Sequencing and characterization of non-coding small RNAs controlling development in *Arabidopsis thaliana* roots

by

Natalie Wynn Breakfield

University Program in Genetics and Genomics
Duke University

July 26, 2011

Approved:

____
Philip Benfey, Supervisor

____
Uwe Ohler

____
David Sherwood

____
Amy Schmid

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy in the University Program in
Genetics and Genomics in the Graduate School
of Duke University

2011

<u>ABSTRACT</u>

Sequencing and characterization of non-coding small RNAs controlling development in *Arabidopsis thaliana* roots

by

Natalie Wynn Breakfield

University Program in Genetics and Genomics
Duke University

Date: July 26, 2011

Approved:

____
Philip Benfey, Supervisor

____
Uwe Ohler

____
David Sherwood

____
Amy Schmid

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the University Program in
Genetics and Genomics in the Graduate School
of Duke University

2011

Abstract

Small noncoding RNAs (ncRNAs) are key regulators of plant development through modulation of the processing, stability and translation of larger RNAs. I generated cell type specific small RNA datasets comprising over 200 million aligned Illumina sequence reads covering all major cell types of the root as well as four distinct developmental zones. These data were analyzed for three major types of small RNAs, namely microRNAs (miRNAs), repeat associated small interfering RNAs (ra-siRNAs), and trans-acting siRNAs (ta-siRNAs). 133 of the 243 known miRNAs were found to be expressed in the root, and most showed tissue- or zone-specific expression patterns. My collaborators and I identified 70 new high-confidence miRNAs, and knockdown of three of the newly identified miRNAs resulted in altered root growth phenotypes. Ra-siRNAs specify methylation by the RNA directed DNA methylation (RdRM) pathway, requiring the generation of additional methylation datasets. Preliminary analysis shows cell-type specific methylation patterns that correlate with small RNA and mRNA expression. Analysis of ta-siRNAs revealed new ta-siRNA generating loci, and a novel triggering miRNA for the trans-acting siRNA-generating TAS1 loci. In summary, our study demonstrates the power of isolating individual cell types and developmental zones in combination with deep sequencing and computational analyses to obtain detailed

profiles of ncRNAs, as well as to significantly extend the compendium of known functional RNAs.

Dedication

To my husband Mark
and my children Ian, Andrew, and Grace:
I could not have done this
without your support.
To my grandmother, Anna Rose Leonard:
you always believed in me. Thank you.

Contents

List of Tables

List of Figures

List of Abbreviations

AGO	ARGONAUTE
Col-0	Columbia-0 ecotype of *A. thaliana* – wild type
COR	pCORTEX:GFP marker line – labels the cortex
dpi	days post imbibition
DCL	DICER LIKE
DDM	DECREASED DNA METHYLATION
DRM	DOMAINS REARRANGED METHYLTRANSFERASE
dsRNA	double stranded RNA
FACS	Fluorescence activated cell sorting
GFP	Green fluorescent protein
hc-siRNA	heterochromatin associated small interfering RNA (aka ra-siRNA)
HEN1	HUA ENHANCER 1
LS1	Longitudinal section 1 – corresponds to early meristematic zone
LS2	Longitudinal section 2 – corresponds to late meristematic zone
LS3	Longitudinal section 3 – corresponds to elongation zone
LS4	Longitudinal section 4 – corresponds to maturation zone
miRNA	microRNA

mRNA	messenger RNA
nat-siRNA	natural cis-antisense small interfering RNA
ncRNA	non-coding RNA
nt	nucleotide
PET	PET111 marker line – labels the columella
Pol II	RNA polymerase II
Pol IV	RNA polymerase IV
Pol V	RNA polymerase V
PPR	pentatricopeptide repeat containing protein
QC	quiescent center
RLM-RACE	RNA ligase mediated-rapid amplification of cDNA ends
ra-siRNA	repeat associated small interfering RNA (aka hc-siRNA)
RdDM	RNA dependent DNA methylation
RDR	RNA dependent RNA polymerase
RISC	RNA induced silencing complex
SCR	pSCARECROW:GFP marker line – labels the endodermis
SGS	SUPPRESSOR OF GENE SILENCING
siRNA	small interfering RNA
TAIR	The Arabidopsis Information Resource
TAS	trans-acting small interfering RNA gene

ta-siRNA	trans-acting small interfering RNA
TE	transposable element
TPM	transcripts per million
UTR	untranslated region
WER	pWEREWOLF:GFP marker line – labels the epidermis
WOL	pWOODEN LEG:GFP marker line – labels the stele
WR	whole root
WT	wild type (in this study, Columbia-0 ecotype)

Acknowledgements

Thank you to my advisor, Philip Benfey, who gave me the opportunity to work on this fantastic project. Working in his lab has allowed me the chance to meet and interact with many wonderful people, and I appreciate the amazing opportunity my time at Duke has been. Thank you also to Marilyn Yoder and Christopher Taylor: your confidence in me made me able to believe in myself, and without your support over the years I would have given up on science a long time ago.

Thanks to my committee Uwe, Dave, Zhen-ming, Amy, and Jun for all your help and mentorship through my years at Duke. Finally, thanks to all the Benfey lab members for all your helpful comments on papers, talks, experiments, and life in general. I am so fortunate to have been part of the lab family. I will miss and always remember each of you!

I would like to thank Jalean Petricka for her generous help with experiments, Jaimie Van Norman and Siobhan Brady for help with figure illustrations. Thanks to Heather Belcher and Beth Harvat for assistance with cell sorting, and to Ramiro Rodriguez for help with target validation protocols. Thanks to Kevin Shianna of the Duke Genomic Analysis Facility for sequencing the first Illumina test run, and Lisa Bukovnik in the Institute for Genome Science and Policy Core Sequencing Facility for sequencing the subsequent runs.

Chapter 1. Introduction

Overview of the thesis

The work presented in this thesis involves the identification and analysis of various classes of small RNAs in different cell types and developmental zones of *Arabidopsis thaliana* roots. It is composed of the following parts: introduction to the system and data sets (Chapter 1), the analysis of microRNAs (miRNAs) (Chapter 2), the analysis of trans-acting small interfering RNAs (ta-siRNAs) and repeat associated small interfering RNAs (ra-siRNAs) (Chapter 3), and the implications and future directions of this work (Chapter 4).

Chapter 2 has formed the basis of a manuscript with David Corcoran in Uwe Ohler's lab at Duke University. The manuscript has been reviewed, revised, and resubmitted to Genome Research entitled "High-resolution experimental and computational profiling of tissue-specific known and novel miRNAs in Arabidopsis". (Note: It was accepted for publication on August 5, 2011.) The ta-siRNA portion of Chapter 3 is being written as a manuscript with Jiandong Ding, Fudan University, Shanghai, China. The ra-siRNA portion of Chapter 3 will be the basis of a manuscript with Ryan Lister in Joseph Ecker's lab at the Salk Institute, La Jolla, CA. I also had the opportunity to co-write a review published at the Plant Cell entitled "Intercellular Communication during Plant Development" (Van Norman et al., 2011).

Introduction to the system

Plants play an irreplaceable role in our ecosystem as they provide humans and animals with food and oxygen. Roots form the interface between the plant and its soil environment, providing anchorage for the plant, an interface between the plant and associated microbes, and a place to uptake needed nutrients (Khan et al., 2011). The plasticity of plant roots allows them to change their developmental programs and architecture to explore differing soil and water conditions (Rubio *et al.*, 2004). Root developmental plasticity possibly could be exploited to improve the nutrient content of the plant, and also be harnessed to address some of the environmental problems caused by humans. For instance, plant roots with associated root-colonizing microbes have been used for soil remediation of heavy metal and organic pollutants, and the cortex of the root has been identified as important in controlling organic contaminant degradation (Mench et al., 2009). Additionally, native grassland perennial roots are currently being studied as a possible carbon sink that can help contain global warming (Tilman et al., 2006; To et al., 2010).

Arabidopsis thaliana is the most widely used model angiosperm plant. It has many advantages as a model plant, including a relatively small genome (125 Mb) containing about 26,000 genes (Initiative, 2000), a small plant itself requiring modest growth space, and a relatively short generation time of about 8 weeks. The genome has been fully sequenced and is well annotated (Initiative, 2000), and a knock out collection of insertional mutants for most genes exists (Sessions et al., 2002; Alonso et al., 2003). It self-fertilizes but can also be easily crossed. Many genetic tools and references are available to the *Arabidopsis* research community, especially through the The Arabidopsis Information Resource (TAIR) website (Swarbreck et al., 2008).

The *A. thaliana* root offers an exceptional opportunity for the study of development at cellular resolution because of its radially symmetric structure (Figure 6). The different cell types of the root are arranged in concentric cylinders around a central core of vasculature. In the stem cell niche located close to the tip of the root, stem cells divide to generate specific cell types that are constrained to longitudinal cell files. Additionally, the developmental age of any given cell can be ascertained from its distance from the stem cell niche along the longitudinal axis, with the youngest cells at the root tip and the oldest cells furthest from the tip. This effectively reduces the four dimensions of development (three spatial dimensions plus time) to two dimensions (the radial axis corresponds to cell type and the longitudinal axis to age), making the root a useful and powerful system in which to identify and study molecules and regulatory networks that control development (Benfey and Scheres, 2000).

Earlier papers from the lab have pioneered the technique that uses promoter:GFP marker lines to mark cell types in the root (Birnbaum et al., 2003). The roots are digested to remove the cell walls and separate the cells, and then the GFP positive cells are collected by fluorescence activated cell sorting (FACS), allowing each cell type to be isolated for use in whole genome microarrays. A follow up study highlighted the importance of analyses using cell type specific populations by describing many patterns of gene expression that were masked in whole root microarrays but revealed when cell type specific gene expression was analyzed (Brady et al., 2007). Additionally, the temporal expression patterns of these genes were also examined allowing developmental programs to be analyzed in space and time. The genes involved in these patterns give further insight into the functions of these individual cell types and the developmental program.

It has been calculated that 90% of the genome is transcribed in humans, and the same is believed to be true for plants (Matzke et al., 2009). In *A. thaliana*, only about 45% of the genome encodes proteins (Initiative, 2000), leaving a large proportion of non-coding RNA being transcribed. In the transcriptional networks described above, only a portion of the protein coding genes were represented on the microarrays used for the experiments, which was the state of the art at the time of the study. New high-throughput sequencing technologies are allowing the exploration of the non-coding RNA transcriptome. In this study, I use one of these new technologies, Illumina sequencing, to explore the small non-coding RNAs found in the *A. thaliana* root, with the ultimate goal of integration of these data into the transcriptional networks controlling root development.

Small RNAs

The phenomenon of small RNA-mediated post transcriptional gene silencing was discovered in petunia (Napoli et al., 1990), followed by the discovery of small regulatory non-coding RNAs (ncRNAs) in *Caenorhabditis elegans* (Lee et al., 1993). Despite their relatively recent discovery, ncRNAs have been implicated as key regulators in both development and disease for most known eukaryotes (Farazi et al., 2011). Plants have a diverse array of small RNAs, including transposon-derived small interfering RNAs (siRNAs), trans-acting small interfering RNAs (ta-siRNAs), natural antisense RNAs (nat-siRNAs), heterochromatin or repeat associated siRNAs (hc-siRNAs or ra-siRNAs), and microRNAs (miRNAs) (Chen, 2010) (Figure 1). In plants, small ncRNAs can both transcriptionally and post-transcriptionally repress expression of their targets, and these targets are recognized by extensive complementarity to the small RNA (Cuperus et al., 2010)

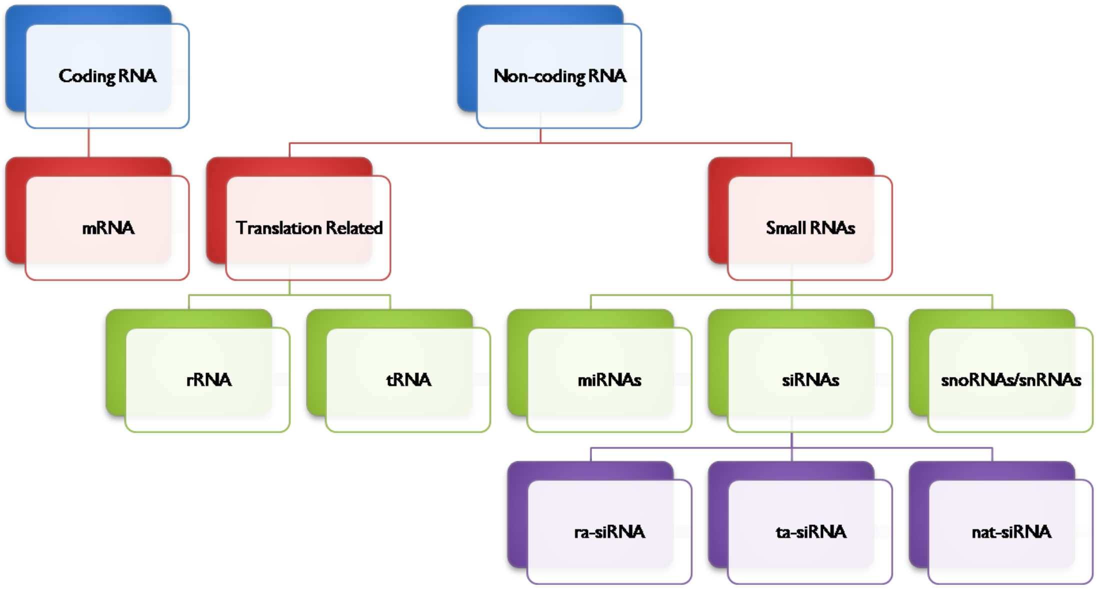

Figure 1. Schematic showing relationship of small RNAs to other RNAs.

These small RNAs share much of the same processing machinery, and mostly differ in the way they produce their double stranded RNA intermediate (Figure 2). There are 4 DICER-LIKE (DCL) and 10 ARGONAUTE (AGO) proteins in *A. thaliana*, which can be used to produce small RNAs with different first nucleotide specificity and varying functionality. The four DCL enzymes produce slightly different sized small RNAs. DCL1 and DCL4 make 21 nt small RNAs, while DCL2 makes 22 nt and DCL3 makes 24 nt products (Czech and Hannon, 2011). Adding an additional layer of complexity, AGO proteins have tissue and developmental stage-specific expression patterns (Mallory and Vaucheret, 2010). Looking at the whole root, AGO1 and AGO4 are relatively highly expressed, AGO2, AGO7 and AGO10 show intermediate expression, while AGO3, AGO5, AGO6, AGO8 and AGO9 show very little expression.

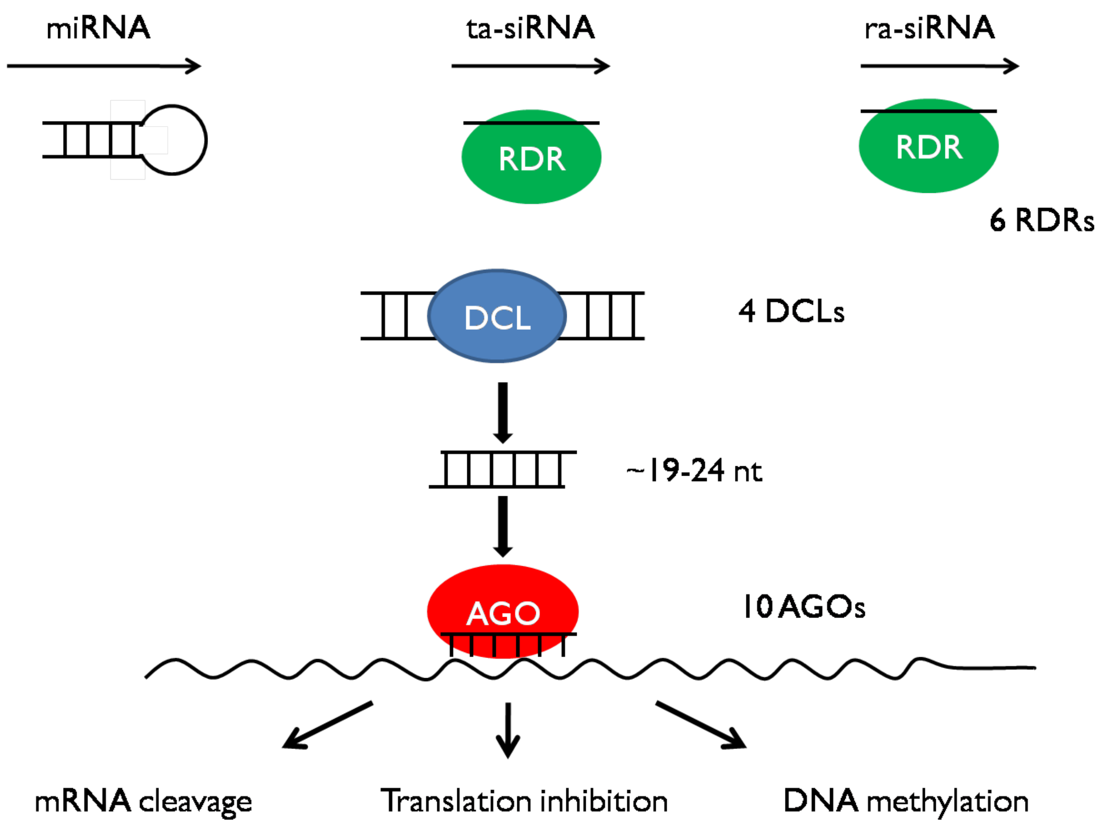

Figure 2. Schematic of small RNA biogenesis.

The small RNAs associated with a specific AGO protein often shows a bias for specific first nucleotide (Montgomery et al., 2008; Mallory and Vaucheret, 2010; Czech and Hannon, 2011). AGO1 is often associated with small RNAs with a 5′ uridine, while AGO2 and AGO4-associated small RNAs often begin with adenine. AGO5 is often found with a 5′ cytosine. There are some exceptions, however. MiR390, the trigger miRNA for TAS3 ta-siRNA, is always found with AGO7, even though it has a 5′ adenine (Montgomery et al., 2008). The size and first nucleotide of a small RNA can allow you to predict the processing DCL and the AGO, leading to a prediction of function. For example, canonical miRNA precursors are processed by DCL1 which results in a 21 nt species with a 5′ uracil that is typically incorporated with AGO1 (Mi et al., 2008). In addition, ra-siRNA often are 24 nt long with a 5′ adenine, characteristic of AGO4 association (Mi et al., 2008)

Several groups have undertaken studies to identify miRNAs, along with other small ncRNAs, using sequencing technologies in *A. thaliana* shoots, flowers, and mature pollen (Llave et al., 2002; Gustafson et al., 2005; Fahlgren et al., 2007; Grant-Downton et al., 2009). These studies identified miRNAs that control leaf, floral, shoot and vascular development. Additionally, miRNAs have been shown to modulate miRNA and siRNA biogenesis by targeting genes within the small RNA processing pathways. Plant miRNAs have been implicated in hormone signal transduction pathways and in responses to plant pathogens and environmental stresses. Furthermore, miRNA expression patterns were precisely defined in both space and time during the development of floral and seed tissues (Valoczi et al., 2006). Taken together, these studies led us to hypothesize that miRNAs and siRNAs exist that are specific not only to the root as a whole, but also to precise spatial and temporal locations during root development.

Small interfering RNAs (siRNAs)

Trans-acting siRNAs (ta-siRNAs)

Ta-siRNAs are a specialized class of small RNAs that combine features of miRNAs and siRNAs, and arise from a trans acting siRNA generating gene (TAS) transcript (Czech and Hannon, 2011) (Figure 3). A miRNA, often 22 nt, performs the initial cleavage event on the TAS gene transcript. After cleavage, the long primary transcript is converted to dsRNA and processed to 21 nt mature ta-siRNAs by the RNA-DEPENDENT RNA POLYMERASE 6/SUPPRESSOR OF GENE SILENCING 3 (RDR6/SGS3) processing complex. The cleavage site of the initial miRNA sets the register for the generation of the ta-siRNAs in 21 nt intervals down the transcript. The resulting mature ta-siRNAs act in a similar way to miRNAs, by targeting messenger RNAs (mRNAs) *in trans* for either slicing or translational repression. Some of these primary ta-siRNAs can in turn generate secondary ta-siRNAs, which then affect the translation or stability of their mRNA targets.

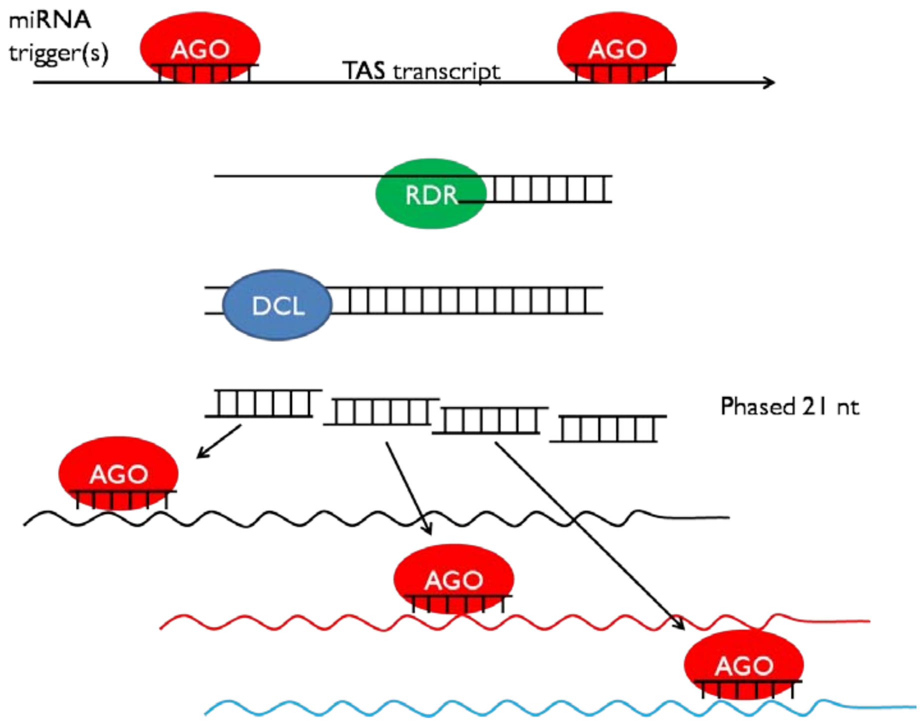

Figure 3. Schematic of ta-siRNA biogenesis.

There are currently four annotated primary ta-siRNA families in *A. thaliana*. The three TAS1 loci (a-c) are targeted/cleaved by miR173 (Vazquez et al., 2004), with resulting ta-siRNAs that target both unknown and pentatricopeptide repeat (PPR) genes (de Felippes et al., 2010). TAS2 is also targeted by miR173 (Yoshikawa et al., 2005), and its ta-siRNAs target a group of PPR genes (Allen et al., 2005). TAS3 is targeted by miR390, and its resulting ta-siRNAs target three auxin response factors which play an important role in lateral root development (Marin et al., 2010). TAS4 is targeted by miR828 (Rajagopalan et al., 2006) and its ta-siRNAs target several MYB transcription factors involved in anthocyanin biosynthesis (Hsieh et al., 2009).

Repeat associated siRNAs (ra-siRNAs)

An interesting class of small RNAs is not associated with mRNA slicing or translational repression, but directs methylation. Plants and fission yeast both have a small RNA dependent system of methylation, which is called RNA directed DNA methylation (RdDM) in plants. In RdDM, cytosines are *de novo* methylated not only at CG locations, but also at CHG and CHH (where H is A, T, or C) locations through RdDM (Matzke et al., 2009). Additionally, RdDM mediates the maintenance of CHH methylation (Teixeira and Colot, 2010). DNA methylation is spread throughout the genome, but is concentrated on centromeric and regions containing many repeats, which is why this class is known as repeat associated (ra-siRNA) or heterochromatin associated small interfering RNAs (hc-siRNA) (Verdel et al., 2009; Teixeira and Colot, 2010) Approximately 30% of the cytosine methylation in *A. thaliana* seems to be directed by siRNAs (Lister et al., 2008). Transposable elements (TEs) are highly repressed, but in a genome wide study of TEs and methylation, 26% were found to not be methylated (Ahmed et al., 2011).

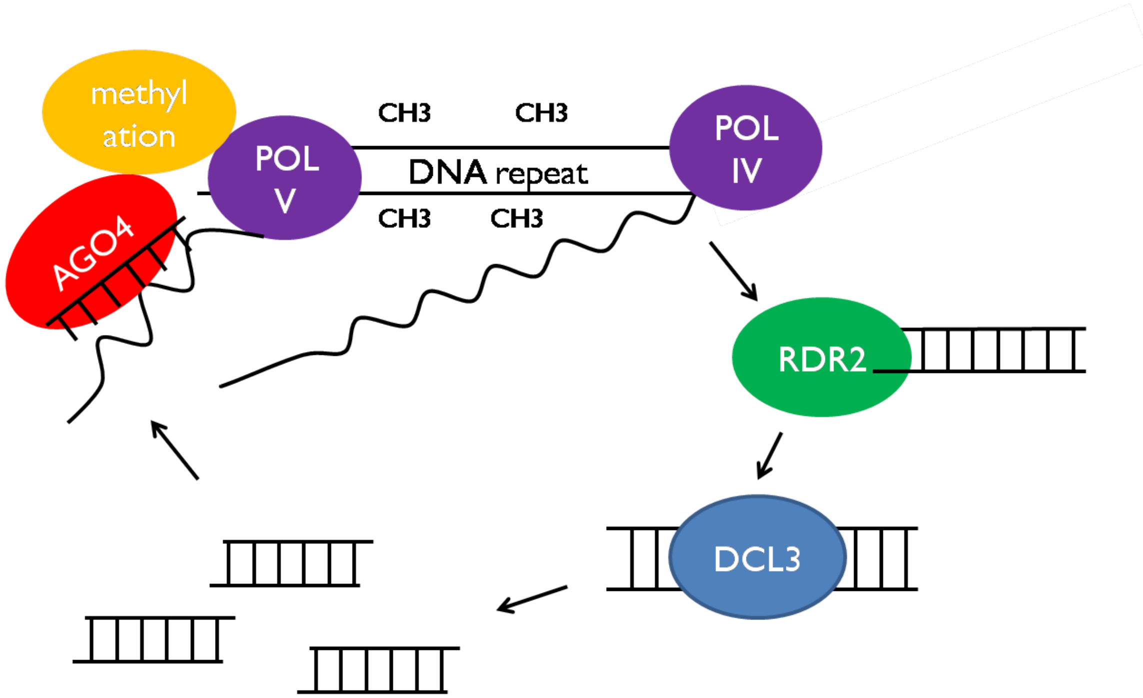

Figure 4. Schematic of ra-siRNA biosynthesis.

The regulation of methylation in plants is very complex. Plants contain two additional polymerases that work in this pathway, RNA polymerase IV and V (Pol IV and Pol V) (Figure 4). Originally, a dsRNA that presumably arises from RNA polymerase II (Pol II) transcription (either overlapping transcription or the transcription of an inverted repeat) is diced by DCL3 and methylated by HUA ENHANCER 1 (HEN1), before loading into ARGONAUTE 4 (AGO4)-Pol V complex. When the complex finds its complementary region, the methyltransferase DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) *de novo* methylates the targeted region. This process is facilitated by Pol V transcription. The siRNA trigger is amplified by the Pol IV transcription of the region, which is converted to dsRNA by RDR2, and is then diced and methylated by DCL3/HEN1 before loading into the AGO4-Pol V complex to continue the cycle (Matzke et al., 2009).

The role of this methylation during development seems to be protective. The RNAi machinery is necessary for the silencing of transposable elements (TEs) in germ cells and the embryo (Mosher and Melnyk, 2010). Lower levels of methyltransferases, e.g. DECREASED DNA METHYLATION 1(DDM1) in the pollen grain, lead to increased transcription from TEs. This leads to siRNA accumulation as described above, resulting in increased methylation and ultimately silencing of TEs. In fact, many of the players in RdDM are very highly expressed in the seed (Teixeira and Colot, 2010). The role of RdDM in the root during development is relatively unexplored, however.

Natural antisense transcript-derived siRNAs (nat-siRNAs)

This class of siRNA is less characterized than many other siRNAs, and indeed, it is not examined in this study. The dsRNA arises from the transcription of two overlapping genes on opposite strands (Chapman and Carrington, 2007). Nat-siRNAs are induced in stress responses (Henz et al., 2007), so the examination of these small RNAs in our non-stress datasets would be less informative.

MicroRNAs (miRNAs)

MiRNA biogenesis begins with a DNA dependent RNA polymerase II-produced primary transcript, which folds into a stem loop structure (Voinnet, 2009) (Figure 5). The stem portion of the primary transcript is cleaved by a protein complex that contains a DCL endonuclease to produce the miRNA precursor.

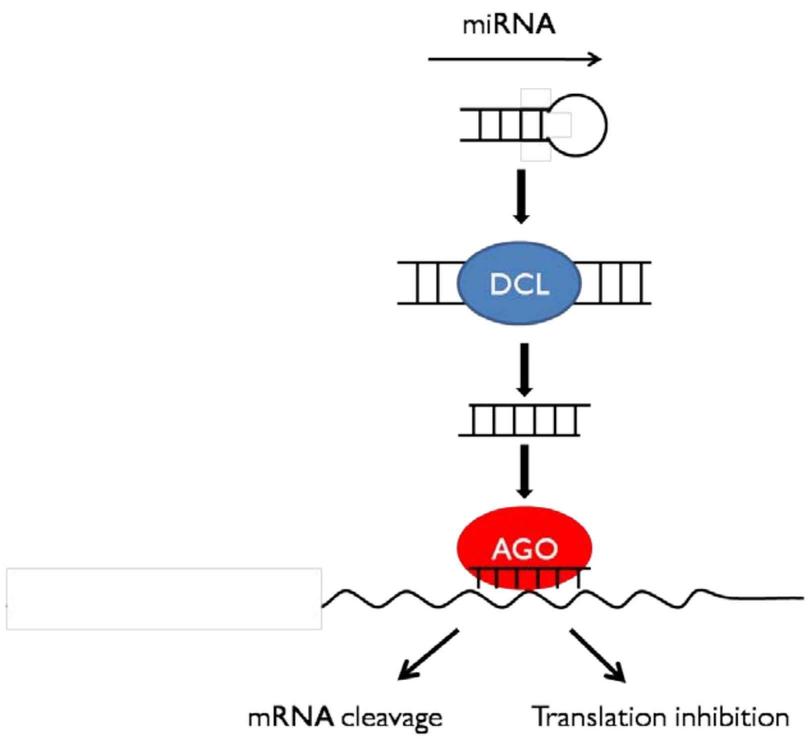

Figure 5. Schematic of miRNA biogenesis.

The precursor is further cleaved to remove the loop portion of the stem loop by the same processing enzymes. The remaining duplex contains the mature miRNA and a complementary miRNA* sequence. The mature miRNA sequence is loaded into the RNA induced silencing complex (RISC), which contains an AGO protein. It is in this complex that the miRNA finds its messenger RNA target by base-pair complementarity and represses target expression by either causing mRNA degradation or by blocking translation (Huntzinger and Izaurralde, 2011).

Introduction to the data sets

We determined the cell type specificity of small ncRNAs in the root by employing cell sorting technology followed by deep sequencing. We used 5 independent lines expressing GFP in the stele, endodermis, cortex, epidermis, and columella, which cover the major cell types of the root tip. Isolation of individual cell types by fluorescence activated cell sorting was then coupled with Illumina small RNA deep sequencing to generate 2 independent libraries for each cell type. We also generated a root developmental time course by hand sectioning 100 Columbia-0 roots into early and late meristematic, elongation, and differentiation zones. In addition, we made 4 libraries consisting of whole root tissue (WR): two unsorted and two mock sorted. An overview of the data sets is given in Figure 6. The figure shows from left to right a schematic drawing of an A. thaliana seedling, the root denoting developmental zones isolated by hand sectioning, and the root tip displaying the cell types analyzed in this study. The colors indicate the regions covered by the GFP marker lines used for isolation of cell types by the cell sorting.

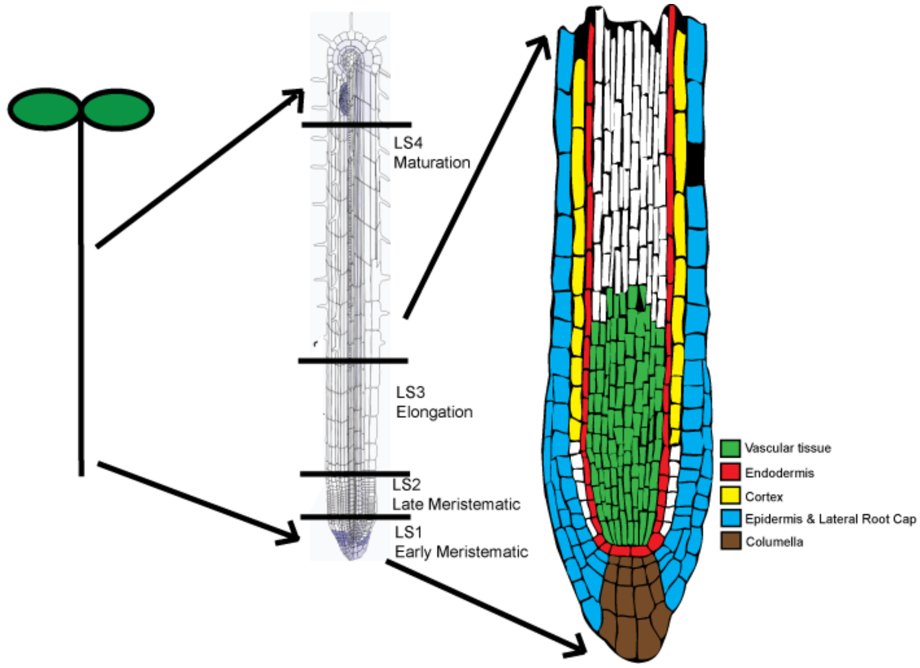

Figure 6. Overview of the root developmental zones and cell types used in this study.

Approximately 200 million short reads were mapped to the *A. thaliana* (TAIR9) genome (Table 1). The reads were allowed to be aligned with up to 25 different locations so to include those miRNAs that are part of large families. The designations 1 & 2 refer to the two biological replicates in the cell types and WR samples. LS1 (Longitudinal Section 1) corresponds to early meristematic zone, LS2 to late meristematic zone, LS3 to elongation zone, and LS4 to maturation zone. The number of raw reads per sample was first normalized to transcripts per million (TPM) to correct for varying number of reads in the different sequencing lanes. This normalization is open to bias from PCR amplification, RNA ligase preference, and reverse transcription reaction, but is the standard procedure at this time (Linsen et al., 2009; Meyer et al., 2010). For samples with a biological replicate, average expression values were used for further analyses since the biological replicates demonstrated high reproducibility of the radial data sets (r= 0.81 to 0.97; Table 2).

Table 1: Mapping statistics of the small RNA deep sequencing data.

Sample	Raw Reads	Reads to Align	Reads that Aligned	Genomic Locations Aligned
STELE1	4021579	3821931	3036345	1476598
STELE2	4710869	4574820	4157643	1234045
ENDODERMIS1	4711441	4621250	4346736	577504
ENDODERMIS2	2629534	2213750	2141417	819856
CORTEX1	6313351	3943684	3679116	1554711
CORTEX2	11477250	11004253	10121942	1240968
EPIDERMIS1	7572641	6661001	6484143	1774942
EPIDERMIS2	9796097	9400938	9066571	2183620
COLUMELLA1	11795139	11172736	10794988	4541389
COLUMELLA2	11477250	16500876	14132689	7299249
WR sort 1	21826690	19532443	14676490	3943574
WR sort 2	20311309	17779283	13397704	4070513
WR1	2161398	1990783	1777967	700129
WR2	3710275	2914143	2555789	1374730
LS1 (meristem)	31147802	27738885	23153324	9377709
LS2 (meristem)	32008086	26526594	23248312	5893945
LS3 (elongation)	30922856	25215859	21161585	6361560
LS4 (maturation)	30898835	24758702	21345448	6666730

Table 2: Sequencing library reproducibility.

Biological replicates compared	**Correlation (r)**
Whole Root	0.921
Whole Root(sorted)	0.969
Stele	0.951

Cortex	0.814
Columella	0.954
Endodermis	0.950
Epidermis	0.886

A measure of the high quality of these data can be seen in the size distributions of the sequence reads (Figure 7). As expected for known small ncRNAs, 77% of the reads fall within the range of 19-24 nts.

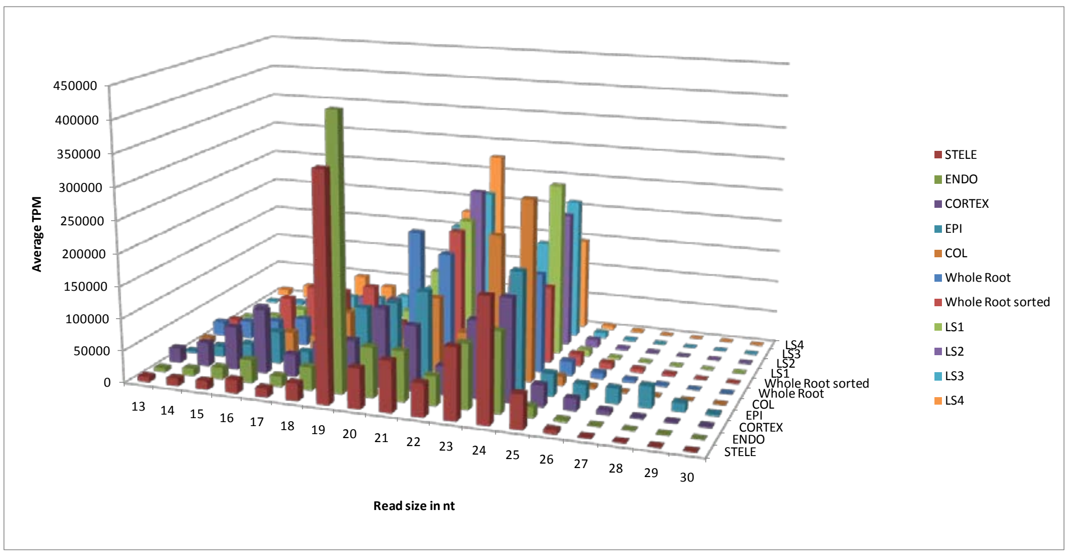

Figure 7: Size distributions of reads in the small RNA sequencing libraries.

Small RNAs of different read lengths show a bias in their most 5′ nucleotide because the 5′ base of small RNAs is a characteristic of the AGO with which it associates (Mi et al., 2008). Of the reads that are 21 nts in length, 62% have a 5′ uracil (Figure 8).

This is characteristic of DCL1 cleavage and AGO1 association found in most known miRNAs (Llave et al., 2002; Park et al., 2002; Reinhart et al., 2002; Rajagopalan et al., 2006; Fahlgren et al., 2007; Fahlgren et al., 2009; Grant-Downton et al., 2009). In addition, 58% of the 24 nt length reads have a 5′ adenine, characteristic of AGO4 association (Mi et al., 2008).

Roots have recently been shown to contain an abundant class of 19 nt small RNAs derived from fragmentation of tRNAs (Hsieh et al., 2009). This size class is abundantly represented in our data (15% of all reads 19-24 nts), and is highly enriched in the stele and endodermis (Figure 8). Taken together, our data represent a rich resource of root small RNA expression profiles.

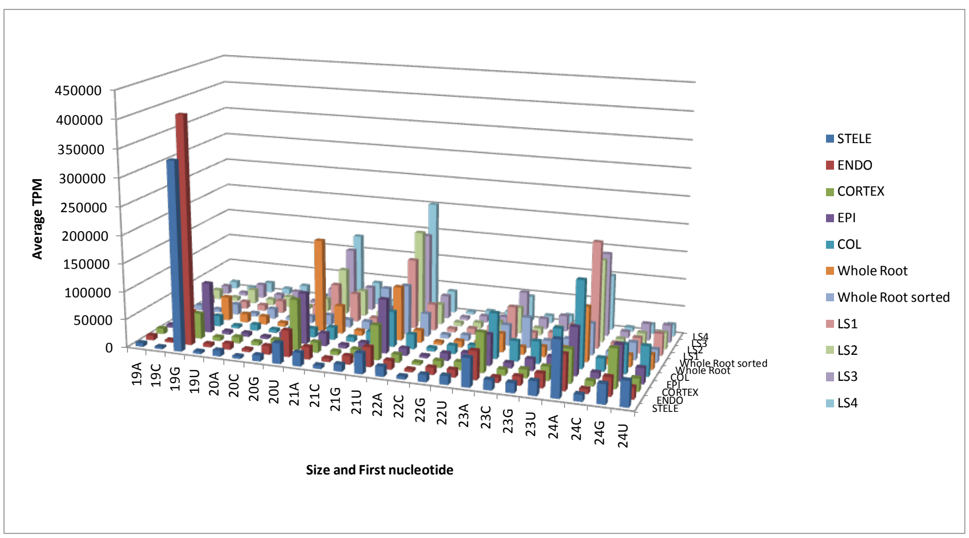

Figure 8: Highly abundant small RNA species are reflected in the read size and the identity of the first nucleotide.

Methods

Cell sorting and Illumina small RNA library preparation

Cell type specific sorting was performed using GFP labeled lines (Birnbaum et al., 2003). The stele was marked by p*WOODEN LEG*::GFP (WOL) (Mahonen et al., 2000), endodermis and quiescent center by p*SCARECROW*::GFP (SCR) (Birnbaum et al., 2003), the cortex by p*CORTEX*::GFP (COR) (Lee et al., 2006), the epidermis and lateral root cap by p*WEREWOLF*::GFP (WER) (Lee and Schiefelbein, 1999; Sena et al., 2004), and columella by enhancer trap PET111 (PET) (Nawy et al., 2005).

At least 1 million GFP positive cells (or mock sorted cells in the case of whole root sorted samples) were collected directly into miRVana (Ambion) lysis buffer and stored at -80 degrees until extraction. The total RNA extraction protocol was used. For the longitudinal sections, 100 six-day-old Columbia-0 wild-type roots were hand dissected into 4 pieces: two meristematic zone sections, one elongation zone section, and one approximately 2 mm maturation zone section. The sections were placed in the miRVana kit lysis buffer immediately after dissection, and then the 100 roots for each section were combined before total RNA extraction.

High molecular weight RNA was precipitated using 5% PEG8000 and 0.5M NaCl, and low molecular weight (LMW) RNA was ethanol precipitated from the supernatant as described in Lu et al. (Lu et al., 2007). LMW RNA used in small RNA library construction had a RIN score of 8 or above by Agilent Bioanalyzer. Illumina small RNA sequencing libraries were prepared as described in the v.1 protocol. To validate each library before using for sequencing, less than 1 nanogram was traditionally cloned and approximately 100 colonies were Sanger sequenced. Libraries that contained less than 5% adapter:adapter sequences (indicating no small RNA insert) with small RNA sequences that could be aligned to the *A. thaliana* genome were used for Illumina sequencing on an Illumina Genome Analyzer II. Two biological replicates of each cell type, two biological replicates of the whole root sorted, two biological replicates of the whole root unsorted, and one of each longitudinal section were constructed and sequenced.

Small RNA Data Set Processing and Alignment

All small RNA libraries were stripped of their respective 3′ adapter sequences using the FASTX toolkit (Blankenberg et al., 2010). Reads that were less than 13nt in length or contained an ambiguous nucleotide were discarded. The remaining reads were aligned to the *A. thaliana* genome (TAIR9) allowing up to 3 mismatches using the Bowtie (Langmead et al., 2009) algorithm. We allowed each sequence to map to multiple locations with a maximum of 25 locations per sequence because many families of small ncRNAs share precisely the same sequence. Reads that failed to map to the genome, or mapped to the genome in more than 25 locations within their optimal mismatch-stratum were discarded. Mapped locations where at least one read mapped with no mismatches were used for further analysis; this removed mapped locations where all of the reads mapping to that particular location contained one or more mismatches. Only intergenic and intronic sequences were considered for further analysis. Of the locations where 10 or more reads mapped and at least one read was a perfect match, less than 18% of the total reads contained a mismatch.

miRNA Expression Profiles

All reads that mapped to a known mature miRNA as identified in miRBase v16 (Griffiths-Jones et al., 2008), or overlapped at least 90% with a known miRNA were counted toward the expression value. The count value was normalized by the size factor as defined by Anders and Huber (Anders and Huber, 2010). As to not be biased by miRNA family size, one representative member of each of the families was used in the calculation of this normalizing factor. The normalizing factors were calculated using the 'DESeq' (Anders and Huber, 2010) library in the R statistical software package (Team, 2009).

Identification of Putative miRNA Mature Sequences

Initially, we constructed a set of putative hairpin sequences based on the aligned short sequence reads. We then used those reads to derive a subset of putative miRNA precursor sequences that are used as input to a probabilistic binary classifier. In the remainder of this section, we describe all the analysis steps in more detail.

Aligned reads that overlapped were grouped together, and the most abundant read within that group was considered the putative mature miRNA; only putative mature miRNAs with more than 10 reads were considered for further analysis. For the purposes of training PIPmiR , exact matches to miRBase v16 (Griffiths-Jones et al., 2008) mature sequences were used as the positive set, regardless of whether that was the most abundant form of reads at that location. In addition, no non-annotated mature miRNA sequences that overlapped a known precursor were used in training the model.

miRNA Expression Analysis

Heat maps were produced using the microarray expression viewer (MeV) which is part of the TM4 microarray software suite (Saeed et al., 2006). Z-scores for miRNA expression and z-scores for their average target(s) expression value were calculated before entering into MeV for heat map production. Z-score was defined as [($\log_2$ normalized expression)-(mean of $\log_2$ normalized expression of group)]/standard deviation. After loading data, the gene expression profiles were hierarchically clustered using Euclidean distance. A figure of merit was generated with 100 iterations to obtain the approximate number of clusters appropriate for k-means clustering. K-means clustering was performed using Euclidean distance and up to 50 iterations.

For known miRNA target expression analysis, the list of validated targets was downloaded from the ASRP website (Backman et al., 2008), and the average expression value was calculated per family by averaging the expression values of the targets of that family from lab microarray data (Brady et al., 2007). Pearson product moment correlation coefficients (r) and p-values were calculated using R, and the miRNA:mRNAs were ordered from lowest to highest p-values.

miRNA validation, target prediction, and phenotyping

Verification of the expression of novel miRNAs was performed using whole root Col-0 total RNA and stem loop RT-PCR (Varkonyi-Gasic et al., 2007) using miR156 as the normalizing control (primers on Table 3). Some miRNAs were further validated (designated "cloning" on Table 2) using stem loop endpoint PCR and these PCR products were cloned into pCR-Blunt-TOPO vector (Invitrogen) and their sequences confirmed by Sanger sequencing (Varkonyi-Gasic et al., 2007).

Targets for both known and novel mature miRNAs were predicted using the WMD3 web microRNA designer using default settings (Ossowski et al., 2008). 5′ RNA ligase mediated -rapid amplification of cDNA ends (RLM-RACE) was performed using Gene Racer kit (Invitrogen).

Table 3. Selected primers used in this study.

Gene and Assay	Primer Name	Primer Sequence (5' to 3')
novel33140 - genotyping	CS824777_LP2	CGATCCTTTCCGAAAATCAA
novel33140 - genotyping	CS824777_RP2	TTTGCCGTCGTTTGAGACTA
novel406654 - genotyping	SAIL_739_F11_LP	TTGTGAGATTTGATGGGAAGC
novel406654 - genotyping	SAIL_739_F11_RP	ACCTTTTGTCAAAATGCAACG
T-DNA - genotyping	SAIL_LB3	TAGCATCTGAATTTCATAACCAATCTCG ATACAC
control gene - qPCR	UBQ10F	GGCCTTGTATAATCCTGATGAATAAG
control gene - qPCR	UBQ10R	AAAGAGATAACAGACGGAAACATAGT
novel21861 target - qPCR	At1g61140_qPCR_F2	GGCAGCTTGTCATGTTCTGA
novel21861 target - qPCR	At1g61140_qPCR_R2	GAGGTGAGACTGTCGGCTTC
novel21861 target - qPCR	At2g33400_qPCR_F	AGTGGAAAGCGGTTAAAGCA
novel21861 target - qPCR	At2g33400_qPCR_R	TGGGGATCTCTTCGATGAAC
novel33140 - stem loop qPCR	novel33140_RT	GTTGGCTCTGGTGCAGGGTCCGAGGTAT TCGCACCAGAGCCAACTGCCCC
novel33140 - stem loop qPCR	novel33140_forward	GCGGCGGTTGGTAGTGGATAAG
novel21861 - stem loop qPCR	novel21861_RT	GTTGGCTCTGGTGCAGGGTCCGAGGTAT TCGCACCAGAGCCAACATGCCG
novel21861 - stem loop qPCR	novel21861_forward	GCGGCGGATCTGAAGAAAATAG
novel406654 - stem loop qPCR	novel406654_RT	GTTGGCTCTGGTGCAGGGTCCGAGGTAT TCGCACCAGAGCCAACCACCAA
novel406654 - stem loop qPCR	novel406654_forward	GCGGCGGTGGACAAGGTTAGAT
control miRNA - stem loop qPCR	mir156_RT	GTCGTATCCAGTGCAGGGTCCGAGGTAT TCGCACTGGATACGACGTGCTC
control miRNA - stem loop qPCR	mir156_forward	GCGGCGGTGACAGAAGAGAGT
universal reverse - stem loop qPCR	universal_reverse	GTGCAGGGTCCGAGGT

Homozygous T-DNA insertion mutants (usually producing knockout mutants) were identified using primers in Table 3 for the novel miRNA candidates (McElver et al., 2001; Sessions et al., 2002; Alonso et al., 2003). The insertion line for novel 33140 was CS824777 – SAIL_582_B05, and for novel406654 was CS833058 - SAIL_739_F11. Target mimicry lines were produced as described (Franco-Zorrilla et al., 2007). Root lengths were assayed by growing mutant seeds vs. control seeds vertically on 1X Murashige and Skoog salt mixture, 1% sucrose, 2.3 mM 2-(N-morpholino)ethanesulfonic acid pH 5.8 and 1% agar plates. Root lengths were measured using ImageJ (Abramoff et al., 2004), and the Student t-test was used to determine statistical significance.

Microarray experiment and analysis

20 root tips (containing the meristematic and approximately half of the elongation zone) were hand dissected into miRVana lysis buffer (Ambion) for each of three biological replicates of 6 dpi novel33140 mutant and Col-0 WT roots. Roots were ground in liquid nitrogen before miRVana total RNA extraction procedure. 3'IVT Express kit (Affymetrix) was used to prepare samples through the fragmentation step for the ATH1 Affymetrix GeneChip. Hybridization was performed by the Duke Microarray Core Facility. Results were analyzed using a mixed model ANOVA, as described (Levesque et al., 2006). Correlations between biological replicates were greater than 0.97 R^2.

Ra-siRNA additional datasets

For DNA methylation analysis, an additional 1,000,000 cells for each of the five cell types and two independent transformants were sorted into phosphate buffered saline, and spun down for 6 minutes at 200 rpm. Most of the supernatant was removed (no pellet was visible) and AP1 lysis buffer (DNeasy kit, Qiagen) added before storage at -80 degrees Celsius. For RNA-seq mRNA transcript analysis, the sorting and centrifugation was as above, but RLT buffer (RNeasy kit, Qiagen) plus 10% volume:volume beta-mercaptoethanol was added before storage. The frozen sorted cells in their respective buffers were shipped to Ryan Lister in Joseph Ecker's lab at the Salk Institute (LaJolla, CA) where he prepared the DNA or RNA, the Illumina sequencing libraries, and performed the sequencing.

In addition, 100 wild type Col-0 roots were hand dissected into four developmental zones: columella, meristem, elongation, and early maturation zone. The sections were immediately placed in RNAlater solution (Ambion), and all four sections were collected for each root. After all 100 roots were collected, the samples were spun down and most of the RNAlater solution removed. The sections were then ground in liquid nitrogen and AP1 lysis buffer was added before storage at -80 degrees Celsius. These samples were also sent to Ryan Lister where they await library preparation and sequencing.

Chapter 2. miRNA analysis

As described in Chapter 1, miRNAs are a class of small RNAs that arise from endogenous transcripts that form hairpins which are then cleaved to form the mature small RNAs. These mature small RNAs then act *in trans* in conjunction with AGO protein-containing complexes to regulate mRNA targets through slicing or translational repression.

The variable size and structure of plant miRNA precursors has made the identification of new miRNA genes a challenge both computationally and experimentally. To date, two types of approaches have been used for the high-throughput identification of new plant miRNA genes. The first approach is purely computational and involves the systematic folding of all intergenic regions into miRNA-like hairpin structures while including additional features such as conservation or simultaneous prediction of stem-loops and their targets (Bonnet et al., 2004; Jones-Rhoades and Bartel, 2004; Adai et al., 2005). The downside to this approach is that there are a vast number of putative miRNA genes with little experimental validation. The second major approach has been to analyze all sequences from small RNA deep-sequencing datasets to see if they conform to a given set of rules. If a sequence meets the necessary rule requirements and the surrounding sequence is able to fold into a stem-loop like structure, then it is automatically classified as a new miRNA (Hendrix et al., 2010). Currently there is only a single tool, miRDeep (Friedlander et al., 2008), that uses a probabilistic model based upon features from both small RNA deep sequencing data and genomic data; however this tool was developed for animal miRNA genes and is not able to identify the types of complex miRNA precursors seen in plants.

Known miRNAs Display Cell Type and Developmental Zone Enrichment

A new miRNA classifier, PIPmiR, was developed in conjunction with this project. Since it is primarily the work of my co-author David Corcoran, only brief details of its construction are included here but the full details can be found in the associated publication (Breakfield et al., 2011). PIPmiR was used to identify and obtain expression values for both known and novel miRNAs, and the analysis in this chapter was done in partnership with David.

The PIPmiR classifier uses the short read data in combination with genomic features to specifically identify miRNAs present in the sequence libraries, and to distinguish miRNAs from other small RNAs. These features were selected based on known biogenesis, processing, and targeting properties of plant miRNAs (Table 4).

Table 4. Feature list for PIPmiR classifier.

Feature	Description
GC% of 1st half of miRNA	Percent of G or C nts in the first half (5′ end) of the mature miRNA sequence

GC% of 2nd half of miRNA	Percent of G or C nts in the second half (3′ end) of the mature miRNA sequence
Nucleotide 10	10th nt from the 5′ end of the mature miRNA
Nucleotide 11	11th nt from the 5′ end of the mature miRNA
First nucleotide	5′ most nt of the mature miRNA sequence
Last nucleotide	3′ most nt of the mature miRNA sequence
Mature length	Number of nts in the mature miRNA sequence
Star length	Number of nts in the miRNA star sequence
Normalized Energy	Minimum free energy of the precursor fold-back structure divided by the length of the precursor sequence
Consecutive mismatches	The maximum number of consecutive non-base pairings between the mature miRNA and the miRNA* sequence
miRNA-miRNA* match	The percent of nts in the mature sequence that have a base-pairing to a nucleotide in the miRNA* sequence
miRNA*-miRNA ratio	The ratio of the number of reads that exactly map to the miRNA* sequence divided by the number of reads that map to the mature miRNA sequence. Any value > 1 for this feature is set to a value of 1
Overhang ratio	The number of reads that overlap 85% or less of the mature miRNA divided by the number of reads that map exactly to the mature sequence. Any value > 1 for this feature is set to a value of 1
miRNA and miRNA* compared to entire precursor	The number of reads that map exactly to the mature miRNA or the miRNA* divided by the number of reads that map to the precursor sequence.
Opposite strand ratio	The number of reads that map to the precursor sequence, but on the opposite strand, divided by the number of reads that map to the precursor on the transcribed strand. Any value > 1 for this feature is set to a value of 1

We profiled the expression of the 243 currently annotated mature miRNAs that originate from 213 precursors as defined by MirBase v.16 (Griffiths-Jones et al., 2008) and found 133 mature miRNAs derived from 127 precursors represented in our datasets (Appendix A). The expression of each miRNA gene in a cell type or longitudinal section was normalized by a scaling factor designed for sequencing based read count data (Anders and Huber, 2010).

We found both broad and very specific expression patterns for known miRNAs (Figures 9-11). The $\log_2$ expression z-scores are visualized in a heat map, with yellow representing higher expression (deviation from the mean) and blue representing lower expression. Overall, clustering of miRNA expression profiles resulted in 9 distinct radial expression patterns (Figure 10), and 10 distinct longitudinal expression patterns (Figure 11). There were clusters of miRNAs showing enriched expression in all cell types and longitudinal zones. We also observed patterns that show fluctuations in miRNA expression across the developmental zones (Figure 10D, H, I).

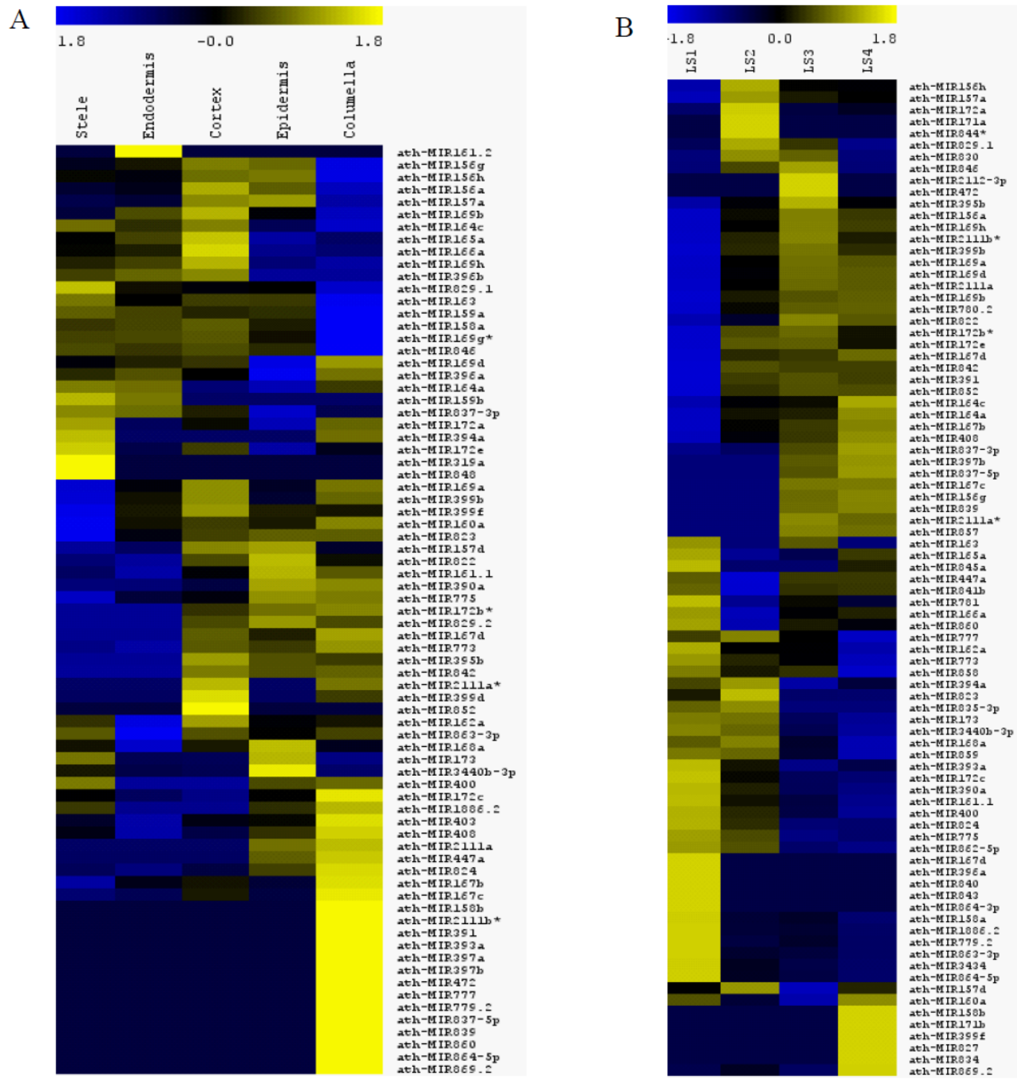

Figure 9: Expression enrichment patterns of known miRNAs in cell types (A) and developmental zones (B).

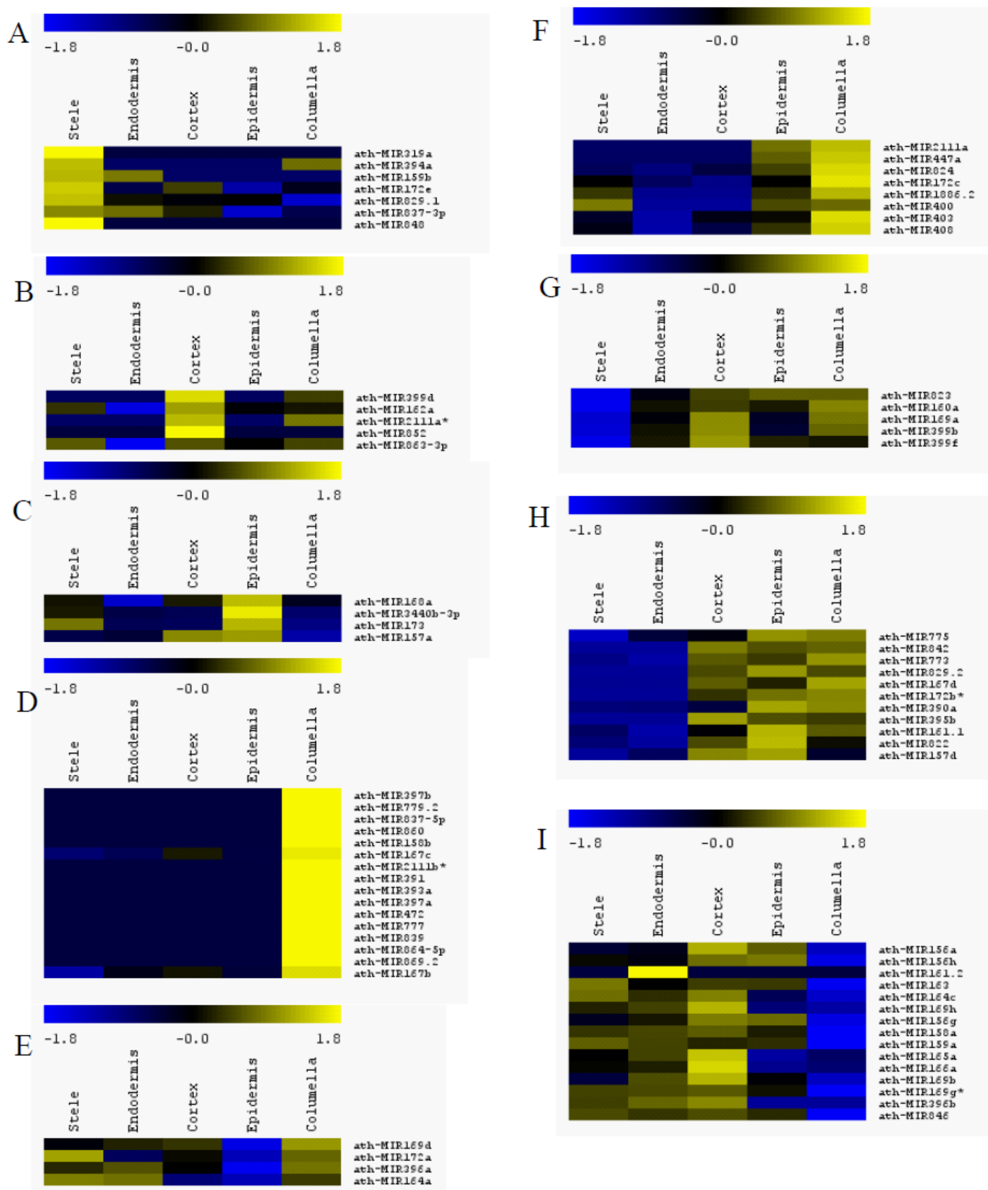

Figure 10: Clusters of miRNA expression in the radial data sets. A-I represent the nine clusters.

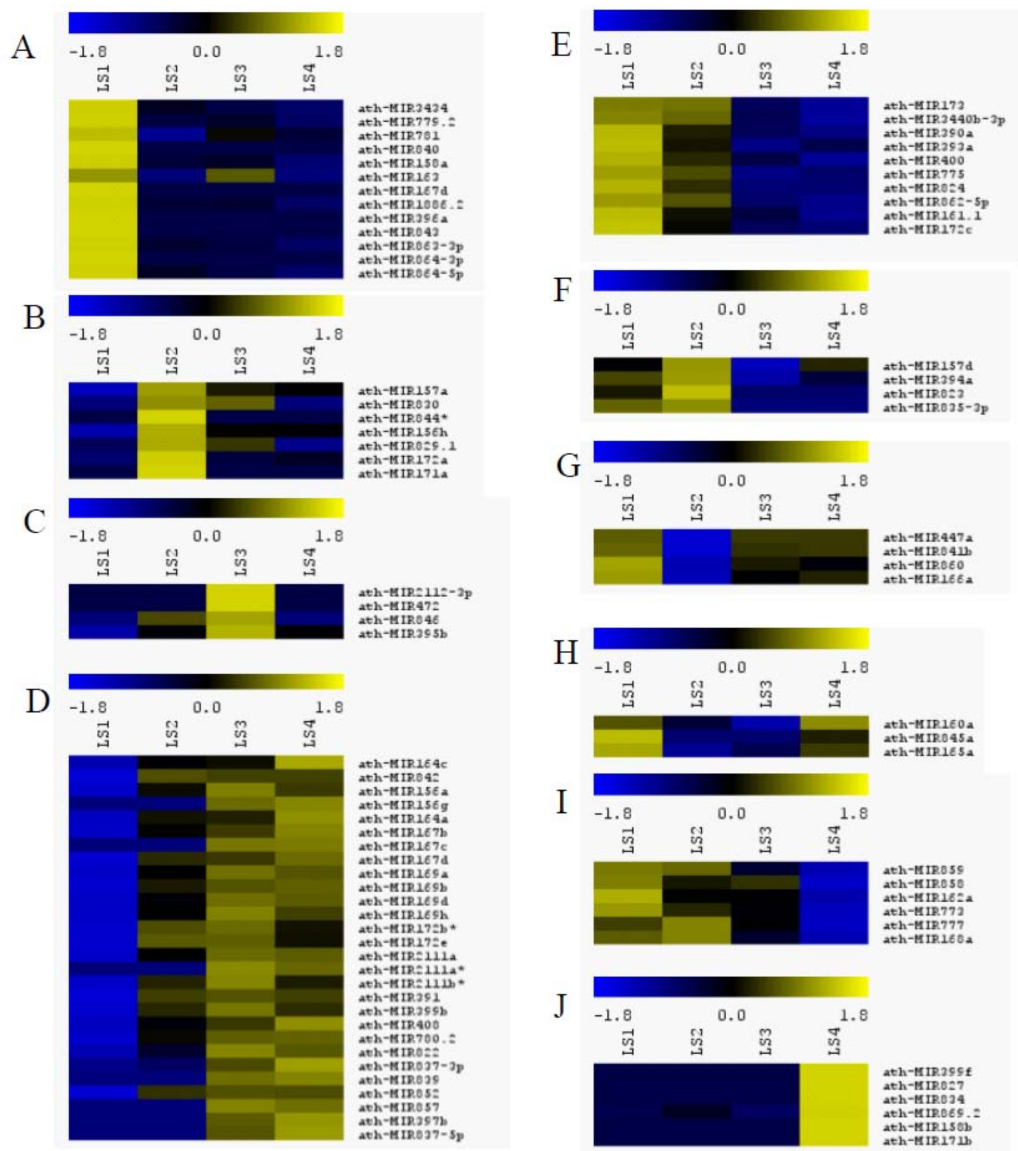

Figure 11: Clusters of known miRNA expression in the developmental zone data sets. A-J represent the ten clusters.

As an additional way to calculate specificity, information content of their expression profiles was calculated using the expression values for the known miRNAs, and the top 50 most specific miRNAs are represented in a similar fashion to sequence logos (Figure 12). The height of the bar represents the information content for that particular miRNA, and the size of each colored area reflects the proportion of expression of the corresponding cell type or developmental zone for the miRNA.

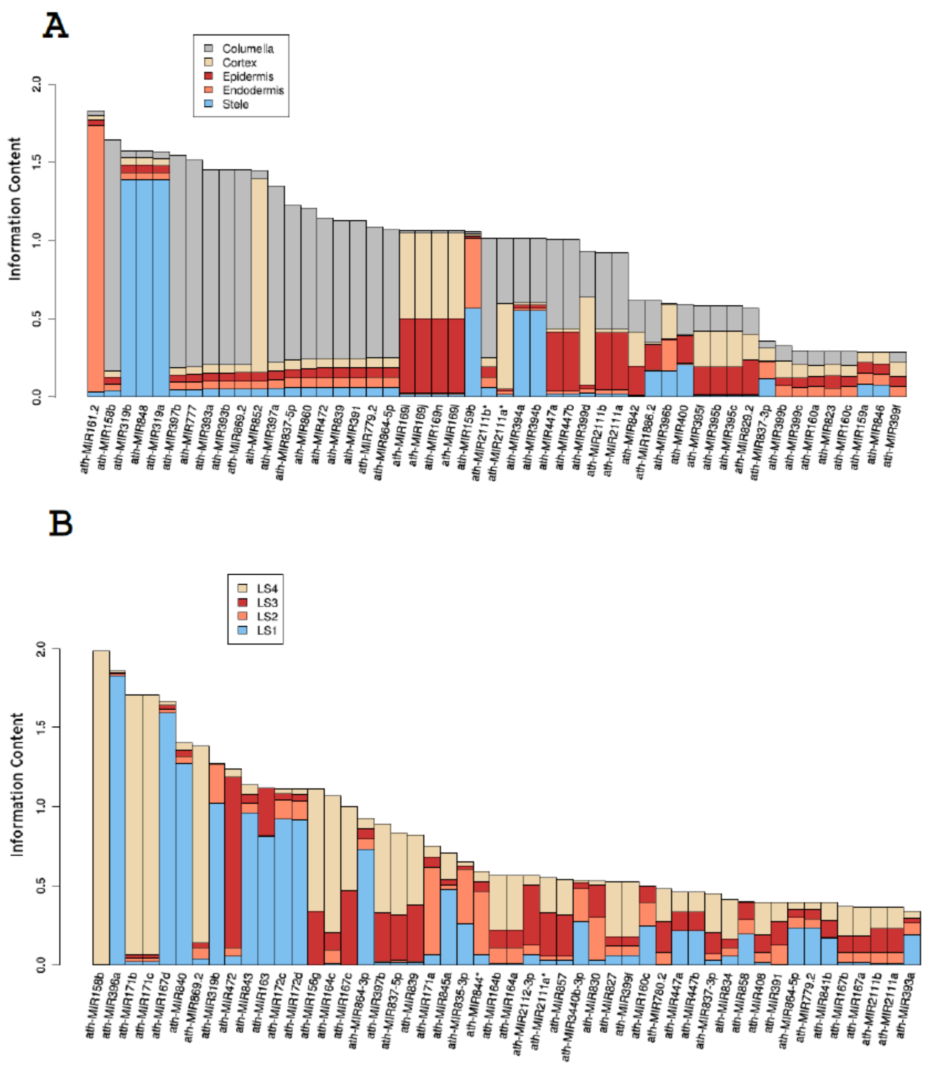

Figure 12: Information content of expression profiles of known miRNAs by cell type (A) or developmental zone (B).

The miRNAs with the lowest variance across tissues, i.e. those that would be most useful for qPCR normalization, are the miR169 and miR162 families (Table 5), while the miRNAs with the lowest variance across longitudinal sections were the miR156 and miR157 families (Table 6). The miRNAs with the highest variance could be considered the ones with the most cell type specificity. These variance calculations also take into account the variability within the biological replicates.

Table 5: Log$_2$ average expression of known miRNAs ranked by their variance across cell types.

miRNA	Stele	Endodermis	Epidermis	Cortex	Columella	Variance
MIR169a	9.826758713	10.21113463	10.15535812	10.50668546	10.45898825	0.074376345
MIR162b	8.448571292	7.865346957	8.32379777	8.676674298	8.384711725	0.088173577
MIR162a	8.448571292	7.865346957	8.32379777	8.676674298	8.384711725	0.088173577
MIR169m	7.039905387	7.141859094	6.682352307	7.505514976	6.581750386	0.138155473
MIR169h	7.078497301	7.163463946	6.64738118	7.486333359	6.52478118	0.154398405
MIR169k	7.103841802	7.163463946	6.64738118	7.486333359	6.529963971	0.154586246
MIR169b	7.624126939	8.019033709	7.775814913	8.309797611	7.254787958	0.159146936
MIR172e	8.442459413	7.611818472	7.335341566	8.01884952	7.741574397	0.177719976
MIR408	7.902232492	7.379819214	8.157672144	7.729105866	8.718688721	0.251602472
MIR169c	7.534566265	8.030498749	7.799359257	8.304114945	6.98019808	0.25627878
MIR172b	9.949332826	9.022148689	8.687739023	9.397713589	9.731285812	0.263159806
MIR172a	9.949332826	9.022148689	8.687739023	9.397713589	9.731285812	0.263159806
MIR173	13.44226632	12.69091696	13.68591584	12.67619385	12.52223454	0.273665293
MIR403	7.543658347	7.056698511	7.745559579	7.622739104	8.573074511	0.302058815
MIR169f	9.326187593	9.501959219	8.299537576	9.59624807	10.05841418	0.42235236
MIR824	7.562830145	7.396763987	8.286427598	7.666997798	8.99786717	0.435523265
MIR169g	9.310293265	9.489510786	8.238342369	9.586591993	10.05257386	0.451337918
MIR169e	9.315610855	9.534692165	8.275912741	9.622726864	10.09008095	0.452367991
MIR169d	9.315610855	9.534692165	8.275912741	9.622726864	10.09008095	0.452367991
MIR164c	6.968895961	6.644089358	5.981022413	7.059916738	5.40655214	0.495115428
MIR164b	10.25595383	10.15641814	8.626852847	8.962724375	9.892008759	0.544008078
MIR164a	10.22636935	10.15641814	8.605876629	8.951862577	9.877948433	0.545242046
MIR3440b-3p	7.399983625	6.862063561	8.579067692	6.787896189	6.663099157	0.624262144
MIR169i	8.104644321	8.293148543	6.57514218	6.728365613	7.953868448	0.661660514
MIR169j	8.079314039	8.293148543	6.549607088	6.7112921	7.89969391	0.662383223
MIR169n	8.079314039	8.293148543	6.549607088	6.7112921	7.89969391	0.662383223
MIR169l	8.171965488	8.293148543	6.540993958	6.7112921	7.914861675	0.697746636

MIR822	11.43399699	11.17429899	13.37600376	12.65119189	12.27243895	0.807233544
MIR156g	10.00706276	10.37162734	10.92161842	11.04950979	8.712767885	0.88003918
MIR168a	14.91614644	13.1829749	16.22653509	14.96707968	14.51936615	1.190718102
MIR168b	14.91614644	13.18265379	16.22546216	14.96468573	14.51220641	1.190819936
MIR829.1	14.35095042	12.88301038	12.76122345	12.68010857	11.08569363	1.338909082
MIR167c	8.119256779	8.347626029	8.526335639	9.288325473	11.02875075	1.401340742
MIR160b	5.41988289	5.856743776	6.12307395	6.973923017	8.418238718	1.402427717
MIR773	5.13606527	4.856743776	6.831216817	7.1094307	7.608406826	1.521669966
MIR775	6.214434599	7.508820472	9.390733407	7.896403505	9.188618143	1.694816214
MIR157c	17.28022953	17.46009284	19.42653235	19.24683112	16.43515051	1.710833878
MIR157a	17.27976756	17.45977784	19.42655979	19.24828142	16.43445492	1.71255393
MIR157b	17.27976756	17.45977784	19.42655979	19.24828142	16.43445492	1.71255393
MIR396a	7.322180799	7.717277189	4.680522682	6.95033165	8.017524136	1.75414879
MIR156b	18.19663836	18.47053518	19.59350788	20.35525995	16.88019175	1.788991987
MIR156d	18.19522672	18.4697997	19.59769427	20.35810746	16.88356348	1.790600344
MIR156c	18.19401749	18.4687296	19.59268007	20.35427636	16.87751324	1.791109402
MIR156a	18.19401749	18.4687296	19.59268007	20.35427636	16.87751324	1.791109402
MIR156f	18.19285828	18.46766235	19.59167929	20.35362073	16.87677658	1.791202518
MIR156e	18.19277264	18.46764588	19.59166617	20.35361757	16.87677184	1.79122179
MIR161.1	12.62029532	11.9434487	15.40344538	13.53779829	14.48669878	1.936622162
MIR157d	10.60479966	11.27432243	14.1743516	13.86451397	11.82178667	2.525951546
MIR166a	15.94353643	16.21419642	14.26906786	18.28676174	14.60471308	2.53208738
MIR158a	17.20453389	17.39925011	16.92698651	17.64902381	13.78385669	2.535508003
MIR166b	15.86893607	16.1188985	14.2126765	18.25577591	14.5475056	2.556404737
MIR166c	15.87078763	16.1202504	14.21446435	18.25762654	14.54923903	2.556448783
MIR166d	15.86963008	16.11986937	14.21356835	18.25645962	14.54734205	2.556816629
MIR166e	15.87151383	16.12083258	14.21384451	18.25752266	14.54931989	2.556880358
MIR166g	15.8722066	16.12180215	14.21473563	18.25820555	14.54917648	2.557278885
MIR166f	15.87080848	16.12083258	14.21306767	18.25739973	14.54926007	2.557359901
MIR172d	7.857288202	6.782929627	7.912164215	6.30919865	10.54230287	2.690096048
MIR172c	7.857288202	6.782929627	7.912164215	6.30919865	10.54230287	2.690096048
MIR156h	8.29481445	7.996287477	9.635871772	9.50851073	5.424349537	2.881180709
MIR167a	8.535751414	10.23760583	9.846630095	10.79389602	13.28517105	3.047410887
MIR167b	8.535751414	10.23760583	9.846630095	10.79693392	13.28771627	3.05129297
MIR165a	13.29047111	14.17266224	11.24354377	15.84620928	12.01978576	3.277929087
MIR165b	13.26046223	14.15319514	11.19147632	15.83565151	11.97920944	3.337028347
MIR390a	9.127824327	9.163515241	13.30515915	9.943151487	12.89867143	4.212565549
MIR390b	9.127824327	9.163515241	13.30531809	9.943151487	12.89976875	4.21386126

Table 6: Log$_2$ expression of known miRNAs ranked by their across developmental zone variance.

miRNA	LS1	LS2	LS3	LS4	Variance
ath-MIR156h	7.008216436	7.618685579	7.31942479	7.303309245	0.062156829
ath-MIR157d	12.78597429	13.07531745	12.43588243	12.85903452	0.070527895

ath-MIR172b	8.140088092	8.784011867	8.226595766	8.285258519	0.083842066
ath-MIR172a	8.140088092	8.784011867	8.226595766	8.285258519	0.083842066
ath-MIR157c	17.66079384	18.39987072	18.11896602	18.06288	0.092787126
ath-MIR157a	17.66058128	18.40217263	18.12223256	18.06508999	0.093496945
ath-MIR157b	17.66058128	18.40217263	18.12223256	18.06508999	0.093496945
ath-MIR777	3.860786072	4.013823521	3.717538987	3.277119721	0.100757728
ath-MIR172e	6.08448233	6.856650017	6.868903578	6.665879324	0.135612046
ath-MIR773	6.380372564	6.024711837	5.891713461	5.323020571	0.193062193
ath-MIR846	9.551712319	10.15724539	10.45026923	9.553404595	0.202410878
ath-MIR169a	7.928781969	8.573055746	8.941754144	8.855329721	0.210231529
ath-MIR842	5.617701256	6.589649227	6.539411717	6.540672243	0.220919385
ath-MIR162a	10.5326742	9.94151171	9.92802716	9.356144652	0.230764911
ath-MIR162b	10.53249508	9.94151171	9.928544286	9.355040408	0.231120071
ath-MIR823	7.179057144	7.73595831	6.727082759	6.719491286	0.231392771
ath-MIR822	11.95788476	12.41318237	13.03613178	12.87700051	0.236933526
ath-MIR168a	14.55957864	14.69806077	14.08439307	13.6214539	0.239586645
ath-MIR168b	14.5558932	14.69518174	14.07736628	13.61471279	0.241447269
ath-MIR156d	16.78734011	17.54621173	17.96543631	17.713815	0.257442158
ath-MIR156c	16.7837209	17.5455511	17.96509372	17.71263732	0.258881731
ath-MIR156a	16.7837209	17.5455511	17.96509372	17.71263732	0.258881731
ath-MIR156b	16.78455076	17.54793385	17.96720637	17.71505728	0.259547439
ath-MIR156e	16.78091511	17.54462423	17.96462302	17.71236338	0.260015188
ath-MIR156f	16.78091982	17.54462885	17.96462762	17.71236843	0.260015197
ath-MIR394a	5.396209112	5.730980634	4.451948876	4.874465946	0.318664635
ath-MIR394b	5.396209112	5.734300992	4.451948876	4.878570344	0.319382598
ath-MIR859	1.557005324	1.492991357	0.897360024	0.326029321	0.33312729
ath-MIR781	3.338365038	1.969429401	2.625280479	2.349876063	0.334192726
ath-MIR863-3p	6.08448233	5.013823521	4.95264246	4.789826132	0.34866113
ath-MIR160b	7.796603853	6.977858023	6.407135029	7.777240433	0.453605757
ath-MIR166b	14.21940946	12.58126111	13.47385913	13.6509599	0.461238469
ath-MIR166f	14.21989618	12.58160679	13.4747741	13.65084742	0.461253397
ath-MIR166e	14.21992399	12.58160679	13.4747741	13.65087554	0.46127025
ath-MIR166g	14.2239089	12.58226911	13.47534935	13.6516346	0.462919319
ath-MIR166d	14.22339578	12.58166439	13.47443474	13.65171892	0.463043692
ath-MIR166c	14.22345126	12.58163559	13.47465607	13.6516346	0.463077675
ath-MIR160a	7.827300651	7.153064783	6.587357996	8.10869998	0.468248467
ath-MIR166a	14.33714776	12.66454892	13.54754229	13.73822866	0.479049545
ath-MIR860	7.54852717	5.763845268	6.841339939	6.635884584	0.54014825
ath-MIR1886.2	7.209850297	5.683674919	5.720727264	5.419138726	0.659637879
ath-MIR169h	5.377184287	6.522098499	7.259303798	6.879594622	0.660491118
ath-MIR169k	5.377184287	6.522098499	7.265866486	6.879594622	0.663782172
ath-MIR169m	5.421191469	6.578608139	7.300372048	6.943680441	0.664196291
ath-MIR390a	13.03178863	12.14197411	11.42788829	11.19634266	0.682573242
ath-MIR390b	13.03210541	12.14197411	11.4280102	11.19665091	0.682604661
ath-MIR862-5p	2.200861514	1.827410396	0.704714946	0.506601567	0.691055584
ath-MIR165b	13.11524665	11.25302423	11.65558324	12.47831416	0.695370661
ath-MIR165a	13.13826476	11.27477868	11.67589922	12.5004141	0.696737621

ath-MIR775	9.340822084	8.83284136	7.579476789	7.718919128	0.735178376
ath-MIR169i	6.514107366	7.543076589	8.55065297	8.225285529	0.810075795
ath-MIR169j	6.505372556	7.541182045	8.54706529	8.21477257	0.81161689
ath-MIR169n	6.505372556	7.541182045	8.54706529	8.21477257	0.81161689
ath-MIR169l	6.496584539	7.538335556	8.541216214	8.210708638	0.814256311
ath-MIR161.1	15.72914956	14.59885718	14.04679954	13.55706045	0.871315503
ath-MIR158a	17.45383142	15.66339612	15.76118635	15.28427389	0.929868988
ath-MIR400	7.028680539	6.040891764	5.25050685	4.671804158	1.04387505
ath-MIR829.1	9.663568265	11.58695426	10.72658724	9.304072374	1.077832501
ath-MIR173	12.90670122	12.8453467	11.30515887	10.88852617	1.084720063
ath-MIR3434	3.284925779	1.554391902	1.219288119	0.942700682	1.109215513
ath-MIR172b*	5.660293133	7.821178443	8.04329217	7.37115189	1.164898199
ath-MIR399c	2.804932838	4.737613726	5.34030352	4.662312709	1.203483056
ath-MIR395f	4.016436943	5.26405015	6.703103896	5.270887767	1.205877532
ath-MIR395b	4.016436943	5.26405015	6.703103896	5.270887767	1.205877532
ath-MIR395c	4.016436943	5.26405015	6.703103896	5.270887767	1.205877532
ath-MIR824	9.564032591	8.513910924	7.230515375	7.281776083	1.243720079
ath-MIR779.2	5.179057144	3.013823521	3.201140772	2.722919474	1.248450883
ath-MIR864-5p	5.291714944	3.378820337	3.10681339	2.794178157	1.265348803
ath-MIR399b	2.76645869	4.770329301	5.385200058	4.807156011	1.312608876
ath-MIR393a	4.746829883	3.42886102	2.10681339	2.647957416	1.313537656
ath-MIR393b	4.746829883	3.42886102	2.10681339	2.647957416	1.313537656
ath-MIR169b	4.351421191	6.387281916	6.884624036	6.992785913	1.513694701
ath-MIR169c	3.983270079	6.151823755	6.88888187	6.89588493	1.893788411
ath-MIR169e	6.036785588	7.999406748	9.203991386	9.004570777	2.099330897
ath-MIR169d	6.036785588	7.999406748	9.203991386	9.004570777	2.099330897
ath-MIR169g	5.962146787	7.925805558	9.170572834	8.982334998	2.164385317
ath-MIR169f	5.96639626	7.938096194	9.17522857	8.993022556	2.168436445
ath-MIR837-3p	5.721912251	6.134336328	7.976844808	8.841203492	2.204489632
ath-MIR167a	8.647117744	10.94572501	11.57909258	12.42329334	2.619717626
ath-MIR167b	8.618376517	10.94474002	11.58040945	12.42557439	2.665967642
ath-MIR408	1.644468165	3.839794121	4.590846982	5.540672243	2.753353767
ath-MIR3440b-3p	8.216109288	7.834389316	5.348571136	4.338001963	3.569455685
ath-MIR164a	7.182714167	10.41029857	10.60622836	12.2535357	4.500042889
ath-MIR164b	7.190000521	10.42308587	10.62059433	12.26930114	4.515572006
ath-MIR172d	9.336724679	6.399881953	4.81024936	4.435947013	4.971310018
ath-MIR172c	9.348168213	6.399881953	4.81024936	4.435947013	4.994924247
ath-MIR869.2	2.601399444	3.378820337	2.219288119	7.622486729	6.209123704
ath-MIR164c	3.048858421	6.447187668	6.829574776	9.786416955	7.606395054

Many known miRNAs show inverse expression with their validated targets

In the simplest case, miRNAs and the mRNA of their targets are expected to have reciprocal expression patterns. To determine how often this occurs, we compared our miRNA expression profiles with target mRNA expression profiles from a previous microarray based root study (Brady et al., 2007). Naturally, this method will only show miRNA:target relationships where the miRNA slices the mRNA target, but not where the miRNA represses the translation of the mRNA. First, we obtained validated targets for known miRNAs expressed in our datasets from the *A. thaliana* Small RNA Project (ASRP) website (Gustafson et al., 2005; Backman et al., 2008). Expression values of all targets for a particular miRNA were averaged within each cell type and compared with the $\log_2$ normalized expression of the miRNA in the same cell type. The z-scores are visualized in a heat map, with yellow representing higher expression (deviation from the mean) and blue representing lower expression (Figure 13).

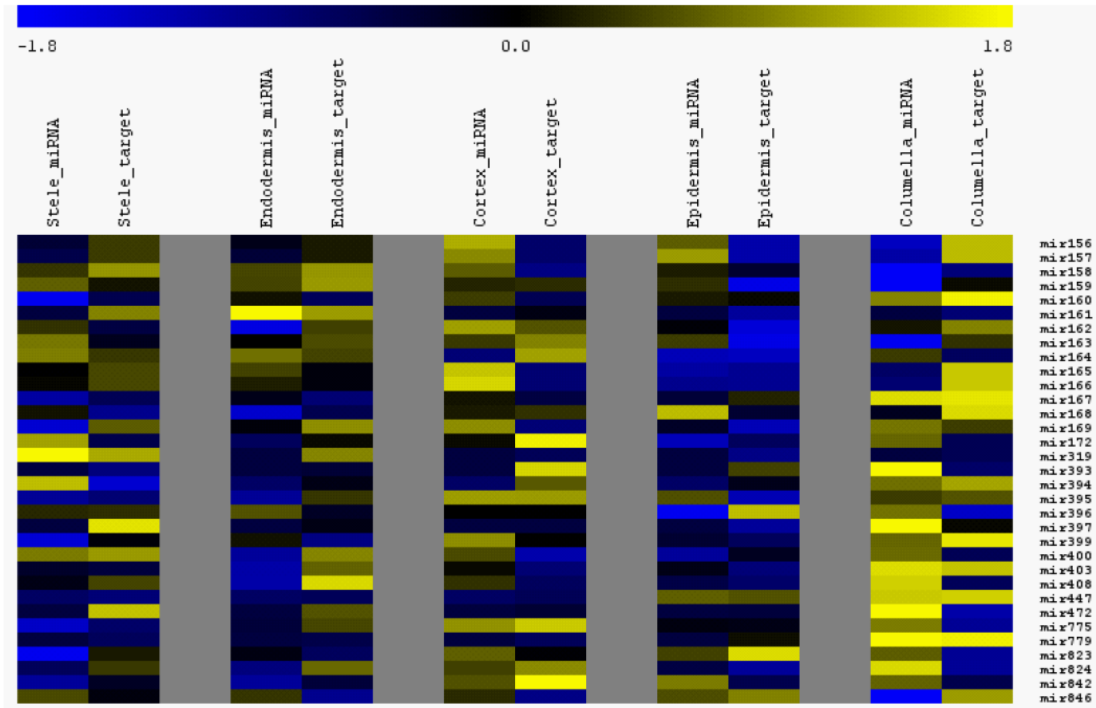

Figure 13: Expression of miRNAs and their validated targets varies by cell type.

In order to identify statistically significant miRNA:target correlations, the Pearson product-moment correlation coefficient (r) and p-values of the expression of known miRNAs (from this study) and their validated targets (list from ASRP website (Backman et al. 2008) and expression values from Brady et al. (2007)) were calculated and ordered according to p-values (Table 7). Analysis of the correlations show three miRNA:mRNA target sets (miR156, miR157, and miR396) with significant p-values for a strong negative correlation as expected. Additionally, two sets (miR447 and miR779) show an unexpected strong positive correlation of unknown relevance. Interestingly, two of the five sets with the least correlation targeted parts of the miRNA biogenesis or processing machinery (miR162 and miR168), while the other three sets target enzymes or transcription factors either known to be regulated translationally (miR172) or likely to be important in root development (miR159 and miR397).

Table 7: Correlation of the expression of known miRNAs and their validated targets ranked by p-value.

miRNA family	target family	r	p-value	target AGI	target name
miR447	2-PGK	0.99	0.0005	AT3G45090	2-phosphoglycerate kinase-related
				AT5G60760	2-phosphoglycerate kinase-related
miR157	SPL	-0.98	0.0028	AT2G33810	SPL3
				AT5G43270	SPL2
				AT1G53160	SPL4
				AT1G27360	SPL11
				AT1G27370	SPL10
				AT1G69170	SPL
				AT2G42200	SPL9
				AT3G57920	SPL15
				AT3G15270	SPL5
miR779	LRR-PK	0.95	0.0134	AT5G53890	PSKR2
miR156	SPL	-0.93	0.0206	AT2G33810	SPL3
				AT5G43270	SPL2
				AT1G53160	SPL4
				AT1G27360	SPL11
				AT1G27370	SPL10
				AT1G69170	SPL
				AT2G42200	SPL9
				AT3G57920	SPL15
				AT3G15270	SPL5
miR396	GRF	-0.88	0.0493	AT2G36400	AtGRF3
				AT3G52910	AtGRF4

				AT5G53660	AtGRF7
				AT4G24150	AtGRF8
				AT2G22840	AtGRF1
				AT4G37740	AtGRF2
miR167	ARF	0.81	0.0967	AT5G37020	ARF8
				AT1G30330	ARF6
miR408	LAC	-0.68	0.2078	AT2G02850	ARPN
				AT2G30210	LAC3
miR319	TCP	0.68	0.2098	AT1G53230	TCP3
				AT3G15030	TCP4
				AT1G30210	TCP24
				AT2G31070	TCP10
				AT4G18390	TCP2
miR472	CC-NBS-LRR	-0.66	0.2224	AT1G12210	RFL1
				AT1G12220	RPS5
				AT1G12280	LRR and NB-ARC domains-containing
				AT1G12290	Disease resistance protein (CC-NBS-LRR class) family
				AT1G15890	Disease resistance protein (CC-NBS-LRR class) family
				AT1G51480	Disease resistance protein (CC-NBS-LRR class) family
				AT4G10780	LRR and NB-ARC domains-containing disease resistance protein
				AT4G27190	NB-ARC domain-containing disease resistance protein
				AT5G05400	LRR and NB-ARC domains-containing disease resistance protein
				AT5G43730	Disease resistance

					protein (CC-NBS-LRR class) family
				AT5G43740	Disease resistance protein (CC-NBS-LRR class) family
				AT5G47260	ATP binding
				AT5G63020	Disease resistance protein (CC-NBS-LRR class) family
miR161	PPR	0.62	0.2634	AT1G62860	pseudogene of pentatricopeptide (PPR) repeat-containing protein
				AT1G06580	PPR
				AT1G62670	RPF2
				AT1G62910	PPR
				AT1G62930	TPR
				AT1G63080	ta-siRNA generating locus
				AT1G63150	ta-siRNA generating locus
				AT1G64580	PPR
				AT5G16640	PPR
				AT5G41170	PPR
				AT2G41720	EMB2654
				AT1G63070	PPR
				AT3G16710	PPR
				AT4G26800	PPR
				AT5G65560	PPR
miR160	ARF	0.56	0.3309	AT2G28350	ARF10
				AT4G30080	ARF16
				AT1G77850	ARF17
miR846	JR/MBP	-0.55	0.3379	AT1g52050	JR/MBP
				AT1g52060	JR/MBP
				AT1g52070	JR/MBP
				AT1g57570	JR/MBP
				AT2g25980	JR/MBP
				AT5g28520	JR/MBP

				AT5g49870	JR/MBP
miR158	PPR	0.45	0.4527	AT1G64100	PPR
				AT3G03580	TPR
miR393	TIR1/bHLH	-0.41	0.4929	AT1G12820	IPS1
				AT3G62980	TIR1
				AT3G23690	bHLH
				AT3G26810	AFB2
miR824	MADS-Box	-0.39	0.5129	AT3G57230	AGL16
miR403	AGO2	0.39	0.5159	AT1G31280	AGO2
miR399	E2-UBC	0.35	0.56	AT2G33770	PHO2
miR775	GT	0.35	0.5645	AT1G53290	Galactosyltransferase family protein
miR166	HDZIPIII	-0.33	0.5874	AT1G52150	ATHB-15
				AT4G32880	ATHB-8
				AT2G34710	PHB
				AT1G30490	PHV
				AT5G60690	REV
miR164	NAC	0.32	0.6013	AT5G07680	ANAC079/ANAC080
				AT5G61430	ANAC100
				AT3G15170	CUC1
				AT5G53950	CUC2
				AT1G56010	NAC1
				AT5G39610	ANAC092
miR394	F-Box	-0.32	0.6052	AT1G27340	Galactose oxidase/kelch repeat superfamily protein
miR395	APS,AST	0.3	0.6192	AT3G22890	APS1
				AT4G14680	APS3
				AT5G43780	APS4
				AT5G10180	AST68
miR169	HAP2	-0.29	0.6356	AT3G20910	NF-YA9
				AT5G12840	HAP2A
				AT1G72830	HAP2C
				AT1G17590	NF-YA8
				AT1G54160	NF-YA5
				AT3G05690	HAP2B
				AT5G06510	NF-YA10

miR163	SAMT/FA MT	-0.28	0.6489	AT1G66690	SAMT
				AT1G66720	SAMT
				AT3G44860	FAMT
miR165	HDZIPIII	-0.26	0.6786	AT1G52150	ATHB-15
				AT4G32880	ATHB-8
				AT2G34710	PHB
				AT1G30490	PHV
				AT5G60690	REV
miR400	PPR	-0.23	0.7078	AT5G39710	EMB2745
				AT1G06580	PPR
				AT1G22960	PPR
				AT1G62670	RPF2
				AT1G62720	PPR
				AT1G62910	PPR
				AT1G62930	TPR
				AT1G63070	PPR
				AT1G63080	TPR
				AT1G64580	PPR
				AT3G16710	PPR
				AT4G19440	TPR
				AT5G16640	PPR
				AT5G46680	PPR
miR842	JR/MBP	0.22	0.7177	AT5G38550	JR/MBP
miR823	CMT3	-0.06	0.9299	AT1g69770	CMT3
miR397	LAC	0.03	0.9564	AT2G38080	IRX12/LAC4
				AT5G60020	LAC17
				AT2G29130	LAC2
miR172	AP2	-0.03	0.9675	AT4G36920	AP2
				AT3G54990	SMZ
				AT2G39250	SNZ
				AT2G28550	TOE1
				AT5G60120	TOE2
				AT5G67180	TOE3
miR168	AGO	-0.02	0.9701	AT1G48410	AGO1
miR159	MYB	0.01	0.9879	AT2G32460	MYB101
				AT5G55020	MYB120

				AT2G26950	MYB104
				AT4G26930	MYB97
				AT3G11440	ATMYB65
				AT5G06100	MYB33
miR162	DCL	0	0.9961	AT1G01040	DCL1

We found that known miRNAs and their targets have reciprocal expression patterns in some cell types or developmental zones, but these simple patterns are not found in all cases. Overall, most known miRNAs have enrichment in one or more cell types and/or developmental zones. These spatio-temporal patterns will be very useful in the elucidation of specific functional roles for root miRNAs in biological circuits.

PIPmiR identifies novel miRNAs expressed in the root.

PIPmiR was applied to each of our root datasets to identify new miRNA genes. Across all libraries, we identified 420,974 unique putative mature miRNA sequences from intergenic or intronic regions that had a valid pre-miRNA fold structure. Putative mature miRNAs that overlapped by at least 90% were grouped into a single mature sequence. The PIPmiR algorithm identified 184 putative mature sequences from 168 precursors that were classified as a miRNA in at least 1 of the datasets. Of these potential mature miRNAs, 145 from 135 precursors had a median classification score > 0 across all of the datasets in which they were observed (Appendix B).

All of these candidates bear the hallmarks of products of the small RNA processing pathways in plants, and exceed the evidence available for many recent miRBase additions based on high throughput sequencing. However, the presence of a small RNA pathway substrate does not imply its consistent physiological function. Here, we took advantage of our biological replicates and applied a more rigorous filter to identify a more stringent set of novel miRNAs for further analysis. We additionally required that each of the putative miRNAs be expressed in both biological replicates for at least a single tissue and have a positive score in each of those replicates. To account for the single replicates of the developmental zones, we required a minimum raw transcript count of 25 and a positive score. Using these criteria, the pipeline reported 70 high confidence novel miRNAs from 68 precursors. Sixty-two of the 70 previously unknown miRNAs (89%) are reported to have at least one target using the WMD3 target predictor (Ossowski et al., 2008). This is highly similar to the number of known miRNAs with predicted targets (220 of 243, 91%).

Novel miRNAs are validated

Twenty-nine of the miRNAs we identified could immediately be classified as new miRNAs based on standard annotation criteria (Meyers et al., 2008) which accepts as evidence the presence of both miRNA and miRNA* sequences on predicted hairpin structures (Table 8). A total of 12 novel miRNAs on our complete list were recently published (Borges et al., 2011; Yang et al., 2011) while our manuscript was under review. These are indicated on Table 8 and Appendix B by bold type.

Table 8. Validated novel miRNAs.

miRNA	Chromosome (Strand)	Precursor Start	Precursor End	Mature Sequence	Expression Location	Additional Validation
novel201442	Chr5(+)	3802713	3803061	UUAGGGUAGUUAACGGAAGUUA	LS1*, LS2*, LS3*, LS4*	
novel33140	Chr3(-)	5497004	5497100	UUGGUAGUGGAUAAGGGGGCA	WT, WER, LS1, LS2, LS3, LS4	qPCR
novel232253	Chr2(-)	8392448	8392607	UUGGAUUUAUAGUUGGAUAAG	LS1*	
novel18686	Chr4(+)	14982319	14982544	AGGGACUUUGUGAAUUUAGGG	LS3*, LS4*	
novel21507	Chr5(+)	6926004	6926445	UGUUAAGGAGUGUUAACGGUG	WOL, SCR, WT, LS2*	
novel233188	Chr2(-)	12270178	12270372	AAUGCGCAACUCUAUAUUUCC	LS1*, LS2*, LS3*, LS4*	
novel15045	Chr3(-)	15336704	15337169	AUACCAAAACUCUCUCACUUU	LS2*, LS3, LS4*	
novel85988	Chr2(+)	10356781	10357162	UAGUCCACUGUGGUCUAAGGC	LS1*	qPCR
novel207342	Chr1(-)	1653220	1653623	UGAGAGAAGGAAUUAGAUUCA	LS1*, LS3*, LS4*	
novel28598	Chr3(+)	14196773	14197146	UCUCGCGCUUGUACGGCUUU	LS1*	
novel172045	Chr5(+)	14099736	14100204	ACAGUGGUCAUCUGGUGGGCU	LS1, LS2*, LS3, LS4	
novel316512	Chr3(+)	17418775	17419219	AUUUGAGUCAUGUCGUUAAG	LS2*	
novel327967	Chr4(+)	4889420	4889913	AUUUGAGUCAUGUCGUUAAG	LS2*	
novel122522	Chr1(-)	22084263	22084682	UCAAGUUUGAUGACGAUUCCA	LS1*, LS2, LS3*, LS4	
novel27285	Chr4(+)	7844513	7844667	ACUCAUAAGAUCGUGACACGU	WOL, SCR, WT, WER, COR, PET*, LS1*, LS2*, LS3*, LS4	qPCR
novel406654	Chr1(+)	11499383	11499882	UUUGGAAAUAUUUGGCUUGACU	LS4*	
novel9277	Chr1(+)	11499383	11499882	AUCUGAAGAAAAUAGCGGCAU	WT, LS2, LS3*, LS4*	Cloning, qPCR
novel31892	Chr2(+)	19686956	19687038	UUGUUUUGGAUCUUAGAUACA	LS4*	Cloning, qPCR
novel182421	Chr1(-)	29525201	29525285	UGUUUGUUGACAUCGGUCUAG	LS1*, LS2*, LS3*, LS4*	
novel63198	Chr1(-)	7423520	7423606	UUAGAGUUUUCUGGAUACUUA	LS1, LS2, LS3, LS4*	
novel32940	Chr3(+)	17178446	17178607	UUGUGCGGUUCAAAUAGUAAC	WT*, LS1*, LS2*, LS3*, LS4*	
novel3806	Chr1(-)	19026914	19026999	UGGGUUGAGUUGAGUUGAGUUGGC	WER, LS1*, LS2	
novel89582	Chr1(+)	11786291	11786370	AUAAAUCCCAACAUCUUCCA	LS1*, LS2*, LS3*, LS4*	
novel291414	Chr1(-)	16766270	16766423	UGACAAGGCCAAGAUAUAACA	LS2*, LS3*, LS4*	
novel214285	Chr1(-)	16766270	16766423	CCGUAUCUUGGCCUUGUCAUU	LS3*, LS4*	
novel27349	Chr1(+)	1727262	1727414	ACUGAAGUAGAGAUUGGGUUU	PET, LS1,	qPCR

miRNA	Chromosome (Strand)	Precursor Start	Precursor End	Mature Sequence	Expression Location	Additional Validation
					LS2, LS3*, LS4	
novel21861	Chr5(+)	9789581	9789665	UGGACAAGGUUAGAUUUGGUG	SCR, WER, COR, PET, LS1, LS2, LS3, LS4	qPCR
novel261634	Chr4(-)	5292625	5292819	CAGGUGGUUAGUGCAAUGGAA	LS1*	
novel27882	Chr2(+)	2804	3175	UCUCGCGCUUGUACGGCUUU	LS1*, LS2	
novel29385	Chr4(-)	6791338	6791428	UGGAAGAAGGUGAGACUUGCA	LS2, LS3, LS4	qPCR
novel396525	Chr1(+)	17386794	17387007	AUAGUCAAUUUUAUCGGUCUG	LS3*, LS4*	
novel328342	Chr4(+)	5321226	5321642	AUUUGAGUCAUGUCGUUAAG	LS2*	

This table lists novel mature miRNAs identified in the deep sequencing data that either had a miRNA* sequence or were validated by qPCR. Cloning refers to stem loop PCR with traditional cloning and sequencing. Expression location represents tissues in which the miRNA is expressed in all biological replicates. (*) Star sequence was present in that particular tissue in addition to the mature sequence. Abbreviations: WOL (stele), SCR (endodermis), COR (cortex), WER (epidermis), PET (columella), LS (longitudinal section). LS1 corresponds to early meristematic zone, LS2 to late meristematic zone, LS3 to elongation zone, and LS4 to maturation zone. Bold indicates these novel miRNAs were published while our article was under review.

We further validated 8 new miRNAs using stem-loop PCR. In this technique, the stem loop reverse transcription (RT) primer is about 50 nt long, with one end complementary to the 3′ end of the small RNA of interest and the other end folding back upon itself to form a stem loop. After RT, the product can be amplified with primers complementary to the 5′ end of the small RNA of interest and the stem loop region to show that the small RNA interest was expressed in the RNA sample tested (Varkonyi-Gasic et al., 2007).

Seven of these miRNAs had T-DNA insertion lines readily available (Sessions et al., 2002; Alonso et al., 2003). For the eighth, Ignacio Rubio-Somoza in Detlef Weigel's laboratory generated a target mimicry line (Franco-Zorrilla et al., 2007). To assess whether the disruptions in our new miRNAs resulted in root development phenotypes, we compared growth rates of the primary root of the knockout lines with a wild-type or empty vector control.

Mutations in three of the new miRNAs, novel33140, novel406654 and novel21861, resulted in a root growth phenotype under standard conditions (Figure 14). Novel33140 and novel21861 had longer roots than wt, while novel406654 had shorter roots. Knockdown of novel 33140 and novel406654 are shown in Figure 15. Novel33140 was found in the epidermis and all developmental zones, with highest expression in the late meristematic zone. Novel21861 was expressed in the late meristematic, elongation and maturation zones, with the highest expression in the late meristematic zone. No T-DNA insertion lines were available for novel21861, but a line using suppression by target mimicry also had an enhanced root growth phenotype. Novel406654 was found in all radial cell types and all developmental zones, with the highest expression in the epidermis, and a knockout line exhibited retarded root growth. While further analysis is needed to determine the exact role of these miRNAs in root development, these results suggest that new miRNAs we identified have functional relevance.

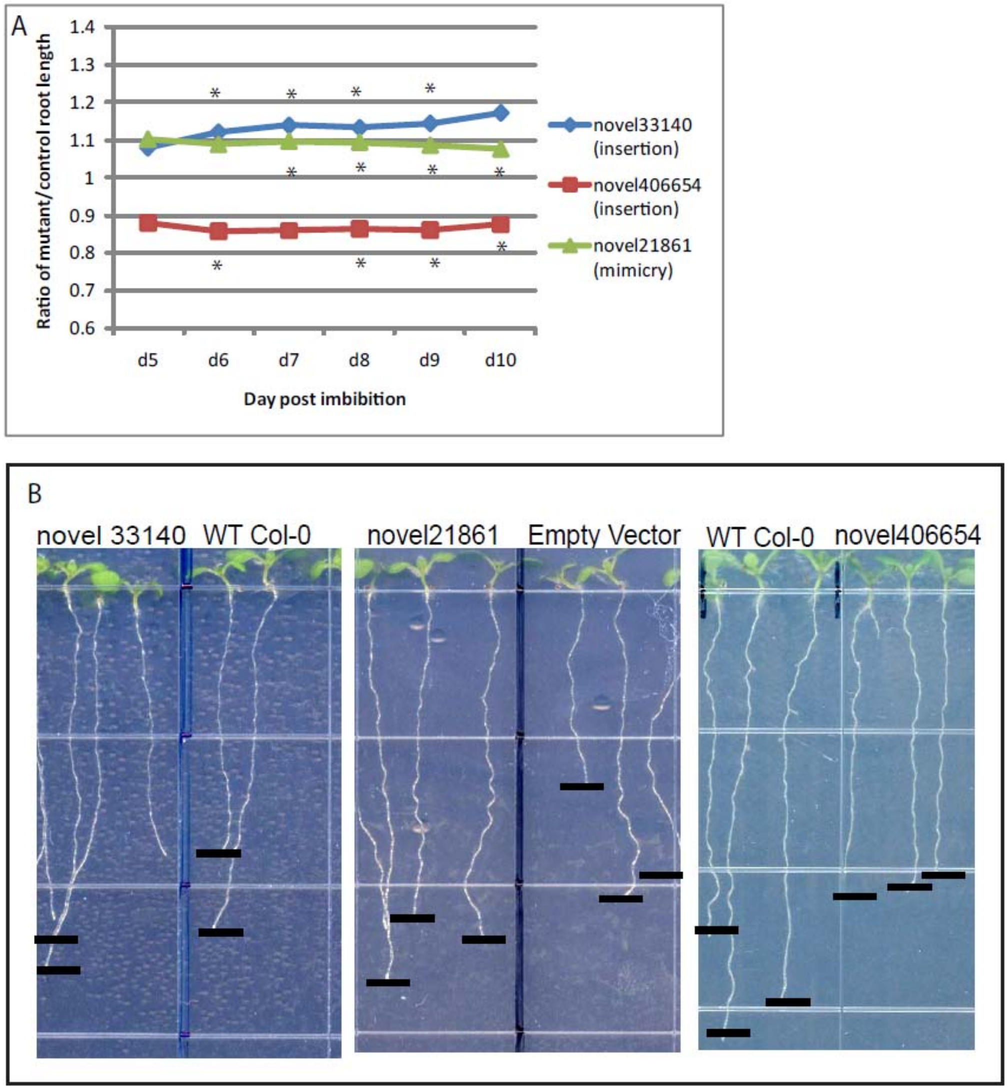

Figure 14. Mutants in three novel miRNAs show root length phenotypes. (A) Ratios of root lengths with * denoting p-values <0.05 (B) Pictures of roots at 8 dpi.

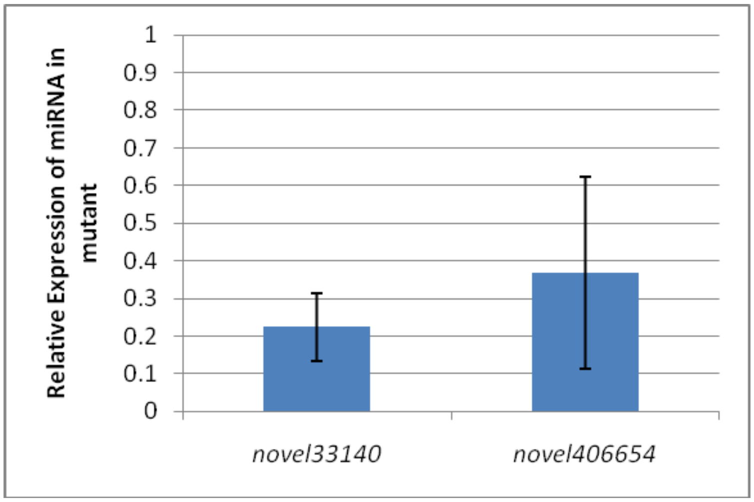

Figure 15. Relative expression showing knockdown of two novel miRNAs in their respective T-DNA insertion mutants.

Novel33140 (miR10) characterization

This novel miRNA was one of the most promising from our original analysis of the datasets so most of the characterization centered on it. The original phenotype of the T-DNA insertion mutant showed longer roots, as seen in Figure 14. This was surprising because many mutants show shorter roots, but very few have longer roots. Additionally, the mutant had more cortex cells in the meristematic region of the root, as shown in Figure 16. This measurement is a proxy for meristem size (Dello Ioio et al., 2007; Tsukagoshi et al., 2010).

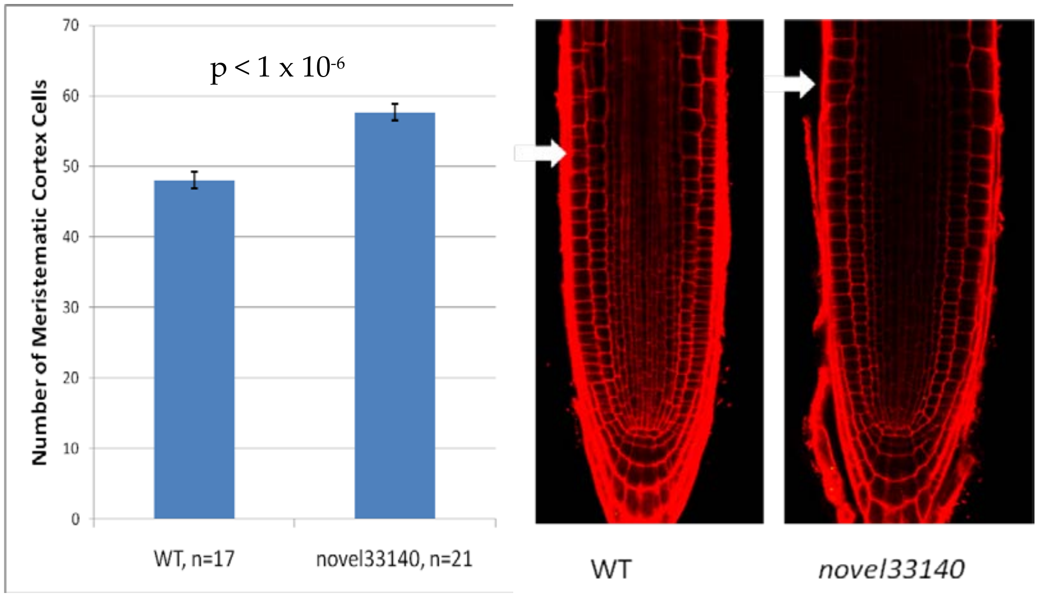

Figure 16. ***Novel33140*** **has an increase in the number of cortex cells in the meristematic region.**

Novel33140 (miR10) targets

In plants, target prediction of miRNAs is reliable due to extensive complementarity between the miRNA and mRNA targets (Llave et al., 2002). Using three different target predictors, WMD3 (Ossowski et al., 2008), psRNA target (Dai and Zhao, 2011), and UEA small RNA toolkit (Moxon et al., 2008), two genes were identified as putative targets of novel33140 (originally named in my notes as miR10). At1g80740, chromomethylase 1, was the first putative target identified, and a great deal of effort went into trying to validate this target using 5′ RLM-RACE (Llave et al., 2011). According to the root map, this target is very lowly expressed in the root. Ultimately, using a *fiery* mutant (Gy et al., 2007) that stabilizes mRNAs cleaved by miRNAs, I was able to isolate 5′RLM-RACE clones that showed that At1g80740 was not cleaved in the location that was predicted by complementarity to novel33140, indicating that it probably was not targeted by this miRNA (Figure 17). This analysis also demonstrated that the annotated intron between exons 15 and 16 is often transcribed. The second putative target, AT4G35730, was identified later and 5′RLM-RACE was not attempted since I could check the mRNA levels on the microarrays described below. Neither putative target changed significantly in the microarrays, indicating that they are probably not targets of novel33140 either.

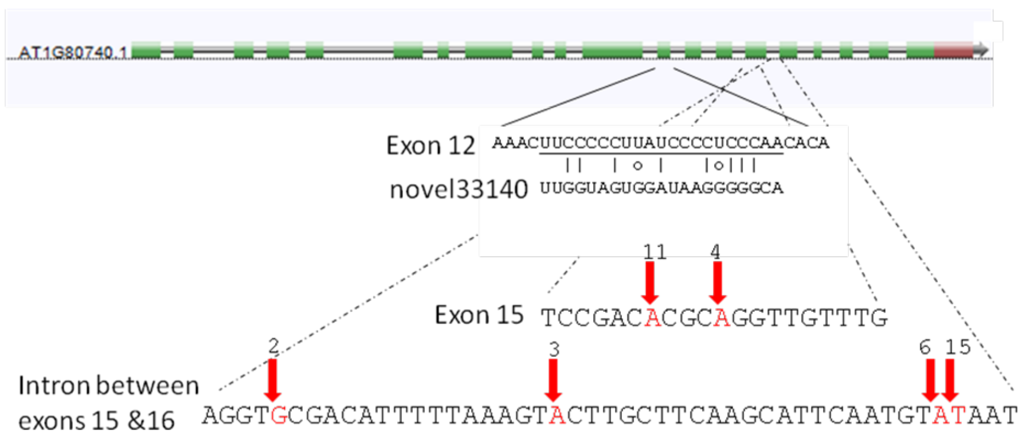

Figure 17. 5′RACE results for At1g80740. The target site for novel33140 was in exon 12, but all 41 RACE products were in exon 15 or the annotated intron between exon 15 and 16.

To identify the targets of novel33140, ATH1 Affymetrix microarray analyses were performed. The first set was done on whole roots of Col-0 WT plants vs. novel33140 mutant. This revealed only three upregulated genes and no downregulated genes at 1.5 fold change and q-value <0.05. None of the upregulated genes had complementary regions to novel33140. Since novel33140 is mainly expressed in the meristematic zone (LS1 and LS2), we thought the miRNA targets could be swamped out in a whole root sample. Consequently, Col-0 and novel33140 mutant roots were hand dissected to isolate the meristematic zone for microarray analysis. There were seven downregulated targets (1.5 fold change, q-value <0.001). The targets that were upregulated in the mutant are shown in Table 9.

Table 9. Putative miR10 targets from microarrays, with >1.5 fold change and q<0.05.

Gene number	Gene Name	miR10 expression	WT expression	q-value	Fold-change
At4g37160	SKU similar 15 (SKS15); glycosyl phosphatidyl-inositol anchored glycoprotein	3.110	1.951	0.005	1.594
At5g63530	FARNESYLATED PROTEIN 3 (FP3)	2.425	1.361	0.002	1.776
At1g23160	Auxin-responsive GH3 family protein	1.196	0.752	0.022	1.589

At4g37160 belongs to a family of proteins that has been implicated in controlling directional root growth (Sedbrook et al., 2002), while At5g63530 is known to affect copper metal binding (Waters and Grusak, 2008). The gene At1g23160 is part of a group of enzymes thought to conjugate plant hormones to various amino acids and sugars for signaling pathways (Shockey et al., 2203). None of the putative novel33140 targets had the canonical complementarity region normally found in plant miRNAs. These results could indicate that novel33140 could act on its target(s) through a less stringent targeting site more similar to animal miRNAs, the target(s) of novel33140 is not represented on the ATH1 microarray, or novel33140 does not act as a miRNA at all. However, the detection of the novel33140 small RNA sequence in the Col-0 whole root samples and decrease in the novel33140 mutant (Figure 14), along with its precursor folding and expression of miRNA and miRNA* in our datasets, supports it as an actual miRNA.

Novel21861 (miR12) characterization

The target mimicry line for novel21861 (originally named miR12) also had longer roots (Figure 14) so this line was further characterized as well. A target mimicry line is not expected to show knockdown of the small RNA itself since it works by sequestering the miRNA of interest through binding to a modified IPS1 transcript (Franco-Zorrilla et al., 2007). Consequently, the confirmation of target mimicry lines is generally done by showing the upregulation of targets. For novel21861, the WMD3 target predictor predicted At1g61140 (Embryo developmental arrest 16) and At2g33400 (unknown protein) as shown in Figure 18. These two genes consistently showed upregulation in the mimicry line as would be expected for a miRNA target (Figure 19), while 2 other genes predicted by WMD3 as having a novel21861 target site had inconsistent expression.

The novel21861 target At1g61140 was annotated as an embryo developmental arrest gene (EDA16) but it contains no domains that give a hint of its possible function. Interestingly, this gene is highly expressed in most of the root, except the meristematic zone which is also where novel21861 is most highly expressed. At2g33400 is an unknown protein that was not on the ATH1 array, initially restricting our understanding of its expression. Preliminary examination of RNA-seq data (presented in Chapter 3) shows this gene not expressed in the columella, but expressed in the other cell types examined.

```
5' CUCUAGCUAUGAAAAUGCCGGUAGUUUCUUCAGAAUAUUC 3' At1g61140
                 |||||| || ||||||||||
3'               UACGGCGAUAAAAGAAGUCUA      5' novel21861
5' AUCUGUAGCUAUGAGCCGCUGCUUUCUUCAGAUCCGACAG 3' At2g33400
               o ||||||o |||||||||||
3'             UACGGCGAUAAAAGAAGUCUA        5' novel21861
```

Figure 18. Targets of novel21861 and their complementary regions.

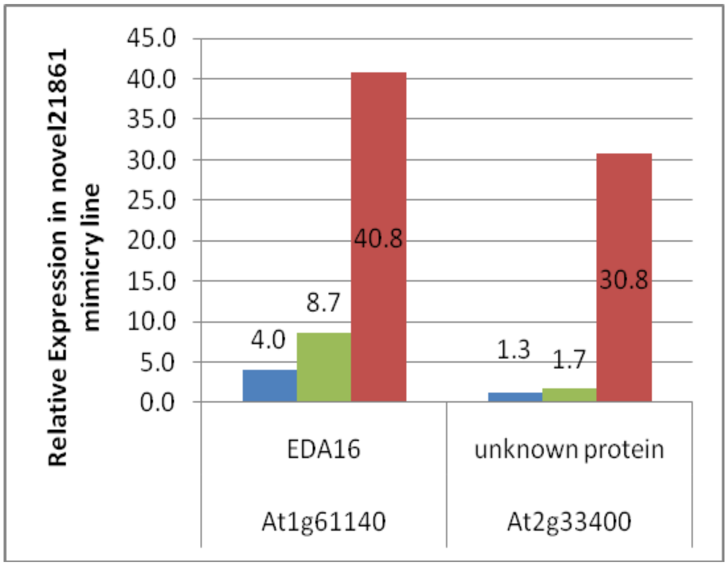

Figure 19. Relative expression showing upregulation of two novel21861 target genes in the target mimicry line.

Novel21861 root phenotypes

Since the mimicry line for novel21861 had a root length phenotype, we wanted to characterize the mutant. As with the mutant for novel33140, I counted the cortex cells in the meristematic zone using confocal microscopy, but found no change from the empty vector control line. The line appeared to have some problems in the radial pattern organization by eye, however, so I quantified other aspects of the radial pattern (Figure 20). The number of quiescent center (QC) cells and the number of columella tiers were not statistically different from the WT control. The mimicry lines show more stele initial cells than the empty vector control ($p<0.05$), but the relevance of this unknown. Crosses of the mimicry line to stele marker lines that also mark initial cells (specifically S32 marking the protophloem and J2501 marking the metaxylem) were done. Changes in the expression of the marker lines in the initials could show if the differences in the stele initials are in the initials of these two cell types.

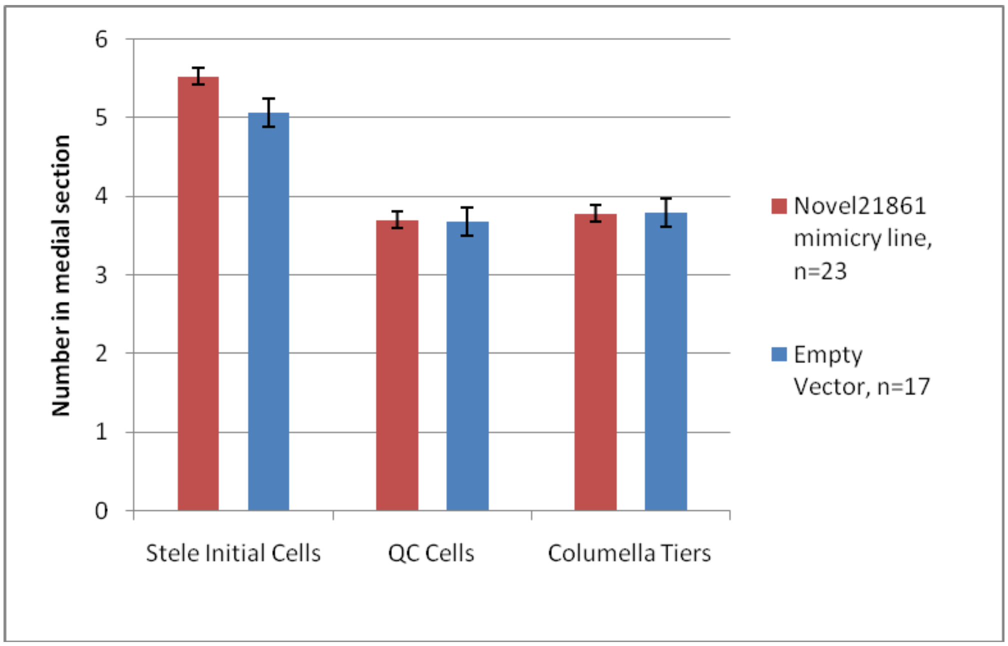

Figure 20. Novel21861 target mimicry line has more stele initial cells than empty vector control.

Chapter 3. Analysis of ra-siRNAs and ta-siRNAs

As described in Chapter 1, ra-siRNAs are small interfering RNAs that direct methylation, mostly at repeat containing regions of the genome. There is a strong direct relationship between the presence of ra-siRNAs and methylation at particular regions of the genome, but a weaker link between this ra-siRNA/methylation and mRNA expression. Additionally, we thought it would be useful to examine the cell type specific response in the root.

Ra-siRNA analysis

In order to analyze the function of ra-siRNAs in individual cell types and developmental zones in the root, some new datasets were generated. I sorted 1,000,000 cells of each cell type for bisulfite sequencing to show which DNAs are methylated, and also 1,000,000 cells of each cell type for RNA-seq to show the mRNA transcripts in each cell type. Additionally, 100 roots were hand dissected into four longitudinal zones, similar to the dissection described for the small RNA datasets (Figure 6) with the exceptions that these were for bisulfite sequencing, and the first section contained only columella for reasons described below. These cells were shipped to Ryan Lister in Joseph Ecker's lab at the Salk Institute for Biological Science, LaJolla, CA. Ryan prepared the libraries and did the bisulfite and RNA-seq using the Illumina platform.

One concern was that the methylation patterns we would see in these cells would be due to the insertion event that created the GFP marker line. Another concern was that the methylation pattern might not be maintained stably for a number of generations. To address both concerns, I identified, screened for the homozygous line, bulked, and sorted an independent transformant for both the pSCR:GFP (designated SCR1.1.2 or endo r2) and the pCOR:GFP (designated COR3.1.1) lines. The pSCR:GFP sorting line that is normally used for cell sorting in our lab has been bulked at least 15 generations and the pCOR:GFP sorting line was calculated to be in about the 8th generation, while the SCR1.1.2 and COR.3.1.1 lines were both in about the 4th generation. The independent transformants were bisulfite sequenced, and compared to the methylation found in the regular lab sorting line. These data have not been analyzed with rigorous statistics as yet, but they appear to correlate quite well by eye when viewed in the Ecker Lab's genome browser (Figure 21).

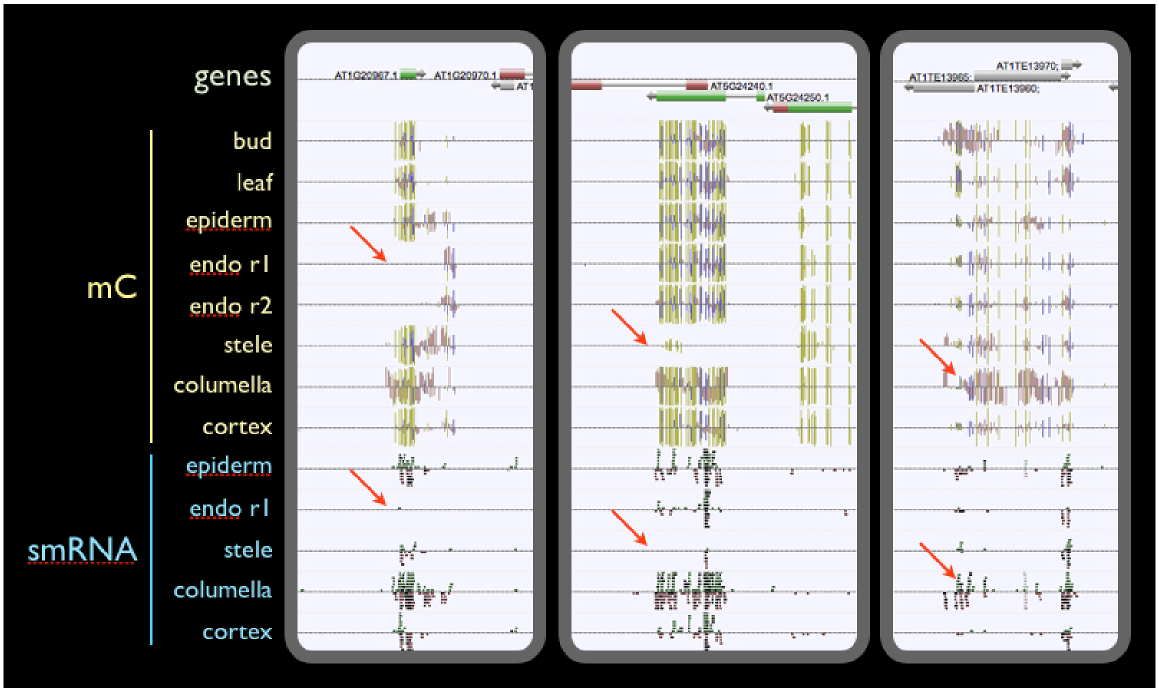

Figure 21. Examples of genomic regions showing cell type specific methylation and small RNAs.

When the preliminary analysis was done for the amount of methylation in the different cell types, the columella sample had twice as much methylation as any of the other cell types (Figure 22). Further breakdown of the subtypes of methylation show that the columella has much greater CHH methylation than any of the other cell types. In order to confirm this finding and exclude the possibility that this difference was due to the insertion event in the sorting line, we wanted to sort either an independent transformant or an additional columella marker. Neither was available, so I hand dissected the columella when doing the longitudinal sections. These samples have not been bisulfite sequenced yet, but they should allow us to determine if the columella has increased CHH methylation as compared to the other cell types.

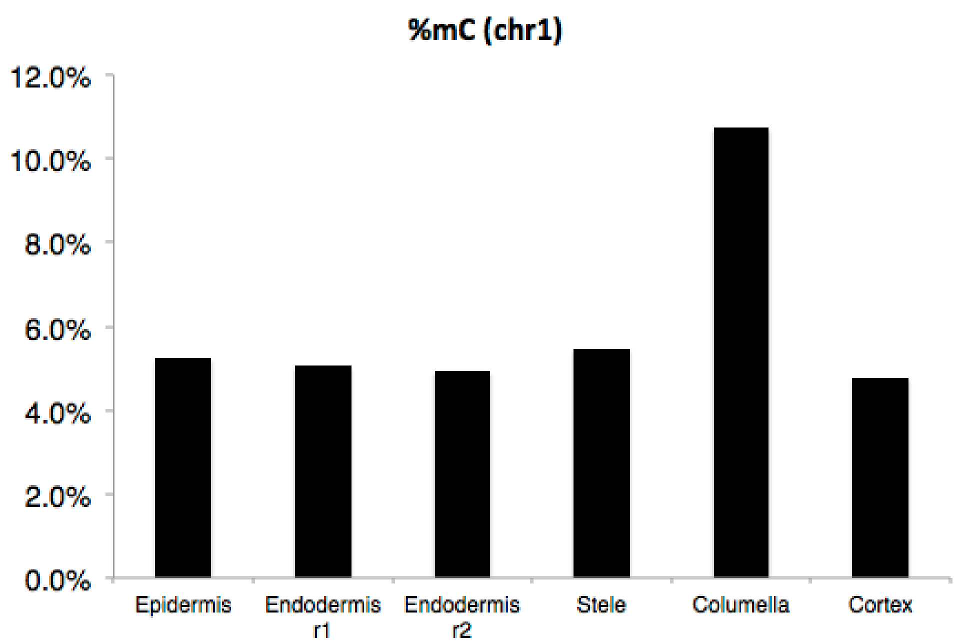

Figure 22. Percent of methylation on chromosome 1 in the different cell types.

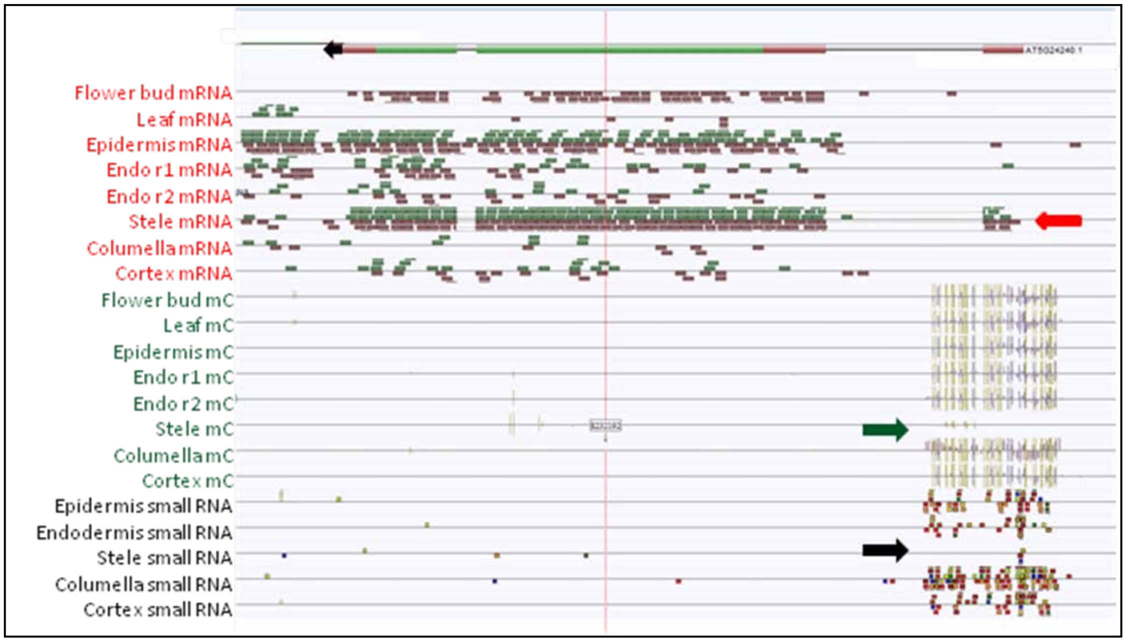

Figure 23. Screen shot of At5g24240 showing differential methylation, small RNAs, and mRNA expression in the stele.

An example of a gene showing cell type specific differences is shown in Figure 23. The methylation pattern in the promoter of At5g24240, a phosphotidylinositol 3- and 4- kinase family/ubiquitin protein, is similar in all the cell types except the stele, which shows low methylation (green arrow). Additionally, the same methylated region has many small RNAs in all the cell types except the stele which contains very few small RNAs (black arrow). Interestingly, the At5g24240 transcript containing the full 5′UTR appears only to be expressed in the stele (red arrow), while a shorter At5g24240 transcript missing an intron in the 5′UTR is expressed in the epidermis. The increase in length corresponds with the transcription of the region that lacks methylation and small RNAs.

Ta-siRNA analysis

The small RNA datasets were analyzed for ta-siRNAs by Jiandong Ding, a graduate student at Fudan University, Shanghai, China. Jiandong developed a ta-siRNA predictor which he used to examine the expression of known ta-siRNAs and identify novel ta-siRNAs. His results are being assembled into a manuscript, and a selection of these results is presented below.

When examining TAS1 loci, we saw that around 40% of the secondary ta-siRNA products were being produced upstream of the miR173 cleavage site. Upon further examination, we found a conserved binding site for one of the novel miRNAs from Chapter 2, novel24173, upstream of the miR173 cleavage site in all three TAS1 loci (Figure 24). This site seems to be specific for TAS1 loci, as it is not found in other loci such as TAS2.

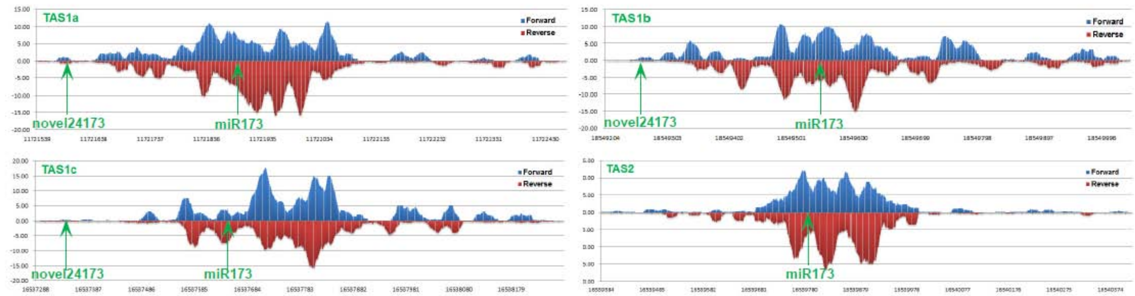

Figure 24. TAS1a, b, and c and TAS2 loci showing miR173 and novel24173 targeting sites.

We also identified other potential TAS loci, including 8 PPR genes, which warrant further study (Table 10). It is noteworthy that TAS4 was not found to be expressed in the root. The ta-siRNAs from this locus target genes involved in anthocyanin production, so perhaps this is not completely unexpected.

Table 10. TAS loci identified in root datasets.

Genomic Location	Strand	Gene number	Annotation
Chr1:18549001-18550042	-	AT1G50055	TAS1b
Chr1:23178001-23179000	-	AT1G62590	PPR
Chr1:23276001-23276500	-	AT1G62860	PPR
Chr1:23299501-23300000	+	AT1G62910	PPR
Chr1:23302001-23302500	+	AT1G62914	PPR
Chr1:23307001-23307500	+	AT1G62930	PPR
Chr1:23385501-23387000	-	AT1G63070	PPR
Chr1:23389501-23390500	-	AT1G63080	TAS2 generating locus
Chr1:23413501-23414000	+	AT1G63130	TAS2 generating locus
Chr1:23420001-23420500	+	AT1G63150	TAS2 generating locus
Chr1:23490001-23490500	+	AT1G63330	PPR
Chr1:23507501-23509000	+	AT1G63400	PPR
Chr2:11721501-11722500	-	AT2G27400	TAS1a

Chr2:16537501-16538000	-	AT2G39675	TAS1c
Chr2:16539501-16540000	-	AT2G39681	TAS2
Chr3:5861491-5862437	+	AT3G17185	TAS3

Chapter 4. Discussion, Future Directions, and Conclusion

Discussion

In Chapter 2 we present analyses of small RNA deep sequencing datasets generated from the major cell types and developmental zones of the *A. thaliana* root. Using these data, we describe the expression profiles not only of known miRNAs, but also identify a substantial number of new mature miRNAs. We observed that many known and new miRNAs have developmental zone specificity, with a large number of known miRNAs enriched in the early maturation zone. This suggests that the miRNAs function to repress targets after the developmental program is initiated.

High resolution profiling of individual cell types can provide insights into the role of miRNAs that are masked by profiling entire organs (Figures 9-11). A known miRNA family, miR165/166, was recently shown to play a role in the specification of the xylem in roots (Carlsbecker et al., 2010), leading us to examine the expression pattern of these miRNAs in our datasets. Carlsbecker *et al.* found that these miRNAs are highly expressed in the endodermis and quiescent center with weaker expression in the cortex and epidermis. We found a similar pattern of miR165/166 expression in our datasets, but our cortex expression was higher than in the endodermis. This could be due to the expression of the cortex GFP marker used for sorting, which labels the cortex cells in the elongation to maturation zones. While previous studies focused on the meristematic zone, we have additional information about miR165/166 localization. Our data show that the highest levels of miR165/166 are in the early meristematic zone, with relative exclusion from the late maturation zone. This indicates that the action of this miRNA could be to set up a gradient of target expression necessary for proper pattern formation.

In addition to describing the root expression profile of currently known miRNA genes, including their variants, we developed a computational pipeline capable of identifying precursor sequences, which are then used in a probabilistic framework to accurately distinguish new miRNAs from other small ncRNAs. We identified 183 new miRNA candidates across all of our datasets. When we required consistency in classification score and expression across biological replicates, we arrived at a subset of 70 new mature miRNAs from 68 precursor sequences for which there is high confidence.

Reduction in activity for three new miRNA genes, novel33140, novel21861 and novel406654, resulted in significant differences in root growth rate when compared to wild-type controls. Reduction in activity for two of the miRNAs resulted in increased root growth, whereas reduction in the third miRNA resulted in reduced root growth. As has been previously reported, regulators that affect root growth are likely to have specific expression patterns across developmental zones (Tsukagoshi et al., 2010), and we see this for the two new miRNAs with enhanced root growth phenotypes. These two new miRNAs show enrichment in specific developmental zones, namely the two meristematic sections. Further analysis is needed to clarify the specific roles of these miRNAs in the regulation of root growth.

The ra-siRNA and the ta-siRNA analyses are not as developed as the analysis of miRNAs, making it more difficult to draw definite conclusions. It appears that there are genomic locations where methylation, small RNAs, and gene expression are correlated, and the thorough analysis of these relationships will have to be explored. Additionally, the ta-siRNA analysis indicated that there are new ta-siRNA loci, as well as new miRNA triggers on known ta-siRNA loci, all of which may be root specific. These findings need to be validated and characterized.

Future Directions

The most exciting results of the miRNA project are the identification of known miRNAs that are likely to be important in root development. The miR397 family showed none of the expected correlation between the miRNA expression and the validated targets, all laccases (LACs). LACs are a class of copper containing enzymes that have been implicated as having a role in lignin synthesis, cell wall integrity, and response to stress (Abdel-Ghany and Pilon, 2008). To further explore this, T-DNA insertion lines were ordered for both miR397 loci, and their validated targets. Additionally, 35S promoter overexpression lines were constructed for LAC17, LAC4 (also known as irregular xylem 12), and LAC2. Two of these LACs, namely LAC17 and

LAC4, seem to be regulated by the SHORTROOT transcription factor, according to microarray results (Levesque et al., 2006). The phenotyping and characterization of these lines could be interesting.

Additional work on the functions of the targets of novel21861 will need to be finished, including obtaining and phenotyping mutants in these genes. The promoter:GFP fusions for novel33140 and novel21861 should be completed as they should show longitudinal zone specificity and therefore be useful for the root map project.

Future directions for the ra-siRNA project include finishing the bisulfite sequencing for the longitudinal sections and the independent pCORTEX:GFP transformant. The full analysis of the correlations between small RNAs, methylation, and mRNA expression need to be completed, and this should reveal new insights. For example, which genes are differentially methylated in individual cell types? Do the small RNAs correlate with this as well? Are the small RNAs only 24 nt or other sizes also? Does this correspond to cell type specific changes in gene expression? Are there alternative splicing events associated? Can these changes explain cell type function?

The difference in columella methylation will need to be examined in the first longitudinal zone section. Is it really different? If so, why? If it is because these cells are more differentiated, then the maturation zone section should also show an increase in

CHH methylation. Are there other types of methylation that are specific to a cell type or a group of genes that are differentially regulated with alternative splicing or start sites? These new datasets, including the RNA-seq data, will be powerful tools to answer these types of questions.

Future directions for the ta-siRNA project are quite exciting. The observation that TAS1 could be triggered by novel miRs is interesting, and the validations to show this would be really incredible to do. The time from this discovery by Jiandong to the time I was writing/defending/leaving was just too short to take on this new project. The predictions of new ta-siRNA loci will also need to be validated, especially since they are most likely root specific.

Overall Conclusions

In conclusion, our small RNA deep sequencing datasets provide detailed expression profiles of classes of ncRNAs in the *A. thaliana* root. Expression patterns for known and novel miRNAs and ta-siRNAs have been examined, and the role of ncRNAs in methylation in the root have been explored. In summary, our study demonstrates the power of isolating individual cell types and developmental zones in combination with deep sequencing and computational analyses to obtain detailed profiles of ncRNAs, as well as to significantly extend the compendium of known functional RNAs.

Appendix A. Expression of known miRNAs.

mature	form	form_sequence	Stele_ave	Endo_ave	Cortex_ave	Epidermis_ave	Columella_ave	WR_sort_ave	LS1	LS2	LS3	LS4	WR_ave
ath-MIR156a	ath-MIR156a	TGACAGAAGAGAGTGAGCAC	299878	362779	1340441	790650	120403	136624	112825	191310	255877	214800	359092
ath-MIR156b	ath-MIR156b	TGACAGAAGAGAGTGAGCAC	300424	363233	1341355	791104	120627	137113	112889	191626	256252	215161	359208
ath-MIR156c	ath-MIR156c	TGACAGAAGAGAGTGAGCAC	299878	362779	1340441	790650	120403	136624	112825	191310	255877	214800	359092
ath-MIR156d	ath-MIR156d	TGACAGAAGAGAGTGAGCAC	300130	363048	1344005	793403	120909	136673	113108	191397	255938	214976	359600
ath-MIR156e	ath-MIR156e	TGACAGAAGAGAGTGAGCAC	299620	362506	1339829	790095	120341	136554	112605	191187	255794	214760	359087
ath-MIR156f	ath-MIR156f	TGACAGAAGAGAGTGAGCAC	299637	362510	1339832	790102	120342	136554	112606	191188	255795	214760	359089
ath-MIR156g	ath-MIR156g	CGACAGAAGAGAGTGAGCAC	1029	1325	2120	1940	420	316	0	0	1041	2354	1233
ath-MIR156h	ath-MIR156h	TGACAGAAGAAAGAGAGCAC	0	0	0	0	0	0	0	0	0	0	215
ath-MIR156h	XXX-MIR156hx1	TTGACAGAAGAAAGAGAGCAC	314	255	728	796	43	22	129	197	160	158	115
ath-MIR157a	ath-MIR157a	TTGACAGAAGATAGAGAGCAC	159121	180267	622745	704656	88565	59854	207188	346422	285322	274242	145954
ath-MIR157b	ath-MIR157b	TTGACAGAAGATAGAGAGCAC	159121	180267	622745	704656	88565	59854	207188	346422	285322	274242	145954
ath-MIR157c	ath-MIR157c	TTGACAGAAGATAGAGAGCAC	159172	180307	622119	704643	88608	59901	207219	345870	284677	273822	145862
ath-MIR157d	ath-MIR157d	TGACAGAAGATAGAGAGCAC	1557	2477	14915	18489	3620	1576	7063	8631	5541	7429	2636
ath-MIR158a	ath-MIR158a	TCCCAAATGTAGACAAAGCA	151036	172861	205535	124604	14104	11480	179526	51898	55538	39905	119176
ath-MIR158b	ath-MIR158b	CCCCAAATGTAGACAAAGCA	0	0	0	0	118	0	0	0	0	793	0
ath-MIR158b	XXX-MIR158bx1	TCCCCAAATGTAGACAAAGCA	1021	0	5371	0	0	0	0	0	662	0	0
ath-MIR159a	ath-MIR159a	TTTGGATTGAAGGGAGCTCTA	1067	631	325	397	0	0	0	0	0	0	464
ath-MIR159b	ath-MIR159b	TTTGGATTGAAGGGAGCTCTT	175	56	0	0	0	0	0	0	0	0	40
ath-MIR160a	ath-MIR160a	TGCCTGGCTCCCTGTATGCCA	0	72	153	83	498	717	227	142	96	276	93
ath-MIR160a	XXX-MIR160ax1	ATGCCTGGCTCCCTGTATGCC	137	38	0	0	0	0	0	0	0	0	47
ath-MIR160b	ath-MIR160b	TGCCTGGCTCCCTGTATGCCA	43	58	126	70	342	463	222	126	85	219	72
ath-MIR160c	ath-MIR160c	TGCCTGGCTCCCTGTATGCCA	0	72	153	83	495	713	227	142	96	0	92

ath-MIR160c	XXX-MIR160cx2	ATGCCTGGCTCCCTGTATGCC	137	38	0	0	0	0	0	0	0	0	49
ath-MIR161.1	XXX-MIR161.2x1	TTGAAAGTGACTACATCGGGG	6296	3939	11893	43341	22958	30333	54318	24814	16924	12053	4345
ath-MIR161.2	ath-MIR161.2	TCAATGCATTGAAAGTGACTA	0	1770	0	0	0	0	0	0	0	0	6955
ath-MIR162a	ath-MIR162a	TCGATAAACCTCTGCATCCAG	349	233	409	320	334	252	1481	983	974	655	213
ath-MIR162b	ath-MIR162b	TCGATAAACCTCTGCATCCAG	349	233	409	320	334	252	1481	983	975	655	213
ath-MIR163	ath-MIR163	TTGAAGAGGACTTGGAACTTCGAT	179	31	77	73	0	0	83	0	31	0	359
ath-MIR164a	ath-MIR164a	TGGAGAAGCAGGGCACGTGCA	1198	1141	495	390	941	1586	145	1361	1559	4883	2448
ath-MIR164b	ath-MIR164b	TGGAGAAGCAGGGCACGTGCA	1223	1141	499	395	950	1611	146	1373	1574	4937	2469
ath-MIR164c	ath-MIR164c	TGGAGAAGCAGGGCACGTGCG	125	100	133	63	42	96	8	87	114	883	582
ath-MIR165a	ath-MIR165a	TCGGACCAGGCTTCATCCCCC	10019	18467	58909	2425	4153	7830	9016	2478	3272	5794	3584
ath-MIR165b	ath-MIR165b	TCGGACCAGGCTTCATCCCCC	9813	18219	58480	2339	4037	7802	8873	2441	3226	5706	3539
ath-MIR166a	ath-MIR166a	TCGGACCAGGCTTCATTCCCC	63021	76026	319789	19743	24915	34547	20697	6492	11973	13665	12086
ath-MIR166b	ath-MIR166b	TCGGACCAGGCTTCATTCCCC	59845	71166	312994	18986	23946	33175	19075	6128	11377	12863	11612
ath-MIR166c	ath-MIR166c	TCGGACCAGGCTTCATTCCCC	59922	71233	313396	19010	23975	33215	19129	6130	11384	12869	11614
ath-MIR166d	ath-MIR166d	TCGGACCAGGCTTCATTCCCC	59874	71214	313142	18998	23943	33177	19128	6130	11382	12870	11621
ath-MIR166e	ath-MIR166e	TCGGACCAGGCTTCATTCCCC	59952	71261	313373	19002	23976	33261	19082	6130	11384	12862	11617
ath-MIR166f	ath-MIR166f	TCGGACCAGGCTTCATTCCCC	59922	71261	313347	18992	23975	33260	19082	6130	11384	12862	11617
ath-MIR166g	ath-MIR166g	TCGGACCAGGCTTCATTCCCC	59981	71309	313522	19013	23974	33263	19135	6133	11389	12869	11626
ath-MIR167a	ath-MIR167a	TGAAGCTGCCAGCATGATCTA	371	1207	1775	921	9982	14446	401	1972	3060	5493	847
ath-MIR167b	ath-MIR167b	TGAAGCTGCCAGCATGATCTA	371	1207	1779	921	10000	14470	393	1971	3062	5501	846
ath-MIR167c	ath-MIR167c	TAAGCTGCCAGCATGATCTTG	0	0	0	0	0	0	0	0	2381	2681	0
ath-MIR167c	XXX-MIR167cx2	TAAGCTGCCAGCATGATCTTGT	278	326	625	369	2089	3542	317	577	0	0	76
ath-MIR167c	XXX-MIR167cx4	TTAAGCTGCCAGCATGATCTT	0	0	0	0	0	0	0	0	0	0	62
ath-MIR167d	ath-MIR167d	TGAAGCTGCCAGCATGATCTGG	0	0	0	0	0	0	16	0	0	0	0
ath-MIR167d	XXX-MIR167dx3	TGAAGCTGCCAGCATGATCTG	0	0	60	21	198	280	0	64	81	191	0
ath-MIR168a	ath-MIR168a	TCGCTTGGTGCAGGTCGGGAA	30918	9300	32029	76679	23484	17714	24147	26580	17371	12603	19179

ath-MIR168b	ath-MIR168b	TCGCTTGGTGCAGGTCGGGAA	30918	9298	31976	76622	23367	17644	24086	26527	17287	12544	19174
ath-MIR169a	ath-MIR169a	CAGCCAAGGATGACTTGCCGA	908	1185	1455	1140	1408	1078	244	381	492	463	1914
ath-MIR169b	ath-MIR169b	CAGCCAAGGATGACTTGCCGG	197	259	317	219	153	142	20	84	118	127	466
ath-MIR169c	ath-MIR169c	CAGCCAAGGATGACTTGCCGG	185	261	316	223	126	118	16	71	119	119	420
ath-MIR169d	ath-MIR169d	TGAGCCAAGGATGACTTGCCG	637	742	788	310	1090	1269	66	256	590	514	1408
ath-MIR169e	ath-MIR169e	TGAGCCAAGGATGACTTGCCG	637	742	788	310	1090	1269	66	256	590	514	1408
ath-MIR169f	ath-MIR169f	TGAGCCAAGGATGACTTGCCG	642	725	774	315	1066	1258	63	245	578	510	1394
ath-MIR169g	ath-MIR169g	TGAGCCAAGGATGACTTGCCG	635	719	769	302	1062	1250	62	243	576	506	1381
ath-MIR169g*	XXX-MIR169gx1	GCAAGTTGACCTTGGCTCTGT	237	258	338	98	0	0	0	0	0	0	159
ath-MIR169h	ath-MIR169h	TAGCCAAGGATGACTTGCCTG	135	143	179	100	92	86	42	92	153	118	493
ath-MIR169i	ath-MIR169i	TAGCCAAGGATGACTTGCCTG	0	0	204	94	0	0	0	0	0	0	277
ath-MIR169i	XXX-MIR169ix2	TAGCCAAGGATGACTTGCCTGA	275	314	106	95	248	243	91	187	375	299	697
ath-MIR169j	ath-MIR169j	TAGCCAAGGATGACTTGCCTG	0	0	203	95	0	0	0	0	0	0	277
ath-MIR169j	XXX-MIR169jx2	TAGCCAAGGATGACTTGCCTGA	270	314	105	94	239	234	91	186	374	297	697
ath-MIR169k	ath-MIR169k	TAGCCAAGGATGACTTGCCTG	138	143	179	100	92	86	42	92	154	118	493
ath-MIR169l	ath-MIR169l	TAGCCAAGGATGACTTGCCTG	0	0	202	94	0	0	0	0	0	0	277
ath-MIR169l	XXX-MIR169lx2	TAGCCAAGGATGACTTGCCTGA	288	314	105	93	241	238	90	186	373	296	700
ath-MIR169m	ath-MIR169m	TAGCCAAGGATGACTTGCCTG	132	141	182	103	96	91	43	96	158	123	485
ath-MIR169n	ath-MIR169n	TAGCCAAGGATGACTTGCCTG	0	0	203	95	0	0	0	0	0	0	277
ath-MIR169n	XXX-MIR169nx2	TAGCCAAGGATGACTTGCCTGA	270	314	105	94	239	234	91	186	374	297	697
ath-MIR171a	ath-MIR171a	TGATTGAGCCGCGCCAATATC	0	0	0	0	0	0	0	2	0	0	0
ath-MIR171b	ath-MIR171b	TTGAGCCGTGCCAATATCACG	0	0	0	0	0	0	0	0	0	19	0
ath-MIR171c	ath-MIR171c	TTGAGCCGTGCCAATATCACG	0	0	0	0	0	0	0	0	0	19	0
ath-MIR172a	ath-MIR172a	AGAATCTTGATGATGCTGCAT	989	520	675	412	850	552	282	441	300	312	1339
ath-MIR172b	ath-MIR172b	AGAATCTTGATGATGCTGCAT	989	520	675	412	850	552	282	441	300	312	1339
ath-MIR172b*	ath-MIR172b*	GCAGCACCATTAAGATTCAC	0	0	0	0	0	0	0	0	0	0	0

ath-MIR172b*	XXX-MIR172b*x1	GCAGCACCATTAAGATTCACA	0	0	36	115	162	146	51	226	264	166	0
ath-MIR172c	ath-MIR172c	AGAATCTTGATGATGCTGCAG	232	110	79	241	1491	433	652	84	28	22	296
ath-MIR172d	ath-MIR172d	AGAATCTTGATGATGCTGCAG	232	110	79	241	1491	433	647	84	28	22	296
ath-MIR172e	ath-MIR172e	GGAATCTTGATGATGCTGCAT	348	196	259	161	214	251	68	116	117	102	342
ath-MIR173	ath-MIR173	TTCGCTTGCAGAGAGAAATCAC	11131	6612	6545	13179	5883	5452	7679	7359	2530	1896	8325
ath-MIR1886.2	ath-MIR1886.2	TGAGATGAAATCTTTGATTGG	13	0	0	11	74	67	148	51	53	43	12
ath-MIR2111a	ath-MIR2111a	TAATCTGCATCCTGAGGTTTA	0	0	0	8	18	27	0	2	5	4	0
ath-MIR2111a*	ath-MIR2111a*	GTCCTCGGGATGCGGATTACC	0	0	67	0	24	41	0	0	2	2	0
ath-MIR2111b	ath-MIR2111b	TAATCTGCATCCTGAGGTTTA	0	0	0	8	18	27	0	2	5	4	0
ath-MIR2111b*	ath-MIR2111b*	ATCCTCGGGATACAGTTTACC	0	0	0	0	4	9	0	1	2	1	0
ath-MIR2112-3p	ath-MIR2112-3p	CTTTATATCCGCATTTGCGCA	0	0	0	0	0	0	0	0	1	0	0
ath-MIR319a	ath-MIR319a	TTGGACTGAAGGGAGCTCCCT	57	0	0	0	0	0	0	0	0	0	95
ath-MIR319a	XXX-MIR319ax5	TTGGACTGAAGGGAGCTCCC	0	0	0	0	0	0	0	0	0	0	0
ath-MIR319b	ath-MIR319b	TTGGACTGAAGGGAGCTCCCT	59	0	0	0	0	0	279	68	0	0	95
ath-MIR319b	XXX-MIR319bx3	TTGGACTGAAGGGAGCTCCC	0	0	0	0	0	0	0	0	0	0	0
ath-MIR3434	ath-MIR3434	ACTTGGCTGATTCTATTATT	0	0	0	0	0	0	10	3	2	2	0
ath-MIR3440b-3p	ath-MIR3440b-3p	TGGATTGGTCAAGGGAAGCGT	169	116	111	382	101	105	297	228	41	20	73
ath-MIR390a	ath-MIR390a	AAGCTCAGGAGGGATAGCGCC	559	573	984	10122	7636	3452	8375	4520	2755	2347	616
ath-MIR390b	ath-MIR390b	AAGCTCAGGAGGGATAGCGCC	559	573	984	10123	7642	3452	8376	4520	2755	2347	616
ath-MIR391	ath-MIR391	TTCGCAGGAGAGATAGCGCCA	0	0	0	0	5	5	0	20	26	20	0
ath-MIR393a	ath-MIR393a	TCCAAAGGGATCGCATTGATCC	0	0	0	0	25	6	27	11	4	6	0
ath-MIR393b	ath-MIR393b	TCCAAAGGGATCGCATTGATCC	0	0	0	0	25	6	27	11	4	6	0
ath-MIR394a	ath-MIR394a	TTGGCATTCTGTCCACCTCC	65	0	0	0	20	37	42	53	22	29	97
ath-MIR394b	ath-MIR394b	TTGGCATTCTGTCCACCTCC	65	0	0	0	21	38	42	53	22	29	97
ath-MIR395b	ath-MIR395b	CTGAAGTGTTTGGGGGGACTC	0	0	36	14	11	29	16	38	104	39	0
ath-MIR395c	ath-MIR395c	CTGAAGTGTTTGGGGGGACTC	0	0	36	14	11	29	16	38	104	39	0

ath-MIR395f	ath-MIR395f	CTGAAGTGTTTGGGGGGACTC	0	0	36	14	11	29	16	38	104	39	0
ath-MIR396a	ath-MIR396a	TTCCACAGCTTTCTTGAACTG	160	210	124	26	259	141	47	0	0	0	264
ath-MIR396b	ath-MIR396b	TTCCACAGCTTTCTTGAACTT	21	38	62	0	0	0	0	0	0	0	0
ath-MIR397a	ath-MIR397a	TCATTGAGTGCAGCGTTGATG	0	0	0	0	14	22	0	0	0	0	0
ath-MIR397b	ath-MIR397b	TCATTGAGTGCATCGTTGATG	0	0	0	0	45	53	0	0	6	12	0
ath-MIR399b	ath-MIR399b	TGCCAAAGGAGAGTTGCCCTG	0	33	229	11	124	152	7	27	42	28	0
ath-MIR399b	XXX-MIR399bx1	CCTGCCAAAGGAGAGTTGCCC	0	0	0	55	0	0	0	0	0	0	0
ath-MIR399c	ath-MIR399c	TGCCAAAGGAGAGTTGCCCTG	0	27	189	35	108	138	7	27	41	25	0
ath-MIR399d	ath-MIR399d	TGCCAAAGGAGATTTGCCCCG	0	0	17	0	3	3	0	0	0	0	0
ath-MIR399f	ath-MIR399f	TGCCAAAGGAGATTTGCCCGG	0	23	115	26	22	23	0	0	0	1	0
ath-MIR400	ath-MIR400	TATGAGAGTATTATAAGTCAC	52	0	0	26	41	21	131	66	38	25	60
ath-MIR403	ath-MIR403	TTAGATTCACGCACAAACTCG	187	133	197	215	381	618	0	0	0	0	151
ath-MIR408	ath-MIR408	ATGCACTGCCTCTTCCCTGGC	239	167	212	286	421	616	3	14	24	47	0
ath-MIR447a	ath-MIR447a	TTGGGGACGAGATGTTTTGTTG	0	0	0	13	60	75	125	0	73	72	0
ath-MIR447b	ath-MIR447b	TTGGGGACGAGATGTTTTGTTG	0	0	0	13	60	75	125	0	73	72	0
ath-MIR472	ath-MIR472	TTTTTCCTACTCCGCCCATACC	0	0	0	0	6	18	0	0	5	0	0
ath-MIR773	XXX-MIR773x1	TTTGCTTCCAGCTTTTGTCTCC	35	29	138	114	195	250	83	65	59	40	0
ath-MIR775	ath-MIR775	TTCGATGTCTAGCAGTGCCA	74	182	238	671	584	884	648	456	191	211	165
ath-MIR777	ath-MIR777	TACGCATTGAGTTTCGTTGCTT	0	0	0	0	37	41	15	16	13	10	0
ath-MIR779.2	ath-MIR779.2	TGATTGGAAATTTCGTTGACT	0	0	0	0	5	10	36	8	9	7	0
ath-MIR780.2	ath-MIR780.2	TTCTTCGTGAATATCTGGCAT	0	0	0	0	0	4	0	7	18	19	9
ath-MIR781	ath-MIR781	TTAGAGTTTTCTGGATACTTA	0	0	0	0	0	0	10	4	6	5	0
ath-MIR822	ath-MIR822	TGCGGGAAGCATTTGCACATG	2767	2311	6433	10631	4947	7358	3978	5454	8400	7523	1709
ath-MIR823	ath-MIR823	TGGGTGGTGATCATATAAGAT	0	31	119	193	185	299	145	213	106	105	15
ath-MIR824	ath-MIR824	TAGACCATTTGTGAGAAGGGA	189	169	203	312	511	166	757	366	150	156	555
ath-MIR827	ath-MIR827	TTAGATGACCATCAACAAACT	0	0	0	0	0	0	0	0	0	1	0

ath-MIR829.1	ath-MIR829.1	AGCTCTGATACCAAATGATGGAAT	20896	7554	6563	6942	2173	2372	811	3076	1694	632	2048
ath-MIR829.2	ath-MIR829.2	CAAATTAAAGCTTCAAGGTAG	0	0	11	27	11	6	0	0	0	0	0
ath-MIR830	ath-MIR830	TAACTATTTTGAGAAGAAGTG	0	0	0	0	0	0	0	2	2	0	11
ath-MIR834	ath-MIR834	TGGTAGCAGTAGCGGTGGTAA	0	0	0	0	0	0	0	0	0	1	0
ath-MIR835-3p	ath-MIR835-3p	TGGAGAAGATACGCAAGAAAG	0	0	0	0	0	0	3	4	0	0	0
ath-MIR837-3p	ath-MIR837-3p	AAACGAACAAAAAACTGATGG	731	432	95	0	10	28	53	70	252	459	424
ath-MIR837-5p	ath-MIR837-5p	ATCAGTTTCTTGTTCGTTTCA	0	0	0	0	8	13	0	0	4	8	0
ath-MIR839	ath-MIR839	TACCAACCTTTCATCGTTCCC	0	0	0	0	5	5	0	0	7	9	0
ath-MIR840	ath-MIR840	ACACTGAAGGACCTAAACTAAC	0	0	0	0	0	0	7	0	0	0	0
ath-MIR841b	ath-MIR841b	TACGAGCCACTGGAAACTGAA	0	0	0	0	0	0	13	0	8	8	0
ath-MIR842	ath-MIR842	TCATGGTCAGATCCGTCATCC	0	0	147	68	94	102	49	96	93	93	69
ath-MIR842	XXX-MIR842x1	CATGGTCAGATCCGTCATCCC	35	0	0	0	0	0	0	0	0	0	25
ath-MIR843	ath-MIR843	TTTAGGTCGAGCTTCATTGGA	0	0	0	0	0	0	4	0	0	0	0
ath-MIR843	XXX-MIR843x1	TTAGGTCGAGCTTCATTGGAA	0	0	0	0	0	0	0	0	0	2	0
ath-MIR844*	ath-MIR844*	TTATAAGCCATCTTACTAGTT	0	0	0	0	0	0	0	1	0	0	0
ath-MIR845a	ath-MIR845a	CGGCTCTGATACCAATTGATG	0	0	0	0	0	0	3	0	0	1	0
ath-MIR846	ath-MIR846	TTGAATTGAAGTGCTTGAATT	1086	693	1147	544	0	0	0	0	0	0	538
ath-MIR846	XXX-MIR846x4	TTTGAATTGAAGTGCTTGAAT	0	0	188	603	1439	1335	751	1142	1399	751	0
ath-MIR848	ath-MIR848	TGACATGGGACTGCCTAAGCTA	59	0	0	0	0	0	0	0	0	0	0
ath-MIR852	ath-MIR852	AAGATAAGCGCCTTAGTTCTG	0	0	23	0	0	0	0	0	0	0	0
ath-MIR852	XXX-MIR852x1	AGATAAGCGCCTTAGTTCTGA	0	0	0	0	62	48	0	31	49	46	0
ath-MIR857	ath-MIR857	TTTTGTATGTTGAAGGTGTAT	0	0	0	0	0	0	0	0	2	2	0
ath-MIR858	ath-MIR858	TTTCGTTGTCTGTTCGACCTT	0	0	0	0	0	0	7	3	4	0	0
ath-MIR858	XXX-MIR858x1	TTCGTTGTCTGTTCGACCTTG	0	0	0	0	0	0	0	0	0	6	0
ath-MIR859	ath-MIR859	TCTCTCTGTTGTGAAGTCAAA	0	0	0	0	0	0	3	3	2	1	0
ath-MIR860	ath-MIR860	TCAATAGATTGGACTATGTAT	0	0	0	0	7	7	187	54	115	99	22

ath-MIR862-5p	ath-MIR862-5p	TCCAATAGGTCGAGCATGTGC	0	0	0	0	0	0	5	4	2	1	0
ath-MIR863-3p	XXX-MIR863-3px1	TGCGATTGAGAGCAACAAGAC	34	0	31	13	27	0	68	32	31	28	23
ath-MIR864-3p	ath-MIR864-3p	TAAAGTCAATAATACCTTGAAG	0	0	0	0	0	0	3	0	0	0	0
ath-MIR864-5p	ath-MIR864-5p	TCAGGTATGATTGACTTCAAA	0	0	0	0	4	12	39	10	9	7	7
ath-MIR869.2	ath-MIR869.2	TCTGGTGTTGAGATAGTTGAC	0	0	0	0	24	60	6	10	5	197	134

Appendix B. Expression of all potential novel miRNAs.

mature	form	form_sequence	Stele_ave	Endo_ave	Cortex_ave	Epidermis_ave	Columella_ave	WR_sort_ave	LS1	LS2	LS3	LS4	WR_ave
novel10199	**novel10199**	TGACAGAAGAGAGAGAGCAC	5931	5870	4239	5922	342	451	1166	1308	1171	989	3465
novel10539	novel10539	TCGTGAACACAAGCACTGACC	0	0	0	0	0	0	16	16	2	2	24
novel106029	novel106029	TACGGAAATCGTTAACGGCGTTA	0	0	0	0	99	25	16	0	4	0	0
novel109417	novel109417	GACGTGACAGTCGATAACTCA	0	0	0	0	0	0	0	0	0	0	0
novel111346	novel111346	AGGCTTTTAAGATCTGGTTGC	0	0	0	13	76	46	38	10	28	12	0
novel111695	novel111695	AGTGGGACAACAATGGAGATG	0	0	0	0	0	0	0	0	0	0	0
novel116556	novel116556	AGGCTTTTAAGATCTGGTTGC	0	0	0	14	93	48	38	12	28	12	0
novel120197	novel120197	ATTAGGATCGGAATCAACGG	0	0	57	0	0	0	0	0	0	0	0
novel122522	novel122522	TCAAGTTTGATGACGATTCCA	0	0	0	0	7	7	31	15	8	6	0
novel124604	novel124604	TACTATTTGAATCGTACTGC	0	0	0	0	0	14	0	0	0	0	0
novel124881	novel124881	TTTGGGTGAAGGAGTGTGTGGTG	0	0	0	6	210	57	32	8	20	8	0
novel127485	novel127485	TGGCAGGAAAGACATAATTTT	0	0	0	6	0	0	5	8	2	4	0
novel137096	novel137096	TAAGCGATTTTAGGCGGTTAGTG	0	0	0	0	24	0	0	0	0	5	0
novel15045	novel15045	ATACCAAAACTCTCTCACTTT	0	0	0	0	0	0	0	9	5	2	0
novel160168	novel160168	TGTCTTCGGACGATGGCGTGGTG	0	0	0	0	20	10	0	0	0	0	0
novel161368	novel161368	TTCGGTCATTTTAGGCATTGTCG	0	0	0	0	144	18	0	0	0	2	0
novel16336	novel16336	TGAGTAGGATCTTAAGTGGTATA	273	193	105	139	500	118	115	81	77	42	63
novel16394	novel16394	GCCGGTTCTGTTGTTGGCATG	0	0	59	0	52	35	0	0	0	0	0
novel16394	novel45692	CCGGTTCTGTTGTTGGCATG	0	0	133	0	0	0	0	0	0	0	0
novel172045	novel172045	ACAGTGGTCATCTGGTGGGCT	0	0	0	0	5	5	5	11	7	6	0
novel180251	novel180251	TTGGCCCATGTCGTTATGATT	0	0	0	0	0	0	0	2	1	1	0
novel18190	novel18190	AGAGGTACATCATGTAGTCTG	0	0	0	0	0	0	14	5	5	3	11

novel182421	**novel182421**	TGTTTGTTGACATCGGTCTAG	0	0	0	0	7	13	5	2	4	8	0
novel182589	novel182589	GTAGACGCGGTAGGTTATTCA	0	0	0	0	0	0	0	0	5	2	0
novel18522	novel18522	TGCAGATCGGCTCCTGGTTCC	0	0	0	12	26	20	8	4	4	6	0
novel18686	novel18686	AGGGACTTTGTGAATTTAGGG	0	0	0	0	0	0	0	0	9	3	0
novel194944	novel194944	GGTGTAAGTGACTGTGTTGCC	0	0	0	0	0	0	0	3	0	0	0
novel194949	novel194949	TGTTTGAGTTGATGATTAGCA	0	0	0	0	0	4	0	0	0	2	0
novel195865	**novel195865**	TTTTACTGCTACTTGTGTTCC	0	0	0	0	0	0	4	2	2	1	0
novel198387	novel198387	TTAGGCTAAGATTTGTGAAGA	0	0	0	0	0	0	35	0	9	6	0
novel201442	novel201442	TTAGGGTAGTTAACGGAAGTTA	0	0	0	0	0	10	10	4	5	7	0
novel205213	novel205213	ATCATTATTGGTTGTCTCTTA	0	0	0	0	0	0	6	0	2	2	0
novel206489	novel206489	CGTGGTGTGATCTGGTTGGTG	0	0	0	0	0	0	0	2	0	2	0
novel207342	novel207342	TGAGAGAAGGAATTAGATTCA	0	0	0	0	0	0	6	0	4	7	0
novel207342	novel30631	ATGAGAGAAGGAATTAGATTC	0	0	0	0	0	0	0	0	0	0	10
novel207384	novel207384	CAAGCTGAGAATCGGTCCGAC	0	0	0	0	0	0	0	1	0	1	0
novel20766	novel20766	TGTTAAGGAGTGTTAACGGTG	175	212	48	0	123	235	0	0	0	0	180
novel20824	novel20824	TGGGACTAGCTGGAGACAATA	0	0	0	17	0	0	8	0	0	2	0
novel208717	novel208717	ATAAGTCAAGTAGTAGTCCTT	0	0	0	0	0	0	4	0	0	0	0
novel20963	novel20963	TTGTTTCCTGTAGAAAGCATT	19	0	23	0	0	0	3	9	12	6	53
novel20963	novel29474	TTTCCTGTAGAAAGCATTTGC	0	40	0	46	0	0	0	0	0	0	0
novel214285	**novel214285**	CCGTATCTTGGCCTTGTCATT	0	0	0	0	0	0	0	0	3	5	0
novel214456	**novel214456**	TACTGTATATATGTAAGTGAC	0	0	0	0	0	0	2	0	0	1	0
novel214827	novel214827	AGGGATCGAATAGCAATGCAT	0	0	0	0	0	0	4	0	0	0	0
novel21507	novel21507	TGTTAAGGAGTGTTAACGGTG	172	200	47	55	120	220	0	33	0	0	177
novel216165	novel216165	CAAATTCAAAATCGGTTTCTA	0	0	0	0	0	0	3	0	2	0	0
novel21693	novel21693	TGGAAGAAGATGATAGAATTA	0	0	0	0	0	0	7	5	3	2	0
novel217147	novel217147	TAGACCGAGTACAACAAACAA	0	0	0	0	0	0	0	2	4	5	0

novel21861	novel21861	TGGACAAGGTTAGATTTGGTG	225	263	198	427	193	100	82	56	22	9	65
novel219738	novel219738	AAAGGTCAAAATTTGGAGTCA	0	0	0	0	0	0	6	4	8	8	0
novel222429	novel222429	ATGGGACATCGAGCATTTAAT	0	0	0	0	0	0	13	4	0	2	0
novel222663	novel222663	ATTTAGTTCGTGTAGTTGGTT	0	0	0	0	0	0	21	5	9	0	0
novel227820	novel227820	AAAGGATTTACAAGGGTTTTA	0	0	0	0	0	0	3	0	0	0	0
novel232253	novel232253	TTGGATTTATAGTTGGATAAG	0	0	0	0	0	0	6	0	0	0	0
novel233188	novel233188	AATGCGCAACTCTATATTTCC	0	0	0	0	0	0	7	6	1	2	0
novel234117	novel234117	TGAAGTCAGGGATGTTAAGCA	0	0	0	0	0	0	10	0	3	0	0
novel234172	novel234172	TCAAAGATTAAGCTATGGAGC	0	0	0	0	0	0	0	0	0	1	0
novel23586	novel23586	TCGCTACGTTCTTTCTGATTC	101	0	0	0	0	0	0	0	0	0	0
novel24173	novel24173	CGTAGTTGCAGAGCTTGACGG	20	0	0	48	51	23	0	12	4	0	19
novel24186	novel24186	TAGCTAAGGATTTGCATTCTC	38	40	0	0	0	0	0	0	0	0	0
novel24186	novel35393	AGCTAAGGATTTGCATTCTCA	65	0	42	36	77	44	97	34	58	46	69
novel241998	novel241998	CAAGAGAAGAAGTTATAGATA	0	0	0	0	0	0	3	1	0	0	0
novel242609	novel242609	TCAAAGATTAAGCTATGGAGC	0	0	0	0	0	0	0	0	0	1	0
novel247457	novel247457	TACATCACTGTACAGTTGCT	0	0	0	0	0	0	3	0	0	0	0
novel254239	novel254239	AATTGTCAGATCGTTAGTGCA	0	0	0	0	0	0	0	3	0	1	0
novel257605	novel257605	TTCCACTAAGTCGCGTTACC	0	0	0	0	0	0	0	0	5	5	0
novel257609	novel257609	ATGAGGTTAGGTAGCAAACAT	0	0	0	0	0	0	28	0	0	3	0
novel261634	novel261634	CAGGTGGTTAGTGCAATGGAA	0	0	0	0	0	0	15	0	0	0	0
novel26231	novel26231	AGGCCAAGGAAGTTTGAGGCA	9355	2901	7143	4653	1171	2672	0	0	0	0	225
novel264653	novel264653	ATGATGATGATGATGATGAAA	0	0	0	0	0	0	13	0	0	0	0
novel265478	**novel265478**	TACTGTATATATGTAAGTGAC	0	0	0	0	0	0	2	0	0	1	0
novel265779	novel265779	TGAAGTCAGGGATGTTAAGCA	0	0	0	0	0	0	10	2	3	0	0
novel265916	novel265916	TGAAGTCAGGGATGTTAAGCA	0	0	0	0	0	0	10	0	3	0	0
novel265917	novel265917	TGAAGTCAGGGATGTTAAGCA	0	0	0	0	0	0	8	0	3	0	0

novel267146	novel267146	TAAGACGAGTGGCCTAGGCT	0	0	0	0	0	0	21	15	0	0	0
novel272477	novel272477	TGGGAGCTGAATGAACAGCGG	0	0	0	0	0	0	0	0	0	1	0
novel272530	novel272530	TAATATTGGTTAATCATCTCG	0	0	0	0	0	0	0	0	2	0	0
novel27285	**novel27285**	ACTCATAAGATCGTGACACGT	169	114	241	246	252	83	464	154	184	120	257
novel27349	novel121550	TGAAGTAGAGATTGGGTTTCC	0	0	0	33	0	24	0	0	0	0	0
novel27349	novel27349	ACTGAAGTAGAGATTGGGTTT	0	48	0	0	37	11	35	64	31	21	26
novel273671	novel273671	TCACCTCAGACAGTGTAGGCT	0	0	0	0	0	0	4	0	0	0	0
novel276003	novel276003	CAAGACGAGTGGCCTAGGCT	0	0	0	0	0	0	18	15	0	0	0
novel276680	novel276680	TATATAAGCCAAGTCCAACT	0	0	0	0	0	0	11	0	10	0	0
novel27783	novel27783	AGAATGCAAATCCTTAGCTAT	0	0	0	13	12	4	4	9	8	3	0
novel27882	novel27882	TCTCGCGCTTGTACGGCTTT	0	65	0	0	0	0	605	1076	0	0	78
novel27909	novel27909	TCCGGGCTAGAAGCGACGCA	0	6249	0	0	0	0	0	0	0	0	0
novel280991	novel280991	TTCGGAGAACGAAGAAGAAGA	0	0	0	0	0	0	2	0	0	0	0
novel284645	novel284645	AAGAAGAAGAAGAAGAAGCAA	0	0	0	0	0	0	0	2	0	1	0
novel28598	novel28598	TCTCGCGCTTGTACGGCTTT	0	67	0	0	0	0	616	0	0	0	80
novel28617	novel28617	TCCGGGCTAGAAGCGACGCA	0	6249	0	0	0	0	0	0	0	0	0
novel28889	novel28889	CCATCCCCACAGACGGCGCCA	0	48	31	0	0	0	0	0	0	0	0
novel29060	novel29060	TGTTAAGGAGTGTTAACGGTG	0	94	0	0	124	124	0	0	0	0	179
novel291414	novel291414	TGACAAGGCCAAGATATAACA	0	0	0	0	0	0	0	6	5	4	0
novel29385	**novel29385**	TGGAAGAAGGTGAGACTTGCA	0	0	0	0	0	0	0	13	20	12	0
novel294059	novel294059	TGAGGTTGAAGCTGGTGATGA	0	0	0	0	0	0	0	2	0	0	0
novel29551	novel29551	AGAATGCAAATCCTTAGCTAT	0	0	0	13	12	4	4	9	8	3	0
novel296095	novel296095	ACAAGGGATTTTAAAATGGTT	0	0	0	0	0	0	0	5	0	0	0
novel296113	novel296113	AACCTACTTAACGACATGACT	0	0	0	0	0	0	0	6	0	0	0
novel297345	novel297345	TGGCATGGAAGAAGGTGAGAC	0	0	0	0	0	0	0	11	12	4	0
novel300613	novel300613	AAAGAAAGAGAGAGAGAGAG	0	0	0	0	0	0	0	0	1	0	0

novel30125	novel30125	GATGAAGGTCTTTGGAACGGTA	0	73	0	0	0	0	0	0	0	0	12
novel30125	novel54578	CGATGAAGGTCTTTGGAACGGTA	67	132	71	146	2122	461	290	72	60	29	53
novel304515	novel304515	GAAAGAGGTTCACCCCTAGG	0	0	0	0	0	0	0	6	0	2	0
novel30517	novel30517	ACACTTAGTTTTGTACAACAT	0	0	0	0	0	0	0	0	4	4	10
novel30518	**novel30518**	TTGTACAAATTTAAGTGTACG	0	0	0	0	0	0	0	0	0	0	10
novel306276	novel306276	AAAAACCGCTAGACATCGCT	0	0	0	0	0	0	0	2	0	0	0
novel306507	novel306507	ATAATTGACTATCAAAAAGAG	0	0	0	0	0	0	0	3	0	0	0
novel309333	novel309333	AAAGAAAGAGAGAGAGAGAG	0	0	0	0	0	0	0	0	2	0	0
novel31114	novel31114	TTAGAAGGATTTTTAACGGTG	0	0	0	0	0	0	0	0	0	0	10
novel31386	novel31386	ATGACGGTCAAGACCTCGTCC	0	0	0	0	0	0	0	0	0	0	74
novel314145	novel238846	AACCTATTTAACGACATGACT	0	0	0	0	0	0	14	0	0	2	0
novel314145	novel314145	AACCTATTTAACGACATGAC	0	0	0	0	0	0	0	3	7	0	0
novel315080	novel315080	ATCAAGTACGGTGCGGTTCTA	0	0	0	0	0	0	0	4	0	8	0
novel316512	novel316512	ATTTGAGTCATGTCGTTAAG	0	0	0	0	0	0	0	27	0	0	0
novel316536	novel316536	TTAAGATTTGGCTAAGATACC	0	0	0	0	0	0	0	3	6	0	0
novel316881	novel316881	TATGATCATCAGAAAACAGTG	0	0	0	0	0	0	0	3	2	4	0
novel317020	novel317020	ACAAGGGATTTTAAAATGGTT	0	0	0	0	0	0	0	5	0	0	0
novel317292	novel317292	AGATTTGGCTAAGATACCACT	0	0	0	0	0	0	0	6	0	0	0
novel317920	novel317920	TCTTATGTTAAGAGATATGTC	0	0	0	0	0	0	0	2	0	0	0
novel31892	**novel31892**	TGTTTTGGATCTTAGATACAC	0	0	0	0	0	0	0	0	0	0	19
novel31892	novel407358	TTGTTTTGGATCTTAGATACA	0	0	0	0	0	0	0	0	0	9	0
novel32263	novel32263	TGGAAGATGCTTTGGGATTTATT	0	0	0	41	25	5	41	14	16	8	23
novel32468	novel32468	TCAAGGACAGATTGGAGTTAA	0	0	89	125	0	69	292	161	121	78	73
novel32763	novel32763	ATGACGGTCAAGACCTCGTCC	0	0	0	0	0	0	0	0	0	0	74
novel327967	novel327967	ATTTGAGTCATGTCGTTAAG	0	0	0	0	0	0	0	27	0	0	0
novel328180	novel328180	AGAGGTGACCATTGGAGATG	0	0	0	0	0	0	0	8	0	5	0

novel328342	novel328342	ATTTGAGTCATGTCGTTAAG	0	0	0	0	0	0	0	27	0	0	0
novel32940	novel32940	TTGTGCGGTTCAAATAGTAAC	0	0	0	0	0	8	49	14	35	36	22
novel330595	novel330595	ATTTGAGTCATGTCGTTAAG	0	0	0	0	0	0	0	27	0	0	0
novel33090	novel33090	TCAAGGACAGATTGGAGTTAA	0	0	0	125	0	0	292	161	0	78	73
novel33140	**novel33140**	TTGGTAGTGGATAAGGGGGCA	0	0	0	53	0	4	97	45	8	5	17
novel336487	novel336487	TGACATCCAGATAGAAGCTTT	0	0	0	0	0	0	0	0	3	2	0
novel338060	novel338060	TCTGGTAGATGTTTCTCGGCT	0	0	0	0	0	0	0	1	0	0	0
novel339006	novel339006	TCCTGTAACAACAATACGACG	0	0	0	0	0	0	0	2	0	0	0
novel339195	novel339195	TGGCGAGATACTTCAGACGGG	0	0	0	0	0	0	0	3	0	0	0
novel339760	novel339760	TCTTATGTTAAGAGATATGTC	0	0	0	0	0	0	0	2	0	1	0
novel339866	novel339866	ACAAGGGATTTTAAAATGGTT	0	0	0	0	0	0	0	5	0	0	0
novel341082	novel341082	AAATTTGGGACGAAAAAAGCA	0	0	0	0	0	0	0	6	0	3	0
novel343312	novel343312	TGAGAGAATGTGCTTGGGGAC	0	0	0	0	0	0	0	2	0	0	0
novel344549	novel344549	GTGGGTTGCGGATAACGGTA	0	0	0	0	0	0	0	98	0	0	0
novel344564	novel344564	TCGTGAAGTGTTTTGTGGGTT	0	0	0	0	0	0	0	2	0	2	0
novel344909	novel344909	GAAATAGCGAAGATATGATTA	0	0	0	0	0	0	0	8	0	0	0
novel345073	novel345073	TGGGAGGTGATCATATAAGAT	0	0	0	0	0	0	0	2	1	0	0
novel34539	novel34539	TTCTAACTCGCTGACGTCATA	0	0	0	311	655	240	0	13	0	0	14
novel34709	novel34709	TACTTATACTTAAGGTCGTTG	0	0	0	0	0	0	10	6	12	11	0
novel35215	novel35215	ATGGAAGAGTGATATGGATAA	0	0	0	0	0	0	0	0	7	16	0
novel352378	novel352378	ATTTGAGTCATGTCGTTAAG	0	0	0	0	0	0	0	27	0	0	0
novel35322	novel35322	ACAAAGTTTTATACTGACAAT	0	0	0	0	0	0	21	9	7	3	13
novel35394	novel35394	ATCATGAGAATGCAAATCCTT	0	0	0	0	0	0	0	0	0	0	10
novel357618	novel357618	ACACCATTGTACGGATGTTTC	0	0	0	0	0	0	0	0	3	0	0
novel366595	novel366595	CTTTTACTGATGGAGATGCT	0	0	0	0	0	0	0	0	4	0	0
novel368302	novel368302	GTTCGAGGCACGTTGGGAGG	0	0	0	0	0	0	0	0	16	0	0

novel371999	novel371999	GAGGAGGAGGAGGTGGTGGTG	0	0	0	0	0	0	0	0	2	0	0
novel3767	novel3767	AGCTAAGGATTTGCATTCTCA	102	0	41	35	76	43	97	34	58	45	69
novel377476	novel377476	AAATAGATGGCGGTTTTTTC	0	0	0	0	0	0	0	0	3	0	0
novel3806	novel3806	TGGGTTGAGTTGAGTTGAGTTGGC	26	0	18	74	321	91	43	20	0	0	0
novel381053	novel381053	TGTCACAGGAGTCGCGGGATCTA	0	0	0	0	0	0	0	0	1	0	0
novel38564	novel38564	TTGAATGTGAATGAATCGGGC	0	0	0	8	18	12	15	3	2	2	0
novel386471	novel386471	AAGGTGAGGGTGGTATTTGCG	0	0	0	0	0	0	0	0	3	0	0
novel389398	novel389398	AAATTCTCGTTGACTAAGCCA	0	0	0	0	0	0	0	0	3	3	0
novel3943	novel3943	TGACAGAAGAGAGAGAGCAG	1696	968	0	0	0	0	0	0	0	0	2626
novel396525	novel396525	ATAGTCAATTTTATCGGTCTG	0	0	0	0	0	0	0	0	1	6	0
novel399381	novel399381	AACGGGTGTGGGATTAGTCGG	0	0	0	0	0	0	0	0	3	0	0
novel4011	novel4011	TGTTAAGGAGTGTTAACGGTG	175	212	48	0	123	235	0	0	0	0	180
novel401922	novel401922	ATTGAATATGTTGGTTACTAT	0	0	0	0	0	0	0	0	0	6	0
novel4040	novel4040	ACTTTCAAAGGTGACTCTCGATA	0	0	0	0	41	8	6	0	0	1	0
novel404796	novel404796	ATTGAATATGTTGGTTACTAT	0	0	0	0	0	0	0	0	0	6	0
novel406654	novel406654	TTTGGAAATATTTGGCTTGACT	0	0	0	0	0	0	0	0	0	5	0
novel407282	novel407282	AGAGGTGACCATTGGAGATG	0	0	0	0	0	0	0	0	0	5	0
novel41001	novel41001	TTGCTGAACTCAAGAGCCGGCTG	0	0	0	13	14	0	0	0	0	0	0
novel411515	novel411515	TATAATAGAGGATTTATGGCA	0	0	0	0	0	0	0	0	0	1	0
novel412409	novel412409	TAATGTTTGTGGATACGATGGTG	0	0	0	0	0	0	0	0	0	2	0
novel413612	novel413612	AGTGAAAGATTTTTGGTACGT	0	0	0	0	0	0	0	0	0	1	0
novel417289	novel417289	AAATTCTCGTTGACTAAGCCA	0	0	0	0	0	0	0	0	0	8	0
novel422464	novel422464	GTAGACGCGGTAGGTTATTCA	0	0	0	0	0	0	0	0	0	4	0
novel47052	novel47052	CACTTCATCTGCAGAAGCTGGC	0	0	0	0	0	0	0	2	0	0	0
novel48705	novel48705	TGTGTTTGGGCCGAATTGGGCATG	0	0	23	21	0	0	0	0	0	0	0
novel52098	novel52098	TTGGTGGACAAGATCTGGGAT	0	0	0	6	6	6	0	6	0	0	0

novel57897	novel57897	TGACAGAAGAGAGAGAGCGG	117	0	0	0	0	0	0	0	0	0	0
novel59948	novel59948	AGCTGACTGAGAAGAAGTGGA	0	0	0	0	0	0	0	0	0	0	20
novel63067	novel63067	TTCGTTGTCTGTTCGACCTTG	0	0	0	0	0	0	3	2	2	3	0
novel63198	novel63198	TTAGAGTTTTCTGGATACTTA	0	0	0	0	0	0	10	4	6	5	0
novel69398	novel69398	GCTAAGAGCGGTTCTGATGGA	0	0	0	13	28	13	15	6	2	2	0
novel69399	novel69399	GCTAAGAGCGGTTCTGATGGA	0	0	0	13	28	13	15	6	2	2	0
novel72768	novel72768	TCGGTTCTGATGTGGGCATG	0	0	114	0	0	0	0	0	0	0	0
novel7317	novel7317	ATGGGCTTCGAGCGAAGGCT	1221	1205	1452	1377	1726	1190	0	0	0	0	0
novel75569	novel75569	ACTCAGGATGGATAGCGCCAA	0	0	0	0	0	0	7	0	0	0	0
novel82806	novel82806	ACAAGGGATTTTAAAAGGTTTTA	0	0	0	0	11	0	0	2	4	2	0
novel8445	novel8445	TAGAGTCTCTGTATTTGACGGTG	38	0	0	17	24	0	4	7	0	2	0
novel85988	novel85988	TAGTCCACTGTGGTCTAAGGC	0	0	0	0	6	11	2	0	0	0	0
novel86630	novel86630	AAGTAGACACATAAGAAGGAG	0	0	0	0	0	0	13	11	5	4	0
novel89582	**novel89582**	ATAAATCCCAACATCTTCCA	0	0	0	0	0	0	6	7	2	3	0
novel91448	novel91448	ACTCAGGATGGATAGCGCCAA	0	0	0	0	0	0	5	0	0	0	0
novel9277	novel9277	ATCTGAAGAAAATAGCGGCAT	20	0	15	0	0	0	0	37	22	34	51
novel95369	novel95369	TTCTCGAGCACGCACGAGCAC	0	0	46	0	0	0	0	0	0	0	0

References

Abdel-Ghany, S.E., and Pilon, M. (2008). MicroRNA-mediated systemic down-regulation of copper protein expression in response to low copper availability in Arabidopsis. Journal of Biological Chemistry **283,** 15932-15945.

Abramoff, M.D., Magelhaes, P.J., and Ram, S.J. (2004). Image Processing with ImageJ. Biophotonics International **11,** 36-42.

Adai, A., Johnson, C., Mlotshwa, S., Archer-Evans, S., Manocha, V., Vance, V., and Sundaresan, V. (2005). Computational prediction of miRNAs in Arabidopsis thaliana. Genome Res **15,** 78-91.

Ahmed, I., Sarazin, A., Bowler, C., Colot, V., and Quesneville, H. (2011). Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in Arabidopsis. Nucleic Acids Res**,** 1-13.

Allen, e., Xie, Z., Gustafson, A.M., and Carrington, J.C. (2005). Micro-RNA directed phasing during trans-acting siRNA biogenesis in plants. Cell **121,** 207-221.

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D.E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C., and Ecker, J.R. (2003). Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science **301,** 653-657.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biology **11,** R106.

Backman, T.W., Sullivan, C.M., Cumbie, J.S., Miller, Z.A., Chapman, E.J., Fahlgren, N., Givan, S.A., Carrington, J.C., and Kasschau, K.D. (2008). Update of ASRP: the Arabidopsis Small RNA Project database. Nucleic acids research **36,** D982-985.

Benfey, P.N., and Scheres, B. (2000). Root Development. Current Biology **10,** R813-814.

Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., and Benfey, P.N. (2003). A Gene Expression Map of the Arabidopsis Root. Science **302,** 1956-1960.

Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., and Nekrutenko, A. (2010). Manipulation of FASTQ data with Galaxy. Bioinformatics **26,** 1783-1785.

Bonnet, E., Wuyts, J., Rouze, P., and Van de Peer, Y. (2004). Detection of 91 potential conserved plant microRNAs in Arabidopsis thaliana and Oryza sativa identifies important target genes. Proc Natl Acad Sci U S A **101,** 11511-11516.

Borges, F., Pereira, P.A., Slotkin, R.K., Martienssen, R.A., and Becker, J.D. (2011). MicroRNA activity in the Arabidopsis male germline. J Exp Bot **62,** 1611-1620.

Brady, S.M., Orlando, D.A., Lee, J.-Y., Wang, J.Y., Koch, J., Dinneny, J.R., Mace, D., Ohler, U., and Benfey, P.N. (2007). A High-Resolution Root Spatiotemporal Map Reveals Dominant Expression Patterns. Science **318,** 801-806.

Breakfield, N.W., Corcoran, D.L., Petricka, J.J., Shen, J., Sae-Seaw, J., Rubio-Somoza, I., Weigel, D., Ohler, U., and Benfey, P.N. (2011). High-resolution experimental and computational profiling of tissue-specific known and novel miRNAs in Arabidopsis. Genome Research **under review**.

Carlsbecker, A., Lee, J.-Y., Roberts, C.J., Dettmer, J., Lehesranta, S., Zhou, J., Lindgren, O., Moreno-Risueno, M.A., Vate´n, A., Thitamadee, S., Campilho, A., Sebastian, J., Bowman, J.L., Helariutta, Y., and Benfey, P.N. (2010). Cell signalling by microRNA165/6 directs gene dose-dependent root cell fate. Nature **465,** 316-321.

Chapman, E.J., and Carrington, J.C. (2007). Specialization and evolution of endogenous small RNA pathways. Nature Reviews Genetics **8,** 884-896.

Chen, X. (2010). Small RNAs - secrets and surprises of the genome. Plant J **61,** 941-958.

Cuperus, J.T., Montgomery, T.A., Fahlgren, N., Burke, R.T., Townsend, T., Sullivan, C.M., and Carrington, J.C. (2010). Identification of MIR390a precursor processing-defective mutants in Arabidopsis by direct genome sequencing. Proc Natl Acad Sci U S A **107,** 466-471.

Czech, B., and Hannon, G.J. (2011). Small RNA sorting: matchmaking for Argonautes. Nat Rev Genet **12,** 19-31.

Dai, X., and Zhao, P.X. (2011). psRNA target: A plant small RNA target analysis server. Nucleic Acids Res.

de Felippes, F.F., Ott, F., and Weigel, D. (2010). Comparative analysis of non-autonomous effects of tasiRNAs and miRNAs in Arabidopsis thaliana. Nucleic Acids Research**,** 1-10.

Dello Ioio, R., Linhares, F.S., Scacchi, E., Casamitjana-Martinez, E., Heidstra, R., Costantino, P., and Sabatini, S. (2007). Cytokinins determine Arabidopsis root-meristem size by controlling cell differentiation. Curr Biol **17,** 678-682.

Fahlgren, N., Howell, M.D., Kasschau, K.D., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., Law, T.F., Grant, S.R., Dangl, J.L., and Carrington, J.C. (2007). High-Throughput Sequencing of Arabidopsis microRNAs: Evidence for Frequent Birth and Death of MIRNA Genes. PLoS One **2,** e219.

Fahlgren, N., Sullivan, C.M., Kasschau, K.D., Chapman, E.J., Cumbie, J.S., Montgomery, T.A., Gilbert, S.D., Dasenko, M., Backman, T.W., Givan, S.A., and Carrington, J.C. (2009). Computational and analytical framework for small RNA profiling by high-throughput sequencing. RNA **15,** 992-1002.

Farazi, T.A., Spitzer, J.I., Morozov, P., and Tuschl, T. (2011). miRNAs in human cancer. J Pathol **223,** 102-115.

Franco-Zorrilla, J.M., Valli, A., Todesco, M., Mateos, I., Puga, M.I., Rubio-Somoza, I., Leyva, A., Weigel, D., Garcia, J.A., and Paz-Ares, J. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. Nat Genet **39,** 1033-1037.

Friedlander, M.R., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S., and Rajewsky, N. (2008). Discovering microRNAs from deep sequencing data using miRDeep. Nat Biotechnol **26,** 407-415.

Grant-Downton, R., Trionnaire, G.L., Schmid, R., Rodriguez-Enriquez, J., Hafidh, S., Mehdi, S., Twell, D., and Dickinson, H. (2009). MicroRNA and tasiRNA diversity in mature pollen of Arabidopsis thaliana. BMC Genomics **10,** 1-10.

Griffiths-Jones, S., Saini, H.K., Dongen, S.v., and Enright, A.J. (2008). miRBase: tools for microRNA genomics. Nucleic acids research **36,** D154-158.

Gustafson, A.M., Allen, E., Givan, S., Smith, D., Carrington, J.C., and Kasschau, K.D. (2005). ASRP: the Arabidopsis Small RNA Project Database. Nucleic Acids Res **33,** D637-640.

Gy, I., Gasciolli, V., Lauressergues, D., Morel, J.B., Gombert, J., Proux, F., Proux, C., Vaucheret, H., and Mallory, A.C. (2007). Arabidopsis FIERY1, XRN2, and XRN3 are endogenous RNA silencing suppressors. Plant Cell **19,** 3451-3461.

Hendrix, D., Levine, M., and Shi, W. (2010). miRTRAP, a computational method for the systematic identification of miRNAs from high throughput sequencing data. Genome Biol **11,** R39.

Henz, S.R., Cumbie, J.S., Kasschau, K.D., Lohmann, J.U., Carrington, J.C., Weigel, D., and Schmid, M. (2007). Distinct expression patterns of natural antisense transcripts in Arabidopsis. Plant Physiology **144,** 1247-1255.

Hsieh, L.C., Lin, S.I., Shih, A.C., Chen, J.W., Lin, W.Y., Tseng, C.Y., Li, W.H., and Chiou, T.J. (2009). Uncovering small RNA-mediated responses to phosphate deficiency in Arabidopsis by deep sequencing. Plant Physiol **151,** 2120-2132.

Huntzinger, E., and Izaurralde, E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat Rev Genet **12,** 99-110.

Initiative, A.G. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature **408,** 796-815.

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

Khan, G.A., Declerck, M., Sorin, C., Hartmann, C., Crespi, M., and Lelandais-Briere, C. (2011). MicroRNAs as regulators of root development and architecture. Plant Mol Biol.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol **10,** R25.

Lee, J.-Y., Colinas, J., Wang, J.Y., Mace, D., Ohler, U., and Benfey, P.N. (2006). Transcriptional and posttranscriptional regulation of transcription factor expression in Arabidopsis roots. PNAS **103,** 6055-6060.

Lee, M.M., and Schiefelbein, J. (1999). WEREWOLF, a MYB-related protein in Arabidopsis, is a position-dependent regulator of epidermal cell patterning. Cell **99,** 473-483.

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

Levesque, M.P., Vernoux, T., Busch, W., Cui, H., Wang, J.Y., Bilou, I., Hassan, H., Nakajima, K., Matsumoto, N., Lohmann, J.U., Scheres, B., and Benfey, P.N. (2006). Whole-Genome Analysis of the SHORT-ROOT Developmental Pathway in Arabidopsis PLoS Biology **4,** e143.

Linsen, S.E., de Wit, E., Janssens, G., Heater, S., Chapman, L., Parkin, R.K., Fritz, B., Wyman, S.K., de Bruijn, E., Voest, E.E., Kuersten, S., Tewari, M., and Cuppen, E. (2009). Limitations and possibilities of small RNA digital gene expression profiling. Nat Methods **6,** 474-476.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell **133,** 523-536.

Llave, C., Kasschau, K.D., Rector, M.A., and Carrington, J.C. (2002). Endogenous and silencing-associated small RNAs in plants. Plant Cell **14,** 1605-1619.

Llave, C., Franco-Zorrilla, J.M., Solano, R., and Barajas, D. (2011). Target validation of plant microRNAs. Methods Mol Biol **732,** 187-208.

Lu, C., Meyers, B.C., and Green, P.J. (2007). Construction of small RNA cDNA libraries for deep sequencing. Methods **43,** 110-117.

Mahonen, A.P., Bonke, M., Kauppinen, L., Riikonen, M., Benfey, P.N., and Helariutta, Y. (2000). A novel two-component hybrid molecule regulates vascular morphogenesis of the Arabidopsis root. Genes & development **14,** 2938-2943.

Mallory, A., and Vaucheret, H. (2010). Form, function, and regulation of ARGONAUTE proteins. Plant Cell **22,** 3879-3889.

Marin, E., Jaouannet, V., Herz, A., Lorerse, A.S., Weijers, D., Vaucheret, H., Nussaume, L., Crespi, M.D., and Maizel, A. (2010). MiR390, Arabidopsis TAS3 tasiRNAs, and their AUXIN RESPONSE FACTOR targets define an autoregulatory network quantiatively regulating lateral root growth. Plant Cell **22,** 1104-1117.

Matzke, M., Kanno, T., Daxinger, L., Huettel, B., and Matzke, A.J. (2009). RNA-mediated chromatin-based silencing in plants. Curr Opin Cell Biol **21,** 367-376.

McElver, J., Tzafrir, I., Aux, G., Rogers, R., Ashby, C., Smith, K., Thomas, C., Schetter, A., Zhou, Q., Cushman, M.A., Tossberg, J., Nickle, T., Levin, J.Z., Law, M., Meinke, D., and Patton, D. (2001). Insertional Mutagenesis of Genes Required for Seed Development in Arabidopsis thaliana. Genetics **159,** 1751-1763.

Mench, M., Schwitzguebel, J.P., Schroeder, P., Bert, V., Gawronski, S., and Gupta, S. (2009). Assessment of successful experiments and limitations of phytotechnologies: contaminant uptake, detoxification and sequestration, and consequences for food safety. Environ Sci Pollut Res Int **16,** 876-900.

Meyer, S.U., Pfaffl, M.W., and Ulbrich, S.E. (2010). Normalization strategies for microRNA profiling experiments: a 'normal' way to a hidden layer of complexity? Biotechnol Lett **32,** 1777-1788.

Meyers, B.C., Axtell, M.J., Bartel, B., Bartel, D.P., Baulcombe, D., Bowman, J.L., Cao, X., Carrington, J.C., Chen, X., Green, P.J., Griffiths-Jones, S., Jacobsen, S.E., Mallory, A.C., Martienssen, R.A., R. Scott Poethig, Qi, Y., Vaucheret, H., Voinnet, O., Watanabe, Y., Detlef Weigel, and Zhu, J.-K. (2008). Criteria for annotation of plant MicroRNAs. Plant Cell **20,** 3186-3190.

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., and Long, C. (2008). Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5'terminal nucleotide. Cell **133,** 116-127.

Montgomery, T.A., Howell, M.D., Cuperus, J.T., Li, D., Hansen, J.E., Alexander, A.L., Chapman, E.J., Fahlgren, N., Allen, E., and Carrington, J.C. (2008). Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. Cell **133,** 128-141.

Mosher, R.A., and Melnyk, C.W. (2010). SiRNAs and DNA methylation. Trends in Plant Science.

Moxon, S., Schwach, F., Dalmay, T., MacLean, D., Studholme, D.J., and Moulton, V. (2008). A toolkit for analysing large-scale plant small RNA datasets. Bioinformatics **24,** 2252-2253.

Napoli, C., Lemieux, C., and Jorgensen, R. (1990). lntroduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes In trans. The Plant Cell **2,** 279-289.

Nawy, T., Lee, J.Y., Colinas, J., Wang, J.Y., Thongrod, S.C., Malamy, J.E., Birnbaum, K., and Benfey, P.N. (2005). Transcriptional profile of the Arabidopsis root quiescent center. Plant Cell **17,** 1908 - 1925.

Ossowski, S., Schwab, R., and Weigel, D. (2008). Gene silencing in plants using artificial microRNAs and other small RNAs. Plant J **53,** 674-690.

Park, W., Li, J., Song, R., Messing, J., and Chen, X. (2002). CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. Curr Biol **12,** 1484-1495.

Rajagopalan, R., Vaucheret, H., Trejo, J., and Bartel, D.P. (2006). A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. Genes Dev **20,** 3407-3425.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. (2002). MicroRNAs in plants. Genes Dev **16,** 1616-1626.

Saeed, A.I., Bhagabati, N.K., Braisted, J.C., Liang, W., Sharov, V., Howe, E.A., Li, J., Thiagarajan, M., White, J.A., and Quackenbush, J. (2006). TM4 microarray software suite. Methods Enzymol **411,** 134-193.

Sedbrook, J.C., Carroll, K.L., Hung, K.F., Masson, P.H., and Somerville, C.R. (2002). The Arabidopsis SKU5 gene encodes an extracellular glycosyl phosphatidylinositol anchored glycoprotein involved in directional root growth. The Plant Cell **14,** 1635-1648.

Sena, G., Jung, J.W., and Benfey, P.N. (2004). A broad competence to respond to SHORT ROOT revealed by tissue-specific ectopic expression. Development **131,** 2817-2826.

Sessions, A., Burke, E., Presting, G., Aux, G., McElver, J., Patton, D., Dietrich, B., Ho, P., Bacwaden, J., Ko, C., Clarke, J.D., Cotton, D., Bullis, D., Snell, J., Miguel, T., Hutchison, D., Kimmerly, B., Mitzel, T., Katagiri, F., Glazebrook, J., Law, M., and Goff, S.A. (2002). A High-Throughput Arabidopsis Reverse Genetics System The Plant Cell **14,** 2985-2994.

Shockey, J.M., Fulda, M.S., and Browse, J. (2203). Arabidopsis contains a large superfamily of acyl-activating enzymes. Plant Physiology **132,** 1055-1076.

Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T.Z., Garcia-Hernandez, M., Foerster, H., Li, D., Meyer, T., Muller, R., Ploetz, L., Radenbaugh, A., Singh, S., Swing, V., Tissier, C., Zhang, P., and Huala, E. (2008). The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Res **36,** D1009-1014.

Team, R.D. (2009). R: A language and environment for statistical computing. In R Development Team.

Teixeira, F.K., and Colot, V. (2010). Repeat elements and the Arabidopsis DNA methylation landscape. Heredity **105,** 14-23.

Tilman, D., Hill, J., and Lehman, C. (2006). Carbon-negative biofuels from low-input high-diversity grassland biomass. Science **314,** 1598-1600.

To, J.P., Zhu, J., Benfey, P.N., and Elich, T. (2010). Optimizing root system architecture in biofuel crops for sustainable energy production and soil carbon sequestration. F1000 Biol Rep **2,** 65.

Tsukagoshi, H., Busch, W., and Benfey, P.N. (2010). Transcriptional regulation of ROS controls transition from proliferation to differentiation in the root. Cell **143,** 606-616.

Valoczi, A., Varallyay, E., Kauppinen, S., Burgyan, J., and Havelda, Z. (2006). Spatio-temporal accumulation of microRNAs is highly coordinated in developing plant tissues. Plant J **47,** 140-151.

Van Norman, J.M., Breakfield, N.W., and Benfey, P.N. (2011). Intercellular Communication during Plant Development. Plant Cell **23,** 855-864.

Varkonyi-Gasic, E., Wu, R., Wood, M., Walton, E.F., and Hellens, R.P. (2007). Protocol: a highly-sensitive RT-PCR method for detection and quantification of microRNAs. Plant Methods **3,** 12.

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A.C., Hilbert, J.L., Bartel, D.P., and Crete, P. (2004). Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. Mol Cell **16,** 69-79.

Verdel, A., Vavasseur, A., Le Gorrec, M., and Touat-Todeschini, L. (2009). Common themes in siRNA-mediated epigenetic silencing pathways. Int J Dev Biol **53,** 245-257.

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

Waters, B.M., and Grusak, M.A. (2008). Quantitative trait locus mapping for seed mineral concentrations in two Arabidopsis thaliana recombinant inbred populations. The New Phytologist **179**.

Yang, X., Zhang, H., and Li, L. (2011). Global analysis of gene-level microRNA expression in Arabidopsis using deep sequencing data. Genomics **98,** 40-46.

Yoshikawa, M., Peragine, A., Park, M.Y., and Poethig, R.S. (2005). A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. Genes and Development **19,** 2164-2175.

Biography

Natalie Wynn (Leonard) Breakfield was born in Saint Louis, Missouri on January 5, 1974. She grew up in Hillsboro, Missouri, graduating from St. Pius X High School in Festus, Missouri in 1991. It was at St. Pius that she met her husband, Mark Breakfield, who she married in 1994. Natalie received her Bachelor of Science in Biological Science with a minor in Chemistry from Quincy University in 1994. She received her Master of Science in Molecular and Cellular Biology in the lab of Dr. Marilyn Yoder from the University of Missouri at Kansas City in 1996. Her thesis received a special merit award.

Natalie and Mark have been blessed with three beautiful children: Ian, age 15, Andrew, age 14, and Grace, age 11. She was a happy stay-at-home mother for several years, and then worked as a part time lab technician in the lab of Dr. Chris Taylor at the Donald Danforth Plant Science Center. Based on the skills and encouragement she received there, she applied to graduate schools, and after the family made its move to North Carolina, she began work on her doctoral degree at Duke University in 2006.

Natalie is an author on two papers from her Master's work (Doan *et al*. *Acta Cryst*. (2000) **D56**, 351-353; and Oliver *et al*. *Acta Cryst*. (1999) **D55**, 522-524.). She co-authored a review (Van Norman *et al*., *Plant Cell* (2011) **23**, 855-864) and first author publication from her dissertation work (**Breakfield,** Corcoran, *et al*., (2011), accepted on August 5, 2011 by Genome Research), with two other publications in preparation.