. Understanding the potential regulative function of miR840 related to Whirly 3 gene is important to deeply analyze the Whirly 3 protein.

Searching the TAIR (http://www.arabidopsis.org) and TIGR (http://plantta.jcvi.org), the two databases give different annotation versions for this *AtWhirly3* gene, differing at the length of the 3' untranslated region. According to the annotation from TAIR, the target of miR840 is only at the 3' untranslated region (3' UTR) of Why 3 gene, so this means miR840 only down-regulate the Why 3 gene. But the information from TIGR gives another annotation, Why 3 gene (At2g02740) and its neighbor gene, a PPR gene (At2g02750) overlap with their 3' UTR where the target site of the miR840 is. This means the miR840 may also be able to down-regulate both the Why 3 and PPR gene, depending on what annotation you take. Therefore experimental analysis is necessary to determine the miR840 target. My master project here is to perform the biological experiments for analyzing the function of miR840.

To verify the target genes and study the biological function of miR840. 4 vectors have been firstly constructed and then introduced into *Arabidopsis thaliana* by Agrobacterium-mediated transformation (1: overexpressed precursors of miR840; 2: overexpressed mutated miR840 target for directing the miR840 lost its normal function; 3: overexpressed Why 3; 4: overexpressed PPR). The mutated miR840 was constructed by the mechanism of target mimicry which through artificial insertion of a mismatch-loop into the cleavage site at the miRNAs target, for making mutated miRNAs targets. The mutated target sites will direct the miRNAs cannot cleave their targets (miRNAs lost their normal function).

The expression level of the transgenic plants will be analyzed, the changes of the expression levels observed in the transgenic plant lines could imply the possible role of miR840.

The subcellular location of the PPR (At2g02750) protein which is still unknown yet, so the bioinformatic method was used to predicted its subcellular location.

1. INTRODUCTION

1.1 Overview of small RNA directed gene regulation.

In almost all eukaryotes, small non-coding RNAs are endogenous 21-24 nucleotide (nt) RNAs which can play common and crucial roles in regulating gene expressions by targeting specific mRNAs. These small RNAs are generated via Dicer-dependent or independent biogenesis pathways, and then are incorporated into an Argonaute family (AGO) protein to guide sequence specific gene silencing either transcriptionally or post-trancriptionally. Small RNAs can bind to target mRNA by complementary base-pairing, which triggers target mRNA degradation or translational

inhibition, or even result in chromatin modification.

1.1.1 The brief history of small RNA research

The first research on small RNA was carried out by Lee *et al* in 1993. They discovered that *lin-4*, which was known for controlling the timing of *Caenorhabditis elegans* (*C. elegans*) larval development does not code for a protein but generates a pair of small RNAs (Lee *et al.,* 1993). One is approximately 22 nt in length and the other is 61nt. At the same time, Gary Ruvkun's lab demonstrated that the 22 nt small RNA had antisense complementarity to a repeated sequence element in 3' untranslated region (UTR) of the *lin-14* mRNA, which encoded a protein involved in development (Wightman *et al.,* 1993). The *lin-4* RNA regulated *lin-14* expression by reducing the amount of LIN-14 protein without changing the levels of *lin-14* mRNA. It was shown for the first time that endogenous small non-coding RNAs could possibly interact with a messenger RNA (mRNA) and therefore influence the levels of the protein produced from the mRNA post-transcriptionally (Wightman *et al.,* 1993). In 1999, Baulcombe's group discovered the presence of short RNAs of 21-23 nucleotides (now called microRNA) derived from both the sense and antisense strand that were also complementary to the silenced mRNA. For discoveries that revealed an unanticipated world of small RNAs that regulate gene function in plants and animals, Victor Ambros, David Baulcombe, and Gary Ruvkun shared the 2008 Albert Lasker Award for Basic Medical Research.

In 1998 a new and revolutionary technique, RNAi, to silence endogenous genes in *C. elegans* by injected long single-stranded sense or antisense RNA (300-1000 bases), was created by Fire, Mello and their colleges (Fire *et al.,* 1998). It was suggested that the process is mediated via a simple complementary hybridization mechanism between the exogenous RNA and endogenous mRNA. Therefore, the mechanism was called RNA interference (RNAi). For the great discovery from Fire and Mello in the silencing phenomenon, they were awarded with the Nobel Prize in Physiology and

Medicine in 2006.

Since *lin-4* was described in *C. elegans*, scientists didn't find any more *lin-4* like genes for quite a long time. Gary Ruvkun lab reported the second endogenous small RNA (*let-7*) in 2000 (Reinhart *et al.*, 2000). Like *lin-4*, *let-7* is 21nt in length, which also can bind to its target mRNA *lin-41*. Homologs of the *let-7* gene were soon identified, then it was clear that *let-7* is conserved among several species, including vertebrates, ascidians, hemichordates, molluscs, annelids and arthropods. According to these new significant achievements, scientists started to pay more attention to find the new small RNAs and began to clone and sequence small RNAs. A short time later the works of three groups were published in 2001, all of their results showed that there were much more endogenous small RNAs (now called microRNAs) than originally thought. They found and cloned several more conserved microRNAs, not only from *C. elegans*, but also from other invertebrates and vertebrates (Lagos-Quintana *et al.*, 2001; Lau *et al.*, 2001; Lee and Ambros., 2001).

1.1.2 Classes of small RNAs and RNA silencing

During the last years several small RNA classes were described in plants. Especially deep-sequencing approaches revealed for the first time the complexity of small RNA classes. Based on their precursor structures, biogenesis pathway and functional mechanism, small RNAs are classified as miRNAs and siRNAs. Next to miRNAs and siRNAs, also PIWI-interacting RNAs (piRNAs, forms RNA-protein complexes through interactions with piwi proteins which are are highly conserved, and present in both plants and animals), primal small RNAs (priRNAs), transacting short interfering RNAs (ta-siRNAs), natural antisense short interfering RNAs (nat-siRNAs) and repeat associated short interfering RNAs (ra-siRNAs) do exist. (Ghildiyal and Zamore., 2009; Halic and Moazed., 2010)

Both miRNA and siRNA may need the incorporation of RISC (RNA-induced

silencing complex) to silence target gene. The master member in this complex is AGO protein, which contains two key domains, PAZ and PIWI. The crystalline structure of the PAZ domain shows that the PAZ domain forms a binding-domain suggestive of pocket, which can encircle approximate 21 nt small RNA (Song *et al.,* 2003). The PIWI domain can direct the cleavage to the recognized RNA sequence. There are ten members of the AGO family in Arabidopsis with distinct functions, and the AGO1 incorporated gene silencing will be discussed in more detail later in this chapter (Brodersen *et al.,* 2008).

miRNAs in plants and animals arise from single-stranded RNA precursors with stem-loop structures, which are processed by the RNase III type enzyme, Dicer or Dicer-like proteins (Bartel., 2004), the stem-loop is the most common structure of mature microRNA. While most siRNAs (20-24nt in length) are produced from double-stranded RNAs by Dicers, some siRNAs may even be generated through de novo biogenesis with RNA-dependent RNA Polymerase (RDRP) in *C. elegans*, such as piRNA (Ghildiyal and Zamore., 2009). The biogenesis pathway requires Argonaute family protein, which contain two characteristic domains, the PIWI and the PAZ domains which are key domains of AGO protein. There are ten members of the AGO family in Arabidopsis with distinct functions. AGO protein is the master member of RISC (RNA-induced silencing complex) which are both miRNA and siRNA need to incorporate for silencing target gene. The PIWI domain shows similarities with RNaseH family members and the PAZ domain can bind single-strand RNA. Because of piRNA interacting with the Piwi domain, it was named PIWI-interacting RNAs (Ghildiyal and Zamore., 2009).

In the yeast *Schizosaccharomyces pombe*, Danesh Moazed's lab found a kind of special small RNA, which was named primal small RNA, or short priRNAs. PriRNA can bind to Ago1 protein and then produce many small interfering RNA（Halic and Moazed., 2010).

Trans-acting siRNA (Ta-siRNAs) also belong to siRNA and are approximately 21nt in length. As the name indicates, Ta-siRNAs negatively regulate mRNA in a distant locus. 24 nt long nat-siRNAs derive from genes whose natural *cis* anti-sense transcripts partially overlap**,** recognize the target gene and result in RNA mediated gene silencing (Borsani *et al.*, 2005). Ra-siRNAs are also called heterochromatic siRNAs, which target transposons, retroelements or trigger DNA methylation (Xie *et al*., 2004).

The previous studies of the function and regulation mechanisms of small RNAs led to a better understanding that the most important proteins or enzymes required were coherent in their establishment and different regulation level. More evidences indicated that small RNAs in distinct pathways may interact with each other in the process of regulating gene silencing (Jones-Rhoades *et al.*, 2006).

It is generally believe that both of the accumulation of aberrant RNA generates by endogenous transcription and insertion of the repeated transgene sequence into the genome can bring out the double-stranded RNA, and then finally initialize the silencing of the specific mRNA (Voinnet., 2009).

1.2 MicroRNAs in Plants

Because more and more plant miRNAs and their target genes were identified, and their functions also have been widely studied in recent years, there is a growing realization that miRNAs introduce a negative pattern in regulating gene expression. Plan researchers have realized that the miRNAs play a very important role in the many key plant cellular processes, such as proliferation, differentiation, development, etc. The following chapters contain overview of miRNAs in plants, including its genes in chromosome, biogenesis pathway, function and regulated target genes will be

described in detail.

1.2.1 microRNA gene

Taking advantage of direct cloning, genetic methods and bioinformatic approaches, many microRNAs have been discovered in nematodes, flies and humans. With a similar strategy, scientists have identified a large amount of plant miRNAs in Arabidopsis, which is a widely used model organism in plant research. Until 2004, the number of the annotation of potential miRNA genes in Arabidopsis was 118 (Griffiths-Jones S., 2004). These 118 genes can be grouped into 42 families, with each family producing identical or similar mature miRNAs. It was proved that 92 miRNA genes, which belong to 21 families, are conserved in Arabidopsis and some of them are highly conserved in other sequenced plant genomes, such as *Oryza sativa,* and *Populus trichocarpa* (Jones-Rhoades *et al.*, 2006). Within each family, the sequence of the mature miRNA is identical or highly conserved between members of the same miRNA family. But the sequence of the miRNA precursor, secondary structure, and length of the intervening "loop" region always are different.

1.2.2 miRNA biogenesis

In Arabidopsis, there are many proteins involved in the miRNA biogenesis pathway (**Figure 1**), which is divided into three steps: transcription of miRNA precursors, miRNA processing and export, miRNA incorporation into the RISC which is the core complex direct the gene silencing. More details are described below.

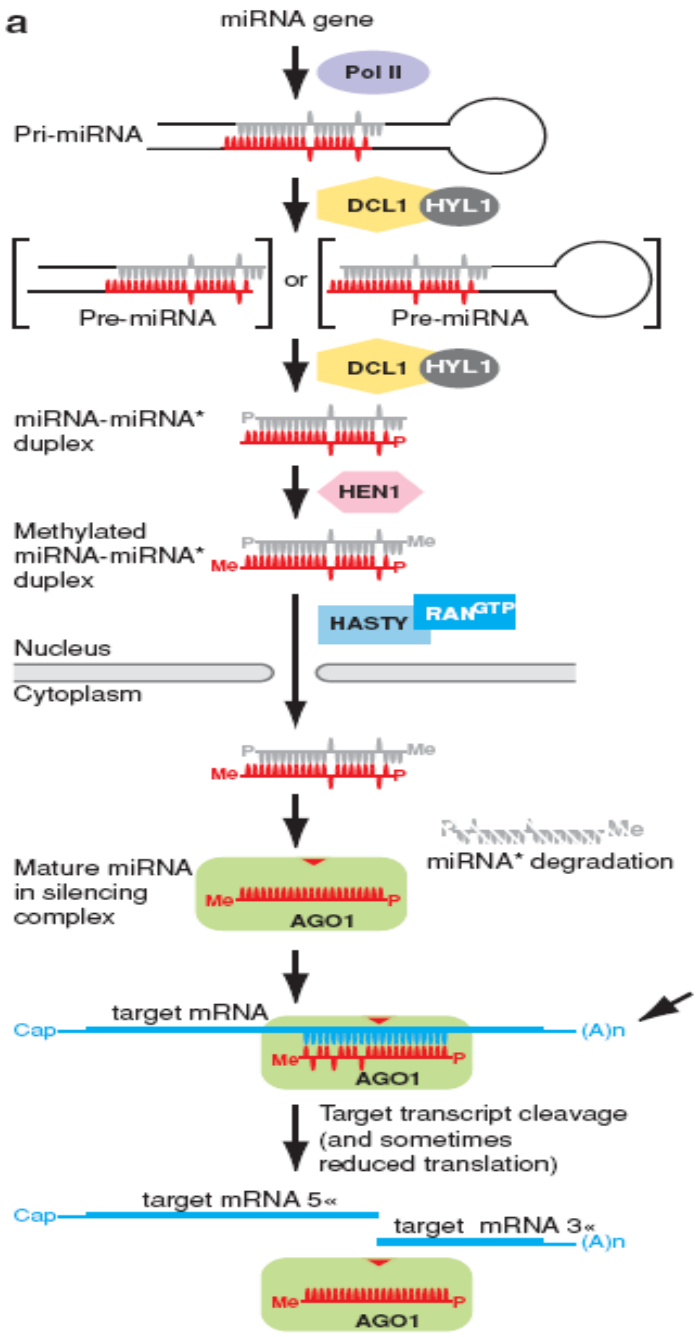


Figure 1. The Biogenesis of miRNAs in Arabidopsis（Jones-Rhoades *et al.*，*2006*）:

After the miRNA gene is transcribed, DCL1 processes the pri-miRNA, maybe together with HYL1 and other factors, to a miRNA:miRNA* duplex with 5' phosphates (P) and two-nucleotide 3' overhangs (Pre-miRNAs). The 3' end of the miRNA:miRNA* duplex is methylated (Me) by HEN1, probably within the nucleus. The miRNA is then exported to the cytoplasm by HST, presumably with the help of additional factors. The mature miRNA cooperates with a silencing complex that includes AGO1, and the miRNA∗ is degraded.

Complex maturation is depicted after nucleocytoplasmic export, but might occur before. Within the silencing complex, the miRNA is capable of targeting complementary RNAs for cleavage by AGO1, and perhaps also for translational repression.

1.2.2.1 Transcription of miRNA precursors

While most animal miRNAs appear to be processed from introns of protein-coding genes, plant miRNA are primarily found in genomic regions, which are driven by their own promoter and transcribed by a RNA polymerase II. Plant miRNA genes are occasionally clustered in the genome (Gudde *et al.,* 2005; Jones-Rhoades & Bartel, 2004) Many data indicated that plant miRNA primary transcripts (sometimes called pri-miRNAs) are longer than needed to encompass the miRNA stem-loops (Jones-Rhoades *et al.,* 2006). Therefore pri-miRNAs appear to be spliced, polyadenylated and capped during maturation.

The pri-miRNAs that are processed in the nucleus into 60-70nt, have a hairpin structure. While base-paired arms form the "stem", the unpaired flanking sequences result in a loop structure. The pri-miRNAs also carry a seven-methyl guanosine (m^7G) cap at the 5' end and a polyadenosine tail (polyA) at the 3′end. (Bartel., 2004; Jones-Rhoades *et al.,* 2006).

1.2.2.2 miRNA processing and export

In plants, Dicer-like proteins which are homologous to the animal DICER proteins, have an important role in the process of miRNA maturing. Dicer-like proteins can recognize the stem-loop structures in pri-miRNAs and generate a duplex containing two strands, termed miRNA and miRNA*. The Dicer-like proteins named DCL1 in Arabidopsis, assisted by another double-strand binding protein, HYL1 (Hiraguri A *et al*., 2005), convert miRNA precursors to miRNA /miRNA*duplexes in the nucleus. After that HUA ENHANCER 1 (HEN1) which is a methyltransferase adds the

2'-O-methyl modification at the 3'end of the duplex to protect the 3' ends from uridylation and associated destabilization (Yu *et al.* 2005). The modified miRNA/miRNA* duplex is thought to be transported to the cytoplasm by transporter HST (HASTY, hasty mutants have reduced accumulation of most miRNAs) (Bollman *et al.*, 2003). Most miRNA molecules enter the cytoplasm through the nuclear pores, which requires other accessory factors such as Ran-GTPase to provide energy. As pre-miRNAs have a short-life *in vivo*, it's still unclear whether miRNAs are already single stranded before transport to the cytoplasm by HST (Jones-Rhoades et al., 2006).

1.2.2.3 miRNA incorporation into the silencing complex (RISC)

The miRNA/miRNA* duplexes are incorporated into an RNA-induced silencing complex such as RISC and involved in the recognition of target mRNA. AGO1 which forms the catalytic core of RISC is a member of the Argonaute protein family. The loading of one strand of the miRNA/miRNA* duplexes into AGO1 guide RISC to down-regulate gene expression by mRNA cleavage or translational repression. So this strand is also named guide miRNA, and the other one (miRNA*) is described as passenger miRNA which was removed before the activation of RISC. It is indicated that the guide miRNA strand of this duplex accumulates to much higher levels *in vivo* than does the miRNA* (Reinhart BJ *et al.*， 2002), the probable reason may be that the guide miRNA strand is loaded into RISC, whereas the passenger strand is excluded from the complex and subsequently confronting degradation.

The first base in the 5' end of the miRNA is crucial for association with Argonaute proteins, and different AGO proteins show diverse favoritism to the first base. Most plant miRNAs contain a 5' terminal U which can be loaded into the PAZ domain of AGO1. Notably, miRNAs and their targets are highly complementary in most cases, the maximum mismatches of the miRNAs and their targets are 5 mismatches (Schwab *et.al*, 2005).

1.2.3 miRNA functional mechanism

miRNAs regulate gene expression through target cleavage and/or translational inhibition. A key factor determining the functional mechanism of miRNAs is the complementarity between miRNAs and their targets (Bartel., 2004). It was observed that a perfect match between miRNAs and their targets promotes target cleavage, while central mismatch enables translational inhibition. This theory can explain why plant miRNAs, which are highly complementary to their targets, function predominantly by target-cleavage.

1.2.3.1 miRNA-directed RNA cleavage

The best-understood mode of regulation by miRNA is directing RNA cleavage. miRNAs regulate gene expression through target cleavage is and/or translational inhibition. A key factor determining the functional mechanism of miRNAs is the complementarity between miRNAs and their targets (Bartel., 2004). In summary, the small RNAs direct the Argonaute component of RISC to a specific complementary molecule, then the cleavage is initialized. When a miRNA guides cleavage, the cut site is at precisely between the nucleotides pairing to residues 10 and 11 of the miRNA (Hutvágner and Zamore., 2002). After cleavage of the mRNA, the miRNA which is finally released, remains intact and can guide the RISC to recognize and cleave additional transcripts. The two most important conditions for miRNA function normally are: they get sufficient complementarity to their target mRNAs, and the cutting site is available.

1.2.3.2 miRNA-directed transcriptional silencing (miRdDM)

In addition to post-transcriptional silencing, non-coding RNAs (ncRNAs), including miRNAs also play an important role in the transcriptional silencing in animals and

plants. In plants, some of miRNAs guide sequence specific DNA methylation through a phenomenon referred to as miRNA-directed DNA methylation (miRdDM) (Bao *et al.*, 2004). MiRdDM and/or the maintenance of DNA methylation pathways are usually associated with transcriptional silencing (TGS). Increasing evidence shows a role for plant miRNAs in guiding chromatin modifications, but the deep mechanism is still need to be explored.

1.2.3.3 miRNA-directed translational repression

It was observed recently that translational inhibition by miRNA is widely present in plants (Brodersen *et al*, 2008). Unlike in animals, our knowledge of the mechanisms governing translational inhibition by plant miRNAs has been rather limited. Among ten AGO proteins from *Arabidopsis*, AGO1 and AGO10 were proposed to be involved in translational inhibition (Brodersen *et al*, 2008). Both *ago1* and *ago10* mutant has some similar developmental defects, and the protein levels of several target increase dramatically.

1.2.4 Regulatory roles of plant microRNAs

1.2.4.1 Target genes of miRNA

In plants, many miRNA target gene products are regulating proteins, suggesting that miRNAs work as a network middleware and have powerful regulating ability. Most downstream genes of these regulatory factors play important roles in the process of plant development and differentiation.

Other predicted target genes encode ubiquitin conjugating enzymes, suggesting that miRNA may play a role in regulating protein stability. Jones-Rhoades *et al.* (2006) reported DCL1 and AGO1 were also miRNA targets, leading to the suggestion that

plant miRNAs can regulate their own biogenesis in turn. Hence it is further implied that better advanced and complicated mechanism under miRNA regulating plants gene expression need to be investigated.

1.2.4.2 Prediction of miRNA targets

Plant miRNAs have a high sequence complementarity to their target mRNAs, which provides the basis for target prediction and validation. In contrast to the situation in animals, where only a few targets were validated by experiments, many predicted miRNA targets in plants have been validated experimentally (Rhoades *et al.*, 2002).

As a pioneer work of plant miRNA target prediction, Rhoades *et al.* (2002) used a pattern search algorithm to predict Arabidopsis miRNA targets based on high sequence complementarity (Rhoades *et al.*, 2002). In this approach, the miRNA complemetarity was searched against the Arabidopsis genome with less than four mismatches, considering G:U base pairing as a mismatch. Using this method, 49 miRNA targets were predicted, of which many were validated successfully. Later, evidences were presented by some other authors that some plant miRNAs can actually target mRNA sequences up to mismatches (Palatnik *et al.*, 2003). From the study of miR319, researchers found that this miRNA could be detected for all down-regulated *TCP* genes, which encode the plant-specific transcription factors with a bHLH motif (Palatnik *et al.*, 2003). The binding site for miR319 in these *TCP* mRNAs showed up to five mismatches, considering G:U base pairing as a mismatch.

A more sophisticated approach that allows more mismatches and bulged nucleotides has been developed by comparing the conserved miRNA binding sites in homologous sequences of two different species (Arabidopsis and Rice), which means a miRNA binding site must be discovered in homologous sequences of both Arabidopsis and rice. Similarly, an approach proposed by Wang *et al.* (2004) applied a nucleotide alignment algorithm in which mismatches were given a lower penalty than a bulge. In

addition, a penalty for gap opening and gap extension was also included.

Microarray analysis also has been used for miRNA target prediction in plants. This approach, together with the structural analysis of validated miRNA:binding-site hybrids, has resulted in a set of rules that could be used for evaluation of a putative miRNA target. In general, the pairing in the 5' part of the miRNA is more important, with only one mismatch allowed in the regions corresponding the nucleotides 2 to 12, which includes the presumptive cleavage site between positions 10 and 11. By contrast, a mismatch loop containing up to two nucleotides could be tolerated in the 3'end of the miRNA, and a perfect match in this part would compensate the presence of up to two mismatches in the 5'end (**Figure 2.**) (Schwab *et al.*, 2005).

As a fruit of the increasing knowledge on miRNA targets structure and characteristics, currently there are several high quality databases for fairly accurate prediction of miRNAs targets. Simply through searching against the Internet, researchers now can easily do the prediction work at office in a timely manner. The next important step scientists are facing to is to validate the targets, especially those have significant functions in plant development and disease-resistance.

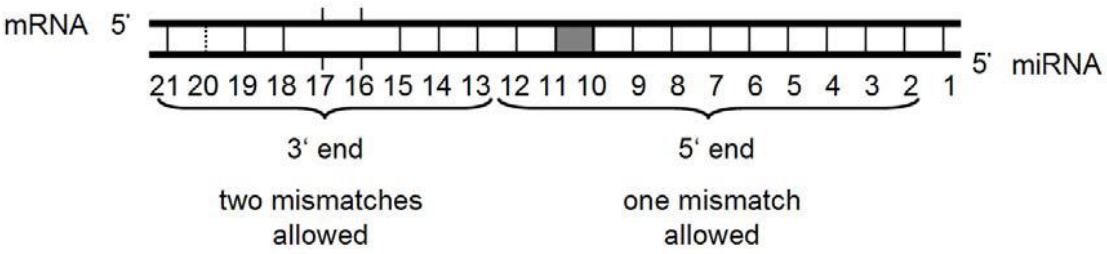


Figure 2. Characteristic miRNA binding site structure in plants. The maximal mismatches allowed in the 3'end of the miRNA is 2, whereas in the 5'end is only one, which means the 5' part of the miRNA is more important. The cleavage site between site nr. 10 and site nr.11 was labeled. (Schwab *et al.*, 2005)

1.2.4.3 Validation of miRNA targets

Unlike animal miRNAs, most plant miRNAs act like siRNAs, inducing the cleavage

of the mRNA target. This feature makes the validation of miRNAs in plants generally more straightforward than in animals. A commonly used method for accurate mapping of the point cleaved by the RISC is the modified version of 5' rapid amplification of cDNA ends (5'RACE). .In this method, an RNA adaptor is ligated to the 5'end of the cleavage product and the adaptor-ligated RNA is used as template for reverse transcriptase followed by PCR with gene specific nested primes. After cloning and sequencing of the PCR product, the position of the RISC-mediated cleavage can be identified (Llave *et al.,* 2002a). There are a large number of reports describing the validation of miRNA targets in this way (Palatnik *et al.*, 2003; Jones-Rhoades and Bartel., 2004; Mallory *et al.*, 2005; Rajagopalan *et al.*, 2006).

Llave *et al.* (2002) developed a transient A. *tumefaciens* infiltration based method for miRNA target validation. In this system, the cDNA of a target and the corresponding pre-miRNA should be cloned into a binary vector and then transiently expressed in Arabidopsis or *Nicotiana benthamiama* leaves. Through Northern blot, the cleavage products can be identified (Palatnik *et al.*, 2003; Wang *et al., 2*005).

Similar to the infiltration assay (Mallory *et al.*, 2004; Reyes and Chua, 2007), an alternative *in vitro* assay for detection of miRNA targets also were reported (Tang *et al*., 2003), in which the cDNA sequence of a miRNA target was cloned and used for *in vitro* transcription. The transcript is then mixed with standard wheat germ extracts that contain all the components of the miRNA silencing pathway. After incubation, the RNA could be isolated and analyzed by northern blotting.

Despite their usage in demonstrating miRNA-dependent cleavage of mRNA targets, all the methods described above could not provide evidences for the functional role of miRNAs. To analyze the effects of miRNA regulation in plants, the disruption of miRNA binding site was one of the best ways that have been successfully employed. By virtue of the site-directed mutagenesis technique, nucleotides in a miRNA binding site can be mutated by introducing silent mutations. This cDNA, which harbors

altered nucleotides in the miRNA binding site but does not result in changes in the amino acid sequence of the encoded protein, can be expressed under the control of an endogenous or constitutive promoter in transgenic plants, leading to the expression of a miRNA-resistant mRNA. The phenotypic effects observed in the transgenic lines will not only provide confidence for target validation, but also help to understand the functions of a miRNA and its target (Palatnik *et al.,* 2003; Bartel., 2004; Kidner and Martienssen., 2004; Parizotto *et al.,* 2004).

Another lately developed and very effective way for miRNA functional analysis was target mimicry. This method can help researchers to understand the regulation mechanism of miRNA and their target genes. This strategy is based on the change of the recognition between miRNA and its target mRNA. The key component involved in this system is the 'pseudosubstrate', which was generated by inserting an artificial mismatched loop through PCR into the target substrate of miRNA, at the predicted miRNA cleavage site. In transgenic plants that express a large amount of such pseudosubstrates, the miRNA we want to study, though still can bind to these pseudosubstrates, cannot cleave them because the pairing at cleavage site of the miRNA and its target was disrupted due to the cleavage site mutation. Clearly, with the help of this mechanism of target mimicry, a validation of the miRNA target and the function can be studied simultaneously and effectively (Franco-Zorrilla *et al.,* 2007).

1.2.5 Whirly protein family, Pentatricopeptide repeat proteins (PPR proteins) and microRNA 840 (miR840) in Arabidopsis

1.2.5.1 Whirly protein family (Whirly 1, 2, and 3)

The whirly protein family is a small family that contains only 3 proteins in Arabidopsis, Whirly 1 (Why 1), Whirly 2 (Why 2) and Whirly 3 (Why 3). Whirly proteins are transcription factors which are discovered only in the plant kingdom, and

bind to single-stranded DNA (Desveaux *et al.,* 2005). Previous research suggested that the Whirly family preferentially has effects on nucleic acid metabolism, such as regulating transcription and modulating the length of telomeres. In Arabidopsis, all members of the Whirly family are encoded by the nuclear genome and subsequently localize to organelles; Why 1 and Why 3 are targeted to plastids and Why2 is targeted to mitochondria (Krause, *et al,*. 2005; Maréchal, *et al* ,. 2008). Evidence indicated it that Why 1 relevant to chloroplast RNA metabolism in maize (Prikryl *et al.,* 2008), in addition to its function of transcriptional regulation (Desveaux *et al.* 2005). In maize, WHY1 protein can interact with DNA from throughout the plasmid and contribute to the accurate repair of plant organelle genomes. The crystallographic structure of the *Solanum tuberosum* Why 1 protein was beneficial to understand the role of Why 1. It was revealed that Whirly proteins form tetramers resembling whirligigs (Desveaux *et. al*., 2002). However, the sequence specificity of Whiry family members, as well as the precise mode of single-stranded nucleic acid binding, has so far remained unclear.

In order to understand how Whirly proteins function in the organelle genomes of Arabidopsis, Cappadocia *et al.* solved the crystal structure of St-WHY2$_{48-216}$, a close homolog of At-WHY2 that also localizes to mitochondria (Cappadocia *et al.,* 2010). Their data indicated that Why 2 binds ssDNA through a conserved mechanis. It was proposed that Whirly proteins could bind and protect resected DNA ends in a sequence-independent manner at break sites (Cappadocia *et al.,* 2010)

The research of the Why 3 has been less than Why 1 and Why 2 until now. In recent reports, researchers demonstrated that Why 3 (together with Why 1) could effect on plastid genome stability by the protection against illegitimate repeat-mediated recombination (Maréchal *et al.,* 2009) and the latter also related to accurate repair of plant organelle genomes by a nonsequence-specific ssDNA binding mechanism (Cappadocia *et al.,* 2010). Another group (Xiong *et al.,* 2009) identified Why 3 and Why 1 as two components of KPRE-binding factor 1 (KBF1), could mediate the transcriptional repression of the kinesin gene *AtKP1*.

1.2.5.2 PPR proteins

The PPR proteins are characterized as having tandem repeats of a degenerate 35 amino acid signature motif that may form a nucleic acid binding groove (Small and Peeters, 2000). The PPR protein family is one of the largest gene families and expanded greatly in land plants, making up a very large proportion of the unknown function proteins in Arabidopsis, with about 450 and 447 members in Arabidopsis and rice, respectively.

A majority of the PPR proteins have been shown to play essential roles in chloroplasts and mitochondria, possibly by binding to certain organellar transcripts. Previous researches have revealed PPR proteins maybe associated with various molecular events, such as embryogenesis, mostly post-transcriptional modification and bind to RNA specifically (Kotera *et al.*, 2005; Koussevitzky *et al.*, 2007). RNA editing is one of the most important functions of PPR proteins in terms of understanding how it came about. Amongst others roles of PPR proteins are such as splicing, RNA stability and turnover, enhancing and blocking translation of RNA (Schmitz-Linneweber & Small, 2008). Gutiérrez-Marcos *et al,*. (2007) recognized that mutations in *empty pericarp4* (*emp4*), a maize PPR-encoding gene, contribute in a seed-lethal phenotype.

1.2.5.3 miR840 in Arabidopsis

Bartel's lab employed high-throughput pyrosequencing to obtain *Arabidopsis thaliana* small RNAs in 2006. They identified many newly emergent and diverse miRNAs, each expressed in specialized tissues or at low levels under standard growth conditions, just like miR840. **Figure 3** shows the stem-loop of the miR840.

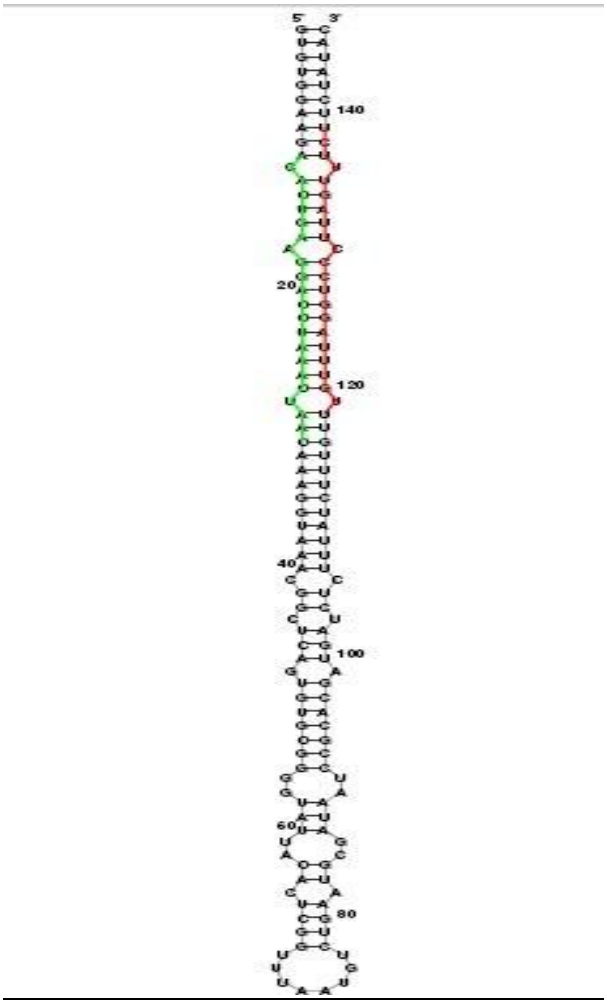


Figure 3. Stem-loop of miR840 (Rajagopalan *et al.* 2006): Green & Red sequence is the miRNA:miRNA* duplex, the green part is miRNA, the red part is miRNA*.

MicroRNA 840 (miR840) is a novel member of *Arabidopsis* microRNA families. The full genomic sequence of miR840 is 146bp in length that is located in chromosome 2: 771337-771562[-]. And the mature sequence (ACACUGAAGGACCUAAACUAAC) is 22nt in length, like other miRNAs,miR840 is processed from hairpin precursors by Dicer-like enzymes and can negatively regulate gene expression by attenuating translation or by directing mRNA cleavage.

The former publication reported that MiR840 miRNA was found within the annotated 3'untranslated region (UTR) of a PPR mRNA, At2g02750 which the neighboring gene is the Why 3, At2g02740, and miR840 c target the 3'UTR of Why 3 for cleavage. (Rajagopalan *et al.*, 2006). The Why 3 protein is being investigated by our group. So

that understanding the actual function of miR840 is important for our group to deeply analyze Why 3 protein.

For the miR840 function analysis, firstly two database TAIR (http://www.arabidopsis.org/) and TIGR (http://plantta.jcvi.org/) were used to predict the annotation of the target gene (Why 3) of miR840. Two databases gave two completely different annotations versions for Why 3. From TAIR, the target of miR840 is only at the 3' untranslated region (3' UTR) of Why 3 gene that means miR840 should just have down regulative function to Why 3 gene. However, the information from the TIGR showed the length of the 3' UTR of Why 3 gene is difference from the result from TAIR. TIGR predicted that Why 3 gene and PPR gene overlap with their 3' UTR where the target site of the miR840 is, this means according to the prediction from TIGR, the miR840 may play a role to down regulate both Why 3 and PPR, but still possible to just regulate one of them.

In summary, the results of bioinformatic method showed the function of miR840 is unclear (may influence the expression level of either Why 3 or the PPR protein, or both of them), depending on depending on what annotation you take. Therefore biological experimental analysis is necessary to determine the miR840 target. **Figure 4** shows the gene model of Why 3 and PPR from TAIR and TIGR.

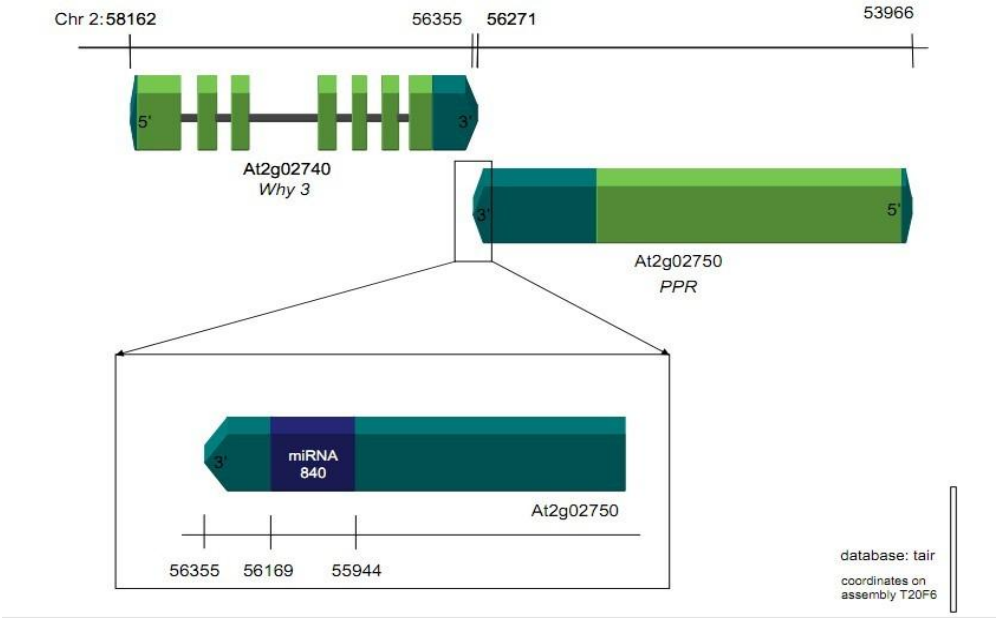


Figure 4A. Gene model for showing the miR840 at the PPR 3'UTR

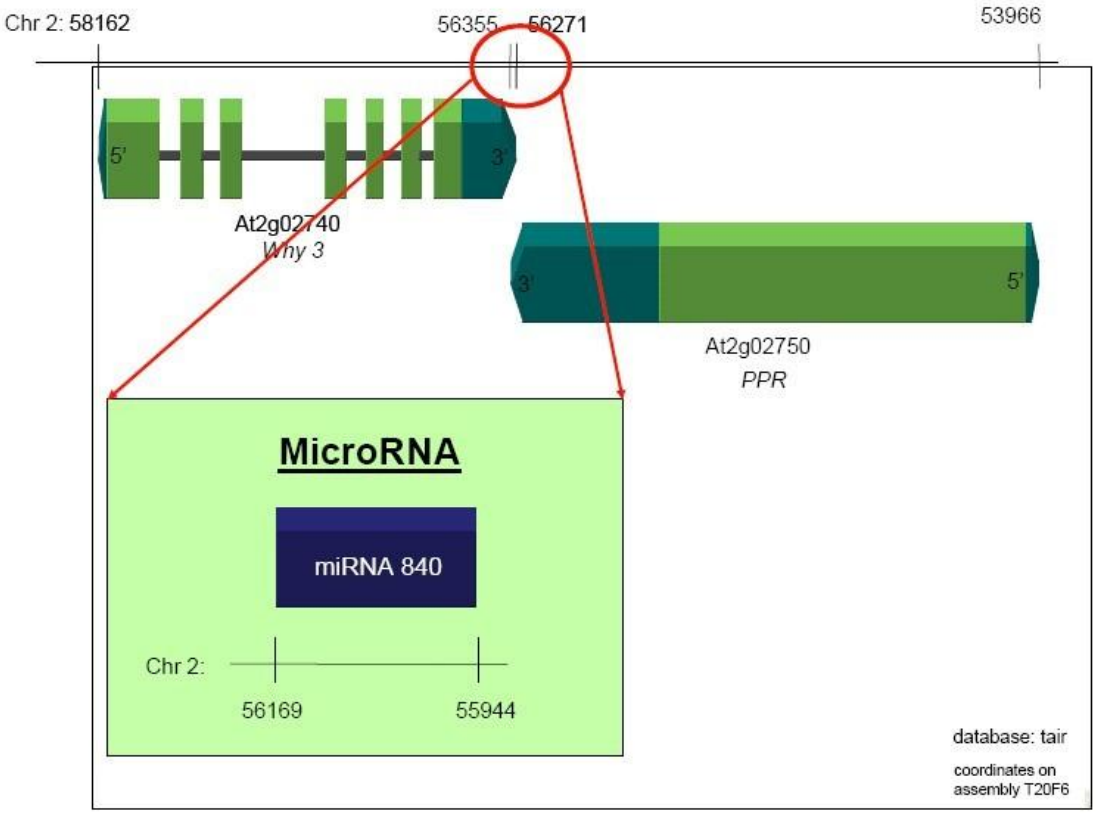


Figure 4B. Gene model for Why 3 and PPR from TAIR, shows the miR840 target the Why 3 3'UTR

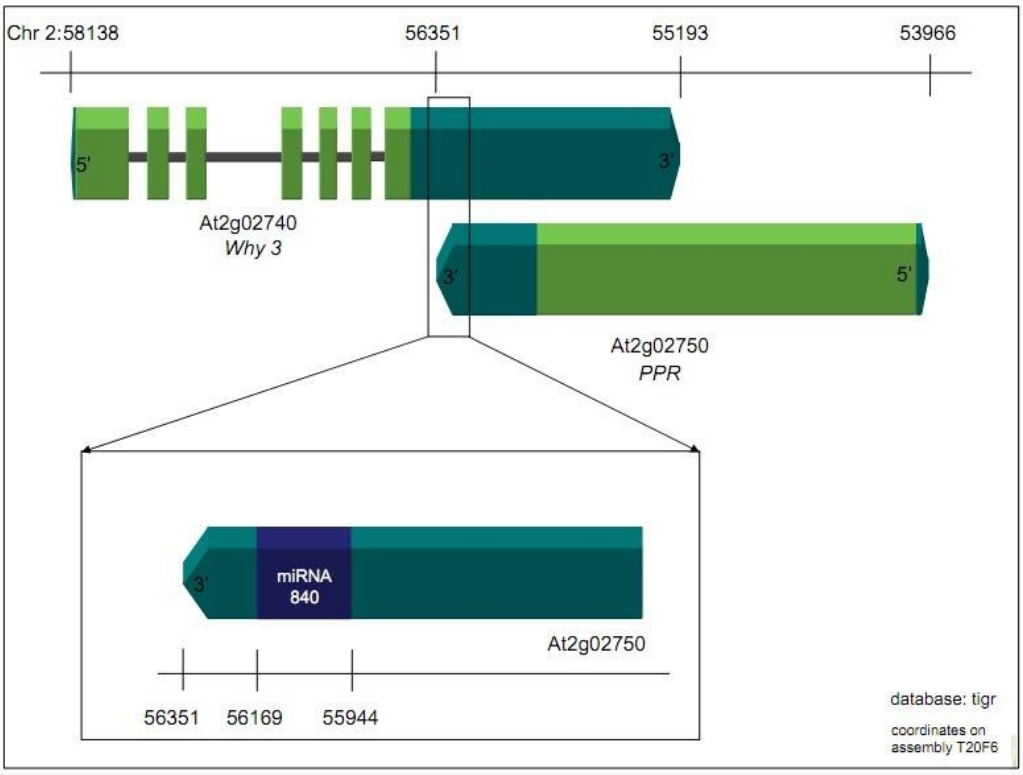


Figure 4C. Gene model for Why 3 and PPR from TIGR, shows the miR840 target the Why 3 the overlapping 3'UTR of Why 3 and PPR.

1.3 Goals of the thesis project

It has been predicted by present bioinformatic methods, that miR840 has Why 3 as a target, but as the annotation at that position is not clear, it may also be the neighboring PPR protein, therefore, understanding the function of miR840, the experimental analysis is requisite. So that my thesis project mainly aims to find the target of miR840 by experimental methods. In order to verify the target gene of miR840, reverse genetics approach will be exploited to construct 4 expression vectors, including miR840 overexpressing vector (Primi), mutated miR840 target overexprssing vector (Primimut), for directing miR840 loss-of-function, Why 3 overexpressing vector (Why 3) and PPR overexpressing vector (PPR), then transformed them into the Arabidopsis for selecting transgenic plants. And then all the expression level of positive transgenic plant lines will be analyzed, through comparative method of expression level for finding the target of miR840. Phenotypes

observation resulting from miR840-overexpression and transgenic expressing miR840-resistant targets in *Arabidopsis* will also be carried out subsequently.

The vectors and transgenic plants preparation have been done by our former lab members before, but not all of the vectors are constructed successfully. So my main actual work was to re-construct two expression vector (Primi and PPR), and perform the preliminary expression level analysis to the other two successful plant lines (Primimut and Why 3).

The second goal of this work is to use bioinformatic method to predict the subcellular localization of PPR Protein. As mentioned above, PPR families can act directly or indirectly on RNA and play many essential roles in mitochondria or plastids, and the subcellular distributions of some PPR proteins have been demonstrated by experimental transformation or bioinformatic methods. However, the distribution of our specific PPR protein has not been pointed out to date. Here in my project, I use the online database to do the location prediction, and the experimental validation will be set up sooner after.

2. Materials and Methods

2.1 materials

2.1.1 Plant (Arabidopsis) growth

Arabidopsis seeds (the same line) were sown on soil in a big square pot (20 cm×20 cm) and leave them for 5-7 days at 4℃ in darkness to allow an even germination. Germination was induced by transfer of the plants to a light chamber with 22 ℃ (24 hours) and a relative humidity of 60 %. The plants were watered daily.

2.1.2. Basta selection and re-pot

After 7-10 days germination in a light chamber (16 h day light), Arabidopsis seedlings were selected by spraying with the herbicide Basta (0,1%). Only the plants that have the tDNA-insertion from the expression vector with the Basta resistance gene, survive this selection. Spraying was repeated after 3 days, 3-4 times. 2 weeks after the Basta selection, the survivors were transferred to an the individual pot (square pot, 9 cm×9 cm)

2.1.3 Chemical and Enzymes

The main chemicals and enzymes used in this study were purchased from New England Biolabs, Ambion, Invitrogen, Promega, Sigma, Bio-Rad and Thermo Fisher. Parts of solutions are made by our lab. All primers are made by Sigma.

2.1.4 Bacterial strains, primers, and vectors (Vector maps are in the Appendix)

2.1.4.1. The bacteria strain used in this project

The bacteria strain used in this project is ***E. coli* DH5α.**

2.1.4.2. Agrobacterium tumefaciens, vectors and primers

Agrobacterium tumefaciens GV3101 (Koncz and Schell, 1986)

Table 1. Entry vectors used in the project

Target	Vector Names (Abbrevation)
MiR840 overexpression (Primi)	AtmiRNA840_pENTR/SD/D-TOPO (Primi_pENTR)
MiR840 taret mutant (Primimut)	AtmiRNA840mut_pENTR/SD/D-TOPO (Primimut_pENTR)
Why 3 Overxpression	Why 3_pENTR/SD/D-TOPO (Why 3_pENTR)

Table 2 Destination vectors used in the project

Target	**Vector Names (Abbrevation)**
MiR840 overexpression (Primi)	pEarleyGate101_35S-Gw-YFP-HA (pEG101)
MiR840 target mutant (Primimut)	pEarleyGate102_35S-GW-CFP-HA (pEG102)
Why 3 Overxpression	pEarleyGate101_35S-Gw-YFP-HA (pEG101)
PPR Overxpression	pEarleyGate103_35S-GW-GFP-His (pEG103)

Table 3. Expression Vectors used and vector maps in the project:

Target	Vector Names (Abbrevation)
MiR840 overexpression (Primi)	Expression Clone/pEarleyGate101_35S-Gw-YFP-HA/Clone pENTR/SD/D-TOPO,PCR Product of AtmiRNA840 (Primi Expression pEG101)
MiR840 target mutant (Primimut)	Expression Clone/pEarleyGate102_35S-GW-CFP-HA/Clone pENTR/SD/D-TOPO,PCR Product of AtmiRNA840mut (Primimut Expression pEG102)
PPR overexpression	Expression-ClonepEarleyGate103_35S-GW-GFP-His/PPR_pENTR/SD/D-TOPO (PPR Expression pEG103)
Why 3 Overxpression	Expression Clone/pEarleyGate101_35S-Gw-YFP-HA/why3pENTR223.1-Sfi (Why 3 Expression PEG101)

Table 4. Entry vector has been recontructed in the project

Target	Name (Abbrevation)
PPR 3 Overxpression	PPR_pENTR/SD/D-TOPO (PPR_ pENTR)

Table 5. Information for primers for PCR

Primer	Sequence
JF3 FWD	5’ -TGCTCCAATTTGCTCCTGCAGCTGGT-3’
JF3 REV	5’-TGCACTCCAGCCTATAAGGTGCGGCAA-3’
JF46 FWD	5’-GGTCATAAACCAGCTCATCCAA-3’
JF46 REV	5’-GTCTCGTAAAGCAGGTGATTC-3’
JF47 FWD	5’-GAAGGACCTATTGATTGTTT-3’
JF47 REV	5’-CTTCAGTGTCTTCCACACTG-3’
JF54 FWD	5’-ACGAGCCAGGGATAGCGC-3’
JF54 REV	5’-ACGAGACAGAGATACCG-3’
JF53 FWD	5’-ACATGAAATGTGGGCTCTCC-3’
JF53 REV	5’-AGACCGACCCAAAAGATCAA-3’
JF116 FWD	5’- ATGATAAACCTAACGAGACAGAGATACCGAGTCTC-3’
JF116_3 REV	5’- ATGATAAACCTAACGAGACAGAGATACCGAGCTC-3’
GFP_as	5’-GCCACGGAACAGGTA-3’
YFP_AS_alt	5’-TGAACTTCAGGGTCAGCTTG-3’
35S_S	5’-ACAATCCCACTATCCTTCGC-3’
UH107	5’-TTCATGTGGTCGGGGTAGCG-3’
M13 FWD	5’- GTAAAACGACGGCCAG -3’
M13 REV	5’-CAGGAAACAGCTATGAC-3’

Table 6 Information for primers were used for amplifying which clone

Primer	Vector
JF3 FWD	Why 3 Expression clone
JF3 REV	
JF46 FWD	Primi Expression vector, Primimut Expression clone
JF46 REV	
JF47 FWD	Primi Expression vector, Primimut Expression clone
JF47 REV	
JF53 FWD	PPRExpression clone
JF53 REV	
JF116 FWD	PPR Entry clone
JF116_3 REV	
GFP_as	PPR Expression clone
YFP_AS_alt	Primi, Primimut and Why 3Expression clones
35S_S	Primi, Primimut, Why 3 and PPR Expression clones
UH107	Primi Primimut and Why 3Expression clones
M13 FWD	Primi, Primimut, Why 3 and PPR Expression clones, PPR entry clone
M13 REV	PPR entry clone

Table 7. Information for primers for Real-Time PCR and functions

Name of Primer	Sequence	Function
ActinSH FWD	5'-ATGGAAGCTGCTGGAATCCAC-3'	Primer pairs for amplifying Actin gene
ActinSH REV	5'-TTGCTCATACGGTCAGCGATG-3'	
JF_RL_PPR_FWD_1	5'-TGGGCTCTCCTCGTGGGCAC-3'	Primer pairs for amplifying PPR gene
JF_RL_PPR_REV_1	5'-CACGGGATCCTTGGGTTTCGGT-3'	
WHY3_RL_FWD	5'-AATTATTCCTTTAAAACCGACGGC-3'	Primer pairs for amplifying Why 3 gene
WHY3_RL_RV	5'-TAATCACTTTGTCGAGACTTCACCG-3'	

2.1.5 Medium for bacteria culture

LB medium:

Bacto-Tryptone	10 g
Bacto-Yeast extract	5 g
NaCl	10 g

ddH_2O to 1L, pH adjusted to 7.0

Final concentration of Antibiotics in growth media

Kanamycin (E.coli and agrobacteria): 50 μg/ml

Rifampicin (agrobacteria): 50 μg/ml

Gentamycin (agrobacteria): 10μg/ml

Stock solutions stored at –20 °C

2.1.6 General medium

1% Agarose gel

Agarose	0,5 g
TAE buffer	50 ml
0,1% Ethidium bromide	0,5 µl

TAE buffer

Tris acetate	40 mM
EDTA	1 mM

pH is 8,5

10x bromophenol blue loading dye

50% glycerol

0,1% bromphenol blue

TPS-buffer

	For 50 ml
100 mM tris/HCL pH 8	5 ml 1 M stock solution
1 M KCL	3,73 g KCL
10 mM EDTA pH 8	1 ml 0,5m stock solution

Solutions needed for preparation of chemical competent Agrobacterium cells

0,15M NaCl solution autoclaved	2,192g/250ml
20mM $CaCl_2$ solution autoclaved	0,735g/250ml ($CaCl_2 \times 2H_2O$)

SOB

2% (w/v) bacto tryptone	12 g
0,5% (w/v) yeast extract	3 g
10 mM NaCl	0,36 g
0,5 mM KCL	0,11 g
10 mM$MgCl_2$	1,2 g
10 mM $MgSO_4$	1,47 g
dH_2O	600 ml

pH: 6,7-7,0

TB

10mM Pipes	0, 302 g
55mM $MnCl_2$	1, 08 g
15mM $CaCl_{2(-2h2o)}$	0,22 g
250 mM KCl	1, 85 g
dH_2O	100 ml

pH 6,7 without $MnCl_2$, then add $MnCl_2$ and filter sterilize.

Borate buffer (pH 9,0)

Sodiumborate	200 mM
EGTA	30 mM
SDS	1% (W/V)

Buffer P1 (pH 8,0)

Tris ·HCL	50 Mm
EDTA	10 mM
RNAse I	50 mg/ml

Buffer P2

NaOH	0,2 M
SDS	1%

Neutralization 3 buffer (pH 4,2)

Guanidine hydrochloride	4 M
potassium acetate (CH_3CO_2K)	0,5 M

2.2 Methods

2.2.1 GATEWAY - Technology INVITROGEN and LR reaction

The Gateway cloning technology employs *in vitro* site-directed recombination to insert a DNA fragment of interest into a donor vector to obtain an entry vector containing the DNA fragment with its orientation and reading frame maintained. The DNA fragment can subsequently be subcloned into various destination vectors to obtain expression vectors ready to be used in the appropriate host.

In a LR reaction, the recombination of an attL containing entry clone with an attR containing estimation vector was performed to create an attB-containing Expression clone. The cloned sequence can then be exported to any type of destination vector, following a similar strategy. An LR clonase reaction exchanges the region between L sites in the entry vector with the region between R sites in the destination vector.

The LR reaction performs as (1.5 mL tube for one sample):

1. 1 µL entry clone + 0,5 µL destination vector at room temperature.
2. Added H_2O up to 4 ul
3. 1 µL LR Clonase enzyme
4. Vortex and spin down
5. Reaction at room temperature. For at least 1 hour.
6. 1/2 of the reactions was transformed into *E.coli* and plated on LB-Kan agar

2.2.2 Polymerase Chain Reaction (PCR)

PCR amplification was done according to Saiki *et al*. (1985). Typically, PCR reactions were prepared as represented in the table 2.

PCR reactions:

Reagents	**Volume (µl)**
10x reaction buffer	2
dNTPs (10mM)	0,4
forward primer (10µM)	1
reverse primer (10µM)	1
Taq DNA polymerase (2u/µl)	0,1
DNA template (50ng/µl dil.)	1
dH2O	14,5
Total	20

PCR reaction was performed by incubating the PCR reaction mix at three different temperatures corresponding to the denaturing, annealing and extension steps for each cycle of amplification. In a typical reaction, the DNA was denatured at 95 ℃, primers annealed at 40-60 ℃, and extension processed at 72 ℃. 30-40 amplification cycles were used.

PCR program:

Step	Temperature	Time
Initial denaturation	94 ℃	10 min
Denaturation	94 ℃	30 sec
Annealing	40-60 ℃	45 sec
Elongation (*rate: approx. 1KB/min*)	72 ℃	Depend on the expected length
45 cycles		
Finishing elongation	72 ℃	10 min
Cooling	12 ℃	forever

2.2.3 Agarose gel electrophoreses

Agarose gel electrophoresis is used to separate and visualize DNA and RNA fragments in different sizes. Due to the negative charge of the phosphate groups, which are part of the backbone of the nucleic acids, DNA and RNA will migrate towards the anode through the agrose gel acting as a molecular sieve under an electrical field.

1x TAE buffer containing 1% agarose was boiled in a microwave until the agarose had melted. The solution was allowed to cool and 0.5ul ethidium bromide (1%, v/v) was added before the solution was poured into a gel holder and allowed to set. DNA and RNA samples containing 1x loading buffer (5% glycerol, 0.05% bromophenol blue as a dye for nucleic acid detection) were loaded onto gels immersed in 1x TAE buffer, along with 5 µl of 1 kb DNA ladder. 1 kb DNA ladder (GeneRuler^{TM}, Fermentas) are used for all the gel pictures in this thesis, as **Figure 5** shown. Voltage and electrophoresis time depended on gel sizes used. DNA/RNA was visualized under UV light using a GelDoc 2000 equipment (BioRad).

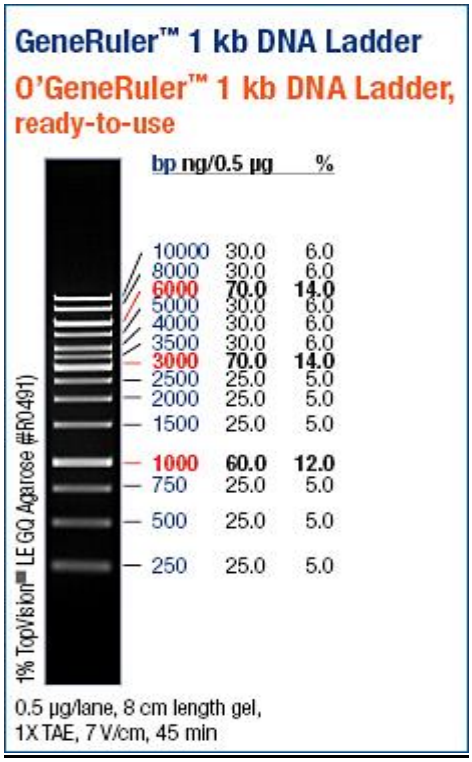


Figure 5. 1 kb DNA ladder (GeneRuler^{TM}, Fermentas)

2.2.4 Preparation, transformation of chemically competent Agrobacterium cells, and Agrobacterium-Mediated Arabidopsis Transformation

2.2.4.1 Preparation of chemically competent Agrobacterium cells

A single colony of Agrobacterium was inoculated in 5 mL YEP liquid media with 20mg/l Rifampicin by incubating at 27℃ overnight with shaking (~250 rpm). Then the culture was transferred to 500ml of YEB with Rif_{20} for overnight growth at the same condition till a bacterium density of 0.5-0.6 at OD_{600} was reached. To collect Agrobacterium cells, 500ml culture was poured into two centrifuge buckets and span for 5 minutes at 7000g at 4^{o}C. Supernatant was discarded and the cell pellets were resuspended in 25 ml ice-cold water containing 0.15M NaCl. The cells were kept on ice for 15 minutes. After spinning the cells again at 7000g for 5 min at 4^{o}C, the supernatant was discarded and cell pellets were resuspended in a total volume of 5ml

ice-cold water containing 20mM $CaCl_2$. The competent cells were aliquoted into small packages (50-100ul each), quickly frozen in N_2-liquid and stored at -80 ^oC.

2.2.4.2 Transformation of Agrobacterium cells

1-2 ul DNA was added to thawed agroaliquot and frozen again in liquid N_2. After thawing again, the cells were heat-shocked for two minutes at 42^oC. After gentle shaking at 28 ^oC for 2 hours in 800 ul YEB, the cells were plated onto appropriate YEB-selection plates.

2.2.4.3 Agrobacterium-Mediated Arabidopsis Transformation

To prepare Agrobacterium for Arabidopsis transformation, the Agrobacterium cells containing recombinant plasmids were spin down at 5000 rpm for 5 min and the cell pellets were re-suspended in 2 ml sucrose solution, Before infection, 1 ul Silwet was added and 1-2 drops of the Agrobacterium solution was injected onto the buds by pipetting for dipping flowers. The plants were covered overnight to maintain humidity and then were grown at 20 ℃ under normal light/dark cycles. Seeds were harvested after plant maturity.

2.2.5 Purification of the DNA fragments from the gel

Purification of DNA fragments ranging from 40 bp to 50 kb from agarose gel followed the instruction as described by the manufacturer of Qia Ex II Kit.

2.2.6 DNA restriction digest

Restriction enzymes can bind and cleave specifically hydrolytic phosphodiesterified bonds of both strands of a DNA molecule. Various kinds of endonucleases differ in their recognition sites, cleavage sites and organism of origin. Some restriction enzymes cut in the middle of their recognition site, creating blunt-ended DNA fragments. However, the majority of enzymes make cuts staggered on each strand,

resulting in a few base pairs of single-stranded DNA at each end of the fragment, known as “sticky” ends. The sticky-ended fragments can be easily ligated to other sticky-ended fragments with compatible single-stranded overhangs, resulting in efficient cloning. In this study, the DNA to be analyzed was incubated in defined buffer conditions with the appropriate restriction enzyme under recommended temperature and incubation time. The optimal temperature for most restriction enzymes is 37 ℃.

In Gateway system, if there are entry vectors with the same resistance as the expression vector remaining in the products after LR reaction, they can also generate colonies after transformation into bacterium, which would mix with the colonies from the LR reaction products. In order to distinguish the LR products from the entry vector, we usually use the restriction digest approach that is based on the fact that LR products can be digested by appropriate restriction enzyme while entry vector cannot. In this method, the digested sample is served as a background control. After transforming the LR reaction and the background control samples into bacterium, a conclusion can be drawn that our LR plate doesn’t contain entry vectors if colonies only are observed in the LR reaction plate.

Reaction for testing entry vector: 1.0 µl of 10 x reaction buffer, 2.0 µg of DNA, 0.2 µl of each restriction enzyme (Hae II), dH_2O to a final volume of 10 µl, then incubated for 1-2h at 37 ℃.

Reaction for identifying the expression vector: 0.1µl BSA, 1.0 µl of 10 x reaction buffer, 2.0 µg of DNA, 0.1 µl of each restriction enzyme (Xho I), dH_2O to a final volume of 10 µl, then incubated for 2h at 37°C.

2.2.7 Amplifying vector and Ligation reaction

To get the new and correct PPR entry vector (adding several bases to the template

entry vector), we amplified the template entry vector by PCR with phusion enzyem. After amplification, the template entry vector was digested by Dpn1, whose function 1 is to selectively digest template entry vectors with methylated sites. All the new amplified entry vectors, because of the lack of methylated site, would not be digested. After ligation of the 5' and 3' of the vectors, they were ready to be used for subsequent experiments.

When the vector ends were blunt or compatible with each other, the vector was dephosphorylated prior to ligation to prevent self-ligation. 5' and 3' of vector were ligated together with the use of T4 DNA ligase, which catalyzes the formation of phosphodiester bond between the 5' phosphate end and the 3' hydroxyl end in double stranded DNA.

Whole reaction process:

We firstly ran normal standard PCR and then resolved 1/2 of PCR reaction (10µl) and the whole negative control on agarose gel to check whether the band has the correct size. 3 µl of PCR reaction from another 1/2 PCR reaction was served as a background control. If the band showed expected size, 0.5 µl of Dpn 1 was added to the rest of the PCR product and incubated at 37°C for 30 min to 2 h. After incubation, 1µl PNK and 1µl 100mM ATP were added to the PCR product which had treated by Dpn 1, and was incubated at 37℃ for 30 min. The band of correct size was extracted from the gel with a QiaEx II kit and then ligation reaction was performed (Process shows below). The ligation reaction was hold at room temperature overnight and kept at 4℃ until used for transformation (including background control).

Ligation reaction::

Ligation Buffer for T4 DNA Ligase	2 µl
TOPO entry Vector (25ng):	3µl
T4 DNA Ligase	0.5µl
Adds dH_2O to a final volume of 20µl	

2.2.8 Chemically Competent E. *Coli* cells preparation (Inoue., *et al* 1990)

250 ml of 'SOB' medium was inoculated with an overnight grown colony of E. *Coli* stain DH5α at RT until OD_{600} = 0.4 to 0.6. A 1:100 inoculation in SOB typically takes 4-5 hours for the cells to grow up to OD_{600} = 0.4 to 0.6 at RT. The cells were ice-clod for 10 minutes, span down at 2500×g for 10 min at 4 °C, resuspended gently in 80 ml of ice-cold 'TB', and then were kept on ice for 10 minutes. Span again at 2500×g for 10 min at 4 °C and resuspended the cells gently in 20 ml of ice-cold 'TB'. Before freezing the cells, DMSO was added to a final concentration of 7% and the cell suspension were kept on ice for 10 minutes. The cell suspension was aliquoted into 100-200 µl packages, quickly frozen in liquid nitrogen and stored at -80 °C.

2.2.9 Transformation of chemically competent cell (E. *coli*)

2 – 5µl of the LR reaction was added into 50µl of the competent cells, mixed gently by ticking against the tube and left on ice for 30 min. Then the sample was heat-shocked at 42 ℃ for 45 sec followed by 10 min recovery on ice. Afterwards, immediately 800 ul of sterile LB medium without antibiotic selection was added to the transformed cells and incubated at 37 ℃ for 45 min with gentle shaking. The cells were plated onto LB agar plates containing appropriate selective antibiotic and incubated overnight at 37 ℃.

2.2.10 DNA Isolation from plants

Small leaf fragments (ca 2 x 2 mm) were immersed in 50 ul TPS and squeezed with piston to allow complete grinding of the leaf tissues. The samples were incubated for 10 min at 95 ℃ in Thermomixer (ca 850 rpm). After cooling down for 2 min on ice, the samples were centrifuged at 14000 rpm at room temperature (RT). The supernatant was transferred into a new tube and filled up to 200 ul with sterile water. 150 ul Isopropanol (0.7 Vol) was added into the supernatant and incubated for 10 min at RT. After centrifugation at 14000 rpm for 10min under RT, the supernatant was discarded and the DNA pellet was washed with 150 ul ice-cold 70% EtOH. Centrifuged again at 14000 rpm for 5min and discarded supernatant, followed by a final briefly centrifugation to recover rest of the EtOH. Discarded any visible remnant EtOH and air-dried the pellet for approx. 10 min then resuspended the DNA in 10 ul sterile water. To remove any RNA contamination, 0.5 ul RNase A was added into each tube and incubated for 1 hour at 37 ℃. The DNA samples were kept at 4 ℃. 1 ul of each DNA was used per PCR-reaction

2.2.11 RNA Isolation from plant tissues

200-600 mg of each of the frozen tissue (-80 ℃) was grinded thoroughly with a cold mortar and pestle. Then the powder was mixed with pre-warmed (65 ℃) mixture of 800 ul Borate buffer and 600 ul Phenol and vortex, and incubated for 20 min at 30 ℃. After centrifugation at 14000 rpm for 10 min, the aqueous phase was transferred to a new EP tube, mixed with 600 ul phenol: chlorophorm: isoamylalcohol (25:24:1), vortexed and incubated for another 10 min at 30 ℃, followed by a second centrifugation for 10-20 min at 14000 rpm. The aqueous phase was extracted again with 600 ul Chloroform: Isomyl alcohol followed by incubation and centrifugation as described above. A final concentration of 2 M LiCl was added to precipitate RNA over night at 4 ℃ and then the RNA pellet was dried for 2-5 min in speed. Each of the RNA sample was resuspended in 20-50 ul H_2O (MilliQ) and their quality and

concentration were checked by gel electrophoresis and NanoDrop spectrophotometer, respectively.

2.2.12 Plasmid Isolation (Plasmid Miniprep)

A single colony from the selective plate was inoculated in a culture of 3 ml LB medium containing the appropriate selective antibiotic. After incubation for approximately 14-18 h at 37 ℃ with vigorous shaking (approx. 300 rpm), the bacterial cells were collected by centrifugation at 6000 rpm for 10 min at 4 ℃. To resuspend the cells, 200 ul of Buffer P1 was added and the bacterial pellet was vortexed thoroughly. For cell lysis, 200 ul of Buffer P2 were added, mixed thoroughly by vigorously inverting the sealed tube several times, and incubated at room temperature (15–25 ℃) for 3 min. After that, 200 ul of chilled buffer neutralization 3 was added, mixed immediately and thoroughly by vigorously Inverting several times, and incubated on ice for 5 min. The samples were centrifuged at maximum speed for 5 min and the supernatant containing plasmid DNA was recovered. Promptly we added Isopropanol (70% of the whole supernatant volume) and shaked several times and incubated on ice for 15 min. After centrifugation at full speed for 10 min at 4 ℃, all the liquid was discarded and 1 ml 70% EtOH was added for washing the DNA pellets. The centrifugation was repeated once and all the liquid was discarded. After drying the pellet in the hood, 20 ul TE buffer was added to resolve the DNA.

2.2.13 DNA/RNA Concentration measurements by NanoDrop

Concentrations of DNA/RNA samples were measured by using Nanodrop ND-1000 Spectrophotometer and relative program. The program was run and spectrophotometer was initialized according to program instructions. In order to minimize background effects measurements were done comparing with blank solution. After that, 1 μL of the sample was loaded and “measure” comment was clicked

2.2.14 Colony PCR

Single Bacteria colony can be used directly as template for PCR without isolating pure DNA by Colony polymerase chain reaction (colony PCR). Through colony PCR, large amount of material and time-consuming process involving extraction of genomic DNA can be saved. Each colony chosen for colony PCR was moved to a new antibiotic plate with a pipette tip and residual part of the tip were mixed into the PCR.

The colony PCR use the same reaction mix to normal PCR, just add 0,2 Triton X for breaking the bacteria cell membrane. And the program of colony PCR is also the same to the normal PCR mentioned above.

2.2.15 DNA sequencing

Sequencing was done to verify new DNA sequences in expression vectors. The method was carried out by the REDUCED-AMOUNT-PROTOCOL FOR SEQUENCING which is similar to the PCR reaction with several important differences, following the Sanger sequencing mechanism. In the sequencing reactions, forward and reverse primer was added in separate tubes. The reaction mix contained the four dNTP's and in addition a portion of ddNTP's labelled with four different fluorescent dyes. The dye labelled ddNTP's terminates the elongation and when they were incorporated randomly, multiple molecules of varying length were generated.

Sequencing reaction:

5X BigDye Sequencing buffer	2 µl
5X Primer (10 µM stock)	0,5 µl
5X BigDye c3.1 Ready Reaction Mix	0,5 µl
DNA (approx. 0,1 µg Plasmid)	X µl
Add dH_2O to 10 µl	

Cycle Sequencing:

96 ℃	1 min
96 ℃	10 sec
50 ℃	5 sec
60 ℃	4 min
30 cycles	
12 ℃	Forever

2.2.16 DNase treatment and Removal Reagents (DNA-freeTM Kit, ABI)

To remove trace DNA contamination in the RNA samples isolated, we added 0.1 volume 10 X DNase I Buffer and 1 µl rDNase I to the RNA, mixed gently and incubated them at 37 ℃ for 20–30 min. To remove excess DNase, 0.1 volume of the resuspended DNase Inactivation Reagent was added, mixed well and incubated for 2 min at room temperature, with mixture occasionally. After the reaction, the tube was centrifuged at 10,000 x g for 1.5 min and the RNA was transferred to a fresh tube.

2.2.17 cDNA synthesis using Rermentas ReverAid TM h Minus M-MuLV Reverse Transcriptase (protocol optimized for amplification of transcripts larger than 500 bp, by Anja Martin)

Each of the RNA sample treated with DNase (for removing the genomic DNA in the sample) was mixed with 1µl oligo$(dT)_{xx}$ and incubated at 70 ℃ for 5 minutes, followed by quick chilling on ice. Then 4µl of 5X reaction buffer, 1µl RiboLock and 2µl dNTPs (10mM) was added and incubated at 37°C for 5 min. Afterwards, 0.5µl M-MuLV Reverse Transcriptase was added and incubated at 45 ℃ for reaction for 60

min, followed by enzyme inactivation at 70 ℃ for 10 min. The synthesized cDNA was ready for Real-time PCR analysis.

2.2.18 Real-time PCR reaction

Real-time PCR is commonly used for both diagnostic and basic research. In research settings, real-time PCR is mainly used to provide quantitative measurements of gene transcription. The technology may be used in determining how the genetic expression of a particular gene changes over time, such as in the response of tissue and cell cultures to an administration of a pharmacological agent, progression of cell differentiation, or in response to changes in environmental conditions.

The procedure follows the general principle of polymerase chain reaction; its key feature is that the amplified DNA is detected as the reaction progresses in real time. This is a new approach compared to standard PCR, where the product of the reaction is detected at its end. Two common methods for detection of products in real-time PCR are: (1) non-specific fluorescent dyes that intercalate with any double-stranded DNA, and (2) sequence-specific DNA probes consisting of oligonucleotides that are labeled with a fluorescent reporter which permits detection only after hybridization of the probe with its complementary DNA target. Frequently, real-time PCR is combined with reverse transcription to quantify messenger RNA and Non-coding RNA in cells or tissues.

In the project the real-time PCR system is Bio-Rad MiniOpticon system. The background control gene is Actin and the reaction master mix is SYBR Green PCR Master Mix from ABI.

The reaction mix:

SYBR Green PCR Master Mix	10 ul
1: 10 diluted cDNA	3 ul
Primer	FWD: 2 ul + REV: 2 ul
dH_2O	5 ul

The real-time PCR program:

1. 95 ^oC	3 min
2. 95 ^oC	10 sec
3. 55 ^oC	10 sec
4. 72 ^oC	30 sec
5. Go to 2, 39 cycles	
6. 95 ^oC	10 sec
7. Melt curve 65 ^oC to 95 ^oC. increment 0,5^oC 0,05 sec + Plate read End	

3. Results

At the beginning, thanks for our former group member Stephanie Hainbuch who has constructed all the four original expression vectors (Primi, Primimut, Why 3 and PPR) for the project, she also transformed them into the plants and finally collected the T2 generation seeds of the four 4 transgenic plant lines.

According to the my original project plan, my main work is to directly extract RNA from the 4 transgenic plant lines and analyze their expression level of the, including some protein level experiments later, to validate the target gene of miR840. However, things are not always going well for some reasons, the plan cannot be totally completed. The problems happened during the project and the actual processes/steps which has been done will be generally introduced in the following section.

Firstly, the transgenic plants died twice during their vegetation process due to the some reasons from the green house, so the whole plan was delayed several months to get the transgenic plant leaves from the original T2 seeds which have been collected by Stephanie before.

Secondly, through the screening work for the positive transgenic plants, the Primi line shows no positive plant. So that the Primi line have to be prepare again from the step of expression vector construct.

There was another problem has been found in the PPR line. The PPR expression vector which has been transformed into the plants, but it lost several bases in the gene coding region, this means the PPR line was also failed. Later we found the problem came from the its entry vector, the gene coding region lost several bases in the entry vector, so that the work of the PPR started from amplifying the entry vector. Now the expression vector has been constructed, later but soon it will be sequenced, the correct

vector will be transformed in to the plant.

The plan for Primimut and Why 3 lines were performed better than Primi and PPR, because their transgenic plants are positive. Preliminary expression level analysis has been done to Primimut and Why 3, the result also achieved the basic expected target. To be on the safe side, during the Primi and PPR vectors were constructing, the new Why 3 and PPR expression vectors also have been prepared again from the LR reaction.

All the results will be showed in details in the following parts. **Figure 6** is the workflow diagram to generally show the outline of the process of project according to the each plant line and the future research plan is stated in the discussion part.

Workflow Diagram

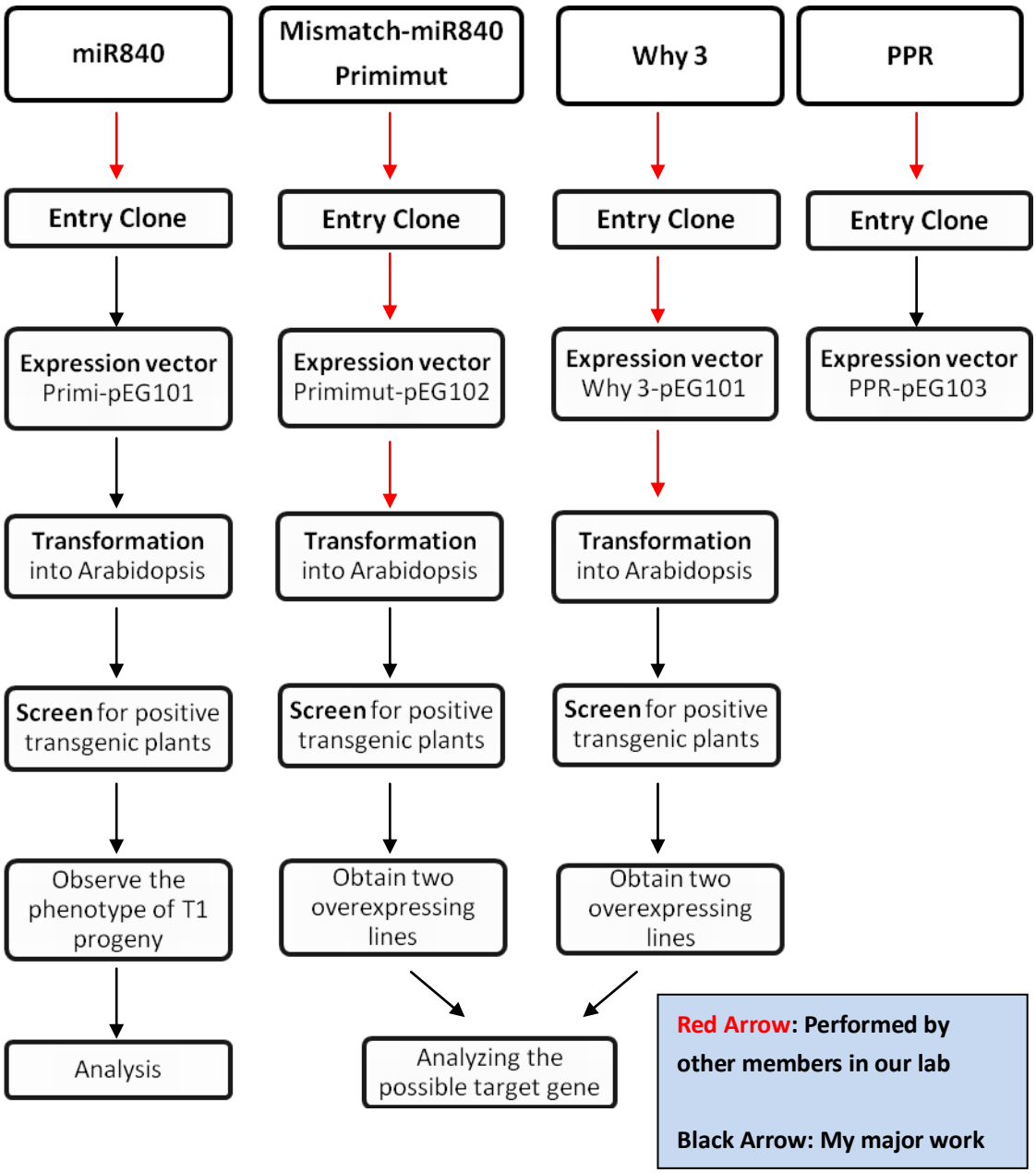


Figure 6. Workflow diagram of the process of project.

3.1 Overexpression analysis of miR840 (Primi) transgenic plants and re-construction of Primi transgenic plant line

3.1.1 Screen for miR840-overexpressing transgenic plants

Thank for the job from our former lab member Stephanie Hainbuch, miR840 has been already successfully cloned and constructed into an entry clone vector, pENTR/SD/D-TOPO **(Appendix 2.1)**. And by LR ligation, miR840 was constructed into the destination vector, pEarleyGate101_35S vector (**Appendix 2.3**), and transformed into Arabidopsis by Agrobacterium-mediated transformation.

Based on the detection with PCR method, all of the former prepared miRNA840 overexpression (Primi) transgenic T2 seeds have been screened but there is no positive transgenic plant obtained ultimately. The result demonstrated that miR840 gene has not integrated into the genome of transformed plants and if we still want to get the miR840-overexpression transgenic plants, we need to do reconstruction of the Primi vector and retransformation for getting the positive transgenic plant line.

3.1.2 Re-construction of miR840 expression vector (Primi Expression pEG101) for overexpression analysis.

3.1.2.1 Selection of cell clone positively expressing miR840

By sequencing analysis again, we confirmed the miR840-TOPO entry clone was correct and could be used for next LR ligation. By this method, we reconstructed miR840 fragment into the destination vector, which was mentioned above, pEarleyGate101_35S vector, i.e. pEG101. (**Appendix 2.3**) by LR reaction.

After transformation into chemically competent *E. coli* DH5α, we used colony polymerase chain reaction to screen for positive clones using the by specific primers JF46 fwd + UH107. Unfortunately, none of the colonies picked for screening contained the correct insert **(Figure 7). 1 kb DNA ladder (GeneRuler™, Fermentas) are used for ALL the gel pictures in this thesis, as Figure 5 shown.**

Then we directly picked four colonies randomly named A, B, C, D and incubated them in LB culture medium overnight. After plasmid isolation, we measured the DNA concentration by NanoDrop ND-1000 Spectrophotometer. The concentrations of the four plasmids is followed **(Table 9)**.

Table 9: concentrations of plasmids

Number	A	B	C	D
Concentration (ng/μl)	346.6	476.0	647.2	279.3

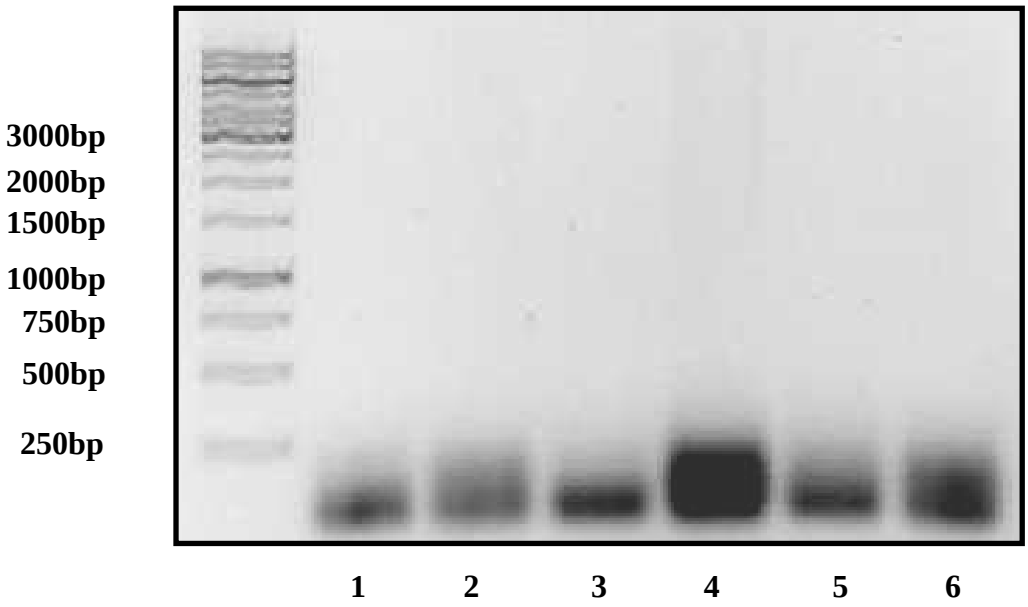


Figure 7. Screen for the positive clones containing miR840 fragment by colony PCR. The size of expected band should be 1234 bp, but all the colonies didn't show positive band.

1 kb ladder.

To verify whether there was correct insert in the four plasmids, the specific restriction enzyme XhoI was used. **Figure 8** shows that there was the expected specific band of around 1100bp (as shown at the arrow), so the plasmid B was then sequenced using primers 35S_S and YFP_as in order to verify its identity and direction of insertion. Thereby, plasmid B has been confirmed to be a positive clone.

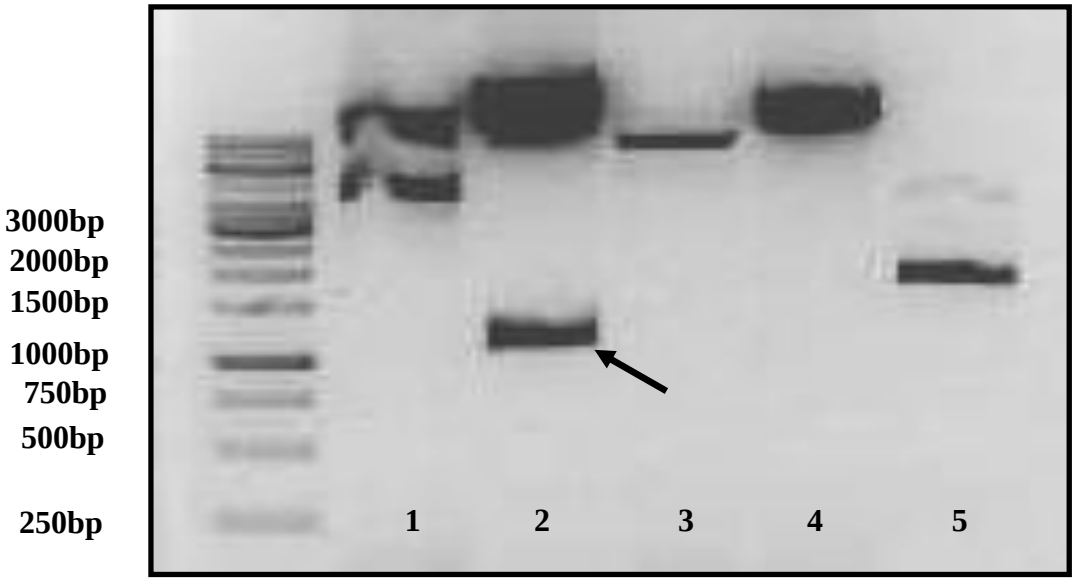


Figure 8. Digestion reactions of PEG101 vectors with subcloned fragments. Lane 1: colony A; Lane 2: colony B; Lane 3: colony C; Lane 4: Colony 6: Lane E: empty donor vector for checking the cutting site is available and correct., 1 kb Ladder. The expected band size is: 1060 bp. Colony B showed the most similar size (1100-1200 bp). 1 kb Ladder.

3.1.2.2 Optimization of PCR conditions.

To obtain the anticipated results and get the specific band by indicated primers for miR840 gene, we redesigned several pairs of primers to optimize our experimental content. The fragments of expected sizes were amplified by different primers (**Figure 9**). By comparison, the combination of primers JF46 fwd + UH107, expect size is 514

bp, as shown at the black arrow (**Fig. 9**) was the best choice for the next procedure because the band from JF46 fwd + UH107 showed a very clear band and the size is fit to distinguish (the optimal band size should be easy to distinguish from other bands, and not too large or too small).

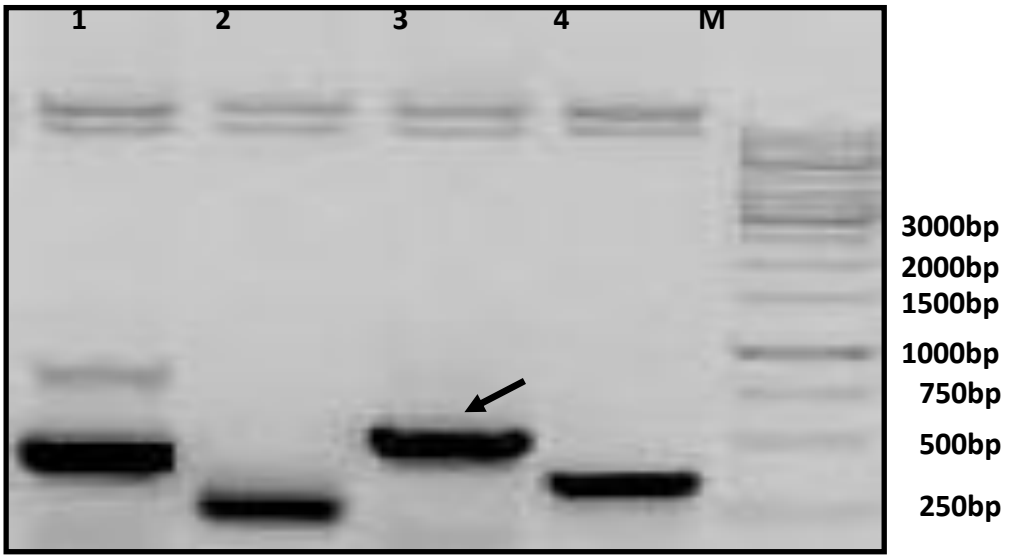


Figure 9. PCR amplification of fragments from PEG101-miR840 by designed primers to optimize the result, Lane 1: JF46 FWD + YFP_as (expect size: 417 bp); Lane 2: JF47 REV + YFP_as (expect size: 245 bp); Lane 3: JF46 FWD + UH107 (expect size: 517 bp); Lane 4: JF47 REV + UH107 (expect size: 417 bp). M: 1 kb Ladder

3.1.2.2 Screening for the positive clone by colony PCR

After transformation of chemically competent Agrobacterium cells, five of the colonies were selected for screening for positive by colony PCR, and the primer pairs are JF46 FWD + UH107 (expect size 514 bp). Result demonstrated that all of them were correctly transformed and can be infected into Arabidopsis (**Figure 10**).

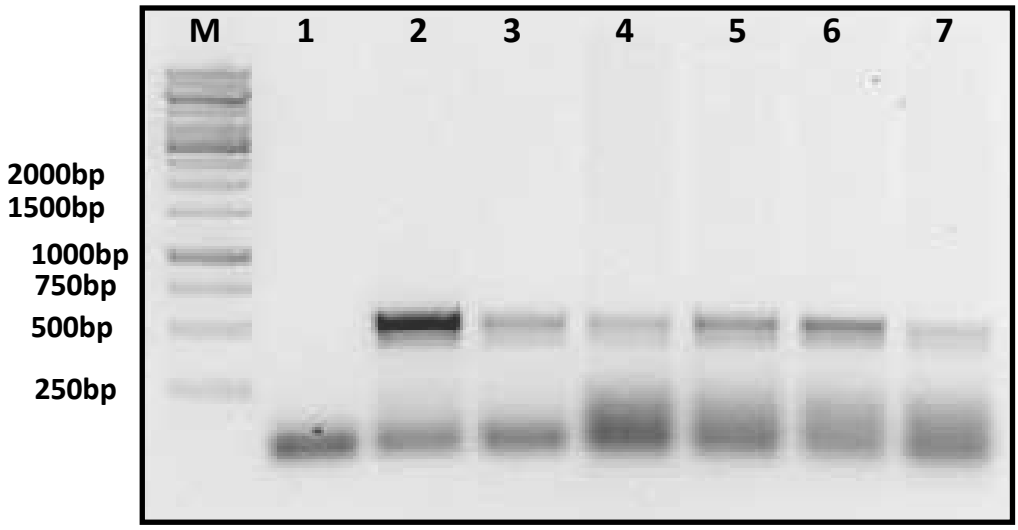


Figure 10. Verification of the positive clone by colony PCR, primer pairs were JF46 FWD + UH107 (expect size 514 bp). Lane 1, negative control; Lane 2, positive control; Lane 3-7, five colonies picked for screening. Every the colony showed positive band and the size of the all the bands matched up to the expected size. M: 1 kb Ladder

3.1.2.3 Screen for T1 miR840-overexpression lines

After successful transformation into Arabidopsis, seeds from the T0 generation were sowed. All of the transgenic plants were grown under eight hours short day conditions in the first five weeks and then transferred into sixteen hours long day (LD) conditions. When the plants were grown for about three to four weeks, the herbicide BASTA was used to select the plants which were infected successfully. After the treatment with herbicide BASTA, there were more than 100 individuals who contained Basta resistance gene can survived, then firstly twenty to fifty individuals were picked randomly to be screened by PCR, finally almost all of them are positive transgenic individuals.

When the transgenic plants were grown for about seven to eight weeks, eight to ten

rosette leaves were collected from each one to extract genomic DNA. We gathered the leaves of Primi transgenic plants in the tenth and eleventh week. Using primers JF46 fwd + UH107 (expected band size: 514), the first 21 lines were screened and almost all of them had the expected band except for line12 (lane 14 in **Figure 11**). The result indicated that miR840 gene has been successfully transformed into Arabidopsis.

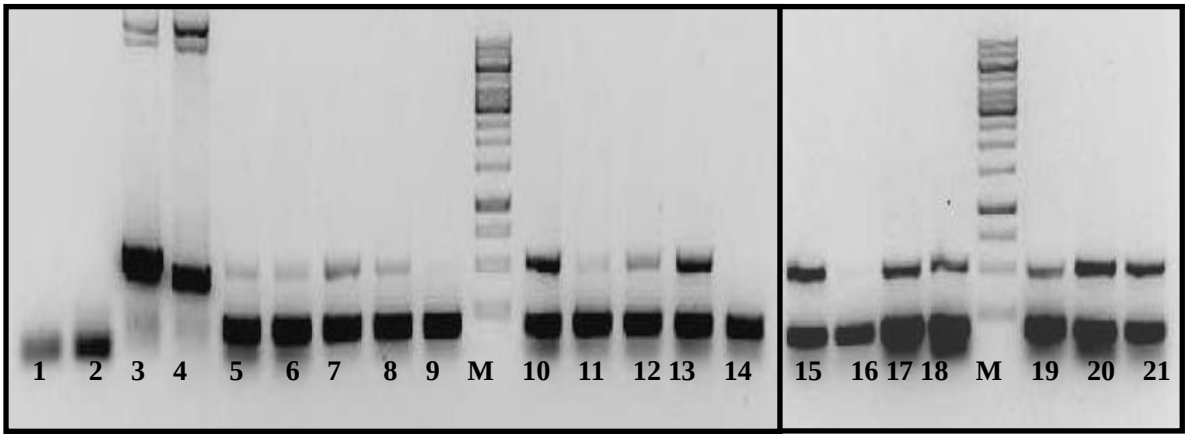


Figure 11. Analysis of miR840-overexpression lines by screen PCR. Lane 1: negative control. Primers: JF46 fwd & UH107; Lane 2: negative control. Primes: JF46 fwd & YFP; Lane 3: positive control by amplifying the Primi expression vector. Primer: JF46 fwd & UH107; Lane 4: positive control by amplifying the Primi expression vector. Primer pairs: JF46 fwd & YFP_as; Lane5-9, 10-14, 15-18 and 19-21: PCR amplification of fragments from each T1 plants, respectively; M: 1kb DNA ladder.

3.1.2.4 The phenotype of T1 transgenic plants

When the T1 seeds were grown firstly in 5 weeks under short day condition (8 h day light), and then about 10 weeks under long day condition (16 h day light), we found that the phenotype of two transgenic lines was quite different from wild type plants (**Figure 12**). The stem of the Primi-1 transgenic plant was shorter than WT and its rosette leaves were smaller than wild type and Primi-2 plant (data not shown). Comparing to the wild type plants, the vegetative cycle and flowering time of Primi-2 plants was obviously later and the petal was also different from WT. Two most

obvious phenotype of Primi-2 are: 1. the color of rosette leaves was purple, indicating that the chlorophyll content of leaves was changed or other stress happened. 2. the stem growth of Primi-2 has been extremely suppressed, we can see Primi-2 only had several small stems, the main stem still nearly not growed. **(Fig. 12)**.

Figure 12. The phenotype comparison of wild type and Primi transgenic plants, T1 progeny.

3.2 Construction and Verification of Primimut expression vector for overexpression analysis

3.2.1 Construction of pEG102-mismatch-miR840 expression vector (Primimut Expression pEG102)

As mentioned above, our former group member has constructed the clone of mutated miR840 target (mismatch-miR840) target into pEG102 expression vector via LR reaction, transformed it into Arabidopsis, and the T2 seeds also have been collected by her. But during the PPR and Primi vectors re-preparation process, to be on the safe side, the Primimut expression clone has also been prepared from the LR reaction directly again by myself, though later the Primimut line has been verified positive (the same situation to Why 3 line which is discussed in **3.4**).

3.2.1.1 Screening for the positive clone by colony PCR

The construction of the expression vector was performed by LR reaction through the way mentioned above and new Primimut expression vector (expression clone pEG102-mismatch-miR840) was transformed into *E. coli* DH5α. Using UH107 and JF46 fwd as primer pairs, fifteen colonies were screened by colony PCR. There was a positive band in the lane13 and its size was also similar to the expected size (expected band size: 421 bp), as shown at the black arrow, indicating that it may be a possible positive clone (**Figure 13**).

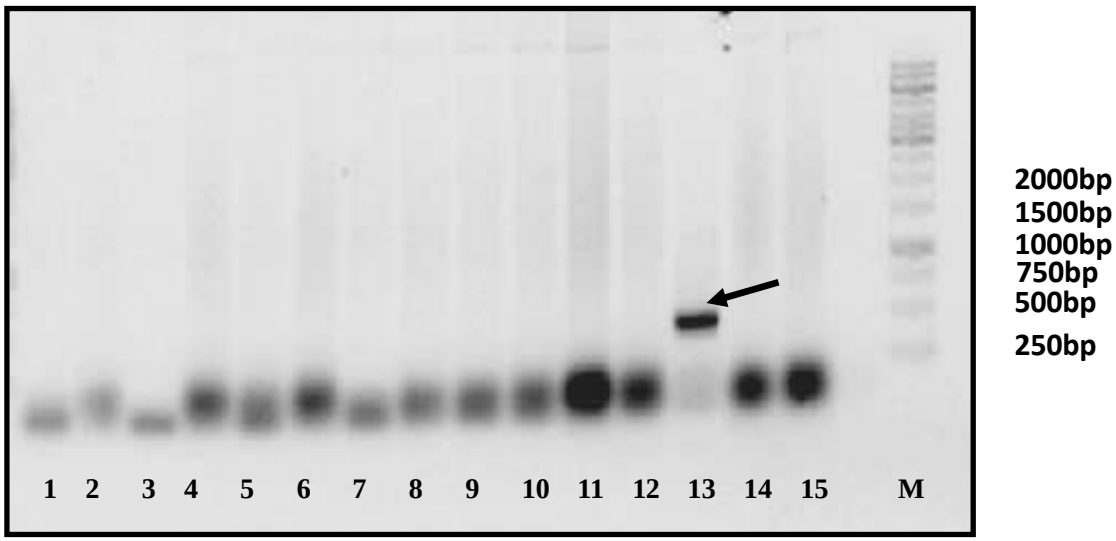


Figure 13.Colony PCR to screen for the positive clone (Primer pairs: JF46 FWD + UH107, expected band size: 421 bp). M: 1 kb Ladder

The possible positive clone was picked and then incubated in LB culture medium overnight. The concentration of plasmid was check by NanoDrop after plasmid isolation, the result is 274 ng/µl.

3.2.1.2 Digestion reaction of the Primimut expression vectors with subcloned Primimut fragment

To further identify whether the clone was correct and inserted in the right orientation, the plasmid was digested with the specific restriction enzyme XhoI. With the positive clone, the plasmid would be cut into two parts, one was the fragment contained the target gene and the other one should be the backbone of pEG102. The result showed that the size of target fragment was a little larger than expected (expected size: 1084 bp), but the backbone seemed correct (expected size: 10005 bp, as shown at the black arrow, **Figure 14**). The plasmid was then sent to sequence by using primers 35S_S and YFP_as and the sequencing result is correct.

As **Fig.14** showed, the lane 1 and 2 are digest tests for expression vector of Why 3 and Primi, also by XhoI, the expected bands sizes of Why 3 expression vector: 1620 bp (gene fragment part) and 100056 bp (backbone). The results of Why 3 and Primi are not so in reasonable sizes, the Primi expression vector finally has been got a correct one **(Fig.8)** and in the following part, the Why 3 will be deeply discussed.

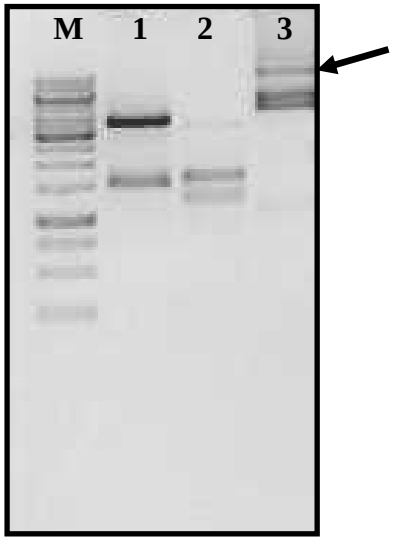


Figure 14. Digestion reaction of possible positive clone. Lane 1: expression vector of Why 3 overexpression; Lane 2: expression vector of miR840 overexpression; Lane3: expression vector of Primimut. All the vectors were digested by XhoI; M: 1 kb ladder. The arrow indicated the digested backbone of PEG102.

3.2.1.3 Optimization of PCR conditions

In order to get the specific band of Primimut in a good quality from PCR, we redesigned several pairs of primers to amplify Primimut fragment. After amplification, the result showed that 35S_S + JF46 rev primers were the best choice. Then this pair of primers was used in a gradient PCR to get the most appropriate annealing temperature (**Figure 15**). All tested temperatures can produce the specific single fragment. Ultimately, 55^{o}C was chosen as the annealing temperature for amplifying Primimut expression clone to continue the following procedures.

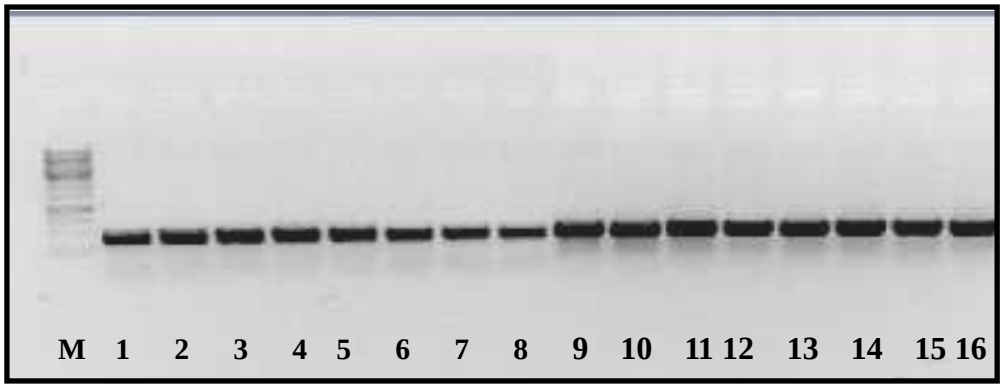


Figure 15. Optimization of Experimental Conditions on amplification of Primimut. Lane 1-8, primers 35S_S and JF46 REV, the expected size is 378bp. Temperature from lane 1-8: 60^{o}C, 59 ^{o}C, 57 ^{o}C, 54,3 ^{o}C, 50,9 ^{o}C, 48,1 ^{o}C, 46,1 ^{o}C, 45^{o}C; Lane 9-16: primers JF46 FWD and YFP_as, temperature from L 60^{o}C-45^{o}C, the expected size is 421bp. Temperature from lane 9-16: 60^{o}C, 59 ^{o}C, 57 ^{o}C, 54,3 ^{o}C, 50,9 ^{o}C, 48,1 ^{o}C, 46,1 ^{o}C, 45^{o}C. M; 1 kb Ladder.

3.3 Overexpression analysis of PPR transgenic plants and re-construction of PPR expression vector

3.3.1 Analysis of overexpressing plants

Since the PPR gene has been transformed into Arabidopsis in our lab and we have already gained the T2 seeds, we wished to identify the positive transgenic plant and detect the expression level. To detect the respective fragment we used 35S_S and GFP_as primers (the length of expected amplification product: 2003 bp), JF 53 FWD and JF53 REV as the primers of positive control (the length of expected product: 323 bp). (**Figure 16**).

The result showed that no plants with the expected amplification product were obtained, and finally the sequencing result showed the PPR gene coding region lost several bases, the problem started from the entry vector which has lost those several bases. So unfortunately, all of the procedures have to be designed again and reconstructed a new overexpression vector in order to get the right insertion.

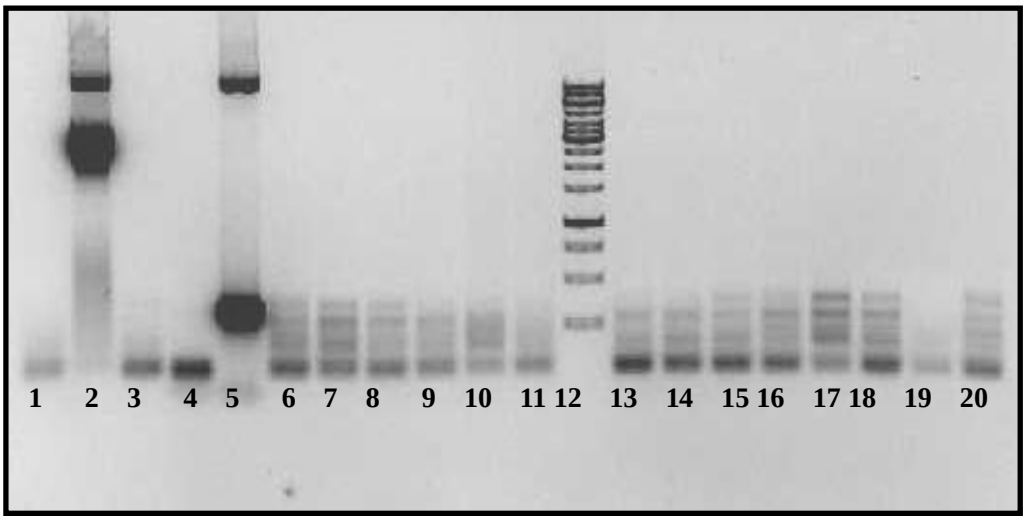


Figure 16. Identification of the PPR positive transgenic plants by PCR. Lane 1: Negative

control (no template); Lane 2: Positive control-Vector (35S_S and YFP); Lane 3: Negative

control-WT (no primer); Lane 4: Positive control-WT (JF 53 FWD and JF53 REV); Lane 5: Positive control-Vector (JF 53 FWD and JF53 REV); Lane 6-11: Samples (1-6); Lane 12: Ladder: 1 kb; Lane 13-20: Samples (7-14); 1 kb Ladder.

3.3.2 Construction of the new PEG101-PPR expression vector (PPR Expression pEG103) for overexpression analysis.

3.3.2.1 Amplification of original PPR entry vector for creating new entry vector by phusion PCR.

Because the original entry vector lost 4 bp in the PPR gene coding region, for getting the complete gene coding region, through the designed new primers and used the original entry vector as the template, the phusion PCR was run to amplifying the original entry vector for creating the new and correct entry vector.

Figure 17A showed that the new PPR entry vectors were created successfully by phusion PCR amplification (expected band size: 4440 bp). And then new PPR entry clone vectors were treated by DpnI for digesting the original templates, then the new entry vectors were ligated through the ligation reaction (expected band size: 4440 bp, **Fig.17B**) which has been described in **2.2.7**.

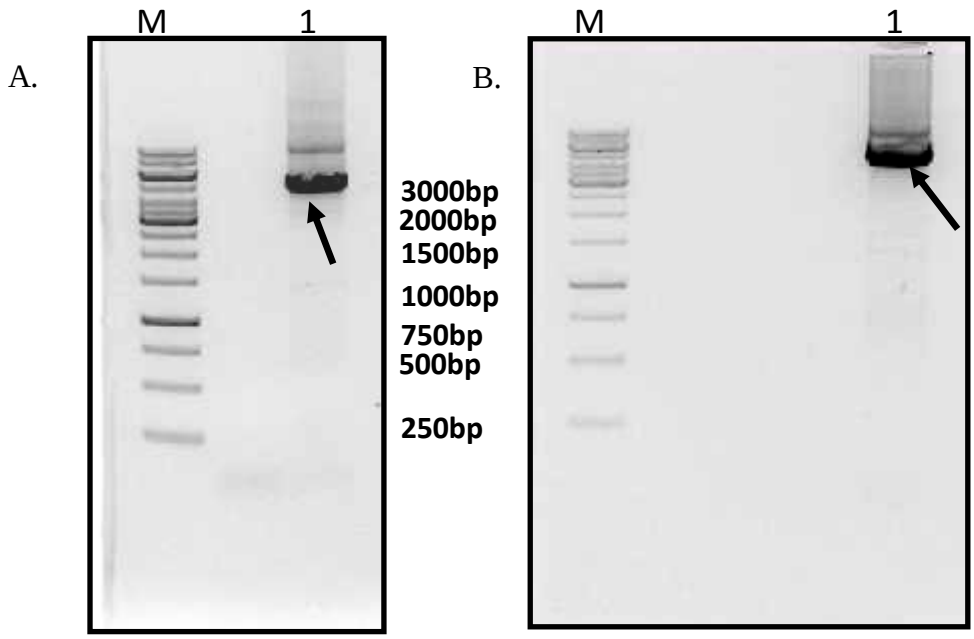


Figure17. A. Construction of new pENTR-PPR entry clone vectors by the phusion PCR amplification. Lane 1 was the product of phusion PCR (Primers: JF116 FWD + JF116REV_3, the length of expected band: 4440 bp). The arrow indicated the objective band of PPR gene. **B. After ligation reaction to the new PPR entry clone vectors.** Lane 1 indicated the constructed vectors which have been ligated (expected band size: 4440 bp). Lane M: 1 kb ladder.

After transformation into chemically competent *E. coli* DH5α, two clones were selected randomly and incubated overnight in LB culture medium. The concentrations of the plasmids were 289, 0 ng/µl and 420, 8 ng/µl, respectively. By sequencing with M13 fwd and M13 rev primers, both of them were shown to contain correct insertions.

3.3.2.2 Construction of the PPR expression vector by new PPR entry vector

We used one of the plasmids to clone the PPR gene from the new PPR entry vector into the destination vector pEG103 to make the new PPR expression vector by LR reaction. Then the positive LR-products were transformed into *E. coli* DH5α, and later colony PCR was done for screening the positive colonies, but it didn't function. So 4 colonies were picked randomly and these colonies were treated by Plasmid Miniprep (extract plasmid). And then we directly run a PCR to check the plasmid DNA from the Miniprep by using JF116_3 FWD and JF116_3 REV Primers, the elongation time we set is 3, 5 min, the expected band size is around 3 KB. (**Figure 18**). The result showed that all the colonies showed positive bands and band sizes similar to expected size. Because the time for the project is limited, the sequencing analysis hasn't been performed, this step will be done later.

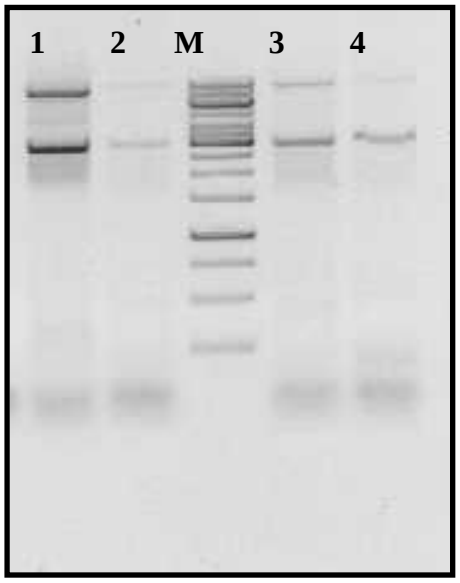


Figure 18. Analysis of PCR for screening positive clones. Lane 1 & 2 + Lane 3 & 4: plasmid DNA from selected colonies for PCR. M: 1 kb ladder.

3.4 Construction and Verification of Why 3 Expression vector for overexpression analysis

As mentioned above, the Why 3 transgenic plant line (T2) has been previously prepared by the former group member and finally has been verified that this line was positive. The prepared Why 3 transgenic line was directly used to perform the expression level analysis and the result is stated in **3.5**.

The Why 3 expression vector also prepared again from the LR reaction during the other 3 vectors preparation process.

3.4.1 Construction of why3-PEG101 expression vector (Why 3 expression vector) through LR reaction

The process of LR reaction was introduced in Materials and Methods. After LR reaction, the LR products have been transformed into *E. coli* DH5α, fifteen colonies were selected to analyze by colony PCR by using the primers JF3 FWD + UH107 (expected band size: 712 bp), one of them produced the expected band with the appropriate size, as shown at the arrow (**Fig. 19A**).

Two copies of plasmid from the positive colony were then isolated and digested by XhoI to confirm the insertion of the fragment (expected fragment size: 1620 bp and the expected backbone size: 10005 bp), both of the plasmid copies produced the same expected bands with the appropriate size, as shown at the arrows (**Figure 19B**). And then we sequenced this plasmid using the GFP + 35S_S primer pairs, the result confirmed that it could be used in next transformation

Fig.19A

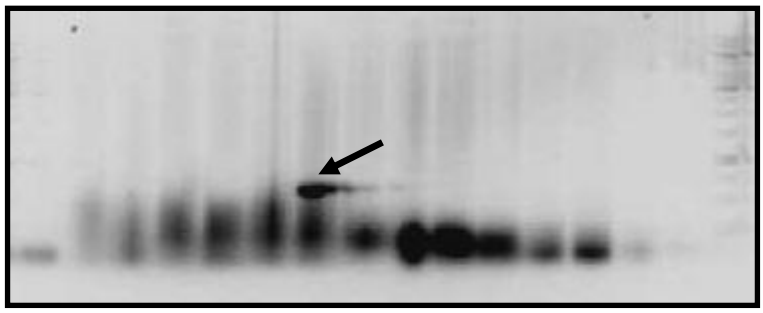


Fig.19B

Figure 19. A. Analysis of colony PCR for screening positive clone, 1 kb Ladder. **B. Digestion reactions of why3-PEG101 vectors**, both the results are positive; 1 kb Ladder.

3.4.2 Optimization of PCR conditions.

To obtain the best results and get the specific objective band, several pairs of primers were used to optimize our experimental content. The objective fragments of expected sizes were amplified by different primers and all the primers pairs specifically produed the expected bands with the appropriate sizes. As the arrow showed in **Figure 20A**, the 35S_S and JF3_FWD primers were the final choice (expected band

size: 833 bp). Then this pair of primers was used in gradient PCR to obtain the most appropriate annealing temperature (**Figure 20B**). All of them can produce obvious specific single fragment. Ultimately, we chose 55°C as the best temperature to continue the following procedures.

Fig.20A.

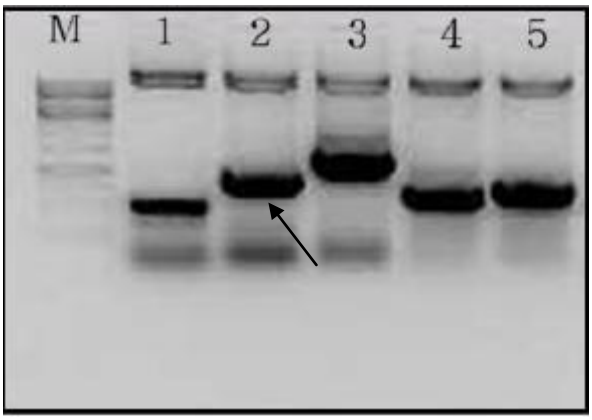


Fig.20B

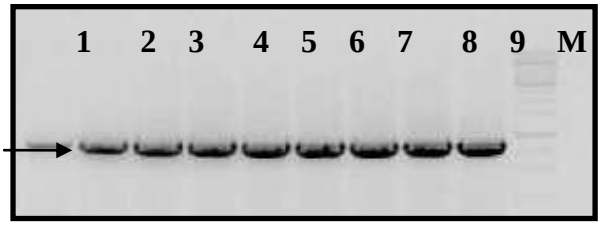


Figure 20. Optimization of Experimental Conditions on amplification of Why3.

A. PCR products by different primers followed. Lane 1. JF3 FWD + JF3 REV, expected band size: 348 bp; Lane 2: 35S_S + JF3 REV, expected band size: 833 bp; Lane 3: 35S_S + UH107, expected band size: 1198 bp; Lane 4: JF3 REV and YFP_as, expected band size: 651 bp; Lane 5: JF3 FWD + UH107, expected band size: 712 bp. **B. Gradient PCR product by 35S_S and JF3 REV primers.** Temperature was designed from 60°C-45°C. The arrow indicated the objective band. M: 1 kb Ladder

3.5 The detection of expression level of Why3 and Primimut transgenic plants

As mentioned above, after the screening of all the 4 transgenic plants lines which have been prepared previously, we found Primimut and Why 3 lines are positive. So these two lines were directly performed the expression level analysis.

3.5.1 RNA extraction of WT, Primimut and Why 3 lines

Since we have already had the progeny of Arabidopsis transformed with Why3 and Primimut, the next setup was to detect the expression at the transcription level. Using the method of RNA isolation mentioned above (**2.2.11**), we successfully extracted RNA from Why3 and Primimut transgenic plants without obvious degradation. The RNA from wild type (WT) was obtained simultaneously as the control (**Figure 21**). The RNA concentration was also detected by Nanodrop test (**Table 9**).

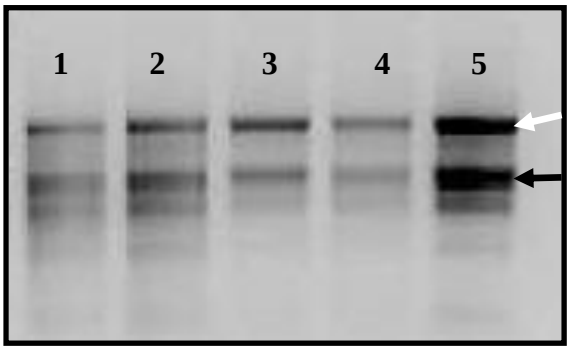


Figure 21. Gel electrophoresis to detect the quality of isolated RNA. Lane 1, wild type sample; Lane 2 & 3: 2 Why3 samples from different Why 3 individual (Why 3_1 & Why 3_2); Lane 4 & 5: Primimut samples from different Primimut individual (Primimut_1 & Primimut_2). The white arrow showed the 28S rRNA, and the black arrow indicated the 18S rRNA.

Table 9. The concentration of RNA.

Line	WT	Why3-1	Why3-2	Primimut-1	Primimut-2
Concentration (ng/μl)	506.1	566.3	342.1	253.3	1054

3.5.2 cDNA synthesis

The cDNA synthesis was performed according to the methods shown in the **2.2.17**, and then we run a gel to detect the quality of cDNA (**Figure 22**). As shown in the figure, the 5 clear bands (Lane 1-4 & Lane 6) showed that all of the five cDNA samples were produced and indicating that quality was sufficient. However, the all the no reaction controls (RNA samples have been treated by DNase before cDNA synthesis, and then treated them also in the cDNAs synthesis step without adding reverse transcriptase, for checking quality of the DNase step) which have also produced not particular evident bands. Considering the genomic DNA had been eliminated thoroughly in the RNA samples, this may be caused by the contamination of pipette tips or other operation, so we can still use the cDNA samples to do the Real-time PCR.

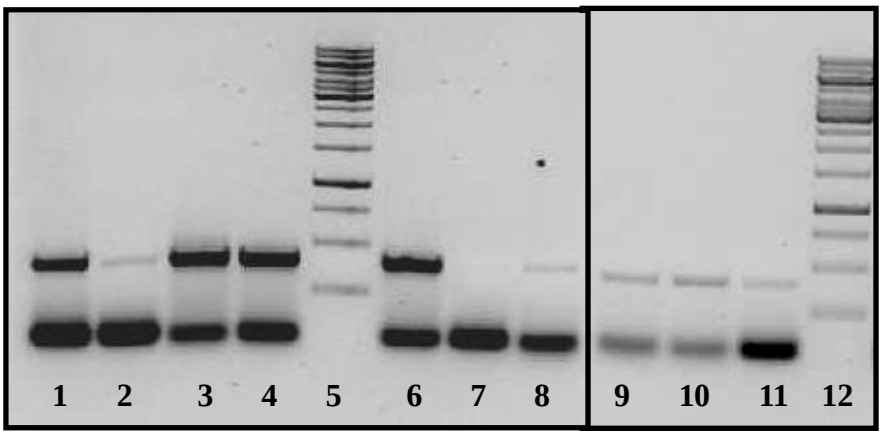


Figure 22. Gel electrophoresis to detect the quality of the cDNA. Lane1, WT cDNA sample; Lane 2: cDNA from Why3 sample 1(Why3_1); Lane 3: cDNA from Why3 sample 2

(Why3_2); Lane 4: cDNA from Primimut sample 1 (Primimut_1) 6, Lane 6: cDNA from Primimut sample 2 (Primimut_2); Lane 5: 1 kb ladder; samples; lane 7-11, no reaction control for the five samples, respectively (WT, Why3_1, Why3_2, Primimut_1, Primimut_2); Lane 12: 1 kb Ladder.

3.5.3 Real-time PCR to detect the transcription expression level.

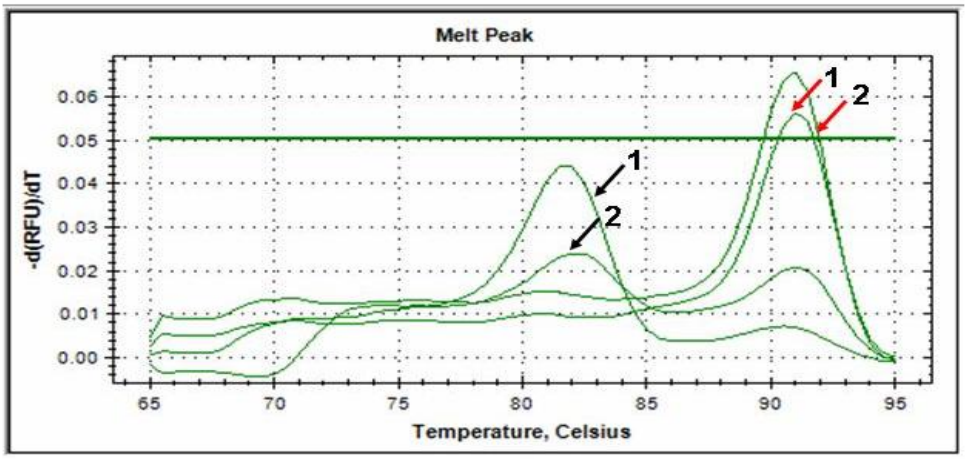
The experiments above showed that the cDNA synthesized could be used for our future work, so we took the Real-time PCR reaction to detect the transcription expression level of transgenic and wild type plants. The procedure was performed according to the methods above (**2.2.18**)

3.5.3.1 Overview of the quality of Real-Time PCR result

Since the time for the project was very limited, the Real-Time PCR just has been performed preliminarily, the PCR program and other factors related to the quality of the result have not been optimized. So this result of Real-Time in my project is not accurate and not the final result.

For showing the result of the Real-Time PCR is not precise and reliable enough, as shown in the **Figure 23,** we can obviously see the both duplicated samples have shown two peaks in the Meal Peak curve analysis, that means the result is not precise, the relative problem must be checked and the PCR must be optimized. In this figure, red arrow pointed out the possible correct melting peaks, but the black arrow pointed out the likely wrong peaks from the unknown.

melting curves of PPR



melting curves of Why3

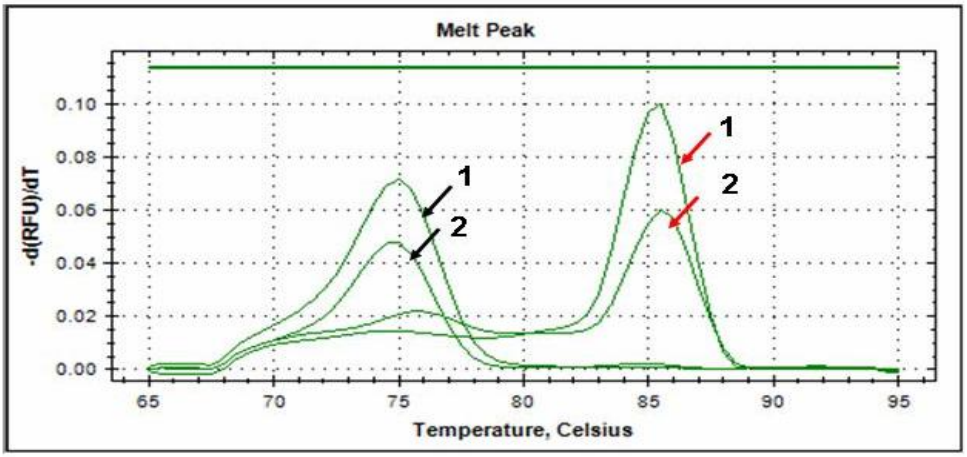


Figure 23. Melting curves of objective gene by real-time PCR. The number in the same figure indicated double repeated experiment in the same transgenic line. the arrows with red color indicated the right peak, and the black arrows showed the wrong peaks.

Although the result of Real-time PCR is not precise enough, from data of the result, we can still see the there is an obviously tendency which miR840 may have regulative function to Why 3 and PPR. In the following section, **3.5.3.2**, the tendency will be analyzed.

3.5.3.2 Tendency relative expression of Why3 and PPR in Primimut transgenic plants

In order to identify the possible target gene of miR840, the expression level of Why3 and PPR in Primimut transgenic plants is very important. To get the correct information, we designed specific primers refer to the **Table 7**. Interestingly, the results indicated that the transcriptions of PPR and Why3 were both significantly up-regulated in Priminut transgenic plants (**Figure 24**). These results implicate that the mutation of miR840 in Priminut transgenic plants may induce, and, not repress the expression of PPR and Why3, suggesting that there is a tendency which both of the two genes maybe the target of miR840 and, moreover, would be down-regulated by miR840.

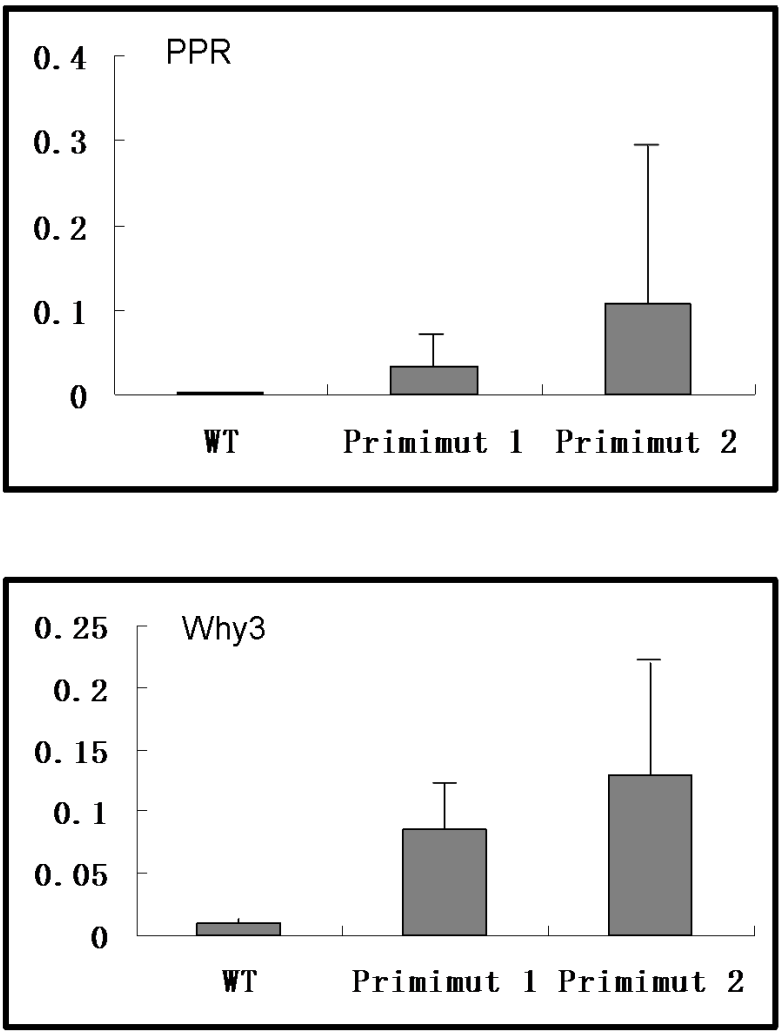


Figure 24. Real-time PCR analysis of transcript levels of Why3 and PPR in Primimut transgenic plants. Expression was normalized to that of Actin. Transcript levels from the wild type (WT) were set as the control. Error bars indicate ±SD (n=3)

3.5.3.1 Tendency analysis for relative expression of Why3 in transgenic plants

To identify the transcript level of Why3 in transgenic plants, the cDNA of why3_1and why3_2 (leave sample of Why3_1 and why3_2 was taken from each different plant in Why3 line) was used as the template and WT was used as the control. As shown in **Figure 25**, the relative expression was much higher in both of why3_1 and why3_2 from transgenic line Why 3 than in wild type, especially in Why3_1.

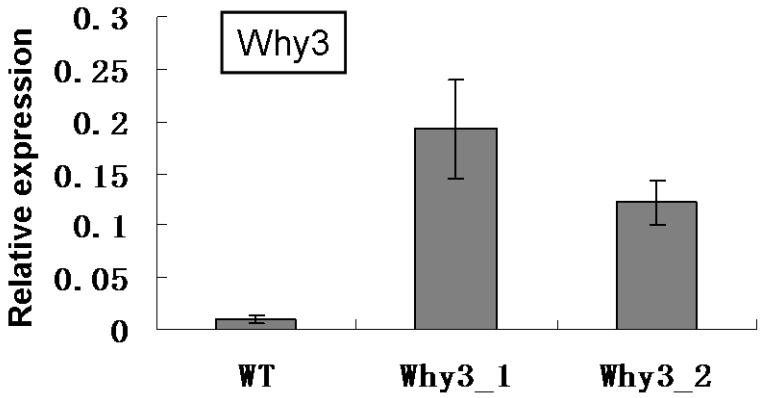


Figure 25. Relative expression of Why3 in transgenic plants and wild type by real-time PCR. Expression was normalized to that of Actin. Error bars indicate ±SD (n=3). The primers referred were mentioned in methods.

3.6 Computational predictions on subcellular localization of the PPR protein

As mentioned above, prediction of the subcellular localization of PPR protein was performed by online website. Two websites we mainly used are TargetP Server and Predota. The predictive results from the two websites were slightly different owing to the distinct alignment and design procedure, however, we can still find the relationships with the their predicted results

The TargetP 1.1 Server (http://www.cbs.dtu.dk/services/TargetP) is the most well-known database as a prediction service. The method used has been verified on Arabidopsis, and the assignment is based on the presence of any of the N-terminal presequences: chloroplast transit peptide (**cTP**), mitochondrial targeting peptide (**mTP**) or secretory pathway signal peptide (**SP**). If the input sequences contain the N-terminal signal, a potential cleavage site can also be predicted.

The Predotar online service (http://urgi.versailles.inra.fr/predotar) was designed for computational screening of large batches of proteins for identifying targeting sequences. The obvious merit is that it has a very lower rate of false positives compared with similar programs.

The result from the TargetP which is mainly for (ER) signals showed that the PPR protein most possibly located in mitochondrial with the value of mTP 0.628. The same result was also found in the Predotar online service, the value of mitochondrial locating prediction was 0.37. Both of the online service gave us the same estimate, indicating that the PPR protein contains a mitochondrial targeting sequence and most probably located in the mitochondrial. Certainly, this prediction needs verification by experiments.

4. Discussion

4.1 General discussion

In the last few years, molecular genetics has been widen and greatly enriched with the discovery of small RNA molecules. Previous work confirmed that all of these small RNAs would not been translated into proteins. They interact with their target genes leading to site-specific cleavage or translational repression of the mRNA transcripts. Having essential functionality and capability, small RNAs serve as powerful tools for molecular biologists in manipulating expression of the gene of interest. In addition, recent evidence indicated that these small RNAs are playing important roles in developmental timing, differentiation and disease development. Although there is accumulating information on the biology of small RNAs, the understanding of small RNAs is far from complete, undoubtedly providing limitless opportunities for various fields of researches on small RNAs.

MicroRNAs (miRNAs) are a large group of small RNAs. In both plant and animal, miRNAs function as small single stranded molecules of only ~19-24 nucleotides in length. These are processed from stemloop structures harbored in primary MIRNA transcripts (Bartel 2004). A miRNA binds to its mRNA target primarily through sequence complimentary with the 3'-untranslated region (3'UTR), triggering the event of either translational repression or degradation of the mRNA transcript (Start et al., 2005).

Most Arabidopsis miRNA families consist of several members (1-14), and the majority of precursors can efficiently produce small RNAs. The molecular mechanism of miRNA function requires base-pairing between miRNAs and their target mRNA genes. Since only limited complementarity is sufficient for target recognition in animals, animal miRNAs tend to have hundreds of targets, which are typically

inhibited at the translational level (Brennecke et al. 2005; Lewis et al. 2005). Where perfect complimentary base pairing mostly exists in plants, it is believed that most plant miRNAs trigger mRNA degradation by cleavage (Axtell & Bowman, 2008; Mallory & Bouche, 2008), though recent studies have suggested that miRNA repression mechanisms in plants are widely spread (Brodersen et al., 2008). Their analysis revealed that miRNA-guided silencing in plants had a very widespread inhibitory component faraway from endonucleolytic cleavage. However, the most basic mechanism of miRNA even requires precise complementarity.

MiRNA expression has also been studied by northern blotting, in situ hybridization, as well as microRNA microarrays (Aukerman and Sakai 2003; Axtell and Bartel 2005; Jones-Rhoades and Bartel 2004; Mallory et al. 2004a; Palatnik et al. 2003; Reinhart et al. 2002). All three methods reflect abundance of complete miRNA families and cannot distinguish individual miRNA family members, since they rely on hybridization of mature miRNAs to complementary nucleic acids. To test the authenticity of the observed patterns of GUS activity for the different putative miRNA promoter, a novel method called stem-loop RT-PCR might be suitable. This method can amplify individual miRNA stem-loops with high specificity and sensitivity (Chen et al. 2005), however we have not yet tested its performance on our miR840 in our lab.

4.2 Overexpression analysis of miR840 (Primi) transgenic plants

The best and convenient way to illuminate the function of specific gene is forward genetics. By knocking out or inserting a destroyed sequence into the specific gene, we can get the relevant loss-of-function (LOF) mutant. However, LOF is not all-powerful forever, it also has own bug. If the gene of interest is a member of a big gene family, knock out of a single gene may not generate the expected phenotype because gene redundancy produces homologous genes. Since the precursors of miRNAs are always

small and there are many members in a family, this problem also happens in microRNA family. Having investigated in all of the databases carefully, we have not found an appropriate knock-out mutant for miR840. So we developed the recombinant overexpression vector to find out the function of miR840. By constructing functional miR840 precursor into expression vector PEG101, we got miR840-overexpression transgenic plant.

4.2.1 Phenotype analysis of Primi plants T1 progeny.

Considering we have already obtained the T0 seeds of Primi transgenic plants, my first work was to identify the positive transgenic plants and analyze the relevant phenotype in the first several months. However, even though all of the T0 seeds were used up, I was not able to identify one positive transgenic plant. Therefore, the generation of an overexpressing line had to be started again

Fortunately, the reconstruction of miR840 expression vector was very well and the T2 seeds have been obtained up to date. When the positive Primi transgenic plants T1 progeny was transferred under long day conditions, we found there were some obvious phenotypes quite different from wild type plants. As is mentioned above, Primi-1 and Primi-2 lines presented different phenotypes. For Primi-1 line, the stem is shorter and the rosette leaves are smaller than that of wild type. The phenotype was very similar to the transformed plants with the genes relevant to auxin response. Previous evidence showed that overexpressing auxin responsive gene in rice could lead to short stem, small leaves and later growth cycle. The vegetative cycle is a little bit later than wild type. For the Primi-2 line, the physiological phenotype is more obvious. The vegetative cycle is quite later than that of wild type.

4.2.2 The probable mechanism of phenotype

Considering the target gene of miR840 maybe Why3 or PPR gene or even both of

them, miR840 may down regulate their transcription expression levels. We would like to compare the phenotype of Primi plants with the Why3 or PPR mutunt. According to the previous results, the physiological phenotype of miR840-overexpressing lines has not matched with neither Why3 nor PPR mutunt. This illogical result may raise a new issue or a new path to follow, whether there were other genes regulated by miR840. By comparative analysis on the genome database, we found some important transcription factors locating near miR840.

There is a hypothesis that the overexpression of transgene may not only increase the transcript level of objective gene, but also would disrupt other genes, so this may back to the basic question mentioned above, whether the phenotype is due to the overexpression of miR840. If not, whether there is any relationship between miR840 and the disrupt gene.

4.3 Overexpression analysis of mismatch-miR840 (Primimut) transgenic plants

Meanwhile, insertion of a wrong sequence led to abnormal miR840 which could not develop his normal function. We constructed it into PEG102 expression vector and named Primimut. So the function of miR840 may be demonstrated by comparing the two overexpression system.

Considering we had the T1 progeny of Primimut transgenic plants, the first setup was to detect the expression level of Primimut, however, the result of real-time PCR is not ideal. The melting peaks were also generated in the negative control reaction and the Ct value of normal reaction was not able to analyze. There may be several reasons led up to failure, the first one was a possible of the pipette tips or the water. The second reason maybe the non-specificity of the qPCR reaction primers. We have redesigned

new primers in order to get the real result, but the analysis could not be repeated due to time constraints.

4.4 Overexpression analysis of Why3 transgenic plants

Several reports showed that the Whirly proteins are transcription factors which have an effect on DNA metabolism, such as regulating transcription level and modulating the length of telomere. In Arabidopsis, all members of Whirly family are encoded by the nuclear genome. Evidences indicated that Why1 and Why3 were targeted to plastids and Why2 was targeted to mitochondria. However, the mechanism of Why3 protein has not been discovered completely to date. Maréchal revealed that Why3 may play an important role on plastid genome stability by protecting against illegitimate repeat-mediated recombination in 2009. Meanwhile, Why3 may regulate the transcriptional repression of the kinesin gene *AtKP1* as one of the two components of KPRE-binding factor 1 (KBF1) (Xiong et al. 2009). Our purpose focused on the potential physiological function of Why3 protein by reverse genetics. Overexpression of Why3 gene may lead to abnormal phenotype.

As shown in the **Figure 25,** the results indicated there is a obvious tendency that the relative expression of Why3 was up-regulated significantly in the transgenic plants, suggesting that the Why3-overexpression vector has been successfully constructed and the positive transgenic plants have been obtained. Both Why3_1 and Why3_1 lines can be used in next experiments.

4.5 Overexpression analysis of PPR transgenic plants

Our research of PPR the gene is under way and would be completed soon. Since there was evidence that PPR proteins maybe interrelated with various molecular events, including in RNA splicing or cleavage, RNA editing, RNA stability, enhancing and blocking translation of RNA in the chloroplasts and mitochondria, we would like to find out by subcellular localization of the PPR protein encoded by protein-coding gene At2g02570. At the same time, the PPR gene constructed into pEG103 could reveal the temporal and spatial expression in Arabidopsis. We can also observe possible abnormal phenotypes of the transgenic plants by overexpressing PPR gene.

However, in our previous work, we did not identify the positive transgenic plants in T1 progeny. Then we sequenced the expression vector and found that the insertion was not correct basically. So I had to reconstruct the PPR-overexpression vector. The next setup would be to transform the plasmid into Arabidopsis and analyze the phenotype.

Taken together, all of the four experiments were done to focus on the function and target selection of miR840. As mentioned above, Searching the TAIR (http://www.arabidopsis.org) and TIGR (http://plantta.jcvi.org) databases give two different annotation versions for this *AtWhirly3* gene, differing at the length of the 3' untranslated region. Although Rajagopalan et al. predicted the target gene was Why3, this bioinformatic contradiction may even give two completely different results for potential target gene of miR840. Transgenic plants were designed to solve this mystery. If the phenotype of overexpressing Primimut transgenic plants could correspond to the phenotype of anyone of the Why3 or PPR transgenic plants, we may reach some tentative conclusions about target selection.

4.6 The possible target gene identified by real-time PCR

To investigate the expression of PPR and Why3 genes in Priminut transgenic plants, real-time PCR analysis was conducted. As shown in the **Figure 24**, the tendency to transcript level of both of PPR and Why3 were up-regulated in Priminut transgenic plants, indicating that the miR840 can influence both of their transcript level. As a member of microRNA family, this result probably means that miR840 can target both of then function as transcriptional silencing. When considering the predicted evidence from TIGR (http://plantta.jcvi.org) database, this result can be explained by bioinformatics.

However, this real-time PRR analysis was preliminary and not very precise because the NTC (no template control) showed signal, we regarded the reason may be the primers formed dimers during the real-time PCR is running. We have designed the new primers for performing the real-time PCR again to get more accurate result.

4.7 Future work

Generally, for the future, more genetic analyses including the expression level of these three genes are planned. In addition, overexpression lines of Primi, Primimut, Why3 and PPR gene are ready for analysis (in preparation). Especially, the expression level of miR840 overexpression lines will be analyzed later, comparing to the expression level of WT, for checking whether the expression level of PPR and Why in Primi obviously less than in WT. If the result shows when the miR840 overexpress, the Why 3 or/and PPR expression are reduced, then we can get another proof (except Primimut lines compare to Why 3 and PPR overexpress line) to show the miR840 has a down-regulation function to Why 3 or/and PPR.

We would observe and analyze the phenotype of all the transgenic plants, especially the analysis the phenotype of Primi transgenic plants. The protein expression level of Why3 and PPR in Primi and Primimut overexpression lines also will be detected by Western Blot technique depending on the results of the expression level analysis. The details for several aspects which we can perform the further research is shown below:

Further identification of the transcription expression of miR840. Since the microRNA was shown to mostly function on the transcriptional or post-transcriptional level, the northern-blot analysis and real-time PCR will be needed for further identification whether the expression of Primi is significantly higher than in wild type, in other words whether this physiological phenotype is due to the overexpression of miR840, which is mentioned above.

Identification of expression level of Why3 and PPR. Since the initial experiments and row data from comparing Primimut to WT could only show the tendency that both the why3 and PPR gene may be down-regulated by miR840 at the transcriptional level from the Primimut-overexpressing plants. So that we firstly need to optimize the Real-time PCR program and steps for getting the accurate result and data, if there is credible result to show there is a relationship between miR840 and Why/PPR from comparing the expression level of Primimut to WT, then we can find the direct evidence to further validate our result,. If Why3 or PPR or both of them were down-regulated in Primi transgenic plants, we could obtain the evidence of the target gene of miR840. However, not only the function of miR840 at the transcriptional level is important, the question about whether the miR840 regulate the target gene at translation level is also interesting quesion, So the identification of the translational level of Why3 and PPR by western blot should be planned.

Comparing the phenotype of the two Primi-overexpressing plants (**Figure 12**) and with that of wild type, there is a quite high possibility that the Primi-overexpressing phenotype could come from the destruction of a totally different gene which has some

important function in leaves color, stem growth and/or life cycle. So that the analysis of T-DNA integration sites (genomic analysis) is important in the next step.

To further elucidate the biochemical property and physiological function of miR840, expression of miR840 in response to various stresses by northern-blot or real-time PCR is the most direct and convenient method. In future, we would analyze the transcript levels of miR840 in young Arabidopsis seedlings treated with environmental and hormonal factors, such as salt, drought, cold, auxin, ethylene, or ABA.

It is possible that there are other factors can influenced by overexpression of miR840, one of the highest possibilities is the Auxin signal pathway related proteins. The similar phenotype was observed in the two transgenic lines, Primi-1 and Primi-2, much lower stem and later flowering time, suggesting that the overexpression of miR840 may influence some important Auxin signal related proteins, just like auxin-binding proteins, and response factors. Considering the possibility that 35S-mediated-transgene may induce the wrong insertion in other genes, we need to utilize some common methods to identify if there was important Auxin-related gene was disrupted, such as southern blot and Tail-PCR. By the method of Tail-PCR, we can quickly identify which gene was disrupted through designed universal primer.

We are looking for some genes located near the miR840 gene in the same chromatin, and the next step is to verify the transcript level of these genes by real-time PCR. Of course, these genes by ourselves may not be the objective gene. The microarray is the next work to screen for the target genes up- or down-regulated by miR840. By extracting the RNA of Primi-2 and analyzing the data by microarray technique, we can directly find which gene would be due to the phenotype.

References

Axtell, M.J., and Bartel, D.P. (2005). Antiquity of microRNAs and their targets in land plants. Plant Cell 17, 1658–1673.

Aukerman M.J., and Sakai, H. (2003)*Plant Cell* 15, pp. 2730–2741.

Bao N, Lye K W, Barton M K .(2004). MicroRNA binding sites inArabidopsis class III HD-ZIP mRNAs are required for methylation of the template chromosome. Dev Cell, Volume 7(5), pp.653–662

Bartel, D.P. (2004) MicroRNAs: Genomics, biogenesis, mechanism, and function
Cell , Volume 116, pp.281-297

Bollman K.M., M.J. Aukerman, M.Y. Park, C. Hunter, T.Z. Berardini, and R.S. Poethig. (2003) HASTY, the Arabidopsis ortholog of exportin5/MSN5, regulates phase change and morphogenesis, Development ,Volume 130, pp.1493-1504

Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., and Zhu, J.K. (2005). Endogenous siRNAsderived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. Cell, Volume 123, pp. 1279-1291.

Brennecke J, Stark A, Russell RB, Cohen SM. (2005) Principles of MicroRNA–Target Recognition. PLoS Biol 3(3): e85.

Brodersen, P., L. Sakvarelidze-Achard, *et al.* (2008). "Widespread Translational Inhibition by Plant miRNAs and siRNAs." Science 320(5880): 1185-1190.

Cappadocia L., Alexandre M., Parent J. S., Étienne L., Jurgen S. and Normand B. (2010) Crystal Structures of DNA-Whirly Complexes and Their Role in Arabidopsis Organelle Genome Repair, Plant Cell, Volume 22(6), pp. 1849-1867

Cesar L., Xie Z. X., Kristin D. K. and James, C. C. (2002) Cleavage of Scarecrow- like mRNA Targets Directed by a Class of Arabidopsis miRNA, Volume 297, pp. 2053-2056

Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, Barbisin M, Xu NL, Mahuvakar VR, Andersen MR, *et al.* (2005) Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res.*;**33**:e179. [

Desveaux D., Després C., Joyeux A., Subramaniam R., and Brisson N. (2000) PBF-2 is a novel single-stranded DNA binding factor implicated in PR-10a gene activation in potato, Plant Cell, Volume 12, pp. 1477–1489

Desveaux D., Maréchal A., and Brisson N. (2005) Whirly transcription factors: defense gene regulation and beyond, Trends in Plant Science, Volume 10, pp. 95–102

Fire, A and Mello, CC. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*， Nature, Volume 391, pp.806-811

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.

Franco-Zorrilla J. M., Adrián V., Marco T., Isabel M., María I. P., Rubio-Somoza I., Leyva A., Weigel D., García J. A., and Paz-Ares J. (2007)
Target mimicry provides a new mechanism for regulation of microRNA activity, Nature Genetics, Volume 39, pp. 1033 – 1037

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

Ghildiyal, M. and Zamore P.D. (2009) Small silencing RNAs: an expanding universe，Nature Review Genetics, Volume 10, pp.94-108

Griffiths-Jones, S., Bateman, A., Marshall M., Khanna A. and Eddy, S.R. (2004)
The microRNA registry, Nucleic Acids Research, Volume 32, pp. 109–111

Grishok A., Tabara H. and Mello C. C. (2000) Genetic requirements for inheritance of RNAi in *C. elegans*, Science, Volume 287, pp. 2494–2497

Halic M. and Moazed D.(2010) Dicer-Independent Primal RNAs Trigger RNAi and Heterochromatin Formation, Cell, Volume 140, pp. 504-516

Hamilton AJ., and Baulcombe, D. (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants，Science, Volume 286, pp.950-952

Hiraguri A., Itoh R., Kondo N., Nomura Y., and Aizawa D. (2005) Specific interactions between Dicer-like proteins and HYL1/DRB-family dsRNA- binding proteins in *Arabidopsis thaliana*, Plant Molecular Biology, Volume 57, pp. 173–188

Javier F. P., Edwards A., Wu X. L., Carla S., Rebecca S., James C. C. and Detlef W. (2003) Control of leaf morphogenesis by microRNAs, Nature, Volume 425, pp. 257-263

Jones-Rhoades M. W., Reinhart B. J., Lim L. P., Burge C. B., Bartel B, Bartel D.P. (2002) Prediction of plant microRNA targets, volume110, pp. 513-520

Jones-Rhoades, M.W., **Bartel, D.P., and Bartel, B.** (2006) MicroRNAs and Their regulatory Roles in Plants，Annual Review of Plant Biology, Volume 57, pp.19–53

Jones-Rhoades, M.W., Bartel, D.P. (2004) Computational identification of plant miRNAs and their targets, including a stress-induced miRNA, Molecular Cell, Volume 14, pp.787-799

Kidner C. A., and Martienssen R. A. (2004) Spatially restricted microRNA directs leaf polarity through ARGONAUTE1, Nature, Volume 428(6978), pp. 81-84

Kotera E., Tasaka M., and Shikanai T. (2005) A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. Nature, Volume 433, pp. 326–330

Koussevitzky S., Ajit N., Todd C., Mockler., Fangxin H., Gilberto S. M., Marci S., Jason L., Ron M. and Joanne C. (2007) Signals from Chloroplasts Converge to Regulate Nuclear Gene Expression, Science, Volume 316, pp.715-719

Krause K., Kilbienski I., Mulisch M., Rödiger A., Schäfer A., and Krupinska K. (2005) DNA-binding proteins of the Whirly family in Arabidopsis thaliana are targeted to the organelles, FEBS Letter, Volume 579(17), pp. 3707-3712

Lee R.C., Feinbaum R.L., and Ambros V. (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*, Cell, Volume 75, pp.843–854

Lewis B.P., ***Burge, C.B., and Bartel, D.P.*** (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*: *15*-20.

Llave, C. (2002). "Cleavage of Scarecrow-like mRNA Targets Directed by a Class of Arabidopsis miRNA." Science 297(5589): 2053-2056.

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

Maréchal A., Parent J. S., Sabar M., Veronneau-Lafortune F., Abou-Rached C., Brisson N. (2008) Overexpression of mtDNA-associated atwhy2 compromises mitochondrial function. BMC Plant Biology, Volume 8, *pp.* 42

Matzke M., Aufsatz W., Kanno T., Daxinger L., Papp I., Mette M. F., and Matzke A. J. (2004) Genetic analysis of RNA-mediated transcriptional gene silencing, Biochimica et Biophysica Acta, Volume 1677(1-3), pp.129-141

Palatnik J. F., Allen E., Wu X., Schommer C., Schwab R., Carrington J. C., Weigel D.

(2003) Control of leaf morphogenesis by microRNAs, Nature, Volume 425, pp. 257-263
Prikryl J., Watkins K. P., Friso G., van Wijk K. J., and Barkan A. (2008) A member of the Whirly family is a multifunctional RNA- and DNA-binding protein that is essential for chloroplast biogenesis. Nucleic Acids Research, Volume 36, pp. 5152–5165

Rajagopalan, H., P. V. Jallepalli, et al. (2004). "Inactivation of hCDC4 can cause chromosomal instability." Nature 428(6978): 77-81.

Rajagopalan R., Hervé V., Jerry T., and David P. Bartel. (2006) A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. Genes Development, Volume 20, pp.3407-3425

Reinhart B.J, Weinstein E.G, Rhoades M.W, Bartel B., Bartel D.P. (2002) MicroRNAs in plants, Genes Development, Volume 16, pp.1616–1626

Reinhart B. J., Slack F. J., Basson M., Pasquinelli A. E., Bettinger J. C., Rougvie, A. E., Horvitz H. R., and Ruvkun G.(2000) The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*, Nature , Volume 403, pp. 901–906

Sawai, S., P. A. Thomason, et al. (2005). "An autoregulatory circuit for long-range self-organization in Dictyostelium cell populations." Nature 433(7023): 323-326.

Schmitz-Linneweber C, Small I. (2008). Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends in Plant Science 13: 663–670.

Schwab R., Palatnik J. F., Riester M., Schommer C., Schmid M,. Weigel D. (2005) Specific effects of microRNAs on the plant transcriptome, Devlopmental Cell, Volume 8(4), pp. 517-527

Small I. D., Peeters N. (2005) The PPR motif - a TPR-related motif prevalent in plant organellar proteins, Trends Biochemical Science, Volume 25(2), pp. 46-47

Song J.J., J. Liu, N.H. Tolia, J. Schneiderman, S.K. Smith, R.A. Martienssen G.J.Hannon, and L. Joshua-Tor. (2003) The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes, Nature Structure Biology, Volume 10, pp. 1026-1032

Voinnet O. (*2009)* Origin, biogenesis, and activity of plant microRNAs. Cell. **Volume** 116, pp. 281-297

Wang X. J., Reyes J. L., Chua N.H., and Gaasterland T. (2004) Prediction and identification of *Arabidopsis thaliana* microRNA genes and their mRNA targets. Genome Biology Volume 5, pp.65

Wightman B., Ha I., and Ruvkun G. (1993) Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*, Cell，Volume 75, pp.855–862

Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D.,Jacobsen, S.E., and Carrington, J.C. (2004). Genetic and functional diversification of smallRNA pathways in plants. PLoS Biol *2*, E104.

Xiong J. Y., Lai C. X, Qu Z., Yang X. Y., Qin X. H., and Liu G. Q. (2009) Recruitment of AtWHY1 and AtWHY3 by a distal element upstream of the kinesin gene AtKP1 to mediate transcriptional repression, Plant Molecular Biology, Volume 71, pp. 437-449

Yamamoto, Y.Y., Sieburth, L., and Voinnet, O. (2008) Widespread translational inhibition by plant miRNAs and siRNAs. Science, Volume 320, pp. 1185-1190.

Yu B., Z. Yang, J. Li, S. Minakhina, M. Yang, R.W. Padgett, R. Steward, and X. Chen. (2005) Methylation as a crucial step in plant microRNA biogenesis, Science, Volume 307, pp. 932-935

Appendix

1. Arabidopsis microRNA families and their target genes.

miR156/157	Squamosa-promoter binding protein-like (SPL)
miR158	Pentatricopeptide repeat (PPR)
miR159/319	MYB transcription factor, TCP transcription factor
miR160	Auxin response factor (ARF)
miR161	Pentatricopeptide repeat (PPR)
miR162	Dicer-like (DCL1)
miR163	S-adenosylmethionine-dependent methyltransferase (SAMT)
miR164	NAC domain transcription factor
miR165/166	HD-ZIPIII transcription factor
miR167	Auxin response factor (ARF)
miR168	Argonaute (AGO1)
miR169	HAP2 transcription factor
miR170/171	Scarecrow-like transcription factor (SCL)
miR172	Apetala2-like transcription factor (AP2)
miR173	TAS1, TAS2
miR390/391	TAS3
miR393	Auxin receptors (TIR1, AFBs), bHLH transcription factor
miR394	F-box
miR395	ATP-sulfurylase (APS), Sulfate transporter (AST)
miR396	Growth regulating factor (GRF)
miR397	Laccase (LAC) 6, 7, 11, 14, 16
miR398	Copper superoxide dismutase (CSD) , Cytochrome-c oxidase
miR399	Phosphate transporter, E2 ubiquiting-conjugating protein
miR400	Pentatricopeptide repeat (PPR)
miR401	Unknown

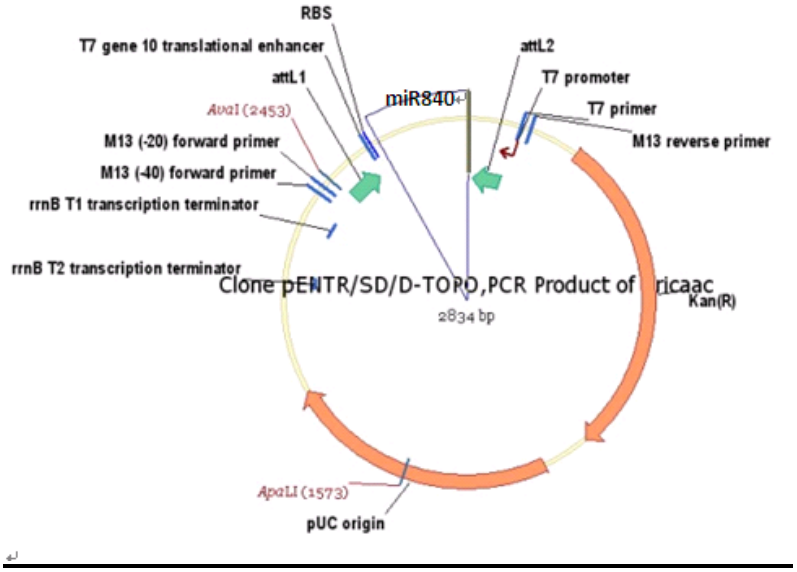
miR402	ROS1-like, putative DNA glycosylase
miR403	Argonaute (AGO2)
miR404	LRR-TM protein kinase
miR405	Unknown
miR406	Spliceosomal proteins
miR407	Short-chain dehydrogenase/reductase
miR408	Laccase (LAC), Plantacyanin-like (PCL)
miR413	Splicing factor, MYB transcription factor, PPR
miR414	DEAD box RNA helicase (DRH1), F-box, nucleosome assembly protein
miR415	Cellulose synthase family, PPR
miR416	F-box
miR417	RNA-directed RNA polymerase, auxin response transcription factor
miR418	Homeobox protein
miR419	ABC transporter family protein, No apical meristem (NAM), histidine kinase
miR420	None predicted
miR426	None predicted
miR447	2-phosphoglycerate kinase-related (2-PGK)
miR472	CC-NBS-LRR
miR771	None predicted
miR773	DNA (cytosine-5-)-methyltransfearse
miR774	F-box
miR775	Galactosyltransferase Avr9 elicitor
miR776	Serine/threonine kinase
miR777	None predicted
miR778	SET-domain
miR779	None predicted
miR780	Cation/hydrogen exchanger

miR781	CD2-binding, MCM
miR782	Pseudogene
miR783	Extra-large G-protein-related
miR822	DC1 domain
miR823	Chromomethylase, CMT3
miR824	AGL16 MADS-box
miR825	Remorin, zinc finger homeobox family, frataxin-related
miR826	AOP2
miR827	SPX domain/C3CH4-type RING zinc finger
miR828	MYB transcription factor
miR829	AP2 domain ethylene response factor
miR830	RanBP1 domain, kinesin motor-related
miR831	None predicted
miR832	Unknown
miR833	F-box
miR834	COP1-interacting protein
miR835	MYB trabscription factor
miR836	None predicted
miR837	GIF transcription factor
miR838	Armadillo/ -catenin
miR839	None predicted
miR840	**WHIRLY3**
miR841	Histone H2A.F/Z
miR842	Jacalin lectin
miR843	F-box
miR844	Kinase
miR845	None predicted
miR846	Jacalin lectin
miR847	Cyclophilin-RNA interacting protein
miR848	None predicted

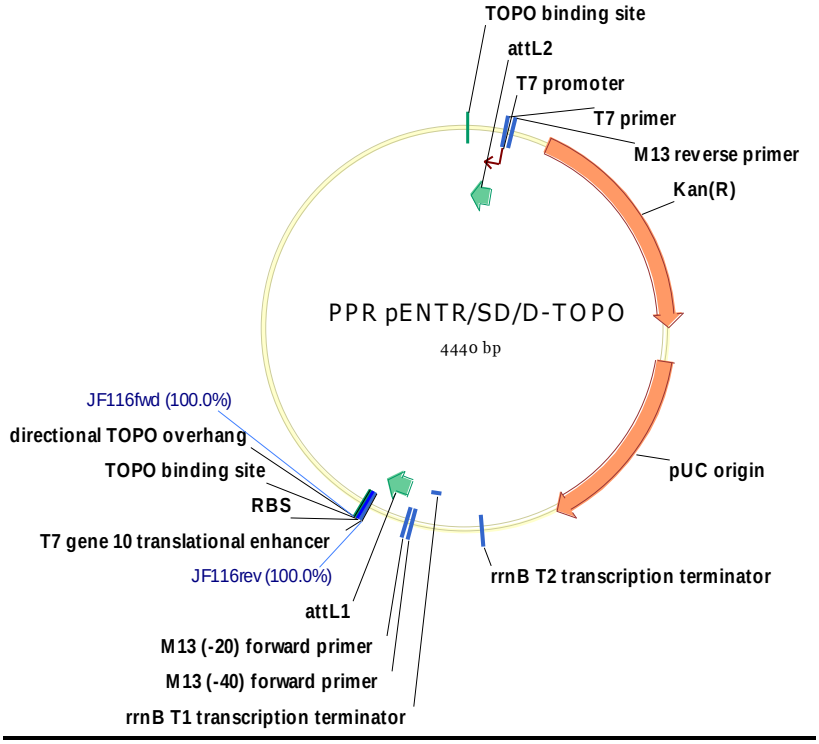
miR849	None predicted
miR850	None predicted
miR851	None predicted
miR852	ATPase
miR853	None predicted
miR854	UBP1b 3'UTR
miR855	UBP1b 3'UTR
miR856	Cation/hydrogen exchanger, Zinc transporter
miR857	Laccase
miR858	MYB transcription factor
miR859	F-box
miR860	Histone deacetylase, ferrochelatase, RNA recognition motif
miR861	None predicted
miR862	None predicted
miR863	None predicted
miR864	Triacylglycerol lipase
miR865	Serine carboxypeptidase, sulfate transporter
miR866	Expressed protein, C2-domain containing protein 4
miR867	PHD finger-related/SET domain, phospholipase/carboxylesterase
miR868	None predicted
miR869	None predicted
miR870	None predicted

Family number according to miRBase Release 10.0

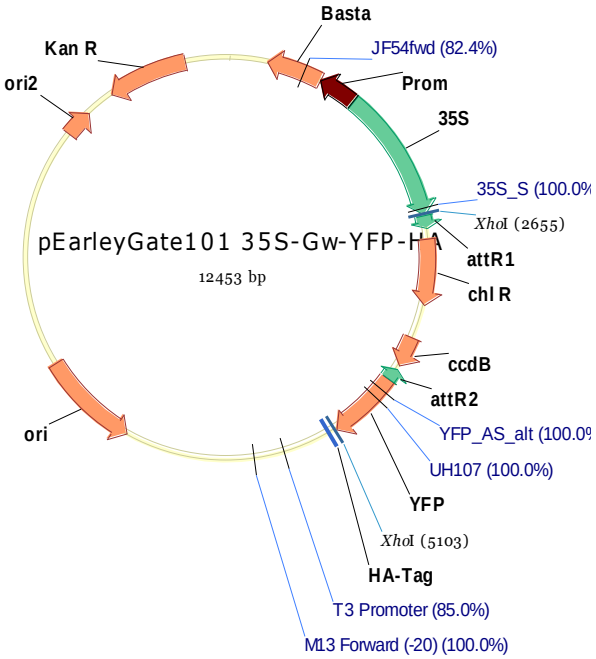
2 Vector maps (constructed by Janina Fuss and Stephanie Hainbuch)



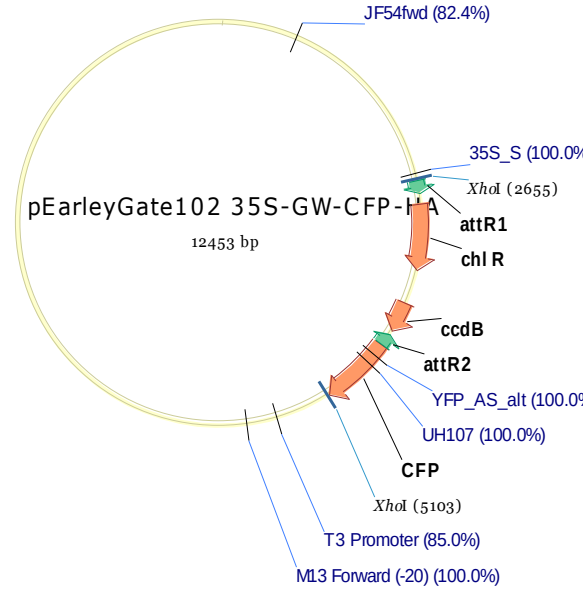
2.1 Primi entry vector: Primi_ pENTR



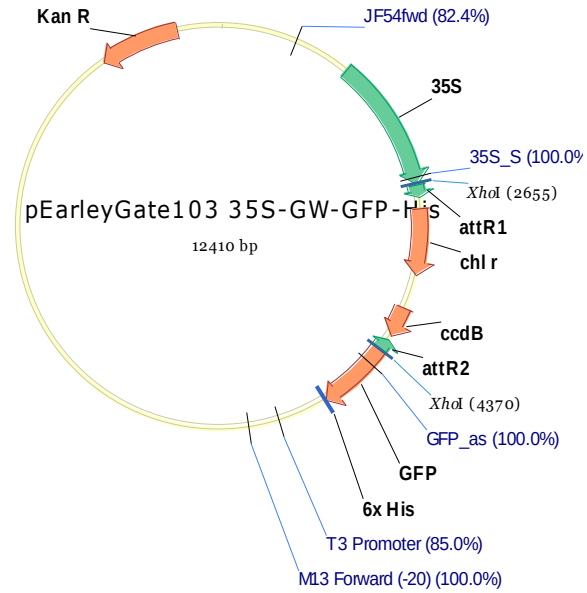
2.2 PPR entry vector: PPR_ pENTR



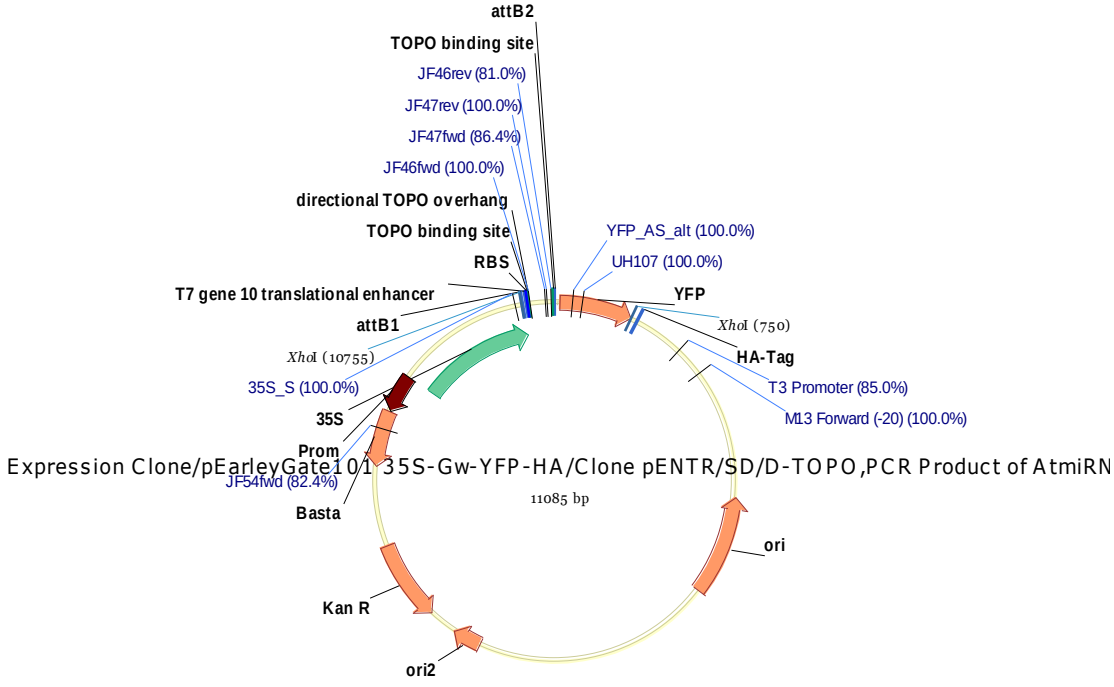
2.3 Destination vector: pEG101



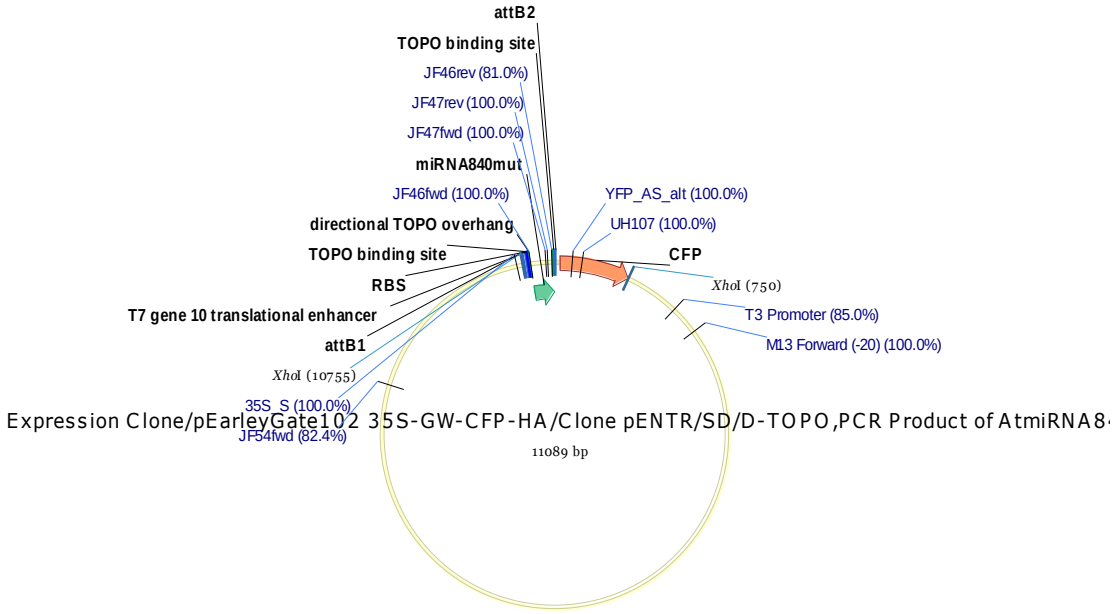
2.4 Destination vector: pEG102



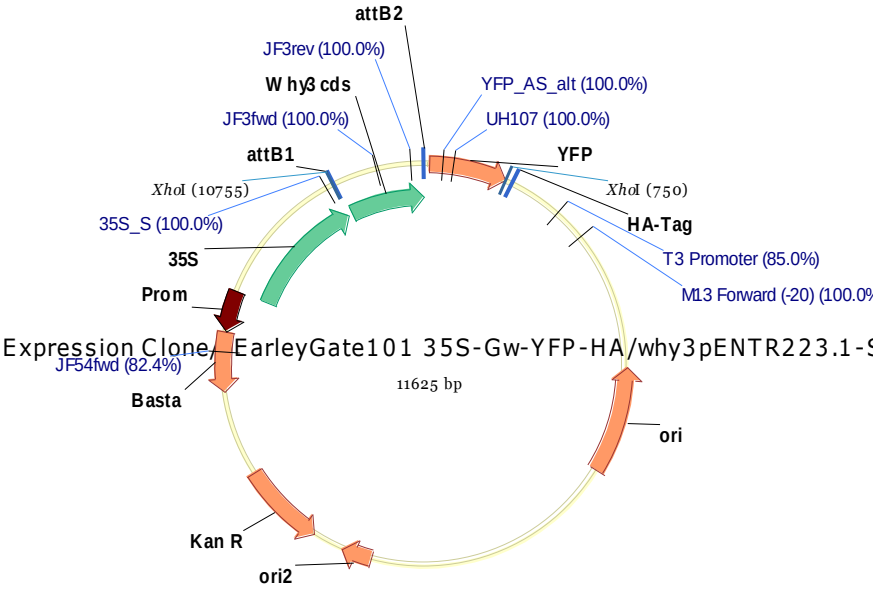
2.5 Destination vector: pEG103



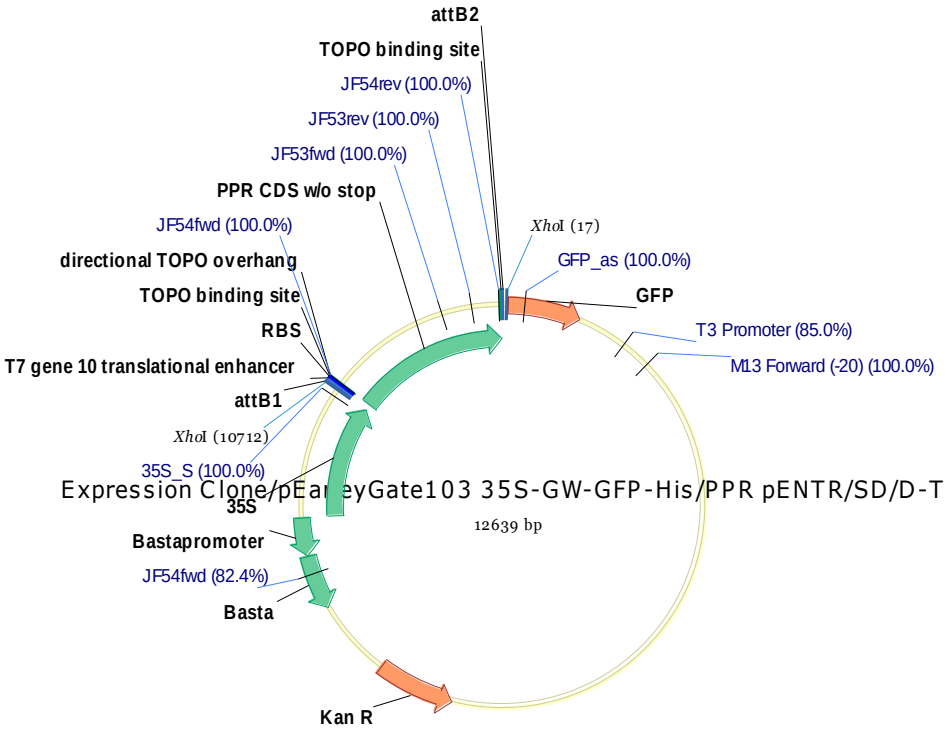
2.6 Expression vector: Primi Expression pEG101



2.7 Expression vector: Primimut Expression pEG102



2.8 Expression vector: Why 3 Expression pEG101



2.9 Expression vector: PPR Expression pEG103