International PhD Program in Biomolecular Sciences
Centre for Integrative Biology
XXVIII Cycle

"Identification and functional characterization of microRNA regulatory elements in *Brassicaceae*"

Tutor

Dr. Claudio Varotto

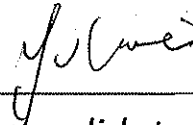*Fondazione Edmund Mach*

Ph.D. Thesis of

Yu Wei

University of Trento

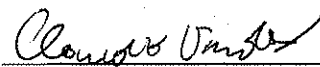*Fondazione Edmund Mach di San Michele all'Adige*

Academic Year 2014-2015

Declaration of Authorship,

I, the undersigned Yu Wei, confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

PhD candidate

Tutor

Table of Contents

Abstract

As a class of short non-coding RNAs, microRNAs can regulate gene expression by a post-transcriptional pathway through repressing or degrading mRNA. In the evolutionary history, many plant microRNAs are highly conserved from green alge to land plants. In recent years, dramatic studies demonstrate that microRNAs play a crucial in plant growth and development, in response to environmental stresses. Some microRNAs can response to plant hormones, while some others are tissue or cell specific. The understanding of how these microRNAs are regulated at the transcriptional level is just initiated.

With the aim to understand the regulatory mechanism of plant microRNA in evolutionary tem, and identify the most relevant *cis*-regulatory elements in some microRNAs for improving the agriculture in the future, this study was carried out.

microRNA390 is one of the many conserved microRNAs, it can indirectly regulate the ARFs expression level by targeting TAS3, and consequently regulate lateral organ and later root development in plants. In order to understand the regulatory mechanism of microRNAs in an evolutionary term, microRNA390a and microRNA390b in *Arabidopsis* were chosen and studied.

In 16 phylogenetically related species within *Brassicaceae*, we analyzed the microRNA promoter sequences and identified overall conserved c*REs* in microRNA390 promoter regions, and accompanied with functional characterization, we obtained a good view of microRNA390 regulatory network.

Based on 454 sequencing technique, took the microRNA sequences of sequenced Arabidopsis as reference, by assembling and aligning the microRNA promoter sequences, calculated the PWM and predicted the putative motifs with both

MEME program and PlantCARE database, subsequently compared the motif similarities by TOMTOM program, we eventually obtained the putative ones met the required *E*-value. In the meantime, we reconstructed the phylogenetic trees of both paralogs by MEGA7 program. We identified 6 and 5 overall conserved c*REs*. Subsequently, we experimentally validated the putative c*REs* by *Arabidopsis* transformation and site-specific mutagenesis.

The results we have obtained were as follows:

(1) There were totally 29 microRNA loci in 9 families identified to be highly conserved, and totally 104 putative motifs were predicted in their promoter regions.

(2) The reconstructed phylogenetic trees based on miRNA390a and miRNA390b promoter sequences respectively were compared with the the phylogenetic relationships (species trees) in known *Brassicaceae* phylogeny. The data derived from both promoter sequences were inconsistent with *Brassicaceae* phylogeny. This implied that there might be multiple copies of specific c*REs* in some specific species, hence the promoter sequences evolution of microRNA is not reflective of species phylogeny.

(3) Took *Arabidopsis thaliana* as model plant, we successfully constructed GUS-fused promoters of miRNA390a and miRNA390b. The GUS histochemical assay indicated that the two paralogs expressed in different tissues in transgenic *Arabidopsis*. miRNA390a expressed in lateral root primordia, true leaves, cotyledons, as well as in the floral organs, yet it was absent from lateral root tip and shoot apical meristem; whereas miRNA390b specifically expressed on lateral root tips, and a more restricted expression pattern was detected on aerial part of true leaves and

floral organ. These differences indicated possible sub-functionalization with respect to their ancestral miR390 during the evolutionary process.

(4) Based on the six putative c*REs* identified in miRNA390a and the reliable WT constructs, we also constructed six GUS-fused promoters that undergone site-specific mutagenesis. The GUS assay demonstrated that the activity of putative *cis*-elements varied with distance to TSS. Mutations of proximal sites (*m2* and *m3*) enhanced expression thereby M2 and M3 were likely to be silencers; while mutations of distal elements (*m5* and *m6*) tended to decrease the promoter expression, hence M5 and M6 probably work as enhancers. These evidences suggest there was a specific modular cooperativity of miR390a c*REs* in regulating gene expression and mediating plant development.

(5) Furthermore, we treated the 7d-old transgenic seedlings with iron-deficiency, both the GUS assay and qRT-PCR data conferred the iron responsiveness of putative iron-deficiency related E-box M3 CCAGATGTGA and the iron-deficiency responsive cis-element 1 M6 GAAATGAAGGAAGCTTAAT.

LIST OF Abbreviations

3'-UTR	3'-untranslated regions
A	adenosine
AdoMet	cofactor S-adenosyl-methionine
AGO	Argonaute
AP1	APETALA1
ARF	auxin response factors
As	Arsenite
AtHsp90	Arabidopsis heat shock protein 90
bHLH	basic helix-loop-helix transcription factors
bZIP	Basic Leucine Zipper
Cd	Cadmium
CBC	nuclear cap-binding complex
cREs	cis-regulatory elements
Cu	Copper
COPT2	copper transporter2
C	cytosine
D-bodies	dicing bodies
DDL	RNA-binding protein DAWDLE
DNA	Deoxyribonucleic acid
DRB4	DOUBLE-STRANDED RNA-BINDING PROTEIN
dsRBD	dsRNA binding domain
FIT1	Fe-deficiency Induced Transcription Factor 1

FUS3	FUSCA3
HEN1	HUA ENHANCER1
HESO1	HEN1 suppressor1
HST	HASTY protein
HYL	HYPONASTIC LEAVES1
IDE	Iron-deficiency responsive cis-element 1
IDF	iron-deficient condition
ISF	iron-sufficient condition
MID (Chapter 1.2)	middle
MID (Chapter 3.1.1.2)	multiplex index
miRNAs	microRNAs
MREs	Metal response elements
MYB	myeloblastosis
ORF	open reading frame sequences
PAZ	Piwi/Argonaute/Zwille
PCR	Polymerase chain reaction
phyA	phytochrome A
pri-miRNA	primary miRNA precursor
qRT-PCR	quantitative real-time PCR
RDR6	RNA-DEPENDENT RNA POLYMERASE6
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNase III	catalytic ribonuclease III
SAM	shoot apical meristem
SDN	small RNA-degrading nuclease

SE	C2H2-zinc finger protein SERRATE
SFF	Standard Flowgram Format
SGS3	Sequences Over-Represented in Light-Induced Promoters 5
siRNAs	SUPRESSORS OF GENE SILENCING
SORLIP5	small interfering RNAs
sRNAs	small RNAs
stRNAs	small temporal RNAs
TAS	TELOMERE-box
*tasi***RNAs**	TRANSPORT-INHIBITOR-RESISTANT1
TELO-box	trans-acting short-interfering RNAs precursor
TIR1	trans-acting short-interfering RNAs
TSS	translation start site
U	uridine
WT	wild type
Zn	Zinc

LIST OF FIGURES

LIST OF TABLES

Introduction

The group of long non-coding miRNAs are fundamental, sequence-specific regulatory elements of eukaryotic genomes. Many plant miRNAs are evolutionarily conserved, and they are known to play essential roles in the regulation of various fundamental processes in plant growth and development, in responding to both biotic and abiotic stresses, as well as in the regulation of the silencing pathway themselves. Generally, promoter sequence evolution is reflective of species phylogeny, and the level of evolutionary conservation of promoter regions could be further attributed to their *cis*-regulatory motifs [1]. Regardless of the relevance, the transcriptional machinery study of plant microRNAs is still at the early stage. Comparative approaches to detect regulatory elements in the promoter regions provided a pathway to understand the regulatory mechanism of plant microRNA [2]. Some computational performances on animals for relevant tasks have achieved promising results. However, due to the phylogenetically distant plant species used, the evolving computational studies in predicting and identifying *cis*-regulatory elements in plant microRNAs promoter regions are just initiated.

microRNA390 is one class of the highly conserved microRNA groups, it is mainly expressed in the lateral root primordia [3], and play important role in mediating the processes in response to several stresses [4], including heavy metal stresses [5].

In an attempt to elucidate the regulatory network of plant microRNAs in *Brassicaceae* and eventually harness it for crop improvement, we aimed to characterize the most relevant motifs in the miRNA390 regulatory region. We

achieved this role by carrying out functional studies with computational approaches coupled with transgenic analysis.

Taking the fully sequenced genome of *A. thaliana* as the reference, by data mining of multiple promoter sequences of multiple microRNA genes, we obtained 29 microRNA genes that presented in more than 10 species within *Brassicaceae*. This similarities of the promoter sequences were limited to the first few hundred nucleotides upstream of the translation start codon, indicating that this region is important for exploring the conservation and the regulatory mechanism of the microRNAs.

To characterize the function of these c*REs*, Pentry/TOPO cloning and site-specific mutagenesis techniques were performed on the putative motif specific sequences. Subsequentially, iron-stress treatment was applied to different mutagenized constructs, and the transgenic plants were assayed by GUS histochemical detection to provide a qualitative view of mutation effects.

We could provisionally assign the roles to several miRNA390a *cis*-regulatory elements (c*REs*) by computational prediction and functional validation through plant transformation. In this study, a putative helix-loop helix protein FIT1 binding motif 5'-CANNTG-3' sequence was found among five conserved c*REs* identified in miRNA390a promoter region. Moreover, another relatively less conserved *cis*-element 5'-GAAATGAAGGAAGCTTAAT-3' was identified in our study, which was known to be involved in iron-deficiency stress in *Arabidopsis* [4].

We expect to extend the methodology to other groups from roside family. In addition, the numerous computational predicted c*REs* of conserved microRNA genes, could possibly serve as a good starting point for future functional characterization of a series of c*REs*, and explore more complete microRNA

regulatory networks, eventually contribute to the crop improvement in different plant families.

This thesis consists of five chapters as follows:

Chapter one provides an overview of current researches on microRNAs and the sequencing technology. It presents the general context of this research and serves as an introduction to the rest of the thesis.

Chapter two describes the design and objectives of this project.

Chapter three presents the computational data mining of conserved microRNAs and the putative *cis*-regulatory elements identified in single or multiple microRNA genes at the *Brassicaceae* family level. The c*REs* identification of miRNA390a and miRNA390b is specifically given.

Chapter four describes the promoter constructions of the two paralogs miRNA390a and miRNA390b in *Arabidopsis*. Six promoter site-mutagenized constructions of miRNA390a putative motifs in WT *Arabidopsis* are also presented.

Chapter five discusses the results of functional exploration (iron-deficient responsiveness) of two iron-responsive c*REs* by GUS assay based on the site-directed constructs of miRNA390a.

Chapter six gives the conclusion, the discussion of advantages and disadvantages, as well as the perspective.

Acknowledgement

Chapter 1.   Current review of miRNA

1.1   miRNAs definition

Small RNAs (sRNAs), including microRNAs (miRNAs) and short interfering RNAs (siRNAs), are fundamental regulatory elements in eukaryotes. Chemically similar, both miRNAs and siRNAs incorporate into the RNA-induced silencing complex (RISC), and inhibit gene expression through sequence-specific interactions with RNA or DNA.

Despite similarities, miRNAs and siRNAs can be distinguished by their precursors and their targets. MiRNAs derive from a short, imperfectly paired stem of much larger fold-back structures and regulate the target of RNA transcripts to which they are partially or fully complementary. siRNAs arise by Dicer-like cleavage from long, perfectly paired double-stranded RNAs, and typically guide mRNA target degradation to which they are completely complementary, including the mRNAs from which they generate. SiRNAs also regulate nuclear events, such as DNA and histone methylation, leading to transcriptional silencing. Generally, the activity guided by a miRNA or siRNA, mainly depends on how precisely si/miRNA anneal to their targets [6, 7].

In particular, miRNAs are a set of small noncoding RNAs, 19-25 nucleotides in length, that regulate gene expression at the post-transcriptional level [8].

The discoveries of new plant miRNAs have been rapidly accumulating in recent years[9]. The miRNAs predicted by computational approaches need to be experimentally verified, thus the expression criteria for miRNAs identification are important.

Some microRNAs, such as lin-4 and let-7 were found by standard positional cloning of genetic loci, but most microRNAs are discovered by cDNA cloning of sequences from size-fractionated RNA samples [6, 7].

If discovered by cDNA cloning, the small RNAs should meet several established identification criteria to be identified as miRNAs, and distinguished from siRNAs [10, 11]:

First, ~22-nt small RNA transcript should be confirmed by blot hybridization to a size-fractionated RNA sample. Quantitative real-time PCR (qRT-PCR), primer extension analysis, ribonuclease protection assay and microarray, as well as *in situ* hybridization are available [12]. Since northern blotting can demonstrate both the mature part (a ~22-nt nucleotide band)and fold-back precursor(in animals they are 60-90 nt, whereas in plants, they are more variable, 70-400 nt) [13, 14], it is one method to confirm the miRNAs [11, 12]. Second, the small RNA sequence should be located in one arm of the hairpin structure, without large internal loops or bulges. Third, both the ~22-nt mature miRNA and its hairpin precursor should be phylogenetically conserved. Fourth, the precursor accumulation accompanied by reduced Dicer function strengthens the evidence.

However, the fulfillment of any single criterion listed above is insufficient for a candidate gene to be confirmed as a novel miRNA. In fact, the first criterion does not exclude small RNAs, and the conservation of fold-back structure (the third criterion) is not a unique characteristic of miRNA biogenesis, nor is the Dicer function (the last criterion). Typically, a combination of the first and the second criteria or the first and the third is considered as an adequate condition for miRNA identification.

If the small RNA is not found by cDNA cloning, expression of the gene must be judged together with the structure of the precursor and conservation. If a putative

small RNA cannot be experimentally verified, the conserved hairpin precursor should accumulate accompanied by the reduced Dicer function. This situation should be treated with special care, as other hairpin-containing RNAs might be processed by Dicer. Small RNAs that do not meet these criteria can be classified as either small interfering RNAs (siRNAs) or as other classes [15].

The miRNA database miRBase (v21, released in June 2014) contains a total of 28645 entries of hairpin precursors which generate 358828 mature miRNA products, including both plant and animal species. However, since several facts indicate that some registered miRNA genes in the public databases such as miRBase lack the established annotation criteria [9, 16], stricter criteria should be applied when the lists in the public databases expand [12, 17, 18].

1.2   Biogenesis of miRNA

microRNAs(miRNAs) were first discovered in the nematode *Caenorhabditis elegans* in 1993 and were considered as small temporal RNAs (stRNAs) [19]. By 2001, miRNAs have been formally named and recognized as a distinct class of RNAs. In 2002, plant miRNAs were identified for the first time in *Arabidopsis* [13].

Most of the plant miRNAs are transcribed from independent, non-protein-coding loci between 2 intragenic regions of genome by RNA polymerase II [20]. Considering the formation of miRNAs, two possible mechanisms are highlighted. The first hypothesis suggests that most plant miRNA genes are formed through inverted duplications of protein-coding genes, which is correlated with the dosage effect [21]. According to the second hypothesis, miRNAs generate from spontaneous transcription of partially self-complementary or inverted sequences, and most of them are present randomly in plant genomes [22, 23].

As Figure 1-1 shows, the primary miRNA (pri-miRNA) duplex generated by RNA polymerase II is then stabilized by RNA-binding protein DAWDLE (DDL) for

its conversion in nuclear dicing bodies (D-bodies), where the Dicer-like 1 (DCL1) is interacted with HYL1 (HYPONASTIC LEAVES1), nuclear cap-binding complex (CBC), and C2H2-zinc finger protein SERRATE (SE), and they act together to stabilize pri-miRNA into stem loop precursor miRNA (pre-miRNA) [24, 25].

The DCL family contains four paralogs, but only DCL1 is necessary for the maturation of most miRNA [26], while other members are involved in defending viruses [9, 27, 28]. Different DCL family members give rise to different length of small RNAs: DCL1 produces mainly 21nt, DCL2, DCL3 and DCL4 generate predominantly siRNA size classes, 22, 24 and 21 nt, respectively. RNA virus infection is mainly affected by DCL4 and DCL2, where DCL4 associates with (RNA-DEPENDENT RNA POLYMERASE6) RDR6 and (HUA ENHANCER1) HEN1 is involved in *tasi*RNAs metabolism and acts during post-transcriptional silencing [27, 29].The pri-miRNA catalyzed protein DCL1 contains an RNA helicase and two ribonuclease III (RNase III)-like domains, a central (Piwi/Argonaute/Zwille) PAZ domain and a dsRNA binding domain (dsRBD). The PAZ domain, separated by the two catalytic RNase III domains is a module that binds the end of dsRNA. Furthermore, the distance between the 3' overhang binding pocket of the PAZ domain and RNase III active site is thought to be a molecular ruler that determines the length of the small RNAs [30].

3'-terminal nucleotide of miRNA/miRNA* is methylated by HEN1 protein. A methyl group of HEN1 is transferred from the cofactor S-adenosyl-methionine (AdoMet) onto the 2-hydroxyl of the 3-terminal nucleotide of the miRNA/miRNA* duplex to stabilize it. This methylation step can protect the 3' terminal of unwound mature miRNA from being degraded by exonucleases, such as small RNA-degrading nuclease (SDN) proteins and HEN1 suppressor1 (HESO1), the former can

specifically degrade single-stranded miRNAs. 3' A is supposed to stabilize miRNAs, while 3' C incorporation might induce miRNA degradation.

The methylation pattern of HEN1 could determine which strand of miRNA/miRNA* duplex will be chosen during RISC maturation [31]. However, miRNAs in (HASTY protein) *hst* mutants didn't accumulate inside the nucleus in *Arabidopsis*, indicating the existence of other export mechanisms yet-to-be-identified [32].

Then the duplex is exported from the nucleus to cytoplasm by protein HASTY, a homolog of metazoan exportin 5, which prefers overhung nucleotides [8, 22, 32, 33]. Once inside the cytoplasm, the duplex separates, and the strand with a less stably paired 5' end is selected as the guide strand [34-36]. Following this, by binding to the Argonaute (AGO) protein, the guide miRNA strand locates into the RNA-induced silencing complex (RISC) to carry out the silencing reactions. The RISC complex cleaves the 3'-untranslated regions (3'-UTR) of target mRNAs catalyzed by AGO1 [37-40].

The passenger strand of the duplex, called miRNA*, is usually degraded [16, 17]. A recent study indicated that beyond degrade being the only way, miRNA* was necessary for the maturation of AGO-RISC, which suggests a yet to be identified role of the miRNA* [41].

The selection of the guide strand depends on the thermodynamic stabilities of the two strands: the strand less stable at the 5'-end will be selected as the guide strand. Since the loading on RISC of the guide strand is performed by binding to the AGO proteins, the selection is partially determined by the AGO proteins.

The AGO family was named after the discovery of AGO1 in *A. thaliana*, the mutation of which had pleiotropic effects on leaves resembling small squids

(*Argonautus*) [42]. There are 10 AGOs in *A. thaliana* and they belong to three phylogenetic clades [43, 44]. Each AGO protein contains four domains: amino-terminal (N), conserved PAZ, MID (middle), and PIWI [45]. The role of the N domain remains unclear. The PAZ domain binds to the 3'-end nucleotide of miRNA and the ribose is modified by 2'-hydroxymethylation [46]. The PIWI domain specifically catalyzes the cleavage of phosphodiester bond within the "seed" region (tenth and eleventh nucleotide) of target mRNA [47]. The MID domain binds the phosphate group on the 5'-terminal of the miRNA, and the MID domains of AGO paralogs from *Arabidopsis* favorite different 5'-end nucleotides, which explains why different AGO members prefer different miRNAs [48, 49]: AGO1 binds to 5' U (uridine) that most plant miRNAs carry, and is a critical effector for miRNAs and ta-siRNAs [49-51]; both AGO2 and AGO4 favor A (adenosine). AGO2 is involved in antiviral defense and is responsive to a broad spectrum of plant viruses [52-54]; AGO4 was identified to be involved in transcriptional gene silencing (TGS) [55]; AGO5 prefers C (cytosine) [49], and is speculated to mediate a somatic sRNA pathway [56]; AGO7 interacts with 5'adenosine site, and specifically binds miR390 [57-59]; AGO10 has a bias towards 5' U and preferentially binds miR165/6 [60, 61].

Although the pri-miRNAs were assumed not to encode any proteins, Lauressergues et *al* presented convincing evidence to the contrary, revealing that some pri-miRNAs contain short open reading frame sequences (ORF) which could encode proteins and generate peptides [62]. The peptides can enhance the activity of pri-miR, thus promoting the transcription of associated miRNAs, resulting in more effective downregulation of their target genes [62, 63]. This discovery uncovers as yet unknown function of pri-miR sequences and highlights another potential layer of biological importance of gene regulation [64].
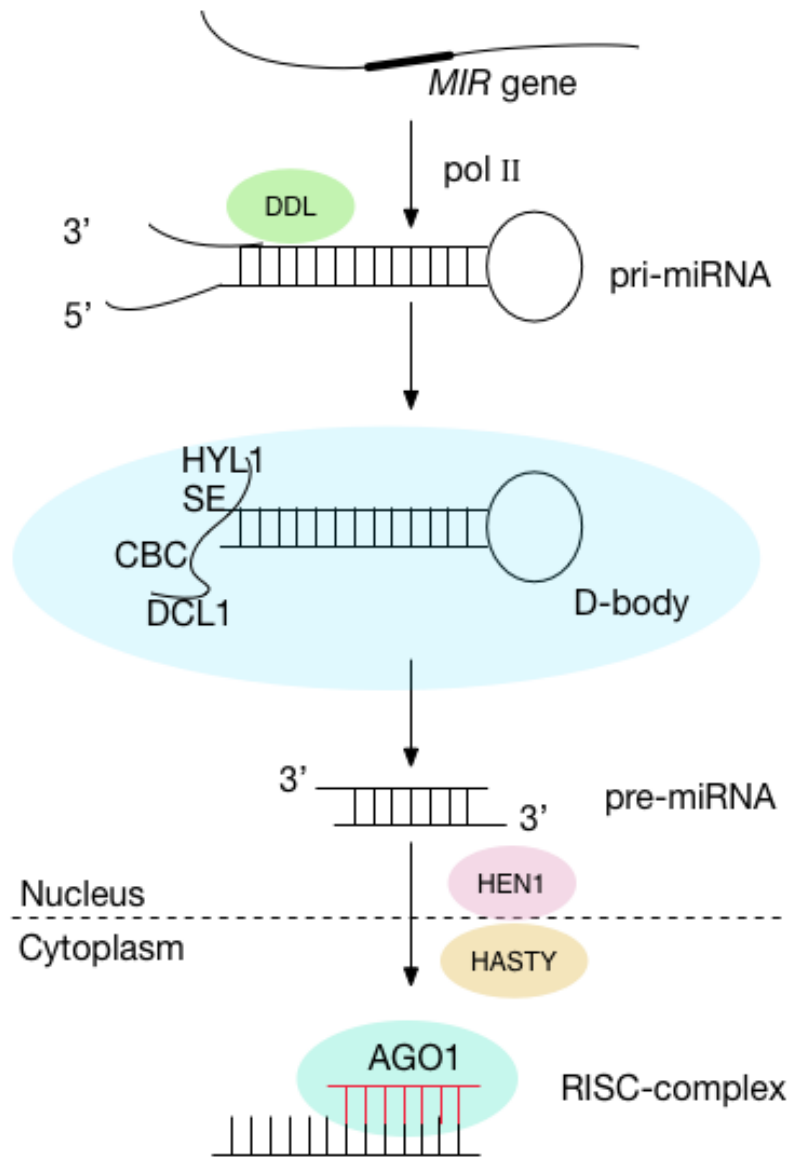
Figure 1-1 Scheme of plant microRNA biogenesis. Plant microRNA are transcribed by RNA poly II from the transcripts located in intergenic regions. pri-miRNA was stabilized by DDL, then binds with SE, HYL1 and CBC, and processed by DCL1 into pre-miRNA. The pre-miRNA duplex is methylated by HEN1 and transported from nucleus to cytoplasm by HASTY. The guide miRNA strand binds AGO1 and is loaded into the RISC complex to carry out the silencing process.

1.3 Plant miRNA and animal miRNA

Both plant and animal miRNAs are involved in regulating the expression of many different genes and play crucial roles in various/diverse biological processes. The biogenesis and functions of miRNAs in plants and animals are similar [58]. However, they may have originated independently due to several differences:

(1) The majority of plant miRNAs are produced from single primary transcript located in the intergenic regions, and the ones derived from the same miRNA gene family often share high similarities [65, 66], while some animal miRNAs are generated from introns and some are derived from polycistronic transcripts in the intergenic regions of chromosome [14, 67].

(2) The precursors of plant miRNAs are longer than those in animals, 70-400 nt comparing to 60-90 nt in animals, and generate more complicated secondary structures with fold-back [13, 14].

(3) For plants, the generation of miRNA from pri-miRNA to pre-miRNA then to the miRNA, involves only the Dicer protein [13, 68, 69]. For animals, the two processes involve both the Drosha and the Dicer [13, 69-71].

(4) Most of the plant miRNAs repress the targets by cleaving mRNA at a single site in the coding regions, while animal miRNAs contain multiple binding sites in the 3' untranslated regions of the target gene [19, 72-76].

(5) The silencing mechanisms also differ. The majority of plant miRNAs are fully complementary to their binding sites, and guide the cleavage of the mRNAs. Most animal miRNAs are only partially complementary to the target mRNAs, and guide the translation inhibition [77, 78].

1.4 MiRNAs are involved in plant developmental regulation

The regulation of gene expression is the fundamental capability of biological phenomena [79]. Through binding to reverse complementary sequences, plant miRNAs can regulate gene expression by guiding target cleavage or translational repression [8, 80].

Various studies have revealed that a large category of plant miRNA targets appear to be genes encoding diverse families of transcription factors or regulatory proteins involved in many plant developmental processes or signal transduction, as well as adaptive responses [80-85]. It is obvious that miRNA-based gene regulation play crucial role in plant growth and development [86, 87].

1.4.1 SAM development

The shoot apical meristem (SAM) is located at the tip of the shoot and contains stem cells that are continuously renewed and generate lateral organs [88]. Some miRNAs are characterized to participate in SAM development [89, 90].

Studies in *Arabidopsis* demonstrated that AGO10, one member of the ARGONAUTE (AGO) family [91], is a key regulator of proper SAM maintenance [92, 93]. The transgenic experiments showed that AGO10 inhibits miR165/166 expression, and these two miRNAs differ in only one nucleotide in mature miRNA sequence. It is known that miR165/166 target the same class III HOMEODOMAINLEUCINE ZIPPER (HD-ZIP III) family transcription factors, which can mediate SAM maintenance and the establishment of leaf polarity [92]. Thus, the interaction between AGO10 and miR165/166 can specifically release the HD-ZIP III gene expression and maintain proper SAM development [60, 61].

Similar findings were reported in maize. As in *Arabidopsis* the phenotype of maize *rev1* (a HD-ZIP III family gene) also contains a miRNA165/166 cleavage site

20

as in *Arabidopsis* [94], indicating that maize and *Arabidopsis* possibly share a common regulatory pathway in SAM.

A microarray study on soybean found 31 miRNAs and 6 putative legume novel miRNAs expressed in the SAM. Among the miRNAs detected, *in situ* hybridization results proved that miR166 and its star strand miR166* were located respectively below and on the abaxial side of the young leaf and the peripheral region of SAM, suggesting that they play different roles in mediating leaf and SAM development. Another conserved miRNA that highly expresses in either soybean SAM or leaves is miR159, it regulates the expression of MYB33 and MYB65, and is found throughout the SAM and leaf primordia. This observation was consistent with the expression level of miR159 in *Arabidopsis* [95]. As the polarity of leaves was initiated in SAM, such miRNAs were identified in both leaf and SAM, implying that they might play essential roles in cell differentiation associated with SAM function [88].

In addition to miR165/166 and miR159 molecular pathways, miR394 and its target, a *LEAF CURLING RESPONSIVENESS (LCR)* from the F-box family, were found to be involved in maintaining stem cell competence in the SAM region [96-98], which indicated that a complex network mediated by a number of miRNAs is regulating SAM development.

1.4.2 Leaf development

Plant leaf is the primary photosynthesizing organ and plays essential role in plant growth and crop plant productivity. A variety of miRNAs are known to be implicated in leaf establishment [86, 88].

Previous studies showed that miR319 modulates plant leaf development via mediating several plant-specific TCP transcription factors [99, 100]. A study on tomato showed that ecotopic increase of miR319 can significantly alter the leaf size

and shape [101]. Similarly, overexpression of miR319 in rice and creeping bentgrass results in wider leaf blade [102, 103]. Furthermore, miR319 participated not only in the regulation of leaf development, but also in shoot and floral organs growth.

Recent studies have shown that miR396 can regulate GROWTH-REGULATING FACTORS (GRFs) genes and their upregulation in *Arabidopsis* leads to dramatically enlarged *Arabidopsis* cotyledons and leaves [104]. The overexpression of miR396 in *Arabidopsis* remarkably represses the gene expression of GRFs, thereby causing narrow-leaf phenotypes [105-108]. What's more, some bHLH transcription factors were identified to be additional targets of miR396, and their interaction could regulate leaf margin and vein pattern formation in *Arabidopsis* [109]. Therefore, miR396 could regulate leaf development via binding to both GRFs and bHLH transcription factors.

In addition, it has been shown that miR393 could help auxin-related leaf and other organs development by mediating the expression of several TIR1/AFB2 clade of auxin receptor proteins (TAARs) [110]. In rice, the overexpression of miR393 brings about altered auxin signalling which then lead to enlarged flag leaves, longer primary root and fewer crown roots [111].

More miRNAs were reported to be involved in leaf development in different plants. In rice, the expression level of miR156 gradually increases during leaf growth after the juvenile stage [112]. miR164 was shown to suppress *Arabidopsis* leaf development by binding NAC gene family members *CUC1* and *CUC2* [86, 113], and miR164 coupled with target NAC2 function in regulating leaf senescence [114]. In addition, negatively regulated by ETHYLENE INSENSITIVE2 (*EIN2*) gene, the level of miR164 decreased gradually with leaf aging in *Arabidopsis* [115].

Beyond the regulatory role in SAM development, miR394 and its target *LCR* were also found to modulate leaf curling-related morphological phenotype [98].

1.4.3 Root development

As a major plant organ, the root system plays a crucial role in nutrient and water uptake as well as in keeping the stability of the entire organism [3, 116].

In *Arabidopsis*, several miRNAs and their targets were demonstrated to participate in lateral root development. Auxin-induced miR164 and its target NAC1 were reported to be involved in lateral root initiation [117]. MiR390 triggered *TRANS ACTING siRNA 3 (TAS3)* mRNA to generate *tasi*RNAs, which then repressed ARFs expression releasing the lateral root growth [3, 118]. Similarly, miR828 cleaved *TAS4* and produces *tasi*RNAs, and one of the *tasi*RNA products bined to (myeloblastosis) MYB proteins to stimulate anthocyanin biosynthesis[119], trichome initiation and root hair patterning [120]. MiR160 was found to be involved in root cap formation through interacting with auxin response factors *ARF10*, *ARF16* [121]. MiR167 , with its two targets in ARF family genes ARF6 and ARF8, and was reported to positively regulate adventitious root formation [122].

In rice (*Oryza sativa*), overexpression of OsmiR393a and OsmiR393b lead to primary root elongation and increasing adventitious roots amount [111]. Another study indicated that two rice homologs of *Arabidopsis TRANSPORT INHIBITOR RESPONSE PROTEIN 1* (*TIR1*) which functioned in seedling root, *OsTIR1* and rice *AUXIN SIGNALING F-BOX2* (*OsAFB2*), were the targets of OsmiR393 [123].

In certain plant species, miRNAs could also cooperate with soil organisms and affects root systems. In *Medicago truncatula* for example, miR169 was reported to modulate nodule development via targeting MtHAP2-1 gene [124], and miR166 could regulate root and nodule development by guiding *HD-ZIP III* cleavage [125].

MiR482 and miR1512 were also reported to participate in nodule development and can increase nodule numbers [126].

1.4.4 Floral organs development

In the transition from vegetative to reproductive phase, floral organs of *Arabidopsis* were initiated in whorls, which was controlled by the activity of floral meristem [127].

In *Arabidopsis*, miR156 and miR172 were well-studied miRNAs that involved in flower development. MiR156 was shown to specifically to target *SQUAMOSA PROMOTER BINDING LIKEs* (*SPLs*) genes [128, 129], and upregulation of miR156 lead to delayed floral transition. As to miR172, its overexpression results in elongated flowering times in both monocotyledons and dicotyledons [130], and the *APETALA2*(*AP2*) family gene presented throughout the floral meristem is reported to be the target of miR172 [131, 132].

MiR167 was indicated to target two *Arabidopsis* auxin response factors ARF6 and ARF8, and affecting in natural flower development [133]. MiR319 could regulate the petal and stamen growth, via modulating the expression level of its target TCP [134].

1.5 The regulation of miRNA390 in plants

miRNA390 was proved to be involved in lateral root development through a miRNA390a-TAS3-ARF2/ARF3/ARF4 regulatory pathway [3, 118]. In particular, as the schematic in Figure 1–2 showed, associated with AGO7, miRNA390 triggered the cleavage of TAS3 precursor RNA, the cleavage product was then processed by RNA-DEPENDENT RNA POLYMERASE6 (RDR6) and SUPRESSORS OF GENE SILENCING (SGS3) into double-stranded RNA, and substantially transcribed into trans-acting short-interfering RNAs (*tasi*RNAs) by DICER-LIKE4

(DCL4) and DOUBLE-STRANDED RNA-BINDING PROTEIN (DRB4) mediated events.

Some studies demonstrated that miRNA390 associated with AGO7 was expressed in root tissues by a miRNA390-*tasi*RNAs pathway which negatively regulated the ARFs, thus releasing the growth of lateral root [3, 118]. Marin et *al* confirmed that microRNA390a is only expressed in the lateral root primordia in *Arabidopsis*, via p*MIR390a:GUS-GFP* reporter fusion construct and a miRNA390-GFP sensor that was degraded in cells which expressed miRNA390 [3]. In their supplemental file, it was shown that the p*MIR390b:GUS-GFP* reporter was only detected in the shoot apical meristem, and was absent from root tip and lateral root primordia, which indicates that in *Arabidopsis* the two paralogs function differently. The same expression pattern of miRNA390a promoter was obtained by Yoon *et al* with a p*MIR390a:GUS* reporter fusion construct [118]. However, a different GUS activity of miRNA390b promoter was present in lateral root tip. It was speculated that the main discrepancy could be generated by different developmental materials, because the researches focused on different lateral root stages [3].

Beyond the expression on root, it was also reported that miRNA390 and TAS3 *tasi*RNAs defined a circuit of affecting leaf patterning and developmental timing through regulating auxin signal by targeting ARFs [26, 135-139]. Notably, one of the targets ARF2 was suggested to be a positive regulator of leaf senescence [3, 140], which indicated that miRNA390 could promote leaf senescence by modulating the level of *tasiRNA*-ARF2 [141]. In maize, miRNA390 and *tasi*RNAs-ARFs accumulated discretely and adaxially in the initiating leaf primordia, only miRNA390a presented below the incipient leaf [142]. In contrast to maize, in *Arabidopsis*, both miRNA390 precursors were limited to the region below SAM, while mature miR390 and its product *tasi*RNAs moved from the biogenesis site

below the SAM into meristem and thus expressed broadly throughout the SAM [139, 143].

In the study of the expression patterns of miRNA390 on incipient leaf in *Arabidopsis*, Chitwood *et al* characterized *MIR390A* and *MIR390B* reporters and showed that both precursors were present beneath the SAM and older leaf primordia but not within the meristem or the youngest leaf primordia [139]. In addition, both reporters were more active on the adaxial than on the abaxial side. Furthermore, they detected identical localization of precursors with *in situ hybridization* technique. Different from the location of precursors, it was demonstrated that the expression of mature miRNA390 accumulated in the SAM and youngest leaf primordia, suggesting an extension of accumulation with respect to its precursor.

However, the results demonstrated by Chitwood were differed from the description given by Marin that GUS activity of p*MIR390b:GUS-GFP* was located in the shoot apical meristem in the supplemental material [3]. One possible reason could be on the description and the methods applied differently.
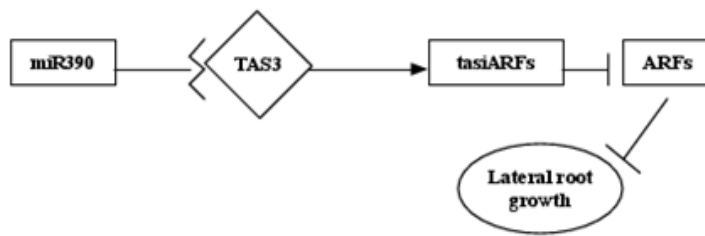
Figure 1–2 Schematic of miRNA390 function pathway in plant

1.6 Promoter structure of plant miRNAs and *cis*-regulatory elements as well as transcription factors (TFs)

In eukaryotes, transcriptional regulation is crucial for a series of fundamental processes as well as in responding to various biotic and abiotic stresses, which range from cell growth, differentiation and plant development [144].

The initiation of plant miRNA transcription is mediated by RNA polymerase II from a DNA template into long primary transcripts, then cleaved to miRNA precursors [14, 145]. Xie *et al* used 5'-RACE and de novo motif discovery tool BioProspector to identify motifs within TSS region (-50, 10) and their results further prove that RNA Pol II is responsible for miRNA transcription [146]. Moreover, this process requires the interactions mediated by a few regulatory components located in the upstream promoter region [147]. In other words, promoter region is defined as upstream genomic regulatory sequence from the first exon containing a start codon (ATG) of a gene, with a maximal length of 5kb in plants, and can be roughly divided into a proximal and a distal part [148, 149].

Class II promoters generally consist of a core promoter and upstream elements. The core promoter is the regulatory region most proximal to the TSS, and is approximately 70 bp upstream with respect to the TTS [150]. It contains at least one TATA box and one initiator centered on the TSS as well as distal specific *cis*-acting elements that are recognized by various TFs [150, 151]. The TATA box is a T/A-rich sequence recognized by TATA-binding protein (TBP), and in general 25-35 bp upstream of the TSS [152]. Upstream of the core promoter, 90% of experimentally supported binding elements were reported to be located within 800bp upstream of the TSS [148].

27

TFs are DNA-binding proteins that interact with other transcriptional components such as chromatin remodeling or modifying proteins to dominate RNA polymerases accessing to the gene promoter [153]. Some classes of TFs are shared by plants and animals, whereas there are many classes which appear to be specifically evolved in plants, including the WRKY [154], NAC [155], and AP2/EREBP [156, 157] families; the representative of plant-specific B3 domain transcription factors [156]; the trihelix DNA binding proteins[158]; the Dof domain proteins [159]; and the auxin response factors (ARFs) as well as the Aux/IAA proteins that are interacting with ARFs [160, 161].

The transcription factors recognize and bind to a variety of c*REs* [162], which are short (8 to 10 nucleotides long) sequences of DNA in the promoter region of TFs' target gene [163]. These regulatory regions organize in a modular way and form various more discrete regions called *cis*-regulatory modules (CRMs), which contain multiple transcription factor binding sites (TFBSs).

Whether the miRNA genes are expressed or not, their promoter sequences are the same in different plant tissues. Thus the activity of a promoter depends mostly on the functionality of the accessible c*REs* and the relative amount of their active binding TFs [164]. The type, number, position and combination of these c*REs* is associated with the spatiotemporal and inducible promoter expression, and they are thus involved in activating or silencing of plant miRNA gene expression at different processes of development or under specific conditions [164, 165]. Therefore, the understanding of plant promoter architecture and characterization of c*REs* is essential for clarifying the transcriptional regulatory code and for uncovering the regulatory mechanism of plant miRNAs [144, 163, 166, 167].

1.7   Public databases for *cis*-regulatory elements study

Cell responds to its environment by changing gene expression [168]. By activating or inhibiting the transcription machinery, transcriptional regulatory network regulates the gene expression and mediates the interactions with environment, which represents one aspect of plant cell signaling. Fundamentally, the transcriptional regulation is mediated by the recruitment of transcription factors to its *cis*-regulatory elements [169].

In the post genomic era, an increase of sequenced genome annotation has revealed the importance of c*REs* identification in promoter region to understand the transcriptional machinery of gene expression [170-172]. The identification of motifs will allow a better understanding of regulatory networks, and eventually will lead the technical processes on genetically modified organisms [171].

Several public databases for motif prediction have become available [169]. An overview of some database sites are shown in Table 1-1 (the type is indicated as "D"). Among them, the databases of AGRIS, PLACE, and PlantCARE are wildly applied. The databases PLACE (Plant *cis*-acting regulatory DNA elements), PlantCARE, RSAT (Regulatory Sequence Analysis Tools) and TRANSFAC are belong to one type [173-175], which will display the putative TFBS or motifs corresponding to the submitted DNA sequences in eukaryotes.

The resources AGRIS (Arabidopsis Gene Regulatory Information Server) and AtPAN are classified into another type, in which the putative TFBS in the promoter region of an *Arabidopsis* given gene will be present when a specific gene identifier is submitted [176-178].

In this study, the databases PlantCARE and AGRIS were applied. PlantCARE contains a number of known plant c*REs*, enhances and repressors. The majority c*REs*

29

data is extracted from the literatures, supplemented with computational predicted data. Moreover, it also provides *in silico* tools for promoter analysis [179].

1.8   Approaches in plant *cis*-regulatory elements analysis

With the current c*REs* databases, as time goes, more data of higher quality will be available, the work of new c*REs* identification and the improvements of annotation will be processed. In the meantime, the techniques for exploring novel and more expressive motif models will be called [180]. Nevertheless, a series of computational and experimental methods have been applied into c*REs* identification. The following sections will review the approaches in predicting and characterizing the c*REs* in both aspects.

1.8.1 Computational

Compare with the molecular techniques in identifying c*REs*, the computational methods do not call for prior knowledge of the TFs, and they are relatively low-cost and less labor intensive. In the meantime, they are more efficient to process high-throughput data within a relatively short period and generate large amount of results [181].

On the other hand, computational techniques for analyzing the miRNA data present certain disadvantages. When the predicting systems need to be trained, the negative control, the sensitivity and specificity, the appropriate input dataset, all these issues could cause imbalances [182]. Moreover, relying on a single bioinformatics tool will result in a high rate of false positives. It could be caused by the limited knowledge of the regulatory mechanisms, other possible factors that affect the transcription regulation, as well as the background noise inherent to process of high-throughput datasets [180]. In addition, the long distance of the promoter to its gene as well as multiple TSS may also be problematic for identifying the promoter region [183]. Therefore, biologists would be suggested to combine a

few complementary tools and take into account the top predicted motifs of each [172].

Over the past few years, a number of bioinformatics tools have been developed to identify the motifs. An overview of some database sites are shown in Table 1-1 (the types are indicated as "P"). Notably, some web-based tools for c*REs* prediction are not plant-specific, and generally applicable to plant studies, for instance, MEME program [184]. The expectation of maximization is an iterative algorithm that employs Position weight matrix (PWM) to scan the site specific frequencies of a sequence motif. MEME, one of such programs, can discover the most relevant motifs from the microRNA promoter sequences. In this program, researchers can define a range of desired motif widths, which is an important feature for motif prediction accuracy [185]. MEME requires one set of FASTA format sequences as the input file, and one background model file.

In this study, MEME program was applied.

1.8.2 Experimental

Regardless of the development of computational approaches for c*REs* characterization, the experimental validation of the promoters functions and their c*REs* is necessary, and should be carried out via plant transformation and transgene expression analysis [186]. However, due to low efficiency in operation and limited possibility for applying on high-throughput data, the traditional experimental methods are not often applied for identifying novel c*REs* anymore [181].

Currently, biological methods include:

reporter gene assay, a promoter sequence being analyzed for c*REs* activity is attached with a reporter gene and introduced into target plant cell for expression analysis [187, 188], and this approach was used in this study;

DNAaseI footprinting and gel mobility shift assay, both methods rely on the interaction between TFs and their DNA binding sites that contained in DNA fragments [189, 190];

in vitro **DNA selection assay** (SELEX, also known as SAAB (selected and amplified binding sites)), a protein is used for isolating high affinity binding sites via *in vitro* selection and amplification [191];

ChIP(chromatin immunoprecipitation)-microarray, the DNA is associated with a TF of interest, the DNA from this DNA-protein complex is used to probe a genomic DNA microarray [192];

the yeast one-hybrid (Y1H) assay that detecting physical interactions between TFs and the binding sites [193, 194].

In the meantime, the biological methods for elucidating the functions of predicted c*REs* have several drawbacks, e.g. the expensive protocols, low specificity, time-consuming and labor-intensive experiments. Therefore, the experimental and computational techniques are proposed to couple with each other to overcome these drawbacks [182].

As time goes, to adapt the evolving bioinformatics, the molecular and genetic approaches to elucidate the identified c*REs* functions are also in the process. Given the still limited knowledge of the plant TFs and their c*REs*, there is still space for c*REs* discovery in the future [169].

Table 1-1 Selected web-based resources for *cis*-regulatory element study

Resource	Type[a]	URL	Reference
AGRIS	D	http://arabidopsis.med.ohio-state.edu/	[195]
AtCOECis	D,P	http://bioinformatics.psb.ugent.be/ATCOECIS/	[196]
Athamap	D	http://www.athamap.de/	[197]
AtPAN	D	http://atpan.itps.ncku.edu.tw/	[177]
BAR Promoter	P	http://bar.utoronto.ca/ntools/cgi-bin/BAR_Promomer.cgi	[198]
DATF	D	http://datf.cbi.pku.edu.cn/	[199]
DOOP	D	http://doop.abc.hu/	[200]
ELEMENT	P	http://element.mocklerlab.org/	[201]
JASPAR	D	http://jaspar.genereg.net	[202]
MEME	P	http://meme.sdsc.edu/meme4_1/intro.html	[184]
Motifindexer	P	http://dinesh-kumarlab.genomecenter.ucdavis.edu/downloads.html	[203]
MotifSampler	P	http://ccmbweb.ccv.brown.edu/gibbs/gibbs.html	[204]
PLACE	D	http://www.dna.affrc.go.jp/PLACE/	[173]
PlantCARE	D	http://bioinformatics.psb.ugent.be/webtools/plantcare/html/	[179]
PlantPAN	D,P	http://plantpan.mbc.nctu.edu.tw/	[205]
PlantProm DB	D	http://linux1.softberry.com/berry.phtml?topic=plantprom& group=data&subgroup=plantprom	[206]
Plant TF DB	D	http://planttfdb.cbi.pku.edu.cn/	[207]
Plant Promoter DB	D	http://ppdb.agr.gifu-u.ac.jp/ppdb/cgi-bin/index.cgi	[208]
RSAT	D,P	http://www.rsat.eu/	[174]
TAIR pattern match	P	http://www.arabidopsis.org/cgi-bin/patmatch/nph-patmatch.pl	[209]
TRANSFAC	D	http://www.gene-regulation.com/pub/databases.html	[175]
WeederWeb	P	http://159.149.109.9/modtools/	[210]

[a] Type: D, database; P, prediction

1.9   DNA sequencing

DNA sequencing and its corresponding discipline, the genomics, are the combination of molecular biology and nucleotide chemistry [211].

Following the Human Genome project, the first so-called "Next generation sequencing" platform, 454 sequencing, was launched by 454 Life Science (now Roche) in 2005 [212-214], and Illumina sequencing released by Solexa the next year [215], followed by (ABI) SOLiD sequencing [216], which are the three major next generation sequencing methods. After years of evolution, these three platforms provide good performance on high-throughput, read length, accuracy, application, and cost [217].

In this project, Roche 454 method was utilized, and it also has been for years the major NGS platform. Briefly, based on a pyrosequencing method, with the help of nucleotide reagents, the complementary dNTP (dATP, dTTP, dGTP, dCTP) will correspondingly be cycled while the template strand is sequenced, and release pyrophosphate which equals to the intensity of incorporated nucleotide. A Standard Flowgram Format (SFF) file containing the basecalled sequences and corresponding quality scores for all high-quality, individual reads will be generated [212, 218].

The major advantages of 454 are its speed and read length (up to 1.0 kb) [219], accompanied by the automatic library construction, as well as the low manpower engaged. While a major limitation is its error rate in terms of homopolymers longer than six bp due to saturation in the detector [220].

Chapter 2.   Design, hypotheses and objectives of the project

2.1   Aim of this study

Given the fact that an increasing number of protein-coding genes are known to be regulated at the post-transcriptional level by microRNAs, the characterization of the regulatory elements in microRNA promoter region will elucidate the variation of plant miRNA regulation and eventually could be harnessed for crop improvement.

Following the development of an in-house high throughput method for isolation of microRNA promoters from non-model species, this study aims to identify the regulatory elements from microRNA390 gene across 16 *Brassicaceae* species and experimentally validate their putative roles, and eventually uncover the regulatory machinery of microRNA genes.

This will be achieved by a combined methodology of applying both bioinformatics comparative approaches and functional studies to elucidate the regulatory networks of functionally conserved microRNAs.

2.2   Hypotheses

2.2.1 Hypothesis1

Due to the conservation of plant microRNAs across different species over extended periods of time, a series of common conserved *cis*-regulatory elements in the promoter regions are shared by different species of *Brassicaceae*.

2.2.2 Hypothesis2

Given the miRN390-tasiRNA-ARFs pathway and known responsiveness of miRNA390 to heavy metals, putative heavy metal or auxin responsive *cis*-regulatory elements could be located in the promoter region of miRNA390a and miRNA390b.

2.3 Objectives

2.3.1 Objective1

To select highly conserved microRNA families and identify the conserved c*REs* present in multiple copies either in single or multiple microRNA genes.

Based on 454 sequencing data, using the fully sequenced genome of *A. thaliana* as the reference, characterize the promoter regions of conserved microRNAs from 16 plant species within *Brassicaceae* and predict the well conserved *cis*-regulatory elements.

2.3.2 Objective2

Deep mining of the *cis*-regulatory elements in both *Arabidopsis* miRN390a and miRN390b 1500bp upstream promoter region. Define the best candidates for functional analyses based on the degree of conservation in evolutionary terms within *Brassicaceae* family.

2.3.3 Objective3

In *Arabidopsis*, use Agrobacterium transformation method, construct GUS-fused promoters of both miRNA390a and miRNA390b, respectively, and build reliable reference lines for the following c*REs* functional characterization.

2.3.4 Objective4

Construct promoters undergone site-directed mutagenesis, and experimentally validate the putative c*REs* of miRNA390a by supplementing the

transgenic plants with specific environmental treatments and then detecting GUS gene expression.

Chapter 3. Identification of putative *cis*-regulatory elements of conserved miRNAs

3.1 Materials and methods

16 species from *Brassicaceae* were selected for this study (see Appendix Table S0-1). All plants have grown in the greenhouse from seeds obtained from wild population in Trentino under standard long-day condition.

Promoter regions encompassing mature miR390a and miR390b paralogs from all analyzed species were amplified by nested PCR-based genome walking with the Genome Walking Kit (Takara, Europe).

DNA deep-sequencing data was generated by the Roche 454 platform.

3.1.1 Computational analysis

The SFF files generated by 454 GS-FLX sequencer were used as input and processed with a variety of algorithms.

3.1.1.1 Quality overview

The sequencing reads generated were tested for sequence quality by FastQC with the default option [221].

3.1.1.2 Raw data processing

The raw data of SFF files was first analyzed by Roche "GS De Novo Assembler" (Newbler) 2.3 [222].

"SFF" files are Roche 454's "Standard Flowgram Format" files, containing the sequence data produced from a 454 run. Each SFF file contains a Manifest header at the start to describe the contents, as well as flow intensity signal values for each base in a single read.

The SFF files are in the binary format which needs to be converted into text format, such as Fasta, to carry out the data processing. Roche's own "sffinfo" utility can achieve this.

Roche 454 sequencers produce single SFF file for each region of a run, which means that in study 2 SFF files were generated.

As the first step, the two SFF files were joined into a single SFF file by "sfffile", with the following command:

$ sfffile -mcf MIDConfig.parse HZTVIVW01.sff HZTVIVW02.sff

Next, the MID (multiplex index, a short barcode sequence used to label samples/species when multiplexing) tags were used to split the single file into 16 sub-files.

Each SFF file contains the information of both the sequenced reads (FASTA) and the positional quality scores of the sequenced reads (QUAL). As the FASTQ files were required for the subsequent analysis, the FASTA and QUAL files were extracted separately with 454's "sffinfo" tool from individual SFF file, after which the two were merged into one FASTQ file with the Perl script.

$ sfffile -mcf MIDConfig.parse -s HZTVIVW.sff

$ sffinfo –s allfiles.sff > allfiles.fasta

 $ sffinfo –q allfiles.sff > allfiles.qual

$perl script.pl allfiles.fasta allfiles.qual > allfiles.fastq

3.1.1.3 5' adapter trimming

Since a pair of specific PCR primers were fused with each sample during sequencing, if the primers are distant or conversely in the sequence during data processing, the assembler may overlap the real contigs with these primers producing noise. It is therefore necessary to remove the primers of one end before processing into assembly. Meanwhile the primers of the other end are temporarily kept as identifiers of different samples.

The FASTQ files of 16 species were processed as input to trim the 5' adapters with CUTADAPT program according to its manual (*GS* Reference Mapper can also handle this) [223].

The MID was placed between the sequencing key and the Primer 1 sequence-specific primer, thus the sequencing reads were longer than the actual sequences. Primer 1 MID needed to be trimmed before the read mapping. 5' end of reads that matched any substring of primer1 MID were removed.

CUTADAPT is a suite of scriptable tools for small and large tasks arising in high-throughput sequencing projects, and supports 454 data trimming. In practice, the 5' end adapters were removed with the "-b" parameter that aimed to trim either 5' or 3' adapters.

Python script usage:

```
$ python2.7 bin/cutadapt –help
Usage: $ cutadapt [options] <FASTA/FASTQ FILE> [<QUALITY FILE>]
$ python2.7 bin/cutadapt -b CTGTGTGCTCACTCTCTTCTGTCA
Bin.fastq > Bin_cut_def.fastq
```

3.1.1.4 Assembly

In order to identify the similarities among input reads, it is necessary to build contigs. This was achieved by assembling the input reads.

Assembly is the process to merge the short reads into long contigs (a set of contiguous sequences, ideally a full transcript) by searching the best sequence overlaps between reads.

Newbler (Roche's *GS De Novo Assembler*) will identify pairwise overlaps between reads, align multiple sequences of overlapping reads, and then generate longer contigs in FASTA format.

With the default values, the "runassembly" option in Newbler was run with command line to assemble the FASTQ files and to generate the 16 FASTAQ files of each species. The command line used was as follows:

$ runAssembly file.fastq

3.1.1.5 Reference sequences and microRNA promoter sequences generation

In order to map the contigs obtained from assembly process to reference sequences, a dataset was generated by extracting 1500 bp upstream of 299 *Arabidopsis* miRNA hairpins annotated in miRBase (Release 19: August 2012) from *Arabidopsis* 5 chromosomes downloaded in NCBI.

This step was conducted with the Perl script.

Perl script usage:

$ perl cutseq.pl chro_ID_site.txt chro.fas > Refseq_up1500_extract.fas

3.1.1.6 NCBI Blastn

NCBI Blastn was used to perform the multiple alignments between contigs and reference sequences. When aligning, the FASTA files of clean contigs, which were generated from FASTQ by EMBOSS were taken as the subjects and the extracted miRNA hairpin sequences as the queries.

The "formatdb" command in the BLAST suite was used to format the database of *Arabidopsis* miRNAs 1500 bp upstream *of* promoter sequences, thereby ensuring that this database could be searched by BLASTALL program. The basic parameters used here were: -p=<F> (nucleotide); -i=< the input file for formatting>; -o=<T>, true, parse seqid and create indexes. The command line was as follows:

$formatdb -p F -i Refseq_up1500_extract.fas -o T

With the formatted *Arabidopsis* database, The BLASTALL program was applied in the following step. The parameters used were: -p=<program name>, blastn (for nucleotides); -d=<database> (formatted file in previous step); -i=<query file> (FASTA format); -m=<alignment view option>, 8=<output format as tabular>; -e=<expectation value>, 0.05; -o=<BLAST report output file>; -b=<number of concatenated queries for blastn>, 1. The command line was as follows:

$ blastall –p blastn –d Refseq_up1500_extract.fas –i *Bin_* rename.txt –m8 –e 0.05 –o *Bin*_blastn.txt –b1

Afterwards, we applied the Perl script to calculate multiple occurrences of the combination of microRNA positions and corresponding contigs within individual species (e.g. contig00059§chr5_8526969_8528149_+), the contigs with less than two occurrences were discarded.

3.1.1.7　Contigs' extraction from cleaned data and renaming

Linux command line "grep" stands for "global regular expression print" and aims to process the standard input for lines containing a match to the given *PATTERN* [224]. In this study, this command was used to extract the titles of contigs.

Subsequently, the names of microRNAs and matched contigs were combined, based on the BLASTN results from 3.1.1.6 step. The script used was as follows:

```
$ grep ">" 454 Allcontig.fas > outfile
```

To ensure all the sequences were in the same orientation, we reverse-complemented all the sequences in the reverse strand with the following python script.

```
Usage: $ python revecomp.py Seq.fasta > revcomp.fasta
```

3.1.1.8　Removing duplicates

Cd-hit-454 is a program to identify and extract duplicated 454 sequencing reads, including near identical duplicates. It can be used to reduce the redundancy of the reads [225].

In the analysis, the parameters for this program were set as follows[225]: -c=<sequence identity threshold>, 0.95; -i=<input filename in FASTA format>; -o=<output filename>, and with other parameters as default. cd-hit-454 was run on 16 renamed FASTA input files. The command line was as follows:

```
cd-hit-454 usage:
```

```
$ cd-hit-454 -i sample.fas -o out.fas -c 0.95
```

3.1.1.9   3' end adapters trimming

Different from 3.1.1.3, the parameter "-g" that aims specifically to remove any 3' adapter here was used with other parameters as default.

The python script of CUTADAPT used was as follows:

$ python2.7 bin/cutadapt -g CTGTGTGCTCACTCTCTTCTGTCA

*Bin_* revcom.fasta > Sp02 _revcomp_para_g.fas

3.1.1.10   MicroRNAs extraction by family/locus

In order to join each microRNA from the same family/locus into one single file, we extracted all the same microRNAs sorted by the loci. The script used was as follows:

Perl script usage:

$perl extractor_multifas_splitbymatch.pl

There were totally 86 microRNAs families and 132 microRNA loci (given the false positive, only miRNA genes detected in more than two species were calculated).

3.1.1.11   Weight calculation of all the miRNAs

Before applying to the motif searching program, sequence weights (numbers in the range of 0< weight <1) of each FASTA file separated by loci were requested and calculated by Perl script, and sequences would contribute to motifs in proportion to their weights [226]. Sequences shorter than 100 bp were discarded by the script.

The weights of each sequence were determined by calculating the frequency of a specific group of string identifiers present in the caption.

The Perl script used for this step was as follows:

Usage: $ perl meme_make_weights.pl

3.1.1.12 *De novo* **motifs prediction**

MEME [1] (Multiple EM for Motif Elicitation) program was applied for searching the *de novo* motifs from the conserved miRNAs selected within *Brassicaceae* [184, 227]. The descriptions of MEME was given in 1.8.1.

A FASTA file was extracted from TAIR9 (*Arabidopsis* genome annotation database) 1000 bp upstream of the TSS of coding sequences (cds). With "fasta-get-markov" tool in the MEME Suite, a Markov Background Model was then generated for motif mining [228].

Then the weights of each FASTA file were taken as input, the TAIR9 background model as the annotation database, with parameter- -dna=< sequences use DNA alphabet>; -mod=<distribution of motifs>, anr; -revcomp=< allow sites on + or - DNA strands >; -miniw=<minimum motif width >, 6; -maxw==<maximum motif width>, 10; –bfile=<name of background Markov model file; -nmotifs=< maximum number of motifs to find); -evt=< stop if motif *E*-value greater than <evt>>, 0.001; -oc<output dir>=<name of directory for output files will replace existing directory>. Other parameters were set as default. MEME program was run on each microRNA loci group with the following command line, taking the sequence group of miR156 as example:

$ meme MIR156 -dna -mod anr -revcomp -minw 6 -maxw 9 –bfile TAIR9_upstream_1000_translation_start-model -nmotifs 15 –evt 0.001 -oc meme_156a

---

[1] MEME suite was downloaded from http://meme-suite.org/

In the following step, MAST (Motif Alignment and Search Tool [227], in The MEME Suite) was applied to search sequences with motifs of the same nucleotides with the following command line:

MAST parameters: -c<count>=<only use the first <count> (default: 0)>; -o<dir>=<directory to output mast results; directory must not exist>.

MAST command line:

```
$ mast MIR156-meme.txt MIR156_fas.txt -c 15 -o mast_MIR156
Batch file processing with Perl script "run_program.pl".
Perl script usage:
$ perl run_program.pl
$ sh meme.sh
```

3.1.1.13 Validation of *in silico* motifs predicted with known database

Given the possible false positive motifs predicted by MEME and the difficulty to validate all the numerous motifs experimentally, several preliminarily filtering steps were necessary to filter the most relevant motifs.

First, we searched the c*REs* of the promoter sequences against PlantCARE database[2] (The descriptions of PlantCARE was given in 1.7). Those motifs detected by PlantCARE were filtered manually and further defined by Bioedit, a common biological sequence alignment editor. The overlapped motifs detected from both MEME and PlantCARE were considered and chosen as good candidates. The putative TSS locations were annotated according to the 63 miRNAs identified by Xie [146].

---

[2] PlantCARE database is available online http://bioinformatics.psb.ugent.be/webtools/plantcare/html/

3.1.1.14 Similarity comparison of the MEME motifs by Tomtom

To verify the success of common motifs discovered by MEME and PlantCARE, we searched the motifs against a database of known TF motifs with Tomtom program [229], which is available in the MEME suite. The motifs with a statistically significat *P*-value of less than 0.05 and *E*-value of less than 10 were considered relevant.

Tomtom is a free online programme used for comparing the motifs, the MEME motif sequences which shared more than 2 nucleotides with PlantCARE were taken into account. The MEME matrix file of each microRNA was used as the input query motifs, and the database of PBM motifs for *Arabisopsis* was applied as the target database, other parameters were set as default.

3.1.1.15 Motifs filtered by STAMP

For microRNA families such as miR160, miR167 and miR319, each of their paralogs have been detected in more than 10 species.

In order to filter the putative motifs identified in these microRNA families, the alignment and logo of each motif were obtained with another *Arabidopsis* database AGRIS [176], and the motifs were processed by the STAMP program[3] online [230]. This program contains several databases and uses Weblogo tool to generate the motif logos.

The information content for trimming motif edges was set to less than 0.4, and other parameters were considered as default settings, the alignment and logos of individual motifs were generated.

---

[3] STAMP, http://www.benoslab.pitt.edu/stamp/

3.1.1.16 Phylogenetic analysis of miR390 *cis*-regulatory elements

To estimate the correlation between the phylogenetic distance of miR390 promoter sequences and the phylogenetic relationship of the tested species within *Brassicaceae*, we conducted Molecular Evolutionary Genetics Analysis (MEGA) software[4] to reconstruct the phylogenetic tree, and then compared the results with the known *Brassicaceae* phylogeny.

Parameters:

Test of phylogeny=bootstrap method; Model/Method= Maximum Composite Likelihood; Nr of bootstrap replication=1000; Gaps/Missing data treatment=Pairwise deletion. Other parameters were set as default.

The whole pipeline of the computational analysis was shown in Figure 3-1.

---

[4] MEGA, http://www.megasoftware.net/
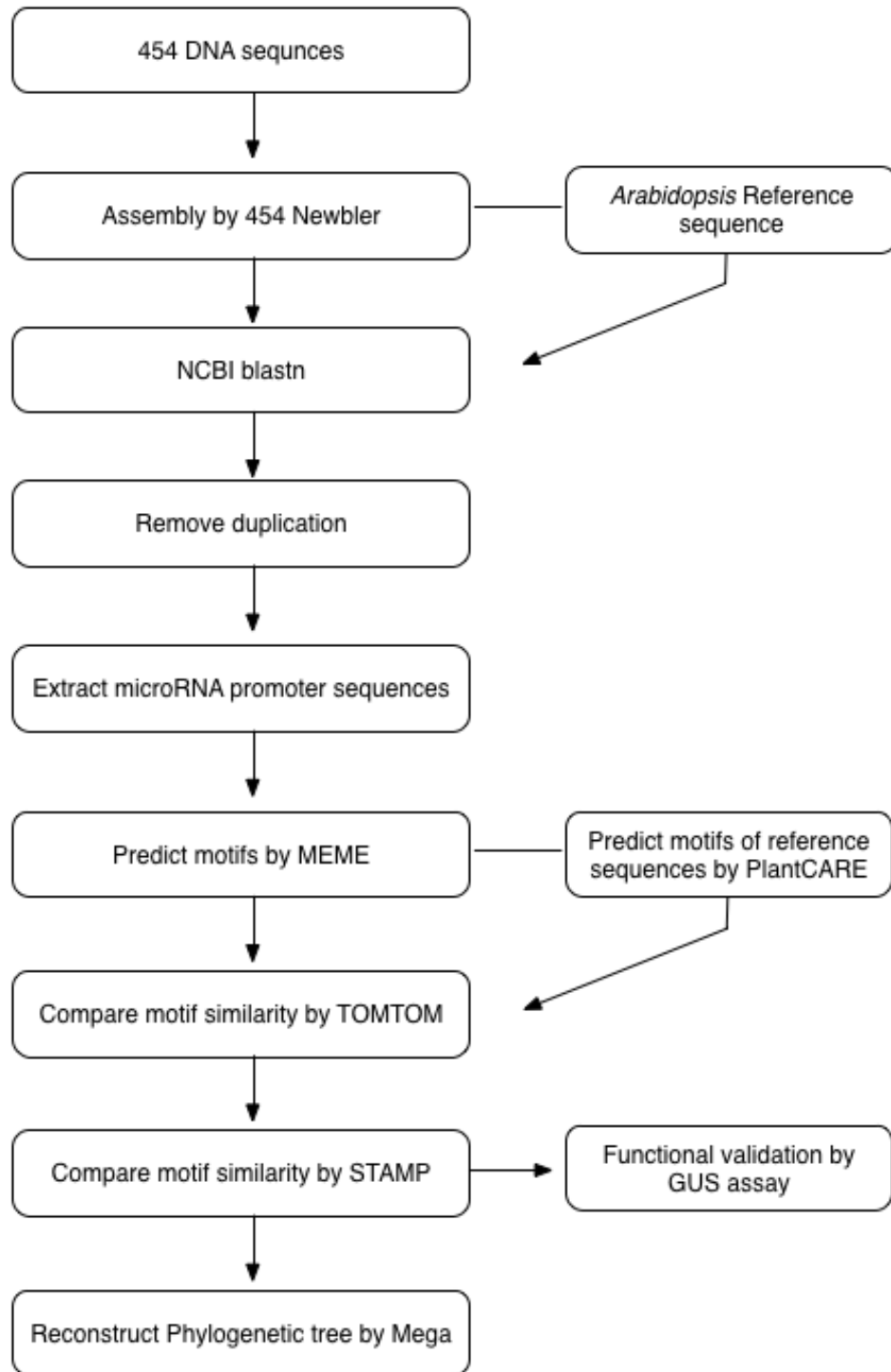
Figure 3-1 Flowchart of microRNAs c*REs* identification and validation

3.2   Results

3.2.1 Sequences generated

Figure 3–2 showed that the distribution of the average quality scores per sequence from the unique dataset ranged from 10 to 39, the quality of the majority of sequences between 0-559bp was relatively good with a quality score of more than 20. The distribution of the quality scores per sequence across all sequences within the set rationally ranged among 22 to 30. Since the dataset of DNA sequences was generated by relatively short reads and without trimming the adapters, it is not surprising that the quality score range was broad.

As shown in Table 3-1, a total of 459060 sequences were obtained from 454 sequencing platform, and the sequence length ranged from 2 to 1182. Since there were 5' and 3' adapters ligated in the original sequences, this could explain the fact of the less than 10bp short reads.

Table 3-1 Summary of numbers and length scope of sequences from different species

Species ID	Sequence Nr.[1]	Sequence length	%GC
Bin	27871	2-875	39
Sof	22495	2-882	37
Npa	34770	2-897	38
Mva	30755	2-882	41
Iam	36128	2-928	38
Tha	36773	2-912	38
Aal	37057	2-896	39
Chi	29987	2-1182	35
Npr	22866	2-1079	39
Aha	14165	2-889	38
Bni	21728	2-918	39
Lpe	19986	2-900	37
Agr	28119	2-878	37
Tar	31142	2-1110	35
Dso	46224	2-994	36
Cim	18994	2-883	38

[1]Nr., number.
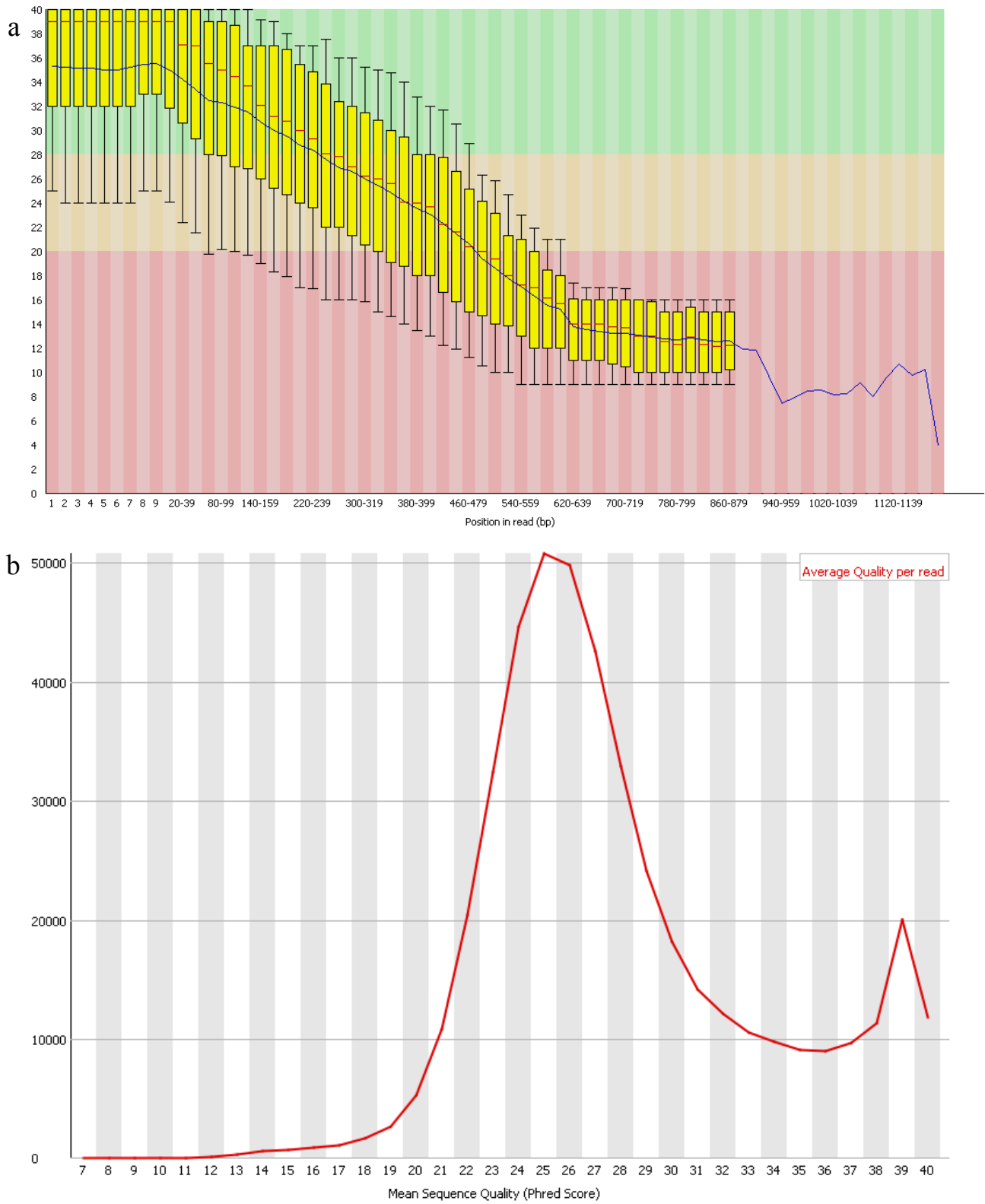
Figure 3–2 Fastqc analysis of fastq sequences from total datasets of 16 species. a. Quality scores distribution per base position across all bases within the data set. b. Distribution of the quality scores per sequence across all sequences within the set always ranges between 22 to 30 QS (quality score) values.

51

3.2.2 MicroRNA loci detected in this study

After reducing redundancy by cd-hit-454 program, there were in total 86 microRNAs families (supplementary Table S0-4) and 132 microRNA loci detected in different species.

As Table 3-2 showed, the item ratio of the number of sequences and the number of species substitutes the redundancy of individual datasets of miRNA genes. The value 1 indicated there were no duplicates in the miRNA loci dataset. The duplicates of most genes were limited, while some miRNA families still contained an amount of redundancy. The genes selected for the following analysis with or without duplicates would be all manually aligned in Bioedit program, and the redundant sequences would be removed.

There were 23 microRNA loci present in more than 10 species. The gene with the shortest average length was miR855, about 636bp, and it was detected in eight species. The one with the maximum average length was miR5642a, 1471 bp, and it was present in 6 species. The average length of most microRNA genes was about 1000bp, which was a rational range in consensus with the promoter isolation region obtained experimentally.

Table 3-2 summary of the miRNAs (by loci) detected

MiRNA Gene	Nr.[2] of Species	Nr. of Seq[3]	Seq/Spe[4] Ratio	Average length of Seq	MiRNA Gene	Nr.[2] of Species	Nr. of Seq[3]	Seq/Spe[4] Ratio	Average length of Seq
[1]MIR171b	17	20	1.18	942.85	MIR156b	5	5	1.00	902.20
MIR160a	16	21	1.31	927.29	MIR156f	5	5	1.00	897.00
MIR319a	16	24	1.50	811.50	MIR159b	5	6	1.20	971.17
MIR399c	16	23	1.44	905.09	MIR164c	5	5	1.00	1080.80
MIR162b	15	23	1.53	897.17	MIR172e	5	5	1.00	880.80
MIR166g	15	18	1.20	882.72	MIR2937	5	6	1.20	826.67
MIR168a	15	18	1.20	928.61	MIR399f	5	5	1.00	813.60
MIR169a	15	18	1.20	967.89	MIR5015	5	11	2.20	668.18
MIR169f	15	21	1.40	844.57	MIR5641	5	6	1.20	809.83
MIR166a	14	21	1.50	933.10	MIR5645a	5	6	1.20	833.83
MIR166b	14	20	1.43	1010.75	MIR5665	5	6	1.20	840.00
MIR166e	14	15	1.07	1055.53	MIR161	4	4	1.00	1185.25
MIR169b	14	18	1.29	902.78	MIR169j	4	4	1.00	881.00
MIR172a	14	19	1.36	645.21	MIR171c	4	4	1.00	978.75
MIR319b	14	19	1.36	725.47	MIR172b	4	5	1.25	757.40
MIR319c	14	18	1.29	1012.11	MIR414	4	4	1.00	799.75
MIR395e	14	24	1.71	838.33	MIR4239	4	4	1.00	802.75
MIR156d	13	15	1.15	987.00	MIR5014a	4	4	1.00	839.00
MIR160c	13	15	1.15	1203.93	MIR5631	4	4	1.00	894.50
MIR167a	13	15	1.15	887.27	MIR5644	4	4	1.00	833.50
MIR169b	13	17	1.31	959.88	MIR5645e	4	4	1.00	956.00
MIR169g	13	25	1.92	775.76	MIR5645f	4	5	1.25	1212.80
MIR162a	12	14	1.17	925.93	MIR5660	4	4	1.00	838.00
MIR164b	12	13	1.08	1043.38	MIR5998b	4	10	2.50	692.70
MIR167b	12	15	1.25	1082.00	MIR826	4	4	1.00	840.75
MIR170	12	17	1.42	990.88	MIR836	4	4	1.00	1014.50
MIR394b	12	16	1.33	811.25	MIR845b	4	5	1.25	743.40
MIR396b	12	14	1.17	845.00	MIR169e	3	3	1.00	1072.00
MIR160b	11	14	1.27	902.93	MIR169l	3	3	1.00	1360.00
MIR171a	11	15	1.36	888.40	MIR169n	3	3	1.00	915.00
MIR165a	10	11	1.10	1026.64	MIR173	3	3	1.00	1294.00
MIR168b	10	10	1.00	1114.50	MIR1888a	3	4	1.33	811.75
MIR391	10	11	1.10	929.36	MIR2111b	3	3	1.00	1007.33
MIR395a	10	11	1.10	932.00	MIR2936	3	3	1.00	922.33
MIR395f	10	11	1.10	828.18	MIR2938	3	3	1.00	902.00
MIR397a	10	12	1.20	1120.50	MIR3932a	3	4	1.33	967.25

MIR					MIR				
MIR390a	9	12	1.33	866.50	MIR394a	3	4	1.33	872.25
MIR5655	9	15	1.67	691.33	MIR398b	3	3	1.00	1306.00
MIR159a	8	9	1.13	798.11	MIR399e	3	3	1.00	1011.67
MIR166c	8	9	1.13	880.67	MIR4227	3	5	1.67	786.60
MIR169c	8	14	1.75	942.36	MIR4228	3	3	1.00	879.67
MIR169d	8	10	1.25	1227.30	MIR5014b	3	3	1.00	794.67
MIR390b	8	9	1.13	855.89	MIR5020a	3	3	1.00	1086.33
MIR393a	8	10	1.25	1113.10	MIR5029	3	4	1.33	880.25
MIR393b	8	8	1.00	864.88	MIR5630b	3	3	1.00	1047.33
MIR395b	8	9	1.13	1231.89	MIR5635a	3	3	1.00	992.33
MIR395c	8	9	1.13	861.00	MIR5635d	3	3	1.00	1100.67
MIR395d	8	8	1.00	1080.25	MIR5638a	3	3	1.00	992.67
MIR5643b	8	12	1.50	728.92	MIR5651	3	3	1.00	1167.00
MIR827	8	11	1.38	1171.91	MIR5658	3	3	1.00	904.67
MIR855	8	46	5.75	635.76	MIR5664	3	3	1.00	840.67
MIR156a	8	9	1.29	868.44	MIR5997	3	3	1.00	909.33
MIR166d	7	8	1.14	840.13	MIR828	3	3	1.00	937.33
MIR408	7	9	1.29	1079.56	MIR830	3	3	1.00	946.00
MIR5634	7	7	1.00	673.57	MIR831	3	3	1.00	986.33
MIR5642a	7	7	1.00	1470.57	MIR832	3	4	1.33	1005.75
MIR824	7	11	1.57	1083.82	MIR837	3	3	1.00	970.67
MIR157d	6	7	1.17	850.14	MIR842	3	5	1.67	842.60
MIR159c	6	6	1.00	1387.00	MIR843	3	3	1.00	986.67
MIR172c	6	6	1.00	987.50	MIR852	3	3	1.00	1045.33
MIR172d	6	7	1.17	741.86	MIR856	3	3	1.00	963.67
MIR396a	6	7	1.17	1008.29	MIR860	3	3	1.00	908.33
MIR399a	6	7	1.17	756.14	MIR863	3	3	1.00	1041.00
MIR401	6	13	2.17	714.38	MIR868	3	3	1.00	869.33
MIR5027	6	7	1.17	707.14	MIR869	3	4	1.33	935.50
MIR857	6	7	1.17	924.86	MIR870	3	3	1.00	863.00

[1]MIR, microRNA; [2]Nr., number; [3]Sep, species; [4]Seq, sequence. The microRNAs were listed based on the number of species that they were detected.

3.2.3 Motifs detected from different conserved miRNAs

To have a better understanding of the regulatory mechanisms of plant miRNAs, the identification of the regulatory elements in promoter regions was necessary.

In the evolutionary history of miRNA genes, there were several plant miRNA families and their targeting sequences on mRNAs remain conserved. According to this, we predominantly processed nine miRNAs families with 29 loci in total as shown in Table 3-3. Majority of the miRNA loci were present in more than 10 species in this study. The most relevant motifs detected by both MEME and PlantCARE, were filtered by similarities by TOMTOM and STAMP programs.

The motifs identified with E-value less than 0.001 from eight families of miR156, miR160, miR166, miR167, miR171, miR319, miR395 and miR408 were shown in Table 3-4, Table 3-5, Table 3-6, Table 3-7, Table 3-8, Table 3-9, Table 3-10 and Table 3-11, respectively. The motifs detected in miRNA390a and miRNA390b were analyzed and discussed separately in 3.2.4.

There were 6 motifs identified in miR156 family, 13 in miR160 family, 25 in miR166 family, 13 in miR167 family, 8 in either miR171 or miR319 family, 18 in miR395 family, 3 in miR408 family, and 94 putative motifs in total. Some motif sequences were predicted to be functionally similar, within or out of the gene family. For instance, as listed in Table 3-5, miR160a-M3 and miR160c-M3 were both identified as putative SORLIP5 motif. In addition, some microRNA genes have multiple paralogs in *Arabidopsis*, such as miR166 and miR395. But the distribution of *cis*-regulatory elements in different miRNA loci was unequal, suggesting the predominant roles of some miRNA loci.

Examples of motif alignments of miR160a, which were generated by STAMP were shown in Appendix. In Figure S2, we can see that the first motif identified of miR160a was GGAGGAATAG, and it was predicted by STAMP as (APETALA1) AP1 binding site (CTAAAAATGG) with E-value 3.9411e-04. AP1 element was presumably positive regulators of (Arabidopsis heat shock protein 90) AtHsp90-1 gene expression under Arsenite (As) treatment [231]. Previous study predicted

miR160 was down-expressed under As treatment in rice [232]. This clue implies a putative correlation between the predicted AP1 motif and the miR160. In Figure S3, the second motif identified of miR160a was GATTTGCATG, with E-value 2.5701e-08 it had the best alignment with RY-repeat (CATGCATG), which is an essential motif for seed-specific expression in both legumin and soybean [233, 234]. In addition, the RY repeat is known to be a target of seed-specific regulator FUSCA3 (FUS3), the latter was shown to regulate miR160 expression [235]. Briefly, these evidences demonstrate that the RY repeat may play crucial role in seed-specific expression that mediated by miR160. As showed in Figure S4, the third motif identified of miR160a was GATGGAGAGA, it was aligned with (Sequences Over-Represented in Light-Induced Promoters 5) SORLIP5 (GAGTGAG) with E-value 5.9209e-04. SORLIP5 is a (phytochrome A) phyA-regulated motif and was overrepresented in the promoter regions of genes that involved in seed development [236, 237]. In Figure S5, the forth motif identified of miR160a was GTGTTTGGGT, with E-value 5.7262e-04, it had the best alignment with GT motif (CATATTAACCACACA), which was involved in light-responsive process in pea [238]. The last identified motif of miR160a was GGGGTTGCTT as shown in Figure S6, and it was predicted to be the (TELOMERE-box) TELO-box (AAACCCTAA) with E-value 1.9702e-04. The TELO-box is known to participate in the control of a β-glucoronidase gene expression in *Arabidopsis* root primordia [239].

It is well known that miR160a targets ARF10, ARF16, ARF17 [240], and ARFs futher regulate the expression of auxin-inducible genes thus lead the proper development of plants. ARF10 was reported to regulate floral organ identity and seed germination [241]. In addition, the auxin receptor (TRANSPORT-INHIBITOR-RESISTANT1) TIR1 can degrade Aux/IAA level in response to auxin and then regulates lateral root development [242].

Taken together, the identification of these putative motifs in the promoter region of miR160a suggests that miR160a may play essential roles in response to both biotic and abiotic stresses, as well as in plant development, via regulating the expression level of its targets ARFs.

Table 3-3 Summary of identified conserved miRNAs genes

MiRNA ID	Gene bank accession	Nr. of Species	miRBase accession	Location	Strand	Start	End	RPM
miR156a	AT2G25095	8	MI0000178	Chr2	-	10676451	10676573	2.22e+04
miR156b	AT4G30972	5	MI0000179	Chr4	+	15074899	15075081	2.3e+04
miR156d	AT5G10945	13	MI0000181	Chr5	-	3456632	3456749	2.2e+04
miR160a	AT2G39175	16	MI0000190	Chr2	+	16340279	16340363	2.51e+04
miR160b	AT4G17788	12	MI0000191	Chr4	+	9888982	9889070	2.35e+04
miR160c	AT5G46845	14	MI0000192	Chr5	-	19009094	19009182	2.38e+04
miR166a	AT2G46685	14	MI0000201	Chr2	+	19176108	19176277	2.52e+04
miR166b	AT3G61897	16	MI0000202	Chr3	+	22922206	22922325	2.52e+04
miR166c	AT5G08712	8	MI0000203	Chr5	+	2838635	2838773	2.61e+04
miR166d	AT5G08717	7	MI0000204	Chr5	+	2840622	2840734	2.44e+04
miR166e	AT5G41905	15	MI0000205	Chr5	-	16775520	16775662	2.37e+04
miR166g	AT5G63715	15	MI0000207	Chr5	+	25504798	25504919	2.44e+04
miR167a	AT3G22886	15	MI0000208	Chr3	+	8108072	8108209	1.64e+05
miR167b	AT3G63375	14	MI0000209	Chr3	+	23406168	23406276	1.63e+05
miR171a	AT3G51375	13	MI0000214	Chr3	+	19073434	19073556	3.13e+04
miR171b	AT1G11735	17	MI0000989	Chr1	-	3961348	3961464	7.39e+03
miR171c	AT1G62035	4	MI0000990	Chr1	-	22930089	22930204	4.17e+04
miR319a	AT4G23713	16	MI0000544	Chr4	+	12352956	12353131	4.87e+03
miR319b	AT5G41663	14	MI0000545	Chr5	-	16660469	16660640	6.24e+03
miR319c	AT2G40805	14	MI0001086	Chr2	+	17029701	17029899	3.23e+04
miR390a	AT2G38325	10	MI0001000	Chr2	+	16061954	16062060	6.37e+04
miR390b	AT5G58465	9	MI0001001	Chr5	+	23636947	23637066	6.37e+04
miR395a	AT1G26973	9	MI0001007	Chr1	-	9363196	9363288	464
miR395b	AT1G26975	8	MI0001008	Chr1	+	9364471	9364570	447
miR395c	AT1G26985	8	MI0001009	Chr1	+	9367080	9367179	447
miR395d	AT1G69792	9	MI0001010	Chr1	-	26269979	26270078	447
miR395e	AT1G69795	15	MI0001011	Chr1	-	26272776	26272870	464
miR395f	AT1G69797	10	MI0001012	Chr1	+	26273858	26273969	447
miR408	AT2G47015	7	MI0001080	Chr2	+	19319814	19320031	7.7e+03

Table 3-4 motifs identified in microRNA156 family

microRNA ID	Motif ID	Motif Sequences detected by MEME	STAMP alignment	STAMP ID[1]	STAMP E-value[2]
miR156a	miR156a-M1	Rev: GAGAAGAGAA	GASAAAARMC---ACTAAAAATGG	AG_v2	6.3984e-04
miR156a	miR156a-M2	For: AGAGTCTTAC	AGAGTCTTAC AGATTKTT--	CCA1	1.8368e-03
miR156a	miR156a-M3	Rev: TCTTACATGC	TCTTACATGC ---CACATG	AtMYC2	8.0428e-05
miR156b	miR156b-M1	For: CGGAGATT	CGGAGATT------AGATTKTT	CCA1	6.1244e-04
miR156b	miR156b-M2	For: ACACGTGTCT	ACACGTGTCC -CACGTG---	G-box	2.5126e-08
miR156d	miR156d-M1	For: GTTAGGGTT	RTTAGGGTTT -TTAGGGTTT	TELO-box	2.4647e-14

Rev, reverse strand; For, forward strand. [1]STAMP ID is the identifier of STAMP best consensus sequence matching motif. [2]STAMP E-value represents expectation value of the STAMP alignment.

Table 3-5 motifs identified in microRNA160 family

microRNA ID	Motif ID	Motif Sequences detected by MEME	STAMP alignment	STAMP ID[1]	STAMP E-value[2]
miR160a	miR160a-M1	Rev: GGAGGAATAG	GGAGAAAKRG CCAAAAATGG	AG_v3	3.9411e-04
miR160a	miR160a-M2	For: GATTTGCATG	GATWTGCATG --CATGCATG	RY-repeat	2.5701e-08
miR160a	miR160a-M3	For: GATGGAGAGA	TCTCTSAATC ---CTCACTC	SORLIP5	5.9209e-04
miR160a	miR160a-M4	For: GTGTTTGGGT	------ACCCAAACAC CATATTAACCACACA-	GT	5.7262e-04
miR160a	miR160a-M5	For: GGGGTTGCTT	--AGGGTTNCTT TTAGGGTTT---	TELO-box	1.9702e-04
miR160b	miR160b-M1	For: AGAGAGAAAG	STTTCTCTTT--------CTTTCCATTTTTAGTAAC	CArG3	4.4310e-05
miR160b	miR160b-M2	For: GAGGAATGGT	-----GAGGAATGGT---TAATCCATGAAAGGTAAG	CArG2	1.8004e-03
miR160b	miR160b-M3	For: TATATAGAGG	-TATATAGAKG ATATATACA--	SORLREP3	3.9563e-06
miR160b	miR160b-M4	For: CATCACCACA	ACAGCGTGCA-------CATGCATG	RY-repeat	1.3724e-02
miR160c	miR160c-M1	For: GATCTTGGCT	GATCTTGGCT ----GTGGCT	SORLIP1	8.0302e-05
miR160c	miR160c-M2	For: GTTGAAGAGG	CMTCTTCAWC CCACGTCA--	TGA1	1.9541e-03
miR160c	miR160c-M3	For: GGAAAGAGAG	YTCWYTTTCC CTCACTC---	SORLIP5	1.0933e-03
miR160c	miR160c-M4	Rev: GGTTTTGATT	--------AAYCGTAACC--TTGGTTTTGATCAAAACCAA	PII	1.7147e-04

Rev, reverse strand; For, forward strand. [1]STAMP ID is the identifier of STAMP best consensus sequence matching motif. [2]STAMP E-value represents expectation value of the STAMP alignment.

Table 3-6 motifs identified in microRNA166 family

microRNA ID	Motif ID	Motif Sequences detected by MEME	STAMP alignment	STAMP ID[1]	STAMP E-value[2]
miR166a	miR166a-M1	For: GTCCCCCACG	CRTGKGKGNC CATGTG----	AtMYC2	2.7237e-04
miR166a	miR166a-M2	Rev: GGAGAAGAAG	CTTCTTCTCY ---CTTATC	Ibox	6.5288e-04
miR166a	miR166a-M3	Rev: GTGAGTGGTG	KTGAGTRGTG ATGAGT----	ATB2_AtbZIP53_AtbZIP44_GBF5	6.8055e-05
miR166a	miR166a-M4	For: AAAAGGGG	---RAAAGGGG CCAAAAATGG	AG_v3	7.5347e-04
miR166a	miR166a-M5	For: CAAGACAAG	CAAGACAAGA -GAGACA---	ARF	6.7282e-05
miR166a	miR166a-M6	Rev: TTGAGACATG	TTGARACATG --GAGACA--	ARF	8.0128e-06
miR166b	miR166b-M1	For: CAGATCCGGC	CAGATCCGGC --GATCCGCG	octamer	1.3387e-05
miR166b	miR166b-M2	Rev: GAAGAGAAAG	GRWGAGAGRG ---GAGTGAG	SORLIP5	1.8380e-04
miR166b	miR166b-M3	Rev: GATGAAGAGA	---YCTCTTCWTC--- TTTCCTATTCTGTTTT	AG_v4	3.9028e-03
miR166b	miR166b-M4	Rev: TGGGTCCCAT	TTNGTCCCMC ---GGCCC--	SORLIP2	2.3647e-02
miR166b	miR166b-M5	For: TTTGGTCAAA	-----TTTGGTCAA------ TTGGTTTTGATCAAAACCAA	PII	2.1404e-07
miR166b	miR166b-M6	Rev: TAAACATTTG	CCCATGTTTA- -CCATTTTTAG	AP1	1.5583e-08
miR166b	miR166b-M7	For: TTTACTTTTT	-TTTACTTKTT TTWTACTAGT	SORLREP1	3.9935e-06
miR166b	miR166b-M8	For: GAGACATGTC	GAGACATGTC GAGACA----	ARF	7.2848e-08
miR166c	miR166c-M1	Rev: AAGAAGAAGA	TCTTCTTCWT TCTACGTCAC	LS7	1.0607e-03
miR166c	miR166c-M2	Rev: GTCTCTTTT	-GTCTCTTTTC TGTCTC-----	ARF	1.8221e-04
miR166c	miR166c-M3	For: GGGATCGAAT	AATCGNTCCC CGTCGATCT	Nonamer	3.5907e-04
miR166d	miR166d-M1	For: TAGTTGTTTCTT	---TAGTTGTTTCTT TCATAGATTTTTTTT	CCA1_v3	1.6559e-05
miR166e	miR166e-M1	For: GAGAGAGAGA	CTCTYTCTC --CTCACTC	SORLIP5	3.0126e-04
miR166e	miR166e-M2	Rev: GAAAGCAATG	-CMTWGCTTTC---- TCATAGATTTTTTTT	CCA1_v3	6.2480e-06
miR166e	miR166e-M3	For: AAAGTCGAAA	AAAGTCAAAA ---GTCAA--	W-box	5.5782e-05
miR166e	miR166e-M4	Rev: CACCTCTTTT	AANAGARGTG ----CAGGTG	RAV1-B	6.7686e-04
miR166g	miR166g-M1	For: TAATTTTGGT	ACCWAAAWTA- -CCAAAAATGG	AG_v3	1.6906e-07
miR166g	miR166g-M3	Rev: AATATCTACA	TGTATATATA TGTATATAT	SORLREP3	2.3759e-14
miR166g	miR166g-M4	Rev: TTGAATGCAA	KTACATTSAA -TRCATTTA	L1-box	3.6862e-06

Rev, reverse strand; For, forward strand. [1]STAMP ID is the identifier of STAMP best consensus sequence matching motif. [2]STAMP E-value represents expectation value of the STAMP alignment.

Table 3-7 motifs identified in microRNA167 family

microRNA ID	Motif ID	Motif Sequences detected by MEME	STAMP alignment	STAMP ID[1]	STAMP E-value[2]
miR167a	miR167a-M1	For: AGAGAAAGAG	TTCTTTCTNT ---TGTCTC	ARF	3.0543e-03
miR167a	miR167a-M2	Rev: GTGAAGAAGA	YCTYATTCAC- -TTCATTGACG	JASE1	8.9697e-05
miR167a	miR167a-M3	For: GGGTCGAAAG	KGGTCCAMAG- -----CAAAGT	T-box	1.0941e-03
miR167a	miR167a-M4	Rev: AGGCGTTAAA	TGTCACKCCT TCTCCCGCC	E2F-varient	2.0332e-04
miR167a	miR167a-M5	For: GACAGATTAG	----CTAATCTGTC-- TTTCCTATTCTGTTTT	AG_v4	5.8683e-07
miR167a	miR167a-M6	For: TCTGTAACTC	---GAGTTACAGA AACSAGTTA----	MYB2	7.4366e-04
miR167b	miR167b-M1	For: GAGATGAGAG	---SAGNTNAGAG TGGTAGGTTAGA	MRE	5.8024e-05
miR167b	miR167b-M2	For: TTTAAGAGGC	CTTKTTMCCC ----TTAACC	Box	5.5697e-03
miR167b	miR167b-M3	For: GAAAGAGATG	--GCYTCTTTAA TAGATTGTTT--	CCA1_v2	7.8037e-03
miR167b	miR167b-M4	For: CACGTACACA	GAAASAGAKG GAGACA----	ARF	4.4014e-03
miR167b	miR167b-M5	For: CTTGACCGTG	TGTGGACTTG- ---TGACGTGG	TGA1	7.8650e-03
miR167b	miR167b-M6	Rev: GTTTGAATGC	MACGGTCAAG ----GTCAA-	W-box	6.4518e-06
miR167b	miR167b-M7	Rev: GTTTGAATGC	------------------GTTTGAATGC GGATTCAAGATACATGCCCCCTTGAATCC	EIN3	1.1066e-04

Rev, reverse strand; For, forward strand. [1]STAMP ID is the identifier of STAMP best consensus sequence matching motif. [2]STAMP E-value represents expectation value of the STAMP alignment.

Table 3-8 motifs identified in microRNA171 family

microRNA ID	Motif ID	Motif Sequences detected by MEME	STAMP alignment	STAMP ID[1]	STAMP E-value[2]
miR171a	miR171a-M1	Rev: TCTCTTCTCT	AGARRWGAGA ----ATGAGT	ATB2_AtbZIP53_AtbZIP44_GBF5	2.1477e-03
miR171a	miR171a-M2	For: CTTTCTTTCT	CTTTMYYTCT -TTTCCCGC	E2F	7.2077e-03
miR171b	miR171b-M1	For: ATTAAATTTG	--ATTAAATTWG CCATTAAATTGG	AGL15	1.8886e-10
miR171b	miR171b-M2	For: AATTGTTTCT	--AATTGTTTCT TAGATTGTTT--	CCA1_v2	5.7146e-07
miR171c	miR171c-M1	For: CCGTTAAACC	-----------GGTTTARCKG ATTACCCCTAGGGTTTAAG--	LFY	5.0869e-04
miR171c	miR171c-M2	Rev: TGTTTTTTCT	AAAAAAAACA----- AAAAAAAATCTATGA	CCA1_v3	4.3104e-07
miR171c	miR171c-M3	For: CCAACAACCC	GGGAWGTTGG ----TGTTG	RAV1-A	8.1864e-05
miR171c	miR171c-M4	Rev: TTGTGATGGA	TTGTGANGNA- -AGTCATGCAA	SORLREP5	6.0612e-05

Rev, reverse strand; For, forward strand. [1]STAMP ID is the identifier of STAMP best consensus sequence matching motif. [2]STAMP E-value represents expectation value of the STAMP alignment.

Table 3-9 motifs identified in microRNA319 family

microRNA ID	Motif ID	Motif Sequences detected by MEME	STAMP alignment	STAMP ID[1]	STAMP E-value[2]
miR319a	miR319a-M1	Rev: TTCTTTACTT	TTCTTTWNTT ACGTTTTAT	SORLREP2	1.0560e-03
miR319b	miR319b-M1	For: ATAGCATAAA	CGGTCCACTC ---CTCACTC	SORLIP5	7.8420e-04
miR319b	miR319b-M2	For: ACCCTAGATG	----MATCTAGGGT----- ATTACCCCTAGGGTTTAAG	LFY	1.0279e-04
miR319b	miR319b-M3	For: AAAAAGAGAG	--CTCTCTTCTT---- TTTCCTATTCTGTTTT	AG_v4	5.1349e-03
miR319b	miR319b-M4	Rev: AAGTGGGTGG	CCACCMACTY -MACCWAMC	MYB	9.4115e-05
miR319b	miR319b-M5	For: CTCCACGTTT	YCCACGTTT GCCACGTG	ABFs	9.6741e-06
miR319c	miR319c-M1	Rev: TTTGAAACAA	TTTSAAACTW -TTGAC----	W-box	8.2708e-03
miR319c	miR319c-M2	Rev: ATAACTACAT	ATGYACTTAT -TRCATTTA	L1-box	3.1684e-06

Rev, reverse strand; For, forward strand. [1]STAMP ID is the identifier of STAMP best consensus sequence matching motif. [2]STAMP E-value represents expectation value of the STAMP alignment.

Table 3-10 motifs identified in microRNA395 family

microRNA ID	Motif ID	Motif Sequences detected by MEME	STAMP alignment	STAMP ID[1]	STAMP E-value[2]
miR395a	miR395a-M1	For: GGTTAAAGGA	TCCTYTRAGC ----TTAACC	Box	1.2548e-03
miR395a	miR395a-M2	Rev: GAGATTCTGT	---GAGATTCTGT--- TTTCCTATTCTGTTTT	AG_v4	4.0665e-05
miR395b	miR395b-M1	Rev: TGAATCGAGT	------AYTCGAWTCA TAATTGACTCAATTA	PRHA	2.2364e-03
miR395b	miR395b-M2	Rev: TTTTGAAACG	-CGTTTCAAAA ACGTTTTAT--	SORLREP2	4.3345e-04
miR395c	miR395c-M1	For: TGAAAGTTGA	-------TGAAAGATGA TAATCCATGAAAGGTAAG	CArG2	8.1690e-06
miR395c	miR395c-M2	For: ATTTGACCGA	TCGGTCAAAT ---GTCAA--	W-box	5.8084e-06
miR395c	miR395c-M3	Rev: TGAAACGACA	-------WGAAAAGWMA TAATCCATGAAAGGTAAG	CArG2	1.6734e-04
miR395d	miR395d-M1	Rev: GTGAACTTGG	GTAAACTTGG CTAAAAATGG	AP1	9.9561e-05
miR395d	miR395d-M2	For: GGGTGGTCGA	-TCGACCACCC GTCGGCCA---	CBF1	2.3909e-04
miR395d	miR395d-M3	For: TTCTTGTGAG	CTCACWATAA CTCACTC---	SORLIP5	1.7918e-05
miR395e	miR395e-M1	For: TAGGGTTTGA	TCAAACCCTA- --AAACCCTAA	TELO-box	2.8384e-11
miR395e	miR395e-M2	For: ACTTAGACGT	ACYTAGACGT TCTATGACGT	LS5	1.3607e-04
miR395e	miR395e-M3	For: AGGTTAAGGG	CCYTTMASCT ---TTAACC	Box	4.2094e-06
miR395e	miR395e-M4	Rev: GGCTCCTTAT	-ANAAGRAGCA AATTAGGAG--	SORLREP4	6.1630e-04
miR395f	miR395f-M1	Rev: GCTTCTTCTC	GWGAAGAAGS GTGACGTAGA	LS7	3.3551e-04
miR395f	miR395f-M2	For: AACATGCGGG	CCCGCATGTT ----CATGTG	AtMYC2	1.0424e-04
miR395f	miR395f-M3	For: AGGAGCCAAG	CTTGGYTCMT --TGGTTAG	AtMYB2	2.5795e-03
miR395f	miR395f-M4	Rev: GATCTTATTC	GATCTTATTC ---CTTATC	Ibox	6.7282e-05

Rev, reverse strand; For, forward strand. [1]STAMP ID is the identifier of STAMP best consensus sequence matching motif. [2]STAMP E-value represents expectation value of the STAMP alignment.

Table 3-11 motifs identified in microRNA408 family

microRNA ID	Motif ID	Motif Sequences detected by MEME	STAMP alignment	STAMP ID[1]	STAMP E-value[2]
miR408	miR408-M1	For: ACGTGGTAGG	ASCTGGTRGG----- ---TGGTAGGTTAGA	MRE	2.9762e-04
miR408	miR408-M2	For: CAAGCCACCA	--CACSNTMCCA TCTAACCTACCA	MRE	4.9701e-05
miR408	miR408-M3	For: CCGAAACGTG	CCCGGTWNGG- ---GGTWGGAK	MYB1	4.9186e-03

Rev, reverse strand; For, forward strand. [1]STAMP ID is the identifier of STAMP best consensus sequence matching motif. [2]STAMP E-value represents expectation value of the STAMP alignment.

3.2.4 Identification of putative miRNA390a and miRNA390b *cis*-regulatory elements

3.2.4.1 Highly conserved motifs and one *IDE* of miRNA390a promoter region identified throughout nine species

Among the conserved microRNAs identified by computational analyses, motifs present in single or multiple microRNA genes were predicted.

Given the fact that miR390 was the well conserved yet less studied one, we chose it for deep mining the microRNA regulatory network, and performed subsequent functional analysis of the predicted motifs.

By aligning the similar motif sequences detected by MEME and PlantCARE, and comparing the motif similarities by TOMTOM and STAMP, we identified 5 highly conserved motifs in miR390a promoter region within 1500bp upstream. An additional Iron-deficiency responsive *cis*-element (IDE) in *Arabidopsis* annotated by previous study, was also studied in our alignment [4].

Figure 3–6 was the visualization of motif sequences alignment in miR390a 800bp upstream promoter region, it was generated by Bioedit Sequence Alignment Editor in FASTA format. miR390a was detected in several species, as well as the identification of its *cis*-regulatory elements. As shown in the schematic alignment, the first 2 motifs identified were the putative TATA-box 5'-**TATAAATA**-3' and TSS 5'- **ACACGTCAT**-3', both were fully conserved throughout all the nine species *Arabidopsis.thaliana* (*Ath*), *Sisymbrium officinale* (*Sof*), *Neslia paniculata* (*Npa*), *Matthiola valesiaca* (*Mva*), *Arabis alpine* (*Aal*), *Neslia paniculata* (*Npr*), *Thlaspi arvense* (*Tar*), *Descurainia sophia* (*Dso*), *Aethionema grandiflora* (*Agr*). The phylogenetic relationship of them was indicated in Figure 3–3B, which was reconstructed by MEGA based on *Brassicaceae* phylogeny [243, 244].

The third motif identified was the putative E-box 5'-**CCAGATG**TGA-3', consensus with the core structure of **CANNTG** (N can be any nucleotide) that reported to be the Fe-deficiency Induced Transcription Factor 1 (*FIT1*)(At2g28160) binding site [245]. One microarray study predicted a putative basic helix-loop-helix (bHLH) transcription factor, which could bind to hexanucleotide E-box site CANNTG, and involved in iron deficiency response in *Arabidopsis* [246]. bHLH transcription factors with around 147 members [247], are one of the three largest families of transcription factors in *Arabiopsis*, and represents ~9% of the total. The other two are AP2/EREBP (APETALA2/ethylene responsive element binding protein) and MYB-(R1) R2R3. bHLHs play crucial roles in plant phenotypic diversity and in the transcriptional regulation of key growth and developmental processes [248].

The core DNA sequence recognized by bHLHs is a consensus hexanucleotide sequence known as the E-box (5'-CANNTG-3') [247], within which and the best understood one is G-box (5'-CACGTG-3'), there are also non-G-box core motifs [247]. In the alignment of our DNA sequencing dataset of the promoter regions of different species, the identified putative E-box (5'- **CCAGATG**TGA -3') was well conserved across *Sof*, *Npa*, *Mva*, *Aal*, *Npr*, *Tar*. Among them, *Neslia paniculata* (*Npr*) was the closest species with respect to *A. thaliana* in phylogeny [245], and it contained a fully consensus putative E-box as in *A. thaliana*. While in the secondly close species *Descurainia Sophia* (*Dso*), the core motif sequence occurred to be 5'-N**NAGANG**TG-3'; in the phylogenetically distant species *Aethionema grandiflora* (*Agr*), the sequence turned to be 5'-T**CACCTA**TGA-3'.

Another two motif candidates identified were 3'-T**GGATCC**C-5' and 5'-GTT**TGACT**TT-3', respectively. Both were fully conserved through *Sof*, *Npa*, *Mva*, *Aal*. The *Arabidopsis* close species *Npr* encompassed a fully consensus sequence of

65

3'-T**GGATCC**C-5' as well, yet this motif turned to be 5'- T**AGATTC**T -3' in the secondly close species *Dso*. By searching against the AGRIS database, this sequence was predicted by STAMP program to be a putative EIN3 motif. Previous studies on *Arabidopsis* suggested that EIN3 (ETHYLENE INSENSITIVE) element was the specific binding site of EIN3 transcription factor, which was ethylene-activated and involved in salt tolerance [249, 250]. Recently, it was reported that the class of EIN3 elements played essential role in regulating ethylene-activated *PIF3* gene expression.

As regards to the structure of the other motif 5'-GTT**TGACT**TT-3', although it slightly changed into 5'-GTT**TGACA**TT-3' in *Arabidopsis* close species *Neslia paniculata* (*Npr*), and was absent from phylogenetically distant species *Aethionema grandiflora* (*Agr*) with respect to *Arabidopsis*. The core sequence of this motif was completely complementary symmetrical. In general, the transcription factor of a reverse complementary symmetry binding element is speculated to be a homodimer, and the mechanism of the binding action between the TF and its binding site is possibly special [251]. In this study, this sequence was annotated as putative W-box due to the consensus **TGAC** core sequence [154]. The W-box (C/T) TGAC(C/T) is recognized by WRKY transcription factor, and plays important role in the plant development and defense responses [252-254].

Moreover, there was an iron-deficiency responsive elements (IDE) 3'-**ATTAAGCTTCCTTCATTTC**-5'reported in *Arabidopsis* [4], with a core sequence 5'-**GAANNGAAGNANGCTTNAT**-3'. As shown in Figure 3–6 and Table 3-12, this motif was varied within different species in the alignment. A complete 5'-**CCTTCATTTC**-3' was shared by phylogenetically close species *Sof*, *Mva*, *Aal*, and *Tar*, while the sequences of other species were slightly differed at some nucleotides. While surprisingly, the closest species *Neslia paniculata* (*Npr*)

was not among the consistent ones. This relatively low conservation also can be seen from its motif logo in Table 3-12.

3.2.4.2 Three highly conserved motifs of miRNA390b promoter region identified throughout eight species

With the same pipeline applied for mining motifs of miRNA390a and other miRNAs, we analyzed miRNA390b promoter region up to 1500bp upstream with respect to the microRNA flanking region. There were 5 putative motifs identified within 300bp upstream of microRNA. The results of alignment performed by Bioedit program was shown in Figure 3–6, and the logos and conservation through eight species are shown Table 3-13.

The first identified motif sequence 3'- **ATGCATCTTC**-5' was a putative TSS, with an *E*-value at 3.6074e-06. With slight changes in *Neslia paniculata* (*Npa*) by 5'-**ATGCATCTGC**-3', in *Matthiola valesiaca* (*Mva*) by 5'-**TTGCAGCTTC**-3' and in *Noccea praecox* (*Npr*) by 5'-**ATGCTTCTTC**-3', this motif was well conserved in other 5 species.

The second motif 5'-TTAT**AGTCTC**- 3' proximal to the TSS was predicted to be a putative ARF binding site motif. Its core sequence **AGTCTC** was different from the Auxin response factor (ARF) core binding motif **TGTCTC** by one nucleoside [161].

The third motif 3'-**AATGCA**TT-5' was detected to be RY-repeat (**CATGCA**), a *cis*-element in response to ABA-induction, and required for seed-specific expression [255]. In the multiple alignment, partially absent in sp04, this motif is fully conserved through other species.

67

The forth motif identified in miR390b promoter region was a putative AuxRE (Auxin responsive elements)-like 5'- AG**TGTCCC**-3'. Its core sequence **TGTCCC** was consistent with the AuxRE core sequence **TGTCNN**. In the alignment with other species, this motif was directly oriented and fully conserved through eight species. It is well known that ARF plays an essential role in plant development via mediating a variety of cellular and developmental processes [256], and miR390-TAS3-ARF2/ARF3/ARF4-Auxin mediate a regulatory pathway in lateral root development, and the ARF4 could in turn negatively regulates the expression level of miRNA390a [3]. Hence, we were expecting to identify several putative ARF binding sites that determine the feedback of ARFs in both paralogs. As expected, together with the other putative ARF binding motif, the identification of the putative AuxRE-like 5'-AG**TGTCCC**-3' in miR390b promoter region, provided a relatively possible clue to uncover the Auxin responsive mystery of miR390b.

Another predicted motif sequence 5'-CCAT**ACCCAC**-3' was a putative DRE 3'-AT**GTCGGT**A-5', and almost fully conserved in all the eight species. The DRE (dehydration response element) with a core sequence **ACCGAC** was implied to play crucial role in regulation gene expression in response to several environmental stresses [257].

However, aligned with miR390a promoter sequences, there were none common motifs detected, which indicating the two paralogs possibly play separated roles in plant development.

3.2.5 Phylogenetic analysis based on miR390 promoter regions

MEGA is a powerful program for estimating evolutionary distances and inferring phylogenetic relationships of DNA and protein evolution. The evolutionary distances are computed using the Maximum Composite Likelihood (MCL) method for estimating distances between sequence pairs [258].

Since some regulatory elements in the miR390 promoter regions were highly conserved across analyzed species in *Brassicaceae*, whether there was correlation between the phylogenetic distance of promoter sequences and the real phylogenetic relationships of the tested species within *Brassicaceae* drew our attention.

As Figure 3–3 showed, the phylogenetic tree of *Brassicaceae* miR390a promoter sequences was reconstructed by MEGA7 using the Neighbor-Joining method (Figure 3–3 A), and the known *Brassicaceae* phylogeny model reconstructed using nuclear markers was redrawn based on other studies (Figure 3–3 B) [243, 244]. Although *Aethionema grandiflora* in both models was the common ancestor of other branch taxon in *Brassicaceae*, the two models showed significant differences as compared to other species. Take the *Noccea praecox* for example, this species exhibited parallel evolution with *Descurainia Sophia* in *Brassicaceae* phylogeny model. While in the phylogenetic tree of miR390a promoter sequences, the *Noccea praecox* was distant from *Descurainia sophia* and diverged earlier than the latter in the evolutionary history, which was conflict with the referred model. In addition, in *Brassicaceae* phylogeny model, *Arabis alpina* split after *Thlaspi arvense*and and *Noccea praecox*, while in the phylogenetic tree of promoters, *Arabis alpina* and *Noccea praecox* turned to be closely related and split after *Thlaspi arvense*.

Similar cases occurred to miR390b. As Figure 3–4 showed, although the two reconstructed phylogenetic trees of miR390b indicated similar evolutionary events in *Arabidopsis halleri* and *Noccea praecox*, and both rooted from *Matthiola valesiaca*, other branches of the two topologies varied apparently. For instance, the *Berteroa incana* diverged from *Matthiola valesiaca* accompanying with other parallel phylogenetic divergence (Figure 3–4 B). While in the phylogenetic tree of

miR390b homologs (Figure 3–4 A), it was indicated that the duplication of all other species besides *Matthiola valesiaca* were occurred after *Berteroa incana*.

Comparing the two phylogenetic trees reconstructed from the promoter regulatory regions of miR390a and miR390b, the phylogenetic relationship of *Noccea praecox* and *Descurainia sophia* within the two paralogs were inconsistent, which confirmed the low resolution of the promoter regions for tracing the evolutionary history of numerous species. What's more, the bootstrap values of most nodes were lower than 70, indicating a low reliability of the phylogenetic relationships [259].

Taken the four topologies together, we concluded that the miR390 promoter sequences could provide overall lower resolution for tracing the phylogeny of tested *Brassicaceae* species. A major reason of this low resolution could be addressed to the miR390 gene duplication events in some species in the evolutionary history, and resulting in hidden paralogs, which could resolve a relatively low reliable phylogenetic tree [243]. Generally, low or single-copy nuclear genes were supposed to be effective phylogenetic markers [260, 261]. In addition, the core promoter regions of conserved plant miRNAs were relatively short (~1kb) with respect to protein-coding genes [262], which were insufficient to infer the evolutionary relationship of their host species. Therefore, to estimate the phylogeny of all tested *Brassicaceae* species, the sequenced whole-genome will be necessary, yet currently this information is still not fully available [243].
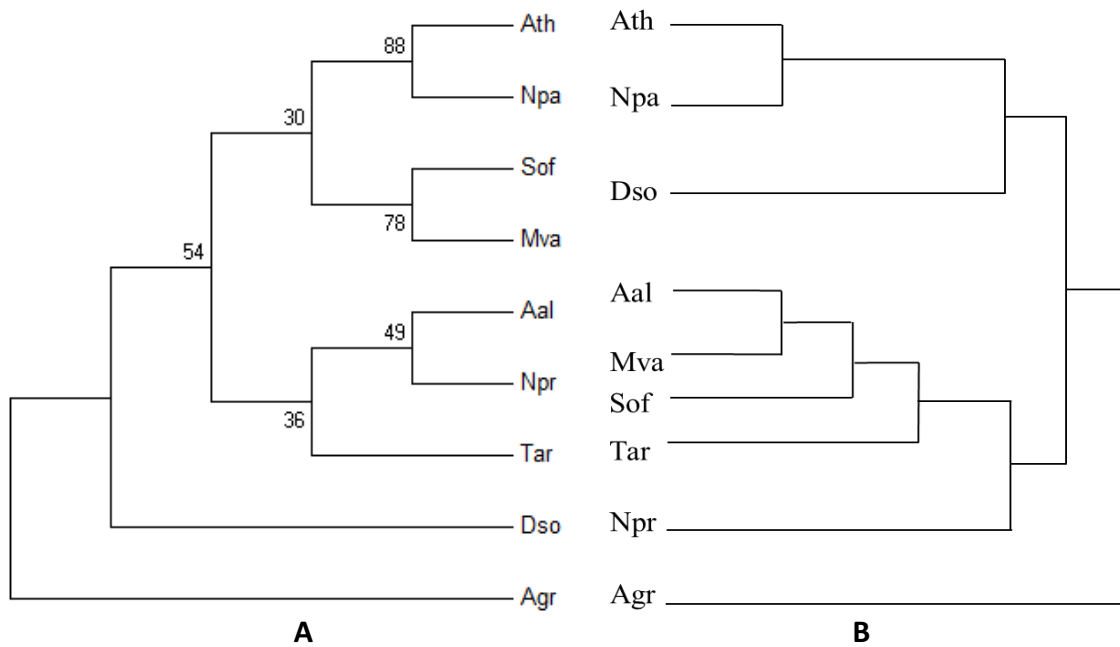
Figure 3–3 Phylogenetic reconstruction of miR390a in *Brassicaceae*. Phylogenetic reconstruction of miR390a in *Brassicaceae* family by MEGA was compared with a phylogenetic tree drawn by concatenating the known *Brassicaceae* phylogeny model. A) MEGA phylogenetic tree of miR390a promoter sequences; B) A summarized phylogenetic tree of 9 species in *Brassicaceae*
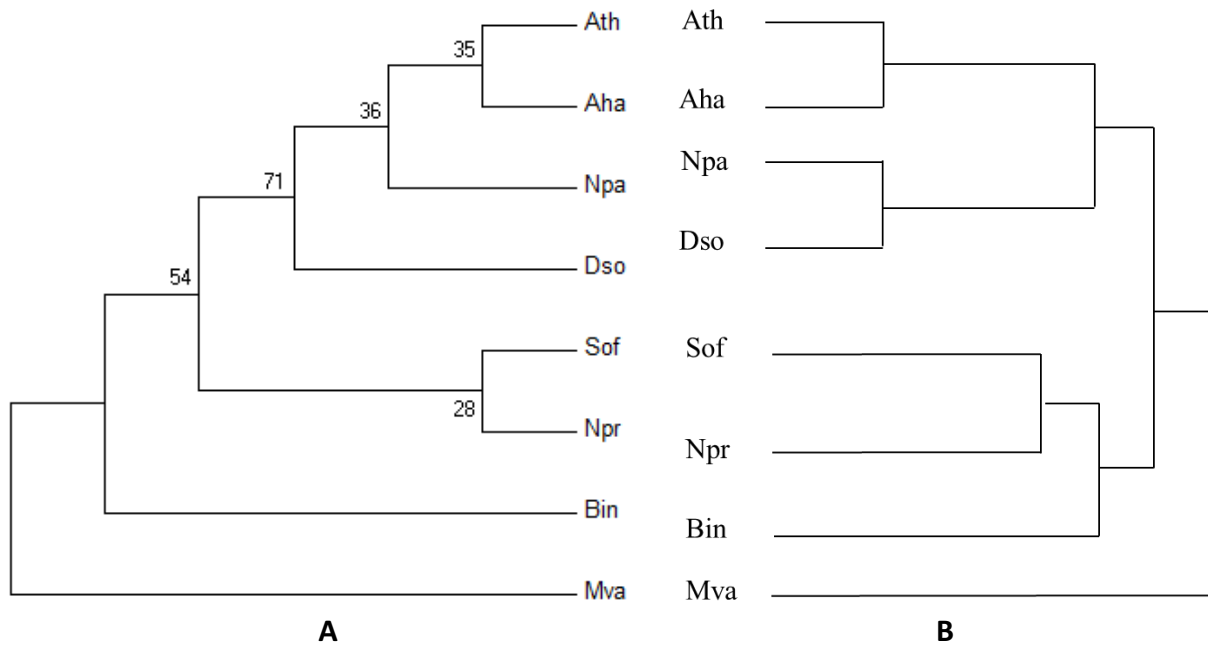
Figure 3–4 Phylogenetic reconstruction of miR390b in *Brassicaceae*. Phylogenetic reconstruction of miR390b promoters in *Brassicaceae* family by MEGA was compared with a phylogenetic tree drawn by concatenating the known *Brassicaceae* phylogeny model. A) MEGA phylogenetic tree of miR390b promoter sequences; B) A summarized phylogenetic tree of 8 species in *Brassicaceae*

71

3.3 Discussion

Multiple putative motifs in a series of conserved plant microRNAs were predicted by computational mining approach [148, 263, 264]. The motifs presented in single or multiple genes, are possible evidences of individual loci' origin that they derived from the same miRNA family, or the miRNA families they belong to are evolutionarily close. This suggests a complex mechanism of the miRNA gene regulation in the evolutionary term. To discover the specific roles of these motifs, we need further to carry out functional analysis by experiments.

The results of specific identification of miR390 motifs showed that there were one more putative motif detected in miR390a than in miR390b, and the whole alignment of miR390a in multiple species was more conserved as well, indicating a major role of miR390a in plant development, which is partially consistent with one observation in root development [3]. As expected, there were two expected putative AuxRE (Auxin responsive elements)-like core TGTCNN identified in miRNA390b promoter region, one contained core sequence TAGTCTC, and the other was TGTCCC. In miRNA390a regulatory region, a similar AuxRE-like core TGTCNN located at -609 ~ -604 (data not shown), while it was not conserved through multiple species, thereby the AuxRE-like hypothesis of miRNA390 was negated. However, the auxin signaling pathways in plants involve many transcription factors and the co-regulators [265]. Bioinformatics studies of *Arabidopsis* and rice suggested bZIP- and MYB-related binding sites as potential AuxRE coupling elements [266]. For instance, a BZR1 and PIF transcription factors binding motif G-box (CACGTG) that belongs to bZIP-related group, was shown to heterodimerize with ARF6 [267, 268]. Which means, in this study, despite the negative identification of AuxRE in miRNA390a promoter region, there could be other possible AuxRE-related motifs unidentified.

In addition, in the overall alignment of miRNA390a and miRNA390b promoter sequences, there were no common motifs identified, indicating they might diverge during the evolutionary history, and gradually play different roles in regulating the gene expression in plant development. Moreover, despite the limited reliability when referred to the phylogeny of the whole *Brassicaceae* family model, the divergent phylogenetic trees of the two miR390 paralogs still could be the part of the evidences for elucidating the possible divergences in the evolutionary history.

However, the identification and conservative knowledge about the c*REs* of both miR390a and miR390b in our study could only partially address the conservation of the different regulatory sequences within *Brassicaceae*. Given the possible duplication events in both microRNA genes, the less conserved c*REs* in tested *Brassicaceae* species could also be highly conserved and play essential roles in phylogenetically close families, which need to be better studied by both computational and experimental techniques in the future.

Table 3-12 Summary of motifs detected in miRNA390a promoter

Motif ID	Motif sequence	Species	Motif conservation in dataset	Position	Orientation	Stamp *E*-value	Logo by STAMP
MIR390a Putative TATA	TATTTATA	*Ath* *Sof* *Npa* *Mva* *Aal* *Npr* *Tar* *Dso* *Agr*	TATTTATA TATTTATA TATTTATA TATTTATA TATTTATA TATTTATA TATTTATA TATTTATA TATTTATA	-126, -119	3'-5'	4.1045e-05	
MIR390a Putative TSS	ACACGTC	*Ath* *Sof* *Npa* *Mva* *Aal* *Npr* *Tar* *Dso* *Agr*	ACACGTC ACACGTC ACACGTC ACACGTC ACACGTC ACACGTC ACACGTC ACACGTC ACACGTC	-161, -155	3'-5'	7.2698e-05	
MIR390a Putative iron-responsive E-box	CCAGATGTGA	*Ath* *Sof* *Npa* *Mva* *Aal* *Npr* *Tar* *Dso* *Agr*	CCAGATGTGA CCAGATGTGA TCAGATGTGA CCAGATGTGA CCAGATGTGA CCAGATGTGA CCAGATGTGA TAAGANGTGA TCACCTATGA	-324, -310	3'-5'	7.2698e-05	
MIR390a Putative EIN3 motif	TGGATCCC	*Ath* *Sof* *Npa* *Mva* *Aal* *Npr* *Tar* *Dso*	TGGATCCC TGGATCCC TGGATCCC TGGATCCC TGGATCCC TGGATCCC TGGATCCA TAGATTCT	-337, -327	5'-3'	6.1244e-04	
MIR390a Putative W-box	GTTTGACTTT	*Ath* *Sof* *Npa* *Mva* *Aal* *Npr* *Tar* *Dso*	GTTTGACTTT GTTTGACTTC GTTTGACTTT GTTTGACTTT GTTTGACTTT GTTTGACATT GTTTGACTTT NNTGAATNC	-351, -340	5'-3'	5.8084e-06	
MIR390a IDE1	GAAATGAAGGAAGCTTAAT	*Ath* *Sof* *Npa* *Mva* *Aal* *Npr* *Tar*	GAAATGAAGGAAGCTTAAT GAAATGAAGGAGACTTTCA AAAATAAAGGTTACATTAT GAAATGAAGGAGACTTCAT GAAATGAAGGAAACGTCTA GAAAAGAATATATTACTCT GAAATGAAGGAGGCTTAAT	-402, -382	5'-3'	1.8853e-03	

The positions were counted with respect to miRNA flanking sequence. N, nucleotides missing or not consensus.

Figure 3–6 Alignment of the motif regions in miRNA390b promoter
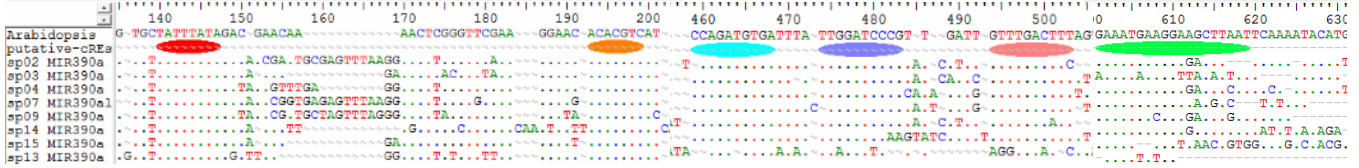
Different colors represent different motifs: red, putative TATA-box; yellow, putative TSS; blue, putative iron-related E-box; purple, unidentified motif one; pink, unidentified motif two; green, IDE. The position of each motif is only the aligned schematic one. Due to the space reason, the alignment is organized by overlapping redundant sequences, and only the motif sequences are shown.

Figure 3–6 Alignment of the motif regions in miRNA390b promoter

Different colors represent different motifs: yellow, putative TSS; pink, putative ARF binding motif; purple, AuxRE-like motif; blue, putative RY-repeat; brown, putative DRE. The position of each motif is only the aligned schematic one. Due to the space reason, the alignment is organized by overlapping redundant sequences, and only show the motif sequences.

75

Table 3-13 Summary of motifs detected in miRNA390a promoter

Motif ID	Motif sequence	Species	Motif conservation in dataset	Position	Orientation	Stamp *E*-value	Logo by STAMP
MIR390b Putative TSS	ATGCATCTTC	*Ath* *Bin* *Sof* *Npa* *Npr* *Aha* *Dso* *Mva*	ATGCATCTTC ATGCATCTTC ATGCATCTTC ATGCATCTGC ATGCTTCTTC ATGCATCTTC ATGCATCTTC TTGCAGCTTC	-56, -47	5'-3'	3.6074e-06	
MIR390b Putative ARF binding motif	TTATAGTCTC	*Ath* *Bin* *Sof* *Npa* *Npr* *Aha* *Dso* *Mva*	TTATAGTCTC ATATAGTCCC TTATAGTCTC TTATAGTCAG TTATAGTCTC TTATAGTCTC TTATAGTCTC TTATAGTCTC	-169, -163	3'-5'	6.6516e-04	
MIR390b Putative AuxRE-like motif	AGTGTCCC	*Ath* *Bin* *Sof* *Npa* *Npr* *Aha* *Dso* *Mva*	AGTGTCCC AGTNTACA AGTGTCCC AGTGTCCC AGTGTCCC AGTGTCCC AGTGTCTT AGTGTCCC	-176, -171	3'-5'	9.7816e-05	
MIR390b Putative RY-repeat	AATGCATT	*Ath* *Bin* *Sof* *Npa* *Npr* *Aha* *Dso* *Mva*	AATGCATT AATGCATT AATGCATT AATGCATT AATGCATT AATGCATT AATGCATT AATGCNNN	-253, -243	3'-5'	1.1575e-06	
MIR390b DRE	CCATACCCAC	*Ath* *Bin* *Sof* *Npa* *Npr* *Aha* *Dso*	CCATACCCAC CCATACCCAC CCATACCCAC CCATACCCAC CCATACCCAC CCATACCCAC CCATACCCAT	-263, -253	5'-3'	4.2041e-04	

The positions were counted with respect to miRNA flanking sequence. N, nucleotides missing or not consensus.

Chapter 4.   GUS-fused constructs of WT miRNA390 promoters and WT miRNA390a promoter undergone site-directed mutagenesis

4.1   Methods

4.1.1 Plant materials

Wild-type (WT) *Arabidopsis thaliana* (*A. thaliana*) ecotype Col-0 (Columbia-0) was used for constructs of miR390a and miR390b promoters fused with GUS reporter gene.

WT *A. thaliana* ecotypes Col-0 (Columbia-0) were also used for the deletion construct where GUS reporter genes were driven by promoters underwent six site specific mutagenesis of c*REs* identified.

4.1.2 Create p-ENTRY clone of miRNA390a and miRNA390b upstream promoter regions

For the p*MIR390a:GUS-GFP* and p*MIR390b:GUS-GFP* constructs, the 3' end of upstream promoter regions of *A. thaliana* miRNA390a (AT2G38325) and miRNA390b (AT5G58465) fused with GUS and GFP reporter genes were amplified (primers MIR390a-AthPT-F/R and MIR390b-AthPT-F/R in Table S0-2) 2.0kb and 1.9kb respectively, then cloned into the pENTRY/D-TOPO vector (Invitrogen) to generate the Gateway entry clones. Primers were designed either on conserved regions of upstream and downstream of *A. thaliana* miRNA390a and miRNA390b or on the highly conserved regions of the mature miRNA390 and miRNA390* (Appendix Table S0-2).

The pENTRY clones of p*MIR390a:GUS-GFP* and p*MIR390b:GUS-GFP* were first validated by PCR with high fidelity DNA polymerase. Then the inserts in the plasmids were confirmed by DNA sequencing (96° C, 1 minute; 55 cycles of 96° C 10 seconds, 55° C 5 seconds and 60° C 4 minutes; and 4 minutes at 60° C).

Figure 4–1 Simplified view of the miR390a promoter construct

4.1.3 Performing the LR recombination reaction to generate the destination plasmids

The pENTRY cloning vectors carrying the inserts of genes were introduced into the destination vector pKGWFS7 by LR reaction. Validation was performed with colony PCR with the primers for constructing promoter [269].

Colony PCR was performed with a pair of promoter specific primers for each construct. Positive colonies were picked up and used for inoculation in LB medium.

Three individual colonies were randomly selected from LB (Luria-Bertani) agar (supplemented with spectinomycin) plates of each construct.

4.1.4 Transformation of destination vectors into *A. thaliana*

The surface-sterilized *Arabidopsis* seeds were soaked in sterile water and stratified for 2 days at 4°C, then moved into 1/2 inch plastic pots containing potting soil under long-day conditions (day/night cycle 6/8 h) at 22 ℃ in white light of approximately 150 μ-mol $\times$ m^{-2} $\times$ s^{-1}, 70% relative humidity.

For Agrobacterium-mediated transformation of *A. thaliana*, floral dip method was performed on 5-week-old plants. In this method, transformation was accomplished simply by dipping developing *Arabidopsis* inflorescence for ~15

seconds into a 2.5% sucrose solution containing 0.02% (vol/vol) Silwet L-77 and re-suspended in Agrobacterium culture carrying the genes to be transferred.

144 plants were subjected to the transformation of each construct. In total, there were 1152 *A. thaliana* plants performed for p*MIR390a:GUS*-GFP, p*MIR390b:GUS-GFP*, and six site specific mutagenesis constructs of different c*REs*.

4.1.5 T1 transgenic seedlings screening

2mL T_1 transgenic seeds (~3000) of each construction were surface sterilized.

The seeds were by soaked in 20 mL 70% (v/v) ethanol for 10-15 min in a 50 mL sterile tube (114*28mm, Sarstedt, Germany). Next, the tubes were vortexed on a mixing rotator (Labnet, USA) for 15-20 sec, then centrifuged up to 1200-1300 rpm (Hettich ROTANTA 460R, Sigma-Aldrich, USA) and stopped the centrifuge to let the seeds spin down. The seeds were washed by 20 mL 5% bleach (NaCIO) for 10-15 min, then were centrifuged as before. Afterward, washed the seeds with sterile distilled water twice by hand-shaking and then centrifuged again. Then seeds were pipetted in Petri dishes on Murashige and Skoog (1962) (MS) basal medium (Sigma) supplemented with 10 g/L sucrose, solidified with 0.8% agar (Sigma) (PH=5.7/5.8) and 50μg/mL of kanamycin (selection marker), and then incubated at 4° C in the dark for 2 days to break seed dormancy.

The plates were then transferred to a culture room under long-day conditions (day/night cycle 16/8 h) at 22 ℃ in white light of approximately 150 μ-mol $\times$ m^{-2} $\times$ s^{-1}, 70% relative humidity.

After 7 days, the positive seedlings were selected as "transgene present" and transferred into 1/2 inch plastic pots containing potting soil under long-day conditions (day/night cycle 16/8 h) at 22 ℃ in white light of approximately 150 μ-mol $\times$ m^{-2} $\times$ s^{-1}, 70% relative humidity.

4.1.6 Confirmation of T-DNA integration in transgenic *Arabidopsis*

To confirm the T-DNA insertion in *Arabidopsis* genome, the young leaves of 2-week-old T1 transgenic plants were used for PCR genotyping assay with CTAB method [270, 271]. Quantification and quality of the DNA was assessed on 1% agarose gels by electrophoresis.

4.1.7 T2 plants segregation analysis

To screen the segregation of T2 transgenic plants, I sowed the seeds onto kanamycin (50 μg/mL) containing MS plates.

Seeds harvested from T2 transgenic plants were sterilized by soaking each 50-60 lines into 70% (v/v) ethanol for 10-15 min in a 1.5-mL microcentrifuge tube (Eppendorf, Germany), followed by 5% bleach (NaCIO) for less than 10 min, both these two steps were vortexed on a mixing rotator (Labnet, USA), and then washed at least twice with 1000mL sterile distilled water by hand-shaking several times. Then seeds were pipetted out on MS plated contained 10 g/L sucrose, solidified with 0.8% agar (Sigma) (PH=5.7/5.8) and 50 μg/mL of kanamycin, then vernalized at 4° C in the dark for 2 days. The plates were then transferred to a growth room under conditions (day/night cycle 8/16 h) at 22 ℃ in white light of approximately 150 μ-mol $\times$ m^{-2} $\times$ s^{-1}, 70% relative humidity.

T2 lines that segregated at a 3:1 on kanamycin containing plates were recorded as "transgenic single copy" and submitted to GUS histochemical assay, and based on the results of GUS staining, further selected the good lines to grow into soil to produce T3 plants.

4.1.8 Histochemical detection and GUS staining on seedlings and inforescence

For histochemical GUS assay, 12 7d-old kanamycin resistant young seedlings of 60 T2 transgenic lines (including "transgenic single copy" and "transgenic multiple copy" that segregated not at 3:1) were subjected to GUS assay.

Seedlings were immersed in GUS staining buffer containing 100 mM sodium phosphate(NaH_2PO_4) buffer, 10 mM EDTA, 0.5 mM potassium ferricyanide (K3[Fe(CN)6]), 0.5 mM potassium ferrocyanide (K4[Fe(CN)6]), and 0.5 mg/ml 5-bromo-4-chloro-3-indolyl-β-glucuronic acid (X-Gluc), pH 7.0. The reactions were incubated at 37°C for 18-24h. Afterwards chlorophylls were removed first by 95% (V/V) ethanol and later by 75% (V/V). Samples were stored in 70% ethanol at 4°C in the dark. Photos were taken with Digital Camera (Germany, Leica DFC490).

For the GUS staining performed on floral organs, inflorescences 4-6 cm in length with flowers partially open were taken from 35 to 45-day-old transgenic *Arabidopsis* plants [272]. When operating, inflorescences were removed by grasping the stem with fine forceps and cutting the stem by a scissors or a fresh razor blade below the forceps [273]. Then ~15 inflorescence of T3 homozygotes from each construct were immersed into petri plates that contained ISF/IDF solutions, and treated with 6h. After the treatment, all inflorescences were soaked in GUS staining buffer and treated with the same procedure as on seedling.

4.1.9 Electrophoresis

1 % w/v of agarose gels were prepared in 1x TAE buffer. Gels were melted in a microwave, allowed to cool to approximately 50°C before 0.1 ^g/mL of ethidium bromide was added and mixed. The molted agarose was immediately poured into a gel tray and allowed to solidify at room temperature for 20-40 minutes. DNA samples were mixed with 1/10 volume of 10 x loading buffer and loaded into gel wells by pipetting. DNA markers were run alongside DNA samples to enable

detection and approximate sizing of fragments. Electrophoresis was performed at 5-10 V/cm in 1x TAE buffer. DNA was visualized on a UV transilluminator (Gel Doc 1000 system with Molecular Analyst version 2.1.1 software, Bio-Rad) and photographed.

A. thaliana were used as the wild type background for the six deletion constructs. All seeds sewn in this study were surface sterilized in 20mL 70% (v/v) ethanol for 10-15 minutes and then shaken in 20mL 5% (v/v) bleach for 10minutes, followed by washing twice in distilled water.

Then seeds were pipetted in Petri dishes on Murashige and Skoog (1962) (MS) basal medium (Sigma) supplemented with 10 g/L sucrose, solidified with 0.8% agar (Sigma) (PH=5.7/5.8) and 0.001% kanamycin (selection plate), and then incubated at 4° C in the dark for 2 days to break seed dormancy. The plates were then transferred to a growth room under conditions (day/night cycle 6/18 h) at 22 ℃ in white light of approximately 120-150μE $\times$ m^{-2} $\times$ s^{-1}, 70% relative humidity.

Seedlings (7 days post germination) were transplanted from MS media plates into soil, and kept under high humidity conditions in green house until booting, prepared to be used for deletion constructs.

4.1.10 Site-directed mutagenesis of identified elements in miRNA390a promoter

Identified elements in miRNA390a promoter were mutated by QuickChange Site-Directed Mutagenesis kit (Strtagene), and the pENTRY vectors containing wild-type promoter sequence of miR390a. The schematic of mutagenized construct was shown in Figure 4–8. Each construct was verified by PCR amplification and amplicons were run on an agarose gel.

4.2 Results

4.2.1 Constructions analysis and PCR confirmation

Colony PCR was carried out to screen the insertion of the recombinant plasmid, and it was performed with a pair of promoter specific primers for each construct. Positive colonies were picked up and used for inoculation in LB medium.

Three individual colonies were randomly selected from LB agar plates of each construct, and the gene insertion was confirmed by the colony PCR as shown in Figure 4–3. Later, we purified plasmid DNA of p*MIR390a:GUS-GFP* and p*MIR390b:GUS-GFP* in pKGWFS7 and performed for PCR, as shown in Figure 4-4.

Followed Agrobacterium-mediated transformation of *A. thaliana*, in order to detect the T-DNA insertion of the *Arabidopsis* transgenic lines, the genotyping PCR was performed. 12 T1 transgenic lines were randomly selected for p*MIR390a:GUS-GFP* and p*MIR390b:GUS-GFP* to extract the genomic DNA, and the results was shown in Figure 4–5. There were clear lanes of T-DNA in both constructs.
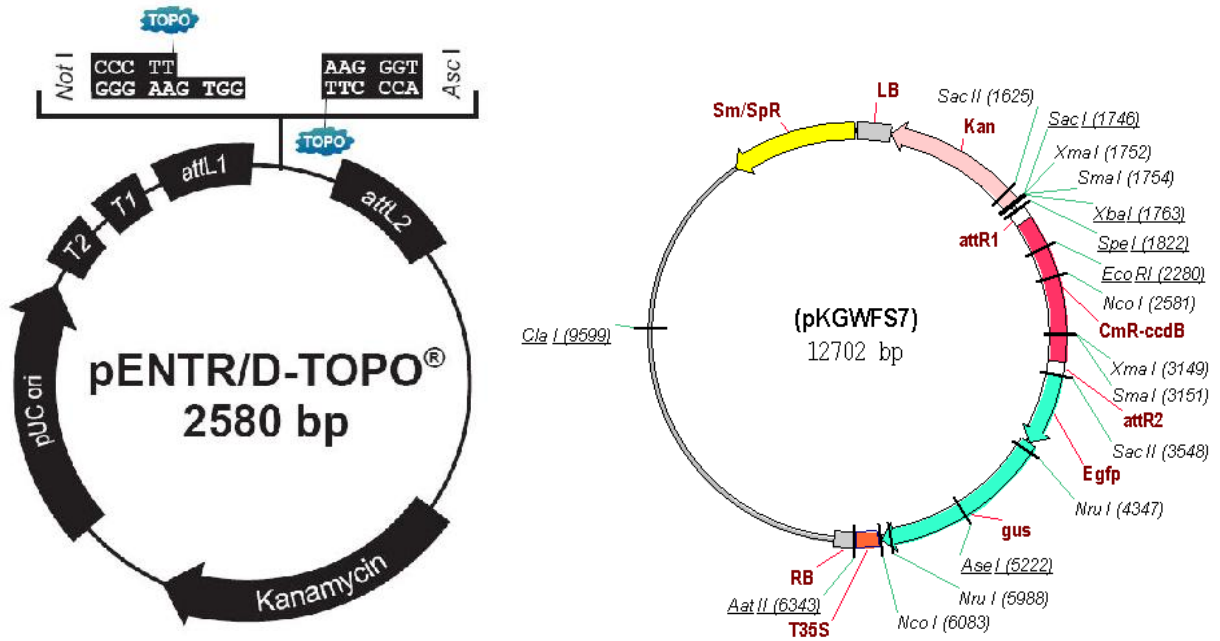
Figure 4-2 Vector Maps of pENTRY and pKGWFS7. (a) pENTR/D-TOPO (2580bp) gateway entry vector used for gateway cloning. (b) pKGWFS7 (12700 bp) binary destination vector used for promoter construction in plants. This vector contains selection markers kanamycin (KnR)
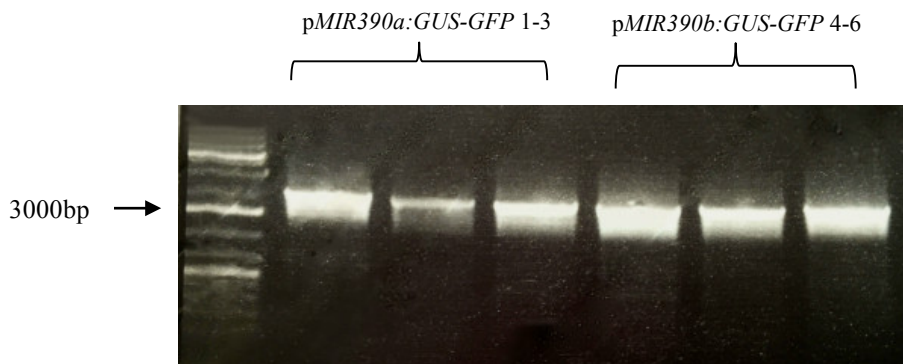
Figure 4–3 p-ENTRY Colony PCR of p*MIR390a:GUS-GFP* and p*MIR390b:GUS-GFP* in Pentry/D-TOPO. From left to right: 1-3, colonies for p*MIR390a:GUS-GFP* replicates; 4-6, colonies for p*MIR390b:GUS-GFP* replicates.
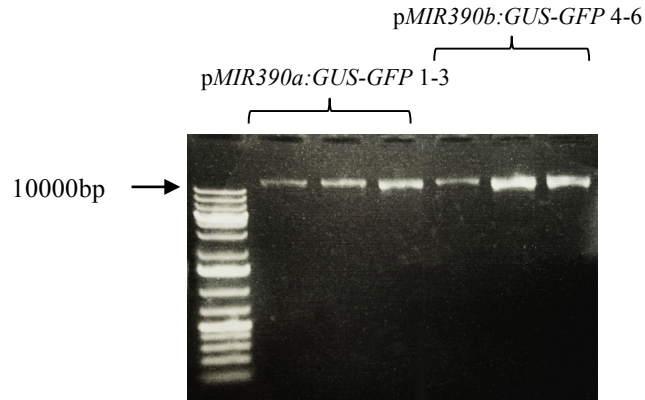
Figure 4-4 Purified plasmid DNA of p*MIR390a:GUS-GFP* and p*MIR390b:GUS-GFP* in pKGWFS7. From left to right: 1-3, plasmids for p*MIR390a:GUS-GFP* 1-3 replicates; 4-6, plasmids for p*MIR390b:GUS-GFP* replicates.
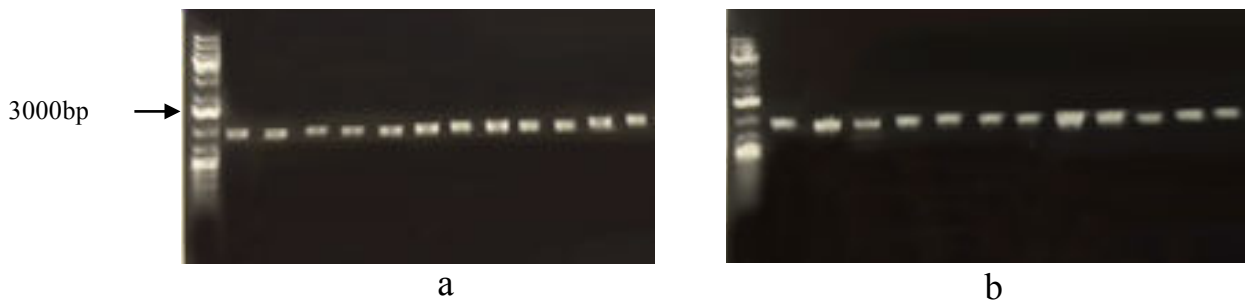
Figure 4–5 Genotyping PCR for confirming p*MIR390a:GUS-GFP* (a) and p*MIR390b:GUS-GFP* (b) T-DNA insertions.
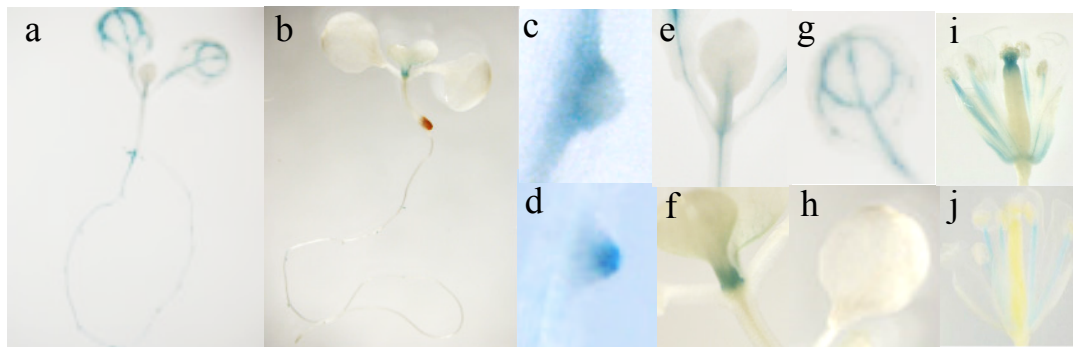
Figure 4–6 GUS staining of transgenic *Arabidopsis* carrying p*MIR390a:GUS-GFP* and p*MIR390b:GUS-GFP*. (a, b)7d-old seedling of pMIR390a (a) and pMIR390b (b). (c, d) GUS staining of p*MIR390a:GUS-GFP* in lateral root primordia but not in lateral root tip(c), GUS staining of p*MIR390b:GUS-GFP* (d). (e, f, g, h) GUS staining of p*MIR390a* in the region below SAM (e) and true leaf(g), GUS staining of p*MIR390b* in SAM (f) but not in true leaf (h). (i, j) GUS staining in 38d-old flowering tissue of p*MIR390a* (i) and p*MIR390b*.

4.2.2 GUS histochemical detection of expression patterns of p*MIR390a:GUS-GFP* and p*MIR390b:GUS-GFP* on 7d-old seedlings

In this study, to confirm the constructs we built was consensus with previous studies, with -2.0Kbp p*MIR390a:GUS-GFP* and -1.9Kbp p*MIR390b:GUS-GFP* reporter fusion constructs, due to inefficiency of the green florescent protein (GFP), we detected the GUS activity of both promoters on 7d-old T2 transgenic *Arabidopsis*. As Figure 4–6 a, c, e, g, i showed, there were clear expression of miRNA390a promoter in the primary root cylinder, lateral root primordia, the aerial parts of the true leaf, cotyledon and some GUS activity below the SAM, as well as in the sepals, petals, filaments and stigma in floral organ.

However, with respect to miRNA390a, as Figure 4–6 b, d, f, j showed that, the expression of miRNA390b promoter only limited to primary root cylinder, lateral root tip, SAM, and the filament on floral organ, which indicated a more restrictive specialization on plant development compared with miRNA390a in plant productive phase.

To summarize, using the two constructs of promoter regions fused with GUS reporter gene of two paralogs miRNA390a and miRNA390b, we established reliable references for further mutational analysis of different highly conserved regulatory elements identified in previous bioinformatics work.
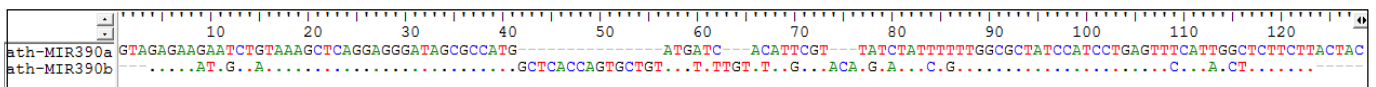
Figure 4–7 Alignments of A. *thaliana* miRNA390a and miRNA390b stem-loop sequences on forward strand.

Table 4-1 GUS staining patterns summary of p*MIR390a:GUS-GFP* and p*MIR390b:GUS-GFP* in 7d-old seedlings and 38d-old inflorescence

Gene	Lateral root primordia	Root cylinder	Lateral root tip	Shoot apex	True leaf	Stigma	Filament	AR	Petal	Sepal
miR390a	√	√	**X**	**X**	√	**X**	√	√	√	√
miR390b	√	√	√	√	√	**X**	√	**X**	**X**	**X**

AR, apical region of pistil.

4.2.3 GUS histochemical detection of expression patterns of miRNA390a and miRNA390b on 38d-old inflorescence

The fact that there were no reports of the role miRNA390a and miRNA390b during *Arabidopsis* inflorescence development drew our attention. To uncover the spatial and temporal expression patterns of miRNA390 promoter in the floral organ, we performed GUS histochemical assay of the inflorescence of 38d-old transgenic plants. As shown in Figure 4–6i, GUS activity of p*MIR390a:GUS-GFP* inflorescence was present in filament, petal, sepal, as well as the apical region of pistil, but not in anther and stigma. On the contrary, weaker GUS activity of p*MIR390b:GUS-GFP* (Figure 4–6j) was only observed in filament, not in other tissues. These differences indicated possible sub-functionalization of the two paralogs in *Arabidopsis* flower development.

4.2.4 Site-directed mutagenesis of identified elements in miRNA390a promoter

With the reliable constructs of WT miRNA390a and miRNA390b promoters fused with GUS reporter gene, to characterize the most relevant and putative regulatory elements, the six elements characterized by data mining of miRNA390a promoter sequence data were mutated by site-directed mutagenesis.

As shown in Table 4-2 and Figure 4–8, the first element identified was the putative TATA box, the mutations were performed using site-directed mutagenesis

kit to mutate TATA box sequence of TATAAATA to **CACAGACG** positioned between -126 and -119 upstream of the mature miR390a.

The second element to be mutated was putative TSS between -161 and -155 upstream of the hairpin structure. The sequence was changed from GACGTGT to **TGGT**G**AA**.

The third mutagenized element was a putative FIT1 binding element located between -320 and -311, and was changed from TCACATCTGG to **G**CA**TCACG**GA.

The forth one was a putative EIN3 element, and the sequence was changed from CGGGATCCAA to **GT**A**GAAG**C**GA**.

The fifth putative W-box element AAAGTCAAAC to be mutated was highly conserved, and present between -350 and -341. It was mutagenized to AAA**TGG**AGAC.

The final muted element mutated was between -402 and -382, an experimentally verified IDE element. To further confirm its response under short period of iron-deficient condition, we mutagenized this IDE from GAAATGAAGGAAGCTTAAT to **TCAAATACATGAT**AA**TG**C.

Sequencing was conducted on vectors on pENTRY clones containing these six mutated miR390a promoter to confirm the correct mutations.

Table 4-2 summary of *cis*-elements sequences performed for site specific mutagenesis

Construct ID	Motif ID	Description	Motif original sequence (5' to 3')	Site Specific mutagenesis sequence
m1	MIR390a-M1	Putative TATA-box	TATAAATA	CACAGACG
m2	MIR390a-M2	Putative TSS	GACGTGT	TGGTGAA
m3	MIR390a-M3	Putative iron responsive E-box	TCACATCTGG	GCATCACGGA
m4	MIR390a-M4	Unidentified M4	CGGGATCCAA	GTAGAAGCGA
m5	MIR390a-M5	Unidentified M5	AAAGTCAAAC	AAATGGAGAC
m6	MIR390a-M6	IDE	GAAATGAAGGAAGCTTAAT	TCAAAATACATGATAATGC
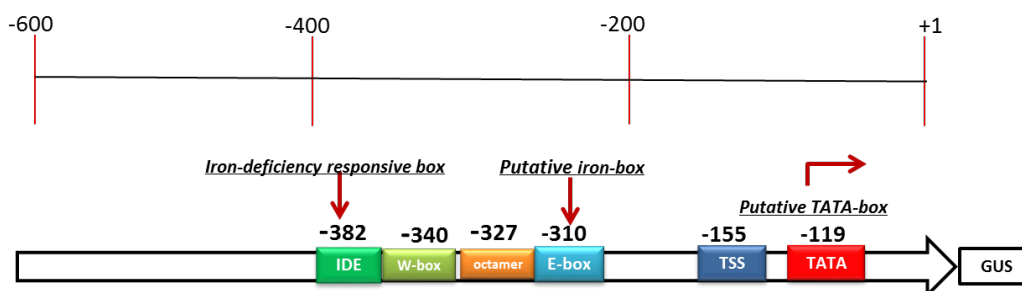
Figure 4–8 Simplified view of miR390a promoter site-mutagenized constructs. Boxes highlighted in different colors represents different motifs identified.

4.2.5 Establishment of transgenic lines

To establish transgenic *Arabidopsis* lines of p*MIR390a:GUS-GFP*, p*MIR390b:GUS-GFP* and mutation constructs, T2 kanamycin-resistant plants segregated at a 3:1 ratio represented a single-locus insertion were identified as candidates for producing later generations. 96 T2 lines per construct were grown for T3 progeny. Among them, at least 60 lines (12 seedlings per line) were performed for GUS staining. The T2 lines segregated at a 3:1 ratio and expressed a clear and main pattern (the expression pattern shared by most of the lines, statistics were not shown here) were selected for obtaining T3 seeds. The kanamycin-resistant T3 seeds population were selected and stored as homozygous lines for future use as shown in Table 4-3.

Table 4-3 Summary of *A. thaliana* transgenic plants of different constructs screening

Item	p*MIR390a:GUS-GFP* (Nr.[1])	p*MIR390b:GUS-GFP*(Nr.)	*m1*[2] (Nr.)	*m2*(Nr.)	*m3*(Nr.)	*m4*(Nr.)	*m5*(Nr.)	*m6*(Nr.)
T0 Transformation	144	144	144	144	144	144	144	144
T1 seedlings positive ratio	120/3000	160/3000	150/3000	150/3000	200/3000	150/3000	200/3000	150/3000
T2 seedlings screened	96	96	96	96	96	96	96	96
T2 seedlings performed for GUS staining	60*12	60*12	60*12	60*12	60*12	60*12	60*12	60*12
T3 seeds	~20000	~15000	~25000	~25000	~30000	~20000	~30000	~30000

[1]Nr, number. [2]*m1-m6* represent the 6 site-mutagenized constructs performed on WT miR390a promoter.

4.2.6 Characterization of GUS expression of 7d-old seedlings from 6 mutagenized constructs in the promoter of miR390a

As Figure 4–9 showed, the main GUS activity of 7d-old transgenic plants of p*MIR390a:GUS-GFP* (Figure 4–9 a-1) was present in the true leaf (Figure 4–9 a-2), the cotyledon (Figure 4–9 a-2), the region below shoot apical meristem (Figure 4–9 a-3), as well as the lateral root primordia (Figure 4–9 a-4). Compare to the expression patterns of p*MIR390a:GUS-GFP* transgenic plants, the true leaves and cotyledons of *m1* (Figure 4–9 b-1) showed similar but slightly weaker GUS expression patterns on true leaves. The transgenic seedlings of *m2* (Figure 4–9 c-1), *m3* (Figure 4–9 d-1), *m4* (Figure 4–9 e-1) and *m6*-pattern2 (Figure 4–9 h-1) exhibited similar expression patterns as the reference line. The GUS staining of *m2* and *m3* mutants were stronger with respect to p*MIR390a:GUS-GFP,* especially in the cotyledons and the primary root cylinders (Figure 4–9 c-2 and Figure 4–9 d-2); the GUS activity of *m4* did not vary much but of *m6*-pattern2 turned to be slightly weaker when compared with the control line respectively. Among all the lines, he GUS activity of *m3* (Figure 4–9 d-1) was the strongest. On the contrary, the GUS expressions of m5 (Figure 4–9 f-1) and *m6*-pattern1 (Figure 4–9 g-1) were relatively weaker when compare with any other constructs.

A maximum of 60 T2 transgenic lines of each construct were statistically calculated and subjected to GUS staining, and all T2 transgenic lines owned the same transgene number with a segregation ratio of 3:1. Only the majority expression pattern (the pattern of ≥70% within 60 tested lines) were taken into account. Alternatively, the two majority expression patterns were considered (e.g. in the case of *m6*) and subjected to the subsequential stress treatment.
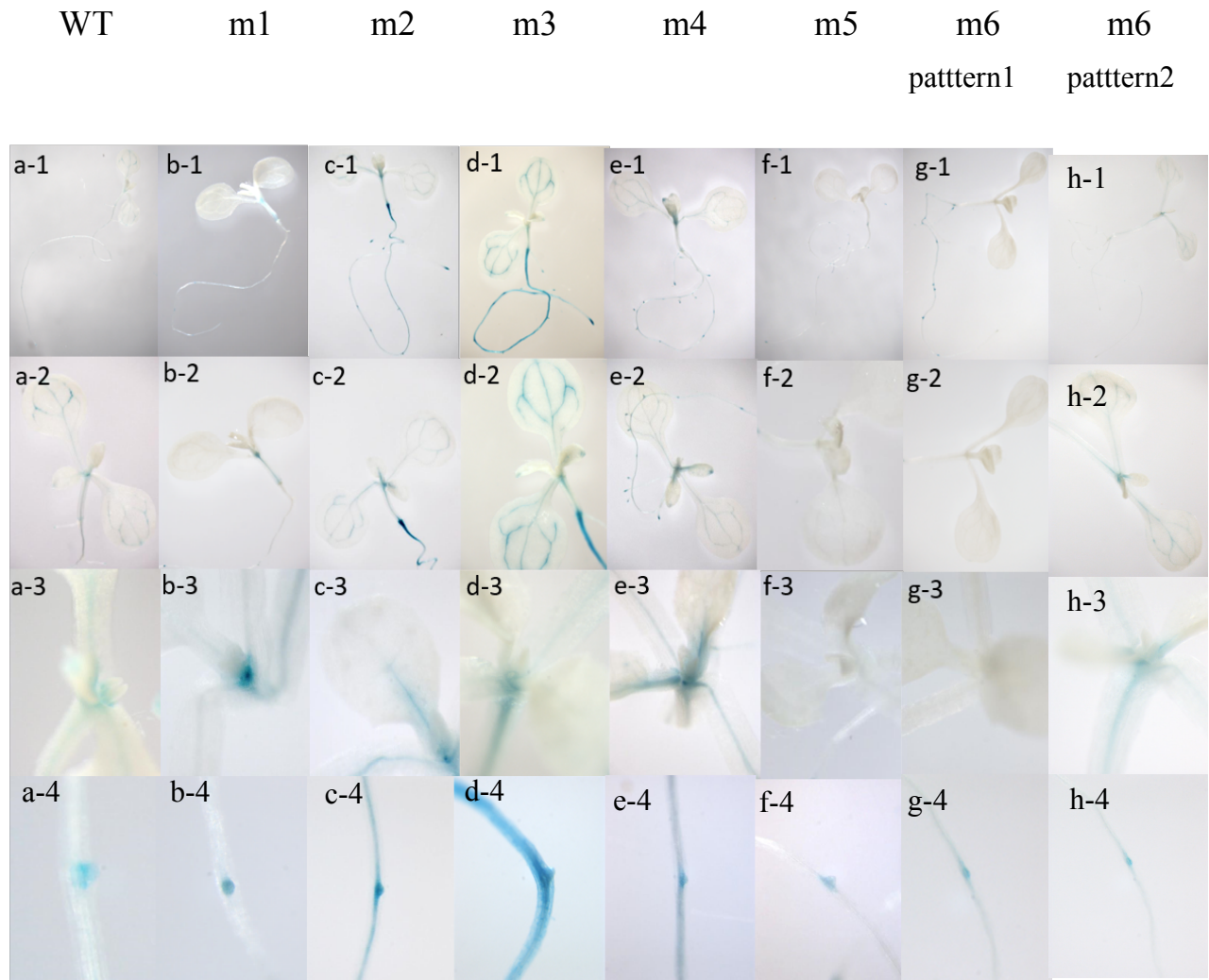
Figure 4–9 GUS assay of wild type and 6 site-mutagenized constructs T2 transgenic seedlings. a1-a4, b1-b4, c1-c4, d1-d4, e1-e4, f1-f4, g1-g4, h1-h4 substitute the whole seedling, the aerial part, the shoot apical meristem and the lateral root primordia from the constructs of p*MIR390a:GUS-GFP*, m1, m2, m3, m4, m5, m6-pattern1 and m6-pattern2, respectively.

4.3 Discussion

4.3.1 Possible sub-functionalization of miR390a and miR390b in plant development

Besides the GUS histochemical detection on young seedlings of miRNA390a and miRNA390b in *Arabidopsis*, we also observed clear GUS accumulation in the inflorescence of 38d-old transgenic plants of the two paralogs, which partially indicated that miRNA390 might also function in plant reproductive phase. What's

more, the two paralogs miRNA390a and miRNA390b having different expression patterns probably implied differential functionalization within the floral organs.

The two miRNA390 genes showed different spatial expression patterns in different tissues: miRNA390a promoter expressed on lateral root primordia, true leaf and cotyledon, yet absent from lateral root tip and shoot apical meristem. Whereas miRNA390b promoter specifically expressed on lateral root tip, alongside more restrictive expression on aerial part of transgenic *Arabidopsis*. Consensus with other evidence that miRNA390a contributes mainly in the early stages (stage 0 to 2/3) of lateral root initiation [274]. By performing 5'-RACE, Marin *et al* concluded that both miRNA390a and miRNA390b were produced by leaves [3], identical with what-we also observed weaker GUS activity of miRNA390b in the cotyledon, but stronger expression in true leaves and shoot apical meristem.

Generally, the overlapped and different expression patterns of the two genes might indicate the possible sub-functionalization with respect to the original copy of miRNA390 before splitting into two paralogs in *Arabidopsis*. Although originated from the identical miRNA390 sequence, as Figure 4–7 illustrates, their fold-backs differ in sequence and base pairs [275]. In addition, there was none common c*REs* was identified in their upstream promoter regions, which hinted the possible specialization on sub-functionalization.

Thus, we speculate that the ancestors of the two paralogs of miRNA390 had an expression pattern that was the overlap between the unions of the two gene expression patterns. Possibly, before becoming into miRNA390a and miRNA390b, the original miRNA390 was present in all the tissues: the primary root cylinder, the lateral root primordia, the lateral root tip, the shoot apical meristem, as well as the petals and filaments in the inflorescence. However, after years of evolution, miRNA390a and miRNA390b split and duplicated. Thereby, if the two genes were

still redundant, they would not carry out single role very well. Thus in order to maximize utilities, the two genes gradually underwent partial sub-functionalization in the evolutionary history. In particular, miRNA390a specialized its expression in early lateral root primordia, true leaf and cotyledon, as well as some organs of the inflorescence; whereas miRNA390b regulates lateral root tip and the shoot apical meristem development in *Arabidopsis*. So that they would integrally conduct the specific regulatory roles in plant development. Such overlapping expression patterns of miR390 might hint the similar evolutionary history as taking place on miR168 two paralogs within *Brassicaceae* [2].

4.3.2 Site-mutagenized promoter construction of miR390a in *Arabidopsis*

Considering the specific locations of these six elements on the upstream of promoter region of miR390a, it appears that the activity of six elements varied with distance to TSS. With respect to the expression of p*MIR390a:GUS-GFP*, mutations of proximal sites (putative TSS and iron-responsive E-box) in the wild type background enhanced the expression of miR390a promoter. These two mutations displayed similar, but stronger GUS activity in roots and leaves as well as in flowers. Hence the putative TSS and iron-responsive E-box could be the silencers that bound by repressors, and reduce transcription and the activity of gene promoter. On the contrary, mutations of distal elements (putative W-box M5 and IDE M6) tended to decrease the promoter activity, specifically in the aerial organs of the leaves. This demonstrated these *cis*-elements were possible enhancers that could increase transcription rates and enhance the promoter activity. Most likely, they are involved mainly in the regulation of aerial part of plant development. With regard to putative EIN motif M4, as Figure 4–9 e1-e4 showed, the accumulation of GUS activity of *m4* did not distinguish so much with respect to p*MIR390a:GUS-GFP*. The similar case occurred to proximal putative TATA-box M1 that *m1* (Figure 4–9 b1-b4) exhibited

similar expression pattern as the background control. Nonetheless, the GUS expression statistic could demonstrate the expression patterns of various promoters, while the GUS amount change among the mutagenized construct and control was not enough to elucidate the differences. Hence, the exact roles of miR390a-M1 and -M4 are still under identified, and more techniques are called to verify their putative roles in the subsequential work.

Chapter 5.   Functional characterization of miRNA390a *cis*-regulatory elements under iron-stress treatments

5.1   Methods

5.1.1 Iron-deficient treatment

T3 homozygous seeds of p*MIR390a:GUS-GFP* and transgenic plants of *m3* and *m6* were used, 2 transgenic lines were used for each selected pattern of the three constructs.

MicrAmp® Black 96-well plates were used as the geminating base. 2-3 seeds were sowed on each hole of the plates that filled with germination Agar, which was made with the same method as previous study [276].

Then put the plates into different sets (StarLab 1000μl Tip Rack) which filled with in modified hoagland hydroponic solution[4, 276] containing n 1.25 mM $Ca(NO_3)_2$, 1.25 mM KNO_3, 0.5 mM $MgSO_4$, 0.25 mM KH_2PO_4, 46 μMH_3BO_4, 9.15 μM $MnSO_4$, 0.77 μM $ZnSO_4$, 0.32 μM $CuSO_4$, 0.5 μM $Na_2MoO_4$and 50 μM Fe (III)-EDTA (Fe-sufficient)(pH 5.8).

The sets were left in 4°C fridge to germinate for three days, then were placed under conditions of 23-25 °C, 70% relative humidity, 100 μmol m^{-2} s^{-1} light intensity and 14 h photoperiod for 7 days, and treated with different iron-deficiency stress.

After the seedlings reaching true leaf emerging stage (7d-old), the entire hydroponic medium in different sets (StarLab 1000μl Tip Rack matched with MicrAmp® Black 96-well Base, Figure S1) was replaced with 650mL the same new full-nutrient solution as before (control) or 650mL new iron-deficient (without 50

µM Fe (III)-EDTA) solution, then the seedlings on different 96-well plates were placed in different sets (IDF/ISF) for 6h treatments.

5.1.2 GUS histochemical assay

The procedure of GUS staining performed on both 7d-old seedlings and 35-40d-old inflorescence is similar as 4.1.8. The only difference is, for functional analysis of mutagenized lines under iron-deficiency condition, the T3 rather than T2 seedlings with single-copy were used.

5.2 Results

5.2.1 Iron-deficiency treatment performed on p*MIR390a:GUS-GFP* to screening possible clue triggering the gene expression

Knowing that there are were iron-relevant *cREs*, the putative E-box (MIR390a-M3) and IDE (MIR390a-M6), in order to screen possible clue triggering the miR390a gene expression, we performed iron-deficient treatment on T3 transgenic plants of p*MIR390a:GUS-GFP*.

Since miR390a (At2g38325) was not circadian clock regulated[277], stress treatments with different time courses will not be affected by circadian rhythm.

7d-old T3 transgenic seedlings of p*MIR390a:GUS-GFP* were subjects to different time course iron-sufficient (control) or iron-deficient conditions, 6h iron-sufficient (ISF), 3h or 6h iron-deficient (IDF).

As shown in Figure 5–1, under iron-sufficient/deficient treatments, there was an obvious GUS activity amount change in primary root apical meristem and the lateral root primordia of 7d-old T3 p*MIR390a:GUS-GFP* transgenic seedlings. Compare to the GUS activity produced by the transformants under iron-sufficient condition (Figure 5–1 a1-a3), the GUS staining of the same transgenic line was stronger under either 3h (Figure 5–1 b1-b3) or 6h iron-deficient (Figure 5–1 c1-c3)

conditions. In addition, the same difference occurred between two iron-deficient time-courses. Most likely, the longer of iron-free treatment, the stronger the GUS activity was. This observation was apparent in the root part (especially in lateral root primordia as shown in Figure 5–1 a1-c1, and primary root apical meristem as shown in Figure 5–1 a3-c3). But in the aerial parts (data not shown here), the differences among different treatments were not so distinguishable.

This observation of reporter gene activity change of p*MIR390a:GUS-GFP* implied the iron-deficient condition could be a triggering stimuli of the promoter driven GUS activity. Hence, a possible iron-deficient responsive mechanism in miR390a promoter region drew our attention. Regard to which *cREs* possibly took responsibility for such a responsiveness, it still needed to be explored in the following experiments, that by repeating the same treatment on *Arabidopsis* transgenic lines where the GUS reporter gene was driven by site specific mutagenized promoter.
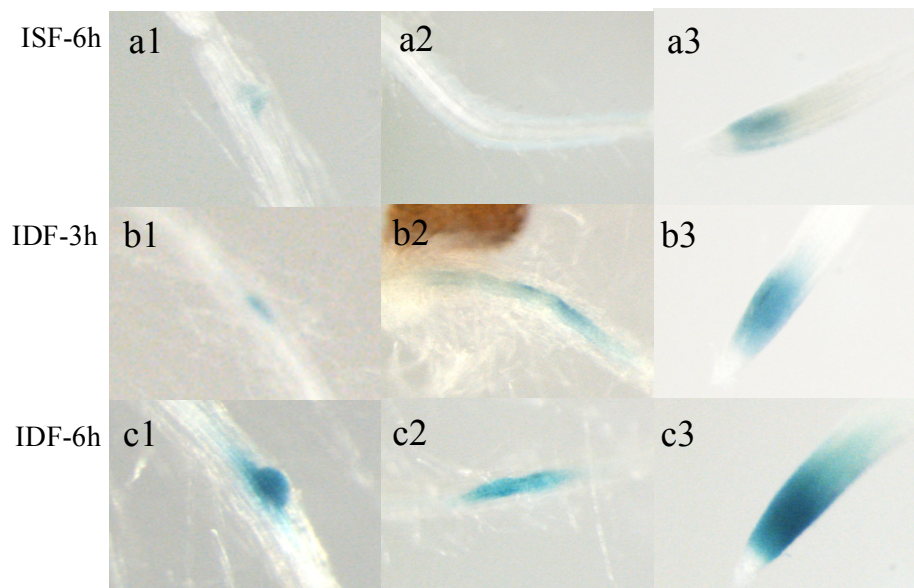
Figure 5–1 GUS expression patterns of p*MIR390a:GUS-GFP* homozygous treated with different time-course of iron-sufficient(ISF)/deficient(IDF) conditions. a1-c1, lateral root primordia; a2-c2, root cylinder; a3-c3, primary root apical meristem.

5.2.2 GUS histochemical assay of 7d-old seedlings under iron-sufficient/deficient treatment

Based on the results of iron-deficient test on p*MIR390a:GUS-GFP* transgenic lines, we speculated there were important iron responsiveness information that related with the putative iron responsive E-box MIR390a-M3 and the IDE motif MIR390a-M6. Subsequently, taking the reliable p*MIR390a:GUS-GFP* as reference line, stress treatments were performed on *m3* and *m6* with the promoters undergone site-directed mutagenesis. Particularly, 7d-old T3 homozygous transgenic seedling were treated with iron-sufficient/deficient condition and subsequently subjected to the GUS histochemical assay.

Each operation was performed on two T3 transgenic lines with the same expression pattern of mutagenized constructs, and 12 seedlings of each line were tested. Two T3 homozygous lines of *pMIR390a:GUS-GFP* were still taken as the background control. The results were shown in Figure 5–2.

Similar to the reference line, when supplied with ISF condition, the majority *m3* lines (Figure 5–2 c1-c4 under ISF, d1-d4 under IDF) exhibited consistent expression patterns in the lateral root primordia, the cotyledon as well as the true leaf.

While for *m6*, *m6*-pattern2 (Figure 5–2 g1-g4 under ISF, h1-h4 under IDF) showed similar but weaker expression patterns when compare with either the reference line or *m3*, respectively. *m6*-pattern2 didn't get any staining in the aerial parts (Figure 5–2 e1-e4 under ISF, f1-f4 under IDF), but rather weaker GUS activity in the lateral root primordia.

Compared with the expression level under iron-sufficient treatment (Figure 5–3 a1-a4), the T3 transgenic plants of p*MIR390a:GUS-GFP* under IDF condition (Figure 5–3 b1-b4) showed stronger GUS activity in lateral root primordia, and a

98

slight increase in the region below SAM. On the contrary, when supplied with IDF condition, the transgenic seedlings of *m3* (Figure 5–3 d1-d4) and *m6*-pattern1 (Figure 5–3 f1-f4) as well as *m6*-pattern2 (Figure 5–3 h1-h4) presented rather weaker expression level with respect to their expressions under ISF condition (*m3*, Figure 5–3 c1-c4; *m6*-pattern1, Figure 5–3 e1-e4; *m6*-pattern 2 g1-g4), the GUS staining was reduced in the surrounded area of lateral root primordia, and some aerial parts of *m3* also showed decreasing GUS activity.

Notably, the GUS activity level of p*MIR390a:GUS-GFP* T3 transgenic plants that subjected to hydroponic stress treatment (Figure 5–3 a1-a4) was weaker than the T2 progeny that grown in MS plate (Figure 4–9 a1-a4), which could be attributed to different gene copy numbers of them, that T3 was the homozygous while T2 progeny was heterozygous.
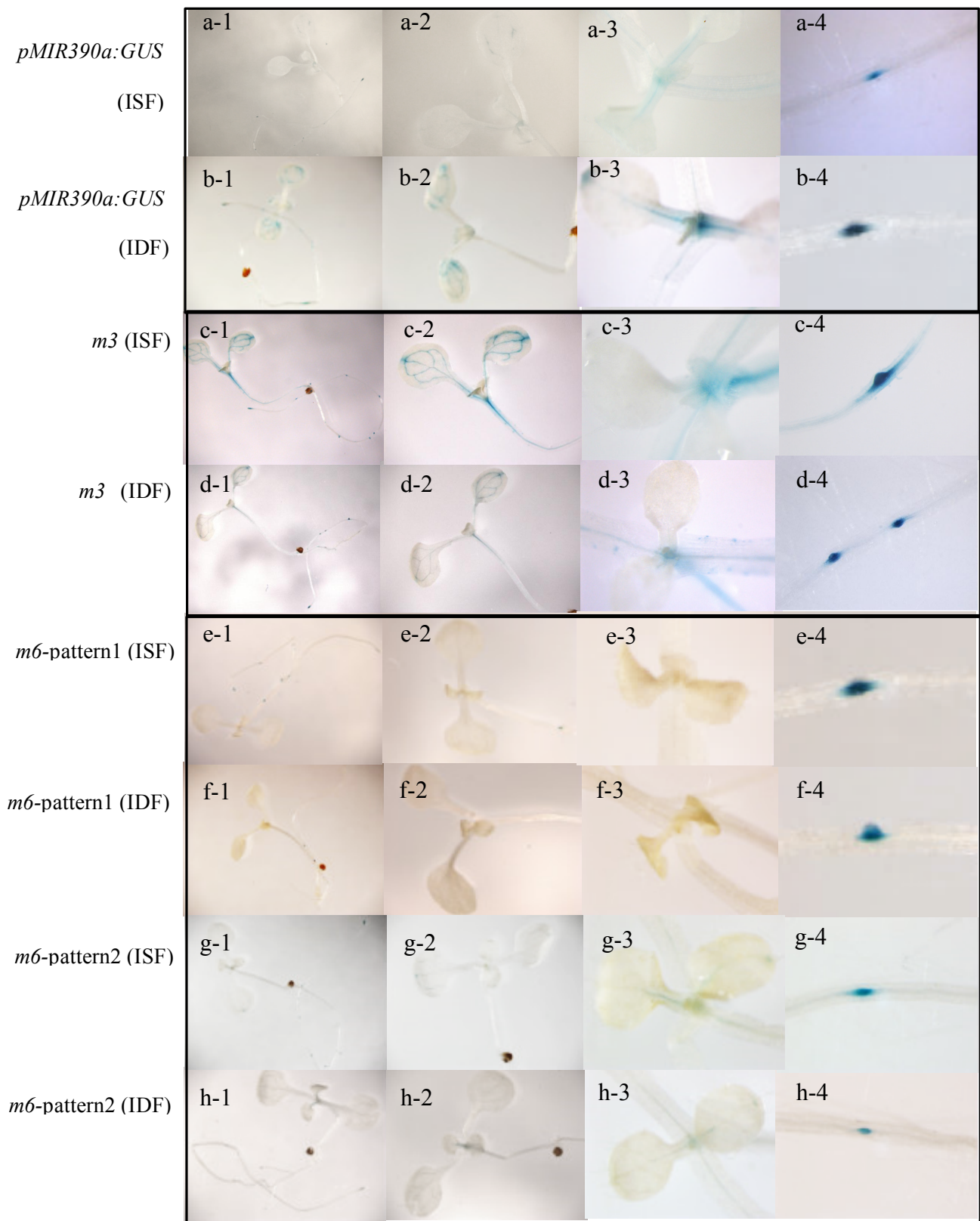
Figure 5–2 GUS assay of p*MIR390a:GUS-GFP* and site-mutagenized constructs *m3* as well as *m6* 7d-old T2 transgenic plants under Iron-sufficient (ISF) /deficient (IDF) stresses. The series a-h are: p*MIR390a:GUS-GFP* under ISF, p*MIR390a:GUS-GFP* under IDE; *m3* under ISF, *m3* under IDE; *m6*-pattern 1 under ISF, *m6*-pattern1 under IDF; *m6*-pattern2 under ISF, *m6*-pattern2 under IDF. 1-4 from each line substitute the whole seedling, the aerial part, the shoot apical meristem and the lateral root, respectively.

5.2.3 GUS histochemical assay of 38d-old inflorescence under iron-sufficient/deficient treatment
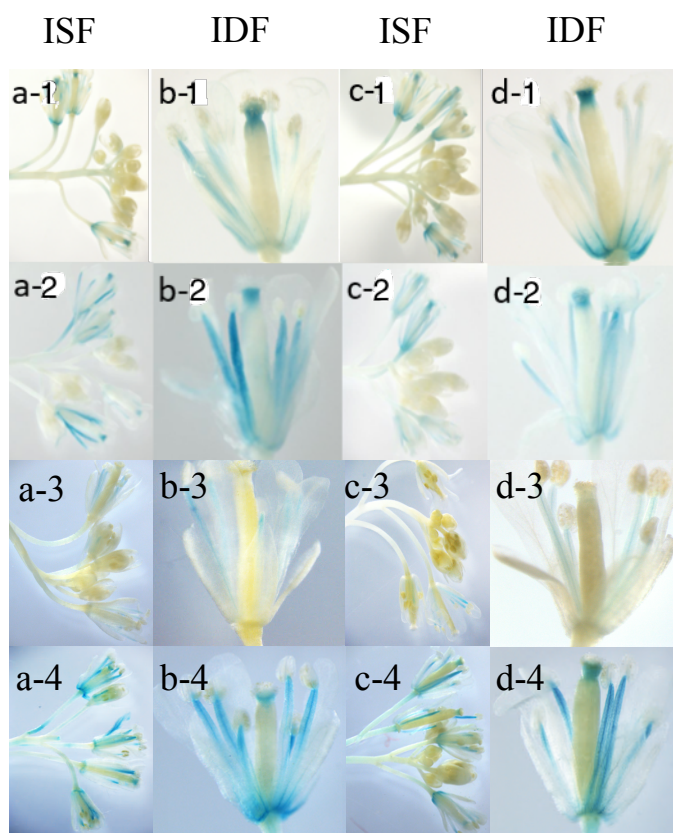
Figure 5–3 GUS assay of p*MIR390a:GUS-GFP* and site-mutagenized constructs *m3* and *m6* 38d-old T2 transgenic plants inflorescence under Iron-sufficient/deficient stresses. a1-a4, GUS staining in inflorescence of p*MIR390a:GUS-GFP*, *m3* and *m6*-pattern1, *m6*-pattern2, respectively, under iron-sufficient condition; b1-b4, GUS staining in single flowers of p*MIR390a:GUS-GFP*, *m3* and *m6*-pattern1, *m6*-pattern2, respectively, under iron-sufficient condition; c1-c4, GUS staining in inflorescence of p*MIR390a:GUS-GFP*, *m3* and *m6*-pattern1, *m6*-pattern2, respectively, under iron-deficient condition; d1-d4, GUS staining in single flowers of p*MIR390a:GUS-GFP*, *m3* and *m6*-pattern1, *m6*-pattern2, respectively, under iron-deficient condition.

Knowing that miRNA390a was capable to affect leaf patterning and developmental timing through a TAS-*tasi*RNAs-ARFs-auxin pathway [26, 135-139], besides detecting the GUS staining performance of p*MIR390a:GUS-GFP* in *Arabidopsis* transgenic leaves, we also performed the GUS assay in the inflorescence (Figure 4–6 i, j) and detected strong GUS activity in the floral organs. In the meantime, a decrease of GUS staining in the lateral root primordia region T3 transgenic plants (homozygous with single transgene copy) of both *m3* and *m6* two patterns under iron-deficient treatment was observed (Figure 5–2) respectively. Therefore, to test if the floral organs of the two constructs were also affected under iron-deficient condition, we further applied GUS assay in 38d-old inflorescence from both mutations-*m3* and *m6* T3 transgenic plants. Taking the p*MIR390a:GUS-GFP* transgenic lines as the reference, each test was conducted on two T3 homozygous transgenic lines with the same expression pattern.

As Figure 5–3 showed, the inflorescence from p*MIR390a:GUS-GFP* (Figure 5–3 a-1, b-1, c-1, d-1) under either iron-sufficient or iron-deficient condition exhibited the same expression pattern, which mainly in the filament, the apical region of pistil, the petals and sepals.

With respect to the reference line, *m3* (Figure 5–3 a-2, b-2, c-2, d-2) showed stronger GUS activity when supplied with ISF condition, identical fact occurred to these constructs under IDF condition.

Compared with either p*MIR390a:GUS-GFP* or *m3* lines, *m6*-pattern1 (Figure 5–3 a-3, b-3, c-3, d-3) expressed a rather weaker GUS activity when subjected to ISF or IDF stress, that was mainly in the filaments of the inflorescence; *m6*-pattern2 (Figure 5–3 a-4, b-4, c-4, d-4) showed similar expression patterns as the control like. Unlike the consistent but rather fainter activity in 7d-old seedlings with respect to the reference line, the inflorescence of *m6*-pattern2 exhibited stronger GUS activity.

This observation hinted a specific role in plant productive phase that differs from in vegetative phase of the putative iron-deficient responsive box M6.

In contrary to the trend occurred in p*MIR390a:GUS-GFP*, both T3 transgenic lines of *m3* and m6 respectively exhibited a slight decrease of GUS activity when subjected to IDF condition, especially in the filaments of *m3* (Figure 5–3 d-2) and *m6*-pattern1 (Figure 5–3 d-3), the sepals of *m6*-pattern2 (Figure 5–3 d-4).

5.3 Discussion

The change of GUS activity of p*MIR390a:GUS-GFP* T3 homozygous plants under iron-sufficient/deficient condition (Figure 5–1), indicating that it might be the iron-responsive elements present in the miR390a promoter region mediated this regulation. Based on the computational analysis, M3 and M6 motifs identified in miR390a promoter sequences were predicted to be iron responsive elements. Thereby, these two motifs could be the putative candidates in response to iron-deficient condition and need to be further verified in the future.

In order to appraise the iron regulation of M3 and M6 motifs, the transgenic plants encompassing mutagenized M3 and M6 motifs in the miR390a promoter were generated and subjected to iron-deficient condition, respectively. *m3* and *m6* exhibited weaker GUS staining in the area surrounded lateral root primordia under IDF stress. In addition, GUS activity change in *m3* and *m6* inflorescence was also observed when subjected to ISF/IDF treatments. The GUS staining was decreased when the inflorescence was treated with iron-deficient condition, especially in the filaments and sepals.

Under ISF/IDF treatments, the changes of the GUS expression level in both *m3* and *m6* transgenic plants with respect to the control respectively, demonstrating that the putative E-box M3 and the IDE M6 in miR390a promoter are likely to be responsible to the plant iron regulation.

However, since GUS activity were varied apparently depending on the transgenic insertion sites, we could not rule out that the differences between WT control and mutagenized M3 and M6 motifs were caused by a position effect. Thus supplemental functional characterization of these two specific elements will be necessary to uncover this effect. In addition, the results obtained from statistic-based GUS histochemical assay partially addressed their specific functional roles. Whereas to validate the GUS activity changes more precisely, more sensitive methods such as the quantitative real-time PCR would be necessary. It would be suggested to quantify the relative expression level of GUS reporter gene among homozygote transgenic plants from p*MIR390a:GUS-GFP*, *m3* and *m6* in the future.

Chapter 6.   Conclusion, discussion and perspective

In this study, we reported a comparative approach at the *Brassicaceae* family level to analyze the regulatory regions of microRNA genes. Using bioinformatics analysis, we characterized 29 microRNA genes belonging to 9 families. Through MEME motif prediction program, in the meantime by searching against PlantCARE database, we identified five overall conserved motifs plus an extra iron-deficient responsive element (IDE) in miRNA390a promoter region among 9 species, and five conserved motifs in miRNA390b promoter region through 8 species. Different expression patterns of miRNA390a and miRNA390b detected by GUS assay imply their different sub-functional roles in plant development. In addition, the experimental elucidation of the predicted motifs of miRNA390a indicated their different regulatory roles in root and flower development.

6.1   Conclusion

Drawing conclusion from this study, we have the following main points:

First, the multiple motifs predicted in single or multiple microRNA genes uncover the conservation of a variety of microRNAs in the evolutionary history of *Brassicaceae* family. These motifs could serve as a guide for characterizing the functional relevance in the regulation of microRNA.

Second, the identification of conserved motifs in miRNA390a and miRNA390b promoter regions across several species, suggests a possible common regulatory mechanism shared by different species. Given that there were no c*REs* shared by the two paralogs, miRNA390a and miRNA390b may play sub-specific role in plant development.

Third, GUS staining results suggested that miRNA390a expressed on lateral root primordia, true leaves, the cotyledons, as well as the floral organs, yet absent from lateral root tip and shoot apical meristem; whereas miRNA390b specifically expressed on lateral root tips, more restrictive expression was detected on aerial part of true leaves and floral organs in transgenic *Arabidopsis* plants. These expression differences between two paralogs, as well as the different motifs identified in the promoter regions, indicate possible sub-functionalization with respect to ancestral miRNA390 during plant evolution.

Forth, the site-specific mutations on six putative motifs identified in the miRNA390a promoter exhibited different GUS expression patterns: the elements' activity varied with distance to TSS. Mutations of proximal sites (m2 and m3) enhanced expression, while mutations of distal elements (m5 and m6) tended to decrease the promoter expression. These evidences suggest the possible modular cooperativity of miRNA390a cREs involved in development of plant root and aerial organs.

Fifth, the GUS histochemical expression changes detected on the site-specific mutations strongly demonstrated that the miRNA390a-M3 (putative iron-responsive E-box) and the miRNA390a-M6 (IDE) are putatively iron-deficient responsive.

6.2 Discussion

6.2.1 Advantages

The study is essential seen in four aspects.

First, the overall view of the regulatory region of a number of microRNAs across different species, could serve as a guide for future functional characterization of these microRNAs. The various microRNAs and their binding sites on targets were conserved through the evolutionary history [58]. Hence, with the clues of c*REs*

identification at *Brassicaceae* family level, the experimental validation could be carried out and draw a more complete map of microRNA regulatory mechanism in the evolutionary terms.

Second, the systematic analysis of the c*REs* regulatory network at the family level of *Brassicaceae*, provided an evolutionary view of plant microRNAs regulation machinery. Phylogenetic shadowing method has already been applied in both phylogenetically distant species (*Arabidopsis* and rice) [263, 278] and several close species within *Brassicaceae* [279], and the regulation machinery of microRNAs is covered but relatively limited. Therefore, this study exhibited an overall higher resolution to explore the microRNA silencing mystery in the evolutionary term.

Third, the reconstructed phylogenetic trees based on miRNA390a and miRNA390b promoter sequences, elucidated inconsistent phylogenetic relationships (species trees) with respect the data in known *Brassicaceae* phylogeny, which implied there might be multiple copies of specific c*REs* in some specific species. This result hinted a clue for better understanding the miRNA390 gene within *Brassicaceae*.

Forth, we identified the putative TSS and iron-deficiency responsive E-box that contribute to the proper plant development under either iron-sufficient or deficient conditions. Given the fact that the metallome of plant cells is very complex, and plants have crosstalk mechanism to maintain metal homeostasis [280, 281], it's not surprising that there are interactions between Fe and other metals. For example, analysis of the copper transporter2 (*COPT2*) promoter sequences indicated several *cis*-elements were responsive to both low iron and low copper, suggesting the interaction between Fe and Cu [282]. The bHLHs that regulated Fe homeostasis in plants, was also involved into Cadmium (Cd) tolerance [283]. Moreover, it was

indicated that some zinc (Zn) transporters was also indirectly regulated by Fe [284]. Considering the iron-deficient responsive element E-box, which is a putative binding site of FIT1, as well as the other known IDE, we presume that they are involved in other metal responsiveness beyond Fe. Therefore, these motifs could be possible tools to harness crop improvement in the future.

Fifth, the evidences of disguising c*REs* modulation in the promoter regions, as well as GUS histochemical assays point to possible sub- or neo-functionalization of miRNA390a and miRNA390b in the productive phase with respect to ancestral miR390, together could be the starting point to have a deeper look of their sub-functions in other plant developmental stages and the fruit organs.

6.2.2 Disadvantages

Regardless of the numerous advantages present in this study, there are still a few issues remain to be resolved.

First, although GUS histochemical assay was an efficient and visible method to trace the spatial and temporal expression patterns of miRNA390a, it's still a qualitative approach. Therefore, a complementary quantitative method should be applied to further confirm the observations of GUS staining. qRT-PCR is an efficient way to address this issue. However, due to the current impossibility in house, this technique had not been performed on the confirmation thus needs to be conducted in the future.

Second, to have a complete comparative view of the two paralogs, it could be instructive to carry out the experimental validation by both GUS histochemical assay and quantitative confirmation also on miRNA390b. Notably, considering possible false positive, a deeper identification of the putative motifs of other microRNA genes should be conducted by mining against more literatures.

Third, as concern to the feedback of ARF4 in the negative regulation of miRNA390a expression level [3], there were two putative AuxRE (Auxin responsive elements)-like core TGTCNN identified in miRNA390b promoter region, while not in miRNA390a. Considering the (Basic Leucine Zipper) bZIP- and MYB-related binding sites also could be potential AuxRE coupling elements [266], regardless of the negative identification of AuxRE in miRNA390a promoter sequences, there could be other possible AuxRE-related motifs yet to be identified in miRNA390 promoter regions.

Forth, as mentioned above, given the possible multiple roles in responsiveness to other metals of the putative E-box and IDE, it will be worth to explore the responsiveness of microRNA390 under other heavy metal conditions, such as Cu, Cd and Zn. Additionally, it would be significant to consider miRNA390b into various environmental screening as well. On one hand, we could exclude possible false positives that occurred to motifs detected from both paralogs; on the other hand, we could screen other environmental responsive clues for miRNA390 family.

Fifth, despite the relatively clear view of the roles of predicted TSS, TATA-box and E-box, as well as the IDE, the other two distant elements M4 and M5 are unidentified yet. Thus, more literature mining and experimental work will be necessary to trace their precise roles.

6.3   Perspective

Due to the complexity of the post-transcriptional network of microRNAs, there are still numerous promoters and regulatory elements waiting to be discovered. In addition, the combinational control of c*REs* and TFs in regulating microRNAs genes expression under various conditions is just initiated. Our comparative study on *Brassicaceae* just uncovered the tip of the iceberg of microRNA regulatory mechanism, while other groups within rosid family as well as many other plant

families remain to be explored. Fortunately, the increasing high throughput sequencing technologies such as DNA-seq, and their improving sensitivity and accuracy of detecting known microRNAs, will not only enrich our knowledge of the global gene regulation, but also facilitate the process of analyzing the microRNA regulatory regions. Meanwhile, more computational algorithms for better predicting the *cis*-regulatory elements are also in development, and hopefully they will be more efficient in circumventing the false-positive elements that are selected by current available tools. Accordingly, the motif functional characterization tools, such as simpler and more rapid transformation systems and validation approaches, still remain to be developed.

Eventually, these developments would lead to a better understanding of both transcriptional and post-transcriptional gene regulation, and bring about significant crop improvement with the available techniques.

Reference

1.   Corcoran, D.L., *Transcriptional Regulation Of MicroRNA Genes And The Regulatory Networks In Which They Participate*. 2008: ProQuest.

2.   Gazzani, S., et al., *Evolution of MIR168 paralogs in Brassicaceae.* Bmc Evolutionary Biology, 2009. **9**(1): p. 62.

3.   Marin, E., et al., *miR390, Arabidopsis TAS3 tasiRNAs, and Their AUXIN RESPONSE FACTOR Targets Define an Autoregulatory Network Quantitatively Regulating Lateral Root Growth.* The Plant Cell, 2010. **22**(4): p. 1104-1117.

4.   Kong, W.W. and Z.M. Yang, *Identification of iron-deficiency responsive microRNA genes and cis-elements in Arabidopsis.* Plant Physiology and Biochemistry, 2010. **48**(2-3): p. 153-159.

5.   Zhou, Z.S., J.B. Song, and Z.M. Yang, *Genome-wide identification of Brassica napus microRNAs and their targets in response to cadmium.* Journal of Experimental Botany, 2012. **63**(12): p. 4597-4613.

6.   Carrington, J.C. and V. Ambros, *Role of microRNAs in plant and animal development.* Science, 2003. **301**(5631): p. 336-338.

7.   Brodersen, P., et al., *Widespread translational inhibition by plant miRNAs and siRNAs.* Science, 2008. **320**(5880): p. 1185-1190.

8.   Voinnet, O., *Origin, biogenesis, and activity of plant microRNAs.* Cell, 2009. **136**(4): p. 669-687.

9.   Meyers, B.C., et al., *Criteria for annotation of plant MicroRNAs.* The Plant Cell, 2008. **20**(12): p. 3186-3190.

10.   Ambros, V., et al., *A uniform system for microRNA annotation.* Rna, 2003. **9**(3): p. 277-279.

11.   Kim, V.N., *MicroRNA biogenesis: coordinated cropping and dicing.* Nature reviews Molecular cell biology, 2005. **6**(5): p. 376-385.

12.   Mishra, A.K., G.S. Duraisamy, and J. Matoušek, *Discovering MicroRNAs and Their Targets in Plants.* Critical Reviews in Plant Sciences, 2015. **34**(6): p. 553-571.

13.   Reinhart, B.J., et al., *MicroRNAs in plants.* Genes & Development, 2002. **16**(13): p. 1616-1626.

14.   Bartel, D.P., *MicroRNAs genomics, biogenesis, mechanism, and function.* Cell, 2004. **116**(2): p. 281-297.

15.   Kim, V.N., *Small RNAs: classification, biogenesis, and function.* Mol cells, 2005. **19**(1): p. 1-15.

16.   Kozomara, A. and S. Griffiths-Jones, *miRBase: integrating microRNA annotation and deep-sequencing data.* Nucleic acids research, 2010: p. gkq1027.

17.   Chiang, H.R., et al., *Mammalian microRNAs: experimental evaluation of novel and previously annotated genes.* Genes & Development, 2010. **24**(10): p. 992-1009.

18.   Taylor, R.S., et al., *Evolutionary history of plant microRNAs.* Trends in Plant Science, 2014. **19**(3): p. 175-182.

19.   Lee, R.C., R.L. Feinbaum, and V. Ambros, *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14.* Cell, 1993. **75**(5): p. 843-854.

20.   Budak, H. and B.A. Akpinar, *Plant miRNAs: biogenesis, organization and origins.* Functional & integrative genomics, 2015. **15**(5): p. 523-531.

21.   Fahlgren, N., et al., *High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes.* PLoS ONE, 2007. **2**(2): p. e219.

22.   Pashkovskiy, P.P. and S.S. Ryazansky, *Biogenesis, evolution, and functions of plant microRNAs.* Biochemistry (Moscow), 2013. **78**(6): p. 627-637.

23.   de Felippes, F.F., et al., *Evolution of Arabidopsis thaliana microRNAs from random sequences.* Rna, 2008. **14**(12): p. 2455-2459.

24.   Manavella, P.A., et al., *Fast-forward genetics identifies plant CPL phosphatases as regulators of miRNA processing factor HYL1.* Cell, 2012. **151**(4): p. 859-870.

25. Liu, Q., et al., *Complementation of HYPONASTIC LEAVES1 by double-strand RNA-binding domains of DICER-LIKE1 in nuclear dicing bodies.* Plant physiology, 2013. **163**(1): p. 108-117.

26. Allen, E., et al., *microRNA-directed phasing during trans-acting siRNA biogenesis in plants.* Cell, 2005. **121**(2): p. 207-221.

27. Blevins, T., et al., *Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing.* Nucleic acids research, 2006. **34**(21): p. 6233-6246.

28. Moissiard, G. and O. Voinnet, *RNA silencing of host transcripts by cauliflower mosaic virus requires coordinated action of the four Arabidopsis Dicer-like proteins.* Proceedings of the National Academy of Sciences, 2006. **103**(51): p. 19593-19598.

29. Liu, B., et al., *Oryza sativa dicer-like4 reveals a key role for small interfering RNA silencing in plant development.* The Plant Cell, 2007. **19**(9): p. 2705-2718.

30. MacRae, I.J., et al., *Structural basis for double-stranded RNA processing by Dicer.* Science, 2006. **311**(5758): p. 195-198.

31. Baranauskè, S., et al., *Functional mapping of the plant small RNA methyltransferase: HEN1 physically interacts with HYL1 and DICER-LIKE 1 proteins.* Nucleic acids research, 2015: p. gkv102.

32. Park, M.Y., et al., *Nuclear processing and export of microRNAs in Arabidopsis.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(10): p. 3691-3696.

33. Chen, X., *Small RNAs and their roles in plant development.* Annual Review of Cell and Developmental, 2009. **25**: p. 21-44.

34. Schwarz, D.S., et al., *Asymmetry in the assembly of the RNAi enzyme complex.* Cell, 2003. **115**(2): p. 199-208.

35. Khvorova, A., A. Reynolds, and S.D. Jayasena, *Functional siRNAs and miRNAs exhibit strand bias.* Cell, 2003. **115**(2): p. 209-216.

36. Winter, J., et al., *Many roads to maturity: microRNA biogenesis pathways and their regulation.* Nature cell biology, 2009. **11**(3): p. 228-234.

37. Nilsen, T.W., *Mechanisms of microRNA-mediated gene regulation in animal cells.* TRENDS in Genetics, 2007. **23**(5): p. 243-249.

38. Pillai, R.S., S.N. Bhattacharyya, and W. Filipowicz, *Repression of protein synthesis by miRNAs: how many mechanisms?* Trends in cell biology, 2007. **17**(3): p. 118-126.

39. Eulalio, A., E. Huntzinger, and E. Izaurralde, *Getting to the root of miRNA-mediated gene silencing.* Cell, 2008. **132**(1): p. 9-14.

40. Erdmann, V.A. and J. Barciszewski, *Non coding RNAs in plants*. 2011: Springer Science & Business Media.

41. Endo, Y., H.o. Iwakawa, and Y. Tomari, *Arabidopsis ARGONAUTE7 selects miR390 through multiple checkpoints during RISC assembly.* Embo Reports, 2013. **14**(7): p. 652-658.

42. Bohmert, K., et al., *AGO1 defines a novel locus of Arabidopsis controlling leaf development.* The EMBO journal, 1998. **17**(1): p. 170-180.

43. Morel, J.-B., et al., *Fertile hypomorphic ARGONAUTE (ago1) mutants impaired in post-transcriptional gene silencing and virus resistance.* The Plant Cell, 2002. **14**(3): p. 629-639.

44. Vaucheret, H., *Plant argonautes.* Trends in Plant Science, 2008. **13**(7): p. 350-358.

45. Tolia, N.H. and L. Joshua-Tor, *Slicer and the argonautes.* Nature chemical biology, 2007. **3**(1): p. 36-43.

46. Ghildiyal, M. and P.D. Zamore, *Small silencing RNAs: an expanding universe.* Nature Reviews Genetics, 2009. **10**(2): p. 94-108.

47. Yuan, Y.-R., et al., *Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage.* Molecular Cell, 2005. **19**(3): p. 405-419.

48. Frank, F., et al., *Arabidopsis Argonaute MID domains use their nucleotide specificity loop to sort small RNAs.* The EMBO journal, 2012. **31**(17): p. 3588-3595.

49. Mi, S., et al., *Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5′ terminal nucleotide.* Cell, 2008. **133**(1): p. 116-127.

50. Vaucheret, H., et al., *The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development.* Genes & Development, 2004. **18**(10): p. 1187-1197.

51. Qi, Y., A.M. Denli, and G.J. Hannon, *Biochemical specialization within Arabidopsis RNA silencing pathways.* Molecular Cell, 2005. **19**(3): p. 421-428.

52. Jaubert, M., et al., *ARGONAUTE2 mediates RNA-silencing antiviral defenses against Potato virus X in Arabidopsis.* Plant physiology, 2011. **156**(3): p. 1556-1564.

53. Harvey, J.J., et al., *An antiviral defense role of AGO2 in plants.* PLoS ONE, 2011. **6**(1): p. e14639.

54. Wang, X.-B., et al., *The 21-nucleotide, but not 22-nucleotide, viral secondary small interfering RNAs direct potent antiviral defense by two cooperative argonautes in Arabidopsis thaliana.* The Plant Cell, 2011. **23**(4): p. 1625-1638.

55. Zilberman, D., X. Cao, and S.E. Jacobsen, *ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation.* Science, 2003. **299**(5607): p. 716-719.

56. Tucker, M.R., et al., *Somatic small RNA pathways promote the mitotic events of megagametogenesis during female reproductive development in Arabidopsis.* Development, 2012. **139**(8): p. 1399-1404.

57. Cenik, E.S. and P.D. Zamore, *Argonaute proteins.* Current Biology, 2011. **21**(12): p. R446-R449.

58. Pashkovskiy, P. and S. Ryazansky, *Biogenesis, evolution, and functions of plant microRNAs.* Biochemistry (Moscow), 2013. **78**(6): p. 627-637.

59. Montgomery, T.A., et al., *Specificity of ARGONAUTE7-miR390 Interaction and Dual Functionality in TAS3 Trans-Acting siRNA Formation.* Cell, 2008. **133**(1): p. 128-141.

60. Zhu, H., et al., *Arabidopsis Argonaute10 specifically sequesters miR166/165 to regulate shoot apical meristem development.* Cell, 2011. **145**(2): p. 242-256.

61. Zhou, Y., et al., *Spatiotemporal sequestration of miR165/166 by Arabidopsis Argonaute10 promotes shoot apical meristem maintenance.* Cell Reports, 2015. **10**(11): p. 1819-1827.

62. Lauressergues, D., et al., *Primary transcripts of microRNAs encode regulatory peptides.* Nature, 2015. **520**(7545): p. 90-93.

63. Couzigou, J.-M., et al., *miRNA-encoded peptides (miPEPs): A new tool to analyze the roles of miRNAs in plant biology.* Rna Biology, 2015. **12**(11): p. 1178-1180.

64. Waterhouse, P.M. and R.P. Hellens, *Plant biology: Coding in non-coding RNAs.* Nature, 2015. **520**(7545): p. 41-42.

65. Jones-Rhoades, M.W. and D.P. Bartel, *Computational identification of plant microRNAs and their targets, including a stress-induced miRNA.* Molecular Cell, 2004. **14**(6): p. 787-799.

66. Li, A. and L. Mao, *Evolution of plant microRNA gene families.* Cell Research, 2007. **17**(3): p. 212-218.

67. Griffiths-Jones, S., et al., *miRBase: microRNA sequences, targets and gene nomenclature.* Nucleic acids research, 2006. **34**(suppl 1): p. D140-D144.

68. Papp, I., et al., *Evidence for nuclear processing of plant micro RNA and short interfering RNA precursors.* Plant physiology, 2003. **132**(3): p. 1382-1390.

69. Lee, Y., et al., *The nuclear RNase III Drosha initiates microRNA processing.* Nature, 2003. **425**(6956): p. 415-419.

70. Basyuk, E., et al., *Human let‐7 stem–loop precursors harbor features of RNase III cleavage products.* Nucleic acids research, 2003. **31**(22): p. 6593-6597.

71.   Millar, A.A. and P.M. Waterhouse, *Plant and animal microRNAs: similarities and differences.* Functional & integrative genomics, 2005. **5**(3): p. 129-135.

72.   Fang, Z. and N. Rajewsky, *The impact of miRNA target sites in coding sequences and in 3′ UTRs.* PLoS ONE, 2011. **6**(3): p. e18067.

73.   Lai, E.C., *Predicting and validating microRNA targets.* Genome Biology, 2004. **5**: p. /2004/5/9/115-/2004/5/9/115.

74.   Wightman, B., I. Ha, and G. Ruvkun, *Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans.* Cell, 1993. **75**(5): p. 855-862.

75.   Bartel, D.P., *MicroRNAs: target recognition and regulatory functions.* Cell, 2009. **136**(2): p. 215-233.

76.   Rehmsmeier, M., et al., *Fast and effective prediction of microRNA/target duplexes.* Rna, 2004. **10**(10): p. 1507-1517.

77.   Haley, B. and P.D. Zamore, *Kinetic analysis of the RNAi enzyme complex.* Nature Structural & Molecular Biology, 2004. **11**(7): p. 599-606.

78.   Cuperus, J.T., et al., *Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis.* Nature Structural & Molecular Biology, 2010. **17**(8): p. 997-1003.

79.   Li, C. and B. Zhang, *MicroRNAs in Control of Plant Development.* Journal of Cellular Physiology, 2016. **231**(2): p. 303-313.

80.   Mallory, A.C. and H. Vaucheret, *Functions of microRNAs and related small RNAs in plants.* Nature Genetics, 2006. **38**: p. S31-S36.

81.   Sunkar, R., Y.-F. Li, and G. Jagadeeswaran, *Functions of microRNAs in plant stress responses.* Trends in Plant Science, 2012. **17**(4): p. 196-203.

82.   Nova-Franco, B., et al., *The Micro-RNA172c-APETALA2-1 Node as a Key Regulator of the Common Bean-Rhizobium etli Nitrogen Fixation Symbiosis.* Plant physiology, 2015. **168**(1): p. 273-291.

83.   Chen, X., *MicroRNA biogenesis and function in plants.* Febs Letters, 2005. **579**(26): p. 5923-5931.

84.   Mittal, D., et al., *Role of microRNAs in rice plant under salt stress.* Annals of Applied Biology, 2016. **168**(1): p. 2-18.

85.   Khraiwesh, B., J.-K. Zhu, and J. Zhu, *Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants.* Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 2012. **1819**(2): p. 137-148.

86.   Rubio-Somoza, I. and D. Weigel, *MicroRNA networks and developmental plasticity in plants.* Trends in Plant Science, 2011. **16**(5): p. 258-264.

87.   Kamthan, A., et al., *Small RNAs in plants: recent development and application for crop improvement.* Frontiers in Plant Science, 2015. **6**.

88.   Wong, C.E., et al., *MicroRNAs in the shoot apical meristem of soybean.* Journal of Experimental Botany, 2011: p. erq437.

89.   Bhalla, P.L. and M.B. Singh, *Molecular control of stem cell maintenance in shoot apical meristem.* Plant Cell Reports, 2006. **25**(4): p. 249-256.

90.   Chuck, G., H. Candela, and S. Hake, *Big impacts by small RNAs in plant development.* Current opinion in plant biology, 2009. **12**(1): p. 81-86.

91.   Zhang, Z. and X. Zhang, *Argonautes compete for miR165/166 to regulate shoot apical meristem development.* Current opinion in plant biology, 2012. **15**(6): p. 652-658.

92.   Liu, Q., et al., *The ARGONAUTE10 gene modulates shoot apical meristem maintenance and establishment of leaf polarity by repressing miR165/166 in Arabidopsis.* The Plant Journal, 2009. **58**(1): p. 27-40.

93. Tucker, M.R., et al., *Vascular signalling mediated by ZWILLE potentiates WUSCHEL function during shoot meristem stem cell development in the Arabidopsis embryo.* Development, 2008. **135**(17): p. 2839-2843.

94. Juarez, M.T., et al., *microRNA-mediated repression of rolled leaf1 specifies maize leaf polarity.* Nature, 2004. **428**(6978): p. 84-88.

95. Allen, R.S., et al., *Genetic analysis reveals functional redundancy and the major target genes of the Arabidopsis miR159 family.* Proceedings of the National Academy of Sciences, 2007. **104**(41): p. 16371-16376.

96. Baumann, K., *Plant cell biology: Mobile miRNAs for stem cell maintenance.* Nature reviews Molecular cell biology, 2013. **14**(3): p. 128-129.

97. Knauer, S., et al., *A Protodermal miR394 Signal Defines a Region of Stem Cell Competence in the Arabidopsis Shoot Meristem.* Developmental Cell, 2013. **24**(2): p. 125-132.

98. Song, J.B., et al., *Regulation of Leaf Morphology by MicroRNA394 and its Target LEAF CURLING RESPONSIVENESS.* Plant and Cell Physiology, 2012. **53**(7): p. 1283-1294.

99. Efroni, I., et al., *A protracted and dynamic maturation schedule underlies Arabidopsis leaf development.* The Plant Cell, 2008. **20**(9): p. 2293-2306.

100. Schommer, C., et al., *Control of jasmonate biosynthesis and senescence by miR319 targets.* Plos Biology, 2008. **6**(9): p. e230.

101. Ori, N., et al., *Regulation of LANCEOLATE by miR319 is required for compound-leaf development in tomato.* Nature Genetics, 2007. **39**(6): p. 787-791.

102. Yang, C., et al., *Overexpression of microRNA319 impacts leaf morphogenesis and leads to enhanced cold tolerance in rice (Oryza sativa L.).* Plant, cell & environment, 2013. **36**(12): p. 2207-2218.

103. Zhou, M., et al., *Constitutive expression of a miR319 gene alters plant development and enhances salt and drought tolerance in transgenic creeping bentgrass.* Plant physiology, 2013. **161**(3): p. 1375-1391.

104. Kim, J.H., D. Choi, and H. Kende, *The AtGRF family of putative transcription factors is involved in leaf and cotyledon growth in Arabidopsis.* The Plant Journal, 2003. **36**(1): p. 94-104.

105. Rodriguez, R.E., et al., *Control of cell proliferation in Arabidopsis thaliana by microRNA miR396.* Development, 2010. **137**(1): p. 103-112.

106. Wang, L., et al., *miR396-targeted AtGRF transcription factors are required for coordination of cell division and differentiation during leaf development in Arabidopsis.* Journal of Experimental Botany, 2011. **62**(2): p. 761-773.

107. Liu, D., et al., *Ectopic expression of miR396 suppresses GRF target gene expression and alters leaf growth in Arabidopsis.* Physiologia Plantarum, 2009. **136**(2): p. 223-236.

108. Mecchia, M.A., et al., *MicroRNA miR396 and RDR6 synergistically regulate leaf development.* Mechanisms of Development, 2013. **130**(1): p. 2-13.

109. Debernardi, J.M., et al., *Functional specialization of the plant miR396 regulatory network through distinct microRNA–target interactions.* Plos Genetics, 2012. **8**(1): p. e1002419.

110. Mockaitis, K. and M. Estelle, *Auxin receptors and plant development: a new signaling paradigm.* Annual review of cell and developmental biology, 2008. **24**: p. 55-80.

111. Bian, H., et al., *Distinctive expression patterns and roles of the miRNA393/TIR1 homolog module in regulating flag leaf inclination and primary and crown root growth in rice (Oryza sativa).* New Phytologist, 2012. **196**(1): p. 149-161.

112. Xie, K., et al., *Gradual increase of miR156 regulates temporal expression changes of numerous genes during leaf development in rice.* Plant physiology, 2012. **158**(3): p. 1382-1394.

113.   Koyama, T., et al., *TCP transcription factors control the morphology of shoot lateral organs via negative regulation of the expression of boundary-specific genes in Arabidopsis.* The Plant Cell, 2007. **19**(2): p. 473-484.

114.   Li, Z., et al., *Ethylene-insensitive3 is a senescence-associated gene that accelerates age-dependent leaf senescence by directly repressing miR164 transcription in Arabidopsis.* The Plant Cell, 2013. **25**(9): p. 3311-3328.

115.   Kim, J.H., et al., *Trifurcate feed-forward regulation of age-dependent cell death involving miR164 in Arabidopsis.* Science, 2009. **323**(5917): p. 1053-1057.

116.   Meng, Y., et al., *MicroRNA-mediated signaling involved in plant root development.* Biochemical and Biophysical Research Communications, 2010. **393**(3): p. 345-349.

117.   Guo, H.-S., et al., *MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals for Arabidopsis lateral root development.* The Plant Cell, 2005. **17**(5): p. 1376-1386.

118.   Yoon, E.K., et al., *Auxin regulation of the microRNA390-dependent transacting small interfering RNA pathway in Arabidopsis lateral root development.* Nucleic Acids Research, 2009. **38**(4): p. 1382-1391.

119.   Luo, Q.-J., et al., *An autoregulatory feedback loop involving PAP1 and TAS4 in response to sugars in Arabidopsis.* Plant Molecular Biology, 2012. **80**(1): p. 117-129.

120.   Xia, R., et al., *Apple miRNAs and tasiRNAs with novel regulatory networks.* Genome Biology, 2012. **13**(6): p. R47.

121.   Wang, J.-W., et al., *Control of root cap formation by microRNA-targeted auxin response factors in Arabidopsis.* The Plant Cell, 2005. **17**(8): p. 2204-2216.

122.   Gutierrez, L., et al., *Phenotypic plasticity of adventitious rooting in Arabidopsis is controlled by complex regulation of AUXIN RESPONSE FACTOR transcripts and microRNA abundance.* The Plant Cell, 2009. **21**(10): p. 3119-3132.

123.   Xia, K., et al., *OsTIR1 and OsAFB2 downregulation via OsmiR393 overexpression leads to more tillers, early flowering and less tolerance to salt and drought in rice.* PLoS ONE, 2012. **7**(1): p. e30039.

124.   Combier, J.-P., et al., *MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in Medicago truncatula.* Genes & Development, 2006. **20**(22): p. 3084-3088.

125.   Boualem, A., et al., *MicroRNA166 controls root and nodule development in Medicago truncatula.* The Plant Journal, 2008. **54**(5): p. 876-887.

126.   Li, H., et al., *Misexpression of miR482, miR1512, and miR1515 increases soybean nodulation.* Plant physiology, 2010. **153**(4): p. 1759-1770.

127.   Okamuro, J.K., B. Den Boer, and K.D. Jofuku, *Regulation of Arabidopsis flower development.* The Plant Cell, 1993. **5**(10): p. 1183.

128.   Rhoades, M.W., et al., *Prediction of plant microRNA targets.* Cell, 2002. **110**(4): p. 513-520.

129.   Xing, S., et al., *miR156-targeted and nontargeted SBP-box transcription factors act in concert to secure male fertility in Arabidopsis.* The Plant Cell, 2010. **22**(12): p. 3935-3950.

130.   Zhu, Q.-H. and C.A. Helliwell, *Regulation of flowering time and floral patterning by miR172.* Journal of Experimental Botany, 2010: p. erq295.

131.   Park, W., et al., *CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana.* Current Biology, 2002. **12**(17): p. 1484-1495.

132.   Chen, X., *A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development.* Science, 2004. **303**(5666): p. 2022-2025.

133.   Wu, M.-F., Q. Tian, and J.W. Reed, *Arabidopsis microRNA167 controls patterns of ARF6 and ARF8 expression, and regulates both female and male reproduction.* Development, 2006. **133**(21): p. 4211-4218.

134.   Nag, A., S. King, and T. Jack, *miR319a targeting of TCP4 is critical for petal growth and development in Arabidopsis.* Proceedings of the National Academy of Sciences, 2009. **106**(52): p. 22534-22539.

135.   Adenot, X., et al., *DRB4-Dependent TAS3 trans-Acting siRNAs Control Leaf Morphology through AGO7.* Current Biology, 2006. **16**(9): p. 927-932.

136.   Fahlgren, N., et al., *Regulation of AUXIN RESPONSE FACTOR3 by TAS3 ta-siRNA Affects Developmental Timing and Patterning in Arabidopsis.* Current Biology, 2006. **16**(9): p. 939-944.

137.   Garcia, D., et al., *Specification of leaf polarity in Arabidopsis via the trans-acting siRNA pathway.* Current Biology, 2006. **16**(9): p. 933-938.

138.   Hunter, C., et al., *Trans-acting siRNA-mediated repression of ETTIN and ARF4 regulates heteroblasty in Arabidopsis.* Development, 2006. **133**(15): p. 2973-2981.

139.   Chitwood, D.H., et al., *Pattern formation via small RNA mobility.* Genes & Development, 2009. **23**(5): p. 549-554.

140.   Lim, P.O., et al., *Auxin response factor 2 (ARF2) plays a major role in regulating auxin-mediated leaf longevity.* Journal of Experimental Botany, 2010: p. erq010.

141.   Woo, H.R., et al., *Plant leaf senescence and death–regulation by multiple layers of control and implications for aging in general.* Journal of Cell Science, 2013. **126**(21): p. 4823-4833.

142.   Copenhaver, G.P., et al., *Regulation of Small RNA Accumulation in the Maize Shoot Apex.* Plos Genetics, 2009. **5**(1): p. e1000320.

143.   Husbands, A.Y., et al., *Signals and prepatterns: new insights into organ polarity in plants.* Genes & Development, 2009. **23**(17): p. 1986-1997.

144.   Megraw, M., et al., *Small Genetic Circuits and microRNAs: Big Players in Pol-II Transcriptional Control in Plants.* The Plant Cell, 2016: p. TPC2015-00852-REV.

145.   Lee, Y., et al., *MicroRNA genes are transcribed by RNA polymerase II.* The EMBO journal, 2004. **23**(20): p. 4051-4060.

146.   Xie, Z., *Expression of Arabidopsis MIRNA Genes.* Plant Physiology, 2005. **138**(4): p. 2145-2154.

147.   Molina, C. and E. Grotewold, *Genome wide analysis of Arabidopsis core promoters.* Bmc Genomics, 2005. **6**(1): p. 1.

148.   Megraw, M., et al., *MicroRNA promoter element discovery in Arabidopsis.* Rna, 2006. **12**(9): p. 1612-1619.

149.   Werner, T., *Models for prediction and recognition of eukaryotic promoters.* Mammalian Genome, 1999. **10**(2): p. 168-175.

150.   Smale, S.T., *Core promoters: active contributors to combinatorial gene regulation.* Genes & Development, 2001. **15**(19): p. 2503-2508.

151.   LIU, Y.-x., et al., *Genomic analysis of microRNA promoters and their cis-acting elements in soybean.* Agricultural Sciences in China, 2010. **9**(11): p. 1561-1570.

152.   Burley, S. and R. Roeder, *Biochemistry and structural biology of transcription factor IID (TFIID).* Annual review of biochemistry, 1996. **65**(1): p. 769-799.

153.   Udvardi, M.K., et al., *Legume transcription factors: global regulators of plant development and response to the environment.* Plant Physiology, 2007. **144**(2): p. 538-549.

154.   Eulgem, T., et al., *The WRKY superfamily of plant transcription factors.* Trends in Plant Science, 2000. **5**(5): p. 199-206.

155.   Aida, M., et al., *Genes involved in organ separation in Arabidopsis: an analysis of the cup-shaped cotyledon mutant.* The Plant Cell, 1997. **9**(6): p. 841-857.

156. Kagaya, Y., K. Ohmiya, and T. Hattori, *RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants.* Nucleic acids research, 1999. **27**(2): p. 470-478.

157. Riechmann, J.L. and E.M. Meyerowitz, *The AP2/EREBP family of plant transcription factors.* Biological Chemistry, 1998. **379**: p. 633-646.

158. Nagano, Y., *Several features of the GT-factor trihelix domain resemble those of the Myb DNA-binding domain.* Plant physiology, 2000. **124**(2): p. 491-494.

159. Yanagisawa, S., *Dof domain proteins: plant-specific transcription factors associated with diverse phenomena unique to plants.* Plant and Cell Physiology, 2004. **45**(4): p. 386-391.

160. Riechmann, J.L., et al., *Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes.* Science, 2000. **290**(5499): p. 2105-2110.

161. Guilfoyle, T., et al., *How does auxin turn on genes?* Plant physiology, 1998. **118**(2): p. 341-347.

162. Alves, M.S., et al., *Plant bZIP transcription factors responsive to pathogens: a review.* International journal of molecular sciences, 2013. **14**(4): p. 7815-7828.

163. Heintzman, N. and B. Ren, *The gateway to transcription: identifying, characterizing and understanding promoters in the eukaryotic genome.* Cellular and molecular life sciences, 2007. **64**(4): p. 386-400.

164. Sabalza, M., et al., *Promoter diversity in multigene transformation.* Plant Molecular Biology, 2010. **73**(45): p. 363378.

165. Potenza, C., L. Aleman, and C. Sengupta-Gopalan, *Targeting transgene expression in research, agricultural, and environmental applications: promoters used in plant transformation.* In Vitro Cellular & Developmental Biology-Plant, 2004. **40**(1): p. 1-22.

166. Thomas, M.C. and C.-M. Chiang, *The general transcription machinery and general cofactors.* Critical reviews in biochemistry and molecular biology, 2006. **41**(3): p. 105-178.

167. Franco-Zorrilla, J.M., et al., *DNA-binding specificities of plant transcription factors and their potential to define target genes.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(6): p. 2367-2372.

168. Stevens, S.G. and C.M. Brown, *Bioinformatic methods to discover cis-regulatory elements in mRNAs*, in *Springer Handbook of Bio-/Neuroinformatics*. 2014, Springer. p. 151-169.

169. Priest, H.D., S.A. Filichkin, and T.C. Mockler, *Cis-regulatory elements in plant cell signaling.* Current opinion in plant biology, 2009. **12**(5): p. 643-649.

170. Jimenez-Lopez, J.C., et al., *Genome sequencing and next-generation sequence data analysis: A comprehensive compilation of bioinformatics tools and databases.* American Journal of Molecular Biology, 2013. **3**(02): p. 115.

171. Porto, M.S., et al., *Plant promoters: an approach of structure and function.* Molecular biotechnology, 2014. **56**(1): p. 38-49.

172. Tompa, M., et al., *Assessing computational tools for the discovery of transcription factor binding sites.* Nature biotechnology, 2005. **23**(1): p. 137-144.

173. Higo, K., et al., *Plant cis-acting regulatory DNA elements (PLACE) database: 1999.* Nucleic acids research, 1999. **27**(1): p. 297-300.

174. Thomas-Chollier, M., et al., *RSAT: regulatory sequence analysis tools.* Nucleic acids research, 2008. **36**(suppl 2): p. W119-W127.

175. Matys, V., et al., *TRANSFAC®: transcriptional regulation, from patterns to profiles.* Nucleic acids research, 2003. **31**(1): p. 374-378.

176. Davuluri, R.V., et al., *AGRIS: Arabidopsis gene regulatory information server, an information resource of Arabidopsis cis-regulatory elements and transcription factors.* BMC bioinformatics, 2003. **4**(1): p. 25.

177.	Chen, Y.-A., Y.-C. Wen, and W.-C. Chang, *AtPAN: an integrated system for reconstructing transcriptional regulatory networks in Arabidopsis thaliana.* Bmc Genomics, 2012. **13**(1): p. 85.

178.	Hehl, R., et al., *Boosting AthaMap Database Content with Data from Protein Binding Microarrays.* Plant and Cell Physiology, 2016. **57**(1): p. e4-e4.

179.	Rombauts, S., et al., *PlantCARE, a plant cis-acting regulatory element database.* Nucleic acids research, 1999. **27**(1): p. 295-296.

180.	Sandve, G.K., et al., *Improved benchmarks for computational motif discovery.* BMC bioinformatics, 2007. **8**(1): p. 193.

181.	Chaivorapol, C., et al., *CompMoby: comparative MobyDick for detection of cis-regulatory motifs.* BMC bioinformatics, 2008. **9**(1): p. 455.

182.	Kleftogiannis, D., et al., *Where we stand, where we are moving: Surveying computational techniques for identifying miRNA genes and uncovering their regulatory role.* Journal of biomedical informatics, 2013. **46**(3): p. 563-573.

183.	Maston, G.A., S.K. Evans, and M.R. Green, *Transcriptional regulatory elements in the human genome.* Annu. Rev. Genomics Hum. Genet., 2006. **7**: p. 29-59.

184.	Bailey, T.L., et al., *MEME: discovering and analyzing DNA and protein sequence motifs.* Nucleic acids research, 2006. **34**(suppl 2): p. W369-W373.

185.	Hu, J., B. Li, and D. Kihara, *Limitations and potentials of current motif discovery algorithms.* Nucleic acids research, 2005. **33**(15): p. 4899-4913.

186.	Hernandez-Garcia, C.M. and J.J. Finer, *Identification and validation of promoters and cis-acting regulatory elements.* Plant Science, 2014. **217**: p. 109-119.

187.	White, M.A., *Understanding how cis-regulatory function is encoded in DNA sequence using massively parallel reporter assays and designed sequences.* Genomics, 2015. **106**(3): p. 165-170.

188.	Nguyen, V.T., M. Morange, and O. Bensaude, *Firefly luciferase luminescence assays using scintillation counters for quantitation in transfected mammalian cells.* Analytical Biochemistry, 1988. **171**(2): p. 404-408.

189.	Galas, D.J. and A. Schmitz, *DNAase footprinting a simple method for the detection of protein-DNA binding specificity.* Nucleic acids research, 1978. **5**(9): p. 3157-3170.

190.	Garner, M.M. and A. Revzin, *A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system.* Nucleic acids research, 1981. **9**(13): p. 3047-3060.

191.	Tuerk, C. and L. Gold, *Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase.* Science, 1990. **249**(4968): p. 505-510.

192.	Eckardt, N.A., *Light regulation of plant development: HY5 Genomic binding sites.* The Plant Cell, 2007. **19**(3): p. 727-729.

193.	Jones, S.J., *Prediction of genomic functional elements.* Annu. Rev. Genomics Hum. Genet., 2006. **7**: p. 315-338.

194.	Reece-Hoyes, J.S. and A.M. Walhout, *Yeast one-hybrid assays: a historical and technical perspective.* Methods, 2012. **57**(4): p. 441-447.

195.	Palaniswamy, S.K., et al., *AGRIS and AtRegNet. a platform to link cis-regulatory elements and transcription factors into regulatory networks.* Plant physiology, 2006. **140**(3): p. 818-829.

196.	Vandepoele, K., et al., *Unraveling transcriptional control in Arabidopsis using cis-regulatory elements and coexpression networks.* Plant physiology, 2009. **150**(2): p. 535-546.

197.	Steffens, N.O., et al., *AthaMap: an online resource for in silico transcription factor binding sites in the Arabidopsis thaliana genome.* Nucleic acids research, 2004. **32**(suppl 1): p. D368-D372.

198.	Toufighi, K., et al., *The Botany Array Resource: e‐Northerns, Expression Angling, and promoter analyses.* The Plant Journal, 2005. **43**(1): p. 153-163.

199.	Guo, A., et al., *DATF: a database of Arabidopsis transcription factors.* Bioinformatics, 2005. **21**(10): p. 2568-2569.

200.	Barta, E., et al., *DoOP: Databases of Orthologous Promoters, collections of clusters of orthologous upstream sequences from chordates and plants.* Nucleic acids research, 2005. **33**(suppl 1): p. D86-D90.

201.	Michael, T.P., et al., *Network discovery pipeline elucidates conserved time-of-day–specific cis-regulatory modules.* Plos Genetics, 2008. **4**(2): p. e14.

202.	Sandelin, A., et al., *JASPAR: an open‐access database for eukaryotic transcription factor binding profiles.* Nucleic acids research, 2004. **32**(suppl 1): p. D91-D94.

203.	Ma, S., et al., *Discovery of stress responsive DNA regulatory motifs in Arabidopsis.* PLoS ONE, 2012. **7**(8): p. e43198.

204.	Thijs, G., et al., *A Gibbs sampling method to detect overrepresented motifs in the upstream regions of coexpressed genes.* Journal of Computational Biology, 2002. **9**(2): p. 447-464.

205.	Chang, W.-C., et al., *PlantPAN: Plant promoter analysis navigator, for identifying combinatorial cis-regulatory elements with distance constraint in plant gene groups.* Bmc Genomics, 2008. **9**(1): p. 1.

206.	Shahmuradov, I.A., et al., *PlantProm: a database of plant promoter sequences.* Nucleic acids research, 2003. **31**(1): p. 114-117.

207.	Guo, A.-Y., et al., *PlantTFDB: a comprehensive plant transcription factor database.* Nucleic acids research, 2008. **36**(suppl 1): p. D966-D969.

208.	Yamamoto, Y.Y. and J. Obokata, *ppdb: a plant promoter database.* Nucleic acids research, 2008. **36**(suppl 1): p. D977-D981.

209.	Swarbreck, D., et al., *The Arabidopsis Information Resource (TAIR): gene structure and function annotation.* Nucleic acids research, 2008. **36**(suppl 1): p. D1009-D1014.

210.	Pavesi, G., et al., *Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes.* Nucleic acids research, 2004. **32**(suppl 2): p. W199-W203.

211.	Mardis, E.R., *Next-generation sequencing platforms.* Annual review of analytical chemistry, 2013. **6**: p. 287-303.

212.	Margulies, M., et al., *Genome sequencing in microfabricated high-density picolitre reactors.* Nature, 2005. **437**(7057): p. 376-380.

213.	Droege, M. and B. Hill, *The Genome Sequencer FLX™ System—Longer reads, more applications, straight forward bioinformatics and more complete data sets.* Journal of biotechnology, 2008. **136**(1): p. 3-10.

214.	GS FLX+ System: 454 Life Sciences, a.R.C.; [Online]. Available from: http://454.com/products/gs-flx-system/.

215.	Mardis, E.R., *Next-generation DNA sequencing methods.* Annu. Rev. Genomics Hum. Genet., 2008. **9**: p. 387-402.

216.	Valouev, A., et al., *A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning.* Genome Research, 2008. **18**(7): p. 1051-1063.

217.	Liu, L., et al., *Comparison of next-generation sequencing systems.* Biomed Research International, 2012. **2012**.

218.	*GS Junior Instrument Owner's manual.* [Online]. Available from: http://my454.com/downloads/my454/documentation/gs-junior/system-wide-documents/GSJunior_OwnersManual_RevSept2010.pdf.

219.	GS FLX+ System: 454 Life Sciences, a.R.C. 2016; [Online]. Available from: http://454.com/applications/whole-genome-sequencing/.

220. Shendure, J. and H. Ji, *Next-generation DNA sequencing.* Nature biotechnology, 2008. **26**(10): p. 1135-1145.

221. *FastQC.* Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

222. *Raw data processing.* 2013; [Online]. Available from: http://prinseq.sourceforge.net/Raw_data_processing.pdf.

223. *catadapt User guide.* [Online]. Available from: http://cutadapt.readthedocs.org/en/stable/guide.html.

224. *ubuntu manuals.* Available from: http://manpages.ubuntu.com/manpages/natty/man1/grep.1.html.

225. *cd-hit User's Guide.* [Online]. Available from: http://cd-hit.org/.

226. *MEME weight calculate.* [Online]. Available from: http://meme-suite.org/doc/fasta-format.html.

227. Bailey, T.L., et al., *MEME SUITE: tools for motif discovery and searching.* Nucleic acids research, 2009: p. gkp335.

228. *MEME Suite.* [Online]. Available from: http://meme-suite.org/doc/bfile-format.html?man_type=web.

229. Gupta, S., et al., *Quantifying similarity between motifs.* Genome Biology, 2007. **8**(2): p. R24.

230. Mahony, S. and P.V. Benos, *STAMP: a web tool for exploring DNA-binding motif similarities.* Nucleic acids research, 2007. **35**(suppl 2): p. W253-W258.

231. Haralampidis, K., et al., *Combinatorial Interaction of Cis Elements Specifies the Expression of the Arabidopsis AtHsp90-1Gene.* Plant Physiology, 2002. **129**(3): p. 1138-1149.

232. Liu, Q. and H. Zhang, *Molecular Identification and Analysis of Arsenite Stress-Responsive miRNAs in Rice.* Journal of Agricultural and Food Chemistry, 2012. **60**(26): p. 6524-6536.

233. Reidt, W., et al., *Gene regulation during late embryogenesis: the RY motif of maturation‐specific gene promoters is a direct target of the FUS3 gene product.* The Plant Journal, 2000. **21**(5): p. 401-408.

234. Caiyin, Q., et al., *Isolation and sequencing analysis on the seed-specific promoter from soybean.* Frontiers of Agriculture in China, 2007. **1**(1): p. 17-23.

235. Becker, M.G., et al., *Genomic dissection of the seed.* Advances in Seed Biology, 2015: p. 62.

236. Chen, F., et al., *Arabidopsis phytochrome A directly targets numerous promoters for individualized modulation of genes in a wide range of pathways.* The Plant Cell, 2014. **26**(5): p. 1949-1966.

237. Peng, F.Y. and R.J. Weselake, *Gene coexpression clusters and putative regulatory elements underlying seed storage reserve accumulation in Arabidopsis.* BMC genomics, 2011. **12**(1): p. 1.

238. Green, P.J., S.A. Kay, and N.-H. Chua, *Sequence-specific interactions of a pea nuclear factor with light-responsive elements upstream of the rbcS-3A gene.* The EMBO journal, 1987. **6**(9): p. 2543.

239. Tremousaygue, D., et al., *Plant interstitial telomere motifs participate in the control of gene expression in root meristems.* The Plant Journal, 1999. **20**(5): p. 553-561.

240. Mallory, A.C., D.P. Bartel, and B. Bartel, *MicroRNA-directed regulation of Arabidopsis AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes.* The Plant Cell, 2005. **17**(5): p. 1360-1375.

241. Liu, P.P., et al., *Repression of AUXIN RESPONSE FACTOR10 by microRNA160 is critical for seed germination and post‐germination stages.* The Plant Journal, 2007. **52**(1): p. 133-146.

242. Williams, L., et al., *Regulation of Arabidopsis shoot apical meristem and lateral organ formation by microRNA miR166g and its AtHD-ZIP target genes.* Development, 2005. **132**(16): p. 3657-3668.

243. Huang, C.-H., et al., *Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution.* Molecular Biology and Evolution, 2016. **33**(2): p. 394-412.

244. Beilstein, M.A., I.A. Al-Shehbaz, and E.A. Kellogg, *Brassicaceae phylogeny and trichome evolution.* American journal of botany, 2006. **93**(4): p. 607-619.

245. Colangelo, E.P. and M.L. Guerinot, *The essential basic helix-loop-helix protein FIT1 is required for the iron deficiency response.* The Plant Cell, 2004. **16**(12): p. 3400-3412.

246. Heim, M.A., et al., *The basic helix–loop–helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity.* Molecular Biology and Evolution, 2003. **20**(5): p. 735-747.

247. Toledo-Ortiz, G., E. Huq, and P.H. Quail, *The Arabidopsis basic/helix-loop-helix transcription factor family.* The Plant Cell, 2003. **15**(8): p. 1749-1770.

248. Carretero-Paulet, L., et al., *Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in Arabidopsis, poplar, rice, moss, and algae.* Plant physiology, 2010. **153**(3): p. 1398-1412.

249. Solano, R., et al., *Nuclear events in ethylene signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSE-FACTOR1.* Genes & Development, 1998. **12**(23): p. 3703-3714.

250. Peng, J., et al., *Salt-induced stabilization of EIN3/EIL1 confers salinity tolerance by deterring ROS accumulation in Arabidopsis.* Plos Genetics, 2014. **10**(10): p. e1004664.

251. van Helden, J., *Evaluation of phylogenetic footprint discovery for predicting bacterial cis-regulatory elements and revealing their evolution.* BMC bioinformatics, 2008. **9**(1): p. 1.

252. Yu, D., C. Chen, and Z. Chen, *Evidence for an important role of WRKY DNA binding proteins in the regulation of NPR1 gene expression.* The Plant Cell, 2001. **13**(7): p. 1527-1540.

253. Wang, H., et al., *Arabidopsis WRKY45 transcription factor activates PHOSPHATE TRANSPORTER1; 1 expression in response to phosphate starvation.* Plant physiology, 2014. **164**(4): p. 2020-2029.

254. Liu, L., et al., *W-box and G-box elements play important roles in early senescence of rice flag leaf.* Scientific reports, 2016. **6**.

255. Chen, M.-X., et al., *Strong seed-specific protein expression from the Vigna radiata storage protein 8SGα promoter in transgenic Arabidopsis seeds.* Journal of biotechnology, 2014. **174**: p. 49-56.

256. Ulmasov, T., G. Hagen, and T.J. Guilfoyle, *Dimerization and DNA binding of auxin response factors.* The Plant Journal, 1999. **19**(3): p. 309-319.

257. Kasuga, M., et al., *Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor.* Nature biotechnology, 1999. **17**(3): p. 287-291.

258. Tamura, K., et al., *MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0.* Molecular Biology and Evolution, 2007. **24**(8): p. 1596-1599.

259. Hall, B.G., *Building phylogenetic trees from molecular data with MEGA.* Molecular Biology and Evolution, 2013: p. mst012.

260. Zeng, L., et al., *Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times.* Nature Communications, 2014. **5**.

261. Yang, Y., et al., *Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing.* Molecular Biology and Evolution, 2015: p. msv081.

262. Cui, X., et al., *Genomic analysis of rice microRNA promoters and clusters.* Gene, 2009. **431**(1): p. 61-66.

263. Zhou, X., et al., *Characterization and identification of microRNA core promoters in four model species.* PLoS Computational Biology, 2007. **3**(3): p. e37.

264. Devi, S.R., et al., *Identification of abiotic stress miRNA transcription factor binding motifs (TFBMs) in rice.* Gene, 2013. **531**(1): p. 15-22.

265. Mironova, V.V., et al., *Computational analysis of auxin responsive elements in the Arabidopsis thaliana L. genome.* Bmc Genomics, 2014. **15**(Suppl 12): p. S4.

266. Berendzen, K.W., et al., *Bioinformatic cis-element analyses performed in Arabidopsis and rice disclose bZIP- and MYB-related binding sites as potential AuxRE-coupling elements in auxin-mediated transcription.* Bmc Plant Biology, 2012. **12**.

267. Sun, Y., et al., *Integration of brassinosteroid signal transduction with the transcription network for plant growth regulation in Arabidopsis.* Developmental cell, 2010. **19**(5): p. 765-777.

268. Oh, E., J.-Y. Zhu, and Z.-Y. Wang, *Interaction between BZR1 and PIF4 integrates brassinosteroid and environmental responses.* Nature cell biology, 2012. **14**(8): p. 802-809.

269. Karimi, M., D. Inzé, and A. Depicker, *GATEWAY™ vectors for Agrobacterium-mediated plant transformation.* Trends in Plant Science, 2002. **7**(5): p. 193-195.

270. Doyle, J.J., *A rapid DNA isolation procedure for small quantities of fresh leaf tissue.* Phytochem bull, 1987. **19**: p. 11-15.

271. Cullings, K., *Design and testing of a plant‑specific PCR primer for ecological and evolutionary studies.* Molecular Ecology, 1992. **1**(4): p. 233-240.

272. Arteca, R.N. and J.M. Arteca, *Heavy-metal-induced ethylene production in Arabidopsis thaliana.* Journal of Plant Physiology, 2007. **164**(11): p. 1480-1488.

273. *SeedGenes Tutorial.* [Online]. Available from: http://www.seedgenes.org/Tutorial.html.

274. Malamy, J.E. and P.N. Benfey, *Organization and cell differentiation in lateral roots of Arabidopsis thaliana.* Development, 1997. **124**(1): p. 33-44.

275. Cuperus, J.T., et al., *Identification of MIR390a precursor processing-defective mutants in Arabidopsis by direct genome sequencing.* Proceedings of the National Academy of Sciences, 2009. **107**(1): p. 466-471.

276. Conn, S.J., et al., *Protocol: optimising hydroponic growth systems for nutritional and physiological analysis of Arabidopsis thaliana and other plants.* Plant Methods, 2013. **9**(1): p. 1-11.

277. Hazen, S.P., et al., *Exploring the transcriptional landscape of plant circadian rhythms using genome tiling arrays.* Genome Biology, 2009. **10**(2): p. R17.

278. Zhao, X. and L. Li, *Comparative analysis of microRNA promoters in Arabidopsis and rice.* Genomics, proteomics & bioinformatics, 2013. **11**(1): p. 56-60.

279. Warthmann, N., et al., *Comparative analysis of the MIR319a microRNA locus in Arabidopsis and related Brassicaceae.* Molecular biology and evolution, 2008. **25**(5): p. 892-902.

280. Haas, K.L. and K.J. Franz, *Application of metal coordination chemistry to explore and manipulate cell biology.* Chemical reviews, 2009. **109**(10): p. 4921-4960.

281. Waters, B.M. and L.C. Armbrust, *Optimal copper supply is required for normal plant iron deficiency responses.* Plant signaling & behavior, 2013. **8**(12): p. e26611.

282. Puig, S., et al., *Copper and iron homeostasis in Arabidopsis: responses to metal deficiencies, interactions and biotechnological applications.* Plant, cell & environment, 2007. **30**(3): p. 271-290.

283. Wu, H., et al., *Co-overexpression FIT with AtbHLH38 or AtbHLH39 in Arabidopsis-enhanced cadmium tolerance via increased cadmium sequestration in roots and improved iron homeostasis of shoots.* Plant physiology, 2012. **158**(2): p. 790-800.

284. Yang, T.J., W.-D. Lin, and W. Schmidt, *Transcriptional profiling of the Arabidopsis iron deficiency response reveals conserved transition metal homeostasis networks.* Plant physiology, 2010. **152**(4): p. 2130-2141.

Appendix

1. Supplementary tables and figures

Table S0-1 species names and abbreviations

species ID	Corresponding name	Abbreviation
sp01	*Berteroa incana*	*Bin*
sp02	*Sisymbrium officinale*	*Sof*
sp03	*Neslia paniculata*	*Npa*
sp04	*Matthiola valesiaca*	*Mva*
sp05	*Iberis amara*	*Iam*
sp06	*Thellungiella halophila*	*Tha*
sp07	*Arabis alpina*	*Aal*
sp08	*Cardamine hirsuta*	*Chi*
sp09	*Noccea praecox*	*Npr*
sp10	*Arabidopsis halleri*	*Aha*
sp11	*Brassica nigra*	*Bni*
sp12	*Lepidium perfoliatum*	*Lpe*
sp13	*Aethionema grandiflora*	*Agr*
sp14	*Thlaspi arvense*	*Tar*
sp15	*Descurainia sophia*	*Dso*
sp16	*Cardamine impatiens*	*Cim*
sp17	*Arabidopsis thaliana*	*Ath*

Table S0-2 Primers used for MIR390a and MIR390b expression studies.

Gene	Primer	Description	Sequence 5' to 3'
MIR390a	AthPT-F	Forward primer for *pMIR390a::GFP-GUS* construct	CACCGGTGGCCAAACTCTTGATCGTATTC
	AthPT-R	Reverse primer for *pMIR390a::GFP-GUS* construct	CTCTACTTTGTTTTTTTGGGTTGTG
MIR390b	AthPT-F	Forward primer for *pMIR390a::GFP-GUS* construct	CACCGCATGCAGATTATTCTCTCAGAT
	AthPT-R	Reverse primer for *pMIR390a::GFP-GUS* construct	GCTATTTCTCTACTACTCAATCTG

Table S0-3 Primers for Site Specific mutagenesis constructs

Gene	Primer	Description	Sequence 5' to 3'
MIR390a_m1	MIR390a_m1_1	Forward primer for first motif of putative TATA-box *mutagenesis*	CGAACCCGAGTTTTGTTCGTCCACAGACGGCACCTTCTCTCTCCTTCTTCC
MIR390a_m1	MIR390a_m1_2	Reverse primer for first motif of putative TATA-box *mutagenesis*	GGAAGAAGGAGAAGAAGAAGGTGCCGTCTGTGGACGAACAAAACTCGGGTTCG
MIR390a_m2	MIR390a_m2_1	Forward primer for first motif of putative TSS *mutagenesis*	CACGTTTAACGAAGAGGAGGTTGGTGAAGTTCCTTCGAACCCGAGTTTTG
MIR390a_m2	MIR390a_m2_2	Reverse primer for first motif of putative TSS *mutagenesis*	CAAAACTCGGGTTCGAAGGAACTTCACCAACCTCCTCTTCGTTAAACGTG
MIR390a_m3	MIR390a_m3_1	Forward primer for first motif of putative iron responsive E-box *mutagenesis*	CAATCAACGGGATCCAATAAAGCATCACGGAACTATATAGTATCAATAC
MIR390a_m3	MIR390a_m3_2	Reverse primer for first motif of putative iron responsive E-box	GTATTGATACTATATAGTTCCGTGATGCTTTATTGGATCCCGTTGATTG
MIR390a_m4	MIR390a_m4_1	Forward primer for first motif of putative TATA-box *mutagenesis*	GAAGAAACTAAAGTCAAACAATCAAGTAGAAGCGATAAATCACATCTGGACTATATAG
MIR390a_m4	MIR390a_m4_2	Reverse primer for first motif of putative TATA-box *mutagenesis*	CTATATAGTCCAGATGTGATTTATCGCTTCTACTTGATTGTTTGACTTTAGTTTCTTC
MIR390a_m5	MIR390a_m5_1	Forward primer for first motif of putative TATA-box *mutagenesis*	GGCCTCAGAAGAAGAAACTAAATGGAGACAATCAACGGGATCCAATAAATC
MIR390a_m5	MIR390a_m5_2	Reverse primer for first motif of putative TATA-box *mutagenesis*	GATTTATTGGATCCCGTTGATTGTCTCCATTTAGTTTCTTCTTCTGAGGCC
MIR390a_m6	MIR390a_m6_1	Forward primer for first motif of putative TATA-box *mutagenesis*	GCATTATCATGTATTTTGAGCATTATCATGTATTTTGATGTGAATCAAAAGGCCTCAGAAG
M390a_m6	M390a_m6_2	Forward primer for first motif of putative TATA-box *mutagenesis*	CTTCTGAGGCCTTTTGATTCACATCAAAATACATGATAATGCTCAAAATACATGATAATGC
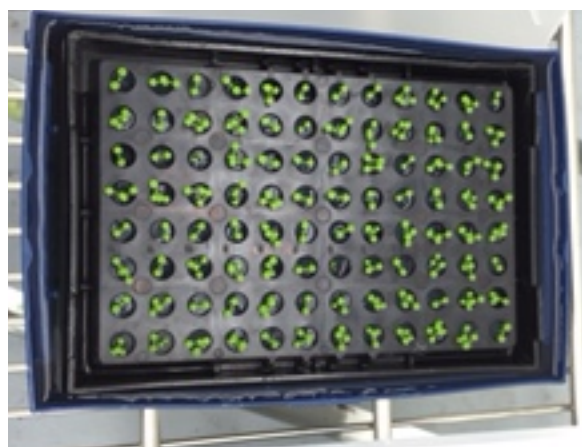
Figure S1 Hydroponic system in-house designed with StarLab Tip Rack and MicrAmp® Black 96-well Base

Table S0-4 summary of the miRNAs (by family) detected

MiRNA Gene	Nr. [2] of Species	Nr. of Seq[3]	Seq/Spe[4] Ratio	Average length of Seq	MiRNA Gene	Nr. [2] of Species	Nr. of Seq[3]	Seq/Spe[4] Ratio	Average length of Seq
MIR156	17	43	2.53	994.14	MIR414	4	4	1.00	799.75
MIR160	17	50	2.94	1003.46	MIR4239	4	4	1.00	802.75
MIR162	17	37	2.18	908.05	MIR5631	4	4	1.00	894.50
MIR166	17	93	5.47	955.33	MIR5638	4	6	1.50	968.50
MIR169	17	124	7.29	947.14	MIR5644	4	4	1.00	833.50
MIR171	17	39	2.29	925.59	MIR5660	4	4	1.00	838.00
MIR172	17	42	2.47	751.62	MIR5998	4	11	2.75	786.64
MIR319	17	61	3.59	843.90	MIR826	4	4	1.00	840.75
MIR395	17	72	4.24	930.00	MIR836	4	4	1.00	1014.50
MIR167	16	32	2.00	1034.66	MIR845	4	6	1.50	896.83
MIR168	16	28	1.75	995.00	MIR173	3	3	1.00	1294.00
MIR399	16	42	2.63	890.79	MIR1888	3	5	1.67	982.00
MIR396	14	21	1.50	899.43	MIR2111	3	4	1.33	1177.25
MIR159	12	21	1.75	1015.81	MIR2936	3	3	1.00	922.33
MIR164	12	19	1.58	1084.63	MIR2938	3	3	1.00	902.00
MIR170	12	17	1.42	990.88	MIR398	3	6	2.00	1609.17
MIR390	12	21	1.75	861.95	MIR405	3	5	1.67	1169.40
MIR394	12	20	1.67	823.45	MIR4227	3	5	1.67	786.60
MIR165	11	13	1.18	1047.08	MIR4228	3	3	1.00	879.67
MIR393	11	26	2.36	1001.42	MIR447	3	5	1.67	1278.80

MIR5645	11	20	1.82	1110.90	MIR5020	3	5	1.67	1309.40
MIR391	10	11	1.10	929.36	MIR5029	3	4	1.33	880.25
MIR397	10	14	1.40	1115.29	MIR5630	3	4	1.33	1178.75
MIR5655	9	15	1.67	691.33	MIR5651	3	3	1.00	1167.00
MIR157	8	12	1.50	1013.67	MIR5658	3	3	1.00	904.67
MIR5643	8	14	1.75	766.64	MIR5664	3	3	1.00	840.67
MIR827	8	11	1.38	1171.91	MIR5997	3	3	1.00	909.33
MIR855	8	46	5.75	635.76	MIR773	3	4	1.33	1073.50
MIR408	7	9	1.29	1079.56	MIR828	3	3	1.00	937.33
MIR5634	7	7	1.00	673.57	MIR830	3	3	1.00	946.00
MIR5642	7	9	1.29	1403.67	MIR831	3	3	1.00	986.33
MIR824	7	11	1.57	1083.82	MIR832	3	4	1.33	1005.75
MIR401	6	13	2.17	714.38	MIR837	3	3	1.00	970.67
MIR5014	6	7	1.17	820.00	MIR842	3	5	1.67	842.60
MIR5027	6	7	1.17	707.14	MIR843	3	3	1.00	986.67
MIR857	6	7	1.17	924.86	MIR852	3	3	1.00	1045.33
MIR2937	5	6	1.20	826.67	MIR856	3	3	1.00	963.67
MIR5015	5	11	2.20	668.18	MIR858	3	4	1.33	1087.00
MIR5635	5	8	1.60	1204.63	MIR860	3	3	1.00	908.33
MIR5641	5	6	1.20	809.83	MIR863	3	3	1.00	1041.00
MIR5665	5	6	1.20	840.00	MIR868	3	3	1.00	869.33
MIR161	4	4	1.00	1185.25	MIR869	3	4	1.33	935.50
MIR3932	4	6	1.50	980.67	MIR870	3	3	1.00	863.00

[1]MIR, microRNA; [2]Nr., number; [3]Sep, species; [4]Seq, sequence. The microRNAs were listed based on the number of species that they were detected.

2. Motif logos of miR160a

Motif Similarity Matches

miR160a-M1

forward   reverse compliment

Name	E value	Alignment	Motif
AG_v3	3.9411e-04	GGAGAAAKRG CCAAAAATGG	
AP1	3.9411e-04	GGAGAAAKRG CTAAAAATGG	
AP1_v2	3.9411e-04	GGAGAAAKRG CCAAAAATGG	
AG_v2	8.9170e-04	−GGAGAAAKRG ACTAAAAATGG	
ARF	1.8844e-03	GGAGAAAKRG −GAGACA−−−	

Figure S2 Sequence similarity alignment of miR160a-M1 by STAMP. Motifs were detected by MEME and matched to 5 known motifs in AGRIS database.

miR160a-M2

forward   reverse compliment

Name	E value	Alignment	Motif
RY-repeat	2.5701e-08	GATWTGCATG --CATGCATG	
SORLREP5	2.6077e-04	---CATGCAWATC AGTCATGCAA---	
L1-box	1.0669e-03	CATGCAWATC --TRCATTTA	
SORLREP3	2.0771e-03	CATGCAWATC- --TGTATATAT	
Ibox	2.0850e-03	CATGCAWATC ----CTTATC	

Figure S3 Sequence similarity alignment of miR160a-M2 by STAMP. Motifs were detected by MEME and matched to 5 known motifs in AGRIS database.

129

miR160a-M3

	forward	reverse compliment

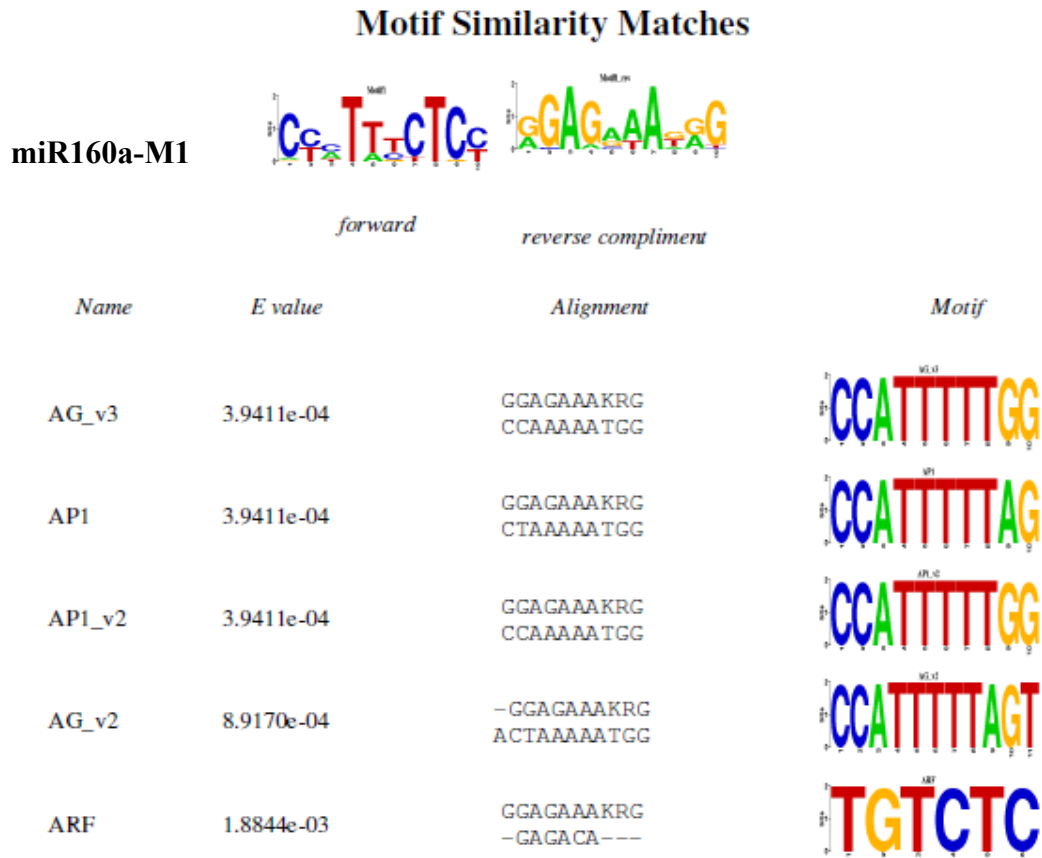| Name	E value	Alignment	Motif
SORLIP5	5.9209e-04	TCTCTSAATC ---CTCACTC	
SORLREP4	3.2555e-03	GATTSAGAGA AATTAGGAG-	
PRHA	4.6480e-03	-GATTSAGAGA---- TAATTGAGTCAATTA	
ARF	5.9700e-03	TCTCTSAATC TGTCTC----	
ARF1	5.9700e-03	TCTCTSAATC TGTCTC----	

Figure S4 Sequence similarity alignment of miR160a-M3 by STAMP. Motifs were detected by MEME and matched to 5 known motifs in AGRIS database.
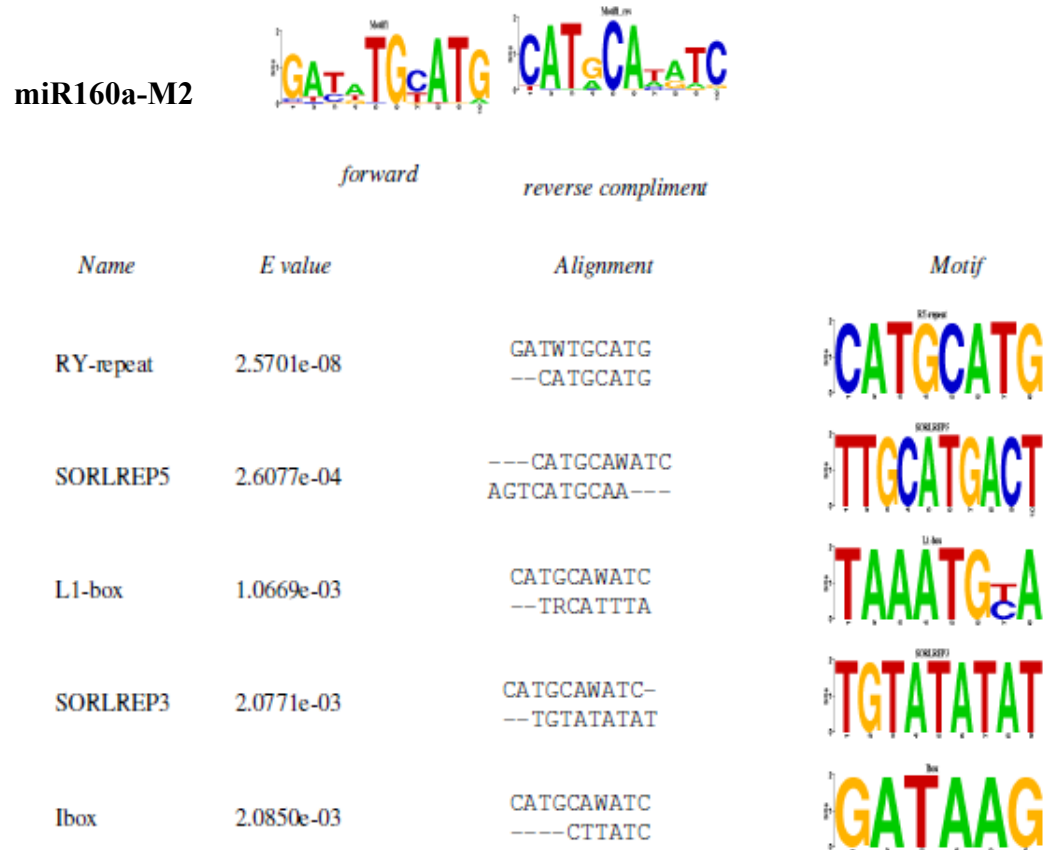
miR160a-M4

forward   reverse compliment

Name	E value	Alignment	Motif
GT	5.7262e-04	------ACCCAAACAC CATATTAACCACACA-	
TELO-box	1.0884e-03	GTGTTTGGGT-- ---TTAGGGTTT	
AG_v3	1.2303e-03	--GTGTTTGGGT CCATTTTTGG--	
AP1_v2	1.2303e-03	--GTGTTTGGGT CCATTTTTGG--	
MYB	2.6334e-03	GTGTTTGGGT --GKTWGGTK	

Figure S5 Sequence similarity alignment of miR160a-M4 by STAMP. Motifs were detected by MEME and matched to 5 known motifs in AGRIS database.
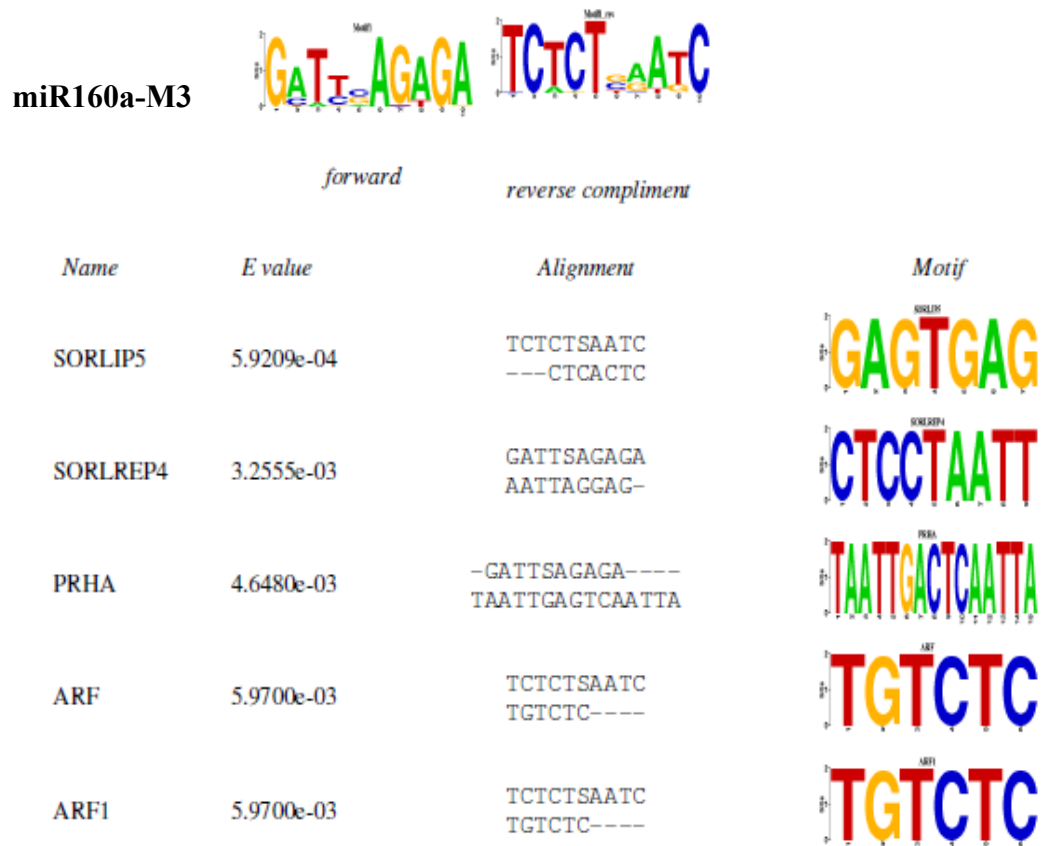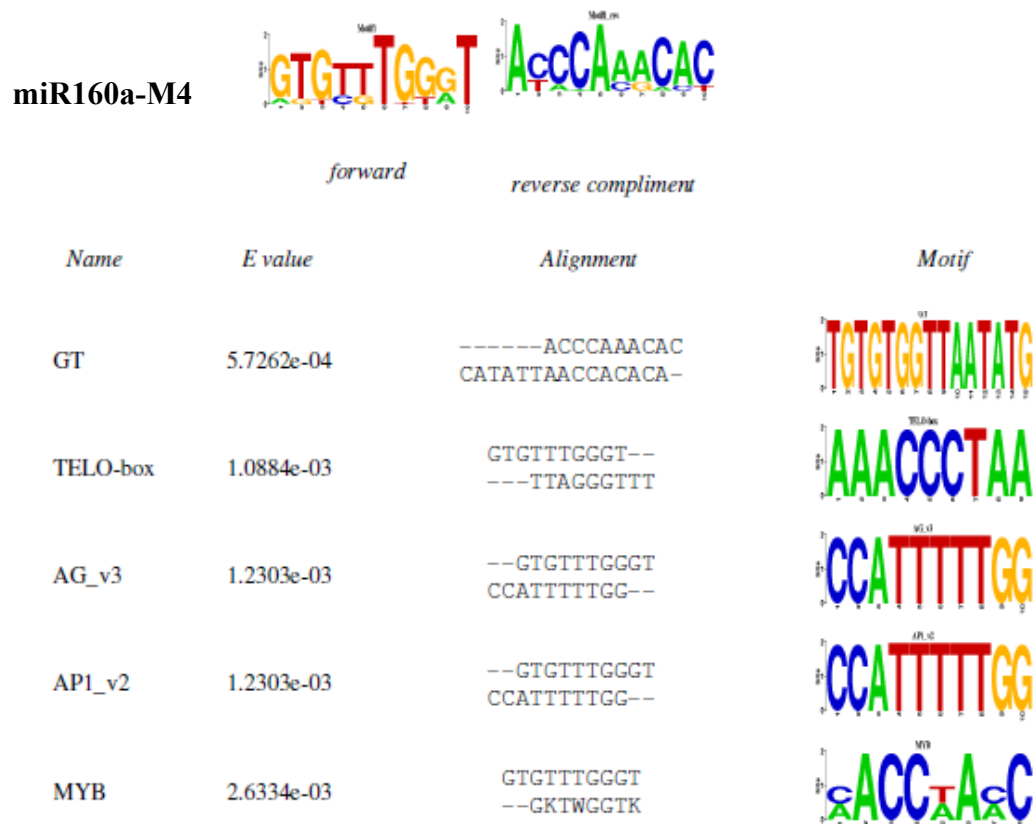
miR160a-M5

forward   reverse compliment

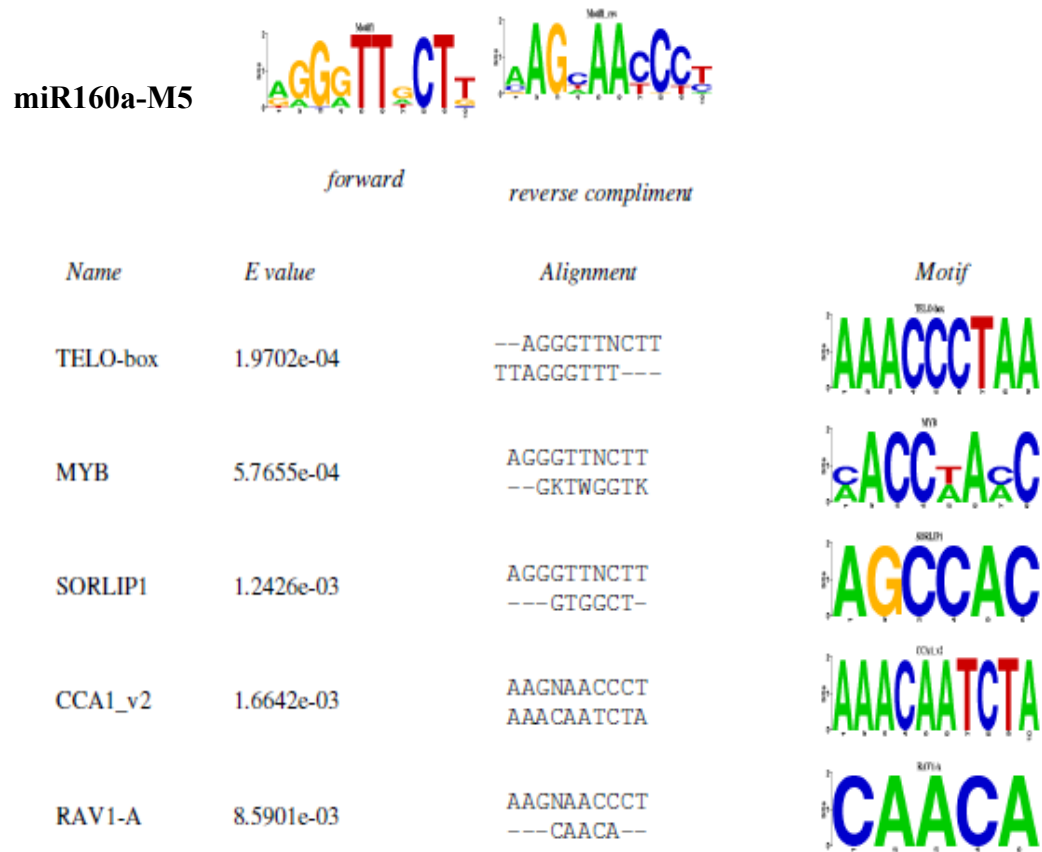| Name	E value	Alignment	Motif
TELO-box	1.9702e-04	--AGGGTTNCTT TTAGGGTTT---	
MYB	5.7655e-04	AGGGTTNCTT --GKTWGGTK	
SORLIP1	1.2426e-03	AGGGTTNCTT ---GTGGCT-	
CCA1_v2	1.6642e-03	AAGNAACCCT AAACAATCTA	
RAV1-A	8.5901e-03	AAGNAACCCT ---CAACA--	

Figure S6 Sequence similarity alignment of miR160a-M5 by STAMP. Motifs were detected by MEME and matched to 5 known motifs in AGRIS database.