TESTING THE COMPLEMENTARITY REQUIREMENTS OF PLANT MICRORNA-TARGET INTERACTIONS

A Dissertation in

Plant Biology

by

Qikun Liu

© 2014 Qikun Liu

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2014
The dissertation of Qikun Liu was reviewed and approved* by the following:

Michael J. Axtell  
Associate Professor of Biology  
Dissertation Advisor  
Chair of Committee

Sarah M. Assmann  
Waller Professor of Biology

Ying Gu  
Assistant Professor of Biochemistry & Molecular Biology

Surinder Chopra  
Professor of Agricultural Sciences

The-hui Kao  
Chair, Intercollege Graduate Degree Program in Plant Biology

*Signatures are on file in the Graduate School
ABSTRACT

Small RNAs are a group of regulatory RNAs of 20-30 nucleotides (nt) that are involved in diverse cellular processes across multiple eukaryotic kingdoms, including animals, plants and fungi. In complex with Argonaute (AGO) effector proteins, they recognize target RNA transcripts (coding and non-coding) based on sequence complementarity, and function as negative regulators at both transcriptional and post-transcriptional levels.

It is widely accepted that plant microRNA-target interactions usually require a high degree of complementarity to trigger target cleavage. However, the sequence requirements for target repression at protein level through slicing independent mechanisms have never been systematically studied.

Utilizing a self-designed dual-luciferase reporter system, I systematically examine the complementarity requirements for microRNA function through *Nicotiana benthamiana* transient assay. I found that changes of the mRNA accumulation accounts for almost all observed regulatory effects. Comparison among those naturally occurring targets indicated that mismatches located near the miRNA 5’ends were more disruptive to miRNA function than those occurred near miRNA 3’ends. Actually, certain naturally occurring targets with unpaired bases only present at the miRNA 3’ ends can even carry stronger regulatory efficiency than perfectly paired sites when position in 3’-untranslated region (UTR), but not open reading frame (ORF). I found that this was largely due to different behavior of perfectly paired sites, when placed in different contexts (ORF vs. 3’-UTR). While up to 3 mismatches could be tolerated at the 3’ end without affecting the target efficacy, base pairing patterns that are typical of the known configuration of animal microRNA-target duplex did not trigger any detectable level of target down-regulation.

Through collaboration we identified a non-canonical plant miRNA-target pairing pattern, where a 6-nt bulge on the target in between nucleotide 6 and 7 relative to the miRNA was tolerated. I demonstrated that this is very unique in terms of both the miRNA involved and the position of the bulge.

*De novo* annotation and quantification of small RNA producing loci were also carried out in *N. benthamiana*. A total of 43652 small RNA producing loci, including 177 MIRNA, 4573 hairpin RNA (hpRNA), and 38902 siRNA loci were identified. 95 MIRNA loci belonging to 37 plant known microRNA families were found. Comparing to the novel MIRNAs, these known MIRNA genes are processed with higher accuracy, and expressed more abundantly. Although small RNA clusters mainly producing 23 to 24 nt siRNA dominated hpRNA and siRNA loci, and 24-nt small RNAs are the most abundant small RNA species, 20 to 21 nt siRNAs are more robustly generated from certain highly active loci. Loci that give rise to longer small RNAs (23-24 nts) are generally less repetitive than those producing short small RNAs (20-22 nts). This is true for all types of annotated loci (MIRNA, hpRNA, and siRNA). MIRNA loci are enriched in genic regions and depleted in repeat regions, whereas siRNA loci, especially 23-24 nt groups, tend to occupy regions enriched for repetitive elements.
TABLE OF CONTENTS

LIST OF FIGURES ............................................................................................................................. vi

LIST OF TABLES ............................................................................................................................... viii

ACKNOWLEDGEMENTS .................................................................................................................. ix

Chapter 1 Introduction ......................................................................................................................1

1.1 Overview of small RNAs ..............................................................................................................1
1.2 Small RNAs in animals ................................................................................................................1
1.3 Small RNAs in plants ..................................................................................................................4
1.4 miRNA function in animals and plants ......................................................................................7
1.5 Objectives ................................................................................................................................10

Chapter 2 Developing a dual-luciferase transient expression system to quantify miRNA-target interactions .........................................................................................................................12

2.1 Introduction ..............................................................................................................................12
2.2 Materials ................................................................................................................................13
  2.2.1 Agro-infiltration components ...............................................................................................13
  2.2.2 Dual-luciferase assay components .......................................................................................13
  2.2.3 RNA extraction, reverse transcription, and qRT-PCR reagent ............................................13
2.3 Methods ..................................................................................................................................13
  2.3.1 Plant growth condition .........................................................................................................13
  2.3.2 Dual-luciferase reporters for miRNA .....................................................................................14
  2.3.3 Agro-infiltration ....................................................................................................................14
  2.3.4 Dual-luciferase Assay ...........................................................................................................15
  2.3.5 Dual-luciferase Assay Data Analysis .....................................................................................18
  2.3.6 qRT-PCR Assay ...................................................................................................................18
  2.3.7 qRT-PCR Assay Date Analysis .............................................................................................18
2.4 Notes .......................................................................................................................................19
2.5 Results .....................................................................................................................................23

Chapter 3 Quantitative analysis of plant microRNA-target interactions ........................................28

3.1 Introduction ..............................................................................................................................28
3.2 Materials and Methods ............................................................................................................29
  3.2.1 GUS-based Competitors .....................................................................................................29
  3.2.2 Dual-luciferase Sensors .......................................................................................................29
  3.2.3 miRNA over-expressers ........................................................................................................30
  3.2.4 Agro-infiltration and dual-luciferase assay at both the mRNA and protein level ..................30
  3.2.5 Small RNA-seq .....................................................................................................................30
  3.2.6 RNA secondary structure analyses ......................................................................................30
3.3 Results ........................................................................................................................................30
  3.3.1 Perfectly paired sites can behave differently in ORF vs. 3’-UTR ..................30
  3.3.2 The 5’ and 3’ regions of target sites are differentially tolerant of consecutive mismatches .............................................................................................................34
  3.3.3 Complementarity is a major determinant for site efficacy ..................................36
  3.3.4 Animal-like miRNA-target sites are ineffective ..................................................37
  3.3.5 5mCSD1 and 5mCSD2 are not functional miR398 targets .................................40
  3.3.6 Multiple pairing patterns with central mismatches are effective miRNA-target-mimics ............................................................ .................................................................42
  3.3.7 Blue Copper-binding Protein (BCBP) target site with a 6-nt bulge in target 5’region is repressed by miR398 ..........................................................46
  3.3.8 Summary of target site properties ...........................................................................53
  3.3.9 Changes in target site secondary structures are unlikely to have affected our measurements ...........................................................................................................57

3.4 Discussion ..................................................................................................................................60

Chapter 4 ShortStack de novo annotation and quantification of small RNA producing loci in N. benthamiana genome ..............................................................................63

  4.1 Introduction .............................................................................................................................63
  4.2 Data analysis ............................................................................................................................64
  4.3 Results ........................................................................................................................................64
    4.3.1 Overview of small RNA analysis ..................................................................................64
    4.3.2 Repetitiveness of N. benthamiana small RNA genes ................................................66
    4.3.3 Genomic features of N. benthamiana small RNA genes ...........................................68
    4.3.4 Identification of MIRNAs with and without known plant homologs 70
  4.4 Discussion ................................................................................................................................72

Chapter 5 Summary and prospects ..................................................................................................74

  5.1 Summary ...................................................................................................................................74
  5.2 Prospects ....................................................................................................................................75

Appendix ...........................................................................................................................................76

References ........................................................................................................................................82
LIST OF FIGURES

Figure 2.1 One-month old *Nicotiana benthamiana* plants ................................................................. 24

Figure 2.2 Schematic diagrams of dual-luciferase sensors ................................................................. 24

Figure 2.3 Conventions for miRNA-target site display and nomenclature ........................................... 25

Figure 2.4 Quantification of naturally occurring target sites of miR164 and miR398 using 3’-UTR sensor .............................................................................................................................................. 26

Figure 2.5 Quantification of naturally occurring target sites of miR164 and miR398 using ORF sensor .............................................................................................................................................. 27

Figure 3.1 Naturally occurring miR156, miR160, and miR165 target sites in ORF sensors ................................................................. 32

Figure 3.2 Naturally occurring miR156, miR160, and miR165 target sites in 3’-UTR sensors .............................................................................................................................................. 33

Figure 3.3 Differential sensitivity of the 5’ and 3’ target site regions to consecutive mismatches .......................................................................................................... 35

Figure 3.4 Complementarity patterns play a major role in target-site efficacy ........................................ 36

Figure 3.5 Little to no evidence for functionality of animal-like target sites ........................................ 38

Figure 3.6 Animal-like sites remain non-functional regardless of sensor expression level ............................................................................................................................................... 39

Figure 3.7 No evidence for functionality of 5mCSD1 or 5mCSD2 sites .............................................. 41

Figure 3.8 Multiple pairing patterns are effective miRNA-target-mimics .............................................. 44

Figure 3.9 *BCBP* is a functional miR398 target ................................................................................... 47

Figure 3.10 Multiple bulged sites show minimal target efficacy .......................................................... 48

Figure 3.11 *BCBP*-like bulged sites are not effective for multiple conserved miRNAs ......................... 49

Figure 3.12 Effects of 6-nt bulge present at different positions ............................................................ 51

Figure 3.13 Correlation of target site scores with protein-level efficacies ............................................ 54

Figure 3.14 Protein-level repression is more effective for ORF sites than for 3’-UTR sites .................. 55
Figure 3.15 No correlation between predicted target accessibility and protein-level target efficacy ........................................................................................................58

Figure 4.1 The fraction of loci with different Dicer-call ..........................................................................................65

Figure 4.2 Many N. benthamiana small RNA loci showed low repetitiveness ..................67

Figure 4.3 Analysis of co-occupancy between small RNA loci and other genomic features ........................................................................................................69

Figure 4.4 Number of MIRNA loci and families identified for both known and novel groups ........................................................................................................................................71

Figure 4.5 Expression level of known MIRNA loci identified ........................................71

Figure 4.6 20-22 nt N. benthamiana MIRNAs are processed more precisely than 23-24 nt MIRNAs ........................................................................................................................................72
LIST OF TABLES

Table 2.1 Loading template for Dual-Luciferase Assay......................................................... 25
Table 2.2 Identification of *N. benthamiana* miRNAs.......................................................... 25
Table 3.1 Identification of *N. benthamiana* miRNAs............................................................. 33
Table 4.1 Summary of *N. benthamiana* small RNA mapping............................................. 65
Table 4.2 Summary of *N. benthamiana* small RNA loci..................................................... 65
Table 4.3 Overlaps between small RNA loci and other genomic features by tallies............. 68
Table 4.4 Overlaps between small RNA loci and other genomic features by number of nucleotides...................................................................................................................... 69
ACKNOWLEDGEMENTS

I feel extremely grateful of having Dr. Michael J. Axtell as my Ph.D. advisor. He served as a role model demonstrating me the importance of hard working in becoming a young successful researcher. He encourages independence and sharing of inspiring thoughts. He never hides his happiness when he sees his students’ achievements, which keeps me motivated and excited to advance. I never feel hesitate to present negative results or mention concerns, because he offers great understanding and mentoring. He made sure the efficiency of accomplishing lab tasks and the accuracy in obtaining results always come first. He trains me to think critically, and logically by taking the action himself. He thinks for his student. I was often reminded for attending meetings and nominated for student awards.

I owe great thanks to my Dissertation committee, Dr. Sarah Assmann, Dr. Surinder Chopra, and Dr. Ying Gu for their constant supports during all these years. They always fueled my study with encouragements and inspiring discussions. They put in time meeting with me, discussing with me about my career choices. I am extremely grateful for their writing me recommendation letters with generous compliments, which jumpstarted me into my next stop. Dr. Gu also spend time helping me edit my job application. I also thank my previous committee member Dr. Hong Ma, who never hesitated to tell me his real thoughts and suggestions.

My lab mates are my wonderful friends. We take care of each other not only inside the lab but also as close friends in life. They are Feng Wang, Ceyda Coruh, Saima Shahid, Seth Polydor, and former lab members, Wan-Ching Lin, Jo Ann Snyder, Sung Hyun Cho, and Rajendran Rajeswaran. I would like to thank particularly Saima Shahid, who took hours answering my questions about bioinformatics analysis.

I would like to thank my wife Wan Yang for her lasting support, understanding, thoughtfulness and love. I would like to express my deep gratitude to my parents, who raised me up, offered me everything they have, and encouraged me pursue my dream oversea. If any achievement, I share with my family.
Chapter 1

Introduction

1.1 Overview of small RNAs

Small regulatory RNAs are produced through heterogenic pathways, and function as master gene regulators through a diverse modes of action (Axtell et al., 2011). They act at both transcription and post-transcriptional level (Tolia and Joshua-Tor, 2007; Chen, 2010; Jones-Rhoades et al., 2006). Small RNA expression profiles are highly dynamic, and are differentially regulated by many factors, including cell types, development stages, environmental signals, such as biotic and abiotic stresses (Narry Kim et al., 2009; Lu et al., 2005; Slotkin et al., 2009; Tanurdzic et al., 2008; Marí-Ordóñez et al., 2013). Interplays between small RNAs within and across different classes are not rare (Okamura et al., 2008; Cho et al., 2012; Marí-Ordóñez et al., 2013). Although accumulating discoveries of novel non-canonical small RNA species kept blurring the boundaries between groups, classification of small RNAs can be fairly successfully achieved based on following criteria: type of precursor (single- vs. double-stranded), machinery involved during biogenesis, type of targets and mode of action (transcriptional vs. post-transcriptional) (Ambros et al., 2003; Narry Kim et al., 2009; Axtell, 2013a).

1.2 Small RNAs in animals

In animals, small RNAs have been categorized into three major groups: microRNA (miRNA), endogenous small interfering RNA (endo-siRNA) and Piwi-interacting RNA (piRNA) (Narry Kim et al., 2009).

1.2.1 miRNA

At present, hundreds of individual miRNA species derived from tens of miRNA gene families have been documented in animals (Narry Kim et al., 2009). Like regular protein coding genes, precursors of miRNAs, termed pri-miRNAs, are transcribed by RNA Polymerase II (Pol-II) from their genomic loci, followed by capping on the 5’ends, and polyadenylation on the 3’ends (Lee et al., 2004a). A self-folding hairpin-like secondary structure, with the mature miRNA sequence embedded inside the stem region constitutes the most unique feature of a pri-miRNA (Pasquinelli et al., 2000; Ambros et al., 2003). The first key step of miRNA maturation takes place in the nucleus, where the local stem-loop structure is released from the long primary transcript through cutting at the stem region by the nuclear RNase III type protein, Drosha (Lee et al., 2003). Animal pri-miRNAs are of high homogeneity with the stem of often ~33-nt measured in length. A double-stranded (ds)RNA binding protein, Pasha, binds specifically at the dsRNA-ssRNA junction, a process that further anchors Drosha to produce a cutting signature (2-nt 3’overhang) ~11-nt above the junction (Han et al., 2004a; Denli et al., 2004; Gregory et al., 2004; Landthaler et al., 2004; Han et al., 2006).
Following Drosha processing, the product, termed pre-miRNA, is then exported from nucleus into cytoplasm by a member of nuclear transport receptor, Exportin-5 (EXP5), in a RanGTP-dependent manner (Lund et al., 2004; Bohnsack et al., 2004). Within the cytoplasm a second RNase III type protein, Dicer, secures the preexisting 2-nt 3’ overhang of the pre-miRNA through its PAZ (Piwi/Argonaute/Zwille) domain, and measures ~22 nts towards the loop region to make a second cut in a ATP-dependent manner, releasing the miRNA:miRNA* duplex (Bernstein et al., 2001; Grishok et al., 2001). As a result, Drosha and Dicer act sequentially to determine one and the other end of the mature miRNA. The two opponent strands unwound from the same miRNA/miRNA* duplex have different destinies, with one strand (miRNA) being loaded into a member of Argonaute-family proteins, guiding the downstream targeting, whereas the other strand is immediately destroyed (Hutvagner, 2001). Selection of guide strand over passenger strand is likely to be thermodynamically driven. The strand with its 5’ end being less stably paired tend to be loaded into AGO, the core catalytic engine of RNA-induced silencing complex (RISC) (Schwarz et al., 2003; Khororova et al., 2003; Han et al., 2006). Animals all contain multiple AGO proteins (Hutvagner and Simard, 2008). AGO proteins from different species are named differently, though they are largely functionally identical. For example, AGO1, ALG1/2, and AGO2 are involved in miRNA functioning pathway in Drosophila melanogaster, Caenorhabditis elegans, and human, respectively (Liu et al., 2004; Meister et al., 2004; Hutvagner and Simard, 2008).

Selection of AGO proteins for small RNA loading is largely determined by the duplex structure (Steiner et al., 2007). Specifically, in D. melanogaster miRNA:miRNA* duplex that is typical of having mismatches and internal loops is directed to AGO1 pathway, whereas perfect siRNA duplex is directed to AGO2 (Tomari et al., 2007; Förstemann et al., 2007; Azuma-Mukai et al., 2008). It should also be noted that, more than being merely a byproduct during miRNA biogenesis, miRNA star is sometimes also loaded into AGO protein, and performs function in target repression (Ghildiyal et al., 2009; Czech et al., 2010).

1.2.2 Endo-siRNAs

Unlike that of miRNA, the definition of endo-siRNA is rather a general term, which includes a broad region of endogenously derived siRNAs that are of varied genomic origin and biogenesis pathway (Okamura and Lai, 2008). However, they share two main signatures in common: First, the precursors are base paired, either inter-molecularly or intra-molecularly; Second, they function through AGO2-mediated endonucleolytic cleavage (Okamura and Lai, 2008; Kawamura et al., 2008). Core endo-siRNA producing enzymes include Dicer2 and AGO2 (Lee et al., 2004b; Okamura et al., 2004).

RNAi has long been used for specifically repressing genes of interest with exogenously introduced dsRNA precursors, where endogenous small RNA processing machinery was hijacked to produce sequence specific siRNA (Liu et al., 2004). During a screen for C.elegans mutants that are defective in RNAi, increased mobilization of endogenous transposon was observed, which suggested the role of endogenous siRNA in suppressing transposon activity (Tabara et al., 1999). It later became clear that read-through
transcripts of activated transposon form intra-molecular dsRNA at terminal inverted repeats giving rise to these siRNAs (Sijen and Plasterk, 2003).

Traditionally, extremely long hairpin structure was omitted from in-depth analysis, based on empirical knowledge that sizes of stem-loops giving rise to canonical miRNAs are uniformly less than 100 nts. In addition to inverted-repeat associate siRNA, long hairpin structures (up to 400 base pairs) can also generate endo-siRNAs (Ruby et al., 2007). Interestingly, phased siRNA are produced periodically in a ~22-nt cycle, indicating a progressive processing pattern carried out by Dicer2 (Okamura and Lai, 2008; Kawamura et al., 2008).

A third subtype of endo-siRNA is produced upon convergent transcriptions of overlapped gene units encoded on opposite DNA strands (Czech et al., 2008). Transcripts in such arrangement, termed cis-natural antisense transcripts (cis-NAT), are commonly seen in both animal and plant genomes, with only a subset of them giving birth to siRNAs. In most cases, regions that produce siRNAs are confined within the 3’-UTRs of the co-expressed genes (Czech et al., 2008). Introns are excluded from the siRNA-producing regions.

Similarly, RNA duplexes formed between co-expressed pseudogene-antisense and gene-sense transcripts (trans-NAT) constitute another source of endo-siRNAs (Tam et al., 2008; Watanabe et al., 2008). Mutations in Dicer and AGO2 significantly increased the expression level of genes that are targeted by endo-siRNAs generated from pseudogene antisense transcripts (Tam et al., 2008; Watanabe et al., 2008). The data suggests that a subset of pseudogenes is still functional, and should not be treated as evolutionary remnants.

1.2.3 PIWI-interacting RNAs

piRNAs are animal germline specific 24- to 29-nt small RNAs (Aravin et al., 2001; 2003). They play critical roles in remaining chromosome integrity during gametogenesis by suppressing transposable element activities (Sarot et al., 2004; Kalmykova et al., 2005). The name, PIWI-interacting RNA, comes from the fact that small RNAs from this clade interact particularly with a germ-cell specific subset of Argonaute proteins, such as PIWI, Aubergine, and AGO3 in D. melanogaster (Cox et al., 2000; Saito et al., 2006; Brennecke et al., 2007; Ghildiyal and Zamore, 2009). piRNAs are derived from a few defined genomic regions, called piRNA-clusters, and are present at extremely high sequence diversities. Although both sense and antisense transcripts were detected from piRNA genomic loci, precursors of piRNA are not likely to be double stranded, because accumulation of piRNAs is not affected in Dicer-deficient mutant (Vagin, 2006). Instead, a so called “Ping-Pong” mechanism was proposed, where piRNAs generated from transposon antisense strand are loaded into PIWI and Aubergine proteins, guiding cleavage of transposon sense strand, a process that defines phasing of a new generation of piRNAs. These newly generated piRNAs are then loaded onto AGO3 and target back to the antisense transcripts and further amplifies the piRNA production cycle (Brennecke et al., 2007; Gunawardane et al., 2007). In line with this hypothesis was the observation that
5’ ends of complementary piRNAs are mostly spaced by 10 nt, a signature of AGO-mediated endonucleolytic processing (Brennecke et al., 2007; Gunawardane et al., 2007).

1.3 Small RNAs in plants

All small RNAs in plants can be exclusively grouped into one of the two major groups: single-strand precursor derived hairpin RNAs (hpRNAs), and double-strand precursor derived siRNAs (Axtell, 2013a). hpRNAs can be further divided into miRNAs and everything else based on pre-established criteria (Meyers et al., 2008). In short, a miRNA is generated as a single dominant functional small RNA species due to the precise excision from the single stranded stem-loop precursor, whereas precursors of other hpRNAs give rise to a mixture of 21- to 24-nt small RNAs of great heterogeneity. siRNAs include three subclasses: 1. Heterochromatic siRNAs are ~24 nt small RNAs produced from intergenic repetitive regions. They function at transcription level by depositing de novo repressive epigenetic markers on DNA. 2. Secondary siRNAs are ~21-nt in length. They are generated upon primary small RNA slicing single stranded precursor RNA followed by conversion into dsRNA in a RNA dependent RNA Polymerase (RDR) dependent manner. 3. Nature antisense transcript siRNAs (NAT-siRNAs), like their animal counterparts, are dicing products from hybridized RNA transcripts that are independently transcribed (Axtell, 2013a).

1.3.1 Plant miRNA

Mechanisms of miRNA biogenesis between animal and plants are largely similar despite that different but functionally identical protein factor are deployed. Upon Pol-II transcription, pri-miRNA become 5’ capped and 3’ polyadenylated (Xie et al., 2005). Since Drosha ortholog is not present in plant genome, the pri-miRNA is processed directly by plant Dicer-like protein, Dicer-like 1 (DCL1) in Arabidopsis, for making both cuts, releasing miRNA-miRNA star duplex inside the nucleus (Reinhart et al., 2002; Park et al., 2002; Kurihara and Watanabe, 2004). Efficient and accurate dicing requires assistance from two DCL1 associated proteins, HYponastic LEAVES1 (HYL1), and SERRATE through stabilization of pri-miRNAs (Han et al., 2004b; Lobbes et al., 2006; Vazquez et al., 2004a; Yang et al., 2006; Kurihara et al., 2006). The 2-nt 3’ overhangs on both ends of the duplex generated as dicing signature are then recognized by a SAM-methyl transferase, HUA ENHANCER1 (HEN1) (Yu et al., 2005). As a result, methyl-group is deposited at the 2’-O position of the 3’-most nucleotide, a process that is known to stabilize the duplex (Yu et al., 2005). Differentiated truncation and uridylation occurs at miRNAs 3’ends in HEN1 deficient plants, a process that poses impacts on miRNA turnover rates, as well as secondary siRNA productions (Zhai et al., 2013). miRNA, in a ss or ds form, is exported from nucleus into cytoplasm by HASTY (HST), the plant ortholog of animal EXP5 (Park et al., 2005). The final stage of miRNA maturation comes with loading of miRNA into one of the AGO family proteins. In Arabidopsis, there are 10 AGO proteins placed into three phylogenetic clades: AGO1/5/10 clade, AGO2/3/7 clade, and AGO 4/6/8/9 clade, with AGO1 being the major player of miRNA action (Vaucheret, 2008; Mallory and Vaucheret, 2010). Sorting of small RNAs into different AGO
complexes is mainly determined by identity of the 5’ terminal nucleotide. The 5’ end nucleotides of most miRNAs are uridine, which channeled miRNAs into to AGO1 (Mi et al., 2008).

1.3.2 Non-miRNA hpRNA

RNAi mediated by siRNA generated from exogenously transformed long invert repeat (IR) precursor has been routinely used as a tool for gene suppression (Stam et al., 2000; Tijsterman et al., 2002). Such self-complementary hairpin structure of extremely long precursor (up to several kb) were also found within Arabidopsis genome (Luff et al., 1999). The biological importance of long hairpin arrangement was best represented by the study of PHOSPHORIBOSYLANTHRANILATE ISOMERASE family genes in Arabidopsis. Inverted repeat arrangement of two homologous PAI in Arabidopsis Wassilewskija ecotype was responsible for DNA methylation at all four homologous loci. In contrast, PAI loci from Columbia ecotype are not methylated due to the absence of such inverted repeat arrangement (Luff et al., 1999). In another example, a mixture of non-phased small RNAs with varied lengths (21-, 22- to 24-nt) were produced from these endogenous long hairpins (IR71 and IR2039), indicating participation of multiple DCL proteins (Dunoyer et al., 2010; R et al., 2006a). The functionality of these long hairpin derived small RNAs in RNA dependent DNA methylation and target post-transcriptional silencing were also observed at both the phenotypical and molecular level (Aufsatz et al., 2002; Zilberman et al., 2004; Dunoyer et al., 2010).

1.3.3 Heterochromatic siRNA

Pioneering studies in plant virology found that the cRNA of viroid RNA became hypermethylated in a sequence specific manner upon integration into tobacco genome, a phenomena dependent on the RNA-RNA replication (Wassenegger et al., 1994). It later became clear that a group of highly abundant siRNAs, usually ~24 nt in length, are involved in this so called RNA-dependent DNA methylation (RdDM) process. These 24-nt siRNAs turned out to be the most abundant class of small RNA species in many land plant lineages, and deep sequencing suggested the majority of them were derived from pericentromeric and centromeric regions where repeat sequences are highly enriched (Kasschau et al., 2007; Nobuta et al., 2008). These chromosomal regions are normally heterochromatic, thus these 24-nt siRNAs are termed heterochromatic siRNA. The production of heterochromatic siRNA and its function in mediating de novo DNA methylation, has been observed in not only flowering plants but also lower lineage that is evolutionarily distant (Cho et al., 2008). In contrast, sequence identity of heterochromatic siRNAs and their producing loci are much less likely to be conserved even comparing closely related species (Ma et al., 2010). Such high fluidity of heterochromatic siRNA enables it to act rapidly in response to transposable element activity, where frequent birth and death constantly occur (Axtell, 2013a).

RdDM is initiated upon transcription at certain genomic loci by a plant specific DNA dependent RNA polymerase, Pol-IV (Herr et al., 2005; Onodera et al., 2005; Mosher et al., 2008; Pikaard et al., 2008), followed by dsRNA conversion catalyzed by RNA
Dependent RNA Polymerase 2 (RDR2) (Xie et al., 2004; Kasschau et al., 2007; Lu et al., 2006). DCL3 acts on the long perfect double stranded precursor to produce 24-nt siRNA duplexes (Xie et al., 2004; Kasschau et al., 2007; Lu et al., 2006), with one strand being channeled into a member of AGO4-clade proteins (AGO4, 6, 9) (Havecker et al., 2010). Following siRNA production, non-coding transcripts from target genomic regions are synthesized by another plant specific RNA polymerase, Pol-V (Pontier, 2005; Zhong et al., 2012; Huang et al., 2009; Wierzbicki et al., 2009). These scaffold RNAs recruit AGO4-associated silencing complex owing to its complementarity to AGO4 loaded siRNA (Wierzbicki et al., 2009). DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) in association with AGO4 was therefore recruited for deposition of DNA CHH methylation (H represents A, C, or T). SAWADEE HOMEODOMAIN HOMOLOG1 (SHH1) and SUVH2/9 recognizing repressive histone and DNA methylation markers respectively, mediated the genome wide occupancy of Pol IV and Pol V at heterochromatic regions (Law et al., 2013; Johnson et al., 2014), which suggested a reinforcing loop among RdDM pathway, DNA methylation, and histone modification (Law et al., 2013; Johnson et al., 2014).

1.3.4 Secondary siRNAs

Early study in plants revealed functional requirements of RDR6 in sense transgene and certain viruses induced Post-Transcriptional Gene Silencing (PTGS) (Dalmay et al., 2000; Mourrain et al., 2000). Through screen of endogenous genes that are de-repressed in rdr6 mutant background, a non-protein coding transcript giving rise to a novel group of 21-nt endogenous siRNAs, named trans-acting siRNA (tasiRNA), was identified (Vazquez et al., 2004b). The reliance of siRNA production on RDR6 and canonical miRNA biogenesis machinery, but not traditional heterochromatic siRNA producing enzymes indicated a novel small RNA biogenesis pathway (Vazquez et al., 2004b; Peragine et al., 2004). Cleavage products from the predicted sites of these siRNAs were also detected, which suggested a trans-acting mode (Vazquez et al., 2004b). People later realized that precursors (TAS genes) of these trans-acting siRNAs (tasiRNAs) are themselves target of miRNAs, whose initial targeting triggered the tasiRNA production (Allen et al., 2005; Yoshikawa, 2005). Mechanistically, in cases as mentioned above, miRNA mediated 3’ or 5’ slicing products managed to escape cellular surveillance system that normally leads to nonspecific RNA degradation. Instead, they are stabilized by SGS3, converted into dsRNAs by RDR6, and diced by DCL4 in a phased manner, generating secondary siRNAs whose 5’end positions are pre-determined by the initial miRNA targeting (Allen et al., 2005; Yoshikawa, 2005). Apparently, only a subset of miRNA sliced transcripts are capable of producing secondary siRNAs with eight of them (TAS1a-c, TAS2, TAS3a-c, TAS4) being identified from Arabidopsis (Howell et al., 2007). It was found that unlike general miRNA population, many tasiRNA triggering miRNAs are of 22 nt in length, and artificially lengthening or shortening miRNA length alters the transitivity (Chen et al., 2010; Cuperus et al., 2010a). On the other hand, asymmetrically paired miRNA:miRNA* duplex (bulge presents on one side) also served as a signal that channels AGO-RISC complex into secondary siRNA producing pathway (Manavella et al., 2012). In Arabidopsis, TAS3 was found targeted by miR390 on both its 5’ and 3’ ends. Although cleavage at the 5’ site is not detectable, miR390 targeting at both site are
required for triggering phased siRNAs from the precursor (Axtell et al., 2006). Other factors including a specific miRNA/AGO pair (Montgomery et al., 2008; Manavella et al., 2012) and improperly processed mRNA 3’ termini could also contribute to secondary siRNA production (Herr et al., 2006). Targets of many tasiRNAs have been well studied, and appeared to be highly conserved in many different plant species (Howell et al., 2007; Xia et al., 2013). Still, it should be noted that tasiRNAs resulting from different DCL4-processing cycles are not equally abundant, with only a subset of them having discrete downstream mRNA targets based on complementarity.

1.3.5 Natural antisense transcript siRNAs (NAT-siRNAs)

As described above in animals, hybridization of separately transcribed mRNAs that share complementary sequence provides another source for siRNAs synthesis. If the sense and antisense hybrids are from opposite strands at the same locus, the resulting small RNA is termed cis-NAT-siRNA. Hypothetically, transcripts from genomic distant loci are also possible to form dsRNAs generating trans-NAT-siRNAs. Production of NAT-siRNAs, which is not very commonly seen, tends to depend on environmental signals or specific tissue type (Borsani et al., 2005; Katiyar-Agarwal et al., 2006; Ron et al., 2010). Unlike other siRNAs of dsRNA precursors, the production of cis-NAT-siRNA is strand biased, and resulted in downregulation of the gene of opposite strand but not its partner (Borsani et al., 2005; Katiyar-Agarwal et al., 2006; Ron et al., 2010). Protein factors required for cis-NAT-siRNAs biogenesis include DCL1, HYL1, HEN1, Pol-IV, SGS3, and RDR2/6, and might also differ from case to case (Borsani et al., 2005; Katiyar-Agarwal et al., 2006; Ron et al., 2010). Genome wide study suggested that while a significant portion of Arabidopsis transcription units can potentially form sense-antisense duplex, only a few of them showed anti-correlation in expression (Wang et al., 2005; Henz et al., 2007). Furthermore, siRNA production from overlapped gene pairs does not seem to be a widespread phenomenon (Wang et al., 2005). It is therefore safe to conclude that function of cis-NAT-siRNAs is confined to limited occasions, the genome-wide trend of asymmetric expression observed between overlapped gene pairs may not result from small RNA mediated silencing, alternative mechanism such as interference due to convergent transcription by Pol-II complex might exist (Axtell, 2013a; Bologna and Voinnet, 2014).

1.4 miRNA function in animals and plants

1.4.1 Functional conservation and diversification

In a typical plant small RNA library, reads coming from annotated MIRNA loci usually accounts for ~20% of the total small RNA abundance, which in turn cover ~10% of the genome (Coruh et al., 2014; Lu et al., 2005). In animals, the amount of confidently annotated MIRNA genes equals ~1 to 2% of that for annotated protein coding genes (Bartel, 2009). However, during past two decades, miRNAs have received probably the most attention among all small RNA subclasses, owing to their important and discrete regulatory role in diverse biological process, including cell differentiation and proliferation, abiotic/biotic stress response, hormone signaling, etc. It is now estimated
that, over two-third genes are regulated by miRNA in human (Friedman et al., 2009). In plants, the majority of miRNA targets were found to be master transcription factors, indicating a broad range of functionalities (Axtell and Bowman, 2008; Jones-Rhoades and Bartel, 2004; Rhoades et al., 2002).

In addition, the view that many miRNAs mediated gene regulation pathway are highly conserved across evolutionary distant lineages is widely accepted in both animals and plants (Pasquinelli et al., 2000; Jones-Rhoades and Bartel, 2004; Palatnik et al., 2003; Friedman et al., 2009; Xia et al., 2013). For example, two miRNAs, let-7 and lin-4, regulating heterochronic gene LIN-14 was among the earliest miRNA-target interactions being discovered (Lee et al., 1993; Reinhart et al., 2000). Besides LIN-14, many other heterochronic genes were found to also contain these miRNA target sites in their 3’-UTR regions (Reinhart et al., 2000). Homology of let-7 was later detected in a wide range of animal species, including vertebrate, which suggested this miRNA-mediated heterochronic regulation pathway is deeply conserved across animal kingdom (Pasquinelli et al., 2000; Lagos-Quintana et al., 2001). In addition, several miRNA families are deeply conserved across a wide range of land plants, including miR156, miR164, miR165, etc (Cuperus et al., 2011; Axtell and Bowman, 2008). miR156 inhibits adult transition in flowering plants by negatively regulating the accumulation of SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) family transcription factors (Gandikota et al., 2007). In Physcomitrella patens, a non-flowering plant with dominant gametophytes entity during its life cycle, miR156 also plays critical role in controlling plant life stage transition. To the contrary, rather than inhibiting developmental transition into adult phase, the same miR156/SPL regulatory circuits were deployed in promoting vegetative growth from protonema to gametophores (Cho et al., 2012).

1.4.2 Target complementarity requirements

In animals, functionality of miRNA targets are largely determined by base paring at nucleotide 2 to 8 counting from the miRNA 5’ ends, which is named miRNA “seed”. Perfect Watson-Crick base pairing at seed region is a strong indicator of target functionality (Bartel, 2009). The notion that seed pairing is crucial for animal miRNA function was stemmed from complementary patterns observed from some of the earliest identified miRNA targets in C. elegans (Wightman et al., 1993), which was later found to be a shared aspect among other miRNAs and across multiple animal species (Lai, 2002). Mismatches or G:U wobbles are poorly tolerated in this region (Doench, 2004; Kloosterman, 2004). Based on the criterion of seed pairing thousands of human genes, representing one third of the total human gene sets, were shown to be regulated by miRNAs (Lewis et al., 2005). In fewer cases, additional complementarity at miRNA 3’ region, composed of 3 to 4 contiguously paired bases starting from nucleotide ~13 relative to miRNA 5’ end, can strengthen target recognition, though the magnitude of enhancement is usually small (Grimson et al., 2007). In addition, a group of so called “centered” sites without confirming satisfactory amount of base pairing in either “seed” region of 3’-supplementary region were found to be functional in animals. The centered regions of these sites, spanning nucleotides 4 to 15, were extensively base paired.
Cleavage products of these targets were detected from in vitro slicing assay, however repression at protein level was more likely to be the case when being tested in vivo (Shin et al., 2010).

Early study in Arabidopsis showed miR171 cleaves three members of Scarecrow-like family genes following perfect complementarity (Llave et al., 2002). Although perfectly paired sites for plant miRNAs are rarely seen, the vast majority of currently validated targets pair to corresponding miRNAs near perfectly, with mismatches usually positioned near the 3’ ends of the 5’-most nucleotides of miRNA (Schwab et al., 2005b; Mallory et al., 2004b; Gandikota et al., 2007; Wang, 2005; Mallory, 2005; Mallory et al., 2004a). By introducing single-nucleotide mismatch at different positions of miRNA targets it was shown that target slicing was most impaired when base pairing near miRNA 5’ end was disrupted (Mallory et al., 2004b). Being very informative, however, it should be noted that the conclusion was solely based on tests at RNA level through in vitro slicing assay. A fine resolution map showing the connection between subtle changes in complementarity and target repression at protein level has never been obtained.

Once a target is recognized through near-perfect complementarity, slicing usually occurs at the phosphodiester bond between nucleotide 10 and 11 relative to the miRNA 5’ end. As a result, perfect pairing within this region is critical for proper target repression. In Arabidopsis, a naturally occurring non-protein coding transcript named INDUCED BY PHOSPHATE STARVATION 1 (IPS1) can specifically sequester miR399 due to the presence of a 3-nt central bulge covering canonical slicing site on the target strand (Franco-Zorrilla et al., 2007). This central bulge renders IPS1 resistant to miRNA cleavage, and released the alternative miR399 target, PHO2, from being targeted and repressed (Franco-Zorrilla et al., 2007). In addition to target mimicry mechanism, target site with slicing region unpaired was also found critical for triggering the production of secondary siRNAs in Arabidopsis (Axtell et al., 2006). Converting such site into a sliceable target by fixing the central mismatches eliminated its capability in triggering siRNA biogenesis suggesting a slicing-independent miRNA-target interaction model (Axtell et al., 2006).

In plants, functional targets bear more than 5 mismatches were rarely seen. Two Arabidopsis miR398 targets, Copper Superoxide Dismutase 1/2 (CSD1/2) were probably the two sites of the least complemenarity among all endogenous targets being identified so far (Dugas and Bartel, 2008). In particular, CSD2 site contains 5 mismatches, including 4 G:U wobbles, and an additional 1-nt bulge in the miRNA 5’ region. Besides being directly regulated by miR398, CSD2 is also indirectly controlled by miR398 through a molecular chaperone, Copper Chaperone for Superoxide Dismutase1 (CCS1), which is also a miR398 native target (Beauclair et al., 2010). Although the regulatory efficiency of CSD2 sites has never been assessed quantitatively, it is possible that the accumulation of large amount of mismatches decreased the target efficacy, and thus requires additional regulatory inputs from the CCS1 pathway.
1.4.3 Mode of action

Modes of miRNA-initiated repression are increasingly recognized as being rather similar in plants and animals; in both cases targets are typically repressed by a combination of mRNA degradation and translational repression (Brodersen et al., 2008; Guo et al., 2010; Bazzini et al., 2012; Li et al., 2013b). In animals, miRNA-mediated silencing typically results in the depletion of target mRNAs (Hendrickson et al., 2009; Guo et al., 2010), through a slicing-independent pathway (Bagga et al., 2005; Jing et al., 2005; Valencia-Sanchez et al., 2006). Target slicing is only favored upon forming a highly paired duplex (Yekta, 2004). Translational repression is also widely seen in animals and involves inhibition of translational initiation (Humphreys et al., 2005; Pillai et al., 2005; Mathonnet et al., 2007), which is followed by subsequent deadenylation and decapping of the mRNA (Wakiyama et al., 2007; Iwasaki et al., 2009). In plants, with many target sites being subject to endonucleolytic cleavage by AGOs, it has also been suggested that miRNA targets with lower complementarity are funneled exclusively towards translational repression without detectable mRNA destabilization (Dugas and Bartel, 2008). However, nearly all documented cases of translationally repressed plant miRNA targets involve canonical, highly complementary target sites (Aukerman and Sakai, 2003; Chen, 2004; Gandikota et al., 2007; Brodersen et al., 2008; Li et al., 2013b; 2013a; Yang et al., 2012). It has been recently suggested that, protein-level regulation is more likely caused by miRNA targeting near the 5’ends of gene coding sequences (Li et al., 2013a).

In an in vitro system, translational repression can be observed for perfectly matched sites only when the AGO1 residues required to catalyze target cleavage are mutated or for sites with central mismatches specifically located in the 5’-UTR; animal-like seed-only sites are not effective (Iwakawa and Tomari, 2013). However, the efficacies of potential target sites with more than five mismatches have not been systematically documented in vivo in a plant system to date. The ease of both predicting (Rhoades et al., 2002) and validating (German et al., 2008; Addo-Quaye et al., 2008; Llave et al., 2002) canonical, highly-complementary plant miRNA target sites could conceivably have led to the neglect of a wider constellation of less-complementary sites in plants.

1.5 Objectives

I aim to investigate plant miRNA-target interactions with special focus on the effects of complementary pattern on target regulatory efficiency. This theme covers the following five chapters:

Chapter 1: I reviewed key literatures of small RNA studies from both animal and plant fields. I extracted and summarized major information defining each different small RNA category. Similarities and difference between animal and plant small RNAs were also compared.

Chapter 2: I developed an experimental tool that can quantify efficiency of miRNA-target interactions at both the protein and RNA levels with high accuracy and reproducibility. The method is based on traditionally widely used Agro-infiltration technique. I coupled it
to a self-designed dual-luciferase reporter system. The design offers high amenability to testing a great variety of miRNA-target pairs.

Chapter 3: I investigated the connection between miRNA-target complementary pattern and its regulatory consequences using the tool I developed from Chapter 2. In particular, the following key questions were addressed: comparing the efficacies of naturally occurring miRNA target sites to perfectly paired sites, position-specific effects of mismatches on miRNA targeting, function of animal-like base paring patterns in plants, comparison of ORF vs. 3'-UTR target site locations, the pairing patterns required for effective miRNA target-mimicry, and evaluation of a non-canonical plant miRNA target site.

Chapter 4: I generated high-throughput sequencing library for *N. benthamiana* small RNAs. Using the computational tool developed by my advisor Dr. Axtell, I *de novo* analyzed *N. benthamiana* small RNA producing loci, and quantified their population. Features of small RNA producing loci, and their co-occupancy with other genomic features were also analyzed.

Chapter 5: I summarize the results from previous chapters, and discuss future prospects.
Chapter 2
Developing a dual-luciferase transient expression system to quantify miRNA-target interactions

2.1 Introduction

miRNA-mediated target regulation participates in diverse cellular activities throughout plant development. Validations of miRNA function are often conducted using stable transgenic lines combining mutant backgrounds, where miRNA and its presumable target gene were mutated to disrupt complementarity in the hope of observing altered response at molecular and phenotypical level (Llave et al., 2002; Palatnik et al., 2003; Mallory et al., 2004b; 2004a; Guo et al., 2005). This type of study provided the earliest and one of the most solid evidences of functional miRNA-target interactions in plants. However, in some cases, direct miRNA-mediated regulation could be confounded by complicated in vivo gene network.

Based on the fact that the 3’ cleavage products generated by miRNA directed slicing are relative stable, in vitro slicing assay, modified 5’RACE assay, and direct sequencing of degradome were all successfully developed (Llave et al., 2002; Palatnik et al., 2003; Mallory et al., 2004b; Tang et al., 2003; Addo-Quaye et al., 2008). Now it is known that many plant miRNAs use complex mechanisms to repress target expression at both the mRNA and protein levels. Examining target cleavage at mRNA level along thus offers limited resolution in recovering the complete picture of miRNA-target interaction (Chen, 2004; Brodersen et al., 2008). Systematic studies of pleiotropic effects due to miRNA overexpression or knockout using genome wide strategy (ribosome-profiling, mRNA-seq, microarray) were also seen (Guo et al., 2010; Baek et al., 2008; Schwab et al., 2005a; Grimson et al., 2007). Similarly, it is quite challenging to separate direction miRNA function from indirect effects due to entangled gene network.

Agro-infiltration mediated transient assay in N. benthamiana has been widely used, and demonstrated to be a powerful tool in testing small RNA functionality (Kurihara et al., 2006; Chen et al., 2010; Montgomery et al., 2008; Allen et al., 2005; Llave et al., 2002; Franco-Zorrilla et al., 2007).

In this chapter, I aim at developing an experimental tool that is suitable for testing various plant miRNA-target interactions in an efficient and highly controlled manner. I coupled the traditional Agro-infiltration system to a dual-luciferase based reporter assay, where Renilla Luciferase (R-Luc) and Firefly Luciferase (F-Luc) are placed on a single vector. Use of the F-Luc/R-Luc ratio allowed internal normalization that corrects for variation in transformation efficiencies between experiments. Two unique restriction sites were embedded to facilitate target site switch. miRNA overexpressor and dual-luciferase reporters are transiently co-expressed in N. benthamiana leaves. miRNA activity and target efficacy were quantified at both the mRNA and protein levels in two different

---

*This chapter is modified from a manuscript submitted to *Methods in Molecular Biology.*
contexts (ORF vs. 3’-UTR). I showed that known functional miRNA targets were all effectively repressed as quantified at both the mRNA and protein levels.

2.2 Materials

Prepare all solution using autoclaved deionized water and analytical grade reagents. All reagents are stored at room temperature unless mentioned otherwise or indicated by the manufacture.

2.2.1 Agro-infiltration components

1 M MgCl₂ stock: Add about 20 mL water in a 100-mL glass bottle. Weigh 10.17 g MgCl₂ and transfer to the glass bottle. Mix, and transfer solution to a 100-mL graduated cylinder. Add water up to 50 mL. Mix, and transfer back to glass bottle. Autoclave.

100 mM 2-(N-morpholino)ethanesulfonic acid (MES) stock: Add about 150 mL water in a 500-mL glass bottle. Weigh 5.33 g MES and transfer to the glass bottle. Mix, and adjust pH to 5.7. Transfer solution to a 250-mL graduated cylinder, and add water up to 250 mL. Filter sterilize.

100 mM acetosyringone stock: Weigh 196 mg acetosyringone and dissolve in 10 mL DMSO. Filter sterilize, and aliquot into 1.5-mL microcentrifuge tubes. Store at -80°C for long term storage.

1-mL syringe without needle, razor blade, microcentrifuge tube, 15 and 50-mL falcon tube, portable hole-puncher.

2.2.2 Dual-luciferase assay components

Reagents for dual-luciferase assay are provided by Luciferase Assay System (Promega cat. # E1500). Equivalent reagents from alternative suppliers should work too, although I have not tested using this protocol.

2.2.3 RNA extraction, reverse transcription, and qRT-PCR reagent

RNA extraction reagents are provided by QIAGEN miRNeasy mini kit.

Reverse transcription reagents are provided by QIAGEN QuantiTect Reverse Transcription Kit.

qRT-PCR reagents are provided by QIAGEN QuantiTect SYBR Green PCR kit (see Note 1).

2.3 Methods

2.3.1 Plant growth condition
*N. benthamiana* plants are grown at 22°C under 24-h-light conditions. Each pot (3×3×2, L×W×D by inches) contains one plant. Two flats (18 pots each) of plants are planted regularly each week. Plants that are exactly one-month old are most suitable for infiltration (3-5 young leaves, Figure 2.1).

2.3.2 Dual-luciferase reporters for miRNA

There are two versions of dual-luciferase reporters (Figure 2.2). To insert target of interest, order desired DNA oligos flanked by *AvrII* and *AgeI* restriction site. For sense strand, *AvrII* site is on its 5’ end, and *AgeI* site is its 3’ end; vice versa, for antisense strand (*see Note 2*).

Dissolve lyophilized oligos in water to obtain 100μM stock. In a small thin-wall PCR tube, make the hybridized inserts by mixing 0.5 nanomoles (5μL each) of sense and antisense oligos, and add water to up to 48.3 μL. Add 1.67 μL of 0.3M NaCl. Mix well.

Program thermal cycler as follow: 97°C 5min, ramp down to 20°C at 0.1°C/sec. Run the program with the mixtures.

Perform *AvrII* and *AgeI* double digestion of dual-luciferase plasmid. Run an Agarose Gel of the digestion product and purify it (*see Note 3*).

Perform T4 DNA ligase mediated ligation of double digested plasmid (10 ng) with hybridized inserts (1μL from step 3). Add a no-insert control to monitor self-ligation due to incomplete double digestion, which result in background colonies upon transforming *E.coli*.

Transform competent *E.coli* cell. Successful transformants should confirm Kanamycin resistance.

Perform colony PCR followed by sequencing check to confirm appropriate insertion (*see Note 4*).

Desired plasmids are further transformed into Agrobacteria (strain GV3101, pMP90, pSoup). Successful transformants should confirm Rifampicin, Tetracycline, Gentamicin, and Kanamycin resistance (*see Note 5*).

2.3.3 Agro-infiltration (*see Note 6*)

Day 1: Primary Inoculation

Around mid-day, prepare 16 mL LB broth containing the following three antibiotics (final concentration): Rifampicin (50 μg/mL), Tetracycline (5 μg/mL), Gentamicin (25 μg/mL).

Aliquot 2 mL LB into a 10-mL glass culture tube covered with metal cap. This is for no-vector control.
Add Kanamycin to the rest 14ml LB to a final concentration of 25 µg/ml, and mix thoroughly.

Aliquot 2mL LB each into 7 other 10-mL glass culture tubes.

A single colony of each strain (sensor, miRNA overexpressor, and no-vector control) is inoculated from plate and grown at 28°C at 280 rpm overnight to reach saturation (see Note 7).

Day 2: Secondary Inoculation

Around 6pm, prepare 50 mL LB broth containing the following three antibiotics: Rifampicin (50 µg/mL), Tetracycline (5 µg/mL), Gentamicin (25 µg/mL).

Aliquot 4mL LB into a 10-mL glass culture tube with metal cap. This is for no-vector control.

Add Kanamycin to the rest 46 mL LB to a final concentration of 25 µg/mL, and mix thoroughly.

Aliquot 4 mL LB into each of 5 other 10-mL culture tubes. There are for different sensor strains.

Aliquot 2mL LB into a 10-mL culture tubes. This is for no-cell control to monitor possible contamination of LB media.

Transfer 24 mL LB into an autoclaved flask covered with aluminum foil. This is for miRNA over-expressor strain.

Secondary working cultures are inoculated from the starter cultures with a 1:100 ratio at 28°C 280rpm.

Day 3: Infiltration

In the early morning (around 9am), check OD$_{600}$ value of the cell cultures (see Note 8).

Prepare a total of 100 mL infiltration media (IM, see Note 9). Mix 88.85 mL water, 1 mL of 1 M MgCl$_2$ stock, 10 mL of 100 mM MES stock, and 150 µL of 100 mM of acetosyringone stock in a glass bottle.

Label 15-mL falcon tubes for each strain. Use 50-mL falcon tube for miRNA over-expressor.
Transfer the cell culture into corresponding falcon tubes, and harvest cells by centrifugation at 3,000g, 22°C for 5 minutes. Alternatively, extend centrifugation to 10 minutes for volumes larger than 10 mL.

Pour off supernatant into an empty flask, and blot dry the opening of falcon tubes on Kimwipes.

Add 2 mL IM to cell pellet and re-suspend by vortexing to obtain stock culture. Once cells are re-suspended, set a timer for 4 hours  (see Note 10).

Adjust and obtain ~5 mL of IM re-suspended cell cultures to OD600 at 0.5 (see Note 11).

Set up a new batch of 15-mL falcon tubes for different treatments, and label accordingly. Mix equal volume (4 mL) of OD600 adjusted sensors and miRNA over-expressor into corresponding tube. Now each tube should contain 8 mL of IM with miRNA over-expressor and sensor mixed at 1:1 ratio.

IMs are kept on bench-top before infiltration. Keep the cap loose to allow the exchange of fresh air.

Record treatments and corresponding plant ID on lab notebook (see Note 12).

Infiltration starts at around 2:30 pm (4 hours from re-suspension). Label ID on all three leaves that will be infiltrated (see Note 13).

Fill 1-mL syringe with well mixed IM (cell sinks during incubation). Use a razor blade to make a tiny wound on the abaxial (lower) surface. Wounded site should be around 5-10 mm away from the mid-vein. Hold the leaf in one hand with abaxial side flipped up, and support the wounded region gently with one of your finger. Gently press syringe against the supporting finger. IM should spread immediately. Make a second wound on the other side of the mid-vein if necessary, and infiltrate until the whole leaf area is filled up. Repeat infiltration for all three plants (9 leaves) of the same treatment.

Cut paper towel into small pieces, and blot excess IM off the leaf surface.

Change a pair of gloves before move on to the next treatment. Alternatively, thoroughly clean the gloves with plenty of water and ethanol.

Place plants back into the growth room. Allow enough spacing. Avoid contacts of leaves of different treatments.

Day 4: No content

Day 5: Sample collection
Around 2:00 pm, set up 4 microcentrifuge tubes on rack for each treatment (3 tubes for protein-level assay, and 1 tube for RNA-level assay). Label accordingly (see Note 14).

Cut off all three labeled leaves from the first plant. Punch 6 leaf discs from each leaf. Collect 3 discs into each of the three protein-tubes (1 disc from each leave). Collect 9 discs into one RNA-tube (3 discs from each leave). Freeze sample tubes in liquid nitrogen immediately.

Before moving onto sample of different treatment, thoroughly clean the puncher head with 70% ethonal to avoid cross contamination.

Proceed with all samples.

Briefly cool plastic pestle (Kimble Chase cat. # 749521-1500) in liquid nitrogen and grind samples into fine powder. Fully grinded samples are stored at -80°C prior to protein and RNA assay (see Note 15).

2.3.4 Dual-luciferase Assay (see Note 16)

Use the provided template (Table 1) to determine the loading pattern of your sample, as well as to calculate the amount of reagents required (see Note 17).

Prepare lysis buffer (PLB), firefly luciferase substrate solution (LARII), and Renilla luciferase substrate solution (Stop&Glo, see Note 18).

If using Luciferase Assay System (Promega), chill PLB on ice prior to the test. Turn on the luminometer, and prime the robotic injector with LARII and S&G. Program GloMax software by highlighting appropriate wells being tested.

Set pipette at 10 µL, and have a new box of 20-µL tips ready.

Remove the first group of samples for protein assay from freezer and freeze and store in liquid nitrogen immediately (see Note 19).

Remove samples from liquid nitrogen onto a rack at room-temperature, add in 500 µL ice chilled PLB. Vigorously shake and vortex samples immediately to completely resolve powders into the solution. Place samples back onto ice right after re-suspension (see Note 20).

Proceed with the rest samples.

Centrifuge at top speed, 4°C, for 30 seconds to clarify undissolved debris. Place samples back onto ice right after centrifugation.

Load 10 µL of the supernatant from each sample into designated position on the 96-well plate (see Note 21).
Load the plate, and start the run.

2.3.5 Dual-luciferase Assay Data Analysis

Remove background value from each sample.

Calculate the Fluc/Rluc ratio for all samples.

Examine control samples for degradation. Perform a linear regression on the sample (#1, #4 in this case) to see if significant non-specific protein degradation occurs (x-value, the order of the reading being obtained; y-value, the Fluc/Rluc ratio).

If degradation occurs, use the degradation curve generated from control samples to normalize Fluc/Rluc ratio for other samples (see Note 22).

Compare the normalized Fluc/Rluc ratio of test samples to that of spacer and perfect site controls to determine the efficiency of miRNA-target interactions.

2.3.6 qRT-PCR Assay

Perform RNA extraction, and Reverse transcription following standard protocol or manufacture’s instruction if using commercial kits (see Note 23).

Perform qRT-PCR amplifying Firefly Luciferase (F-Luc) and Renilla Luciferase (R-Luc) (see Note 24).

2.3.7 qRT-PCR Assay Data Analysis

Examine each reaction to make sure there is a single product of expected Tm. Eliminate samples with aberrant results or no amplification.

Examine the no RT, and no RNA control. Make sure no amplification is present (see Note 25).

Examine the threshold line. Some instruments generate threshold line automatically. Manually make adjustment if necessary (see Note 26).

Export data as an Excel file, and calculate mean amplification efficiency (E) for both R-Luc and F-Luc amplicon (see Note 27).

To calculate relative expression (RE, F-Luc/R-Luc), use the following formula:

\[
RE = \frac{(R-Luc-E)^{(R-Luc-Ct)}}{(F-Luc-E)^{(F-Luc-Ct)}}
\]
To calculate fold-repression, designate your control sample. In this example, the control sample is the “spacer control” (plant #4, 5, 6). Calculate the median RE value from control samples. Normalize other RE values to the median control RE.

2.4 Notes

1. Alternative methods can also be used. If over-expression of the miRNA is a concern, and needs to be confirmed later, please make sure the method you choose can effectively recover small RNA portion.

2. For example, target sequence “ATCGATCGATCGATCGATCGA”. Oligos to order:
Ssense 5’-CTAGGATCGATCGATCGATCGA-3’
Antisense 5’-CCGGTTCGATCGATCGATCGATC-3’
Please note that, restriction sites are highlighted in red. Sticky ends are automatically generated upon hybridizing the two oligos. If using 3’-UTR sensor, a common 21-nt spacer can be used as negative control. If using ORF sensor, each target site requires a distinct synonymous negative control, where the encoded amino acids remain the same with complementarity being maximally disrupted.

3. Any commercial gel purification kit should work. It is better to run single digestion and undigested controls in parallel. Since region in between the two restriction sites is small (~20nt), successful double digestion generates no visible band besides linearized plasmid.

4. Colony PCR primers and program:
F: 5’-GTTTTGGAGCACGGAAAGAC-3’
R: 5’-AAGCTCGGAATTAACCCTCA-3’
PCR program:
Holding Stage:
94°C, 5 minutes
Cycling Stage: (35 cycles)
94°C, 30 seconds
52°C, 30 seconds
72°C, 35 seconds
Holding Stage:
72°C, 10 minutes
4°C, Forever

5. Prepare LB plates ahead containing the following four antibiotics of appropriate concentration: Rifampicin (50 µg/mL), Tetracycline (5 µg/mL), Gentamicin (25 µg/mL), and Kanamycin (25 µg/mL). Rifampicin selects for the chromosome of the Agrobacterium strain. Gentamicin selects for the large Vir plasmid. Tetracycline selects for pSOUP, which is a helper plasmid required for pGreenII replication. Kanamycin selects for pGreenII. It is recommended to also perform colony PCR followed by
sequencing check on selected *Agrobacterium* colonies, though false positive occurs at extremely low rate.

6. In this protocol, 3 candidate miRNA target sites, along with a spacer and a perfect site control will be tested (5 treatments in total). A miRNA over-expressor will be co-infiltrated. The volume of each reagents required can be scaled up if one samples are tested. For a reasonable amount of work load, it is recommended to limit the number of treatments to 10 per experiment (3 biological replicates each).

7. Tubes should be tilted with an angle during incubation. Colonies on plates are generally in good condition for 50 days. Transfer colonies to a fresh plate with appropriate antibiotics periodically or save glycerol stocks for any long-term experiment.

8. OD$_{600}$ should be above 1.5, not exceeding 1.8 (1cm path). Values within this range indicate that the bacteria is at logarithmic growth phase (most suitable for the experiment). Adjust length of incubation time to empirically determine your own parameters if necessary.

9. Each treatment requires sensor strain to be re-suspended in at least 4mL IM, to final OD$_{600}$ at 0.5. Empirically, preparing 7mL of IM for each sensor should be sufficient. Since sensors and miRNA over-expressor will be mixed at 1 to 1 ratio, the amount of IM being prepared for miRNA over-expressor equals to the total amount of IM prepared for all sensors. In this case, a total of 100mL IM is prepared.

10. It is critical to fully re-suspend the cell pellets. Un-dissolved pellets give inaccurate concentrations. Re-suspension is usually done around 10:30 am. The infiltration starts at 2:30 pm.

11. Estimate the amount of stock cultures required. Transfer excess stock culture into 1.5-mL microcentrifuge tubes, and label accordingly. Usually, around 0.8mL of the 2-mL stock culture is retained for adjustment. Final volume should reach at least 4mL for each sensor being tested.

12. Each plant receives only one treatment (small RNA moves!). Three plants are used for each treatment (biological replicates). Three leaves from each plant will be infiltrated. 17 plants will be used in this experiment (3 for each treatment plus two background controls).

- Plant #: 1, 2, 3: miRNA over-expressor + Perfect site (Positive control)
- Plant #: 4, 5, 6: miRNA over-expressor + Spacer site (Negative control)
- Plant #: 7, 8, 9: miRNA over-expressor + Target site 1
- Plant #: 10, 11, 12: miRNA over-expressor + Target site 2
- Plant #: 13, 14, 15: miRNA over-expressor + Target site 3
- Plant #: 16: miRNA over-expressor + No-vector control (Control of background signal)
- Plant #: 17: No-vector control only (Help to verify the miRNA over-expressor)
13. Choose leaves that are healthy, and have no visible damage. Avoid wrinkled, unexpanded leaves, which are usually quite resistant to infiltration. Also avoid leaves that are too large, which are not only resistant, but also consume a lot of IM during infiltration.

14. Have liquid nitrogen ready. Samples are collected exactly 48 hours after infiltration (starts at around 2:30 pm). A portable one-hole puncher will be used (punch size 0.25 inches). Please refer to the link for the type of ideal hole-puncher (http://www.staples.com/Staples-1-Hole-Punch-5-Sheet-Capacity/product_146308).

15. It is critical that samples remain frozen at all times to prevent degradation. Chill samples frequently in liquid nitrogen to avoid elevated temperature. Place sample tube in a microcentrifuge rack to provide enough support during grinding.

16. Dual-luciferase protein assay is carried out on a GloMax 96 microplate luminometer equipped with robotic dual-injector using Luciferase Assay System (Promega, catalog# E4550). You can chose alternative instrument and equivalent product at your own choice.

17. If using a plate reader, it is highly recommended to load two of your samples on the plate in multiple equally spaced positions, so that readings of the same sample at different time point can be used to monitor any non-specific degradation, especially when a large amount samples are being tested (>30). A degradation curve (if there is any) will be generated using those two samples, and averaged for normalization purpose of the rest samples. Also, I usually split the tests into three separate runs, with each run being normalized independently.

18. Here is an example using Luciferase Assay System (Promega):
Prepare Passive Lysis Buffer (PLB, 500 µL/sample):
PLB 5× buffer: (17×3)×500 µL÷5 = 5100 µL
H2O: 5100 µL×4 = 20400 µL
Prepare Luciferase Assay Reagent II (LAR II, 100 µL/sample, plus 1000 µL for priming the robotic injector):
LARII: (21×3+10)×100 µL = 7300 µL
Prepare Stop & Glo Reagent (S&G, 100 µL/sample, plus 1000 µL for priming the robotic injector):
S&G buffer: 7300 µL
S&G 50× reagent: 7300 µL÷50 = 146 µL

19. It is helpful to have a divider from a fiberboard storage box placed in the liquid nitrogen so that samples can be placed in the order that they will be tested. Keep samples in enough but minimum amount of liquid nitrogen bath. Too much liquid nitrogen will make tubes hard to open, and difficult for dissolving into lysis buffer.

20. Experienced person can proceed up to 6 tubes in one run.
21. It is a good idea to print out a template table to guide the loading. Samples (#1 and #4 in this case) being used for monitoring non-specific degradation should be remained on ice in-between each loading.

22. A sample data sheet is attached to illustrate the normalization process. Please double click on cells to view the formula being applied.

23. Add –RT control for each RNA sample during reverse transcription to determine if genomic contamination is present. Random oligos work fine for the dual-luciferase sensor, gene specific primers are not necessary during reverse transcription.

24. A 25-µL reaction system is used. Each cDNA sample is run in triplets as technique replicates. 
Oligos for amplifying dual-luciferase reporters: 
Fluc F: 5’-GTTTTGGAGACGGAAAGAC-3’
Fluc R: 5’-CAAGAGTAAAAGATAGTAAAAACCGG-3’
Rluc F: 5’-TGTTGGACGACGAATTAC-3’
Rluc R: 5’-CATTTTTGTCGGCCATGATT-3’
PCR program: 
Holding Stage: 95°C 15 minutes
Cycling Stage: (45×) 
95°C 15 seconds
52°C 30 seconds
72°C 30 seconds (data collecting stage)
Melt Curve Stage: 
95°C 15 seconds
60°C 1 minute
95°C 15 seconds
(collect data on ramp, temperature increment 0.3°C)

25. Sometimes amplification curve shows up for –RT samples. However, the cycle number at which exponential amplification occurs should be significantly larger than corresponding +RT samples (more than 10). When examining melting curve, the amplicon appears to be different, which indicates non-specific amplification. This is acceptable.

26. The optimal position of threshold line should be set in the exponential phase of amplification curve. Threshold line setting too high or too low will increase the deviation among replicates.

27. I use third party software “LinRegPCR” to calculate amplification efficiency (http://www.hartfaalcentrum.nl/index.php?main=files&sub=LinRegPCR). Other methods can also be used. Traditionally, a dilution series of cDNA samples were used for calculating amplification efficiency. Plot the Ct values (y) against the log_{10} dilution.
values \( (x) \). Calculate the best-fit linear regression equation for the plot. \( E = 10^{(-1/\text{slope})} \), where slope is derived from the linear regression equation.

2.5 Results

To confirm that the transient dual-luciferase assay can truly reflect real miRNA-target interaction quantitatively, I test the functionality of several known miRNA targets. miR164 and miR398 were chosen as miRNA candidates, because both are well conserved miRNAs across land plant lineages, and targets of which were well studied by multiple groups (Mallory et al., 2004a; Schwab et al., 2005a; Beauclair et al., 2010; Dugas and Bartel, 2008; Guo et al., 2005; Sieber et al., 2007; Baker et al., 2005). Target site designs were based on the *Arabidopsis* miR164a and miR398b mature sequences. *N. benthamiana* small RNA-seq from leaves revealed a single miR164-related species whose sequence exactly matched *Arabidopsis* miR164a, and two predominant miR398 isoforms with one and two mismatches, respectively, relative to *Arabidopsis* miR398b (Table 2.2). I thus co-overexpressed *Arabidopsis* miR398b in experiments with miR398 sensors. The frequency and position of mismatches present in these naturally occurring sites varies greatly among different targets of the same miRNA.

Seven genes are experimentally verified miR164 targets in *Arabidopsis*: NAC1, NAC2, CUC1, CUC2, AT3G12977, AT5G07680, and AT5G61430 (Mallory et al., 2004a; Guo et al., 2005; Baker et al., 2005; Addo-Quaye et al., 2008). Analysis using 3’-UTR sensors and endogenous *N. benthamiana* miR164 showed that these target sites had variable efficacies compared to a spacer control, which is a 21-nt insert of random sequences (Figure 2.3 and 2.4). The least effective natural target sites (NAC2 and CUC1/2) had mismatches or G-U wobbles in the central and 5’ regions of the target site (Figure 2.4). In contrast, for the most effective sites (AT5G61430 and AT5G07680), non Watson-Crick pairs were limited to the 3’ region (Figure 2.4). Three miR398 natural target sites from *Arabidopsis* (CSD1, CSD2, and CCS1) (Dugas and Bartel, 2008; Beauclair et al., 2010), and a perfectly matched miR398 site were also repressed at different levels, with the perfect site showing the strongest activity. The CSD2 site, which had the most disrupted 5’ region, was the least effective target site (Figure 2.4). This was consistent using both 3’-UTR (Figure 2.4) and ORF sensors (Figure 2.5). For ORF sensors, each tested site changes the amino acid sequence of the extended F-Luc protein, and thus each requires a distinct synonymous negative control (SNC), which is a target site with maximally disrupted complementarity without changing the amino acid sequence (Figure 2.2, Appendix Table 1.1). The experiments were performed multiple times, and results were highly reproducible. These data showed that the efficacy of plant native miRNA target sites can vary among different targets of the same miRNA, and that some are less than maximally effective. It is also in line with canonical view that amount of unpaired bases near the miRNA 5’ region is a strong indicator of impaired regulatory efficiency. Most importantly it demonstrated that the transient assay indeed reflected bonafide miRNA-target interactions in a highly quantitative manner.
**Figure 2.1** One-month old *Nicotiana benthamiana* plants.

**Figure 2.2** Schematic diagrams of dual-luciferase sensors.
Upper, ORF sensor; Lower, 3’-UTR sensor. \(P_{35S}\), promoter sequence of the CaMV 35S gene; \(P_{\text{NOS}}\), promoter sequence of the nopaline synthase gene.
Table 2.1 Loading template for Dual-Luciferase Assay.
Samples numbered in red (1, and 4) are selected for monitoring non-specific protein degradation. They are loaded repeatedly at different positions. Three technique replicates are split into separate runs, with each run being normalized independently.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>B</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>1</td>
<td>4</td>
<td>1st run</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>D</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>1</td>
<td>4</td>
<td>2nd run</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>F</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>G</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>1</td>
<td>4</td>
<td>3rd run</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.2 Identification of *N. benthamiana* miRNAs
Nucleotides that are different from *Arabidopsis* miRNA sequences are labeled in red.

<table>
<thead>
<tr>
<th>miRNA</th>
<th>Sequence, 5’ to 3’</th>
<th>Relative abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR164a_related</td>
<td>UGGAGAAGCAGGGCACGUGCA</td>
<td>100%</td>
</tr>
<tr>
<td>miR398b_related_1</td>
<td>UAUGUUCUCAGGUCGCCCU</td>
<td>70.8%</td>
</tr>
<tr>
<td>miR398b_related_2</td>
<td>UGUGUUCUCAGGUCGCCCU</td>
<td>16.6%</td>
</tr>
<tr>
<td>miR398b_related_3</td>
<td>UAUGUUCUCAGGUCGCCCU</td>
<td>0.1%</td>
</tr>
</tbody>
</table>

Figure 2.3 Conventions for miRNA-target site display and nomenclature.
Figure 2.4 Quantification of naturally occurring target sites of miR164 and miR398 using 3’-UTR sensor.
Naturally occurring miR164 and miR398 sites in 3’-UTR sensors. Boxplots summarize results from fifteen independent replicates and show the median (thick line), extent of the 1st to 3rd quartile range (box), values extending to 1.5 times the interquartile range (whiskers), and outliers (black circles). Target site sequences are shown in Appendix Table 1.1. Dual-asterisks indicate significant differences at the protein level between two adjacent groups (P<0.05, n=15, ANOVA-TukeyHSD).
Figure 2.5 Quantification of naturally occurring target sites of miR164 and miR398 using ORF sensor. As in Figure 2.4, except using ORF sensors. Synonymous negative controls (SNC) are negative controls with maximal disruption of their cognate sites while maintaining identical amino acid sequences. Data display conventions as in Figure 2.4.
Chapter 3
Analysis of complementarity requirements for plant microRNA targeting

3.1 Introduction

microRNA (miRNA) mediated target down-regulation plays essential role in a variety of cellular activities in organisms across both animal and plant kingdoms. Computational prediction of miRNA targets have been fairly successful in both animals and plants using a combination of criteria including the amount of base pairing and thermodynamic stability of the candidate miRNA/target duplex, conservation of sequence (target and flanking contexts), conservation of regulatory pathway involved, and target site abundance (in animals) (Garcia et al., 2011; Alves et al., 2009; Chorostecki et al., 2012; Lewis et al., 2003). In all the criteria mentioned above, degree of complementarity was considered as the most critical factor. In animals base pairings at miRNA 5’ends were given special weight, whereas in plants, extensive base pairing (≥17 nt) throughout target site was required to avoid high false positive rate. For example, penalty score for mismatches and G:U wobbles was doubled if they occur between position 2 to 13 relative to the 5’end of miRNA (Fahlgren and Carrington, 2010).

In animals, miRNA target sites are almost exclusively present within the 3’-UTR of target mRNAs (Bartel, 2009; Lewis et al., 2005; Krek et al., 2005; Lai, 2002). 5’-UTR and open reading frame (ORF) targeting are also functional, but occur with much less frequency (Grimson et al., 2007; Lytle et al., 2007). Many animal miRNA-target sites form seven consecutive base pairs from position two through eight (numbering from the 5' end of the aligned miRNA); this is the "seed". A perfect “seed” pairing serves a strong indicator for functional miRNA target site (Lai, 2002; Doench, 2004). Mismatches introduced in the miRNA “seed” region showed defects of miRNA targeting as evaluated through both phenotype and GFP sensors (Kloosterman, 2004). The relatively low pairing threshold also explains how animal miRNAs managed targeting such a big portion of protein coding genes. Both perfectly and imperfectly paired ‘seed’ sites are occasionally supplemented by additional pairings in the miRNA 3’ region, resulting in enhanced target recognition (Grimson et al., 2007; Kloosterman, 2004). In fewer cases, centered sites, comprising pairing at positions 4 to 15, also mediate target repression (Shin et al., 2010).

Most experimentally verified miRNA target sites in plants are single sites located in ORFs, with a few exceptions in 5’-UTRs, 3’-UTRs, or in non-coding RNAs (Allen et al., 2005; German et al., 2008; Addo-Quaye et al., 2008). Also in contrast to animal miRNA target sites, extensive complementarity (typically ≤5 mismatches) is the hallmark of all functionally verified miRNA target sites in plants to date. Base pairing at the miRNA 5’ region (from positions 2-13) is critical for plant miRNA-mediated target repression with pairing in the vicinity of the AGO-catalyzed slicing site (positions 9-11) being especially important (Mallory et al., 2004b; Parizotto et al., 2004; Schwab et al., 2005a).

*This chapter is modified and reproduced from the following published works: Liu et al. 2014 Plant Cell, Cécile Brousse and Liu et al. 2014 Nucleic Acids Research.*
Mismatches at the miRNA 3’ region are much less destructive than those at the 5’ or central region (Mallory et al., 2004b; Parizotto et al., 2004; Schwab et al., 2005a; Lin et al., 2009). Using a wheat germ lysate assay, Mallory et al. demonstrated that mismatches induced at target 5’ ends as well as the 3’ most nucleotide had minimal effects on target slicing rate, whereas pairing at miRNA 5’ portion were strictly required for efficient targeting (Mallory et al., 2004b). Further investigating the pairing pattern all known natural miRNA-target pairs (~50) by then revealed extremely low enrichment of mismatches at those critical positions (Mallory et al., 2004b).

On the other hand, it should be noted that most previous studies that explicitly examined complementarity requirements for plant miRNA function have relied heavily or exclusively on mRNA accumulation data (Schwab et al., 2005a; Mallory et al., 2004b; Parizotto et al., 2004), or visual phenotypes (Lin et al., 2009; Mallory et al., 2004b). Quantitative measurements of in vivo target site efficacies, measured not only at the mRNA level but also simultaneously at the protein level, coupled with systematic exploration of different complementarity patterns, have not been reported for any plant to date.

3.2 Materials and Methods

3.2.1 GUS-based Competitors

GUS ORF was amplified by PCR using Phusion High-Fidelity DNA Polymerase (New England Biolabs). PCR products were double digested with XhoI and EcoRI, and were ligated into pGreenII-0229 previously digested by the same enzymes. CaMV 35S promoter (5’-ApaI and 3’-XhoI) and terminator (5’-SpeI and 3’-NotI) were also PCR amplified, and were placed upstream and downstream, respectively, from GUS. A 45-nt artificial 3’-UTR sequence flanked by PstI and SpeI was synthesized (Integrated DNA Technologies), and was placed in-between the GUS ORF and the 35S-terminator. Two unique restriction sites (AvrII and AgeI) were embedded in the artificial 3’-UTR sequence to facilitate the insertion of different target sites. Target site inserts were generated by hybridizing synthetic oligonucleotides flanked by AvrII and AgeI sites, ligated into the 3’-UTR of GUS previously digested by the same enzymes, and confirmed by sequencing. Appendix Table 1.2 contains oligonucleotide sequences.

3.2.2 Dual-luciferase Sensors

Renilla luciferase (R-Luc) ORF was amplified by PCR using Phusion High-Fidelity DNA Polymerase (New England Biolabs). PCR products were double digested with SalI and XbaI, ligated into NOS-cassette of pUC119 vector previously digested by the same enzymes. NOS-R-Luc-cassette was inserted into pGreenII-0000 previously digested with HpaI through blunt-end ligation. Firefly luciferase (F-Luc) ORF with (3’-UTR sensor) or without (ORF sensor) stop codon were also PCR amplified, and were double digested with XhoI and EcoRI. The 35S-F-Luc cassette was created by replacing GUS ORF in pGreenII-0229 with F-Luc ORF. The whole cassette (P35S-F-Luc-3’UTR-ter) was cut by Apal and NotI, and ligated into pGreenII-0000 bearing NOS-R-Luc cassette. Target site inserts were generated by hybridizing synthetic oligonucleotides flanked by AvrII and AgeI sites, ligated into the 3’-UTR of F-Luc previously digested by the same enzymes,
and confirmed by sequencing. (Appendix Table 1.1 for target site sequence, Appendix Table 1.2 for oligonucleotides used).

3.2.3 miRNA over-expressers
All MIRNA sequences were PCR amplified from Col-0 genomic DNA using Phusion High-Fidelity DNA Polymerase (New England Biolabs). The complete hairpin structures and ~50nt both upstream and downstream of the hairpin region were amplified. XhoI and EcoRI sites were incorporated into the 5’ and 3’ end of the PCR product followed by double digestion to replace the GUS ORF in GUS-based competitor.

3.2.4 Agro-infiltration and dual-luciferase assay at both the mRNA and protein level
Please see Chapter 2.

3.2.5 Small RNA-seq
Two replicate RNA samples from one-month old N. benthamiana leaves were extracted and used as input for TruSeq Small RNA Sample Preparation Kit (Illumina), followed by 50 nt single-end sequencing on a HiSeq2500 (Illumina) in rapid run mode. Data have been deposited at NCBI GEO (accession GSE48886). After adapter trimming, reads were aligned to Arabidopsis miRNAs from miRBase 19 using bowtie 1 (Langmead et al., 2009).

3.2.6 RNA secondary structure analyses
Hybridization energy between miRNAs and corresponding targets was calculated using RNAcofold (from ViennaRNA; (Lorenz et al., 2011)) with default settings. Target site sequences were extended 6-nt upstream and downstream to avoid dangling ends. MFE of the predicted secondary structure flanking target site was calculated using RNALfold (ViennaRNA; (Lorenz et al., 2011)) with default settings. Upstream 500-nt and downstream 200-nt sequences were added to the target sequence as input.

3.3 Results

3.3.1 Perfectly paired sites can behave differently in ORF vs. 3’-UTR
Agrobacterium tumefaciens-mediated transient expression assays in Nicotiana benthamiana leaves (agroinfiltrations) have been widely used to study plant miRNA-target interactions (Llave et al., 2002; Allen et al., 2005; Franco-Zorrilla et al., 2007). I designed a dual-luciferase based reporter system to quantitatively assess target site efficacies through agroinfiltration (Figure 2.2). miRNA target sites were inserted into either the ORF (ORF sensor) or the 3’-UTR (3’-UTR sensor) of firefly luciferase (F-Luc) (Figure 2.2). Relative accumulation of the F-Luc sensor was evaluated at both the mRNA (via qRT-PCR with primers flanking the target site) and protein levels (by quantification of luciferase activities) using Renilla luciferase (R-Luc), expressed from the same T-DNA, as an internal control. By testing known functional targets of miR164 and miR398 it was demonstrated that the system was capable of quantitatively reflecting the efficacies of bona fide miRNA-target interactions.
The first question I am interested in is how efficient naturally occurring sites are comparing to perfectly pair sites. I addressed this question by testing the experimentally validated miRNA target sites from *Arabidopsis thaliana*: miR165a (based on REV, PHV and PHB) (Mallory et al., 2004b), miR160a (based on ARF17, ARF16, and ARF10) (Mallory, 2005), and miR156a (based on SPL2 and SPL3) (Schwab et al., 2005a; Gandikota et al., 2007). Perfectly paired sites served as positive controls. These miRNAs and target sites were chosen because they are members of deeply conserved miRNA families, and nearly all mismatches in these target sites are located in the 3' regions of the miRNAs (Figure 3.1). *N. benthamiana* small RNA-seq from leaves revealed a single miR156a and miR160a-related species whose sequences exactly matched their *Arabidopsis* counterparts, and two predominant miR165a isoforms with zero and one mismatch, respectively, relative to *Arabidopsis* miR165a (Table 3.1). In all cases the relevant *Arabidopsis* miRNAs were co-overexpressed. The previously validated miRNA targets were repressed significantly at both the mRNA and protein levels in our system, which demonstrated that our assay can reflect the bona fide effects of miRNA-target interaction in vivo. With sites placed in the ORF, efficacies for these targets were largely comparable to that of a perfectly complementary site (Figure 3.1), which is consistent with a previous study using artificially designed miRNA (Park et al., 2009). In contrast, I found that many native sites acted stronger than perfect sites when being tested in the 3'-UTR sensors (Figure 3.2). For example, *ARF16* and *ARF10*-like miR160 target sites located in the 3'-UTR were significantly more effective than a perfect site, despite the presence of two or three mismatches (Figure 3.2). Further investigation suggested that the difference between ORF and 3'-UTR efficacies was due to decreased performance of perfect sites when placed in the 3'-UTR. The efficacies of two perfect sites (miR156, miR165), especially at the protein level, dropped in 3'-UTR sensors comparing to ORF sensors (Figures 3.1 and 3.2; ~2 fold for miR156, \( p=8.01E-07 \); ~4 fold for miR165, \( p=9.62E-08 \); not significant for miR160; Mann-Whitney U-test). Diminished function of perfect sites when placed in the 3'-UTR has also been observed using an in vitro miRNA-target system (Iwakawa and Tomari, 2013).
Figure 3.1 Naturally occurring miR156, miR160, and miR165 target sites in ORF sensors. Boxplots summarize results from fifteen independent replicates and show the median (thick line), extent of the 1st to 3rd quartile range (box), values extending to 1.5 times the
interquartile range (whiskers), and outliers (black circles). Shaded area indicates that target protein down-regulation is significantly stronger than the corresponding perfect site (P<0.05, n=15, ANOVA-TukeyHSD). Synonymous negative controls (SNC) are negative controls with maximal disruption of their cognate sites while maintaining identical amino acid sequences. Target site sequences are shown in Appendix Table 1.1.

<table>
<thead>
<tr>
<th>N. benthamiana miRNA</th>
<th>Sequence, 5' to 3'</th>
<th>Relative abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR156a_related</td>
<td>UGACAGAAGAGAGUGAGCAC</td>
<td>100%</td>
</tr>
<tr>
<td>miR160a_related</td>
<td>UGCCUGGCUCCCUAUGCCA</td>
<td>100%</td>
</tr>
<tr>
<td>miR165a_related_1</td>
<td>UCGGACCAGCUUCAUCCCC</td>
<td>86.9%</td>
</tr>
<tr>
<td>miR165a_related_2</td>
<td>UCGGACCAGCUUCAUGCCCC</td>
<td>13.1%</td>
</tr>
</tbody>
</table>

**Table 3.1** Identification of *N. benthamiana* miRNAs
Nucleotides that are different from *Arabidopsis* miRNA sequences are labeled in red.

**Figure 3.2** Naturally occurring miR156, miR160, and miR165 target sites in 3’-UTR sensors.
As in Figure 3.1, except for 3'-UTR sensors.
3.3.2 The 5’ and 3’ regions of target sites are differentially tolerant of consecutive mismatches.

To systematically study target site sensitivity to mismatches on both ends of a target site, a series of miR160 targets with 0 to 13, or 0 to 6 consecutive mismatches relative to the miRNA 3’ and 5’ ends, respectively, were tested. Up to three consecutive mismatches at the miRNA 3’ end were tolerated without affecting site efficacy relative to the perfectly matched control (Figure 3.3). Target site efficacy was significantly diminished when four or five mismatches were introduced at the miRNA 3’ end, and no efficacy whatsoever was observed for sites with between six and 13 consecutive mismatches at the miRNA 3’ end (Figure 3.3). In contrast, a single mismatch at the miRNA 5’ end significantly diminished target site efficacy, and two or more consecutive mismatches at the miRNA 5’ end fully abolished it (Figure 3.3). The reduction in efficacy caused by single mismatch at the miRNA 5’ end (about 4 fold at the mRNA level, 2 fold at the protein level) is much stronger than data shown before based on an *in vitro* slicing assay (Mallory et al., 2004b). I attribute the difference in magnitude to the different systems being employed. In particular, I measured *in vivo* intact mRNA and protein levels as opposed to the *in vitro* accumulation of cleaved RNA fragments that were measured by Mallory et al. (Mallory et al., 2004b). Overall, these data are consistent with the canonical viewpoint that mismatches at the miRNA 5’ region are much more deleterious to target site function than those at the miRNA 3’ region (Mallory et al., 2004b; Schwab et al., 2005a; Lin et al., 2009). Importantly, our results demonstrate that this is quantitatively true when measuring efficacy at the level of protein accumulation. Finally, I note that all of the sites with between six and 13 consecutive mismatches at the 3’ end would be considered functional animal target sites (Bartel, 2009), as they all contain intact seed pairing in a 3’-UTR context; there was no evidence for efficacy for any of these target sites at either the mRNA or protein levels.
Figure 3.3 Differential sensitivity of the 5’ and 3’ target site regions to consecutive mismatches.
Engineered miR160 target sites in 3’-UTR sensors. Data display conventions as in Figure 2.4. Dual-asterisks indicate significant difference at the protein level between two adjacent groups (P<0.05, N=9, ANOVA-TukeyHSD).
3.3.3 Complementarity is a major determinant for site efficacy

I next examined whether the differences in target site efficacies could be attributed to inherent functional differences between different miRNAs, instead of the complementarity patterns. To test this, I performed target swap experiments where the patterns of the natural miR398 targets were engineered for miR160, and vice versa. Target site patterns that were highly effective for miR160 (ARF10, ARF16, and ARF17; Figures 3.1 and 3.2) were also highly effective for miR398 (Figure 3.4). In contrast, efficacies of targets with functional miR398 target patterns (CSD1, CSD2, and CCS1; Figures 2.4 and 2.5) were barely detectable with only miR160-CSD2 showing marginal repression at the protein level ($p=4.8E-04$, Mann-Whitney $U$-test; Figure 3.4). These data suggest that target site complementarity patterns are a major factor in determining target site efficacies.

![Figure 3.4](image)

**Figure 3.4** Complementarity patterns play a major role in target-site efficacy. Target base pairing pattern swap between miR160 and miR398 in 3'-UTR sensors. Data display conventions as in Figure 2.4. Shaded area indicates targets with protein level down-regulation significantly stronger than the perfect site ($P<0.05$, $n=9$, ANOVA-TukeyHSD).
3.3.4 Animal-like miRNA-target sites are ineffective

Since seed sites, the most typical pattern of animal miRNA targets, were not effective in our experiments (Figure 3.3), I then tested the functionality of two other target site patterns known to be functional in animals: centered sites (Shin et al., 2010), and 3'-supplemental sites (Reinhart et al., 2000), both of which have a relatively high degree of complementarity. Positive controls were based on known Arabidopsis miR164 and miR398 targets (NAC1 and CCS1, respectively) (Mallory et al., 2004a; Beauclair et al., 2010). I examined the efficacies of various miR164 and miR398 animal-like target sites in both ORF- (Figure 3.5A), and 3'-UTR sensors (Figure 3.5B). While the NAC1 and CCS1 positive controls were highly effective in both contexts, significant down-regulation was rarely detected for the miR164 or miR398 animal-like target sites regardless of ORF vs. 3'-UTR location, or mRNA vs. protein measurements (Figures 3.5A and 3.5B). Among the animal-like sites tested, only the seed site for miR164 and the 3'-supplemental site for miR398 resulted in a statistically significant amount of repression at the protein level, and only in the ORF sensor (Figure 3.5A). The magnitudes of these effects were very small (less than two-fold), not consistent for both miRNAs tested (Figure 3.5A), nor consistent in both the ORF and 3'-UTR contexts.

I considered the possibility that animal-like target sites could have been inactive only because the miRNA-AGO complexes were saturated with the overwhelmingly expressed sensors. This was a particular concern for the miR164 experiments, which relied solely upon endogenous miR164. I addressed this question by performing a series of agroinfiltrations using the miR164 3'-UTR sensors where the concentration of Agrobacterium inoculum was diluted in ten-fold increments. Analysis of R-Luc (which lacks engineered miRNA target sites) mRNA (Figure 3.6A) and protein (Figure 3.6B) accumulation showed that sensor expression responded to inoculum dilution in a linear fashion. The positive control NAC1 site had strong efficacy across the dilution series (Figure 3.6C). In contrast, sensors bearing animal-like sites remained ineffective across the dilution series, with only a few cases showing significant repression, and none more than two-fold (Figure 3.6C). These data argue against the hypothesis that lack of efficacy is due to saturation brought on by over-expression. Overall, I do not find compelling evidence suggesting that animal-like target sites can enable reliable target repression, regardless of location (ORF vs. 3'-UTR) or of measurement (mRNA vs. protein).
Figure 3.5 Little to no evidence for functionality of animal-like target sites. (A) Animal-like sites in ORF sensors. Data display conventions as in Figure 2.5. Synonymous negative controls (SNC) are negative controls with maximal disruption of
their cognate sites while maintaining identical amino acid sequences. Extruded dots represent single nucleotide bulges on targets. Red asterisks indicate statistically significant down-regulation (P<0.05, Mann-Whitney U-test).

(B) Animal-like sites in 3'-UTR sensors. A 21-nt spacer serves as common negative control for 3'-UTR-located target sites. Data display conventions as in Figure 2.4.

Figure 3.6 Animal-like sites remain non-functional regardless of sensor expression level. (A) Relative R-Luc mRNA accumulation levels across a ten-fold dilution series of Agrobacterium concentrations. Mean (n=24) relative R-Luc mRNA levels ± SD are shown, and along with the linear regression. (B) As in (A) except for R-Luc luciferase activities. (C) Relative F-Luc target mRNA accumulation and luciferase activity of 3'-UTR sensors across a dilution series of Agrobacterium concentrations. Data display conventions as in Figure 2.4. Results from two independent replicates are shown.
3.3.5 5mCSD1 and 5mCSD2 are not functional miR398 targets

A previous study showed that two Arabidopsis miR398 targets, CSD1 and CSD2, remained sensitive to miR398 induction even after the miRNA-target complementarity was severely disrupted (Dugas and Bartel, 2008). The disrupted variants, 5mCSD1 and 5mCSD2, lost regulation at the mRNA level but retained miR398-mediated repression of protein accumulation, suggesting that miRNA-induced translational repression in plants required a lower complementarity threshold than miRNA-induced reductions in mRNA levels. However, when placed in our F-Luc sensor, 5mCSD1 and 5mCSD2 sites were ineffective at both the mRNA and protein levels in both the 3'-UTR and ORF contexts (Figures 3.7A and 3.7B). Although it is possible that these sites require their native mRNA flanking sequence contexts to function, the simplest conclusion is that the severely disrupted 5mCSD1 and 5mCSD2 sites themselves are non-functional. The previously reported miR398-dependent repression of 5mCSD1 and 5mCSD2 protein levels could instead have been due to indirect, post-translational mechanisms. The fact that the CCS1 mRNA, which encodes a chaperone required for the stability of the CSD1 and CSD2 proteins, is also a miR398 target (Beauclair et al., 2010), provides a basis for this hypothesis. This result highlights a potential pitfall of assessing miRNA function via native target protein accumulation data alone: post-translational effects caused by other targets of the miRNA in question could confound such assays.
Figure 3.7 No evidence for functionality of 5mCSD1 or 5mCSD2 sites.

(A) Results from 3’-UTR sensors. Data display conventions as in Figure 2.4. 5mCSD1 and 5mCSD2 sites are miR398 targets with five mutations compared to CSD1 and CSD2 sites, respectively, as in Dugas et al. (2008).

(B) Results from ORF sensors. Data display conventions as in Figure 2.5.
3.3.6 Multiple pairing patterns with central mismatches are effective miRNA-target-mimics

Based largely on measurements of mRNA accumulation, pairing at the central positions, 9-11, is thought to be especially critical for plant miRNA function (Schwab et al., 2005a; Franco-Zorrilla et al., 2007; Todesco et al., 2010). Consistent with this hypothesis, I found that single nucleotide mismatches at positions 9 and 10, as well as combinations of mismatches at positions 9, 10, and 11, showed complete elimination of the responsiveness of a miR164-targeted 3'-UTR sensor (Figure 3.8A). Importantly, these results held true regardless of whether efficacy was measured at the mRNA or protein level (Figure 3.8A). Notably, Arabidopsis miR173 is known to direct AGO-catalyzed slicing of several non-coding RNA targets despite mismatches at positions 9 (TAS1a, TAS1d, and TAS2) or at both positions 9 and 10 (TAS1b) (Allen et al., 2005). However, the miRNA-programmed RNA-induced silencing complex in this case is likely to have unique properties because miR173 is a member of ancient superfamily of plant miRNAs that, instead of simply repressing target accumulation, are instead specialized to recruit RNA-dependent RNA polymerases to their targets (Allen et al., 2005; Cuperus et al., 2010b; Chen et al., 2010; Xia et al., 2013; Cuperus et al., 2010a).

Central bulges are also a hallmark of miRNA-target-mimics in plants, which have frequently been based on the At4/IPS1 family of non-coding RNAs and contain a central 3-nt asymmetric bulge (Franco-Zorrilla et al., 2007; Todesco et al., 2010). Sites with two consecutive central mismatches (not bulges) at positions 10 and 11 have also been demonstrated to perform as intermediate-strength target-mimics, and all types of target-mimicry sites appear ineffective at conferring repression when embedded in the 3'-UTRs of protein-coding transcripts (Ivashuta et al., 2011). Consistent with expectations, I observed that a canonical miRNA-target-mimic site for miR164 (MIM164) was unable to confer any repression in our system (Figure 3.8A). To test our series of central mismatch sites for mimicry activity, I took advantage of the fact that target-mimic sites are effective decoys when embedded into the 3'-UTR of GUS (Ivashuta et al., 2011). I thus designed a series of GUS-based competitors with various miR164 target sites in the 3'-UTRs, and co-agroinfiltrated them with a functional NAC1 dual-luciferase 3'-UTR sensor. As a negative control, GUS with a spacer site had no effect on miR164-mediated repression (Figure 3.8B). In contrast, the positive control (GUS with a MIM164 site in its 3'-UTR) had a strong decoy effect, fully preventing detection of miR164-mediated repression (Figure 3.8B). All variants with central mismatches (9m, 10m, 9m&10m, and 9m&10m&11m) were moderately effective target-mimics, though none were as strong as MIM164 (Figure 3.8B). These results are largely consistent with those of Ivashuta et al. (2011), except that, in our hands, a single mismatch site position 10 was a moderately effective mimic instead of a non-effective one. The discrepancy may be due to increased sensitivity in our assay, or perhaps to intrinsic differences between the miRNAs that were tested. I also found that the animal-like 3'-supplementary site pattern, based on a classic let-7 target site from C. elegans lin-41 (Reinhart et al., 2000), was a moderately effective target-mimic (Figure 3.8B). This was predictable, because the pattern of let-7 is similar to canonical plant target-mimicry sites. In contrast, centered and seed sites were ineffective target-mimics (Figure 3.8B). Finally, the NAC1 site, itself a highly potent target site in terms of repression (Figure 2.4) was ineffective as a miRNA-target-mimic.
(Figure 3.8B). I conclude that many complementarity patterns involving central mismatches flanked by two highly complementary regions, beyond the classic At4/IPS1-like pattern and the mismatch 10-11 pattern, are functional miRNA-target-mimics in plants. Our data based on target sites within 3’-UTRs reinforce the conclusion that effective target mimicry is mutually exclusive with effective repression (Ivashuta et al., 2011): all of the sites that function well as target-mimics failed to confer repression, and a site that confers robust repression failed to act as a target-mimic. However, I do not rule out the possibility that target mimic sites in other mRNA contexts may function differently. For instance, single sites with central mismatches mediate translational repression with catalytically active AGO1 in vitro only when placed in the 5’-UTR (Iwakawa and Tomari, 2013).
Figure 3.8 Multiple pairing patterns are effective miRNA-target-mimics.
(A) Results from central-mismatch containing sites in 3'-UTR sensors. Data display conventions as Figure 2.4.
(B) Results from a target-mimicry assay. Data display conventions as in Figure 2.4.
NAC1 3'-UTR site activity was measured in the presence of co-infiltrated GUS-based
competitors. Blue shaded area indicates effective competitors with median values of relative F-Luc repression less than two-fold. Peach shaded area indicates non-functional competitors with median relative F-Luc repression stronger than two-fold.
3.3.7 Blue Copper-binding Protein (BCBP) target site with a 6-nt bulge in target 5’region is repressed by miR398

Given the above results, it is therefore reasonable to speculate that targets with bulges in them are all likely to be slicing resistant. However, I found at least one exception. In collaboration with Dr. Nicolas Bouché at INRA, a novel miR398 target BCBP was identified by lessening restrictions during target prediction (Brousse et al., 2014). It contains a 6-nt bulge between position 6 and 7 relative to the miRNA 5’end. Endogenous level of BCBP mRNA and miR398 were shown be inversely correlated (Brousse et al., 2014). Consistently, mismatches introduced at cleavage site abolished response of BCBP mRNA to miR398. To further confirm the functionality of BCBP site I inserted it into my dual-luciferase reporters in both the 3’-UTR and ORF contexts. The results confirmed functionality of BCBP target site, which is independent of its native sequence context, and position (5’-UTR, Figures 3.9A and 3.9B). Interestingly, mismatches of the target 3’ most 2 nucleotides were required for target repression, particularly at protein level, but not RNA level (Figure 3.9A). This is consistent with my previous data showing that many native targets with mismatches near their 3’ ends act stronger than perfectly paired site when positioned in the 3’-UTR contexts (Figure 3.2).

Then I asked whether bulged site-mediated target repression is unique to miR398, or a generally shared figure among other miRNAs and targets. To address this question two sets of experiments were carried out in parallel. In the first place, I examined functionalities of other naturally occurring bulged pattern identified through the same criteria for BCBP. Despite a few sporadic small-range responses, there is no strong evidence that these alternative bulged sites can mediate efficient target repression (Figure 3.10). On the other hand, BCBP-like base pairing pattern with its 3’ end either paired (Figure 3.11A) or unpaired (Figure 3.11B) were tested for other conserved miRNAs, including miR156, 160, 164, and 165. None of these cases showed positive response (Figures 3.11A and 3.11B). I then examined the uniqueness of BCBP site in terms of the position of its bulge. A series of bulged sites with the 6-nt bulge scanning from target 5’ to 3’ end were tested in the presence over-expressed miR398. Interestingly, while positions at and after nucleotide 15 are generally tolerant of bulges (Figure 3.12), position between nucleotide 6 and 7 seemed to be the only site that allows presence of bulges in target 5’ end region (Figure 3.12). In conclusion, I demonstrated that miR398 is capable of repressing target at both the RNA and protein levels in spite of the presence of 6-nt bulge in the target 5’ region. The function of this target does not require its native sequence context, or its 5’-UTR positioning. However, it appeared to be a rather unique case of miRNA-target interaction. There is no evidence that BCBP bulged pattern is functional for other conserved miRNAs.
**Figure 3.9** *BCBP* is a functional miR398 target.

(A) Results from 3’-UTR sensors. Data display conventions as Figure 2.4.

(B) Results from ORF sensors. Data display conventions as Figure 2.5.
Figure 3.10 Multiple bulged sites show minimal target efficacy. Engineered miR398 target sites in 3’-UTR sensors following different bulged patterns. Data display conventions as in Figure 2.4.
Figure 3.11 *BCBP*-like bulged sites are not effective for multiple conserved miRNAs. (A) *BCBP*-p-like bulged sites are tested using 3'-UTR sensors. Data display conventions as in Figure 2.4. (B) *BCBP*-m like bulged sites are tested using 3'-UTR sensors.
**Figure 3.12** Effects of 6-nt bulge present at different positions. Engineered miR398 target sites in 3'-UTR sensors. Data display conventions as in Figure 2.4. Red dots on target 5’ and 3’ ends represent dangling ends as predicted by RNAcofold.
3.3.8 Summary of target site properties

Many studies have reported scoring schemes for computational predictions of plant miRNA targets (Rhoades et al., 2002; Allen et al., 2005; Schwab et al., 2005a; Alves et al., 2009). However, there is limited knowledge on the correlation between target prediction scores and target site efficacies, especially at the protein level. I thus scored the target sites used in our studies with respect to two different scoring schemes: the position-specific scoring matrix of Allen et al. (Allen et al., 2005), and the minimum free energy (MFE) ratio between the site and a perfectly complementary site. I also split the data into ORF and 3'-UTR categories to further reveal the positional effects on target efficacies. At extreme values, both methods in both sensors identified sharp decreases in efficacy after a critical threshold was reached; an Allen et al. score of four or higher (Figures 3.13A and 3.13C), or an MFE ratio of less than 0.75 were associated with marginal or non-effective sites (Figures 3.13B and 3.13D). When focusing only on 3'-UTR sensors, these scoring methods seemed inversely correlated with efficacy for cases above the threshold: the maximum median efficacies were observed for sites with Allen et al. scores of 1-1.5 (Figure 3.13A), and MFE ratios between 0.8 and 0.85 (Figure 3.13B). Although the differences were small and the p-values relatively large (due both to small sample size and variability within categories), this observation suggests that maximally effective target sites in plants are not necessarily perfectly complementary. However, using ORF sensors, I did not observe an inverse correlation (although the number of observations was extremely low). Instead, in the ORF, perfect sites had a similar distribution of efficacies compared to sites with slightly higher scores/lower MFE ratios. Nevertheless, the data imply that, for computational prediction of miRNA targets, it should not be assumed that the target site scoring regime necessarily predicts the quantitative efficacy of the target site since perfect sites are often equivalent or even sometimes less effective to sites with a few 3' mismatches.

Next, I compared overall efficacies between the ORF and 3'-UTR. ORF-located and 3'-UTR-located site efficacies were correlated in a one-to-one manner at the mRNA level (Figure 3.14A). However, target repression at the protein level was more effective in ORF sensors than in 3'-UTR sensors (Figure 3.14B, P<0.05, ANCOVA). Our data also allowed us to compare magnitudes of repression at the mRNA vs. protein levels. For 3'-UTR located sites, I found that the magnitude of repression measured at the mRNA level was frequently stronger than that measured at the protein level (Figure 3.14C, P<0.05, ANCOVA), while for ORF sites, mRNA and protein-based efficacies were correlated in a one-to-one manner (Figure 3.14C). I did not generally see stronger levels of repression at the protein level compared to the mRNA level (Figure 3.13C). However, this does not necessarily indicate that translational repression is not contributing to miRNA-mediated repression in our system. In the absence of specific measurements of both the translational efficiencies of the F-Luc mRNAs and the F-Luc protein turnover rates, mechanistic interpretations of these data are not feasible.
Figure 3.13 Correlation of target site scores with protein-level efficacies.

(A) Boxplots of relative target efficacy in 3’-UTR sensors at the protein level with respect to target penalty score. Target penalty score was calculated according to Allen et al. (2005) ($P=0.27$, Mann-Whitney $U$-test).

(B) As in (A) except for the MFE ratio between the sites and a perfectly complementary site. ($P=0.06$, Mann-Whitney $U$-test).

(C) As in (A) except for ORF sensors.

(D) As in (B) except for ORF sensors.
Figure 3.14 Protein-level repression is more effective for ORF sites than for 3′-UTR sites.

(A) Comparison of target repression at the mRNA level between 3′-UTR sensors and ORF sensors. Each dot represents the median target repression at the mRNA level in 3′-UTR (x value) and ORF (y value) sensors (n=26). The linear regression is shown (solid line), as is a reference line with slope of 1 (dotted). P-value represents the significance of difference between slope of the experimental regression line and the reference line (ANCOVA).

(B) As in (A), except for repression efficacies measured at the protein level (n=26).

(C) Correlation between target repression at mRNA levels and protein levels. Each dot represents the median mRNA (x value) and protein (y value) change of a unique target site. Data are pooled from miR164, miR398, miR156, miR160, and miR165 experiments.
(n=113). 3’-UTR targets and ORF targets are plotted separately in green and blue, respectively. The corresponding linear regression is shown (solid line), as is the reference line with slope of 1 (dotted). $P$-value represents the significance of difference between slope of the UTR-regression and ORF-regression (ANCOVA).
3.3.9 Changes in target site secondary structures are unlikely to have affected our measurements

Target site accessibility may be important for miRNA function in plants: GC-poor codons tend to flank known target sites (Gu et al., 2012) and transcriptome-wide RNA structure analysis indicates a slight increase in target site accessibility relative to flanking bases (Li et al., 2012a). In our experiments, I wished to minimize any effects of site accessibility by testing all sites in the same two flanking contexts (ORF and 3'-UTR sensor). However, it remained possible that the differing sequences of the sites themselves affected site accessibilities in our various sensors. Therefore, I tested whether inadvertent changes in F-Luc secondary structures involving the target sites might have affected our results. Predicted RNA secondary structures surrounding sensor target sites were generally very weak when compared against a control set of known functional structures (MIRNA hairpins from plants; Figure 3.15A). No correlation was observed between target efficacies and the MFE ratios of experimental vs. control target sites (Figure 3.15B). Additionally, in terms of absolute target site occlusion (the number of predicted intramolecular paired nucleotides within a target site), there were essentially no differences among all the sensors I tested (Figure 3.15C). I conclude that inadvertent changes in sensor site accessibilities are unlikely to have had a strong influence on our results, and that our data primarily reflect the inherent efficacies of the sites themselves.
Figure 3.15 No correlation between predicted target accessibility and protein-level target efficacy.

(A) Predicted RNA secondary structures surrounding sensor target sites are weak compared to plant MIRNA hairpins. Cumulative distributions of predicted thermodynamic stabilities. Black line, plant MIRNA hairpins. Red line, sensors.

(B) Scatterplots of target efficacy at protein level with respect to the MFE ratio between the experimental and negative control target sites (spacer for 3’-UTR sensors, or SNC for ORF sensors).
(C) Scatter-plot of target efficacy at protein level vs. the predicted relative target occlusion. Target occlusion is defined as the fraction of target site bases that are predicted to be intramolecularly paired. Ratios of occlusions were calculated between the experimental and corresponding negative control target sites (spacer for 3' UTR sensors, or SNC for ORF sensors).
3.4 Discussion

Our data demonstrates that, instead of being "all or nothing" regulators, naturally occurring target sites in plants have a range of inherent efficacies. This implies that targets are fine-tuned by miRNAs to acquire desired biological properties. Consistent with this idea, the variation of spike density among different barley cultivars (*Hordeum vulgare*) is a direct consequence of the varied efficacies of miR172 target sites due to slight variations in target site complementarities (Houston et al., 2013).

I found that the presence of mismatches near miRNA 3’ ends maintained optimal efficacy when placed in the 3'-UTR, in contrast to perfectly paired sites, whose performance decreased in in the 3'-UTR relative to the ORF. The differential performance of perfect sites in the ORF and 3’_UTR is consistent with prior observations suggesting that target sites located in the 5’ regions of mRNAs are often more effective (Li et al., 2013a; Iwakawa and Tomari, 2013). Our data suggest that this position-specific effect can be lessened when the target sites contain 3’ mismatches. Indeed, the fact that 3’ end mismatches enhance siRNA efficacy in human cells (De et al., 2013) argues that this is a conserved feature of AGO-bound small RNAs between plants and animals. This observation may also explain why many known target sites in plants possess mismatches at the 3’ end of the site as opposed to being perfectly complementary. One hypothesis to explain the effect of 3’-mismatches is that following AGO1-mediated endonucleolytic cleavage, the RISC complex disassociates more slowly from perfectly-paired target sites and thus delays the process of product release and subsequent turnover of the cleaved mRNA fragments. With perfect sites in 3’-UTR sensors, the protein coding region of the target is entirely contained within the upstream cleavage fragment, remains intact after AGO-catalyzed cleavage, and because of the delay in product release remains translatable for an extended period resulting in less efficient repression. This hypothesis could also account for the stronger effects upon mRNA levels vs. protein levels seen in the 3'-UTR sensors; our qRT-PCR assay for mRNA levels spans the cleavage site, so that cleaved sensor mRNA will not be detected, even if the upstream fragment remains stable and translatable for some time. I also observed other natural miRNA target sites with more extensive mismatches in the central and/or 5’ regions, and correspondingly suboptimal efficacies, such as the miR398 targets (Figures 2.4 and 2.5). This suggests that natural selection either tolerates, or perhaps has even selected for, sub-optimal miRNA efficacies for some targets. Consistent with this idea, a range of different target site efficacies was also observed for miR396 targets (Debernardi et al., 2012).

Single animal-like target sites, including seed, 3'-supplementary, and centered sites, are ineffective in our assay. Importantly, these results are true when quantitatively measured at both the mRNA and protein levels. It should be noted that in animals effects observed for most target sites are generally subtle, and miRNA target sites frequently act synergistically (Grimson et al., 2007; Baek et al., 2008; Guo et al., 2010). These data support the canonical view that plant miRNAs require a higher complementarity threshold to exert repression on targets than do animal miRNAs. Recently, it has been shown that in the green alga *Chlamydomonas reinhardtii* miRNA seed sites can confer ~20% repression (Yamasaki et al., 2013). Thus, the ineffectiveness of seed sites may be a...
trait unique to the land plants. The mechanistic reasons why animals and plants have differing complementarity thresholds for miRNA function are currently unclear.

Effective miRNA-directed target repression at either the protein or mRNA levels in our assay requires pairing at the central nucleotides and minimal mismatches in the 5’ region. In contrast, an expanded set of target site patterns, united by the presence of central mismatches flanked by highly paired regions on both sides, are effective target mimics in plants. This suggests that searches for new naturally occurring target mimics, which have already identified a large number of candidates (Wu et al., 2013; Ivashuta et al., 2011), should be expanded to include a wider variety of complementarity patterns.

The discory of BCBP site harboring six-nucleotide bulge near miRNA 5’ region suggest that our current knowledge about plant miRNA targeting rules are very likely incomplete and still expandable. Although BCBP-like base pairing patterns are not functional for other conserved miRNAs, and the natural position seems to be the only position compatible with the six-nucleotide bulge, its uniqueness still suggests that the current geometric rules of functional miRNA-target interactions are not always universally applicable. Outliers clearly exist, and miR398 might not be the only unconventional example.

It has been suggested that target site near start codon in either 5’-UTR or the open reading frame tend to induce stronger protein-level repression (Iwakawa and Tomari, 2013; Li et al., 2013a). Notably, BCBP site is located only seven-nucleotide upstream its start codon. Although we demonstrated that its RNA-level repression was independent of the target position in 5’-UTR, it worth further testing to see if this site is subject to additional protein-level repression to complement its relative low target efficacy in its native context.

The dual-luciferase sensor system allows rapid, quantitative testing of a large number of target sites. It also has the advantage of not being confounded by redundant systems of miRNA-induced post-translational control that can affect native target protein accumulation levels. In addition, sensor assays can ensure that the expression domains of the target and miRNA have 100% overlap, avoiding artifacts due to distinct cell- or tissue-type specific expression of native miRNAs and targets. However, the assay is based upon ectopic over-expression, in the context of a massive Agrobacterium tumefaciens infection. Thus, the system is potentially subject to artifacts arising both from saturation by over-expressed targets, and from the modulation of small RNA pathways by Agrobacterium infection (Dunoyer et al., 2006; Wang et al., 2011). Our titration assays indicate that saturation does not appear to be an issue (Figure 3.6), but nonetheless I cannot rule out entirely such artifacts. Therefore it is theoretically possible that some properties of this system do not entirely match those that occur under normal cellular conditions. For instance, it is possible that the apparent lack of effectiveness for animal-like targets is caused by low sensitivity or some other artifact particular to agroinfiltration. However, in this regard, it is worth pointing out that despite more than a decade of intense study of miRNAs, there are no reports showing positive data in favor of the hypothesis that animal-like target sites are functional in plants.
While miRNA-target complementarity patterns are clearly pivotal for miRNA function, there are several other factors that may contribute to miRNA target site efficacy. Changes in nucleotide compositions flanking target sites might influence local mRNA secondary structure resulting in variations in target site accessibility. Experimentally determined structural occlusions are slightly depleted within miRNA target sites relative to flanking residues (Li et al., 2012b), and GC-poor codons are enriched around target sites (Gu et al., 2012). Both observations suggest that selection has favored minimizing local mRNA secondary structure surrounding miRNA target sites. Target site accessibility could also be influenced by the positioning of various mRNA-binding proteins, which at present are very hard to accurately predict \textit{a priori}. The relative position on the target mRNA may also play a role in regulating target-site efficacy. I observed that perfectly matched sites were often less effective in the 3'-UTR relative to the ORF, and the same is true in an \textit{in vitro} AGO1 targeting system (Iwakawa and Tomari, 2013). In a systematic analysis of multiple artificial miRNAs (a-miRNAs) all with similar or identical levels of very high target site complementarity, sites located towards the 5' of the mRNA target tended to be more effective (Li et al., 2013a). Our development of a rapid and quantifiable system to measure miRNA target-site function \textit{in vivo} will facilitate further study of the role of mRNA secondary structure and target-site positioning. Mechanisms of miRNA-initiated target repression can also differ between cell types (Grant-Downton et al., 2013), and of course the miRNA and a mRNA bearing a potential target site must accumulate in at least some of the same cell type(s) or tissues in order for a site to be measurably functional. Future studies in this area should focus on incorporating all of these parameters to produce a more refined model for the efficacies of miRNA target sites in plants.
Chapter 4
Annotation and quantification of small RNA producing loci in *N. benthamiana* genome

4.1 Introduction

*N. benthamiana* has been traditionally used as a model host plant for plant-microbe interactions, and virology. Thereafter, small RNA community began to benefit tremendously from this system, because it is highly amenable to rapidly and robustly expressing various small RNAs and target transcripts upon delivery by *Agrobacterium* (Kurihara et al., 2006; Chen et al., 2010; Montgomery et al., 2008; Allen et al., 2005; Llave et al., 2002; Franco-Zorrilla et al., 2007). However, little is known about its small RNA populations. Given its great importance and utility, there is an urgent need in understanding the properties of its small RNA background.

Parallel efforts were made recently on sequencing and assembling the *N. benthamiana* genome (Naim et al., 2012; Bombarely et al., 2012). It thus became feasible to annotate small RNA producing loci using the reference genome. So far, great efforts have been made on computationally predicting and annotating miRNA producing loci utilizing deep-sequenced small RNA libraries, including that from our own lab (Yang and Li, 2011; Xie et al., 2012; Qian et al., 2012; Axtell, 2013b). Comparison of performances of these packages using common sets of data source was also conducted with respect to sensitivity, accuracy, and computational time (Axtell, 2013b). In general, several key principles are shared among these tools: In the first place, small RNA reads are adaptor-trimmed, quality-filtered, and aligned to reference genome using aligner of choice; Secondly, precursor sequences covering small RNA mapped region are retrieved, and subjected to *in silico* secondary structure prediction. Thirdly, additional criteria, including presence of two-nucleotide 3’-overhang of miRNA/miRNA star duplex, gap size between miRNA/miRNA star, length and amount of mismatches at stem region, thermodynamic stability of the stem-loop structure, and size distribution of small RNA population within the locus, are being considered (Meyers et al., 2008). Parameter settings that work well for animal data sets can not be directly applied to plants, because plant miRNA precursors are of much higher heterogeneities (Axtell, 2013b). In an ideal world, mature miRNAs should account for the most abundant small RNA species coming from a defined *MIRNA* locus, and form a distinct signal peak that can be easily distinguished from other non-miRNA backgrounds. However, in the real world this is hardly the truth. Strict parameters are thus applied to ensure high accuracy, although sensitivity might be compromised a bit (Axtell, 2013b).

miRNAs are registered in miRBase, which serves as the repository of miRNA annotations (Kozomara and Griffiths-Jones, 2014). The most recent release contains 30,424 mature miRNAs derived from 24,521 *MIRNA* loci in 206 species. No *N. benthamiana* miRNA sequence is currently registered. However, 164 mature miRNAs from *N. tabacum*, another model plant organism for virology, and also a close relative of *N. benthamiana*, were registered.
4.2 Data analysis

Two replicate RNA samples from one-month old *N. benthamiana* leaves were extracted. Small RNA library was prepared using TruSeq Small RNA Sample Preparation Kit (Illumina), followed by 50 nt single-end sequencing on a HiSeq2500 (Illumina) in rapid run mode. Two RNA samples resulted in two replicate small RNA libraries (1U and 1V). Data were deposited at NCBI GEO (accession GSE48886). Sequencing adaptors were trimmed using 'trim_illumina_sRNA_fastq.pl' version 0.3 (http://axtell-lab-psu.weebly.com/tools.html). The reference genome of *N. benthamiana* (assembly version 0.4.4) was obtained from sol genomics network FTP site (ftp://ftp.solgenomics.net/genomes/Nicotiana_benthamiana/).

As a first step, inverted repeats were identified using einverted, from the EMBOSS package (version 6.6.0) (Rice et al., 2000). ShortStack (version 1.2.4) *de novo* run was then carried out under default settings by passing the following files into the program: Query, adaptor trimmed small RNA libraries, 1U and 1V; Reference genome, *N. benthamiana* v0.4.4 scaffolds; and inverted repeats file obtained above.

4.3 Results

4.3.1 Overview of small RNA analysis

Two replicates of small RNA libraries (1U and 1V) prepared from *N. benthamiana* one-month old leaves were sequenced. A total of 13,859,162 and 15,373,285 reads were generated from 1U and 1V, respectively (Table 4.1). ShortStack annotation generates three types of small RNA loci, MIRNA loci, hairpin RNA loci (hp_RNA), and si_RNA loci. The analysis result in a total of 43652 annotated loci, of which 177 were MIRNA loci, 4573 were hp_RNA loci, and 38902 were siRNA loci (Table 4.2, Figure 4.1).

Loci with greater than 80% small RNA reads between 20 to 24 nts in length were annotated as Dicer-derived, with the rest being annotated as non-Dicer-derived. Within each type of small RNA loci, for example MIRNA loci, all Dicer-derived loci were further categorized into five different subgroups based on the predominant small RNA length (Dicercall of 20 to 24 nt). It appeared that MIRNA loci were dominated by clusters of Dicercall at 21, whereas small RNAs from the other two types of loci are mainly in size of 24 nts (Figure 4.1).
Table 4.1 Summary of *N. benthamiana* small RNA mapping.

<table>
<thead>
<tr>
<th>Source</th>
<th>Accession</th>
<th>Library</th>
<th>Reads mapped</th>
<th>Reads unmapped</th>
<th>Fraction mapped</th>
<th>Fraction unmapped</th>
</tr>
</thead>
<tbody>
<tr>
<td>BTI</td>
<td>PRJNA170566</td>
<td>1U</td>
<td>13,560,920</td>
<td>298,242</td>
<td>97.8%</td>
<td>2.2%</td>
</tr>
<tr>
<td></td>
<td>PRJNA170566</td>
<td>1V</td>
<td>15,045,040</td>
<td>328,245</td>
<td>97.9%</td>
<td>2.1%</td>
</tr>
</tbody>
</table>

Table 4.2 Summary of *N. benthamiana* small RNA loci.

<table>
<thead>
<tr>
<th>DicerCall</th>
<th>Number of Loci</th>
<th>Number of alignments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-HP</td>
<td>HP</td>
</tr>
<tr>
<td>N</td>
<td>10007</td>
<td>10</td>
</tr>
<tr>
<td>20</td>
<td>49</td>
<td>19</td>
</tr>
<tr>
<td>21</td>
<td>627</td>
<td>251</td>
</tr>
<tr>
<td>22</td>
<td>1525</td>
<td>210</td>
</tr>
<tr>
<td>23</td>
<td>375</td>
<td>77</td>
</tr>
<tr>
<td>24</td>
<td>26319</td>
<td>4006</td>
</tr>
</tbody>
</table>

Figure 4.1 The fraction of loci with different Dicer-call.
4.3.2 Repetitiveness of *N. benthamiana* small RNA genes

I then examined the repetitiveness of all three types ShortStack annotated small RNA loci. One factor being studied was the fraction of highly repetitive loci called “shadow loci”, of which no uniquely mapped reads were found. As expected, fraction of shadow loci for *MIRNA* clusters is significantly lower than that of hp_RNAs and siRNAs (Figure 4.2A). This is in line with current knowledge that *MIRNA* genes are usually present in gene rich regions with low degree of repetitiveness, whereas siRNAs are more often derived from repeat-rich heterochromatic regions. In addition, small RNA loci producing longer small RNA are generally less repetitive than those give rise to short small RNAs (Figure 4.2A).

Another important parameter reflecting repetitiveness, called uniqueness index (UI), is obtained by calculating the ratio of repeat-normalized reads to the total mappings. For a given locus, this value could vary from zero (highly repetitive) to one (highly unique). As expected, distribution of *MIRNA* clusters of both 20-22 nt and 23-24 nt types showed high enrichments in high uniqueness index ends (Figures 4.2C and 4.2D). In contrast, small RNA clusters with a Dicer call of “N”, composed of large amount of RNA fragments due to non-specific degradation from rRNAs and tRNAs, showed extremely low uniqueness (Figure 4.2B). Surprisingly, 23-24 nt siRNA clusters also show high uniqueness index, which was not expected for heterochromatic siRNAs (Figure 4.2H). I reasoned this is likely due to incompleteness of genome assembly and annotation, where a large amount of repetitive elements have yet been assembled and annotated. Alternatively, if this is the real case that would suggest *N. benthamiana* 24-nt siRNAs indeed tend to be derived from non-repetitive regions. In order to better dissect the question, I further looked at genomic co-occupancy between different types of small RNA clusters and genomic features, such as exon, intron, intergenic, and repeat region.
Figure 4.2 Many *N. benthamiana* small RNA loci showed low repetitiveness. (A) Barplot showing the fraction of ‘shadow’ loci for each class of small RNA loci. (B-H) Histogram showing distribution of uniqueness indices for different class of small RNA clusters.
4.3.3 Genomic features of *N. benthamiana* small RNA genes

Both absolute number of overlapping loci (Table 4.3), and non-redundant nucleotides were calculated (Table 4.4). The relative enrichment was obtained by calculating the ratio of the amount of observed over-lapping (in nucleotides) to that of expected over-lapping, using following formula

\[ R = O/E = O / [(x/g) * (y/g) * g] = O^*g / x*y \]

Where \( R \) represents relative accumulation, \( O \) is observed over-lapping in nucleotides, \( E \) is expected over-lapping in nucleotides assuming that the two features are randomly positioned, \( x \) is amount of non-redundant nucleotides of feature 1, \( y \) is amount of non-redundant nucleotides of feature 2, and \( g \) is the length of genome in nucleotides. In general, all short sized small RNA loci as well as long sized *MIRNA* loci are enriched in gene regions compare to intergenic region (Figure 4.3). *MIRNA* loci are particularly depleted from repeat rich regions (Figure 4.3). On the other hand, 23-24 siRNA loci are more likely to be present in TE regions (Figure 4.3). This is in line with properties of heterochromatic siRNAs present in other plant species, and also suggests that high uniqueness index observed for *N. benthamiana* 24-nt siRNAs are very likely due to incomplete genome assembly.

<table>
<thead>
<tr>
<th></th>
<th>Total loci in category</th>
<th>Genes</th>
<th>Exons</th>
<th>Introns</th>
<th>Intergenic</th>
<th>Repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Dicer</td>
<td></td>
<td>76379</td>
<td>338652</td>
<td>270148</td>
<td>210373</td>
<td>7951064</td>
</tr>
<tr>
<td>20-22 nt <em>MIRNA</em></td>
<td></td>
<td>150</td>
<td>47:48</td>
<td>17:17</td>
<td>39:41</td>
<td>106:104</td>
</tr>
<tr>
<td>23-24 nt <em>MIRNA</em></td>
<td></td>
<td>20</td>
<td>9:10</td>
<td>2:2</td>
<td>7:8</td>
<td>11:11</td>
</tr>
<tr>
<td>23-24 nt siRNA</td>
<td></td>
<td>482</td>
<td>76:68</td>
<td>5:5</td>
<td>72:71</td>
<td>408:345</td>
</tr>
</tbody>
</table>

*The tallies of overlapping loci of each type are shown in the format [n overlapping of type row] : [n overlapping of type column]*

**Table 4.3** Overlaps between small RNA loci and other genomic features by tallies.
Table 4.4 Overlaps between small RNA loci and other genomic features by number of nucleotides.

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>Genes</th>
<th>Exons</th>
<th>Introns</th>
<th>Intergenic</th>
<th>Repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Dicer</td>
<td>361812614</td>
<td>76867545</td>
<td>284365052</td>
<td>2232559463</td>
<td>1274054173</td>
</tr>
<tr>
<td>20-22 nt MIRNA</td>
<td>1414284</td>
<td>98538</td>
<td>38099</td>
<td>1315746</td>
<td>769028</td>
</tr>
<tr>
<td>20-22 nt hpRNA</td>
<td>27297</td>
<td>10460</td>
<td>1753</td>
<td>8707</td>
<td>16837</td>
</tr>
<tr>
<td>20-22 nt siRNA</td>
<td>106829</td>
<td>39429</td>
<td>7711</td>
<td>31718</td>
<td>67400</td>
</tr>
<tr>
<td>23-24 nt MIRNA</td>
<td>162371</td>
<td>39197</td>
<td>16660</td>
<td>22537</td>
<td>123174</td>
</tr>
<tr>
<td>23-24 nt hpRNA</td>
<td>6895</td>
<td>3783</td>
<td>826</td>
<td>2957</td>
<td>3112</td>
</tr>
<tr>
<td>23-24 nt siRNA</td>
<td>429590</td>
<td>106209</td>
<td>5214</td>
<td>100995</td>
<td>323381</td>
</tr>
</tbody>
</table>

*The tallies of overlapping loci of each type are shown in the format [n overlapping of type row] : [n overlapping of type column]*

Log$_2$ (observed overlap / expected overlap) for each of the pairwise comparison shown. Cell values are shown in white texts.

**Figure 4.3** Analysis of co-occupancy between small RNA loci and other genomic features.
4.3.4 Identification of MIRNAs with and without known plant homologs

One of the questions I am mostly interested in is what miRNA families are present in *N. benthamiana*, and how abundant they are. To address this question, I compared sequences of all 177 annotated MIRNA loci and of their corresponding mature miRNAs to those of currently known plant miRNAs (miRBase release 20) using bowtie 1.0.0 allowing maximally three mismatches (Langmead et al., 2009; Kozomara and Griffiths-Jones, 2014). As a result, of 177 MIRNA loci, 95 were found belong to 37 known miRNA families, whereas 82 loci with no homology to known miRNAs were annotated as novel MIRNA clusters, and they were grouped into 79 miRNA families (Figure 4.4). As expected, the five conserved miRNAs used in Agro-infiltration mentioned above were among the 37 known miRNA families. Following this categorization, the expression level of the 37 known miRNA families in *N. benthamiana* was also examined (Figure 4.5). They were found to express at various levels spanning a range of a few thousand-fold (Figure 4.5). Notably, miR165 family has the highest expression level among all of them (Figure 4.5).

It is evidenced that small RNAs of different length are products of different Dicer family proteins, with varied degree of heterogeneity. For examples, in *Arabidopsis*, DCL1 generates 21-nt miRNAs of high homogeneity, whereas DCL3 gives rise to 24-nt siRNAs with great variation in sequence identity (R et al., 2006b; Xie et al., 2004). Since I identified two types of MIRNA loci that are dominated by short (20 to 22-nt) and long (23 to 24-nt) miRNAs, respectively, I am interested to see if the two groups are quantitatively identical in terms of precision of precursor processing. I thus calculated the processing precision score, which is the ratio between reads of annotated mature miRNA and star sequence to total amount of reads mapped within the same locus. The results showed 23 to 24-nt MIRNA loci are generally processed more heterogeneously compared to canonical 20 to 22-nt MIRNA loci (Figure 4.6), which suggest that different machinery might be involved during biogenesis.
Figure 4.4 Number of MIRNA loci and families identified for both known and novel groups.

Figure 4.5 Expression level of known MIRNA loci identified.
Figure 4.6 20-22 nt *N. benthamiana* MIRNAs are processed more precisely than 23-24 nt MIRNAs.

### 4.4 Discussion

Utilizing a draft genome and small RNAseq data, I were able to *de novo* annotate and quantify small RNA producing loci in *N. benthamiana*, which has been extensively utilized in plant-microbe interaction and small RNA function related studies. The depth of sequencing (>13 million) and high mapping ratio (>97%) provided us solid base for downstream analysis.

It has been shown that siRNAs, especially the canonical heterochromatic 24-nt siRNAs were associated with repeat-rich regions (Axtell, 2013a; Law and Jacobsen, 2010; Pontier, 2005; Kasschau et al., 2007; Nobuta et al., 2008). In cases where multiple mappings occur, current aligners would arbitrarily assign a position among multiple mapped regions assuming that each mapped site contributes equally to the accumulation of the corresponding reads, which is not likely to be the truth. Thus, a better aligning strategy giving more weight to high-confidence loci will certainly minimizes the false positive rate and improves the annotation fidelity. A promising strategy for identifying high-confidence loci would be based on the abundance of uniquely mapped reads. In brief, loci that generate higher amount of uniquely mapped reads are more likely to be the real source of those multiple mapped reads.

Repetitiveness analysis suggests that all annotated MIRNA loci (DicerCall between 20 and 24 nt) are highly unique (Figures 4.2c and d), in terms of the portion of uniquely mapped small RNA sequences. Whereas siRNA loci giving rise to short siRNAs (mainly 20 to 22 nt) has very low uniqueness index (Figure 4.2g). This is consistent with
canonical view that siRNA are usually associated with highly repetitive regions. However, to our surprise, siRNAs from loci of DicerCall 23 to 24 nt all tended to be uniquely mapped (Figure 4.2h). Whether this is in general an unique signature of *N. benthamiana* heterochromatic siRNA or due to the incomplete genomic assembly of repetitive elements deserves further scrutiny. Nevertheless, when compare the locations of small RNA loci to other genome-wide features, I found these heterochromatic siRNA loci (DicerCall of 23 to 24 nt) were more likely to be enriched in repetitive regions (Figure 4.3), and *MIRNA* loci were greatly depleted from these regions (Figure 4.3).

Throughout the analysis, hpRNA and siRNA clusters appear to be largely indistinguishable. They both were dominated by loci of DicerCall 24 (Figure 4.1), which suggests similar, if not the same, small RNA biogenesis pathway was being deployed. It should be noted that computationally predicted hairpin structures are not necessarily present in reality. It is possible that biogenesis of small RNAs from many of the annotated hpRNA loci does not actually involve a single-strand self-fold precursor. The key criterion to determine the type of precursor during small RNA production is whether RNA-dependent RNA Polymerase activity is involved. Without such information, the real relevance of a predicted hairpin precursor should be taken with caution.
Chapter 5
Summary and prospects

5.1 Summary

By coupling Agro-infiltration technique with dual-luciferase reporters I successfully developed a transient assay for quantifying the strengths of miRNA-target interactions. This experimental system offers several advantages: First, it is highly amenable to testing a great number of miRNA-target pairs. Sites of interest can be synthesized and ordered directly and inserted into sensor with two choices of targeting positions by restriction digestion. Secondly, the use of dual-luciferase sensors helps mitigate interference from endogenous gene network, that often confound analysis using native gene as miRNA readout. Thirdly, the high sensitivity of Luciferase sensors allows accurate detection of subtle changes in target efficacy due to altered target complementarity. In addition, the combined in vivo assay at both the RNA and protein levels provide a comprehensive picture of miRNA effects, which are quantitatively assessed at both levels.

Using the dual-luciferase reporter system, I was able to tackle several long-standing questions. One of the most important discoveries was that many of the native targets with hallmarks of mismatches near the miRNA 3’ ends are very often equally, or even more, effective than perfectly paired sites. It suggests that base pairings at this region are not required for efficient target recognition or repression. Instead, plant targets may have evolved to specifically avoid paring near miRNA 3’ ends. In my study, the observation that target efficiencies (protein level) generally drop when targeting positions were shifted from ORF to 3’-UTR, which is particularly evident for perfectly paired site, very well explains the choice plants made during evolution. However, how ORF targeting is mechanistically different from 3’-UTR targeting became another interesting question awaits further exploration.

The presence of central bulge at target cleavage site is a strong indicator of its slicing resistance. I demonstrated that besides bulges, central mismatches could also render targets unresponsive to miRNA targeting. Moreover, target repression is mutually exclusive with target function as mimicry. All sites that are resistant to miRNA slicing were shown to protect and elevate levels of alternative sliceable sites in the presence of miRNA. Function of target mimicry doesn’t require specific positioning, because both ORF and 3’-UTR located sites are able to compete for miRNA targeting.

I also provided evidence that base pairing patterns that are typical for animal miRNA-target interactions are indeed too low for plant counterparts. This is true under a broad range of ratios between miRNA and targets (10^5). Also I ruled out the possibility that targets of lowered complementarity could be exclusively channeled into translation repression. Instead, the indirect control through alternative genes, which in turn is responsive to miRNA, is likely to be the real cause of protein level change of 5mCSD site observed in earlier study.
High throughput sequencing combined with computational analysis was demonstrated to be a powerful tool for de novo identification of *N.benthamiana* small RNA species and corresponding producing loci. A total of 177 MIRNA loci were recovered during this analysis, followed by family categorization. The expression level as well as sequence identities of each miRNA species, especially those related to Agro-infiltration experiments, were analyzed, which offered strong background support for dissecting effects of complementarity.

5.2 Prospects

It was demonstrated that center-mismatched site could mediate protein-level repression when positioned in sensor’s 5’-UTR region (Iwakawa and Tomari, 2013). Experiments using artificial miRNA targeting different positions also indicates the positive link between proximity of target site to ORF 5’ ends and protein-level repression (Li et al., 2013a). However, positional effects on miRNA targeting efficiency with respect to mRNA and protein response have not been studied systematically. Experiments addressing the mechanisms underlying miRNA-target interactions with respect to target positions are also worth exploring.

In my experiments, targets efficacies were all tested in fixed contexts, either 3’-UTR or ORF. Change in flanking sequences were long hypothesized to be an important factor for target efficiency. However, influences of flanking sequences on target response to miRNA were never systematically studied. A system, where nucleotides of miRNA target flanking site can be randomized without affecting encoded protein sequence would offer insightful information.

The basic principle underlying effects of sequence complementarity on target regulatory efficiency has been largely settled. The other two topics discussed above (positional effects and influence from flanking sequences) become the next important ones to tackle.
Appendix
List of Supplemental Table

Supplemental Table 1.1 Sequence of target inserts ...............................................76

Supplemental Table 1.2 Primers used for making and quantifying dual-luciferase sensors .................................................................80
<table>
<thead>
<tr>
<th>MicroRNA</th>
<th>Target</th>
<th>Sensor</th>
<th>Sequence and Complementarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR164</td>
<td>Spacer UTR</td>
<td>5’-AUCGGUACCAGCCUCCCA-3’ Spacer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NAC1 UTR/ORF</td>
<td>5’-AGCGGACCCGCAGCAG-3’ NAC1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NAC1_SNC UTR/ORF</td>
<td>5’-UGCAGCCGACGCAGCAG-3’ NAC1_SNC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Seed UTR/ORF</td>
<td>5’-AGGAGACCGACCCGCAGCAG-3’ Seed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Seed_SNC ORF</td>
<td>5’-AGGAGACCGACCCGCAGCAG-3’ Seed_SNC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Perfect UTR/ORF</td>
<td>5’-CGCAGCAGCAGCAGCAG-3’ Perfect</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Perfect_SNC ORF</td>
<td>5’-CGCAGCAGCAGCAGCAG-3’ Perfect_SNC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3’Supplement UTR/ORF</td>
<td>5’-AGGAGACCGACCCGCAGCAG-3’ 3’Supplement</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3’Sup_SNC ORF</td>
<td>5’-AGGAGACCGACCCGCAGCAG-3’ 3’Sup_SNC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MIN164 UTR/ORF</td>
<td>5’-UGGAGACCGACCCGCAGCAG-3’ MIN164</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MIN164_SNC ORF</td>
<td>5’-UGGAGACCGACCCGCAGCAG-3’ MIN164_SNC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Center UTR/ORF</td>
<td>5’-AGGAGACCGACCCGCAGCAG-3’ Center</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Center_SNC ORF</td>
<td>5’-AGGAGACCGACCCGCAGCAG-3’ Center_SNC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CUC1/2 UTR</td>
<td>5’-AGGAGACCGACCCGCAGCAG-3’ CUC1/2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AT3G12977 UTR</td>
<td>5’-AGGAGACCGACCCGCAGCAG-3’ AT3G12977</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AT5G07680 UTR</td>
<td>5’-AGGAGACCGACCCGCAGCAG-3’ AT5G07680</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NAC6 UTR</td>
<td>5’-AGGAGACCGACCCGCAGCAG-3’ NAC6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AT5G61430 UTR</td>
<td>5’-AGGAGACCGACCCGCAGCAG-3’ AT5G61430</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9m UTR</td>
<td>5’-AGGAGACCGACCCGCAGCAG-3’ 9m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10m UTR</td>
<td>5’-AGGAGACCGACCCGCAGCAG-3’ 10m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9&amp;10m UTR</td>
<td>5’-AGGAGACCGACCCGCAGCAG-3’ 9&amp;10m</td>
<td></td>
</tr>
<tr>
<td>miR156</td>
<td>5'-GAGCGGUGACCAGAGAACACA-3'</td>
<td>ARF10_like</td>
<td>UTR</td>
</tr>
<tr>
<td>miR160</td>
<td>5'-GCCGCCCAUCGAGGCAGCGCA-3'</td>
<td>ARF16_like</td>
<td>UTR</td>
</tr>
</tbody>
</table>

**SPL3**

| UTR/ORF | 5'-UUCGGUACCCUCUCUCUGCU-3' | SPL3 | UTR/ORF | 5'-GCCGCCCAUCGAGCGCA-3' | SPL2 | UTR/ORF | 5'-GCCGCCCAUCGAGCGCA-3' |

**miR156**

| ORF | 5'-GCCGCCCAUCGAGCGCA-3' | SPL3_SNC | ORF | 5'-GCCGCCCAUCGAGCGCA-3' | SPL2_SNC | ORF | 5'-GCCGCCCAUCGAGCGCA-3' |

**miR160**

| ORF | 5'-GCCGCCCAUCGAGCGCA-3' | ARF10 | UTR/ORF | 5'-GCCGCCCAUCGAGCGCA-3' | ARF10 | UTR/ORF | 5'-GCCGCCCAUCGAGCGCA-3' |

**Perfect**

| UTR/ORF | 5'-GCCGCCCAUCGAGCGCA-3' | Perfect | UTR/ORF | 5'-GCCGCCCAUCGAGCGCA-3' | Perfect | UTR/ORF | 5'-GCCGCCCAUCGAGCGCA-3' |

**Perfect_SNC**

| ORF | 5'-GCCGCCCAUCGAGCGCA-3' | ARF16 | UTR/ORF | 5'-GCCGCCCAUCGAGCGCA-3' | ARF16 | UTR/ORF | 5'-GCCGCCCAUCGAGCGCA-3' |

**ARF10**

| ORF | 5'-GCCGCCCAUCGAGCGCA-3' | ARF16 | UTR/ORF | 5'-GCCGCCCAUCGAGCGCA-3' | ARF16 | UTR/ORF | 5'-GCCGCCCAUCGAGCGCA-3' |

**ARF16**

| ORF | 5'-GCCGCCCAUCGAGCGCA-3' | ARF17 | UTR/ORF | 5'-GCCGCCCAUCGAGCGCA-3' | ARF17 | UTR/ORF | 5'-GCCGCCCAUCGAGCGCA-3' |

**ARF17**

| ORF | 5'-GCCGCCCAUCGAGCGCA-3' | ARF17 | UTR/ORF | 5'-GCCGCCCAUCGAGCGCA-3' | ARF17 | UTR/ORF | 5'-GCCGCCCAUCGAGCGCA-3' |

**Perfect**

| UTR/ORF | 5'-GCCGCCCAUCGAGCGCA-3' | Perfect | UTR/ORF | 5'-GCCGCCCAUCGAGCGCA-3' | Perfect | UTR/ORF | 5'-GCCGCCCAUCGAGCGCA-3' |

**Perfect_SNC**

| ORF | 5'-GCCGCCCAUCGAGCGCA-3' | CCS1_like | UTR | 5'-GCCGCCCAUCGAGCGCA-3' | CCS1_like | UTR | 5'-GCCGCCCAUCGAGCGCA-3' |

**CSD1_like**

| UTR | 5'-GCCGCCCAUCGAGCGCA-3' | CSD1_like | UTR | 5'-GCCGCCCAUCGAGCGCA-3' | CSD1_like | UTR | 5'-GCCGCCCAUCGAGCGCA-3' |

**CSD2_like**

| UTR | 5'-GCCGCCCAUCGAGCGCA-3' | CSD2_like | UTR | 5'-GCCGCCCAUCGAGCGCA-3' | CSD2_like | UTR | 5'-GCCGCCCAUCGAGCGCA-3' |

3'-1m

| UTR | 5'-GCCGCCCAUCGAGCGCA-3' | 3'-1m | UTR | 5'-GCCGCCCAUCGAGCGCA-3' | 3'-1m | UTR | 5'-GCCGCCCAUCGAGCGCA-3' |
### Supplemental Table 1.2 Primers used for making and quantifying dual-luciferase sensors

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence, 5’ to 3’</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GUSF</strong></td>
<td>TGACTCGAGATGGTACGTCCTGTAGAAAC</td>
<td>Amplify GUS ORF</td>
</tr>
<tr>
<td><strong>GUSR</strong></td>
<td>CAGGAATTCTCATTTGCTGCCTCCCTGCTG</td>
<td>Amplify GUS ORF</td>
</tr>
<tr>
<td><strong>35SF</strong></td>
<td>GGGCCCACCCCCCTACTCCAAAAATG</td>
<td>Amplify 35S promoter</td>
</tr>
<tr>
<td><strong>35SR</strong></td>
<td>CTCGAGGTCCTCTCCAAATGAAATG</td>
<td>Amplify 35S promoter</td>
</tr>
<tr>
<td><strong>TerF</strong></td>
<td>TCCCAGTCATTTAGCTGAGTTTCCC</td>
<td>Amplify 35S terminator</td>
</tr>
<tr>
<td><strong>TerR</strong></td>
<td>GTGCCGGCGCCCGTATTGTGTTTTAGTCTGGA</td>
<td>Amplify 35S terminator</td>
</tr>
<tr>
<td><strong>UTRs</strong></td>
<td>GTCCGCGTGGTCTATTACACTATCCCTCCTAGGATCGGTATGCAGACCGG</td>
<td>Hybridize with UTRas to obtain artificial 3’UTR</td>
</tr>
<tr>
<td><strong>UTRas</strong></td>
<td>CTAGTTCAAGAGTAAAAGATAGTAAAACCGGTCTGCATACCGATCCTA</td>
<td>Hybridize with UTRs to obtain artificial 3’UTR</td>
</tr>
<tr>
<td><strong>RlucF</strong></td>
<td>GTCGACATGACTTCGAAAGTTTATGATCC</td>
<td>Amplify Renilla luciferase ORF</td>
</tr>
<tr>
<td><strong>RlucR</strong></td>
<td>TCTAGATTATTGTTCATTTTTGAGAACTCGC</td>
<td>Amplify Renilla luciferase ORF</td>
</tr>
<tr>
<td><strong>FlucF</strong></td>
<td>CTCACTCCGAGGAAGACGCCAAAAACATAAAGA</td>
<td>Amplify Firefly luciferase ORF</td>
</tr>
<tr>
<td><strong>FlucR1</strong></td>
<td>GAATTCTTACACG</td>
<td>Amplify Firefly luciferase (no stop codon)</td>
</tr>
<tr>
<td><strong>FlucR2</strong></td>
<td>GATTTGGAGCACGGAAAGAC</td>
<td>Amplify Firefly luciferase (no stop codon)</td>
</tr>
<tr>
<td><strong>GUS_clvF</strong></td>
<td>ATGTATCACCGCGTCTTTGA</td>
<td>Real-time PCR for GUS</td>
</tr>
<tr>
<td><strong>GUS_clvR</strong></td>
<td>CAGGAATTCTCATTTGCTGCCTCCCTGCTG</td>
<td>Real-time PCR for GUS</td>
</tr>
<tr>
<td><strong>Rluc_intF</strong></td>
<td>TGTTGGACGACGAACTTCAC</td>
<td>Real-time PCR for Rluc</td>
</tr>
<tr>
<td><strong>Rluc_intR</strong></td>
<td>CATTTTTGTCGGCCATGATT</td>
<td>Real-time PCR for Rluc</td>
</tr>
<tr>
<td><strong>Fluc_clvF</strong></td>
<td>GATTTGGAGCACGGAAAGAC</td>
<td>Real-time PCR for Fluc</td>
</tr>
<tr>
<td><strong>Fluc_clvR</strong></td>
<td>Same as GUS_clvR</td>
<td>Real-time PCR for Fluc</td>
</tr>
<tr>
<td><strong>MIR398_F</strong></td>
<td>CTGAGGCTCTCAGACAGATTTTGAGGA</td>
<td>miRNA over-expressor</td>
</tr>
<tr>
<td><strong>MIR398_R</strong></td>
<td>GAATTCAACCTGCCTGGTATGAAACATGG</td>
<td>miRNA over-expressor</td>
</tr>
<tr>
<td><strong>MIR156_F</strong></td>
<td>GCAACTGAGGAGGAGAGTGGTAAGGAAAGT</td>
<td>miRNA over-expressor</td>
</tr>
<tr>
<td><strong>MIR156_R</strong></td>
<td>ACCGAATTCGCAAAGACCAACAGACG</td>
<td>miRNA over-expressor</td>
</tr>
<tr>
<td><strong>MIR160_F</strong></td>
<td>GCAACTGAGGAGGACACGCTGGAGAGACATT</td>
<td>miRNA over-expressor</td>
</tr>
<tr>
<td><strong>MIR160_R</strong></td>
<td>ACCGAATTCGCAAAGACCAACAGACG</td>
<td>miRNA over-expressor</td>
</tr>
<tr>
<td><strong>MIR165_F</strong></td>
<td>GCAACTGAGGAGGACACGCTGGAGAGACATT</td>
<td>miRNA over-expressor</td>
</tr>
<tr>
<td><strong>MIR165_R</strong></td>
<td>ACCGAATTCGCAAAGACCAACAGACG</td>
<td>miRNA over-expressor</td>
</tr>
</tbody>
</table>
References


Iwasaki, S., Kawamata, T., and Tomari, Y. (2009). Drosophila argonaute1 and


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


V transcription guides ARGONAUTE4 to chromatin. Nat Genet 41: 630–634.


Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R., and


VITA
Qikun Liu

Education
Ph.D. in Plant Biology
The Pennsylvania State University, University Park, PA. August 2014
Intercollege Graduate Program in Plant Biology
B.S. in Biological Science
Fudan University, Shanghai, China. May 2008
School of Life Sciences

Honors and Awards
• J. Ben and Hellen D. Hill Memorial Fund, The Pennsylvania State University, 2013
• Conference Travel award, Post-transcriptional Gene Regulation in Plants, American Society of Plant Biologists, 2013
• Robert W. Graham Endowed Graduate Fellowship, The Pennsylvania State University, 2008
• Chun-Tsung Endowment, Fudan University, 2007
• Top Prize, Challenge Cup Competition of Science Achievement in China, 2007
• Silver Medal, China National Biology Olympiad (Top 1 in Shanxi Province) 2002

Publications


