MicroRNAs in differentiating tissues

of *Populus* and *Eucalyptus* trees

by

MICHELLE VICTOR

Submitted in partial fulfillment of the requirements for the degree

Magister Scientiae

In the Faculty of Natural and Agricultural Sciences

Department of Genetics

University of Pretoria

Pretoria

November 2006

Under the supervision of Prof. Alexander A. Myburg

and co-supervision of Prof. Henk Huismans

DECLARATION

I, the undersigned, hereby declare that the dissertation submitted herewith for the degree M.Sc. to the University of Pretoria, contains my own independent work and has not been submitted for any degree at any other university.

Michelle Victor

November 2006

PREFACE

Trees represent the majority of terrestrial biomass production on earth and are massive sinks for atmospheric carbon sequestration. Trees exhibit many unique aspects of developmental plant biology, one of which is the formation of wood. Wood is one of the most abundant natural products and is the raw material essential for the multi-billion dollar world-wide pulp and paper industry. The formation of wood (xylogenesis) is a highly ordered developmental process involving the patterned division and differentiation of meristematic tissues into the specialized cell types in wood. Wood (secondary xylem) originates from the primary vasculature of the plant, which subsequently differentiates from a secondary meristem, the vascular cambium, into secondary xylem and phloem tissue types. The progression of this developmental process requires various molecular signals and differential gene expression across the different tissue types. The tight regulation of this xylogenesis is mediated by genes that regulate cambial meristem differentiation, vascular patterning and xylem cell fate. However, our knowledge of the genetic control and molecular mechanisms determining vascular patterning and ultimately wood formation is limited. MicroRNAs (miRNAs) are a group of endogenous ~ 20 to 24 nt RNA molecules that downregulate gene expression at the post-transcriptional level. MicroRNAs have validated roles in plant development where they are known to regulate meristem cell differentiation and tissue patterning. These regulatory molecules have also been shown to spatially regulate gene expression at different developmental stages. Thus, the vascular cambium and its derivatives are excellent candidate tissues for miRNA discovery and may provide insights into the global gene regulatory networks involved in this developmental process.

The **aim of this M.Sc. study** was to isolate microRNAs from actively differentiating tissues of two tree species, and to investigate if these miRNAs exhibit tissue-specific expression

patterns, in order to identify possible mechanisms of gene regulation during primary and secondary meristem differentiation, tissue patterning and vascular development in plants.

Chapter 1 of this dissertation comprises a review of the literature concerning microRNAs in plants, focusing on their role in plant development. It details some of the main aspects of miRNA discovery, biogenesis and mechanism of action in plants and highlights some of the important differences between animal and plant miRNAs and siRNAs. It further describes some of the known interactions of miRNAs and their target genes, and their subsequent roles during plant development. Recent developments in the identification of miRNAs are discussed with specific reference to sequencing technology. Finally, a brief introduction into the biology of wood formation in trees is presented, focusing on vascular patterning.

In *Populus*, microRNAs were recently shown to be involved in the regulation of secondary xylem formation and in the development of reaction wood. However, little is known about miRNA regulation during the development of the primary vascular tissues of poplar trees. An understanding of the differentiation of primary vascular tissues, as well as other early developmental patterning events, will provide insight into the process involved in early tree growth and development. The aim of this study was to identify miRNAs present in early post-embryonic tissues of poplar trees. **Chapter 2** of this dissertation describes the isolation and characterization of microRNAs from *in vitro* whole plantlets of *Populus trichocarpa*. Putative miRNA precursor genes are identified and characterized with the use of computational prediction. A range of predicted target genes are described for these miRNAs, and the possible roles of the miRNAs and their targets during the early development of *P. trichocarpa* are discussed.

MicroRNAs are well characterized as regulators of tissue patterning of plant organs. Although certain aspects of xylogenesis have been studied in poplar, there may be other

aspects of this process under miRNA-mediated control, which have to date not been identified. The use of another tree genus, such as *Eucalyptus*, to investigate this hypothesis will provide further insights into vascular development, wood formation and possibly allow the identification of miRNAs specific to trees. **Chapter 3** of this dissertation describes the isolation and characterization of putative microRNAs from secondary vascular tissues of the *Eucalyptus* tree stem. The gene targets for these miRNAs are identified in the poplar and *Arabidopsis* genomes, and the functions of these targets are discussed. Detailed expression profiles are provided for the putative *Eucalyptus* miRNAs in both xylogenic and non-xylogenic tissues.

Finally, the general findings, conclusions and implications of this M.Sc. study are summarized at the end of the dissertation in a section titled **Concluding Remarks**.

The findings presented in this dissertation represent the outcomes of a study undertaken from March 2004 to October 2006 in the Department of Genetics, University of Pretoria, under the supervision of Prof. A.A. Myburg. The work described Chapter 2 was performed at North Carolina State University and forms part of a collaborative project which was initiated to complement their previous poplar miRNA studies, and to obtain technical training in the isolation and characterization of plant miRNAs. Chapters 2 and 3 have been prepared in manuscript format for submission to peer-reviewed research journals, as is the standard procedure for our research group. Therefore, a certain degree of redundancy may exist between the introductory sections of these chapters and Chapter 1. The following posters and congress presentations were generated based on the preliminary results obtained in this M.Sc. study:

VICTOR, M., Lu, S., Sun, Y-H., Chiang, V.L., and Myburg, AA. 2005. Isolation and characterization of miRNAs and siRNAs in differentiating woody tissues of *Eucalyptus*. IUFRO Tree Biotechnology Conference S6.26p, November 6-11. Pretoria, South Africa (Poster presentation).

VICTOR, M., Lu, S., Sun, Y-H., Chiang, V.L., and Myburg, AA. 2006. Isolation and characterization of miRNAs and siRNAs in differentiating woody tissues of *Eucalyptus*. South African Genetics Society Conference. April 3 -6. Bloemfontein, South Africa (Prize for best poster).

VICTOR, M., Lu, S., Sun, Y-H., Chiang, V.L., and Myburg, AA. 2006. MicroRNA and siRNA profiling in woody tissues of *Eucalyptus* and *Populus*. South African Genetics Society Conference. April 3 – 6. Bloemfontein, South Africa. (Oral presentation)

ACKNOWLEDGEMENTS

I would like to express my gratitude to the following people, organizations and institutes for assisting me in the completion of this project:

- To Prof. A.A. Myburg, for introducing me to the idea of studying the molecular biology of wood formation and microRNAs, for his professional, thorough and creative leadership of this project, and for extensive review of this dissertation.

- To Prof. V.L. Chiang, Dr. S. Lu, Dr. Y-H. Sun, and all the members of the Forest Biotechnology Group at North Carolina State University, for training in the miRNA isolation technique and for all their insights and valuable technical assistance.

- To Dr. E. Asamizu and Dr. S. Tabata from Kazusa DNA Research Institute for providing the *Eucalyptus camaldulensis* genome sequences and performing certain bioinformatic analyses essential for the successful completion of this study.

- To Prof. H. Huismans and Dr. J. Theron for their assistance with the initiation of this research project and review of the dissertation.

- To Martin Ranik and Saul Garcia for outstanding insights, excellent advice, technical assistance and extensive review of this dissertation.

- To all my colleagues in the Forest Molecular Genetics Laboratory: Elna Cowley, Adrene Laubscher, Minique De Castro, Nicky Creux, Frank Maleka, John Kemp, Eshchar Mizrachi, Tracey-Leigh Hatherell, Joanne Bradfield, Honghai Zhou, Luke Solomon, Grant McNair, Alisa Postma, Mmoledi Mphahlele, Marja O'Neill and Kitt Payn for forming a superb research programme and intellectually stimulating research environment.

- To Sappi Forest Products and Mondi Business Paper South Africa, for supplying the plant materials used in this study, as well as for funding contributions to the Wood and Fibre Molecular Genetics Programme (WFMG).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BAC	Bacterial Artificial Chromosome
BLAST	Basic Local Alignment Search Tool
DCL	Dicer-like enzyme
EMBOSS	European Molecular Biology Open Source Suite
Egr-miRNA	*Eucalyptus grandis* microRNA
EST	Expressed Sequence Tag
miRNA	MicroRNA
miRNA*	Small RNA derived from opposite stem-loop arm
NBS	Nucleotide-binding site
NCBI	National Center for Biotechnology Information
PAPS	3'-phosphoadenosine 5'-phosphosulfate sulfotransferase
PTGS	Post Transcriptional Gene Silencing
Ptc-miRNA	*Populus trichocarpa* microRNA
RISC	RNA Induced Silencing Complex
RNAi	RNA interference
rRNA	ribosomal RNA
siRNA	Small Interfering RNA
snRNA	Small nucleolar RNA
TIGR	The Institute for Genomic Research
tRNA	Transfer RNA
RT-PCR	Reverse Transcriptase PCR
qRT-PCR	Quantitative Real Time PCR
UTR	Untranslated Region

CHAPTER 1

LITERATURE REVIEW

MICRORNAS AND THEIR ROLE IN PLANT DEVELOPMENT

1.1 INTRODUCTION

For years the central dogma of molecular biology has held that biological information and function flows from DNA to RNA to protein. This long accepted theory has been challenged by the discovery of a gene regulation mechanism, known as RNA-mediated gene silencing (RNA silencing). RNA silencing is the process whereby small, approximately 20 –24 nt, non-coding RNA molecules direct the sequence-specific down regulation of target genes at the post-transcriptional level. There are two known classes of small RNAs that mediate this process, namely microRNAs (miRNAs) and short interfering RNAs (siRNAs), which act as cellular regulators to control mRNA stability and translation, to protect the genome from invading nucleic acids, to facilitate epigenetic modifications of chromatin and histones, and to direct complex developmental pathways (Baulcombe, 1996; Jones et al., 1999; Ketting et al., 1999; Olsen and Ambros, 1999; Mette et al., 2000).

MicroRNAs were discovered in *Caenorhabditis elegans* (Lee et al., 1993) and since then have been found in plants (Llave et al., 2002a; Mette et al., 2002; Park et al., 2002; Reinhart et al., 2002), *Drosophila* (Elbashir et al., 2001a), mammals (Lagos-Quintana et al., 2002) and viruses (Pfeffer et al., 2004). To date, over 4300 miRNAs have been identified in 49 different species (miRBase release 9.0, October 2006; http://microrna.sanger.ac.uk; http://cgrb.orst.edu/smallRNA/db/) and it has been estimated that approximately one percent of the predicted genes in vertebrates encode miRNAs (Bartel, 2004). MiRNAs are thought to be at least 400 million years old (Pasquinelli et al., 2000; Floyd and Bowman, 2004) and many miRNAs, and their target genes, are conserved across unrelated species. Thus, it is clear that this group of molecules form part of an evolutionary conserved mechanism of gene regulation.

The common theory emerging about the role of miRNAs in biological systems is that they function to down-regulate endogenous genes involved in developmental programmes or

tissue patterning (Bartel and Bartel, 2003; Carrington and Ambros, 2003). The role of miRNAs during development was first seen in *C. elegans*. The first known miRNAs, *lin-4* and *let-7*, were found to regulate the timing of larval development by controlling transitions from embryo to larva to adult phases (Feinbaum and Ambros, 1999; Reinhart et al., 2000; Banerjee and Slack, 2002; Grosshans and Slack, 2002). In plants, developmental pathways that are controlled by miRNAs include leaf, shoot, floral and auxiliary meristem differentiation (McConnell et al., 2001; Aukerman and Sakai, 2003; Chen, 2004), vascular development (Kim et al., 2005; Lu et al., 2005; Ko et al., 2006), the establishment of organ polarity and timing of vegetative to reproductive transition (Reinhart et al., 2002; Rhoades et al., 2002; Mallory et al., 2004a; Wu and Poethig, 2006; Wu et al., 2006), and environmental- and stress-induced responses (Sunkar and Zhu, 2004; Lu et al., 2005).

This review will focus on microRNAs in plants and will aspects of highlight their discovery, biogenesis, mechanism of action and cellular functioning. Some aspects of miRNAs in animals and siRNAs will be discussed, but it is beyond the scope of this review to provide in depth discussions of these topics. In order to understand the mechanism of RNA silencing, and the role of small RNAs, it may be helpful to understand how this novel gene regulatory process was discovered. The next section provides a brief historical overview of the discovery of RNA silencing and an introduction to the key participants (siRNAs and miRNAs) involved in the functionality of the process.

1.2 THE DISCOVERY OF RNA SILENCING

The phenomenon of RNA silencing was discovered during transgene studies in plants. In 1990, the Jorgensen group attempted to upregulate the chalcone synthase (CHS) gene involved in flower pigmentation by introducing an additional copy of the CHS gene. Instead of the expected increase in flower pigmentation, most flowers showed white sectors, or a total lack of pigment, in contrast to the wild-type purple phenotype. The loss of pigmentation was

attributed to reduced CHS mRNA levels of both the transgene and endogenous gene, a phenomenon which they termed cosuppression (Napoli et al., 1990). These initial observations lead to reports of similar events observed during other transgene studies in plants (de Carvalho et al., 1992; van Blokland, 1994; Metzlaff et al., 1997; Tijsterman et al., 2002a). Cosuppression was later renamed post-transcriptional gene silencing (PTGS) after findings indicated that down-regulation occurred due to a reduction in mRNA at the post-transcriptional level (Vaucheret et al., 2001).

One year after the discovery of RNA silencing in plants, reports began to emerge about similar phenomena, termed RNA interference (RNAi) in *C. elegans* and *Drosophila melanogaster* (Fire et al., 1991; Kennerdell and Carthew, 1998; Tuschl et al., 1999) and quelling in the fungi *Neurospora crassa* and *Trypanosome brucei* (Romano and Macino, 1992; Cogoni et al., 1996; Ngo et al., 1998). These studies further proved that gene expression could be down regulated through the introduction of transgenes exhibiting complementary to endogenous genes. There was a definite correlation between the events of PTGS, RNAi and quelling and thus their mechanisms were thought to be similar. Since all of these processes lead to a similar outcome, for the purpose of this review they will be collectively termed RNA silencing.

By the late 1990's, numerous studies of RNA silencing provided insights into the mechanism involved in this process. It was discovered that the effector molecule of RNA silencing was dsRNA, where the addition of exogenous dsRNA molecules were shown to be potent triggers of endogenous gene silencing (Guo and Kemphues, 1995; Baulcombe, 1996; Fire et al., 1998; Waterhouse et al., 1998; Hamilton and Baulcombe, 1999; Mette et al., 2000). The silencing effect was shown to be highly specific, in that it was limited to the homologous target gene, and the silencing signal was systemic and could travel in a stable fashion throughout the organism (Fire et al., 1991; Fire et al., 1998). A general impression emerged

that endogenous mRNA molecules were targeted for destruction by introducing homologous dsRNA molecules.

Further studies proved that the introduced dsRNA molecule was cleaved into smaller (20 – 24 nt) dsRNAs, termed siRNAs, which are the actual mediators of RNA silencing (Hamilton and Baulcombe, 1999; Elbashir et al., 2001c; Elbashir et al., 2001a). These molecules were shown to direct cleavage of its target mRNA molecule in a sequence-specific manner, and resulted in reduced mRNA levels and a corresponding reduction in protein products (Zamore et al., 2000; Elbashir et al., 2001b; Tang et al., 2003). These discoveries confirmed that the small RNAs target corresponding complementary mRNAs in a sequence-specific manner to direct gene silencing through mRNA cleavage and degradation.

In 2000, microRNAs (miRNAs) were discovered to be another functionally important, but seemingly distinct, class of endogenous small, non-coding RNAs. The first identified microRNAs, and the genes encoding them, were discovered during mutant studies in *C. elegans*. The *lin-4* and *let-7* miRNAs were shown to be involved in the heterochronic developmental pathway (Lee et al., 1993; Reinhart et al., 2000), and found to specifically control the timing of larval development (Lee et al., 1993; Reinhart et al., 2000). Elbashir et al (2001b) then identified miRNAs involved in developmental timing events in *Drosophila*. And during 2002, four groups independently reported the existence of plant miRNAs in *Arabidopsis thaliana* (Llave et al., 2002a; Mette et al., 2002; Park et al., 2002; Reinhart et al., 2002). Taken together, these groups contributed to the discovery of over 100 *Arabidopsis* miRNAs, many of which are now known to be conserved in sequence in other plant species such as rice (*Oryza sativa*), maize, tobacco and poplar (Llave et al., 2002a; Mette et al., 2002; Park et al., 2002; Reinhart et al., 2002; Lu et al., 2005). Since then the Sanger Institute has established the miRNA Registry, which is a public database of all known miRNAs that allows users to search for published miRNA sequences, precursor sequences and the putative target

genes of all known miRNAs. As of October 2006 the Sanger Centre MicroRNA Registry recorded 4361 miRNAs in a total of 49 different species, and of these 863 are from 9 different plant species (http://microrna.sanger.ac.uk). These miRNAs are grouped into families based on sequence similarity, where members differ from each other by a maximum of three nucleotides, target different members of the same gene families, and are derived from different genomic locations.

Studies of miRNA function in animals revealed that gene expression could be regulated by a novel mechanism of RNA silencing. This process differed from that seen for siRNAs, where target mRNA is cleaved and degraded, and thus miRNAs were not initially classified as being involved in classical RNA silencing. In animals, the mature miRNAs were shown to bind to the 3' untranslated region (UTR) of their target mRNAs with partial homology, and in this way prevented translation of the target gene (Lee et al., 1993; Wightman et al., 1993; Reinhart et al., 2000; Slack et al., 2000). Thus, mRNA translation is halted, protein levels are decreased, and mRNA levels remained constant. This mechanism of RNA mediated gene silencing was referred to as translational repression. However, studies of plant miRNAs proved that miRNAs could in fact act as siRNAs during RNAi, by targeting mRNAs for cleavage and degradation (Llave et al., 2002b).

Since their initial discovery, numerous miRNAs have been identified, many of which are specifically involved in developmental programmes across kingdoms (Banerjee and Slack, 2002; Bartel and Bartel, 2003; Carrington and Ambros, 2003). These observations lead to the belief that microRNAs were involved in novel gene regulatory mechanisms during complex developmental pathways, where certain genes need to be down-regulated at crucial times. Although it is clear that much still remains unknown, RNA silencing is one of the most rapidly expanding areas of research, especially since the discovery and use of introduced siRNAs for knock-out mutant studies. Today RNA silencing events are known to occur in

almost all eukaryotes, including protozoa, flies, nematodes, insects, parasites, mouse and human cell lines (reviewed by Agrawal et al., 2003). The biological functions of gene silencing include small RNA targeted mRNA degradation, regulation of gene expression during developmental processes (Olsen and Ambros, 1999; Ambros, 2000; Reinhart et al., 2000; Grishok et al., 2001; Ketting et al., 2001; Lau et al., 2001; McConnell et al., 2001; Brennecke et al., 2003), protection against invading dsRNA viruses in plants (Baulcombe, 1996; Voinnet, 2001), genome defense against endogenous retroelements (Ketting et al., 1999), DNA methylation, chromatin and histone modifications and translational repression (Jones et al., 1999; Mette et al., 2000; Aufsatz et al., 2002a).

1.3 BIOGENESIS OF PLANT MIRNAS

1.3.1 Plant miRNA genes

The loci encoding miRNAs, termed *MIR* genes, are located throughout the genome in regions not associated with known protein coding genes, previously considered to be "junk" DNA (Reinhart et al., 2002). This indicates that most plant miRNAs are derived from their own endogenous genes and from one independent transcript (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001). However, some miRNA genes are found in clusters and are transcribed as one multi-cistronic unit from a single promotor element, but this arrangement is more common in animals than in plants (Lagos-Quintana et al., 2001; Lee et al., 2002; Jones-Rhoades and Bartel, 2004; Seitz et al., 2004; Tanzer and Stadler, 2004; Guddeti et al., 2005). Furthermore, in animals, some *MIR* genes are not transcribed from their own promotors, but are derived from the introns of other protein coding genes (Aravin et al., 2003; Gvozdev et al., 2003; Kogan et al., 2003; Lagos-Quintana et al., 2003; Lim et al., 2003; reviewed by Ying and Lin, 2004; Baskerville and Bartel, 2005). Studies revealed that the primary single stranded RNA miRNA transcript, termed the pri-miRNA, is transcribed by

RNA polymerase II enzymes (Kurihara and Watanabe, 2004; Xie et al., 2005). These transcripts are usually ~ 1 kb in length, polyadenlyated, 5' capped, contain introns and typical TATA-box motifs, characteristic of class II transcription (Aukerman and Sakai, 2003; Kurihara and Watanabe, 2004; Xie et al., 2005). Little is known about the regulation of miRNA transcription, but it can be hypothesized that this regulation would be comparable to that of other protein-coding genes.

1.3.2 Plant miRNA processing

The broad concept of mature miRNA biogenesis is outlined in Figure 1.1. It involves two sequential cleavage steps of the primary transcript by RNase III-like enzymes (Dicer), known as Dicer-like (DCL) in plants and Drosha in animals (Bernstein et al., 2001; Lee et al., 2003). In the nucleus, the pri-miRNA molecule is cleaved by Dicer on each arm of the stem-loop to form the smaller pre-miRNA precursor (Figure 1.1). The pre-miRNA molecule folds to form an imperfect hairpin dsRNA stem-loop structure with the mature miRNA located on one arm (Lau et al., 2001; Lee and Ambros, 2001). Subsequent Dicer cleavage of this molecule on each arm releases a miRNA duplex, containing the mature miRNA and its near reverse complement (miRNA*), from the pre-miRNA stem-loop (Lee et al., 2003). The miRNA and the miRNA* remain together after cleavage, leaving 2-nucleotide 3'-overhangs with 3' hydroxyl and 5' monophosphate ends, characteristic of Dicer cleavage products (Elbashir et al., 2001a).

Dicer homologues have been found in most organisms that undergo RNA silencing (Bernstein et al., 2001; Ketting et al., 2001). *Arabidopsis* and rice have four DCL genes (Ray 1996, Jacobsen 1999) and there appear to be different roles for distinct Dicer enzymes in plants. In plants, DCL1 is specifically involved in miRNA accumulation and is responsible for both cleavage steps in the nucleus (Park et al., 2002; Reinhart et al., 2002; Schauer et al., 2002; Papp et al., 2003; Xie et al., 2003; Xie et al., 2004). The remaining three DCL enzymes
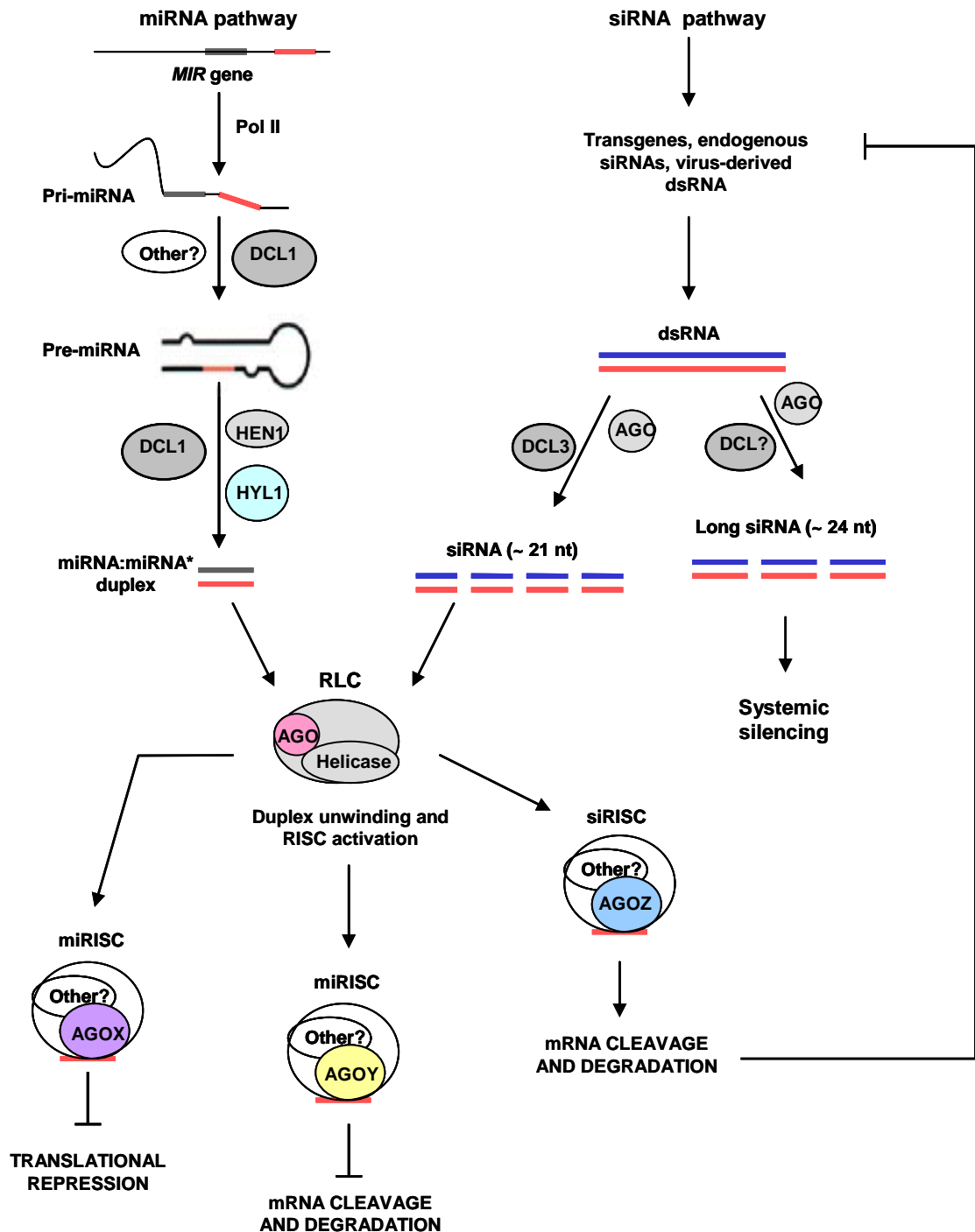
Figure 1.1. Model for the biogenesis of small RNAs in *Arabidopsis*. See text for references. MicroRNA genes are transcribed in the nucleus by an RNA polymerase II to form pri-miRNAs. The pri-miRNAs are cleaved by DCL1 and other cofactors to form pre-miRNA stem-loop precursors. Mature miRNAs are processed from the long pre-miRNA precursor molecules as a miRNA duplex by the action of DCL1, HEN1 and HYL1. In the siRNA pathway, long dsRNA precursor molecules are derived from a variety of sources with the ability to form fold back stem-loop structures and are

9

processed much the same as miRNAs. The long dsRNA molecules are processed by Dicer and various AGOs to form duplex siRNAs. However, two forms of siRNAs are produced, siRNAs that can act much like miRNAs, and long siRNAs that lead to systemic silencing through DNA methylation and transcriptional silencing. SiRNAs and miRNAs are then incorporated into the RLC where duplex unwinding occurs via AGOs and helicases. Only one strand of the mature miRNA or siRNA duplex is then incorporated into the active miRISC or siRISC, respectively. Together with RISC the small RNA is then directed to its complementary mRNA target molecule where it can either result in mRNA cleavage or translational repression. It seems that the functional specificity of the small RNA is determined by the degree of complementarity to its target; the location of the small RNA binding site and the variation of AGO proteins contained within the final RISC. As shown, different AGOs may direct specific functional outcomes, for example AGOX is required for translational repression; AGOY for miRNA directed mRNA cleavage and AGOZ for siRNA mediated cleavage of its original precursor molecule. DCL = Dicer-like; HEN = HUA enhancer 1; AGO = ARGONAUTE; RISC = RNA induced silencing complex; RLC = RISC loading complex (Modified from Bartel et al. 2004).

appear to be involved in the biogenesis of different siRNA molecules (Chan et al., 2004; Vazquez et al., 2004b; Xie et al., 2004; Allen et al., 2005; Gasciolli et al., 2005). The protein structure of the Dicer/DCL enzymes facilitates the processing of dsRNA molecules. They are dimeric proteins with four characteristic domains: a PIWI/ ARGONAUTE/ZWILLE (PAZ) domain, a DExH-box RNA helicase domain, two ribonuclease III motifs, and at least one dsRNA binding domain (Blaszczyk et al., 2001). Structural analysis revealed that each Dicer monomer contains one functional catalytic site, which cleavages dsRNA molecules in ~23 to 28 bp intervals and explains the observed size of siRNAs and miRNAs (Blaszczyk et al., 2001; MacRae et al., 2006). Slight changes in Dicer structure would result in a modification of the spacing between the two catalytic sites, and the size variability of Dicer products seen in *Arabidopsis*, where two species of siRNAs are found, namely long (~24 nt) and short (~21 nt) siRNAs (Hamilton et al., 2002).

The first Dicer cleavage removes the majority of the pri-miRNA stem-loop and defines one end of the mature miRNA within the newly made pre-miRNA. This pre-miRNA then undergoes Dicer cleavage, releasing the mature miRNA as a miRNA: miRNA* duplex molecule. The second Dicer cleavage occurs about two helical turns away from the end of the pre-miRNA molecule, thus determining the opposite end of the miRNA. It is clear that DCL1 in plants cuts preferentially at specific positions in the miRNA stem-loop to release the appropriate mature miRNA molecule (Reinhart et al., 2002). Although the exact mechanism of Dicer recognition is unknown, it is hypothesized that correct Dicer processing of the miRNA is determined by the structure, not sequence, of the precursor molecule to yield the mature miRNA duplex, where flanking secondary structures determine Dicer binding sites (Parizotto et al., 2004). This hypothesis would explain the diversity observed in the sequences of different miRNAs which are processed via the same mechanism.

Two other proteins, HEN1 and HYL1, act in conjunction with Dicer in miRNA biogenesis in plants during precursor processing (Figure 1.1). The *Arabidopsis* HUA ENHANCER 1 (HEN1) protein is a dsRNA methylase with two dsRNA binding domains and a nuclear localization signal (NLS) (Park et al., 2002). This protein has been shown to methylate the 3'-terminal nucleotide of miRNAs and siRNAs and is predicted to protect these molecules from 3' uridylation (Llave et al., 2002b; Park et al., 2002; Li et al., 2005; Yang et al., 2006). The *Arabidopsis* protein, HYPONASTIC LEAVES 1 (HYL1) has two dsRNA-binding domains and has shown to be essential for miRNA accumulation, but not siRNA production (Han et al., 2004; Vazquez et al., 2004a). HYL1 has been shown to preferentially bind dsRNA *in vivo* and is thought to bind miRNA precursors to assist Dicer binding and cleavage (Lu and Fedoroff, 2000; Han et al., 2004). Thus DCL1, HEN1 and HYL1 could act together during the processing of miRNA precursors in the nucleus (Papp et al., 2003). This theory is supported by the fact that DCL1, HEN1 and HYL1 have nuclear localization signals

(Boutet et al., 2003; Papp et al., 2003). After the miRNA:miRNA* duplex is formed in the nucleus, most plant miRNAs are transported to the cytoplasm (Figure 1.1). The export of miRNAs is facilitated by *HASTY (HST)*, which is a member of the nucleocytoplasmic transporter family of proteins (Bollman et al., 2003; Park et al., 2005). *Arabidopsis HASTY* mutants display developmental defects resulting from faulty miRNA production and exhibit reduced miRNA accumulation (Bollman et al., 2003).

1.3.3 The silencing complex

Once the mature small RNA molecules are formed and transported out of the nucleus they associate with a ribonucleoprotein complex, termed the RNA Induced Silencing Complex (RISC) (Hammond et al., 2000). SiRNAs and miRNAs are incorporated into related RISCs, termed siRISC and miRISC, respectively (Hutvagner and Zamore, 2002b; Doench et al., 2003; Tang et al., 2003; Zeng et al., 2003). It was shown that the 2 nt 3'overhangs and 5'-phosphate termini of the small RNAs, formed via Dicer cleavage, are essential requirements for incorporation into RISC (Elbashir et al., 2001b). Once RISC is associated with the a small RNA duplex it is in an inactive form, known as the RISC Loading Complex (RLC) (Elbashir et al., 2001b; Nykanen et al., 2001; Tang, 2005). RISC is activated when small RNA duplex molecules are unwound in the RLC, and only one strand accumulates as the mature miRNA or siRNA, whereas the other arm, termed miRNA* or siRNA* is subsequently degraded (Figure 1.1). It is suggested that RNA helicases associate with RISC and unwind the duplex siRNA/miRNA molecules to allow RISC activation. To date three RNA helicases with possible roles in RNA silencing have been identified: MUT-7 and SMG-2 of *C. elegans,* and SDE3 of *Arabidopsis* (Ketting et al., 1999; Tijsterman et al., 2002b). In plants, SDE3 encodes a RNA helicase partially related to SMG-2 and was found to be essential for transgene induced PTGS (Dalmay et al., 2001). The mature miRNA or siRNA is then sequestered in the

active miRISC or siRISC, which is guided to the target mRNA molecule (Mourelatos et al., 2002; Dostie et al., 2003).

Proteins of the *Argonaute* (AGO) gene family are principle components of RISCs. Members of this family have been linked to gene silencing phenomenon and developmental control in many different organisms (Bohmert et al., 1998; Hammond et al., 2001; Carmell et al., 2002). Argonaute proteins were initially identified in *Arabidopsis*, and to date there are 10 predicted family members, all of which could have different roles in RNA silencing (Fagard et al., 2000; Carmell et al., 2002). AGO4 is shown to be involved in siRNA-directed DNA methylation during transcriptional gene silencing (Zilberman et al., 2003). AGO10/PNH/ZLL and AGO7/ZIPPY are required for proper development (Lynn et al., 1999; Hunter et al., 2003; Moussian et al., 2003), but their function during RNA silencing is unknown. In animals, AGO2 or Slicer directs siRNA mediated cleavage of target mRNAs (Liu et al., 2004). The *Arabidopsis* AGO1 was classified as an essential protein in establishing organ polarity and in leaf development (Bohmert et al., 1998). AGO1 is the *Arabidopsis* homologue of Slicer and binds miRNAs and catalyzes target cleavage *in vivo* (Fagard et al., 2000; Liu et al., 2004; Vaucheret et al., 2004; Baumberger and Baulcombe, 2005). *Ago1* mutants exhibit elevated levels of target mRNA and a null allele of this gene shows a decrease in mature miRNA accumulation (Vaucheret et al., 2004). It has been speculated that the *C. elegans* Argonaute protein, RDE-1, may associate with small RNAs to provide specificity ensuring correct targeting, and stabilization of the small RNA molecules to prevent degradation (Hutvagner et al., 2001).

The Argonaute protein structure includes two conserved regions, the PAZ and Piwi domains (PPD) (reviewed by Carmell et al., 2002). The PAZ domain is a RNA-binding domain that binds single stranded RNAs at the 3' end of the molecule through a hydrophilic cleft (Lingel et al., 2003; Song et al., 2003; Yan et al., 2003; Ma et al., 2004). The Piwi

domain is a putative RNase H and is thought to facilitate miRNA-mediated mRNA cleavage or provide RISC with 'slicer' activity (Song et al., 2004). The PPD domain of Argonaute and the PAZ domain of Dicer may mediate Argonaute-Dicer interactions to allow for the transfer of the siRNA/miRNA molecule from Dicer to RISC (Bernstein et al., 2001; Carmell et al., 2002). This interaction might confer some degree of specificity regarding which pathway the small RNA molecules enter, either sequence-specific mRNA degradation, or translational repression. Documented cases already exist where different Argonaute proteins are involved in different aspects of RNA silencing (Tabara et al., 1999; Hammond et al., 2001; Parrish and Fire, 2001). Argonaute proteins can be divided into functionally distinct subgroups and when associated with a specific RISC could provide the RNA silencing pathway with the specificity to perform distinct functions or modes of action (Hammond et al., 2001; Carmell et al., 2002; Liu et al., 2004; Pillai et al., 2004; Song et al., 2004).

1.3.4 Plant miRNA expression

Some miRNAs are among the most abundant RNAs, where individual animal miRNAs are present at up to 10,000 to 50,000 copies per cell (Lim et al., 2003). Although the expression levels of plant miRNAs have not been quantified to the same extent, it is clear that many are abundantly expressed. Certain miRNAs have been cloned hundreds of times and most are readily detectable by Northern blot (Reinhart and Bartel, 2002; Gustafson et al., 2005). More recently, microarray and real-time PCR technology was adapted to deduce expression profiles of plant miRNAs (Axtell and Bartel, 2005; Chen et al., 2005; Shi and Chiang, 2005). Some miRNAs are broadly expressed, whereas others are expressed most strongly in particular organs or developmental stages (Reinhart et al., 2002; Axtell and Bartel, 2005; Lu et al., 2005). *In situ* hybridization experiments has provided more precise data on the localization of a few plant miRNAs (Chen, 2004; Juarez et al., 2004; Kidner and Martienssen, 2004), or from miRNA-responsive reporters (Parizotto et al., 2004). Little is known about the transcriptional

or post-transcriptional regulation of miRNA expression, although expression patterns of miRNA promoter reporter constructs have been described for miR160 (Wang et al., 2005) and miR171 (Parizotto et al., 2004).

The first indication for such roles came from bioinformatic miRNA and target gene predictions and miRNA cloing from stressed *Arabidopsis* plants, which revealed new miRNAs that had not been cloned previously from plants grown in normal conditions (Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004; Fujii et al., 2005). Levels of several miRNAs are responsive to phytohormones or growth conditions; miR159 levels are enhanced by gibberellin (Achard et al., 2004), miR164 is induced by certain auxin treatments (Guo et al., 2005b), and miR393 levels are increased by a variety of stresses (Sunkar and Zhu, 2004). The dependence of miR395 and miR399 levels on growth conditions is particularly striking. For example, miR395 is undetectable in plants grown on standard medium, but induced over 100-fold in sulphate-starved plants (Jones-Rhoades and Bartel, 2004; Adai et al., 2005; Allen et al., 2005). miR395 targets ATP sulfurylases (APS) that catalyze the first step of inorganic sulphate assimilation and the accumulation of APS1 mRNA is decreased under low-sulphate stress (Sunkar and Zhu, 2004). Similarly, miR399 is strongly and specifically induced only in plants grown under low-phosphate stress (Fujii et al., 2005; Bari et al., 2006). miR399 targets a ubiquitin-conjugating enzyme (UBC), and UBC mRNA levels are decreased during low-phosphate stress, which is important to induce the phosphate transporter gene AtPT1 and attenuate primary-root elongation (Chiou et al., 2005; Fujii et al., 2005). Other miRNAs are likely to have roles during stress based on the function of their targets or their expression patterns. For example, miR398 targets two copper dismutase enzymes that protect cells against harmful oxidative radicals produced during stress (Jones-Rhoades and Bartel, 2004). Furthermore, in poplar trees, miR408 expression is induced by tension and compression stresses in xylem tissues, suggesting that this miRNA has a critical role in the structural and

15

mechanical fitness of woody plants (Lu et al., 2005). Although certain miRNAs are conserved across a wide range of species, there are also species-specific miRNAs that should control species-specific developmental events, such as miRNAs regulating flowering time that will be absent from non-flowering plants. With all the different genes and expression patterns, it is reasonable to propose that every cell type at each developmental stage might have a distinct miRNA expression profile.

1.4 MECHANISM OF ACTION

Differences in the role of miRNAs in plants and animals have provided insights into the mechanism of action of small RNAs. MicroRNAs regulate gene expression at the post-transcriptional level by two known mechanisms, mRNA cleavage and translational repression. Recently it has been shown that miRNAs can also regulate gene expression transcriptionally by leading to chromosomal modifications (Bao et al., 2004; Eshed and Bowman, 2004), however this topic is beyond the scope of this review and will not be discussed here.

Initially it was thought that there were distinct pathways for miRNA-mediated gene regulation in plants and animals, where the predominant function in animals is translational repression, and mRNA cleavage was mostly seen in plants. The reason for this early distinction can be explained by differences observed in miRNA characteristics between plants and animals and not that either pathway was unique. Most animal miRNAs have low levels of complementarity to their targets, bind to multiple sites in the 3'UTR and the majority guide translational blockage (Ambros et al., 2003). In contrast, plant miRNAs exhibit near perfect base pairing to their targets, bind at a single site within the coding region and were only shown to result in mRNA cleavage. The discovery of plant miRNAs that could direct translational repression (Aukerman and Sakai, 2003; Chen, 2004), and animal miRNAs which could lead to mRNA cleavage (Hutvagner and Zamore, 2002b; Yekta et al., 2004), disputed

16

this theory of different mechanisms for miRNA regulation of gene expression in plants and animals.

It has now been shown that one of the main factors determining how miRNAs will exert their function depends on the degree of complementarity between the miRNA and its target mRNA (Doench et al., 2003; Zeng et al., 2003). The general rule of how a small RNA will exert its function emerged that if the miRNA and its target molecule bind with imperfect complementarity the result is translational repression, whereas if the miRNA binds with near perfect complementarity to its target mRNA, cleavage occurs.

Further studies have revealed insights into the connection between miRNA characteristics, and the dynamics of target binding, in order to deduce mechanisms of action. It appears that there is a miRNA core-binding region (bases 2 – 6) that requires perfect complementarity to the mRNA target and only this region is initially presented by RISC to the mRNA target as a prearranged 6 – 7 nt A-form helix to enhance the binding affinity of the small RNA and mRNA (Wightman et al., 1993; Lewis et al., 2003; Lim et al., 2003). From this point it appears that low complementarity of the rest of the small RNA to its target is sufficient for translational repression. However, if there is a large degree of base pairing after the core region, especially in the centre and 5'end of the miRNA, target cleavage will occur (Lewis et al., 2003; Mallory et al., 2004b).

1.4.1 mRNA cleavage

If the miRNA directs mRNA target cleavage, the mechanism is similar to that of siRNA-directed cleavage. The mature miRNA is incorporated into miRISC and binds to its complementary mRNA molecule. Endonucleolytic cleavage of the mRNA occurs between the 10th and 11th base pairs (Elbashir et al., 2001c; Hutvagner and Zamore, 2002a; Llave et al., 2002b; Kasschau et al., 2003; Palatnik et al., 2003; Tang et al., 2003), and is performed by a member of the AGO family, or SLICER, as discussed in the previous section (Liu et al.,

2004). It is proposed that the PAZ and Piwi domains of AGO allow interaction between the miRNA and the target mRNA, where the 3' end of the miRNA is held in the groove of the PAZ domain, which then aligns with the target mRNA held in the Piwi domain. This interaction ensures that the phosphate between the 10th and 11th residues of the mRNA falls near the nuclease catalytic site and cleavage occurs (Lingel et al., 2003; Song et al., 2003; Yan et al., 2003).

The initial mRNA cleavage allows for rapid degradation of the target molecule and the mature miRNA is subsequently released intact, re-incorporated into RISC, and guides cleavage of additional mRNA molecules (Hutvagner and Zamore, 2002a; Tang et al., 2003). The mRNA cleavage products are further degraded starting from the initial cleavage sites. The 5' segment of the mRNA molecule could be degraded by the exosome (van Hoof and Parker, 1999). And degradation of the 3' end of the mRNA is carried out by the *Arabidopsis* EXORIBONUCLEASE4 (AtXRN4), a homologue of the yeast mRNA degrading exoribonuclease, Xrn1p, in a 5' to 3' direction (Souret et al., 2004).

1.4.2 Translational repression

Translation repression as a result of miRNA regulation occurs mainly in animals, and there is only one example of this form of regulation in plants (Aukerman and Sakai, 2003). In animals translation repression occurs when multiple miRNAs bind to the 3'UTR of their target mRNA molecules and inhibit effective translation to result in decreased protein levels (Wightman et al., 1993; Olsen and Ambros, 1999; Reinhart et al., 2000). The exact mechanism of translation repression remains vague, however some theories have emerged. Studies of the *C. elegans* miRNA, *lin-4* identified that bound miRNA molecules could cause slowing/stalling of ribosomes on the mRNA, and in this way translation is repressed after initiation and does not alter the number of ribosomes present on a certain transcript (Olsen and Ambros, 1999). Alternatively, translation may proceed at the normal rate, however, the

synthesised protein product may not contain essential elements encoded in the 3'UTR and could be targeted for degradation and translation is in effect "silent". The reason for multiple miRNA complementary sites on one target mRNA may be explained by the necessity for the cooperative action of multiple RISCs during this event (Lee et al., 1993; Reinhart et al., 2000; Lin et al., 2003). More than one RISC would cause a greater hindrance of the ribosomes along the mRNA transcript. Humphreys et al (2005) show that in animals miRNAs can also affect the initiation of translation by interfering with the function of the mRNA cap structure and the poly(A) tail. The hypothesis is that miRNAs interfere with elongation initiation factors (eIF4E) during initiation by either interfering with cap recognition or blocking cap function once it is bound (Humphreys et al., 2005; Pillai et al., 2005).   This type of gene regulation, where translation is simply repressed may have some advantages over mRNA cleavage as simply releasing the miRNA can stop repression, rather than destruction and synthesis of new mRNAs (Ke et al., 2003).

1.5 MICRORNAS VS SIRNAS

It was initially thought that the miRNA and siRNA pathways were distinct, however, it is now known that miRNAs and siRNAs cannot be distinguished on the basis of their biochemical composition, general mechanism of biogenesis or function. One of the unifying features of all small RNAs is their small (21–26nt) size and their ability to act as the specificity determinants for RNA silencing via homologous sequence interactions (Hamilton and Baulcombe, 1999; Parrish et al., 2000; Parrish and Fire, 2001; Tijsterman et al., 2002b). The general mechanism of miRNA and siRNA biogenesis is highly similar, whereby dsRNA precursor molecules are cleaved via Dicer into small RNAs, both showing characteristic 5' phosphate and 3' hydroxyl termini. In both cases one strand of the small RNA molecule is transferred to a RISC and is subsequently directed to its target mRNAs. Thus, both siRNAs and miRNAs operate at a

post-transcriptional level of silencing; both are processed from larger dsRNA molecules, and both require members of the Dicer and ARGONAUTE protein families (Carmell et al., 2002).

The key distinction between miRNAs and siRNAs lies in their genomic origin and evolutionary conservation. As shown in Figure 1.1, siRNAs are processed from perfect dsRNA duplexes which are almost randomly cleaved into multiple siRNA molecules, whereas a single miRNA is derived from one arm of an imperfect hairpin precursor molecule (Hannon, 2002). Furthermore, miRNAs are encoded by their own endogenous *MIR* genes and regulate other genes elsewhere in the genome, and their sequences are almost always conserved between related organisms. The dsRNA siRNA precursor molecule can be derived from a variety of dsRNA sources, including transgenes, viruses and transposons. SiRNAs then mediate silencing of the same gene from which they originate and their sequences are rarely conserved between related organisms as their sequence depends on the gene from which they are derived. This may explain why miRNAs are evolutionary conserved, whereas siRNAs are not. Since siRNAs target the same mRNA molecule from which they are derived base-pairing between siRNAs and their targets are perfect and changes in target sequences would result in a corresponding change in siRNA sequence (reviewed by Ambros et al., 2003; reviewed by Finnegan and Matzke, 2003). However, a mutation in a miRNA would most likely not be accompanied by a corresponding change in the target gene (reviewed by Bartel, 2004).

It was initially believed that siRNAs and miRNAs directed distinct functional pathways and each could only function in one of the "routes" of RNA silencing. Where siRNAs only result in mRNA degradation and miRNAs can only lead to translational repression. However there is strong evidence to support the hypothesis that these two groups and pathways are functionally interchangeable (Llave et al., 2002b; Rhoades et al., 2002; Doench et al., 2003; Kasschau et al., 2003; Ke et al., 2003; Palatnik et al., 2003; Zeng et al., 2003; Tang, 2005). This means that miRNAs can lead to mRNA degradation and siRNAs can

also function as miRNAs by resulting in translational repression. This was shown by synthetic siRNA studies which showed that siRNAs can indeed function as miRNAs if they target 3'UTRs with partial complementarity (Doench et al., 2003).

It seems that the downstream biological role of siRNAs is more diverse than what is known for miRNAs, which show strong functions in development. The biological role of siRNAs is to function as a host defence system against foreign invading nucleic acids, whereby transgenes, injected RNA, viruses or transposons are targeted for silencing. In plants it appears that there are two functionally distinct groups of siRNAs, namely short siRNAs of ~ 21-22nt, derived from viruses or transgenes and long siRNAs of ~ 24nt, derived from transposons or transgene promoters. The long siRNAs may play a role in transposon and transgene silencing via chromatin modifications (Hamilton et al., 2002). Furthermore, siRNAs might modulate gene expression by guiding sequence-specific promotor methylation in plants (Mette et al., 2000; Aufsatz et al., 2002b), heterochromatin formation in fission yeast (Volpe et al., 2002; Grewal and Moazed, 2003), transposon silencing in *C.elegans* (Ketting et al., 1999; Tabara et al., 1999), cytosine methylation, histone methylation and epigenetic modifications of DNA (Wassenegger et al., 1994; Matzke et al., 2001; Zilberman et al., 2003). These molecules are also known to be the signal/primers for the amplification and systemic spread of RNA silencing (Voinnet and Baulcombe, 1997; Hamilton et al., 2002). Thus, it is clear that some distinctions can be made about the roles miRNAs and siRNAs partake in gene silencing. However, it may be that the differences noted between the two pathways may simply reflect the fact that they were discovered independently, and that they have a different evolutionary histories.

1.6 MICRORNAS IN PLANT DEVELOPMENT

The discovery of miRNAs prompted the next objective to determine which genes are regulated by each of these molecules to deduce their precise role in the cellular environment.

Most miRNA targets were initially identified using computational predictions based on sequence complementarity to sequenced genomes. Target predictions are easier in plants as the miRNAs exhibit near perfect complementarity to target mRNA sequences, with most showing less than four mismatches (Rhoades et al., 2002). Initial studies identified that a large proportion of miRNA targets are transcription factors families, especially those involved in cell fate determination, cell growth and tissue patterning (Rhoades et al., 2002; Mallory and Vaucheret, 2004; Mallory et al., 2005) as shown in Table 1.1. Further analysis of miRNAs and their cellular functioning indicate that they are key participants in a number of gene regulatory networks involved in developmental processes (Pasquinelli et al., 2000; Lagos-Quintana et al., 2001; Johnson et al., 2003; Ke et al., 2003). Mutant studies in *Arabidopsis,* where either specific miRNAs or genes involved in miRNA biogenesis are knocked out, have provided much of the evidence for the elucidation of miRNA functions in plants. Plants defective in miRNA biogenesis, such as *dcl1*, *hen1* and *hyl1* mutants, show decreased accumulation of mature miRNA, and exhibit dramatic developmental phenotypes, ranging from defects in floral development to defects in leaf morphology (Lu and Fedoroff, 2000; Schauer et al., 2002; Boutet et al., 2003; Palatnik et al., 2003). Recent studies revealed roles for miRNAs in more specific aspects of plant development, such as control of leaf morphogenesis (Palatnik et al., 2003), regulation of flowering time and floral organ identity (Aukerman and Sakai, 2003), control of shoot, floral and auxiliary meristem formation (Lu and Fedoroff, 2000; Boutet et al., 2003), the establishment of organ polarity, vascular development and in hormone and stress responses (Jones-Rhoades and Bartel, 2004; Wang et al., 2004b; Lu et al., 2005; Ko et al., 2006). However, recent computational predictions have identified numerous targets for newly identified miRNAs with more diverse biological roles. These include miRNA biogenesis, cellular metabolic functions and stress responses (Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004; Lu et al., 2005). It is now accepted that

Table 1.1. Certain *Arabidopsis* miRNAs and their putative targets*

microRNA family	Predicted target gene family	Predicted Target Gene	Predicted target gene function
miR156	Squamosa-promoter Binding Protein (SBP)-like transcription factors	10 *SPL* genes	Floral meristem identity and flowering time
miR157	SBP-like proteins	9 *SPL* genes	Floral meristem identity and flowering time
	Putative DEAD-box RNA helicase		MicroRNA biogenesis
	Unknown proteins		
miR159a	MYB transcription factors	5 MYB genes, *GAMYB*	Plant growth, anther development and flowering time
miR159b	MYB transcription factors	*3 MYB* genes, *MYB33, MYB65*	As above
miR-JAW	TCP transcription factors	*TCP2, TCP3, TCP4, TCP10, TCP24*	Cell division , leaf development, and embryo patterning
miR160	Auxin response factors	*ARF10, ARF16, ARF17*	Plant response to phyothormones during development
miR161	Pentatricopeptide repeat proteins	*9 genes*	Organelle metabolism
miR162	DICER	*DCL1*	Biogenesis of miRNAs and control of floral meristem development
miR163	SAM-dependant methyltransferases	5 genes	
miR164	NAC domain proteins	*CUC1, CUC2, NAC1,* 2 others	Control of boundry size in meristems and formation and separation of embryonic, vegetative and floral organs
miR165	HD-ZIP transcription factors	*PHB, PHV, REV, ATHB8*	Embryo patterning, postembryonic meristem initiation, axial meristem initiation, vascular development, leaf polarity and meristem size
miR166	HD-ZIP transcription factors	*ATHB15*	As above
miR167	Auxin response factors	*ARF6, ARF8*	Auxin signalling
miR168	ARGONAUTE	*AGO1*	miRNA biogenesis and pathway, stem cell function and organ formation
miR169	CCAAT-binding factor (CBF)-HAP2-like proteins		Promotor DNA binding protein in Eukaryotes
miR170	GRAS domain transcription factors (SCARECROW-like)	*SCL6-II, SCL6-III, SCL6-IV*	Radial patterning in roots
miR171	GRAS domain transcription factors (SCARECROW-like)	*SCL1, SCL6-II, SCL6-III, SCL6-IV*	Radial patterning in roots
miR172	APETALA2-like transcription factors	*AP2, TOE1, TOE2, TOE3*	Specification of flower organ identity and flowering time
miR393	F-box proteins	*TIR1* and 3 others	Component of SCF ubiquitin-ligase complexes for ubiquitin-mediated proteolysis
	bHLH transcription factor	1 gene	Circadian rhythm control by light
miR395	ATP sulfurylase (3)	*APS4* and 2 others	Sulfer metabolism
miR396	Growth Regulating Factor (GRF) transcription factors	*GRL1, GRL2, GRL3, GRL7,GRL8, GRL9*	Growth regulation
	Kinesin-like protein B	1 gene	
miR397	Laccase	3 genes	Normal cell wall structure and integrity of xylem fibers
	Beta-6 tubulin	1 gene	Involved in microtubule based processes and response to cold
miR398	Copper superoxide dismutases	*CSD1, CDS2*	Involved in programmed cell death and apoptosis

*constructed from information obtained from www.sanger.ac.uk/Software/Rfam/miRNA, www.TAIR.org, Rhodes et al (2002) and Bonnet et al (2004).

miRNAs are key regulators in a vast number of very diverse developmental pathways in plants.

The number of processes under miRNA control in animals is rapidly expanding. MicroRNA functions in animals include control of developmental timing events in *C. elegans* (Lee et al., 1993), cell division, apoptosis, and fat metabolism in *Drosophila* (Brennecke et al., 2003; Xu et al., 2003), neuron differentiation in *C. elegans* and mammals (Johnston and Hobert, 2003; Sempere et al., 2004), hematopoietic lineage differentiation in mammals (Chen et al., 2004), angiogenesis (Yang et al., 2005), and mammalian spermatozoa-specific miRNAs have been identified with roles in germ-cell transitions (Ostermeier et al., 2005; Yu et al., 2005). Furthermore miRNAs have been shown to be involved in human cancers (Calin et al., 2004), viral infections and even some HIV encoded miRNAs have been identified (Bennasser et al., 2004; Pfeffer et al., 2004; Omoto and Fujii, 2005). Some well characterized miRNAs and their targets will be discussed below with reference to their roles in specific developmental processes in plants.

1.6.1 Organ polarity and meristem identity

Plant organs are highly complex structures that exhibit polarity in the organization of specific cell types, such as in the vascular tissue of leaves and stems. For example, in most leaves photosynthetic cells are located on the adaxial (top) surface and stomata are located on the abaxial (bottom) surface and in vascular tissues cambial cells differentiate into xylem cells to the inside and phloem cells to the outside of the stem.

Members of the Class III Homeodomain-leucine zipper (HD-ZIP) transcription factor family, *PHAVOLUTA (PHV), PHABULOSA (PHB),* and *REVOLUTA (REV)* are known targets of miR165/166 (Rhoades et al., 2002; Emery et al., 2003; Tang et al., 2003). These genes encode transcription factors known to regulate axiliary meristem initiation and leaf polarity (Emery et al., 2003). miR165/166 negatively regulates *PHV* and *PHB* mRNAs by guiding sequence-specific cleavage at specific stages during leaf development and differentiation (Tang et al., 2003; Zhong and Ye, 2004). Dominant mutations in these genes

24

that disrupt the miRNA binding site render the transcripts resistant to cleavage and result in severe developmental defects (McConnell et al., 2001). These defects include adaxialization of leaves (where cell types normally found on the adaxial surface are seen on the abaxial surface) and radialization of vascular tissues (McConnell et al., 2001; Otsuga et al., 2001a; Juarez et al., 2004). The *Nicotiana PHV* gene regulates adaxial identity, vascular patterning and meristem structure and is also a target of miRNA regulation. In tobacco, miR165/166 is responsible for restricting *PHV* expression to differentiating and/or differentiated tissues and to keep meristem tissues in an undifferentiated state (McHale and Koning, 2004). Figure 1.2 represents a model for the specification of adaxial/abaxial polarity in *Arabidopsis* leaves. Expression of *PHV* and *PHB* in leaf primordium cells close to the meristem results in the initiation of a transcription program that specifies adaxial fate. Clearing of *PHV* and *PHB* transcripts by miR165/166-guided degradation in cells further away from meristem cells specifies abaxial fate (Carrington and Ambros, 2003). *In vivo* hybridization assays showed that miR165 is exclusively expressed in the abaxial side of leaves (Kidner and Martienssen, 2004), which is opposite to that of *PHB* expression (McConnell et al., 2001). Thus, miR165/166 functions to clear *PHB*, *PHV* and *REV* transcripts from abaxial cells to ultimately allow differentiation into abaxial tissues.

Another member of this family, *ATHB15*, has been implicated to have roles during vascular development and is exclusively expressed in vascular tissues (Ohashi-Ito and Fukuda, 2003; Kim et al., 2005; Ohashi-Ito et al., 2005). This gene was recently identified as a target of miR166 by computational predictions (Rhoades et al., 2002). Kim et al (2005) confirmed miR166 directed cleavage of *ATHB15* mRNA is essential for the regulation of vascular development in the *Arabidopsis* infloresence stem. They showed that a *MIR166a* gain-of-function mutant exhibited decreased *ATHB15* mRNA levels and accelerated vascular cell differentiation from the cambial/procambial cells indicating alteration in the vascular
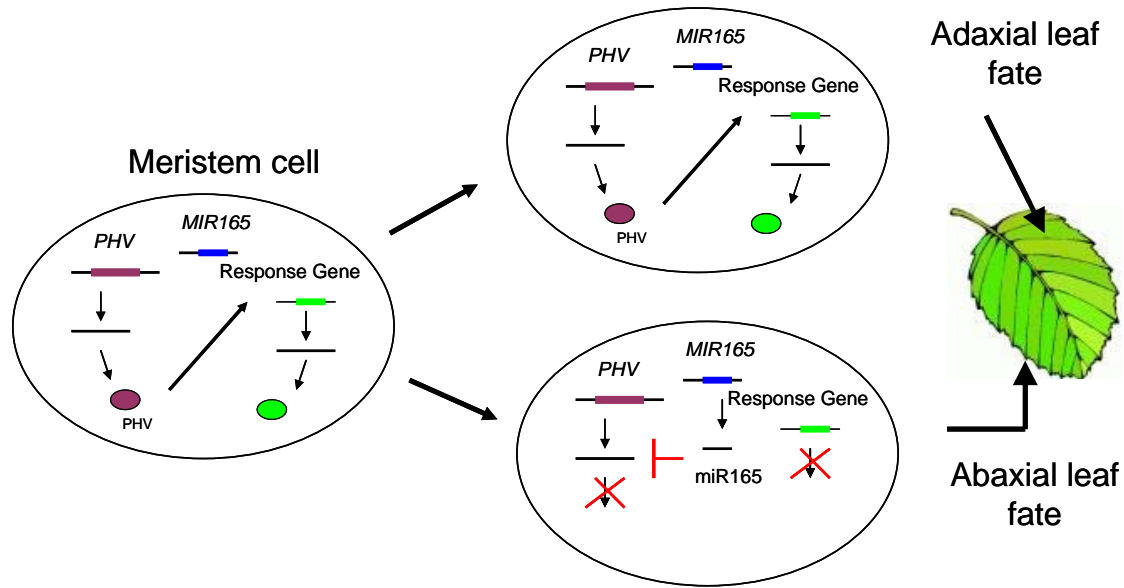
Figure 1.2. Structure and Function of *Arabidopsis* miR165. The miRNA, miR165/166, regulate the *PHV* and *PHB* transcription factors involved in determining leaf fate. During the differentiation of adaxial leaf cells, miR165/166 is not expressed which allows the expression of the *PHV/PHB* genes. During the development of the abaxial side of the leaf, *PHV/PHB* expression is inhibited through miRNA-directed silencing (Carrington and Ambros, 2003).

system. Class III HD-ZIP genes and their regulation by miR165/166 is highly conserved in green plants (Floyd and Bowman, 2004). This regulation of Class III HD-ZIP genes by miR165/166 is conserved in *Arabidopsis*, rice, Maize, gymnosperms, ferns, lyocopods and bryophytes which may be an indication that miR165/166 regulation of this protein family has been occurring of at least 400 million years (Floyd and Bowman, 2004).

Other targets of plant miRNAs include genes responsible for the specification of embryonic, vegetative and floral development. miR164 targets five members of the NAC-domain family of proteins. One member, the *CUP-SHAPED COTYLEDONS (CUC)* genes, regulate organ boundaries and positively regulate *SHOOTMERISTEMLESS1 (STM1)* (Aida et al., 1997; Aida et al., 1999). *CUC1* and *CUC2* regulation by miR164 acts to control lateral

organ boundaries and organ separation (Laufs et al., 2004; Mallory et al., 2004a). Another member of this family, *NAC1,* transports auxin signals for promoting lateral root emergence, and it was recently demonstrated that miR164 itself is induced by cellular auxin levels (Guo et al., 2005a). Thus, through negative regulation of *NAC1* mRNA, miR164 may function in maintaining auxin homeostasis by downregulating auxin signals during lateral root development (Guo et al., 2005a). Conserved miR164 binding sites have been identified in NAC-domain containing genes in *Arabidopsis*, snapdragon (Weir et al., 2004) and petunia (Souer et al., 1996) which suggests that regulation by miR165 may be specific to flowering plants.

1.6.2 Floral organ identity, flowering time and transitions in the plant life cycle

The *Arabidopsis APETELA2 (AP2)* gene is known to be regulated by miR172 (Aukerman and Sakai, 2003; Chen, 2004). The *AP2* transcription factor family is involved in the regulation of flowering time and floral organ identity (Jofuku et al., 2005). Loss of miR172 regulation results in early flowering, and overexpression disrupts floral organ identity (Chen, 2004). miR172 regulation of *AP2* was the first example of miRNA control through translational repression in plants. Although miR172 binds with near perfect complementarity to the *AP2* transcript, translational repression occurs as mRNA levels remain constant but protein levels decrease (Aukerman and Sakai, 2003; Chen, 2004). It was further shown that two members of the *AP2*-like gene family are also targets of miR172. These genes, *TOE1* and *TOE2*, are floral repressors and their expression results in delayed or inhibited flowering, and thus regulation by miR172 leads to flowering as the floral repressors are no longer active (Aukerman and Sakai, 2003). The Maize *AP2*-like homologue, *Glossy15*, has also been shown to be regulated by miR172 (Lauter et al., 2005). *Glossy15* regulates the transition from juvenile to adult leaf identity, where G*lossy15* keeps the leaf in juvenile form and miR172-directed degradation results in transition to adult leaf fate (Lauter et al., 2005). Thus, miR172

has a proposed role in the temporal regulation of flowering time, floral meristem development, and in regulating vegetative phase changes in higher plants.

miR156 targets different members of the *SQUAMOSA PROMOTOR BINDING* (SBP) protein family, which are thought to regulate floral meristem identity genes in *Antirrhinum*. miR156 is upregulated once flowering initiates and regulates members of the *SBP-like (SBL)* gene family which suppress flowering in *Arabidopsis* (Schmid et al., 2003). And the target of miR159 is a MYB transcription factor that can bind to the promoter of the floral meristem identity gene, *LEAFY* (Rhoades et al., 2002). There are a large number of MYBs in *Arabidopsis* and collectively they have roles in controlling cell shape, regulation of secondary metabolism, resistance to disease and hormone signalling (Jin and Martin, 1999; Millar and Gubler, 2005). Mutant studies have shown that miRNA-directed regulation of 11 MYBs is essential for anther development, leaf development, flowering time and hormone signalling (Palatnik et al., 2003; Achard et al., 2004; Han et al., 2004).

The miR170 and miR171 microRNAs target members of the *SCARECROW-like* (SCL) family of transcription factors (Rhoades et al., 2002; Jones-Rhoades and Bartel, 2004). *SCLs* are members of the *GRAS (GIBBERELLIN-INSENSITIVE (GIA), REPRESSOR of GIA, SCARECROW (SCR))* gene family. The *SCL* family is thought to have roles in radial root patterning, signaling by the phytohormone, gibberellin, and light. This was one of the first plant miRNAs which was shown to direct the developmentally controlled cleavage, and not translational repression, of the *SCL* mRNA (Llave et al., 2002b). Furthermore, SCR is known to have roles in regulation of asymmetric cell division and in regulation of cell differentiation in response to gravity (Fukaki and Tasaka, 1999; Tasaka et al., 1999; Helariutta et al., 2000; Nakajima et al., 2001).

1.6.3 MicroRNA Biogenesis

Two predicted plant miRNA targets that are not transcription factors are proteins involved in miRNA biogenesis, namely *DCL1* and *AGO1*. This suggests that in *Arabidopsis*, the miRNA biogenesis machinery is regulated by a negative feedback mechanism. The plant Dicer homologue, *DCL*, has been indirectly implicated to play a role in the development of embryos, leaves, and floral meristems in *Arabidopsis* (Jacobsen et al., 1999). The *dcl* mutants show disruption of embryo development, delay in flowering time, over-proliferation of floral meristems, and accumulation of miRNA precursor molecules (Park et al., 2002). *Arabidopsis* DCL1 is maintained at relatively low levels in wild-type plants as it is negatively regulated by miR162 through mRNA cleavage and degradation (Xie et al., 2003). In *dcl1* mutants, defective dcl1 proteins result in deficiencies of miR162, which leads to abnormally high levels of *DCL1* mRNA. This may be a highly effective mechanism of miRNA transcriptional regulation to ensure that only certain amounts of miRNAs are produced. Once enough miRNA molecules are present in the cell, excess, unbound miRNAs will result in cleavage of *DCL1* mRNA leading to a reduction of DCL1 levels, and in response, a reduction in miRNA levels. The binding of miR162 to DCL1 is unique in that there is a 1nt bulge in the binding site from position 7 to 8, however cleavage still occurs in the center of the binding site (Xie et al., 2003). This type of binding was also seen for miR396 which targets members of the Growth regulating factor (GRL) transcription factor family (Jones-Rhoades and Bartel, 2004).

Arabidopsis AGO1 is required for both axillary shoot meristem formation and leaf development, and is a target for miR168 (Rhoades et al., 2002). In AGO1 mutants the severe developmental defects observed are a result of the failure of biogenesis of multiple miRNAs involved in a number of developmental pathways, as was shown in mutants where a variety of miRNA targets are overexpressed (Vaucheret et al., 2004). AGO1 is the predicted target of miR168, thus its expression is controlled by a feedback mechanism. Furthermore, miR157

might target an uncharacterized DEAD-box RNA helicase, and miR163 might target uncharacterized members of a plant family of methyltransferases, both of which are predicted to function during miRNA biogenesis as shown in Figure 1.1 (Ross 1999).

1.6.4 Stress-response genes and non-transcription factors

As computational predictions of plant miRNAs become more in depth many more putative targets have been identified that could have been missed through cloning strategies simply because of frequency of cloning and with the identification of stress-responsive miRNAs it is reasonable to believe that certain miRNAs can be overlooked in samples not undergoing certain stress conditions. It seems that microRNA regulation of hormone signaling pathways plays a further role in developmental events, and many miRNAs themselves are developmentally or environmentally regulated (Jones-Rhoades and Bartel, 2004; Wang et al., 2004a). Some of these examples have already been mentioned in previous sections, however further examples will be mentioned here.

The phytohormone auxin can either stimulate or inhibit cell growth depending on its concentration and location within the plant. Auxin has roles in stem elongation, phototropic and gravitropic responses, and lateral and adventitious root formation. Transcription factor families, such as Auxin Response Factors (ARFs) and NAC-domain transcription factors, mediate the effects of auxin on plant development. ARFs are DNA binding proteins found specifically in plants and control auxin regulated transcription (Guilfoyle et al., 1998). Computational prediction and experimental analysis of plant miRNA targets revealed that miR160 is complementary to ARF 10, ARF16 and ARF17 (Rhoades et al., 2002; Mallory et al., 2005), while miR167 is complementary to ARF6 and ARF8 (Rhoades et al., 2002; Bartel and Bartel, 2003). Furthermore, miR393 is predicted to target the *TRANSPORT INHIBITOR RESPONSE1 (TIR1)* gene, which is conserved in rice, *Arabidopsis*, poplar, *Medicago* and *Lotus* (Bonnet et al., 2004). TIR1 functions by binding to Auxin/indole-3-acetic acid (IAA)

30

proteins that are found in the E3 ubiquitin-ligase SCF complex, which is involved in protein degradation (Xie et al., 2000; Downes et al., 2003). Furthermore, mmicroRNAs have also been predicted to target members of the ubiquitination pathway involved in protein degradation (Bonnet et al., 2004; Sunkar and Zhu, 2004). miR393 and miR399 are predicted to target ubiquitin conjugating enzyme E2 and ubiquitin ligase/TIR1 (Sunkar and Zhu, 2004).

Another hormone responsive microRNA is miR159. miR159 itself is positively regulated by the phytohormone gibberellic acid (GA) and subsequently negatively regulates *GAMYB*, a positive regulator of *LEAFY*, resulting in the regulation of flowering time and anther development, and negatively regulates the GA response inhibitor proteins of the DELLA family (Achard et al., 2004). There may also be feedback from GAMYB into regulation of the miRNA as seen with AGO1 and the DELLA proteins also mediate auxin and ethylene responses. Thus this regulatory stem loop may allow the coordination of several hormone signaling pathways (Fu and Harberd, 2003).

Three groups have identified miRNAs that are formed in response to mechanical and environmental stress conditions in two plant species (Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004; Lu et al., 2005). miR395 regulates ATP sulfurylases and is itself regulated by cellular sulfur concentrations indicating that this miRNAs may also mediate environmental responses (Jones-Rhoades and Bartel, 2004). miR395 targets three ATP sulfurylases, namely APS1, APS3, and APS4, which are enzymes that catalyze steps in sulphate metabolism (Leustek, 2002). Furthermore, the poplar ptrmiR139 is predicted to target a sulphate transporter protein (AST) that mediates sulfur uptake in plants, however this miRNA is totally different from the *Arabidopsis* miR395, which indicates one of the first examples of a species-specific miRNA (Lu et al., 2005). Newly identified *Arabidopsis* miR399 targets CSD1 and CSD2 are known copper superoxide dismutases. These enzymes are expressed in response to conditions of oxidative stress and protect the cell against oxygen radicals (Kliebenstein et al.,

1998). The identification of these examples strongly suggests that miRNA expression may be modulated by external metabolites as well. Other putative targets include laccases, cytochrome C oxidases, enzymes involved in organelle metabolism and photosynthesis such as PPR transcription factors and plastocyanin, and genes involved in protein metabolism such as spliceosomal proteins and peptide chain release factors (Bonnet et al., 2004; Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004).

1.7 RECENT DEVELOPMENTS

The fact that certain miRNAs have low abundance, and others exhibit temporal-, environmental- and tissue-specific expression patterns, makes the experimental identification of these miRNAs difficult. This is due to the fact that large amounts of sequence data from virtually any condition is necessary, and miRNAs expressed at low levels may not be detected by the classical method of sequencing small RNA libraries. DNA sequencing technologies have improved dramatically in the last decade and numerous whole-genome sequences are now available. The increasing number of large-scale DNA sequencing projects in recent years has driven a search for alternative methods to reduce time and cost. Furthermore, genome sequence data is not enough to identify transcript levels and expression patterns. Recently, companies such as Solexa and 454 Life Sciences, have been developing new approaches that rely on computer algorithms to virtually assemble hundreds of thousands of relatively short DNA fragments to produce the final sequence.

Massively Parallel Signature Sequencing (MPSS) technology (Lynx Therapeutics, Inc.) has revolutionized the process of high-throughput DNA sequencing. MPSS is a sequencing-based technology that uses a unique method to generate sequence information from millions of DNA or cDNA fragments per library. This method also provides a method to simultaneously identify and quantify gene expression levels of many molecules in a single sample. MPSS produces short (16-20 nt) sequence tags from a defined position within an

mRNA, and measures the relative abundance of these tags in a library to estimate gene expression profiles. At an extremely large scale, the MPSS technology approach eliminates the need for individual sequencing reactions and the physical separation of DNA fragments required by conventional sequencing methods (Brenner et al., 2000). This depth provides a quantitative assessment of transcript abundance, while greatly increasing the likelihood of discovering novel transcripts. Gene expression differences can be determined using comparisons across multiple samples, as with DNA microarrays or other gene expression platforms (Nakano et al., 2006).

In an advance for the discovery of alternative DNA sequencing technology, 454 Life Sciences Corp. has developed the 454 sequencing technology (http://www.454.com) which is a massively-parallel sequencing-by-synthesis (SBS) system capable of sequencing ~ 20 megabases of raw DNA or cDNA sequence per four hour run of the current sequencing machine, the GS20 (Margulies et al., 2005). The 454 team was able to assemble almost the complete genome of *Mycoplasma genitalium* with 96% coverage at 99.96% accuracy with the data obtained from one run of the instrument (Margulies et al., 2005).

A number of groups have already applied this technology to miRNA research (Meyers et al., 2004; Lu et al., 2006; Maher et al., 2006; Mineno et al., 2006; Ronemus et al., 2006). Solexa Inc. (Hayward, CA), in conjunction with a number of research groups, has created a series of MPSS databases for four species, including *Arabidopsis*. A modified version of MPSS has been used to perform deep profiling of small RNAs from *Arabidopsis* (Nakano et al., 2006). The *Arabidopsis* MPSS plus (http://mpss.udel.edu/at) database is a web-based public resource which provides sequence and expression data for over 60 miRNAs, including miRNA*, identified through MPSS and 454 sequencing from 17 different tissue libraries (Meyers et al., 2004). Wang et al. (2004b) with the use of MPSS data were able to provide expression profiles of 25 predicted miRNAs. Furthermore, deep sequencing of mutants

33

provides a genetic approach for the dissection and characterization of miRNA populations and the identification of low abundance miRNAs. With the use of MPSS and 454 sequencing technologies, Lu et al. (2006) were able to characterize the complement of miRNAs expressed in *Arabidopsis* inflorescence to considerable depth. Nearly all known miRNAs were enriched in this mutant and thirteen new miRNAs were identified, all of which were relatively low abundance and constitute new families.

With new technologies, the sequencing of small RNAs is no longer a limiting factor in the discovery of novel miRNAs. In plants, over the next few years it is likely that more investments will be made to validate predicted miRNAs and to their targets. With the increasing availability of plant genome sequences, comparative genomics approaches are likely to be fruitful for the discovery of conserved miRNAs. A combination of approaches will be necessary to identify the total number of conserved and non-conserved miRNAs in a genome, to determine how many of these are species-specific, and to characterize the biological targets of these molecules.

1.8 XYLOGENESIS: A POSSIBLE TARGET?

Woody plant species represent the majority of the terrestrial biomass on earth. Wood is one of the most abundant biological resources and is an important raw material for the lumber and paper industry, with increasing demand.   Despite the economic and environmental significance of wood, our current understanding of the molecular processes underlying its formation is still limited. Wood formation (xylogenesis) initiates when the primary vascular tissues of stems are replaced by secondary vascular tissues, which are produced by a secondary (lateral) meristem, the vascular cambium (reviewed by Burton et al., 2000; Plomion et al., 2001). This secondary growth is a highly ordered developmental process, which involves the patterned periclinal division of vascular cambium cells and subsequent differentiation of into distinct highly specialized cell types, namely secondary xylem and

phloem tissues (Lachaud et al., 1999). The phloem, produced towards the outside of the stem, transports organic material throughout the plant, and the secondary xylem (wood), produced towards the inside of the stem, is dedicated to water and mineral transport.

The differentiation of the vascular cambium into these complex and highly specialized tissues involves unique developmental programs and the coordinate expression of hundreds of genes, where each tissue requires the expression of a specific set of genes to ensure the maintenance of their specific functionality. Studies have shown that differential gene expression patterns exist across these tissues (Riechmann et al., 2000; Campalans et al., 2001; Ranik et al., 2006). Further research is beginning to prove the existence of differentially expressed genes with specific functions at different levels of wood formation. Many of the genes involved in xylogenesis have been identified and a few biochemical pathways relating to wood formation (e.g. lignin and cellulose biosynthesis) have been characterized. Although the physiological outline of xylogenesis is relatively well documented, the regulation of key developmental genes and the regulation of differentially expressed genes are not.

Many of the identified *Arabidopsis* miRNA targets have family members with known roles in the process of wood formation. The first of these is the MYB transcription factor family, whose members control, among other things, the biosynthetic pathways that lead to the production of components required for the production of lignin during cell wall synthesis. Many of these MYB proteins are produced specifically in xylem tissue (Patzlaff et al., 2003). Another class of transcription factors that has been identified as miRNA targets in *Arabidopsis* is the homeodomain-leucine zipper (HD-ZIP) family. This family of transcription factors is centrally implicated in vascular development in both trees and *Arabidopsis* (McConnell et al., 2001; Otsuga et al., 2001b; Emery et al., 2003). Three members of this family in *Arabidopsis* are closely related and expressed in vascular tissue, in particular in vascular cambium (Burton et al., 2000). An interesting finding is that the

Arabidopsis HD-ZIP transcription factors, *PHABULOSA*, *PHAVOLUTA, REVOLUTA* and *ATHB15* are known targets of the microRNA, miR165. It was further found that not only are these genes involved in organ polarity, but have additional functions in regulation of the tissue arrangement within the vascular bundles (Emery et al., 2003). Furthermore, plant growth hormones, such as auxins, also play regulatory roles in vascular tissue differentiation (Przemeck et al., 1996; Jang et al., 2000; Carland et al., 2002). For example, auxin has been shown to stimulate the differentiation of undifferentiated tissues into vascular tissues in *Zinnia* (Fukuda et al., 1993). In transgenic plants, auxin overproduction results in increased production of vascular tissue, whereas auxin inactivation decreases vascular tissue differentiation (Schrader et al., 2003). *Arabidopsis* mutants defective in auxin transport display altered vascular differentiation and patterning, although these mutations are pleiotropic and affect many different tissues (Przemeck et al., 1996; Carland et al., 2002). Several auxin responsive factors are known targets of miR160/167 (Guo et al., 2005a; Guo et al., 2005b; Mallory et al., 2005; Wang et al., 2005; Wu et al., 2006). Thus, it may be feasible to hypothesize that these miRNAs could be involved in the regulation of auxins during the vascular development of trees.

It is becoming clear that miRNAs have roles during vascular patterning of both primary and secondary vascular tissues (McHale and Koning, 2004; Kim et al., 2005; Ko et al., 2006). To date, there have been two independent investigations into the role of miRNAs during xylogenesis and vascular development in poplar trees (Lu et al., 2005; Ko et al., 2006). Investigation of miRNAs present in different tissues of the poplar stem, including stress-responsive tissues, such as tension and compression wood, identified novel miRNAs not discovered in *Arabidopsis* (Lu et al., 2005). The expression profiles of these novel miRNAs indicated that these miRNAs most likely have specific roles during the development of these specialized tissues in poplar (Lu et al., 2005). This observation lends support to the notion

that miRNA networks may also modulate species-specific processes such as wood formation in trees. Furthermore, Ko et al. (2006) investigated the role of miR166 regulation of an HD-ZIP III protein, PtaHB1, on vascular development in poplar. The expression of PtaHB1 was closely associated with wood formation and regulated both developmentally and seasonally, with the highest expression during the active growing season. The expression of Pta-miR166 was much higher in the winter than in the growing seasons, suggesting seasonal and developmental regulation of microRNA in this perennial plant species (Ko et al., 2006). Thus, it has been shown that miRNAs have validated roles during certain aspects of wood formation and vascular patterning in poplar. However, there may be other specific aspects of xylogenesis under miRNA control that are yet to be discovered. Furthermore, investigation of miRNAs in wood forming tissues from other tree species could identify species-specific miRNAs which are still to be identified.

The aim of this M.Sc. study is to isolate miRNAs from actively differentiating tissues of two tree species, namely *Populus trichocarpa* and *Eucalyptus*. We isolated miRNAs from post-embryonic tissues of poplar plantlets with the aim of identifying miRNAs involved in early tissue differentiation, possibly with roles in primary vascular tissue formation. Furthermore, we aimed to isolate miRNAs from the secondary vascular tissues of *Eucalyptus* trees in order to identify miRNAs with a role in vascular patterning of secondary tissues, which have not been identified in previous studies, and possibly identifying miRNAs specific to this tree genus.

1.11 REFERENCES

Achard, P., Herr, A., Baulcombe, D.C., and Harberd, N.P. (2004). Modulation of floral development by a gibberellin-regulated microRNA. Development **131,** 3357-3365.

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.

Agrawal, N., Dasaradhi, P.V.N., Mohmmed, A., Malhotra, P., Bhatnagar, R.K., and Mukherjee, S.K. (2003). RNA interference: Biology, mechanism, and applications. Microbiol. Mol. Biol. Rev. **67,** 657- 685.

Aida, M., Ishida, T., and Tasaka, M. (1999). Shoot apical meristem and cotyledon formation during *Arabidopsis* embryogenesis: interaction among the CUP-SHAPED COTYLEDON and SHOOT MERISTEMLESS genes. Development **126,** 1563-1570.

Aida, M., Ishida, T., Fukaki, H., Fujisawa, H., and Tasaka, M. (1997). Genes involved in organ separation in *Arabidopsis*: an analysis of the cup-shaped cotyledon mutant. Plant Cell **9,** 841-857.

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

Ambros, V. (2000). Control of developmental timing in *Caenorhabditis elegans*. Curr. Opin. Genet. Dev. **10,** 428-433.

Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T., and Jewell, D. (2003). MicroRNAs and other tiny endogenous RNAs in *C. elegans*. Curr. Biol. **13,** 807-818.

Aravin, A.A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., and Tuschl, T. (2003). The small RNA profile during *Drosophila melanogaster* development. Dev. Cell **5,** 337-350.

Aufsatz, W., Mette, M.F., van der Winden, J., Matzke, M., and Matzke, A.J.M. (2002a). HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. EMBO J. **21,** 6832-6841.

Aufsatz, W., Mette, M.F., van der Winden, J., Matzke, A.J., and Matzke, M. (2002b). RNA-directed DNA methylation in *Arabidopsis*. Proc. Natl. Acad. Sci. USA **99 Suppl 4,** 16499-16506.

Aukerman, M.J., and Sakai, H. (2003). Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. Plant Cell **15,** 2730-2741.

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

Banerjee, D., and Slack, F. (2002). Control of developmental timing by small temporal RNAs: a paradigm for RNA-mediated regulation of gene expression. Bioessays **24,** 119-129.

Bao, N., Lye, K.W., and Barton, M.K. (2004). MicroRNA binding sites in *Arabidopsis* class III HD-ZIP mRNAs are required for methylation of the template chromosome. Dev. Cell **7,** 653-662.

Bari, R., Datt Pant, B., Stitt, M., and Scheible, W.-R. (2006). PHO2, microRNA399, and PHR1 define a phosphate-signaling pathway in plants. Plant Physiol. **141,** 988-999.

Bartel, B., and Bartel, D.P. (2003). MicroRNAs: at the root of plant development. Plant Physiol. **132,** 709-717.

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

Baskerville, S., and Bartel, D. (2005). Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. RNA **11,** 241-247.

Baulcombe, D.C. (1996). RNA as a target and an initiator of post-transcriptional gene silencing in transgenic plants. Plant Mol. Biol. **32,** 79-88.

Baumberger, N., and Baulcombe, D.C. (2005). *Arabidopsis* ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proc. Natl. Acad. Sci. USA **102,** 11928-11932.

Bennasser, Y., Le, S.Y., Yeung, M.L., and Jeang, K.T. (2004). HIV-1 encoded candidate micro-RNAs and their cellular targets. Retrovirology **1,** 43-47.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature **409,** 363-366.

Blaszczyk, J., Tropea, J.E., Bubunenko, M., Routzahn, K.M., Waugh, D.S., Court, D.L., and Ji, X. (2001). Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage. Structure (Camb) **9,** 1225-1236.

Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., and Benning, C. (1998). AGO1 defines a novel locus of *Arabidopsis* controlling leaf development. EMBO J. **17,** 170-180.

Bollman, K.M., Aukerman, M.J., Park, M.Y., Hunter, C., Berardini, T.Z., and Poethig, R.S. (2003). HASTY, the *Arabidopsis* ortholog of exportin 5/MSN5, regulates phase change and morphogenesis. Development **130,** 1493-1504.

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. USA **101,** 115511-111516.

Boutet, S., Vazquez, F., Liu, J., Beclin, C., Fagard, M., Gratias, A., Morel, J.B., Crete, P., Chen, X., and Vaucheret, H. (2003). *Arabidopsis* HEN1: a genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance. Curr. Biol. **13,** 843-848.

Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B., and Cohen, S.M. (2003). *bantam* encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. Cell **113,** 25-36.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S.R., Moon, K., Burcham, T., Pallas, M., DuBridge, R.B., Kirchner, J., Fearon, K., Mao, J., and Corcoran, K. (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat. Biotechnol. **18,** 630-634.

Burton, R.A., Gibeaut, D.M., Bacic, A., Findlay, K., Roberts, K., Hamilton, A., Baulcombe, D.C., and Fincher, G.B. (2000). Virus-induced silencing of a plant cellulose synthase gene. Plant Cell **12,** 691-706.

Calin, G.A., Liu, C.-G., Sevignani, C., Ferracin, M., Felli, N., Dumitru, C.D., Shimizu, M., Cimmino, A., Zupo, S., Dono, M., Dell'Aquila, M.L., Alder, H., Rassenti, L., Kipps, T.J.,

Bullrich, F., Negrini, M., and Croce, C.M. (2004). MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. Proc. Natl. Acad. Sci. USA **101,** 11755-11760.

Campalans, A., Pages, M., and Messeguer, R. (2001). Identification of differentially expressed genes by the cDNA-AFLP technique during dehydration of almond (*Prunus amygdalus*). Tree Physiol. **21,** 633-643.

Carland, F.M., Fujioka, S., Takatsuto, S., Yoshida, S., and Nelson, T. (2002). The identification of CVP1 reveals a role for sterols in vascular patterning. Plant Cell **14,** 2045-2058.

Carmell, M.A., Xuan, Z., Zhang, M.Q., and Hannon, G.J. (2002). The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes Dev. **16,** 2733-2742.

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

Chan, S.W., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C., and Jacobsen, S.E. (2004). RNA silencing genes control *de novo* DNA methylation. Science **303,** 1336.

Chen, C., Ridzon, D.A., Broomer, A.J., Zhou, Z., Lee, D.H., Nguyen, J.T., Barbisin, M., Xu, N.L., Mahuvakar, V.R., Andersen, M.R., Lao, K.Q., Livak, K.J., and Guegler, K.J. (2005). Real-time quantification of microRNAs by stem-loop RT-PCR. Nucl. Acids Res. **33,** e179.

Chen, C.Z., Li, L., Lodish, H.F., and Bartel, D.P. (2004). MicroRNAs modulate hematopoietic lineage differentiation. Science **303,** 83-86.

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

Chiou, T.J., Aung, K., Lin, S.I., Wu, C.-C., Chiang, S.F., and Su, C.l. (2005). Regulation of phosphate homeostasis by microRNA in *Arabidopsis*. Plant Cell **18,** 412-421.

Cogoni, C., Irelan, J.T., Schumacher, M., Schmidhauser, T.J., Selker, E.U., and Macino, G. (1996). Transgene silencing of the *al-1* gene in vegetative cells of *Neurospora* is mediated by a cytoplasmic effector and does not depend on DNA-DNA interactions or DNA methylation. EMBO J. **15,** 3153-3163.

Dalmay, T., Horsefield, R., Braunstein, T.H., and Baulcombe, D.C. (2001). SDE3 encodes an RNA helicase required for post-transcriptional gene silencing in *Arabidopsis*. EMBO J. **20,** 2069-2078.

de Carvalho, F., Gheysen, G., Kushnir, S., Van Montagu, M., Inze, D., and Castresana, C. (1992). Suppression of beta-1,3-glucanase transgene expression in homozygous plants. EMBO J. **11,** 2595-2602.

Doench, J.G., Petersen, C.P., and Sharp, P.A. (2003). siRNAs can function as miRNAs. Genes Dev. **17,** 438-442.

Dostie, J., Mourelatos, Z., Yang, M., Sharma, A., and Dreyfuss, G. (2003). Numerous microRNPs in neuronal cells containing novel microRNAs. RNA **9,** 180-186.

Downes, B.P., Stupar, R.M., Gingerich, D.J., and Vierstra, R.D. (2003). The HECT ubiquitin-protein ligase (UPL) family in *Arabidopsis*: UPL3 has a specific role in trichome development. Plant J. **35,** 729-742.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001a). RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev. **15,** 188-200.

Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. (2001b). Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. EMBO J. **20,** 6877-6888.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001c). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature **411,** 494-498.

Emery, J.F., Floyd, S.K., Alvarez, J., Eshed, Y., Hawker, N.P., Izhaki, A., Baum, S.F., and Bowman, J.L. (2003). Radial patterning of *Arabidopsis* shoots by class III HD-ZIP and KANADI genes. Curr. Biol. **13,** 1768-1774.

English, J.J., Mueller, E., and Baulcombe, D.C. (1996). Suppression of virus accumulation in transgenic plants exhibiting silencing of nuclear genes. Plant Cell **8,** 179-188.

Eshed, Y., and Bowman, J.L. (2004). MicroRNAs guide asymmetric DNA modifications guiding asymmetric organs. Dev. Cell **7,** 629-630.

Fagard, M., Boutet, S., Morel, J.B., Bellini, C., and Vaucheret, H. (2000). AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. Proc. Natl. Acad. Sci. USA **97,** 11650-11654.

Feinbaum, R., and Ambros, V. (1999). The timing of *lin-4* RNA accumulation controls the timing of postembryonic developmental events in *Caenorhabditis elegans*. Dev. Biol. **210,** 87-95.

Finnegan, E.J., and Matzke, M.A. (2003). The small RNA world. J. Cell Sci. **116,** 4689-4693.

Fire, A., Albertson, D., Harrison, S.W., and Moerman, D.G. (1991). Production of antisense RNA leads to effective and specific inhibition of gene expression in *C. elegans* muscle. Development **113,** 503-514.

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

Floyd, S.K., and Bowman, J.L. (2004). Gene regulation: ancient microRNA target sequences in plants. Nature **428,** 485-486.

Fu, X., and Harberd, N.P. (2003). Auxin promotes *Arabidopsis* root growth by modulating gibberellin response. Nature **421,** 740-743.

Fujii, H., Chiou, T.J., Lin, S.I., Aung, K., and Zhu, J.K. (2005). A miRNA involved in phosphate-starvation response in *Arabidopsis*. Curr. Biol. **15,** 2038-2043.

Fukaki, H., and Tasaka, M. (1999). Gravity perception and gravitropic response of inflorescence stems in *Arabidopsis thaliana*. Adv. Space Res. **24,** 763-770.

Fukuda, H., Sato, Y., Yoshimura, T., and Demura, T. (1993). Molecular mechanisms of xylem differentiation. J. Plant Res. **57,** 97-107.

Gasciolli, V., Mallory, A.C., Bartel, D.P., and Vaucheret, H. (2005). Partially redundant functions of *Arabidopsis* DICER-like enzymes and a role for DCL in producing trans-acting siRNAs. Curr. Biol. **15,** 1494-1500.

Goodwin, J., Chapman, K., Swaney, S., Parks, T.D., Wernsman, E.A., and Dougherty, W.G. (1996). Genetic and biochemical dissection of transgenic RNA-mediated virus resistance. Plant Cell **8,** 95-105.

Grewal, S.I., and Moazed, D. (2003). Heterochromatin and epigenetic control of gene expression. Science **301,** 798-802.

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. Cell **106,** 23-34.

Grosshans, H., and Slack, F.J. (2002). Micro-RNAs: small is plentiful. J. Cell Biol. **156,** 17-22.

Guddeti, S., Zhang, d.C., Li, A., Leseberg, C., Kang, H., Li, X., Zhai, W., Johns, M., and Mao, L. (2005). Molecular evolution of the rice miR395 gene family. Cell Res **15,** 631-638.

Guilfoyle, T.J., Ulmasov, T., and Hagen, G. (1998). The ARF family of transcription factors and their role in plant hormone-responsive transcription. Cell Mol. Life. Sci. **54,** 619-627.

Guo, H.-S., Xie, Q., Fei, J.-F., and Chua, N.-H. (2005a). MicroRNA164 directs NAC1 mRNA cleavage to downregulate auxin signals for lateral root development. Plant Cell **17,** 1376-1386.

Guo, H.S., Xie, Q., Fei, J.F., and Chua, N.H. (2005b). MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals for *Arabidopsis* lateral root development. Plant Cell **17,** 1376-1386.

Guo, S., and Kemphues, K.J. (1995). *par-1*, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. Cell **81,** 611-620.

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 Database Issue,** D637-640.

Gvozdev, V.A., Aravin, A.A., Abramov, Y.A., Klenov, M.S., Kogan, G.L., Lavrov, S.A., Naumova, N.M., Olenkina, O.M., Tulin, A.V., and Vagin, V.V. (2003). Stellate repeats: targets of silencing and modules causing cis-inactivation and trans-activation. Genetica **117,** 239-245.

Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. (2002). Two classes of short interfering RNA in RNA silencing. EMBO J. **21,** 4671-4679.

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

Hammond, S.M., Bernstein, E., Beach, D., and Hannon, G.J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. Nature **404,** 293-296.

Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R., and Hannon, G.J. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. Science **293,** 1146-1150.

Han, M.-H., Goud, S., Song, L., and Fedoroff, N. (2004). The *Arabidopsis* double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. Proc. Natl. Acad. Sci. USA **101,** 1093-1098.

Hannon, G.J. (2002). RNA interference. Nature **418,** 244-251.

Helariutta, Y., Fukaki, H., Wysocka-Diller, J., Nakajima, K., Jung, J., Sena, G., Hauser, M.T., and Benfey, P.N. (2000). The SHORT-ROOT gene controls radial patterning of the *Arabidopsis* root through radial signaling. Cell **101,** 555-567.

Humphreys, D.T., Westman, B.J., Martin, D.I.K., and Preiss, T. (2005). MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. Proc. Natl. Acad. Sci. USA **101,** 16961-16966.

Hunter, C., Sun, H., and Poethig, R.S. (2003). The *Arabidopsis* heterochronic gene *ZIPPY* is an *ARGONAUTE* family member. Curr. Biol. **13,** 1734-1739.

Hutvagner, G., and Zamore, P.D. (2002a). RNAi: nature abhors a double-strand. Curr. Opin. Genet. Dev. **12,** 225-232.

Hutvagner, G., and Zamore, P.D. (2002b). A microRNA in a multiple-turnover RNAi enzyme complex. Science **297,** 2056-2060.

Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. Science **293,** 834-838.

Jacobsen, S.E., Running, M.P., and Meyerowitz, E.M. (1999). Disruption of an RNA helicase/RNAse III gene in *Arabidopsis* causes unregulated cell division in floral meristems. Development **126,** 5231-5243.

Jang, J.C., Fujioka, S., Tasaka, M., Seto, H., Takatsuto, S., Ishii, A., Aida, M., Yoshida, S., and Sheen, J. (2000). A critical role of sterols in embryonic patterning and meristem programming revealed by the fackel mutants of *Arabidopsis thaliana*. Genes Dev. **14,** 1485-1497.

Jin, H., and Martin, C. (1999). Multifunctionality and diversity within the plant MYB-gene family. Plant Mol. Biol. **41,** 577-585.

Jofuku, K.D., Omidyar, P.K., Gee, Z., and Okamuro, J.K. (2005). Control of seed mass and seed yield by the floral homeotic gene APETALA2. Proc. Natl. Acad. Sci. USA **102,** 3117-3122.

Johnson, S.M., Lin, S.-Y., and Slack, F.J. (2003). The time of appearance of the *C. elegans* let-7 microRNA is transcriptionally controlled utilizing a temporal regulatory element in its promoter. Dev. Bio. **259,** 364-379.

Johnston, R.J., and Hobert, O. (2003). A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. Nature **426,** 845-849.

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.

Jones, L., Hamilton, A.J., Voinnet, O., Thomas, C.L., Maule, A.J., and Baulcombe, D.C. (1999). RNA-DNA interactions and DNA methylation in post-transcriptional gene silencing. Plant Cell **11,** 2291-2302.

Juarez, M.T., Kui, J.S., Thomas, J., Heller, B.A., and Timmermans, M.C. (2004). microRNA-mediated repression of *rolled leaf1* specifies maize leaf polarity. Nature **428,** 84-88.

Kasschau, K.D., Xie, Z., Allen, E., Llave, C., Chapman, E.J., Krizan, K.A., and Carrington, J.C. (2003). P1/HC-Pro, a viral suppressor of RNA silencing, interferes with *Arabidopsis* development and miRNA function. Dev. Cell **4,** 205-217.

Ke, X.S., Liu, C.M., Liu, D.P., and Liang, C.C. (2003). MicroRNAs: key participants in gene regulatory networks. Curr. Opin. Chem. Biol. **7,** 516-523.

Kennerdell, J.R., and Carthew, R.W. (1998). Use of dsRNA-mediated genetic interference to demonstrate that *frizzled* and *frizzled 2* act in the wingless pathway. Cell **95,** 1017-1026.

Ketting, R.F., Haverkamp, T.H., van Luenen, H.G., and Plasterk, R.H. (1999). Mut-7 of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell **99,** 133-141.

Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. Genes Dev. **15,** 2654-2659.

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

Kim, J., Jung, J.H., Reyes, J.L., Kim, Y.S., Kim, S.Y., Chung, K.S., Kim, J.A., Lee, M., Lee, Y., Narry Kim, V., Chua, N.H., and Park, C.M. (2005). microRNA-directed cleavage of ATHB15 mRNA regulates vascular development in *Arabidopsis* inflorescence stems. Plant J. **42,** 84-94.

Kliebenstein, D.J., Monde, R.A., and Last, R.L. (1998). Superoxide dismutase in *Arabidopsis*: an eclectic enzyme family with disparate regulation and protein localization. Plant Physiol. **118,** 637-650.

Ko, J., Prassinos, C., and Han, K. (2006). Developmental and seasonal expression of *PtaHB1*, a *Populus* gene encoding a class III HD-Zip protein, is closely associated with secondary growth and inversely correlated with the level of microRNA (miR166). New Phytol. **169,** 469-478.

Kogan, G.L., Tulin, A.V., Aravin, A.A., Abramov, Y.A., Kalmykova, A.I., Maisonhaute, C., and Gvozdev, V.A. (2003). The GATE retrotransposon in *Drosophila melanogaster*: mobility in heterochromatin and aspects of its expression in germline tissues. Mol. Genet. Genomics **269,** 234-242.

Kurihara, Y., and Watanabe, Y. (2004). *Arabidopsis* microRNA biogenesis through Dicer-like 1 protein functions. Proc. Natl. Acad. Sci. USA **101,** 12753-12758.

Lachaud, S., Catesson, A.M., and Bonnemain, J.L. (1999). Structure and functions of the vascular cambium. C R Acad Sci III **322,** 633-650.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science **294,** 853-858.

Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A., and Tuschl, T. (2003). New microRNAs from mouse and human. RNA **9,** 175-179.

Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., and Tuschl, T. (2002). Identification of tissue-specific microRNAs from mouse. Curr. Biol. **12,** 735-739.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science **294,** 858-862.

Laufs, P., Peaucelle, A., Morin, H., and Traas, J. (2004). MicroRNA regulation of the CUC genes is required for boundary size control in *Arabidopsis* meristems. Development **131,** 4311-4322.

Lauter, N., Kampani, A., Carlson, S., Goebel, M., and Moose, S.P. (2005). MicroRNA172 down-regulates *glossy15* to promote vegetative phase change in maize. Proc. Natl. Acad. Sci. USA **102,** 9412-9417.

Lee, R.C., and Ambros, V. (2001). An extensive class of small RNAs in *Caenorhabditis elegans*. Science **294,** 862-864.

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.

Lee, Y., Jeon, K., Lee, J.-T., Kim, S., and Kim, V.N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. EMBO J. **21,** 4663-4670.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., and Kim, V.N. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature **425,** 415-419.

Leustek, T. (2002). Sulfate metabolism. In the *Arabidopsis* book, C.R. Somerville and E.M. Meyerowitz, eds. (Rockville, MD: America Society of Plant Biologists).

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. Cell **115,** 787-798.

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

Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., and Bartel, D.P. (2003). The microRNAs of *Caenorhabditis elegans*. Genes Dev. **17,** 991-1008.

Lin, S.Y., Johnson, S.M., Abraham, M., Vella, M.C., Pasquinelli, A., Gamberi, C., Gottlieb, E., and Slack, F.J. (2003). The *C. elegans* hunchback homolog*, hbl-1*, controls temporal patterning and is a probable microRNA target. Dev. Cell **4,** 639-650.

Lingel, A., Simon, B., Izaurralde, E., and Sattler, M. (2003). Structure and nucleic-acid binding of the *Drosophila* Argonaute2 PAZ domain. Nature **426,** 465-469.

Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science **305,** 1437-1441.

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

Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C. (2002b). Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. Science **297,** 2053-2056.

Lu, C., and Fedoroff, N. (2000). A mutation in the *Arabidopsis* HYL1 gene encoding a dsRNA binding protein affects responses to abscisic acid, auxin, and cytokinin. Plant Cell **12,** 2351-2366.

Lu, C., Kulkarni, K., Souret, F.F., MuthuValliappan, R., Tej, S.S., Poethig, R.S., Henderson, I.R., Jacobsen, S.E., Wang, W., Green, P.J., and Meyers, B.C. (2006). MicroRNAs and other small RNAs enriched in the *Arabidopsis* RNA-dependent RNA polymerase-2 mutant. Genome Res. **16,** 1276-1288.

Lu, S., Sun, Y.-H., Shi, R., Clark, C., Li, L., and Chiang, V.L. (2005). Novel and mechanical stress-responsive microRNAs in *Populus trichocarpa* that are absent from *Arabidopsis*. Plant Cell **17,** 2186-2203.

Lynn, K., Fernandez, A., Aida, M., Sedbrook, J., and Tasaka, M. (1999). The *PINHEAD/ZWILLE* gene acts pleiotropically in *Arabidopsis* development and has overlapping functions with the *ARGONAUTE1* gene. Development **126,** 469-481.

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

MacRae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D., and Doudna, J.A. (2006). Structural basis for double-stranded RNA processing by Dicer. Science **311,** 195-198.

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

Mallory, A.C., and Vaucheret, H. (2004). MicroRNAs: something important between the genes. Curr. Opin. Plant Biol. **7,** 120-125.

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

Mallory, A.C., Dugas, D.V., Bartel, D.P., and Bartel, B. (2004a). MicroRNA regulation of NAC-domain targets is required for proper formation and separation of adjacent embryonic, vegetative, and floral organs. Curr. Biol. **14,** 1035-1046.

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

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., and Chen, Z. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature **437,** 376-380.

Matzke, M.A., Matzke, A.J., Pruss, G.J., and Vance, V.B. (2001). RNA-based silencing strategies in plants. Curr. Opin. Genet. Dev. **11,** 221-227.

McConnell, J.R., Emery, J., Eshed, Y., Bao, N., Bowman, J., and Barton, M.K. (2001). Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. Nature **411,** 709-713.

McHale, N.A., and Koning, R.E. (2004). MicroRNA-directed cleavage of *Nicotiana sylvestris PHAVOLUTA* mRNA regulates the vascular cambium and structure of apical meristems. Plant Cell **16,** 1730-1740.

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

Mette, M.F., Aufsatz, W., van der Winden, J., Matzke, M.A., and Matzke, A.J.M. (2000). Transcriptional silencing and promoter methylation triggered by double-stranded RNA. EMBO J. **19,** 5194-5201.

Metzlaff, M., O'Dell, M., Cluster, P.D., and Flavell, R.B. (1997). RNA-mediated RNA degradation and chalcone synthase A silencing in petunia. Cell **88,** 845-854.

Meyers, B.C., Lee, D.K., Vu, T.H., Tej, S.S., Edberg, S.B., Matvienko, M., and Tindell, L.D. (2004). *Arabidopsis* MPSS. An online resource for quantitative expression analysis. Plant Physiol. **135,** 801-813.

Millar, A.A., and Gubler, F. (2005). The *Arabidopsis* GAMYB-like genes, *MYB33* and *MYB65*, are microRNA-regulated genes that redundantly facilitate anther development. Plant Cell **17,** 705-721.

Mineno, J., Okamoto, S., Ando, T., Sato, M., Chono, H., Izu, H., Takayama, M., Asada, K., Mirochnitchenko, O., Inouye, M., and Kato, I. (2006). The expression profile of microRNAs in mouse embryos. Nucl. Acids Res. **34,** 1765-1771.

Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev. **16,** 720-728.

Moussian, B., Haecker, A., and Laux, T. (2003). *ZWILLE* buffers meristem stability in *Arabidopsis thaliana*. Dev. Genes Evol. **213,** 534-540.

Nakajima, K., Sena, G., Nawy, T., and Benfey, P.N. (2001). Intercellular movement of the putative transcription factor SHR in root patterning. Nature **413,** 307-311.

Nakano, M., Nobuta, K., Vemaraju, K., Tej, S.S., Skogen, J.W., and Meyers, B.C. (2006). Plant MPSS databases: signature-based transcriptional resources for analyses of mRNA and small RNA. Nucl. Acids Res. **34,** D731-735.

Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. Plant Cell **2,** 279-289.

Ngo, H., Tschudi, C., Gull, K., and Ullu, E. (1998). Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. Proc. Natl. Acad. Sci. USA **95,** 14687-14692.

Nykanen, A., Haley, B., and Zamore, P.D. (2001). ATP requirements and small interfering RNA structure in the RNA interference pathway. Cell **107,** 309-321.

Ohashi-Ito, K., and Fukuda, H. (2003). HD-ZIP III homeobox genes that include a novel member, ZeHB-13 (*Zinnia*)/ATHB-15(*Arabidopsis*), are involved in procambium and xylem cell differentiation. Plant Cell Physiol. **44,** 1350-1358.

Ohashi-Ito, K., Kubo, M., Demura, T., and Fukuda, H. (2005). Class III homeodomain leucine-zipper proteins regulate xylem cell differentiation. Plant Cell Physiol. **46,** 1646-1656.

Olsen, P.H., and Ambros, V. (1999). The *lin-4* regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. Dev. Biol. **216,** 671-680.

Omoto, S., and Fujii, Y.R. (2005). Regulation of human immunodeficiency virus 1 transcription by nef microRNA. J. Gen. Virol. **86,** 751-755.

Ostermeier, G.C., Goodrich, R.J., Moldenhauer, J.S., Diamond, M.P., and Krawetz, S.A. (2005). A suite of novel human spermatozoal RNAs. J. Androl. **26,** 70-74.

Otsuga, D., DeGuzman, B., Prigge, M.J., Drews, G.N., and Clark, S.E. (2001a). REVOLUTA regulates meristem initiation at lateral positions. Plant J **25,** 223-236.

Otsuga, D., DeGuzman, B., Prigge, M.J., Drews, G.N., and Clark, S.E. (2001b). REVOLUTA regulates meristem initiation at lateral positions. Plant J. **25,** 223-236.

Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C., and Weigel, D. (2003). Control of leaf morphogenesis by microRNAs. Nature **425,** 257-263.

Papp, I., Mette, M.F., Aufsatz, W., Daxinger, L., Schauer, S.E., Ray, A., van der Winden, J., Matzke, M., and Matzke, A.J.M. (2003). Evidence for nuclear processing of plant microRNA and short interfering RNA precursors. Plant Physiol. **132,** 1382-1390.

Parizotto, E.A., Dunoyer, P., Rahm, N., Himber, C., and Voinnet, O. (2004). *In vivo* investigation of the transcription, processing, endonucleolytic activity, and functional relevance of the spatial distribution of a plant miRNA. Genes Dev. **18,** 2237-2242.

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

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.

Parrish, S., and Fire, A. (2001). Distinct roles for RDE-1 and RDE-4 during RNA interference in *Caenorhabditis elegans*. RNA **7,** 1397-1402.

Parrish, S., Fleenor, J., Xu, S., Mello, C., and Fire, A. (2000). Functional anatomy of a dsRNA trigger: differential requirement for the two trigger strands in RNA interference. Mol. Cell **6,** 1077-1087.

Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Muller, P., Spring, J., Srinivasan, A., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E., and Ruvkun, G. (2000). Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA. Nature **408,** 86-89.

Patzlaff, A., McInnis, S., Courtenay, A., Surman, C., Newman, L.J., Smith, C., Bevan, M.W., Mansfield, S., Whetten, R.W., Sederoff, R.R., and Campbell, M.M. (2003). Characterisation of a pine MYB that regulates lignification. Plant J. **36,** 743-754.

Pfeffer, S., Zavolan, M., Grasser, F.A., Chien, M.C., Russo, J.J., Ju, J.Y., John, B., Enright, A.J., Marks, D., Sander, C., and Tuschl, T. (2004). Identification of virus-encoded microRNAs. Science **304,** 734-736.

Pillai, R., Artus, C., and Filipowicz, W. (2004). Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis. RNA **10,** 1518-1525.

Pillai, R.S., Bhattacharyya, S.N., Artus, C.G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E., and Filipowicz, W. (2005). Inhibition of translational initiation by *let-7* microRNA in human cells. Science **309,** 1573-1576.

Plomion, C., Leprovost, G., and Stokes, A. (2001). Wood formation in trees. Plant Physiol. **127,** 1513-1523.

Przemeck, G.K.H., Mattsson, J., Hardtke, C.S., Sung, Z.R., and Berleth, T. (1996). Studies on the role of the *Arabidopsis* gene *MONOPTEROS* in vascular development and plant cell axialization. Planta **200,** 229-237.

Ranik, M., Creux, N.M., and Myburg, A.A. (2006). Within-tree transcriptome profiling in wood-forming tissues of a fast-growing *Eucalyptus* tree. Tree Physiol. **26,** 365-375.

Reinhart, B.J., and Bartel, D.P. (2002). Small RNAs correspond to centromere heterochromatic repeats. Science **297,** 1831.

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

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

Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., and Bartel, D.P. (2002). Prediction of plant microRNA targets. Cell **110,** 513-520.

Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Z, C., Jiang, Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J.Z., Ghandehari, D., Sherman, B.K., and L. Yu, G. (2000). *Arabidopsis* transcription factors: Genome-wide comparative analysis among eukaryotes. Science **290,** 2105-2110.

Romano, N., and Macino, G. (1992). Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. Mol. Microbiol. **6,** 3343-3353.

Ronemus, M., Vaughn, M.W., and Martienssen, R.A. (2006). MicroRNA-targeted and small interfering RNA-mediated mRNA degradation is regulated by Argonaute, Dicer, and RNA-Dependent RNA polymerase in *Arabidopsis*. Plant Cell **18,** 1559-1574.

Schauer, S.E., Jacobsen, S.E., Meinke, D.W., and Ray, A. (2002). DICER-LIKE1: blind men and elephants in *Arabidopsis* development. Trends Plant Sci. **7,** 487-491.

Schmid, M., Uhlenhaut, N.H., Godard, F., Demar, M., Bressan, R., Weigel, D., and Lohmann, J.U. (2003). Dissection of floral induction pathways using global expression analysis. Development **130,** 6001-6012.

Schrader, J., Baba, K., May, S.T., Palme, K., Bennett, M., Bhalerao, R.P., and Sandberg, G. (2003). Polar auxin transport in the wood-forming tissues of hybrid aspen is under simultaneous control of developmental and environmental signals. Proc. Natl. Acad. Sci. USA **100,** 10096-10101.

Seitz, H., Royo, H., Bortolin, M.-L., Lin, S.-P., Ferguson-Smith, A.C., and Cavaille, J. (2004). A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. Genome Res. **14,** 1741-1748.

Sempere, L.F., Freemantle, S., Pitha-Rowe, I., Moss, E., Dmitrovsky, E., and Ambros, V. (2004). Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. Genome Biol. **5,** R13.

Shi, R., and Chiang, V.L. (2005). Facile means for quantifying microRNA expression by real-time PCR. Biotechniques **39,** 519-525.

Slack, F.J., Basson, M., Liu, Z., Ambros, V., Horvitz, H.R., and Ruvkun, G. (2000). The *lin-41* RBCC gene acts in the *C. elegans* heterochronic pathway between the *let-7* regulatory RNA and the LIN-29 transcription factor. Mol. Cell **5,** 659-669.

Smith, H.A., Swaney, S.L., Parks, T.D., Wernsman, E.A., and Dougherty, W.G. (1994). Transgenic plant virus resistance mediated by untranslatable sense RNAs: expression, regulation, and fate of nonessential RNAs. Plant Cell **6,** 1441-1453.

Song, J.J., Smith, S.K., Hannon, G.J., and Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. Science **305,** 1434-1437.

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

Souer, E., van Houwelingen, A., Kloos, D., Mol, J., and Koes, R. (1996). The no apical meristem gene of *Petunia* is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. Cell **85,** 159-170.

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

Sunkar, R., and Zhu, J.K. (2004). Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. Plant Cell **16,** 2001-2019.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C.C. (1999). The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. Cell **99,** 123-132.

Tang, G. (2005). siRNA and miRNA: an insight into RISCs. Trends Biochem. Sci. **30,** 106-114.

Tang, G., Reinhart, B.J., Bartel, D.P., and Zamore, P.D. (2003). A biochemical framework for RNA silencing in plants. Genes Dev. **17,** 49-63.

Tanzer, A., and Stadler, P.F. (2004). Molecular evolution of a microRNA cluster. J. Mol. Biol. **339,** 327-335.

Tasaka, M., Kato, T., and Fukaki, H. (1999). The endodermis and shoot gravitropism. Trends Plant Sci. **4,** 103-107.

Tijsterman, M., Ketting, R.F., and Plasterk, R.H. (2002a). The genetics of RNA silencing. Annu. Rev. Genet. **36,** 489-519.

Tijsterman, M., Ketting, R.F., Okihara, K.L., Sijen, T., and Plasterk, R.H. (2002b). RNA helicase MUT-14-dependent gene silencing triggered in *C. elegans* by short antisense RNAs. Science **295,** 694-697.

Tuschl, T., Zamore, P.D., Lehmann, R., Bartel, D.P., and Sharp, P.A. (1999). Targeted mRNA degradation by double-stranded RNA *in vitro*. Genes Dev. **13,** 3191-3197.

van Blokland, R., Van der Geest, N., Mol, J.N.M., and Kooter, J.M. (1994). Transgene-mediated suppression of chalcone synthase expression in *Petunia hybrida* results from an increase in RNA turnover. Plant J. **6,** 861-877.

van Hoof, A., and Parker, R. (1999). The exosome: a proteasome for RNA. Cell **99,** 347-350.

Vaucheret, H., Beclin, C., and Fagard, M. (2001). Post-transcriptional gene silencing in plants. J. Cell. Sci. **114,** 3083-3091.

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

Vazquez, F., Gasciolli, V., Crete, P., and Vaucheret, H. (2004a). The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. Curr. Biol. **14,** 346-351.

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

Voinnet, O. (2001). RNA silencing as a plant immune system against viruses. Trends Genet. **17,** 449-459.

Voinnet, O., and Baulcombe, D.C. (1997). Systemic signalling in gene silencing. Nature **389,** 553.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science **297,** 1833-1837.

Wang, J.-F., Zhou, H., Chen, Y.-Q., Luo, Q.-J., and Qu, L.-H. (2004a). Identification of 20 microRNAs from *Oryza sativa*. Nucleic Acids Res. **32,** 1688-1695.

Wang, J.-W., Wang, L.-J., Mao, Y.-B., Cai, W.-J., Xue, H.-W., and Chen, X.-Y. (2005). Control of root cap formation by microRNA-targeted auxin response factors in *Arabidopsis*. Plant Cell **17,** 2204-2216.

Wang, X.J., Reyes, J.L., Chua, N.H., and Gaasterland, T. (2004b). Prediction and identification of *Arabidopsis thaliana* microRNAs and their mRNA targets. Genome Biol. **5,** R65.

Wassenegger, M., Heimes, S., Riedel, L., and Sanger, H.L. (1994). RNA-directed *de novo* methylation of genomic sequences in plants. Cell **76,** 567-576.

Waterhouse, P.M., Graham, M.W., and Wang, M.B. (1998). Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. Proc. Natl. Acad. Sci. USA **95,** 13959-13964.

Weir, I., Lu, J., Cook, H., Causier, B., Schwarz-Sommer, Z., and Davies, B. (2004). CUPULIFORMIS establishes lateral organ boundaries in *Antirrhinum*. Development **131,** 915-922.

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

Wu, G., and Poethig, R.S. (2006). Temporal regulation of shoot development in *Arabidopsis thaliana* by miR156 and its target SPL3. Development **133,** 3539-3547.

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

Xie, Q., Frugis, G., Colgan, D., and Chua, N.H. (2000). *Arabidopsis* NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. Genes Dev. **14,** 3024-3036.

Xie, Z., Kasschau, K.D., and Carrington, J.C. (2003). Negative feedback regulation of Dicer-Like1 in *Arabidopsis* by microRNA-guided mRNA degradation. Curr. Biol. **13,** 784-789.

Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S.A., and Carrington, J.C. (2005). Expression of *Arabidopsis MIRNA* genes. Plant Physiol. **138,** 2145-2154.

Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. (2004). Genetic and functional diversification of small RNA pathways in plants. PLoS Biol **2,** e104.

Xu, P., Vernooy, S.Y., Guo, M., and Hay, B.A. (2003). The *Drosophila* microRNA Mir-14 suppresses cell death and is required for normal fat metabolism. Curr. Biol. **13,** 790-795.

Yan, K.S., Yan, S., Farooq, A., Han, A., Zeng, L., and Zhou, M.M. (2003). Structure and conserved RNA binding of the PAZ domain. Nature **426,** 468-474.

Yang, W.J., Yang, D.D., Na, S., Sandusky, G.E., Zhang, Q., and Zhao, G. (2005). Dicer is required for embryonic angiogenesis during mouse development. J. Biol. Chem. **280,** 9330-9335.

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

Yekta, S., Shih, I.H., and Bartel, D.P. (2004). MicroRNA-directed cleavage of HOXB8 mRNA. Science **304,** 594-596.

Ying, S.Y., and Lin, S.L. (2004). Intron-derived microRNAs - fine tuning of gene functions. Gene **342,** 25-28.

Yu, Z., Raabe, T., and Hecht, N.B. (2005). MicroRNA122a reduces expression of the post-transcriptionally regulated germ cell transition protein 2 (Tnp2) messenger RNA (mRNA) by mRNA cleavage. Biol. Reprod. **73,** 427-433.

Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P. (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell **101,** 25-33.

Zeng, Y., Yi, R., and Cullen, B.R. (2003). MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. Proc. Natl. Acad. Sci. USA **100,** 9779-9784.

Zhong, R., and Ye, Z.H. (2004). *Amphivasal vascular bundle 1*, a gain-of-function mutation of the IFL1/REV gene, is associated with alterations in the polarity of leaves, stems and carpels. Plant Cell Physiol. **45,** 369-385.

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

CHAPTER 2

NOVEL MICRORNAS WITH DIVERSE ROLES DURING EARLY DEVELOPMENT OF *POPULUS TRICHOCARPA* PLANTLETS

Michelle Victor[1], Shanfa Lu[2], Ying-Hsuan Sun[2], Vincent L. Chiang[2] and Alexander A. Myburg[1]

[1] *Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, 0002, South Africa*

[2] *Forest Biotechnology Group, Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC, 27695*

This chapter has been prepared in the format of a manuscript for a peer-reviewed research journal, as is the standard procedure of our research group. I performed all laboratory work and data analysis of the isolation and sequencing of the poplar small RNAs. Technical training in the cloning of microRNAs was done at North Carolina State University under the supervision of Prof. V.L. Chiang, who provided the tissues and certain facilities used in this study. Dr. S. Lu assisted with technical training, guidance and expertise in the procedure for the cloning of microRNAs. Dr. Y-H. Sun did the initial searches against the *Poplar* genome for the identification of miRNA homologues, precursors and putative target loci. Prof. A. A. Myburg provided the funding, facilities, and infrastructure in South Africa. I wrote the manuscript with extensive suggestions on organization and content from Prof. Myburg.

2.1 ABSTRACT

MicroRNAs are a rapidly growing group of endogenous ~ 20 to 24 nt RNA molecules with validated roles in a number of developmental processes in plants and animals. There are over 700-recorded miRNAs from nine different plant species, and this number is rapidly increasing. The discovery of unique stress- and environmentally-regulated miRNAs has opened up the arena for miRNA discovery under virtually any condition. It is clear that the number of miRNAs still to be discovered is far from saturated. In this study we aimed at isolating plant miRNAs from young, developing poplar plants in an attempt to isolate novel miRNAs expressed during early stages of plant development that may have not been identified in the past as most tissues used in previous studies have been mature differentiated tissues. We report on the identification of microRNAs isolated from two-month old *Populus trichocarpa* plantlets. We isolated a total of 72 poplar miRNAs. Of these 16 are putative novel miRNAs, which cluster into nine new families. The remaining 56 identified miRNAs belong to nine previously identified families. For the newly identified miRNAs, we predicted fifty-five putative target genes belonging to a large number of gene families with diverse cellular functions, ranging from transcriptional regulation, cellular metabolism and defense response. This study has allowed the identification of novel miRNAs from a unique set of tissues, and has contributed to the ever-growing number of plant-specific miRNAs.

2.2 INTRODUCTION

MicroRNAs (miRNAs) are a group of endogenous small RNA molecules with important regulatory roles in plants and animals (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001; Park et al., 2002; Reinhart et al., 2002). MiRNAs regulate the expression of protein coding genes through sequence-specific down regulation at the post-transcriptional level (Doench et al., 2003; Nakahara and Carthew, 2004; Tang, 2005). With the discovery of

so many miRNAs in the last ten years it is essential for the science community to avoid redundancy with respect to new miRNAs. The miRBase database of the Wellcome Sanger Institute maintains information on all identified miRNAs to date (Griffiths-Jones, 2004; Griffiths-Jones et al., 2006). MiRBase release 9.0 (October 2006) has raised the total number of identified, published miRNAs to 4361. This collection includes miRNAs of primates, rodents, birds, fish, worms, flies, plants and viruses. There are 863 miRNAs from plants including 242 *Oryza sativa*, 215 from *Populus trichocarpa*, 131 from *Arabidopsis*, and 96 from *Zea mays*. The miRNAs identified from plants and animals are grouped into miRNA gene families containing a varying number of members (Ambros et al., 2003). A specific miRNA family contains members with identical or near identical sequence of the mature miRNA, varying in only up to 3 nucleotides, each being encoded by distinct genomic loci. The precursors of different family members are highly variable and show little sequence homology (Dezulian et al., 2006). It has been postulated that miRNA genes may constitute approximately 1% of coding genes in most organisms, including plants (Lai, 2003; Lim et al., 2003; Bartel, 2004). It is likely that not all miRNAs have yet been discovered, possibly due to the fact that stress- or environmentally-induced miRNAs would not be expressed under most conditions (Sunkar and Zhu, 2004; Dresios et al., 2005; Lu et al., 2005; Ko et al., 2006). It is clear that more work with respect to discovery, identification, classification and functionality is necessary in order to elucidate the entire miRNA component of many organisms.

The identification of miRNAs in a wide range of different species suggests that these molecules play an important role normal cellular functioning. A large number of studies have shown that miRNAs play pivotal roles in directing developmental patterning and the determination of cell fate in a variety of developmental pathways. MiRNAs are known to direct leaf, shoot and floral tissue differentiation (McConnell et al., 2001; Aukerman and Sakai, 2003; Chen, 2004), to establish organ polarity in leaves and stems (Palatnik et al.,

2003; Kim et al., 2005), and to direct the timing of vegetative phase changes (Rhoades et al., 2002; Mallory et al., 2004). Furthermore, miRNAs have been show to be expressed in response to mechanical and environmental stress (Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004; Chiou et al., 2005; Lu et al., 2005) and in response to phytohormones (Eckardt, 2005; Guo et al., 2005; Mallory et al., 2005). Initially, it was apparent that miRNAs control developmental pathways by targeting families of transcription factors with roles in development (Reinhart et al., 2002; Rhoades et al., 2002; Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004). As an increasing number of novel miRNAs are discovered a more diverse range of targets with functions in a variety of cellular processes have been identified. In cases where transcription factors are not targeted for regulation, other proteins with roles in developmental pathways are regulated, such as proteins involved in apoptosis or programmed cell death (Bonnet et al., 2004; Sunkar and Zhu, 2004).

In this study, a small RNA library was constructed from developing *P. trichocarpa* plantlets, where many pathways are actively directing plant morphogenesis in its initial stages. The aim was to identify novel miRNAs that directly contribute to the early postembryonic developmental patterning of growing plants. Furthermore, we aimed to identify known and putative novel miRNAs in poplar using a combination of cloning and computational approaches. Although *P. trichocarpa* has the largest number of miRNAs already identified of all plant species, these have primarily been characterized in mature stem tissues (Lu et al., 2005), or using only computational prediction based on previously identified miRNAs in other species (Dezulian et al., 2006). Thus, it is reasonable to assume that there may still be unidentified miRNAs, which act during early development of the primary, postembryonic plant body of poplar trees.

Here we describe the discovery of nine new poplar miRNA families isolated from two-month old *P. trichocarpa* plantlets, which increases the number of identified miRNAs in this

species, and provide a more comprehensive list of possible cellular pathways under miRNA control. We also isolated nine miRNA families that have been identified in previous studies and are conserved in other plant species. These miRNAs have homologues in *Arabidopsis*, rice, maize, medicago or lotus, suggesting that these have conserved functions, whereas, the non-conserved miRNAs may be involved in species-specific gene regulatory events and pathways.

2.3 MATERIALS AND METHODS

2.3.1 Small RNA isolation

Two-month-old *Populus trichocarpa* (clone Nisqually-1) whole plantlets were grown on MS medium (0.5X MS macronutrients and micronutrients, 1X MS vitamins, 100 mM myo-inositol, 20 M sucrose, 7.5 M agar, pH5.5-5.8) in a growth room at 25°C in 16 hour light/8 hour dark periods. Approximately three to four whole plantlets (~ 5 g fresh weight), including roots, were ground to a fine powder in liquid nitrogen and used immediately for RNA isolation. Total RNA was extracted from the pooled plantlet tissue using the CTAB RNA extraction method as described by Chang et al. (1993), and assayed by MOPS denaturing 1.2% agarose gel electrophoresis to verify RNA quality. Small RNAs were isolated as previously described (Elbashir et al., 2001; Lau et al., 2001) with certain modifications. Briefly, total RNA was resolved on a 12% denaturing (6 M Urea) polyacrylamide gel using a 10 bp DNA ladder (Invitrogen, Carlsbad, CA) as a reference. A gel region including a size range of ~ 15 – 30 nt was excised and the RNA was eluted overnight in 0.3 M NaCl at 4°C with agitation. Small RNAs were recovered by ethanol precipitation with 15 μg glycoblue (Ambion, Austin, TX) at -80°C for more than two hours.

2.3.2 Cloning of small RNAs

The recovered small RNAs were dephosphorylated using alkaline phosphatase for 60 min at 50°C in a 50 μl reaction volume containing 50 mM Tris-HCl (pH 9.3), 1 mM MgCl$_2$, 1 mM spermidine, 0.1 mM ZnCl$_2$, and 10 U calf intestinal alkaline phosphatase (Fisher Scientific, Hampton, NH). The reaction was stopped by phenol/chloroform extraction, and the RNA was recovered by ethanol precipitation with 0.3 M NaCl and 15 μg glycoblue (Ambion). A 5'-phosphorylated 3'-adaptor oligonucleotide (5' – pCTGTA<u>GGCACC</u>ATTCATCACx – 3'; p, phosphate; x, 3'-amino-modifier C-7; underlined bases indicate *BanI* restriction enzyme site; MWG Biotech, Highpoint, NC) was ligated for 60 min at 37°C to the dephosphorylated small RNAs in a 20 μl reaction volume containing 20 μM 3'- adaptor, 50 mM Tris-HCl, 10 mM MgCl$_2$, 10 mM DTT, 1 mM ATP, and 40 U T4 RNA ligase (New England Biolabs (NEB), Beverly, MA). The ligation reaction was stopped with the addition of RNA loading dye and the entire ligation reaction resolved in a 12 % denaturing polyacrylamide gel.

The ligation product (~ 40 nt) was separated from the unligated product (~ 21 – 25 nt) and a gel band spanning 35 – 50 nt was excised, eluted overnight and recovered as described above. The ligated RNA product was then 5'- phosphorylated in a 50 μl reaction at 37°C for 60 min in the presence of 50 mM Tris-HCl, 10 mM MgCl$_2$, 10 mM DTT, 1 mM ATP and 20 U T4 polynucleotide kinase (NEB). The phosphorylation reaction was extracted with phenol/chloroform, and the RNA recovered by ethanol precipitation. A 5' modified adaptor oligonucleotide with 5' and 3' hydroxyl groups (5'–ATGTCGTGa<u>ggcacc</u>tgaaa– 3', uppercase DNA, lowercase RNA, underlined bases indicate *BanI* restriction enzyme site; Dharmacon, Chicago, IL) was ligated to the phosphorylated product as described above in a 20 μl reaction at 37°C for 60 min. The final ligation product (~ 50 – 70 nt) was size-selected on a 10% denaturing polyacrylamide gel, excised and eluted from the gel slice. The purified ligation product was used for first strand cDNA synthesis in a reverse transcription reaction using the

59

RT primer (5' – GATGAAT<u>GGTGCC</u>TAC – 3', underlined bases indicate *BanI* restriction enzyme site) for 60 min at 42°C in a 60 μl reaction containing 5 μM RT primer, 75 mM KCl, 50 mM Tris-HCl, 3 mM MgCl$_2$, 10 mM DTT, 40 U RNase inhibitor, and 60 U Moloney Murine Leukemia Virus (M-Mulv) Reverse transcriptase (NEB).

The entire reverse transcription reaction was used in 12 to 16 separate, PCR reactions using the 5' PCR primer (5' – GTCGGA<u>GGCACC</u>GAAA – 3', underlined bases indicate *BanI* restriction enzyme site) and the 3' PCR primer (5' – GATGAAT<u>GGTGCC</u>TACAG – 3', underlined bases indicate *BanI* restriction enzyme site). The thermal cycling conditions were as follows: initial denaturation at 95°C for 2 min; followed by 31 cycles of 95°C for 30 sec, 50°C for 30 sec, 72°C for 30 sec, and a final elongation step of 7 min at 72°C. The PCR products were pooled and purified by phenol/chloroform extraction and recovered by ethanol precipitation. The entire recovered PCR product was digested with *BanI* restriction enzyme at 37°C for ~ 16 hours (300 μl reaction, 20 mM Tris-acetate, 50 mM potassium acetate, 10 mM Magnesium acetate, 1 mM DTT, and 100 U *BanI*; NEB). The reaction was stopped with a phenol/chloroform extraction and the digested cDNA recovered by ethanol precipitation.

Small fragments and adaptor digestion products were removed using CENTRI.SPIN™ - 20 columns (Princeton Separations, Adelphia, NJ) as per the manufacturer's instructions. Digested products were concatamerized using T4 DNA ligase (20 μl reaction, room temperature for 1 hour, 30 mM Tris-HCl (pH 7.8), 10 mM MgCl$_2$, 10 mM DTT, 1 mM ATP, 10 U T4 DNA ligase; Fisher Scientific). Concatamers were separated on a 2% agarose gel and products ranging from 400 – 700 bp were excised and recovered from the gel slice using the QIAquick® gel extraction kit (QIAGEN, Valencia, CA). The unpaired ends were filled in by incubation with *Taq* DNA polymerase at 72°C for 30 min. The DNA product was directly cloned into the pCR2.1-TOPO vector using the TOPO TA cloning kit (Invitrogen) as per the manufacturer's instructions. Plasmid DNA was isolated from positive colonies using the

R.E.A.L prep 96-plasmid isolation kit (QIAGEN) as per manufacturer's protocol. Sequencing reactions were performed with the M13 forward primer, using the BigDye Version 3.1 sequencing kit (Applied Biosystems, Foster City, CA) and analyzed on an automated DNA 3700 ABI sequencer (Applied Biosystems).

Small RNA sequences were identified in the longer concatamer sequences via manual inspection. Each cloned small RNA sequence is flanked by a repetitive unit comprising identifiable adaptor sequence. Any sequence flanked by appropriate adaptor sequence was treated as a potential miRNA sequence. At this point, only sequences in the size range of 19 to 24 nt were used for further analysis. Small RNAs were named according to the naming convention for microRNAs suggested by Ambros et al (2003) and Griffiths-Jones et al (2004). However, small RNA identity was initially assigned according to the order of identification. A numerical system was used with the prefix ptc-miR.

2.3.3 Prediction of stem-loop structures

True miRNAs can be distinguished from non-miRNAs in that miRNA precursor molecules must be able to form fold-back stem-loop structures (Lee et al., 2002; Zeng and Cullen, 2003). The identified small RNA sequences were used to search the poplar genome to identify putative precursors, which were used for stem-loop structure prediction. Briefly, the cloned small RNA sequences were used in a search against the *Populus* draft genome assembly (http://genome.jgi-psf.org/Poptr1/Poptr1.home), *Populus* ESTs (http://www.ncbi.nlm.nih.gov/ dbEST), *Arabidopsis thaliana* genome annotation version 5.0 (ftp://ftp.tigr.org/pub/data/ athaliana/ath1/), and *Oryza sativa* ssp *japonica* cv Nipponbare genome assembly release 3 (http://www.tigr.org/tdb/e2k1/osa1/index.shtml) using PatScan (Dsouza et al., 1997). Secondary structures were predicted using the mfold programme (www.bioinfo.rpi.edu/ applications/mfold/old/rna/form1.cgi, version 3.2) with default parameters (Mathews et al., 1999; Mansfield et al., 2004). In each case, only the lowest energy structure was selected for

manual inspection, as described by Reinhart et al (2002). Small RNA sequences were folded with flanking sequences in five contexts: 300 bp upstream and 20 bp downstream; 150 bp upstream and 20 bp downstream; 150 bp upstream and 150 bp downstream; 20 bp upstream and 150 bp downstream; 20 bp upstream and 300 bp downstream.

2.3.4 Prediction of poplar miRNA targets

Putative targets of the identified ptr-miRNAs were predicted from 58,036 gene models of the current *P. trichocarpa* draft genome assembly (http://genome.jgi-psf.org/Poptr1/Poptr1. home). PatScan (Dsouza et al., 1997) was first used to identify predicted transcripts containing sequence complementary to the cloned small RNA sequences, allowing up to 3 mismatches or bulges. The scoring scheme developed by Jones-Rhodes and Bartel (2004) was used to screen for authentic miRNA targets . Functions of the predicted targets were assigned manually based on either the eukaryotic orthologous group annotation of the *Populus* genome or the function of the best hit from the BLAST homology search (Altschul et al., 1997) against the *Arabidopsis* annotation version 5.0 peptide sequence database. For predicted targets with unknown gene function, the European Bioinformatics Institute InterPro database (http://www.ebi.ac.uk/ interpro) was used to search for possible gene functions based on identified domains of the target gene.

2.4 RESULTS

2.4.1 Identification of miRNAs from *Populus trichocarpa*

Small RNAs in the size range of approximately 17 – 34 nt were cloned from *P. trichocarpa in vivo* whole plantlets using a method designed to enrich for DICER cleavage products, which contain 5'-phosphate and 3'-hydroxyl groups (Bernstein et al., 2001; Hutvagner et al., 2001). Small RNAs were size selected by gel purification from total RNA and modified adaptors were ligated to the 5' and 3' ends respectively to maintain the unidirectional orientation of the

molecule. Small RNAs were then reverse transcribed, PCR amplified, concatamerized and cloned into a plasmid vector as described previously (Elbashir et al., 2001; Lau et al., 2001; Lagos-Quintana et al., 2002). Plasmid DNA was isolated and sequenced from a total of 384 individual clones, yielding 1311 small RNA sequences. BLASTn analyses of sequences against the GenBank database (http://www.ncbi.nlm.nih.gov/BLAST) resulted in 756 matches against known non-coding rRNA, tRNA, snRNA and retrotransposon sequences, which were considered to be breakdown products of these molecules and these small RNA sequences were excluded from further analyses. Of the remaining sequences, only unique sequences in the size range of 19 to 24 nt were analyzed further.

The ability of precursor miRNAs to form fold-back stem-loop structures containing the mature miRNA on one arm of the stem-loop is a prerequisite for accurate prediction of true miRNAs (Lee et al., 2002; Zeng and Cullen, 2003). Thus, the remaining 141 sequences were used for stem-loop structure prediction. Approximately 300 nucleotides of poplar genome sequence flanking each cloned small RNA sequence was used for analysis in the mfold programme (Mathews et al., 1999; Zuker, 2003) to predict if stable stem-loop structures, representative of pre-miRNA molecules, could form. Possible stem-loop structures were predicted in four orientations and, via manual inspection, only those sequences that could form stem-loops were denoted as pre-miRNA (precursor) sequences. For 124 of the 143 sequences, no stem-loop structures could be identified, and therefore these were excluded from further analysis, as they could not be considered true miRNAs. They could, however, be considered putative endogenous siRNAs (Bartel, 2004) and are listed in Supplemental Table S2.1, along with the putative target genes of these siRNAs. The remaining 19 small RNA sequences were capable of forming stable stem-loop structures in the context of the surrounding genomic sequences. These 19 sequences were considered to be true miRNAs,

63

and denoted as ptc-miRNAs. The cloned miRNA sequences are shown in bold in Tables 2.1 and 2.2, and all precursor sequences are shown in Supplemental Table S2.2.

BLAST analysis of the 19 putative miRNA sequences against the NCBI database and the Sanger Institute miRBase miRNA registry release 9.0 at http://www.microrna.sanger.ac.uk (Griffiths-Jones, 2004; Griffiths-Jones et al., 2006) revealed that ten of the cloned sequences showed identity to known miRNAs identified previously from *P. trichocarpa* and other plant species (Reinhart et al., 2002; Rhoades et al., 2002; Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004; Lu et al., 2005). These miRNAs belong to nine miRNA families namely: ptc-miR156, ptc-miR159, ptc-miR166, ptc-miR167, ptc-miR396, ptc-miR398, ptc-miR408, ptc-miR472 and ptc-miR475, as shown in Table 2.2. The remaining nine small RNA sequences are considered to be new miRNAs (Table 2.1), which have not been identified in other species. These are ptc-miR26, ptc-miR39, ptc-miR44, ptc-miR56, ptc-miR69, ptc-miR81, ptc-miR99, ptc-miR170, and ptc-miR235. The naming of the new ptc-miRNAs was based on the order of cloning. For the ptc-miRNAs belonging to families identified previously, we adhered to the naming convention assigned by the original authors. Cloning frequencies and genome locations of each of the ptc-miRNAs are included in Tables 2.1 and 2.2.

2.4.2 Identification of homologs and gene families of ptc-miRNAs

In order to identify any possible homologues (further family members) of the cloned ptc-miRNAs (known and previously unknown), we searched the *Populus* genome using PatScan for sequences similar to the 19 cloned ptc-miRNAs. Up to three nucleotide substitutions from the cloned sequence was allowed as this is the largest number of mismatches observed between miRNA family members to date. Mfold analysis was performed, as described above, on any possible homologues in order to identify stem-loop structures. This analysis revealed 16 new miRNAs belonging to the nine newly identified miRNA gene families (denoted as ptc-miRs), and 56 family members belonging to the nine previously identified miRNA

families. All ptc-miRNA homologues are listed in Table 2.1 and Table 2.2. The precursor stem-loop structures for all ptc-miRNA homologues of the new families are given in Figure 2.1, and the sequences and genomic locations for each of the 16 ptc-miRNAs are given in Supplementary Table 2.2. Figure 2.2 shows the precursor structures for the 56 ptc-miRNAs previously identified, and their sequences and genomic locations are given in Supplementary table 2.2. As with metazoans, the mature miRNA molecule can be produced from either the 5' or 3' arm of the fold-back precursor. However, as seen in other studies of miRNAs in plants (Reinhart et al., 2002; Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004; Wang et al., 2004; Lu et al., 2005), all of the miRNAs identified with multiple matches to the genome were always present on the same arm of their potential precursors, suggesting that these loci share a common ancestry (Table 2.1 and Table 2.2). Four of the miRNA sequences had single copies in the poplar genome, whereas the rest corresponded to multiple (2 to 4) genomic loci (Table 2.1).

We did not identify any new members for previously identified ptc-miRNA families. The reason for this is that homologue searches were performed on the miRNA families identified in previous studies (Lu et al., 2005; Dezulian et al., 2006) and all possible miRNAs have already been identified. Naming conventions for homologues of known miRNA family members was determined by sequence similarity to previously identified ptc-miRNAs. The identified precursor sequences and genomic locations were also identical for our identified known ptc-miRNAs and those identified in other studies (Lu et al., 2005; Dezulian et al., 2006).

The number of miRNAs present within each family varied greatly between families. For instance, the ptc-miR99 family had only one family member, whereas the ptc-miR166 and ptc-miR396 families contained 17 and 11 members respectively. Although the family identity and miRNA sequence is conserved across species, the number of miRNAs in each family

seems to differ between species. For example, the *Arabidopsis* miR166 family contains seven members, whereas in poplar this family contains 17 members. However, in the miR398 and miR408 families, the number of miRNAs is largely conserved across *Arabidopsis*, rice, poplar, sorghum and maize (miRBase). Furthermore, miR472 and miR475 are specific to poplar and have not been identified in other plant species analyzed.

2.4.3 Characteristics of ptr-miRNA precursors

The pre-miRNA stem-loop structures presented in Figures 2.1 and 2.2 represent only the miRNA/loop/miRNA* portion of each precursor, which was determined manually. The loop segment refers to the sequence between the miRNA and miRNA* regions. We considered the longest possible fold-back structure in each case and only depicted this as the pre-miRNA. However, this does not represent the full-length transcribed miRNA precursor molecule, the extent of which is unknown. In most cases, the sequence beyond the predicted stem-loop structure was not predicted to be capable of forming any significant secondary structures.

We further analysed the structure of the identified plant miRNA precursors. The precursor sequences of each member within a family (where the family contained four or more members) were compared using the multiple alignment programme ClustalW (http://www.ebi.ac.uk/clustalw/; Chenna et al., 2003). As expected, the mature miRNA sequence itself was highly conserved within families, with the maximum deviation between two sequences of two nucleotides. The miRNA* sequence was the next most conserved region between precursor molecules, which was expected due to its requirement to anneal to the mature miRNA within the stem-loop (Figure 2.3). However, these sequences do show less conservation than the miRNA itself due to the occurrence of G: U wobble pairing and mismatches within bulges.

The loop sequences and sequences upstream and downstream from the miRNA and miRNA* sequences varied in length and sequence within miRNA families (Figure 2.3). No

conservation in structure or sequence was observed in these regions for the different precursor molecules in families ptc-miR235, ptc-miR398 and ptc-miR475. An interesting exception was observed for the ptc-miR159 family where the precursors showed a very large degree of sequence complementarity. However, within some of the larger families, such as ptc-miR166, ptc-miR167 and ptc-miR396, there seems to be a degree of sequence, and structural, conservation among family members. For example, ptc-miR166b and ptc-miR166e have almost identical precursors, as is the case for ptc-miR166h and ptc-miR166i. Furthermore, the family appears to be divided into two distinct types of precursors. For ptc-miR166a-e, g-j, and m all have simple stem-loop structures, however for ptc-miR166f, k, l, o, p and q the stem-loop structures are more complex with two or three secondary stem-loops at the terminal of the main arm. This was also observed in the ptc-miR396 family. Despite these observations, analysis using ClustalW and phylogram analysis (results not shown) yielded no correlation between type of loop and precursor sequence neither, within or between families of different species. The reason for this observation could possibly be functional, where the precursor secondary structure might play a role in differential DICER recognition (Zeng et al., 2005).

2.4.4 Prediction of ptr-miRNA targets

Potential targets of the newly identified ptr-miRNAs were computationally predicted using methods previously described which rely on a scoring system designed to allow mismatches and bulges between the mature miRNA and its target sequence (Jones-Rhoades and Bartel, 2004). Target prediction for plant miRNAs is facilitated by the fact that they exhibit near perfect antisense complementarity to their mRNA targets (Rhoades et al., 2002; Bonnet et al., 2004; Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004; Wang et al., 2004; Adai et al., 2005). To identify potential targets, we used the miRNA sequences to search the complete gene model set collected from the currently annotated Populus genome (http://genome.jgi-

psf.org/Poptr1/Poptr1.home.html), which includes transcript sequences of 58,036 gene models. The scoring system is based on assigning penalties to the mismatch pattern between the miRNA and possible target mRNA molecule. Each complementary pair is assigned 0 points; G: U wobble pairs are assigned 0.5 points; each non-G: U wobble mismatches are assigned 1 point, and 2 points are given to each bulged nucleotide in either RNA strand. From here a total score across the miRNA: mRNA duplex of 3.0 or less was considered as a match to an authentic target with very high confidence (Jones-Rhoades and Bartel, 2004). This scoring system was applied during target predication using PatScan (Dsouza et al., 1997).

Once potential targets conforming to the above scoring system were identified, manual inspection of the mismatch patterns between the miRNA and its target molecule was performed to further eliminate the number of non-authentic targets. The criteria used were that only one mismatch was allowed in the region complementary to nucleotides 1 to 9 in the 5' end of the miRNA, but none in the miRNA cleavage site (nucleotides 10 and 11) and two additional mismatches were allowed in nucleotides 12 to 21, but not more than 2 continuous mismatches in these regions (Palatnik et al., 2003; Sunkar et al., 2005)

Using the criteria described above, we identified 55 putative target loci for the nine newly identified ptc-miRNA families (Table 2.3). For the nine previously identified ptc-miRNA families we identified 120 putative target loci (Table 2.3), which correlated well with the putative targets already predicted for these families. Since the majority of family members in each family were computationally predicted, and the putative targets for the new ptc-miRNAs have not been experimentally verified, it is impossible to predict which specific miRNA family member targets a specific gene of the putative target. Thus, Table 2.3 lists all possible target genes of an entire miRNA family. Further experimental validation and knock-out studies are required to elucidate any gene- or tissue-specific regulation by a single miRNA family member. Each miRNA family identified had a number of putative targets, ranging

from one to 21. The protein annotation described for each locus in was obtained either from the *P. trichocarpa* annotations or, if no predicted protein function was observed, the sequence was blasted against the TAIR database (http://www.arabidopsis.org) to identify putative *Arabidopsis* homologues. Using this, in conjunction with the EMBL-EBI database of the European Bioinformatics Institute (http://www.ebi.ac.uk/interpro), putative ptc-miRNA targets were assigned predicted functions. Of all the predicted targets, ~ 26% were members of transcription factor gene families or function in transcription regulation. A further ~ 8 % were PPR proteins, ~ 12 % were disease resistance proteins and ~ 20 % of the targets did not have gene annotations or functions assigned. These unknown targets will be prime candidates for further analysis as they may represent poplar-specific genes, and could represent novel gene pathways regulated by miRNAs specific to this woody plant.

A number of the putative targets identified for the new miRNAs in this study have been confirmed as miRNA targets in other species (Bonnet et al., 2004; Sunkar et al., 2005). Interestingly, in some cases the miRNA family regulating the gene appears to be non-conserved. This is true for ptc-miR396, ptc-miR235, ptc-miR475 and ath-miR400, all of which are predicted to target of pentatricopeptide repeat (PPR) proteins. Similarly ptc-miR39 and ath-miR407 are predicted to target short-chain dehydrogenase/reductase proteins, as is the case for ptc-miR44 and ath-miR390 (Sunkar and Zhu, 2004). Furthermore, different miRNA families appear to target different members of the same gene family or genes with similar functions, as for the ptc-miR69 and ptc-miR472 families, which are both predicted to target different disease resistance proteins.

Of the 55 target loci with known functions predicted for the newly identified ptc-miRNAs, 15 have not been described previously as miRNA targets in *Arabidopsis*, rice or poplar. These are the Choline transporter-like proteins, D111/G-patch domain containing proteins, and LIM-domain containing proteins. And also include protein phosphatase 2C,

ULT1 B-box transcription factor, DNAJ heat shock proteins, microtubule severing proteins, ceramidases and polypyrimidine tract-binding protein (Table 2.3). These targets have a variety of predicted functions, with a large proportion being related to transcription regulation or targets with roles in metabolic and cellular processes, such as electron transport, organelle biogenesis, cell cycle control, and carbohydrate and fatty acid metabolism also feature prominently. These results suggest that the roles of miRNAs in plant development are not restricted to the regulation of transcription factors as initially thought, but that they have diverse biological role in developmental events and maintenance of cellular homeostasis. It should be made clear though that these are only predicted targets based on computational identification and experimental validation of predicted targets for the novel and conserved ptr-miRNAs is necessary to confirm that the genes are indeed targets of the identified miRNAs, and to further confirm that the identified small RNAs are functional miRNAs with distinct cellular regulatory functions.

2.5 DISCUSSION

The discovery of miRNAs and their essential role in the control of gene expression through post-transcriptional gene silencing in a wide variety of organisms has opened up a new area of research in the field of gene regulation. In order to fully understand the mechanism of gene expression regulation, developmental patterning and pathways, cellular homeostasis, and environmental- and stress-responses of any organism, it is essential to understand how miRNAs could affect these pathways. As more novel miRNAs are discovered, especially those exclusively expressed under specific conditions, it is becoming clear that miRNAs regulate a more diverse range of pathways than initially hypothesized.

We have successfully isolated and characterized 72 miRNAs belonging to 18 distinct miRNA families from *P. trichocarpa*. Of these, 56 miRNAs belong to nine known miRNA families. The remaining 16 miRNAs, group into nine new putative families, which have not

yet been described in *P. trichocarpa*, or other plant species, and are considered novel miRNAs. As expected, the predicted targets of the known miRNAs are conserved in poplar and *Arabidopsis*. However, as shown previously, *P. trichocarpa* miRNAs tend to have more diverse targets relative to *Arabidopsis* miRNAs (Lu et al., 2005). We identified 55 putative target genes for the nine new miRNA families which have not been described previously. These targets cover a wide range of gene families, with functions in numerous cellular pathways, including transcription regulation and general metabolic functions. These results contribute to the rapidly growing number of miRNAs identified in plant species, and in particular those involved in early vegetative development.

The majority of plant miRNAs have been discovered using one of two approaches. The first approach is the direct cloning of small RNAs. Small RNA libraries have been made from a number of different plant species, including, *Arabidopsis*, rice, sorghum, maize and poplar, using a method designed to preferentially clone DICER cleavage products (Reinhart et al., 2002; Bonnet et al., 2004; Sunkar and Zhu, 2004; Wang et al., 2004; Lu et al., 2005; Sunkar et al., 2005; Xie et al., 2005). Using this cloning strategy, the first plant miRNAs were identified in *Arabidopsis* (Reinhart et al., 2002). As an increasing number of miRNAs were identified, it became clear that miRNA sequence conservation was high among plant species. This finding led to a second approach for miRNA identification using computational methods to search for miRNAs in sequenced genomes, which rely on their high degree of conservation (Llave et al., 2002; Reinhart et al., 2002; Li et al., 2005; Lu et al., 2005; Dezulian et al., 2006). Using this method, the identification of miRNA paralogs within a given genome and miRNA orthologues in other species could be achieved relatively quickly in comparison to the first method. Either of these approaches alone is not sufficient for the discovery of a large number of novel miRNAs. The cloning method requires mass sequencing of a large number of putative small RNAs. Using this method, only a small proportion of the cloned small RNAs

are authentic miRNAs. The bulk of the sequences cloned are usually break-down products of expressed genes, or derived from tRNA, rRNA or snRNA transcripts. The bioinformatics approach is very useful for the discovery of miRNA orthologs or paralogs, but it is less useful for the discovery of novel miRNAs. Thus, a combination of these two approaches, as utilized in this study, provides a more thorough method for the identification of whole miRNA families in a specific species or tissue.

2.5.1 Newly identified poplar miRNAs are true plant miRNAs

The miRNAs identified in this study conform to the characteristics of plant miRNAs identified in previous studies (Reinhart et al., 2002; Sunkar and Zhu, 2004; Lu et al., 2005; Sunkar et al., 2005). All of the putative miRNAs identified are in the size range from 20 to 22 nucleotides, the most common size range of reported miRNAs (Table 2.1). It is possible that due to the screening process applied in this study, a number of miRNAs could have been overlooked due to the size exclusion criteria, as validated miRNAs have been reported being as short as 17 nucleotides in length (Lu et al., 2005). However, this not the most common size range and these miRNAs were identified using functional analysis. Furthermore, most of the identified miRNAs begin with a Uracil, a trend observed in other plant miRNAs (Reinhart et al., 2002). No new family members of the known miRNAs were discovered here, which is due to the fact that these were identified through exhaustive genome searches and all homologues have already been found (Lu et al., 2005; Dezulian et al., 2006). Of the known poplar miRNA families, only nine out of 32 were cloned in this study. This is due to the fact that only a small proportion of the small RNA population isolated was sequenced. Therefore, miRNAs that are known to be highly expressed, such as ptc-miR159, were cloned multiple times, whilst miRNAs that were expressed at low levels in the early vegetative tissues may not have been sampled in the fraction of small RNAs sequenced in this study.

Of all the known miRNAs, the mature miRNA sequences of family members is highly conserved, with each member showing only up to four nucleotides difference from other members (miRBase, 2006, http://www.microrna.sanger.ac.uk). This phenomenon was also observed for the miRNA families discovered here. Of the miRNAs we cloned, the largest amount of sequence variation between family members was observed in the ptc-miR235 family, exhibiting four nucleotide differences between the two members. However, in the ptc-miR56 family, the sequences were highly conserved, with all four family members having identical sequences (Table 2.1).

Animal miRNA precursor molecules are generally conserved in length and are almost always approximately 70 nucleotides long, whereas in plants they can range from 50 to 500 nucleotides (Reinhart et al., 2002; Bartel, 2004). The size distribution of miRNA precursors discovered in this study falls within this range, with stem-loops ranging from 57 nucleotides for ptc-miR39b to 332 nucleotides for ptc-miR166n (Table S2.2). These results indicate that the small RNA sequences identified here fulfil the criteria of true miRNAs, which are: 1) that the surrounding genomic sequence can form a stable fold-back stem-loop structure, 2) the sequence is highly conserved within families and 3) the sequence shows high complementarity to target genes. The last point is known to be particularly characteristic of plant miRNAs, whereas animal miRNAs show a lower degree of complementarity to their target genes. The only remaining proof of miRNA activity is expression and functional analysis. Expression analysis using real-time PCR will allow confirmation of miRNA expression and the identification of tissue-specific expression patterns of a given miRNA family. Functional analysis entails the investigation into the regulation of target genes by a miRNA family; this includes confirmation of target cleavage using 5' rapid amplification of cDNA ends (RACE). However, this is beyond the scope of this current study and will form the basis of future work on the identified poplar miRNAs.

2.5.2 Newly identified poplar miRNAs could regulate diverse pathways

As more miRNAs are discovered, it is becoming clear that miRNAs regulate target genes with functions in more diverse pathways than initially expected. Transcription factor gene families were originally identified as the primary targets of most miRNAs (reviewed in Jones-Rhoades et al., 2006), and they still make up a large proportion of newly identified target genes. This could be due to the fact that these targets are regulated by the more common, or highly expressed, miRNAs such as miR159 regulating the large family of MYB transcription factors. However, as miRNAs are now being isolated from specific tissues or under specific stress conditions, genes that are not transcription factors are identified as targets of these new miRNAs.

A general impression has emerged regarding miRNA function in the cellular environment, where targeted miRNA expression and subsequent target down-regulation allows organ differentiation and localized silencing in specific plant tissues. For example, in *Arabidopsis* miR166 is involved in the determination of abaxial and adaxial leaf cell fate (Rhoades et al., 2002; Mallory et al., 2004; Zhong and Ye, 2004). And ath-miR172 is involved in floral organ development through regulation of the negative regulator of flowering, AP2 (Aukerman and Sakai, 2003; Chen, 2004). Furthermore, it has been discovered that many miRNAs are selectively expressed under certain environmental conditions, such as in response to hormones, chemical compounds and stress (Achard et al., 2004; Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004; Eckardt, 2005; Guo et al., 2005; Lu et al., 2005). Thus, it could be postulated that miRNAs are also involved in a similar manner during early vegetative development of poplar plantlets. The gene targets identified for the known miRNAs have been extensively characterized and validated as true miRNA targets and will not be discussed in depth here. Instead, we will discuss the new ptc-

miRNA targets and their predicted functions, including novel targets of known miRNA families.

We identified 46 putative target genes with roles in transcriptional regulation, some of which have not been previously identified as miRNA targets in any plant species. Ptc-miR81 targets a B-box transcription factor, whose *Arabidopsis* homologue, *ULTRAPETALA1* (*ULT1*), negatively regulates meristem cell accumulation in inflorescence and floral meristems (Fletcher, 2001; Carles et al., 2005). Ptc-miR99 has one putative target locus that is uncharacterised in poplar, although this gene shows high similarity to the *Arabidopsis ACCELERATED CELL DEATH1 (ACD1)* gene. ACD1 is involved in pathogen defence responses, flower and fruit development, and senescence, with expression localized to fruit and flowers and in response to ethylene (Tanaka et al., 2003). The newly identified ptc-miR69 family is predicted to target three transcription factor genes, where two are members of the zinc-finger (B-box) protein family involved in transcription regulation. The third is a LIM-domain containing protein, and although the exact function of this target gene is unknown in poplar and *Arabidopsis,* the motif is a zinc-ion binding domain characteristic of transcription factors (Michelsen et al., 1993).

Many of the known ptc-miRNA families have validated targets that are involved in transcription regulation. Four members of the MYB transcription factor family, were identified here as targets of ptc-miR159. Ptc-miR166 targets ten members for the HD-ZIP III transcription factor family and ptc-miR396 is known to target eleven loci that are all putative growth regulating factors (GRF) with transcriptional activation functions, and have roles in the regulation of cell expansion in leaf and cotyledon tissues (Kim et al., 2003). A member of the MADS-box transcription factor family is a newly identified putative target of the known family ptc-miR396. This group of transcription factors are involved in leaf and flower development, meristem identity, root development, fruit dehiscence and flowering time (Coen

and Meyerwitz, 1991; Weigel and Meyerwitz, 1994; Riechmann and Meyerwitz, 1997; Hartmann et al., 2000; Theissen et al., 2000). Transcription factors are essential for controlled and localized developmental patterning during plant growth, which makes them prime candidates as miRNA targets. Many of the transcription factors described above have specific roles during developmental events and exhibit tissue-specific expression patterns. Their localized expression is essential for proper plant development, and miRNA regulation of these genes would allow a "fine tuning" of their activity in defined cell types.

A number of putative targets of ptr-miRNAs are involved in RNA binding or processing. Ptc-miR69 targets five D111/G-patch domain containing proteins, where this domain occurs in a number of putative RNA-binding proteins, suggesting a RNA binding function (Aravind and Koonin, 1999). Ptc-miR235 is predicted to target a polypyrimidine tract-binding protein and ptc-miR26 is predicted to target two RRM motif-containing proteins. These targets all contain an RNA recognition motif, the RNA-binding region RNP1, which is predicted to be involved in single-stranded RNA binding and export (Bandiziulis et al., 1989; Birney et al., 1993). Another RNA binding protein, a SWAP/SURP domain containing protein, is a newly identified target of a known ptr-miRNA family, ptc-miR396. The *Arabidopsis* homologue of this gene, *TOUGH* (*TGH*), is involved in RNA processing and vascular development, where mutants display developmental defects, including reduced plant height, polycotyly, and reduced vascularization of the stem (Calderon-Villalobos et al., 2005), and thus seems to have an important role in normal plant development. These results suggest that miRNAs are not only involved in regulation of gene expression during transcription, but may also be acting during post-transcriptional events in order to regulate gene expression.

A number of the putative target loci had predicted roles in different aspects of cellular metabolic functions. An acyl-CoA oxidase, an enzyme catalyzing the first step of fatty acid beta-oxidation, is a predicted target of ptr-miR39. The *Arabidopsis* homologue of this gene is

ACX1, which is induced by wounding, drought stress, abscisic acid, and jasmonate. ptc-miR170 targets a member of the glycosyl hydrolase protein family 10, which functions in carbohydrate metabolism, and the degradation of cellulose and xylans (Beguin, 1990; Davies and Henrissat, 1995). A new target for ptc-miR396 is a neutral/alkaline nonlysosomal ceramidase with a proposed function in sphingolipid signalling. MicroRNA control of these genes would allow targeted gene silencing and only allow expression when the substrate of the enzymes is present and expression is necessary.

Ptc-miR44, ptc-miR156 and ptc-miR170 are predicted to target six different transporter proteins. Among the different families of transporter proteins, only two occur ubiquitously in all organisms. These are the ATP-binding cassette (ABC) superfamily and the major facilitator superfamily (MFS), which together control the influx and efflux of substrates across cellular membranes. Ptr-miR44 targets three ABC transporter proteins, involved in the metabolism and catabolism of secondary metabolites. Ptr-miR156 is known to target a high-affinity nitrate transporter, which is part of the major facilitator superfamily, and enables directed movement of nitrate ions within or between cells in response to chemiosmotic ion gradients (Pao et al., 1998; Saier et al., 1999). Ptc-miR170 is predicted to target two more transporter proteins of the sexual differentiation process protein ISP4 family. ISP4 has oligopeptide transporter domains and enables directed movement of oligopeptides into, out of, within or between cells. In *Arabidopsis*, different members of this protein family show differential expression patterns, and have distinct cellular roles in nitrogen metabolism during germination and senescence, pollen tube growth, pollen and ovule development, seed formation and metal transport (Koh et al., 2002; Stacey et al., 2002a; Stacey et al., 2002b; Stacey et al., 2006).

Together, three identified ptr-miRNAs target 22 different members of the pentatricopeptide repeat (PPR) protein family. These are ptc-miR235, ptc-miR396 and ptc-

miR475. PPR proteins occur predominantly in plants, where they appear to have essential roles in organellar RNA/DNA metabolism (Nakamura et al., 2004). PPR proteins contain a tandem array of repeats, where the number of PPR motifs controls the affinity and specificity of the protein for RNA. It has been hypothesised that each PPR protein is a gene-specific regulator of RNA metabolism in plant organelles, such as chloroplasts and mitochondria, probably by binding organellar transcripts (Lurin et al., 2004).

A number of the newly predicted ptc-miR targets have defined functions in electron transfer and transport. Ptc-miR39 targets a member of the thioredoxin gene family, which is involved in redox control in the chloroplast. A plastocyanin-like protein is a validated target of ptc-miR408, and is involved in electron transport and copper binding. The plant plastocyanins exchange electrons with cytochrome c6, and have predicted roles in lignin polymerizaion (Nersissian et al., 1998). Furthermore, a known target of ptc-miR398 is a respiratory chain complex cytochrome c oxidase Vb subunit involved in electron transfer from cytochrome c to oxygen (Capaldi et al., 1983).

Cellular homeostasis is dependant on the ability of the cellular machinery to synthesize proteins correctly and efficiently, and on its ability to clear incorrect protein products from the cell. There are a number of pathways controlling this process, where molecular chaperones recognise and assist in the refolding of misfolded or unfolded proteins, especially in times of cellular stress, and the ubiquitination pathway, which leads to the targeted degradation of abnormal proteins. A predicted molecular chaperone, belonging to the DNAJ protein superfamily, is another predicted target of ptc-miR396. This target gene product is involved in protein folding through chaperone activity, where it is thought that the DNAJ domain interacts selectively with heat shock proteins, and any protein synthesized in response to heat shock. A predicted E3 ubiquitin ligase is a putative target of ptc-miR44, and is located within the ubiquitin ligase complex where it determines the substrate specificity for

ubiquitination (Stone et al., 2005). Furthermore, ptc-miR472 is predicted to target a cyclin-like F-box domain containing protein, which is predicted to be involved in the ubiquitination of target proteins, where they serve as a link between a target protein and a ubiquitin-conjugating enzyme (Bai et al., 1996; Skowyra et al., 1997; Craig and Tyers, 1999). A new target of the known family, ptc-miR396, is a WD-40 repeat containing microtubule severing protein, katanin p80 subunit B. The most common function of all WD-repeat proteins is coordinating multi-protein complex assemblies (E.g. E3 ubiquitin ligase complex), where the repeating units serve as a scaffold for protein interactions (Smith et al., 1999; Li and Roberts, 2001). Ptc-miR56 is predicted to target an ankyrin-domain containing protein, where this domain is one of the most common protein-protein interaction motifs in nature. The repeat has been found in proteins of diverse functions such as transcriptional initiators, cell-cycle regulators, cytoskeletal, ion transporters and signal transducers (reviewed by Bork, 1993). Ptc-miR26 is predicted to target a gene of the dynamin protein family. Dynamin is a GTP-hydrolysing protein that is a molecular regulator of clathrin-mediated endocytosis by cells (Sever et al., 1999).

A further two loci involved in protein metabolism are predicted targets of ptc-miR69 and ptc-miR81. A leucine rich protein kinase is a predicted target of ptc-miR69, and has a general function in protein amino acid phosphorylation, whereas a protein phosphatase 2C is a putative target of ptc-miR81, and has a predicted function in protein amino acid dephosphorylation. Ptc-miR26 is predicted to target a Ras small GTPase involved in intracellular signalling processes through cell membrane associations (Hancock, 2003). These proteins are implicated in the regulation of cell growth, proliferation and differentiation.

A number of newly identified miRNAs have been shown to be exclusively expressed in response to some external stimulus (Sunkar and Zhu, 2004; Guo et al., 2005; Lu et al., 2005; Ko et al., 2006), or have been shown to target genes involved in stress responses, where

it is important that the target genes of these miRNAs are not expressed until induced by an external stimulus. We have found a number of disease resistance proteins that are predicted targets of ptc-miR69 and ptc-miR472 where, together, these miRNAs target 26 different loci. The target loci have apoptotic ATPase activity, and contain various NBS-LRR domains present in most disease resistance proteins. This class of proteins are thought to have roles in apoptosis and defence responses.

Known targets of ptc-miR159 are two members of the asparagine synthetase family, which are involved in amino acid biosynthesis in plants. In *Arabidopsis*, asparagine synthetase expression is suppressed by light and sucrose, and is stimultated by dark and nitrogen. Control of expression in this way may favour asparagines synthesis at night and subsequently allow efficient conversion of aspartate to aspartate family amino acids during the day (Zhu-Shimoni and Galili, 1998). Newly identified putative targets of ptc-miR167 include three Hly-III related integral membrane proteins. The poplar gene sequences show high similarity to the *Arabidopsis* HHP2 (HEPTAHELICAL TRANSMEMBRANE PROTEIN2). This protein is involved in response to sucrose and hormones and thus could be a growth regulator. Furthermore, two superoxide dismutases 1 (SOD1) are known targets of ptc-miR398. This family of proteins are involved in oxidative stress responses and act to prevent damage by oxygen-mediated free radicals by catalysing the conversion of superoxide into molecular oxygen and hydrogen peroxide (Capaldi et al., 1983).

This study provides the first identification of miRNAs from somatic vegetative plant tissues. The putative ptc-miRNA targets appear to be involved in a more diverse range of pathways than those of previously identified plant miRNAs, with many of these pathways involved in developmental series that seem to form the framework of basal developmental programmes, such as electron transport and metabolism of cellular components, such as carbohydrates, proteins, fatty acids. If these targets are negative regulators of these pathways,

or involved in metabolite catabolism it would be essential to ensure the suppression of these genes during plant growth phases where cell division and protein anabolism is actively occurring, where later these processes would slow down and need to be halted. Furthermore, mature organ development and processes need to be suppressed during early seedling growth, such as floral organ development, fruit development, senescence, and secondary vascular development. Also, since the main priority of the plant is growth, certain genes that would be expressed later in the plant life cycle would need to be turned off in order to conserve energy.

However, it is important to note that the miRNAs isolated here were taken from whole plants and the tissue-specificity of each miRNA is unknown. For proper development of different plant organs it is essential that certain genes are actively expressed in one tissue or one part of a tissue and are silenced in another. Thus, the exact function of each miRNA and its interaction with the target genes cannot be identified with further testing. Additionally, it must be noted that the ptc-miRNA target genes identified here were only computationally predicted and therefore cannot be designated as unequivocal targets of the identified ptr-miRNAs. These targets will have to be experimentally validated in order to determine their exact cellular function and the contribution of miRNA regulation in overall plant development and homeostasis.

2.6 ACKNOWLEDGEMENTS

2.7 FIGURES

ptc-miR26a UCUUCAAUGCUGGUGUUAAA

```
           AU         A    CUUAAAG         CAU      A        ACC                     UA
5' UUUUAA  CUAGUAUU  GAGA           UGUUUUU   GGUAU  UUAGUUAUU  AAAAUGCAGGGGUU   A
   ||||||  ||||||||  ||||           ||||||||  |||||  |||||||||  ||||||||||||||
3' AAAAUU  GGUCGUAA  UUCU           ACGAAAAU  UCAUA  AAUCAAUAA  UUUUACGUCUUUGG   U
     GU         C    ---AAA-         UU-      A      / \  AAA               UG
                                              A   U
                                            A       A
                                            AAAAGUA
```

ptc-miR26b UCUUCAAUGCUGGUGUUAAA

```
                                  31  nt
                                  LOOP            _____
     AAU         UGA   AG          \  /    GAAA  AG   /          \
5' GCCC  UUGAGCUUUAAU  GG  AUUGAAGAUAG  UUGGUU   UC  GGAU        175 nt
   ||||  ||||||||||||  ||  ||||||||||||  ||||||   ||  ||||        COMPLEX
3' UGGG  AGCUUGAAGUUG  UC  UAACUUCUGUC  AACCAG   AG  CCUA        LOOP
     AAC        UAG GG          / \   AAAC AA   _____/
                                G     A
                                C     A
                                G     U
                                GAAAAAAA
-------------------------------------------------------------------------------
```

ptc-miR39a GCAUCCAUGGUUGAAUAGUUAA

```
     UAA-     C       AAU       ACCCAAC       AA
5' CAC   GCAU CAUGGUUG  AGUUAAAGC    UCAU   \
   |||    |||| ||||||||  ||||||||||    ||||    U
3' GUG   CGUA GUAUCAAC  UUAAUUUUG    AGUG   /
     UAAC     A       -C-         AUACUUA   GU
```

ptc-miR39b GCAUCCAUGGUUGAAUAGGUUAA

```
     UU           U    -A-    AUC
3' UUUAG  AUUCA-CCAUG AUG  GG    \
   |||||  ||||| |||||| |||   ||    C
5' AGUU  UAAGUUGGUAC UAC   CC   /
     GG         C    GAA  GCC
-------------------------------------------------------------------------------
```

ptc-miR44 UCUUUCCAACGCCUCCCAUAC

```
     GA      G         A-    C U  U-         UU
5' GCUGG  GUUAC GGUAUGGGAGG  UUGGA AG AC GCUU-GGUU   A
   |||||  ||||| |||||||||||  ||||| || || |||| ||||
3' CGACC  UAAUG CCAUACCCUCC  AACCU UC UG UGAAACCAA   A
     UC      G         GC    U U  UC         UU
-------------------------------------------------------------------------------
```

ptc-miR56a UGGUGGUGCAUGGCCGUUCUU

```
                                          _____
     GGGG-    U                   AA      /          \
5' UUCCG    AAG AUGGU--UGCACCGCUACUUA  GAAUUGACGGAAG        93 nt
   |||||    ||| |||||  ||||||||||||||  |||||||  |||||        COMPLEX
3' GAGGU    UUC UGCCGGUACGUGGUGGUGGGU  CUUAGUU-CUUUC        LOOP
     GGUUG    U                   GA           _____/
```

82

ptc-miR56b UGGUGGUGCAUGGCCGUUCUU

```
                                          18 nt
                                          LOOP                    _____
       AG      CG        AG               \ /     UUU   ---GA-   /          \
5' CCAA   AAGA   GGCUAUG   ACCAUCAUCC  GGAAUCAG   GGAG      AGU        200 nt
   ||||   ||||   |||||||   ||||||||||  ||||||||   ||||      |||        COMPLEX
3' GGUU   UUCU   CCGGUAC   UGGUGGUGGG  UCUUAGUU   UCUC      UCA         LOOP
       GA      UG        G-                        CUU   CCAGAG  _____/
```

ptc-miR56c UGGUGGUGCAUGGCCGUUCUU

```
                                       C U
                                       A   C
       A              AA       A-      \ /                                     C
5' AAC GGGAAGGCCAUGCA   ACUGC   GUAGAUCAA GAGGGAGCUUCA-CAGAGA AG-GAGAG    \
   ||| ||||||||||||||   |||||   ||||||||| ||||||||||||| ||||||| || |||||   A
3' UUG UUCUUCCGGUACGU   UGGUG   UAUCUAGUU CUUUCUCGAAGUCGUCUCU UCUCUCUC    /
       A              GG       GG                              /  \    U
                                                              C    A
                                                              U    A
                                                               CUUA
```

ptc-miR56d UGGUGGUGCAUGGCCGUUCUU

```
           AGAA              G G  G                    GAAAC      U
5' GCUUUACUGG    AGGGACGG--GUGCA UA CA CCAUGAUCGAAGGAG     AGC    U
   ||||||||||    |||||||| |||||| || || |||||||||||||||     |||    C
3' CGAGGUGGUU    UUCUUGCCGGUACGU GU GU GGUACUAGCUUUCUC     UCG   C
          -GA-          G G  G                    AUCAU      C
```

----

ptc-miR69a UCUUGCCUACUCCUCCCAUU

```
       C   C   CUU                  U                   C
5' GAGUC UAG AAGU   UGGAGAUGGGAG-AGUA GCAAGAAGGAAAAAUU    \
   ||||| ||| ||||   ||||| |||||| |||| ||||||||||||||||     A
3' CUCAG GUC UUCG   ACCUU-ACCCUCCUCAU CGUUCUUUCUUUUUAG    /
       C   U   UCU                C        /  \     U
                                          AUAA
```

ptc-miR69b UUGCCUACUCCUCCCAUUCC

```
       C   C   CUU                  U                   C
5' GAGUC UAG AAGU   UGGAGAUGGGAG-AGUA GCAAGAAGGAAAAAUU    \
   ||||| ||| ||||   |||||||||||| |||| ||||||||||||||||     A
3' CUCAG GUC UUCG   ACCUU-ACCCUCCUCAU CGUUCUUUCUUUUUAG    /
       C   U   UCU                C        /  \     U
                                          AUAA
```

----

ptc-miR81 UGCACAAAUUGAGAAAUGGU

```
        G          G  A  G       GUGCU           AACU          UGA          G
5' UAU-CAU CACAAAUUGA AA UG UCUGU       UGGCAGAGCUG    AGG-AUGGA    AAUUG    U
   ||| ||| |||||||||| || || |||||       |||||||||||    ||| |||||    |||||
3' GUAUGUA GUGUUUAAUU UU AC AGAUA       AUCGUUUCGAU    UCUGUACCU    UUGAU    G
        G          A  A  G    -AU--   / \         AGU-          CAA         A
                               UAAC
```

----

ptc-miR99 UUAGAUGACCAUCAACGAAA

```
        -C-        AUA        A          A-    AUU
5' UUGCAA    CAUGAAU    UUUGUUGAU GUCAUCUAGUG  AC     \
   ||||||    |||||||    ||||||||| |||||||||||  ||       G
3' AACGUU    GUACUUA    AAGCAACUA CAGUAGAUUAU  UG     /
        AUA        CAA        C          AC    AAA
```

----

ptc-miR170 CUGUUUGUUUUGUUUGAUGGCUGC

```
    UU                          UU        UG CU    GAUAA    C                  U              U
5' UAA   CAAAGAAGGCAC-CUGUUUGU  UGUUUGA  G  GCUU    UUCU UUUUCUAUGAGAG GGAGGUUGG  \
   |||   |||||||||| |  ||||||||  |||||||  |  ||||    |||| ||||||||||||| |||||||||   U
3' AUU   GUUUCUUUU-U-AGACAAACA  ACGAACU  U  CGGG    AAGA AAGAGAUAUUUUU UCUCUAAUA  /
    CG                          UU        GU CU    -G-A-    A                  G              U
```

----

ptc-miR235a UGGAUGUUAUGCAUGUAGGU

```
                                        _____
    GAUC     CUA      A      UA    /                  \
5' GAA    GAUGG  UGGAUGUU UGCAUG  GGUA         256 nt
   |||    |||||  |||  ||| ||| ||  ||||         COMPLEX
3' CUU    CUACC  ACC--CAA ACG-AC  CCAU          LOOP
    AGUA     AAG      A      CC    _____/
```

ptc-miR235b UGGAAGUUAUGCAUGAACGU

```
    GUA              G                UU
5' GAUCGU   GCUCGAAGUUAU GAAGUUAUGCAUGA-ACG   \
   ||||||   ||||||| |||| ||||||| ||||||  |||     U
3' UUAGCA   UGAGCU GAUA UUUUAG--CGUGCUGUGU   /
    AG-    /   \   G                GG
             104 nt
          COMPLEX LOOP
```

Figure 2.1. Predicted stem-loop structures of non-conserved *P. trichocarpa* ptc-miRNAs. Precursor structures are presented as determined by computational analysis using the mfold programme. Mature cloned miRNA sequences are shown in bold. MiRNA families are separated by dashed lines. The sequences represent only the shortest possible stem-loop structure, where only that portion of the stem-loop starting at the cloned sequence and ending at the complementary miRNA* sequence is shown.

ptc-miR156d UUGACAGAAGAGAGUGAGCAC

```
                                   ACU
      GAGU            A     A     A  \ /  UU      AC
5' U  UGACAGA-AGAG GUG GCAC CAGGGUUUC  GCAUG   \
   |  |||||||| |||| ||| |||| |||||||||  |||||    G
3' A   ACUGUCUAUCUC CAC CGUG GUUUCGAAG  CGUAC   /
      CCCC          C   C    C          UU      UU
```

ptc-miR156e UUGACAGAAGAGAGUGAGCAC

```
      U                      A        UAUUU      A
5' A  UGACAGAA-GAGAG-UGAGCAC CAGAGGCA      GUAUA  A
   |  ||||||||| ||||| ||||||| ||||||||     |||||
3' U ACUGUCUUUCUCUUUACUCGUG GUUUUCGU       CAUAU  A
      C                    C        -UAC-      U
```

ptc-miR156g UUGACAGAAGAUAGAGAGCAC

```
       U                       U      A      UGGAG
5' --UG UGACAGAAGAU-AGAGAGCACAGA GA-UGA AUGCA      \
    || |||||||||||| |||||||||||| || ||| |||||       C
3' CCAC ACUGUCUUCUGAUCUCUCGUGUUU CUCACU UACGU      /
        U                    C      C      UAAUU
```

ptc-miR156h UUGACAGAAGAUAGAGAGCAC

```
      U                      U  C     A      UGGAG
5' -UG UGACAGAAGAU-AGAGAGCAC GA GA-UGA AUGCA      \
    || |||||||||| |||||||||| || || ||| |||||       C
3' CAC ACUGUCUUCUGAUCUCUCGUG UU CUCACU UACGU      /
       U                     U  C     C      UAAUU
```

ptc-miR156i UUGACAGAAGAUAGAGAGCAC

```
        U  A                        U    UUUUG   UA
5' ---UG UG CAGAAG-AUAGAGAGCACAGA GA-UG      CAG    G
     || || |||||| |||||||||||||| || ||       |||
3' UCCAC AC GUCUUCGUAUCUCUCGUGUUU CUCAC      GUC    A
        U  C                      C    UCUAG   UC
```

ptc-miR156j UUGACAGAAGAUAGAGAGCAC

```
   U U                     U     AU     AUG
5'  G UGACAGAAG-AUAGAGAGCAC AGA GA-UG  AUGCA    G
    | |||||||||| |||||||||||||| || ||   |||||
3'  C ACUGUCUUCGUAUCUCUCGUGUUU CUCAC   UACGU     A
    - U                      C     CC     CUC
```

----

ptc-miR159a UUUGGAUUGAAGGGAGCUCU

```
                GA       U    C  CU         AUC    G    C   AAUC  A   C      CCAUCA
5' UGGAGCUCCUU  AGUCCAA  AGAAG UC  GC-UGGGUAG    GAGCU CUGAG UAUG   CC CAG CCUAUCA   \
   ||||||||||  |||||||| ||||| ||  || |||||||    ||||| ||||| ||||   || ||| |||||||     G
3' AUCUCGAGGGA  UUAGGUU UUUUC AG  UGAAUCCAUU   UUCGA GACUC AUAC   GG GUC GGGUAGU    /
             AG       U    U  AU         AU-    G    U   GUUC  C   C      UUUACU
```

ptc-miR159b UUUGGAUUGAAGGGAGCUCU
```

```
        AG            GA                U-   U          AUU      G  A        U  A  U       A  A
5' GGG UGGAGCUCCUU AGUCCAAUAGAGG UCU GCUGGGUAG   AAGCU CU AG UAUG-GA CC CAG CCUAUCU UC   A
   |||  ||||||||||  |||||| |||||  ||| ||||||||||  ||||| || || |||| ||  ||  ||| ||||||| ||
3' CCC AUCUCGAGGGA UUAGGUU-UUUCC AGA UGAUCCGUU   UUCGA GA UC AUACGUU GG GUU GGAUAGG AG   C
        AA            AG               CU   U          AU-      G  C  U        C  C  U       A  C
```

ptc-miR159c UUUGGAUUGAAGGGAGCUCU

```
        AG            GA              U-   U          AUU      G  A      U  A  U       A  A
5' GGG UGGAGCUCCUU AGUCCAAUAGAGG UCU GC-UGGGUAG   AAGCU CU AG UAUG-GA CC CAG CCUAUCU UC   A
   |||  ||||||||||  |||||| |||||  |||  || |||||||  ||||| || || |||| ||  ||  ||| ||||||| ||
3' CCC AUCUCGAGGGA UUAGGUU-UUUCC AGA UGUAUCCGUU   UUCGA GA UC AUACGUU GG GUU GGAUAGG AG   C
        AA            AG             CU   U          AU-      G  C  U        C  C  U       A  C
```

----

ptc-miR166a UCGGACCAGGCUUCAUUCCCC

```
                                                                      G
                                                                      U  A
        A        UU      CU  G    C    UGG------         UUUA        \ /    U  UUUA
5' UUG GGGGAAUG  GUCUGG  CGA GUCA UAA             GAUCUAUGAU   UCUCAAUUUGAUUU CU     A
   |||  |||||||||  ||||||  |||  ||||  |||              |||| |||||    ||||||||||||||| ||
3' AAC CCCCUUAC  CGGACC  GCU UAGU AUU             CUAG-UACUA   AGAGUUAAGUUAAA GA     A
        C        UU      AG  G    A    UAUAUUCAA        UAG-               U  UCUU
```

ptc-miR166b UCGGACCAGGCUUCAUUCCCC

```
        A  A                UC      AC     C   AGG       GAA   AA
5' GGAA GUG GGAUGUGGGGGAAUG  GUCUG  UCGAGA AAC   UAAACU  GG   \
   |||| |||  |||||| |||||||||  |||||  ||||||| |||  ||||||  ||    U
3' CCUU CAC CCUAC-UCCCCUUAC  CGGAC  GGCUCU UUG   AUUUGA  CC   /
        C  A                UU      CA     C   AGA      GG-    GG
```

ptc-miR166c UCGGACCAGGCUUCAUUCCCC

```
          CUU
          \ /  U  U      UU            A       GC--     CAU
5' GGAAGCUUAAUUG GG GAAUG  GUCUGGUUC AGG-CAUG    CACCA   \
   |||||||||||||| || |||||  |||||||||| ||| ||||   |||||    C
3' CCUUCGAAUUAAC CC CUUAC  CGGACCAGG UCCUGUAU    GUGGU   /
               U  C  UU          C     AUAA    UCU
```

ptc-miR166d UCGGACCAGGCUUCAUUCCCC

```
        UU      CU  G    U---   UC              UC--      U--      C
5' GGGGAAUG  GUCUGG  CGA GACUUU   UGU AUCAAUCU-AA   GA-ACU   UCUA   C
   ||||||||  ||||||  |||  ||||||   |||  |||||||| ||   || |||   ||||
3' CCCCUUAC  CGGACC  GCU UUGGGA   GCA UAGUUAGAAUU   CUAUGA   AGAU    U
        UU      AG  G    UUAU   C-              UAUU      UCU      G
```

ptc-miR166e UCGGACCAGGCUUCAUUCCCC

```
        UU      CU  G    --U--   UC            -UC-     -U-      C
5' GGGGAAUG  GUCUGG  CGA GACUUU   UGU AUCAAUCU-AA   GA-ACU   UCUA   C
   ||||||||  ||||||  |||  ||||||   |||  |||||||| ||   |||||| ||||
3' CCCCUUAC  CGGACC  GCU UUGGGA   GCA UAGUUAGAAUU   CUAUGA   AGAU    U
        UU      AG  G    AUUAU   -C            UAUU      UCU      G
```

ptc-miR166f UCGGACCAGGCUUCAUUCCCC

```
                                    C   CC
                                 AGGC UGG  \
                                /|||| |||   A
                               /  UCUG ACC  /
                              A   \  U   CC
```

86

```
                                            /      \
        UU   CUU           UU          |      \   AUUUAAA
5' GGAAGC  GU  UUGAGGGGAAUG  GUUUGGUUC       GGA       \
   ||||||  ||  ||||||||||||  |||||||||               A
3' CCUUCG  CA  AACUCCCCUUAC  CGGACCAGG       CCU       /
        U-   UU-          UU           \      /   GUACUAU
                                        C - U
```

ptc-miR166g UCGGACCAGGCUUCAUUCCCC

```
         A      CU    A     C    ACUUUAAGCAC
5' AGUUG GGGGAAUG  GUCUG UUCGAGA CAUUC          A
   ||||| |||||||||  ||||| ||||||| |||||
3' UCAAC CCCCUUAC  CGGAC AGGCUCU GUAAG          C
         C      UU    C     A    CUUUCUACUUA
```

ptc-miR166h UCGGACCAGGCUUCAUUCCCC

```
                                    CCU
         A      CU         C    \ /  GCACG
5' AGUUG GGGGAAUG  GUCUGGUUCGAGA CAUUCAGAAGA      \
   ||||| |||||||||  |||||||||||||| |||||||||||      C
3' UCAAC CCCCUUAC  CGGACCAGGCUCU GUGAGUUUUCU      /
         C      UU         A          ACUUA
```

ptc-miR166i UCGGACCAGGCUUCAUUCCCC

```
           UU        CU      A    AAGA     CUC
5' GUUGAGGGGAAUG  GGC-UGG  CGAAGCUU AGCA     GUUUU  \
   ||||||||||||  ||| |||  ||||||||| ||||     ||||| U
3' CAACUCCCCUUAC  UCGGACC  GCUUCGGA UUGU     CAAAG  /
           U-        AG      A    CAA-     AAC
```

ptc-miR166j UCGGACCAGGCUUCAUUCCCC

```
             UU        CU      A    AAGA     CUA
5' GUGUGUU-GAGGGGAAUG  GGC-UGG  CGAAGCUU AGCA     GUUUC  \
   ||||||||  ||||||||||  ||| |||  ||||||||| ||||     ||||| A
3' CAUACAAACUCCCCUUAC  UCGGACC  GCUUCGGA UUGU     CAAAG  /
             U-        AG      A    CAA-     UAC
```

ptc-miR166k UCGGACCAGGCUUCAUUCCCC

```
                                        A
                                      -GAG  \
                                      /  |||   A
                                      U  CUC  /
         A     U  UC        U       /   A  G
5' UGUUG GGGGA UG   GUCUGGUUCGA GUCAUUCA    /
   ||||| ||||| ||  |||||||||||| ||||||||  A
3' ACAAC CCCCU AC  CGGACCAGGCU UAGUAAGU    \       AU
         C     U  UU        U        \    -- AC    A
                                      \        ||
                                      UAUAAUG     A
                                           CA
```

ptc-miR166l UCGGACCAGGCUUCAUUCCCC

```
         AAC
          U  U
       U  \ /   A      U  CU        U            AA  U    A
5' GCA GUAU GUUG GGGGA UG   GUCUGGUUCGA GUCAUUCAUGUG  GC UUA    C
   ||| |||| |||| ||||| ||  |||||||||||| |||||||||| |||  || |||
```

87

```
3' UGU UAUAUCAAC CCCCU AC  CGGACCAGGCU UAGUGAGU-UAU  UG AAU    A
        U         C    U UU            U         GA U  U
```

ptc-miR166m UCGGACCAGGCUUCAUUCCCC

```
                                                          AG
                                            U--CUUGAUC     \
                                          /  |||||||     U
                                         /   GAACUAG     /
                                        U        \      UC
         UUCUUUUU   A       UU     CU   G      /       \
5' GGGAGC        UUG GGGGAAUG  GUCUGG  CGA GACUCU       UGUUCU
   ||||||        ||| ||||||||  ||||||  ||| ||||||       ||||||A
3' CCUUCG        AAC CCCCUUAC  CGGACC  GCU CUGGGA       AUUUCU
   -UUAUU-        C     UU        AG   G      U       /      U
                                             U  GAUCUAA A
                                              \ |||||||| |
                                              GUGUACUAGAUU U
                                                          C
```

ptc-miR166n UCGGACCAGGCUUCAUUCCUU

```
                                          _____
           UU              A    A     /              \
5' GGGGUGU  GGAAUGAAGUUUG UCC AGAUC                   259 nt
   |||||||  ||||||||||||| ||| |||||                   COMPLEX
3' UUCCACA  CCUUACUUCGGAC AGG UCUGG                    STEM-LOOP
        UU              C   C     _____/
```

ptc-miR166o UCGGACCAGGCUUCAUUCCUU

```
                                                          AUUACC
                                                 U-GGC        U
                                                /  |||         |
                                                A  CCG         G
        AG      UU      CU              C-   U   /  G ACAGUA
5' GUUG   AGGAAUG  GUCUGG  CGAGG-UCAUUGAGG  CA GAU-UAUACAGAC    /
   ||||   |||||||  ||||||  ||||| ||||||||||  || ||| |||||||||| /
3' CAAC   UCCUUAC  CGGACC  GCUCUUAGUAGCUCC  CU CUAUGUGUGUCUG  A   A
        CU       UU      AG              AU U              G GA  A
                                                           \ ||
                                                           AACU   A
                                                              U
```

ptc-miR166p UCGGACCAGGCUCCAUUCCU

```
                                                    GAUUA
                                           GCCAU        U
                                          /|||||
                                         /  CGGUA        A
                                        /   A   AAUAC
      -UAA--      AG      CGCU     GU        GG /   \       AUGA
5' AGCG       GGUUG AGGAA   GUCUGG  CGAGG-UCAU  AG      UUAUCUG   \
   ||||       ||||| |||||   ||||||  ||||| ||||  ||      ||||||    C
3' UCGU       CCAAC UCCUU   CGGACC  GCUCUUAGUA  UC      AUAGAC    /
   UUACUA        CU   ACCU      AG            AA  C    /       CCGA
                                              A  /  U  A
                                               \ UCGA GC  C
                                                \|||| ||
                                                AGUU UG  C
                                                  C  U
```

ptc-miR166q UCGGACCAGGCUUCAUUCCUU

```
                                 U  C  CU      AUUU     CA  UCC
                              G-UGC AG UAG  UUUAA     UUAAU  GA     C
                             /  ||| || |||  |||||     ||||| ||
                  UAG        U  ACG UG AUU  GGAUU     AAUUG CU     U
     CACU    \ /    A    A    CAUC     /   A  U  U   UU     CUU-  UC UCA
```

```
5' GUGAG     AGUGAAGUUUG UUC AGAUC    UCAUG     A
   |||||     |||||||||||| ||| ||||    |||||     |
3' CAUUC     UUACUUCGGAC AGG UCUAG    AGUAC     G          U
      -C--            C   C     UCA-       U     \UUAUUA    A
                                            \    ||||||
                                           ACA--AAUGAU     U
                                                          U
```

----

ptc-miR167a UGAAGCUGCCAGCAUGAUCUA

```
        AA    A     -C-               U  UU      AGAAA
5' GGGAA  AGUGA GCUG  CAGCAUGAUCUA CU  GGUUAG        G
   |||||  ||||| ||||  |||||||||||| ||  ||||||
3' UCCUU  UCACU CGAC  GUCGUACUGGAU GA  CCAAUC        A
        GC    C     UGA              C  UC      AGGAA
```

ptc-miR167b UGAAGCUGCCAGCAUGAUCUA

```
        A      AA    A   C            U  UU      AGAAGGAUAG
5' GUUC CAA--GGGAA  GGUGA GCUG CAGCAUGAUCUA CU  GGUUAG        \
   |||| |||  |||||  ||||| |||| |||||||||||| ||  ||||||         A
3' CAGG GUUUAUCCUU  UCACU CGAC GUCGUACUGGAU GA  CCAAUC        /
        G       CC    C     A            C  CU      AAGAAAGCGA
```

ptc-miR167c UGAAGCUGCCAGCAUGAUCUA

```
                                         AUUAA
        AC-  C      U                    \  /      CUU
5' UCGUG   CCA UAGCAG UGAAGCUGCCAGCAUGAUCUAA CC-UCCUU   \
   |||||   ||| |||||| |||||||||||||||| |||||||||| || |||||    U
3' AGCAC   GGU GUCGUC ACUUUGAUGGU-GUACUAGAUUAGGUAGGAA   /
        UAC  A      C                              CUA
```

ptc-miR167d UGAAGCUGCCAGCAUGAUCUA

```
        AC-  C      U                              CUUC
5' UCGUG   CCA UAGCAG UGAAGCUGCCAGCAUGAUCUAACU-UCCUUG   \
   |||||   ||| |||||| |||||||||||||| |||||||||||| ||||||    U
3' GGCAC   GGU GUUGUC ACUUUGGUGGU-GUACUAGAUUGGUAGGAAC   /
        CAC  A      C                              UAUU
```

ptc-miR167e UGAAGCUGCCAGCAUGAUCUU

```
        C-        U                    UAACCUC    CUU
5' UCGUG   ACCA-CUAUCAG UGAAGCUGCCAGCAUGAUCU      CCUC   \
   |||||   |||| || |||| |||||||||||| |||||||||         ||||    U
3' AGCAC   UGGUUGA-AGUC ACUUUGACGGU-GUACUAGA      GGAG   /
        AU        G                    CAAGAAA    CUG
```

ptc-miR167f UGAAGCUGCCAGCAUGAUCUU

```
        C-   UC  U   U                 UAGCC-     CUU
5' UCGUG   ACC  UA CAG UGAAGCUGCCAGCAUGAUCU        UUCCUC   \
   |||||   |||  || ||| |||||||||||| |||||||||         ||||||    U
3' AGCAC   UGG  GU GUC ACUUUGACGGU-GUACUAGA        AAGGAG   /
        AU   UU U   C                 CAAAGA      UUG
```

89

ptc-miR167g UGAAGCUGCCAGCAUGAUCUG

```
              C     U                        AA        UAAUUU
5' UCAUGC-ACCA UAGUAG UGAAGCUGCCAGCAUGAUCUG CUUUCCU        U
   |||||||  ||||  |||||||  |||||||||||||||||| |||||||
3' GGUACGAUGGU GUUAUC ACUUUGAUGGU-GUACUAGAU  GAAAGGG        C
              A     U                        CA        CAUAUC
```

ptc-miR167h UGAAGCUGCCAACAUGAUCUG

```
              C  U  U                        AG        UAAUUU
5' UCAUGC-ACCG UA UAG UGAAGCUGCCAACAUGAUCUG CUUUCCU        U
   |||||||  ||||  ||  |||  ||||||||||||||  |||||||  |||||||
3' GGUAUGAUGGU GU AUC ACUUUGACGGU-GUACUAGAU  GAAAGGA        C
              A  U  C                        CA        CAUAUC
```

ptc-miR167i UGAAGCUGCCAGCAUGAUCUG

```
   AA           C      U                        AA        UAAUUU
5'    UCAUGC-ACCA UAGUAG UGAAGCUGCCAGCAUGAUCUG CUUUCCU        U
      |||||||  ||||  |||||||  |||||||||||||||||| |||||||        |
3'    GGUACGAUGGU GUUAUC ACUUUGAUGGU-GUACUAGAU  GAAAGGG        C
   GG           A      U                        GA        CAUAUC
```

----

ptc-miR396a UUCCACAGCUUUCUUGAACUG

```
                                 CCU
        C    C    G    UC      C      \  /  UU     UGU          C
5' UGA CCU UU GUAU UUCCACAGCUUU UUGAACUGCAAU  GAAUAU  UGUUG-AUGUUGC  \
   |||  ||| || ||||  |||||||||||| ||||||||||  ||||||   ||||| |||||||    G
3' ACU GGA AA CAUA AGGGUGUCGAAA AACUUGGCGUUG  UUUAUG   ACAGUAUACAACG  /
        A    C    A    GA          U          UU     UU-        / \   U
                                                              UGU
```

ptc-miR396b UUCCACGGCUUUCUUGAACUG

```
      C    C       UU          C       CCUAUUA--              U   UG  CGA
5' UGA CCU UUUGUAU UUCCACAGCUUU UUGAACUGGA        GAU-UAAUGUUGAUGU GU  UG    U
   |||  ||| ||||||  |||||||||||| |||||||||||          |||  ||||||||| |||| || ||
3' ACU GGA AAACAUA AGGGUGUCGAAA AACUUGGCGU          UUAUGUUACAG-UAUA CA  AC    A
      A    C       GA          U          CGUUUUUAC              C  GU  CGU
```

ptc-miR396c UUCCACGGCUUUCUUGAACU

```
                                               GC    CUU
                                          -AAAGG  UAGG    U
                                          / |||||  ||||
                                         /  UUUCC  AUCU    C
                                        GAGU  /    UU      UUU
            C               A    GGC  C      /    C
5' CAUG UUUUCCACGGCUUUCUUGA CUU    ACU AAGAGACAUGA     \
   |||| |||||||||||||||||| |||    |||  ||||||||| |||    \         AUUU
3' GUAU AAAGGGUGCCGAAAGAACU GAA    UGA UUCUCUG-ACU     -GUUCUUUGAA      U
         A               C    ACA   U                  |         |||||||||||
                                          UAUACAUACGUAAGAAACUU      U
                                                              GUCC
```

ptc-miR396d UUCCACAGCUUUCUUGAACUU

```
      G-      A       A-    --U-  CAU  UAACUU      UG   GU U        A
5' CAUG  UUUUCCAC GCUUUCUUGA  CUUCC   UGC  GCU      GUGUGUGUG UGU G GUGUGUAG  U
   ||||   ||||||| |||||||||||  |||         |||  |||       ||||||||| ||| | |||||||||
3' GUAC  AGAGGGUG CGAAAGAACU   GAAGG   ACG  CGA      UACACAUAU AUA U UAUAUAUC  C
      CA      C       CC    UACC  AU-  UCGAUU      UG   UC C        A
```

ptc-miR396e UUCCACAGCUUUCUUGAACUU

```
                                               13nt LOOP
                                              /  \       A    A
                                              \   UGUAUAU UAG    U
                                              |   ||||||| |||
                        C                A     UU- /   ACAUGUA AUC    C
5' AAGUCCUGGUCAUG UUUUCCACAGCUUUCUUGA CUUCU    GC    \       C    A
   |||||||||||||| ||||||||||||||||||| |||||    ||     GC
3' UUUAGGACCGGUAC AGAGGGUGUCGAAAGAACU GAAGG    CG      \
        /  \   A                      C    UAC \     CU
       U   A                                C    /       AA
        UA                                   \  CUAUAUAUAUA   U
                                              \ |||||||||||
                                               -GAUAUGUAUGU  A
                                                           AU
```

ptc-miR396f UUCCACAGCUUUCUUGAACUU

```
         U      A                          UA
5' CAUGUU UUCCAC GCUUUCUUGAACUUCCUAG-AGC-C    \
   |||||| |||||| |||||||||||||||||||| ||| |    G
3' GUACGA AAGGUG CGAAAGAACUUGAAGGAUCGUCGUG    /
         U      C              /  \      GA
                         17nt LOOP
```

ptc-miR396g UUCCACGGCUUUCUUGAACUG

```
                                                            AGAC
       UG        C  C                AUAUAUUAU    GG     \ /      A
5' UGGGUC-UCU  GUGAUUUU CA GGCUUUCUUGAACUGU          CAAU  CUUUUUACAUGGAAG   U
   |||||| |||  |||||||| || ||||||||||||||||          |||| |||||| |||||||||
3' AUCUAGGAGA  UAUUAAAA GU CCGAAAGAACUUGACA          GUUA GAAGAG-GUACCUUU   G
       UU        A  C                AAAAAGACU    A-                G
```

ptc-miR396h UUCCACGGCUUUCUUGAACUU

```
       UCA                                                      A
5' AACACCUGG      UGCUAUUCCACGGCUUUCUUGAACUUCUUA-AGUUAU    U
   ||| |||||      ||||||||||||||||||||||||||||||| ||||||
3' UUG-GGACU    ACGGUAAGGUGUCGAAAGAACUUGAAGGAU UCGAUA    G
       UUG    /  \                        /  \       U
              C   A                      C   A
               AAA                       \   G   ACC
                 \                        \ C-AGC   U
                  \                         \  |||
                   ---UCG   C
                        GAU
```

ptc-miR396i UCCCACAGCUUUAUUGAACCG

```
       G   U                    A      C  ACAAAAA    AUG
5' UGAUCCU UU GUAU-CUUCCCACAGCUUU UUGAAC GCA        UACA    U
   ||||||| || |||| |||||| |||||||| |||||| |||        ||||
3' ACUGGGA AA CAUAAGAAGG-UGUCGAAA AACUUG CGU        AUGU    C
       G   C                 G     A  \    /     AUA
                                    37 nt
                                    LOOP
```

ptc-miR396j UCCCACAGCUUUAUUGAACUU

```
                              19 nt LOOP
                                \ /
         G                  A    //    AU
5' UGAUCCU UUUGUAU-CUUCCCACAGCUUU UUGAAC AUGUCAU    G
   ||||||| ||||||| ||||| |||||||| |||||| |||||||
```

91

```
3' ACUGGGA AAACAUAAGAAGG-UGUCGAAA AACUUG UACGGUA   G
              G                   G     \\     CU
                                         / \
                                  37 nt LOOP
```

ptc-miR396k UCCCACAGCUUUCUUGAGCUU

```
                                                       U
                                                UAUAU     U
                                                /|||||
                                               / AUAUA     A
                                              / /           U
              CUC                      CAUG    AUACAUACA U
5' CAUGUU  CCACAGCUUUCUUGAGCUUC   GCGCU           \AGG
   ||||||  ||||||||||||||||||||   |||||              |
3' GUACGA  GGUGUCGAAAGAACUUGAAG   GCGGA           CGA
      AAA                      AACG/ \ AAUUAGACA   U
                                  A  G    \ \   C        U
                                   A        \ GUA AUGUGUAG   A
                                            \||| ||||||||
                                            CAU UAUAUAUC    G
                                              A            U
```

----

ptc-miR398a UGUGUUCUCAGGUCACCCCUU

```
           ACA          A     CCA                G     C    UG
5' ACAGCCGGU    CCCCAGAGG GUGGCU   GAGAACACAGGGGGUU GUUUU UAGC   \
   |||| ||||    |||||||||| ||||||   ||||||||||||| ||| ||||| ||||   U
3' UGUC—CCCA    GGGGUUUCC CACUGG   CUCUUGUGUCUC CAG UAGAA AUCG   /
        -C-          C       -A-              / \  G    C    AA
                                             AUAA
```

ptc-miR398b UGUGUUCUCAGGUCGCCCCUG

```
       AA    U-    A    A      U  GU     U CG    C
5' UGUA  AGGG  UCCG CAGG GCGGCCU GG  CACAUG GG  GCAU   \
   ||||  ||||  |||| |||| |||||||| ||  |||||| ||  ||||   C
3' ACAU  UCCU  GGGC GUCC CGCUGGA UC  GUGUAC UC  UGUA   /
       C-    UC    C   C      C  UU     U A-    A
```

ptc-miR398c UGUGUUCUCAGGUCGCCCCUG

```
       UU-    A-    A      AAU       GG          U
5' GUGGAGGA   CCU  CAGG GCGACCUGA   CACAUGU  GCUGCACCC   C
   ||||||||   |||  |||| ||||||||||   |||||||  ||| |||||
3' UACCUCCU   GGG  GUCC CGCUGGACU   GUGUACA  CGA-GUGGG   C
       UUC    CC    C        CUU         A-   / \ U
                                              U  U
                                            UCUA
```

----

ptc-miR408a AUGCACUGCCUCUUCCCUGGC

```
              GAUCA
           AG  \ /A    C     A     A       CU   AA
5' GGGGAGAGGGAGCAG  ACAAG CGGGGAA AGGCAG GCAUGG UGGAGCUA  AACAG   G
   |||||||||||||| ||  |||||  ||||||| |||||| |||||| ||||||||  |||||
3' CUCUUUUCCUUC-UC  UGUUC GUCCUU UCCGUC CGUACC AUCUCGGU  UUGUU   U
              GG   G       C      A      C      U-    CA
```

92

----

ptc-miR472a UUUUCCCUACUCCACCCAUCC

```
         UU       G     C           CG         UUU   U   UG
5' GGAU    UCGGAAG UUAU GGGUGGGUG-AG  GGGAAGAUAAC   GGU UU    A
   ||||    ||||||| |||| |||||||||| ||  |||||||||||   ||| ||
3' CUUA    AGCCUUU GAUA CCUACCCACCUC  CCCUUUUAUUG   UCA GA    G
         CU       G     C           AU          -U-   U   UA
```

ptc-miR472b UUUUCCCUACUCCACCCAUCC

```
        UU       G    GU          G                       U
5' GGGU    UCGGAAG UUA  GGAUGGGUG-AGU GGGAAGAUAACUAAGC    \
   ||||    ||||||| |||  ||||||||| ||| |||||||||||| ||||    C
3' CUUA    AGCCUUU GAU  CCUACCCACCUCA CCCUUUUAUUG-UUUG    /
        CU       G    AC          U                      U
```

----

ptc-miR475a UUACAGUGCCCAUUGAUUAAG

```
                                              UAAC
                                        AGCAA       U
                                        /|||||
                                       /  UCGUU        C
          A            C          GAU   U  / /     UUCU
5' GUUAUA CAUCUUGAUCAAUGG CAUUGUAAGAGUAGAAG   CCA GA  A
   |||||| |||||||||||||||| ||||||||||| ||||   ||| ||  |
3' UAAUAU GUAGAAUUAGUUACC GUGACAUUCUCA-CUUU   GGU CU  A       GUG
        G            C                 AU-  C  C AAUGU    A
                                               \ |||||
                                             AGUUACA    U
                                                AAA
```

ptc-miR475b UUACAGUGCCCAUUGAUUAAG

```
                                   GA-       UAAC
                                 C-CAU   AGCAA    \
                                 /  |||   |||||     U
                                 U  GUA   UCGUU    /
                                  A  /   AAG     UCUC
              C                 G   UGC
5' UAACAUCUUGAUCAAUGG CAUUGUAAGAGUAGAAG       \
   |||||||||||||||||||| ||||||||||| ||||        /
3' AUUGUAGAAUUAGUUACC GUGACAUUCUUA-CUUU      CAG
              C                 A   \  AUA
                               U  CAA    G
                               U  |||
                                A-GUU    A
                                   CUC
```

ptc-miR475c UUACAAUGUCCAUUGAUUAAG

```
                             UCCAU
        A     A     C      A     \ /   AACUCUUU      G
5' AUAAC UCUUGA CAAUGG CAUUGUAAG GU-AAGUGGGAGUCAAU      UGCA    A
   ||||| |||||| |||||| ||||||||| || |||||||||||||||     ||||
3' UAUUG AGAAUU GUUACC GUAACAUUC CACUUUAUCCUCAGUUA      ACGU    A
        G     A     U      C                 CGAAACAGU      A
```

93

ptc-miR475d UUACAGAGUCCAUUGAUUAAG

```
       C         C          CCA          A          C   AAG
5' GU AUAACAUC UGAUCAAUGG     UUGUAAGAGU GA-AGGA-UC AUG      C
   || |||||||| ||||||||||     |||||||||| || |||| || |||
3' CA UAUUGUAG AUUAGUUACC   GACAUUCUCA CU UCCUCAG UAC      A
    A        A          UGA          C / \      U   AAA
                                      UAUAG
```

Figure 2.2. Predicted stem-loop structures of conserved *P. trichocarpa* ptc-miRNAs. Precursor structures are presented as determined by computational analysis using the mfold programme. Mature cloned miRNA sequences are shown in bold. MiRNA families are separated by dashed lines. The sequences represent only the shortest possible stem-loop structure, where only that portion of the stem-loop starting at the cloned sequence and ending at the complementary miRNA* sequence is shown.

(a)

```
ptc-MIR56d   ------------------G--GCTAATTACTGGA---GAAAGG----GAC---GGTG---
ptc-MIR56b   -----------------AA--ACTGAA--CCG-A---GGCAGG----GAC--CAGGGCTC
ptc-MIR56c   -------------------------AA--CAG-----GGAAGG-----AC--AAGAG---
ptc-MIR56a   GAATTCCGGGGGAAGTATGGTTGCAA--AGCTGAAACTTAAAGGAATTGACGGAAGGGCAC
                                      *  *        ***      **    * *
```

```
ptc-MIR56d   ---CAGTAGG-----CAGC--------CATGATCGA-AGGAG--GAAACAGCTTCCG----
ptc-MIR56b   --CCAAATAA-----CTCTT-CTCCAGTGGACTT---GGGAGCTGCGAAATCTTAAGTAGT
ptc-MIR56c   ---CAAACAA-----CGGCA-GTAGATCAAACTCGA-GGGAAAAGCTTCA-CAGAGAAGGA
ptc-MIR56a   CACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGAAACCTCAC------CAGG
               **          *                    *         *

ptc-MIR56d   -----------------------CTCT-ACT----ACTCTTTCTTGATTCTATGGGTGGT
ptc-MIR56b   CATAGGCATTGGATGAGAATGGG---CA-ACTGAGAGCTCTTTCTTGATTCTATGGGTGGT
ptc-MIR56c   GAGCATCTCTCTCTA-AATTCTCTCTGACTGAGAGCTCTTTCTTGATTCTATGGGTGGT
ptc-MIR56a   CCCGGACACTGGAAG-GAT----TGACAGATTGAGAGCTCTTTCTTGATTCAGTGGGTGGT
                                   * *     ************** *******

ptc-MIR56d   GGTGCATGGCCGTTCTTAGTTGGTGGAGC
ptc-MIR56b   GGTGCATGGCCGTTCTTAGTTGG------
ptc-MIR56c   GGTGCATGGCCGTTCTTAGTTGGTGGAG-
ptc-MIR56a   GGTGCATGGCCGTTCTTAGTTGGTGGAGC
             *********************
```
(b)

```
ptc-MIR156g  ----------TGTTGACAGAAGATAGAGAGCACAGATGATGAAAT-GCATGGAGCTTAAT
ptc-MIR156h  ----------TGTTGACAGAAGATAGAGAGCACTGACGATGAAAT-GCATGGAGCTTAAT
ptc-MIR156j  ----------TGTTGACAGAAGATAGAGAGCACAGATGATGATAT-GCAATGGACTC---
ptc-MIR156i  ----------TGTTGACAGAAGATAGAGAGCACAGATGATGTTTT-GCAGTAGACTC---
ptc-MIR156e  GGACATTAGAAATTGACAGAAGAGAGTGAGCACACAGAGGCATAT-TTGTATAAAT---
ptc-MIR156d  AGGGTAAGTGAGTTGACAGAAGAGAGTGAGCACACAGGGTACTTTCTTGCATGACGTTCA
                       ********** ** ****** *        *

ptc-MIR156g  TGCATCTCACT-CCTTTGTGCTCTCTAGTCTTCTGTCATCACC--------------
ptc-MIR156h  TGCATCTCACT-CCTTTGTGCTCTCTAGTCTTCTGTCATCAC---------------
ptc-MIR156j  TGCATCCCACT-CCTTTGTGCTCTCTATGCTTCTGTCATCA---------------
ptc-MIR156i  TGGATCTCACT-CCTTTGTGCTCTCTATGCTTCTGCCATCACCT------------
ptc-MIR156e  --TATACCATTGCTTTTTGCGTGCTCATTTCTCTTTCTGTCACTT-------------
ptc-MIR156d  TGCTTGAAGCT--TTGCGTGCCCACCCTCTATCTGTCACCCCATCACCATCTCTCCC
                 *      *   *   *  * *  *        *     *
```
(c)

```
ptc-MIR159b  TAGGGAGTGGAGCTCCTTGAAGTCCAATAGAGGTTCTTGCTGGGTAGATTAAGCTGCTAA
ptc-MIR159c  ----GAGTGGAGCTCCTTGAAGTCCAATAGAGGTTCTTGCTGGGTAGATTAAGCTGCTAA
ptc-MIR159a  ----GAGTGGAGCTCCTTGAAGTCCAATAGAAGCTCCTGCTGGGTAGATCGAGCTGCTGA
             *************************** * ** ************ ******* *

ptc-MIR159b  GCTATGGATCC-ACAGTCCTATC--TATCAACCGAAG-GATAGGTTTGCGGCTTGCATAT
ptc-MIR159c  GCTATGGATCC-ACAGTCCTATC--TATCAACTGAAG-GATAGGTTTGCGGCTTGCATAT
ptc-MIR159a  GCTATGAATCCCACAGCCCTATCACCATCAGTCATTTTGATGGGCCTGCGGCTTGCATAT
             ****** **** **** ****** ****     *** ** *************

ptc-MIR159b  CTCAGGAGCTTTATTGCCTAATGTTAGATCCCTTTTTGGATTGAAGGGAGCTCTAAAACCC
ptc-MIR159c  CTCAGGAGCTTTATTGCCTAATGTTAGATCCCTTTTTGGATTGAAGGGAGCTCTAAA---
ptc-MIR159a  CTCAGGAGCTTTATTACCTAATGTTAGATCTTTTTTTTGGATTGAAGGGAGCTCTAAACCT
             *************** *************        ***********************
```

(d)

```
ptc-MIR166d  ------------------------------------------------------------
ptc-MIR166e  ------------------------------------------------------------
ptc-MIR166a  ------------------------------------------------------------
ptc-MIR166o  ------------------------------------------------------------
ptc-MIR166p  ------------------------------------------------------------
```

95

```
ptc-MIR166i   ------------------------------------------------------------
ptc-MIR166j   ------------------------------------------------------------
ptc-MIR166m   ------------------------------------------------------------
ptc-MIR166k   ------------------------------------------------------------
ptc-MIR166l   ------------------------------------------------------------
ptc-MIR166g   ------------------------------------------------------------
ptc-MIR166h   ------------------------------------------------------------
ptc-MIR166c   ------------------------------------------------------------
ptc-MIR166b   ------------------------------------------------------------
ptc-MIR166f   ------------------------------------------------------------
ptc-MIR166n   GGGGTGTTTGGAATGAAGTTTGATCCAAGATCCTTGTCTCTCCCGTTAACTTAGTCTCTG
ptc-MIR166q   ------------------------------------------------------------


ptc-MIR166d   ----------------------------------------GGGGAATGT-----------
ptc-MIR166e   ----------------------------------------GGGGAATGT-----------
ptc-MIR166a   --------------ATAGAAATGGAAGCTATTTCTTTTGAGGGGAATGT-----------
ptc-MIR166o   ------------------------------GTTGAGAGGAATGT-----------
ptc-MIR166p   ------------------------GCGTAAGGTTGAGAGGAACGC-----------
ptc-MIR166i   ----------------------------GTTGAGGGGAATGT-----------
ptc-MIR166j   ---------------------------GTGTGTTGAGGGGAATGT-----------
ptc-MIR166m   ------------------------AGCTTCTTTTTTTTGAGGGGAATGT-----------
ptc-MIR166k   --------------------TCTCGCGTATTAACTGTTGAGGGGATTGT-----------
ptc-MIR166l   --------------------GCATGTATTAACTGTTGAGGGGATTGC-----------
ptc-MIR166g   -------------------------------AGTTGAGGGGAATGC-----------
ptc-MIR166h   ------------------------TGCTACACAGTTGAGGGGAATGC-----------
ptc-MIR166b   ----------------------GGAAAGTGAGGATGTGGGGGAATGT-----------
ptc-MIR166c   ----------------TAGCTAGGAAGCTTAACTTTTGTGGTGAATGT-----------
ptc-MIR166f   ---------------------GGAAGCTTGTCTTTTGAGGGGAATGT-----------
ptc-MIR166n   TGAGAGTAGGAAGTGCTGGTGATCATAGGGTTTGGTTCAAGATCCATTTGACTCTTCTCT
ptc-MIR166q   -------GTGAGCACTAGTG---ATAGAGTTTGATTCAAGATCCAT-------------
                                                           *


ptc-MIR166d   --------TGTCTGGCTCGAGGACTTT--------TTGTTCATCAATCT-AATCGAACTT
ptc-MIR166e   --------TGTCTGGCTCGAGGACTTT--------TTGTTCATCAATCT-AATCGAACTT
ptc-MIR166a   -------TGTCTGGCTCGAGGTCACTAATGGGATCTATGATTTTATCTCAATTGATTGA
ptc-MIR166o   --------TGTCTGGCTCGAGGTCATTGAGG----CCATGATTATACAGACATGGCATTA
ptc-MIR166p   --------TGTCTGGGTCGAGGTCATGGAGG----CCATGATTATACATAAATGGCATTA
ptc-MIR166i   --------TGGCTGGCTCGAAG-----------------CTTAAGCA-------AAGA
ptc-MIR166j   --------TGGCTGGCTCGAAG-----------------CTTAAGCA-------AAGA
ptc-MIR166m   --------TGTCTGGCTCGAGGACTCT----------TTCTTGATCAGTCTGATCAAGT
ptc-MIR166k   -------CGTCTGGTTCGATG-----------------TCATTCA------TGAGAA
ptc-MIR166l   -------TGTCTGGTTCGATG-----------------TCATTCA------TGTGAA
ptc-MIR166g   -------TGTCTGATTCGAGA------------------CCAT-----------TCA
ptc-MIR166h   --------TGTCTGGTTCGAGA------------------CCAT-----------TCA
ptc-MIR166b   -------CGTCTGACTCGAGA------------------CAACAG-------GTAA
ptc-MIR166c   -------TGTCTGGTTCAAGG------------------CATG---------GCCA
ptc-MIR166f   -------TGTTTGGTTCAAGG------------------CCTG---------GCCA
ptc-MIR166n   TTATATCTCTCCTGTGTCCTAGCTGGTAATCTGTAGTAGTCTTAATTAT--CCCTTACTT
ptc-MIR166q   -------CTCATGTGTGCTAGCTAG----CTTTTAAATTTTTAATCAGA-TCCCTACTT
                       * *    *


ptc-MIR166d   TCTACCTGTAGAT-CTAGTATCTT-ATTTAAGATTGATCACG---------TATT-AGGG
ptc-MIR166e   TCTACCTGTAGAT-CTAGTATCTT-ATTTAAGATTGATCACG---------TATT-AGGG
ptc-MIR166a   TTTTCTTTTAAATTCTAGTAAATTGAATTGAGAGATATCATGATCAACTTATATTTAATG
ptc-MIR166o   CCTGATGACAGCCGAGAAAATTCAAAGTCTGTGTGTATCTTGT-------ACCTCGATGA
ptc-MIR166p   TCTGATGACAGCCCAGATAATCGA-----TGCACCTGTCTTGA-------ACCTAAATGA
ptc-MIR166i   GTTTT----------CTCTCAA-------GAAACAACTGTTA-------------AG-G
ptc-MIR166j   GTTTC----------CTAACAT-------GAAACAACTGTTA-------------AG-G
ptc-MIR166m   GTTCTATCTTTAGATCTAATATCT-----TAGATCATGTGTTA------------GG-G
ptc-MIR166k   GCTCA-----------AACAT--------AAACGTAATATTG-----------AATG
```

```
ptc-MIR166l   GCTTT-----------AACAT-------TAATGTAGTATTG------------AGTG
ptc-MIR166g   CTTTAAG-----------CAC-------ACATTCATCTTTCG------------AATG
ptc-MIR166h   CCTGAAGA----------GCAC-------GCATTCATCTTTTG------------AGTG
ptc-MIR166b   ACTGAAG----------GAAT-------GGCCGGAGTTTAAG------------AGTT
ptc-MIR166c   CCACATC----------TCTT------GGTG-----AATAT-------------ATG
ptc-MIR166f   CCCCATG----------TCTT------GGAATTTAAAATAT------------CATG
ptc-MIR166n   CTTGTTAATTTTTAGGTTTGATCTTGCAAGTAATTATATTTGGA-AAATATCATAAAATG
ptc-MIR166q   CCTGTTAATTCTTAGGTTTTATCTTGCAAGTTATTATATTTAGT-AAACATCATGAACTG
```

```
ptc-MIR166d   TTGTCGGACCAGGCTTCATTCCCC------------------
ptc-MIR166e   TTGTCGGACCAGGCTTCATTCCCCCCAATCATTGCTTCCAT--
ptc-MIR166a   ATGTCGGACCAGGCTTCATTCCCCCCAATTGTTTCTTCCTTTC
ptc-MIR166o   TTCTCGGACCAGGCTTCATTCCTTCCAACA-------------
ptc-MIR166p   TTCTCGGACCAGGCTCCATTCCTTCCAACCATCATTTGCT---
ptc-MIR166i   CT-TCGGACCAGGCTTCATTCCCCTCAACCAAT----------
ptc-MIR166j   CT-TCGGACCAGGCTTCATTCCCCTCAAACATA----------
ptc-MIR166m   TCGTCGGACCAGGCTTCATTCCCCCCAATTATTGCTTCC----
ptc-MIR166k   ATTTCGGACCAGGCTTCATTCCCCCCAACAATGTTTGA-----
ptc-MIR166l   ATTTCGGACCAGGCTTCATTCCCCCCAACTATATTTGTT----
ptc-MIR166g   ATCTCGGACCAGGCTTCATTCCCCCCAACT-------------
ptc-MIR166h   ATCTCGGACCAGGCTTCATTCCCCCCAACTCAAAA--------
ptc-MIR166b   CTCTCGGACCAGGCTTCATTCCCCTCATCCACACCTTCC----
ptc-MIR166c   TCCTCGGACCAGGCTTCATTCCCCTCAATTAATGCTTCC----
ptc-MIR166f   TCCTCGGACCAGGCTTCATTCCCCTCAATTACTGCTTCC----
ptc-MIR166n   ATCTCGGACCAGGCTTCATTCCTTACACCTTGCTTTTCTTTTA
ptc-MIR166q   ATCTCGGACCAGGCTTCATTCCTTACA---------------
                 *********** ******
```

(e)

```
ptc-MIR167g   --TCATGCACCACTAGTAGTTGAAGCTGCCAGCATGATCTGAACTTTCCTTAATTTTCCT
ptc-MIR167i   AATCATGCACCACTAGTAGTTGAAGCTGCCAGCATGATCTGAACTTTCCTTAATTTTCCT
ptc-MIR167h   --TCATGCACCGCTATTAGTTGAAGCTGCCAACATGATCTGAGCTTTCCTTAATTTTCCT
ptc-MIR167c   --TCGTGCACCACTAGCAGTTGAAGCTGCCAGCATGATCTAAATTAACCTC--CTTCTTT
ptc-MIR167d   CATCGTGCACCACTAGCAGTTGAAGCTGCCAGCATGATCTAACTTC-CTTG--CTTCTTT
ptc-MIR167e   --TCGTGCACCACTATCAGTTGAAGCTGCCAGCATGATCTTAAC---CTCC--CTCCTTT
ptc-MIR167f   --TCGTGCACCTCTATCAGTTGAAGCTGCCAGCATGATCTTAGC---CTTC--CTCCTTT
ptc-MIR167a   ----------GGGAAAAGTGAAGCTGCCAGCATGATCTATCTTTGGTTAGAGAAAGA-
ptc-MIR167b   ---GTTCACAAGGGAAAAGGTGAAGCTGCCAGCATGATCTATCTTTGGTTAGAGAAGGAT
                 *   *   *********** ********
```

```
ptc-MIR167g   ATACGGGAAAGAC-TAGATCATGTGGTAGTTTCATCTATTGATGGTAGCATGGGAAACCC
ptc-MIR167i   ATACGGGAAAGAC-TAGATCATGTGGTAGTTTCATCTATTGATGGTAGCATGGG------
ptc-MIR167h   ATACAGGAAAGAC-TAGATCATGTGGCAGTTTCACCTATTGATGGTAGTATGGAA-----
ptc-MIR167c   ATCAAGGATGGATT-AGATCATGTGGTAGTTTCACCTGCTGATGGCATCACGAAA-----
ptc-MIR167d   ATCAAGGATGGATTTAGATCATGTGGTGGTTTCACCTGTTGATGGCACCACGGAAAAATC
ptc-MIR167e   GTCGAGGAAAGAA-CAGATCATGTGGCAGTTTCACCTGAAGTTGGTATCACGAGA-----
ptc-MIR167f   GTTGAGGAAAGAAACAGATCATGTGGCAGTTTCACCTGTTGTTGGTATCACGAGAAACCC
ptc-MIR167a   -------AAGGACTAACCC---TAGCTAGGTCATGCTGTGACAGCCTCACTC-------
ptc-MIR167b   AGAAGCGAAAGAACTAACCC---TAGCTAGGTCATGCTCTGACAGCCTCACTCCTTCCTA
                 *    *    *    * *     ***    *    *    *
```

(f)

```
ptc-MIR396a   -TGACCCTC----TTGGTATTCTTCCACAGCTTTCTTGAACTG----CACCTA-------
ptc-MIR396b   -TGACCCTC----TTTGTATTCTTCCACAGCTTTCTTGAACTG----CACCTA-------
ptc-MIR396i   -TGATCCTG----TTTGTATCTTCCCACAGCTTTATTGAACCG----CAACAA-------
ptc-MIR396j   -TGATCCTG----TTTGTATCTTCCCACAGCTTTATTGAACCG----CAGCAA-------
ptc-MIR396g   ATGGGTCTC----TTGGTGATTTTCCACGGCTTTCTTGAACTG----TATATA-------
```

```
ptc-MIR396d  -AAGTCCTG----GTCATGCTTTTCCACAGCTTTCTTGAACTT----CCTTGCCATGCTT
ptc-MIR396e  -AAGTCCTG----GTCATGCTTTTCCACAGCTTTCTTGAACTT----CTTTGCCTTGCTT
ptc-MIR396k  ----TCCTGAATTGCCATGTTCTCCCACAGCTTTCTTGAGCTT----CCATGGC--GCTA
ptc-MIR396f  --------------CATGTTTTTCCACAGCTTTCTTGAACTT----CCTAGAG------
ptc-MIR396h  -AACACCTG----GTCATGCTATTCCACGGCTTTCTTGAACTT----CTTAAG-------
ptc-MIR396c  -TAAGGTCG---TTGCATGCTTTTCCACGGCTTTCTTGAACTTGGCACTCAAGAGACATG
                              *    * **** ***** **** *

ptc-MIR396a  ---TTT------GA----ATATTGTTGTTGATGTTGC----CGTGC-ATGTACATAT-GA
ptc-MIR396b  ---TTA------GA----TTAATGTTGATGTTGTTGTGCGATATGCCATGACCATAT-GA
ptc-MIR396i  ---AAA------T-----ACAATGTC-ATATGTACATG-CACG----GCAACATCAACAA
ptc-MIR396j  ---AAA------TGAAT-ACAATGTC-ATATGGTCATGGCATATCGCACAACAACATCAA
ptc-MIR396g  ---TTA------TCAA--TGGCTTTTTACAAGACTGGAAGATG----GTTTCCATGG-AG
ptc-MIR396d  AACTTG------TGTGTGTGTGTGTGTGTGTGTGTAGATCACTATATATCTGTATAGTTA
ptc-MIR396e  AATCTG------TGTATATATAGATCACTACATGTACAGCTCC---TATATATATAAATA
ptc-MIR396k  TACATA------CATATATTTATATATATAGGAGCTGTACATGTAGTG-ATCTATATATA
ptc-MIR396f  --CCTA------GAGGTGCTGCTAGCTATACATATAA----------------------
ptc-MIR396h  ---TTA------TATGTATAGCTAGCAGCACCTCTAGGC--------------------
ptc-MIR396c  AGAGTAAAGGGCTAGGCTTTCTTTTCTATTCCTTTCGTTCTTTGAAATTTTCCTGTTCAA

ptc-MIR396a  CATTGTATT-----------TTTGTTG----CGGTTCAATAAAGCTGTGGGAAGATACAA
ptc-MIR396b  CATTGTATTCATT------TTTGCTG----CGGTTCAATAAAGCTGTGGGAAGATACAA
ptc-MIR396i  CAATATTCAAA---------TAGGTG----CAGTTCAAGAAAGCTGTGGAAGAATACCA
ptc-MIR396j  CATTAATCTAA---------TAGGTG----CAGTTCAAGAAAGCTGTGGAAGAATACAA
ptc-MIR396g  AAGAATTGTCA---------CAAAAA----CAGTTCAAGAAAGCCCTGAAAAATTATTT
ptc-MIR396d  TACACATTTAGCTAGCTAGCACCATGG----AAGCTCAAGAAAGCCGTGGGAGAACATGG
ptc-MIR396e  TATGTATGTA------TAGCGCCATGG----AAGCTCAAGAAAGCTGTGGGAGAACATGG
ptc-MIR396k  CACAGATTAAGC------AAGGCAAAG----AAGTTCAAGAAAGCTGTGGAAAAGCATGA
ptc-MIR396f  ------CTTA--------------AG----AAGTTCAAGAAAGCCGTGGAATAGCATG-
ptc-MIR396h  ------TCTA--------------GG----AAGTTCAAGAAAGCTGTGGAAAAACATGG
ptc-MIR396c  AGAATCCATACATATCAGTCTCTTTAGTACAAAGCTCAAGAAAGCCGTGGGAAAAATATGA
                              * **** ***** ** * *

ptc-MIR396a  A----CAGGATCA--------------------
ptc-MIR396b  A----CAGGATCA--------------------
ptc-MIR396i  A----GAGGGTCA--------------------
ptc-MIR396j  A----GAGGGTCA--------------------
ptc-MIR396g  A----GAGGATCTA-------------------
ptc-MIR396d  CAACTCAGGATTTTTTGGGGCGATCTGTTTCTTG
ptc-MIR396e  CAATTCAGGATTT--------------------
ptc-MIR396k  C----CAGGA-----------------------
ptc-MIR396f  ----------------------------------
ptc-MIR396h  CAGTTTCAGGGTT--------------------
ptc-MIR396c  A-----AGGATTTTG------------------
```

(g)

```
ptc-MIR398c  --------------TGGAGGATTCCTACAGGAGCGACCTGAAATCACATGTG-GGCTGCACC
ptc-MIR398b  ------------GTAAAAGGGTTCCGACAGGAGCGGCCTTGGGTCACATGTG-G-CGGCATC
ptc-MIR398a  AGCCGGTACACCCCAGAGGAGTGGCTCCAGAGAACACAGGGGGGTTGGTTTTCTAGCTGTAAG
                      *   *   *   ***        *        *       *    * *      * * *

ptc-MIR8a    CTCCTGGGTTATCTTGAGCAACATGTGTTCTCAGGTCGCCCCTGCCGGGCTTTCCTCC
ptc-MIR8b    CAA-----------TGTACTTCATGTGTTCTCAGGTCGCCCCTGCCGGGCTT-CCTCT
ptc-MIR8c    CTACAAGATGGAC-AAAGCACTCTGTGTTCTCAGGTCACCCCTTTGGGGCAC-CCTGT
                *            *    ************* *****    ****    ***
```

(h)

98

```
ptc-MIR475b  GTTAAAATATAATGGTCAACCAAGGTTGCGCTTCTGGCTTTCATAGTCATAACATCTTGA
ptc-MIR475d  -------------------------------------------GTCATAACATCCTGA
ptc-MIR475a  -------------------------------------------GTTATAACATCTTGA
ptc-MIR475c  --------------------------------------------ATAACATCTTGA
                                                        ******* ***


ptc-MIR475b  TCAATGGCCATTGTAAGAGTAGAAGGATCCATGAAGCAATAACTCTCTTTGCTGAAATGT
ptc-MIR475d  TCAATGGCCATTGTAAGAGTAGAAGGATCCATGAAGCAA-AAC----------------
ptc-MIR475a  TCAATGGCCATTGTAAGAGTAGAAGGATCCATGAAGCAATAACTCTCTTTGCTAAAATGT
ptc-MIR475c  ACAATGGCCATTGTAAGAGTA-AGTGGTCCATGAGTCAATAACTCTCTTTGCAGAAATGC
              ****************** *   * ******* *** ***



ptc-MIR475b  GCGACCAA--ATAGACTCTTGATATTTCATTCTTACAGTGCCCATTGATTAAGATGTTAT
ptc-MIR475d  ----------ATTGACTCCTGATATTCCACTCTTACAGAGTCCATTGATTAAGATGTTAT
ptc-MIR475a  GTGATAAAACATTGACTCCTGGTATTTCACTCTTACAGTGCCCATTGATTAAGATGGTAT
ptc-MIR475c  ATGACAAAGCATTGACTCCT---ATTTCACCCTTACAATGTCCATTGATTAAGAGGTTAT
              ** ***** *   *** ** ***** * ************* * ***


ptc-MIR475b  AACATTGTTCTCA
ptc-MIR475d  AACA---------
ptc-MIR475a  AATGT--------
ptc-MIR475c  AA-----------
              **
```

Figure 2.3. Precursor alignments of selected ptc-miRNA families. Only alignments for selected ptc-miRNA families that contain 3 or more family members are presented. Alignments were done using the ClustalW multiple alignment program. Stars indicate conservation in all sequences. Mature miRNA sequences are shown in black boxes, and miRNA* sequences are shown in bold. (a) ptc-MIR56 family members; (b) ptc-MIR156 family members; (c) ptc-MIR159 family members; (d) ptc-MIR166 family members; (e) ptc-MIR167 family members; (f) ptc-MIR396 family members; (g) ptc-MIR398 family members; (h) ptc-MIR457 family members.

2.8 TABLES

Table 2.1. Newly identified microRNAs and gene family members from *Populus* plantlets

ptc-miRNA[a]	miRNA Sequence (5'-3')[b]	Length (nt)[c]	miRNA genome location[d]	Arm[e]
	UCUUCAAUGCUGGUGUUAAA (1)	20		
ptc-miR26a	UCUUCAAUGCUGGUGUUAAA		scaffold_628:11840-11822	3'
ptc-miR26b	UCUUCAAUGCUGAUGUUGAA		LG_XIII:4013878-4013860	3'
	GCAUCCAUGGUUGAAUAGUUAA (2)	22		
ptc-miR39a	GCAUCCAUGGUUGAAUAGUUAA		LG_XVI:12106158-12106137	5'
ptc-miR39b	GCAUCCAUGGUUGAAUGGUUAA	22	LG_XVI:12079283-12079262	5'
	UCUUUCCAACGCCUCCCAUAC (2)	21		
ptc-miR44	UCUUUCCAACGCCUCCCAUAC		LG_VIII:8093051-8093071	3'
	UGGUGGUGCAUGGCCGUUCUU (1)	21		
ptc-miR56a	UGGUGGUGCAUGGCCGUUCUU		scaffold_21597:306-327	3'
ptc-miR56b	UGGUGGUGCAUGGCCGUUCUU		scaffold_6643:1755-1774	3'
ptc-miR56c	UGGUGGUGCAUGGCCGUUCUU		scaffold_520:40935-40954	3'
ptc-miR56d	UGGUGGUGCAUGGCCGUUCUU		scaffold_196:124030-124009	3'
	UCUUGCCUACUCCUCCCAUU (3)	20		
ptc-miR69a	UCUUGCCUACUCCUCCCAUU		LG_VIII:8092874-8092893	3'
ptc-miR69b	UUGCCUACUCCUCCCAUUCC		LG_VIII:8092876-8092895	3'
	UGCACAAAUUGAGAAAUGGU (2)	20		
ptc-miR81	UGCACAAAUUGAGAAAUGGU		scaffold_129:575580-575599	5'
	UUAGAUGACCAUCAACGAAA (2)	20		
ptc-miR99	UUAGAUGACCAUCAACGAAA		LG_VII:120974-120993	3'
	CUGUUUGUUUUGUUUGAUGGC (3)	21		
ptc-miR170	CUGUUUGUUUUGUUUGAUGGC		LG_IV: 6840751-6840730	5'
	UGGAUGUUAUGCAUGUAGGU (3)	20		
ptc-miR235a	UGGAUGUUAUGCAUGUAGGU		LG_XVIII:5423599-5423617	5'
ptc-miR235b	UGGAAGUUAUGCAUGAACGU		LG_XVII:2785402-2785420	5'

[a] ptc-miRNA gene name assigned in the order of cloning. In cases where the miRNA family contains multiple members, the gene name is kept the same followed by a letter to distinguish between members.

[b] Sequence of identified ptc-miRNAs given in 5' to 3' orientation. The cloned small RNA sequences are shown in bold, followed by the frequency of cloning as indicated in parentheses. Sequences in normal type were identified through computational analysis as described in the Materials and Methods section of this chapter.

[c] Length of the cloned small RNA in nucleotides (nt). Only the longest cloned sequence of each family is shown.

[d] Genome locations are presented as "genome ID (http://genome.jgi-psf.org/Poptc1/ Poptc1.home.html). Ptc-miR start site-stop site with the ID "for *Populus trichocarpa*.

[e] Location in the predicted hairpin structure as 5' or 3' depending on stem-loop arm position as shown in Figure 2.1.

Table 2.2. *Populus* microRNAs and gene family members from known families

ptc-miRNA[a]	miRNA Sequence (5' - 3')[b]	Length (nt)[c]	miRNA genome location[d]	Arm[e]
	UUGACAGAAGAUAGAGAGCAC (5)	21		
ptc-miR156d	UUGACAGAAGAGAGUGAGCAC		LG_XVIII:1365515-1365535	5'
ptc-miR156e	UUGACAGAAGAGAGUGAGCAC		LG_XVIII:6181085-6181105	5'
ptc-miR156g	UUGACAGAAGAUAGAGAGCAC		LG_II:21820427-21820407	5'
ptc-miR156h	UUGACAGAAGAUAGAGAGCAC		LG_IV:6006028-6006048	5'
ptc-miR156i	UUGACAGAAGAUAGAGAGCAC		LG_XII:4450913-4450893	5'
ptc-miR156j	UUGACAGAAGAUAGAGAGCAC		LG_XV:3242397-3242417	5'
	UUUGGAUUGAAGGGAGCUCUA (5)	20		
ptc-miR159a	UUUGGAUUGAAGGGAGCUCUA		LG_XII:4201697-4201678	3'
ptc-miR159b	UUUGGAUUGAAGGGAGCUCUA		LG_XV:3478615-3478634	3'
ptc-miR159c	UUUGGAUUGAAGGGAGCUCUA		scaffold_1140:8367-8386	3'
	UCGGACCAGGCUUCAUUCCCC (2)	21		
ptc-miR166a	UCGGACCAGGCUUCAUUCCCC		LG_I:6097079-6097060	3'
ptc-miR166b	UCGGACCAGGCUUCAUUCCCC		scaffold_122:445122-445141	3'
ptc-miR166c	UCGGACCAGGCUUCAUUCCCC		LG_XV:460396-460377	3'
ptc-miR166d	UCGGACCAGGCUUCAUUCCCC		LG_XIV:3909159-3909178	3'
ptc-miR166e	UCGGACCAGGCUUCAUUCCCC		LG_XIV:3895190-3895209	3'
ptc-miR166f	UCGGACCAGGCUUCAUUCCCC		LG_XII:2546105-2546086	3'
ptc-miR166g	UCGGACCAGGCUUCAUUCCCC		LG_X:8492342-8492361	3'
ptc-miR166h	UCGGACCAGGCUUCAUUCCCC		LG_VIII:11091614-11091595	3'
ptc-miR166i	UCGGACCAGGCUUCAUUCCCC		LG_V:824488-824469	3'
ptc-miR166j	UCGGACCAGGCUUCAUUCCCC		LG_VII:3395263-3395244	3'
ptc-miR166k	UCGGACCAGGCUUCAUUCCCC		LG_V:15276210-15276191	3'
ptc-miR166l	UCGGACCAGGCUUCAUUCCCC		LG_II:2737702-2737721	3'
ptc-miR166m	UCGGACCAGGCUUCAUUCCCC		LG_II:13668251-13668270	3'
ptc-miR166n	UCGGACCAGGCUUCAUUCCUU		LG_XIX:5391587-5391568	3'
ptc-miR166o	UCGGACCAGGCUUCAUUCCUU		LG_VII:3395015-3394996	3'
ptc-miR166p	UCGGACCAGGCUCCAUUCCUU		LG_V:824242-824223	3'
ptc-miR166q	UCGGACCAGGCUUCAUUCCUU		LG_XIII:4057479-4057498	3'
	UGAAGCUGCCAGCAUGAUCUA (1)	21		
ptc-miR167a	UGAAGCUGCCAGCAUGAUCUA		LG_II:3057594-3057574	5'
ptc-miR167b	UGAAGCUGCCAGCAUGAUCUA		scaffold_14274:285-265	5'
ptc-miR167c	UGAAGCUGCCAGCAUGAUCUA		LG_V:14949534-14949554	5'
ptc-miR167d	UGAAGCUGCCAGCAUGAUCUA		LG_II:3054996-3054976	5'
ptc-miR167e	UGAAGCUGCCAGCAUGAUCUG		scaffold_70:484760-484740	5'
ptc-miR167f	UGAAGCUGCCAGCAUGAUCUU		LG_X:9042289-9042269	5'
ptc-miR167g	UGAAGCUGCCAGCAUGAUCUU		LG_VIII:10761725-10761745	5'
ptc-miR167h	UGAAGCUGCCAACAUGAUCUG		LG_XIII:2740303-2740323	5'
ptc-miR167i	UGAAGCUGCCAGCAUGAUCUG		scaffold_70:484760-484740	5'
	UUCCACAGCUUUCUUGAACUG (4)	21		
ptc-miR396a	UUCCACAGCUUUCUUGAACUG		LG_VI:4075653-4075673	5'
ptc-miR396b	UUCCACGGCUUUCUUGAACUG		LG_XVIII:12249333-12249353	5'
ptc-miR396c	UUCCACGGCUUUCUUGAACUU		LG_VII:6490267-6490247	5'
ptc-miR396d	UUCCACAGCUUUCUUGAACUU		LG_VI:4082543-4082523	5'
ptc-miR396e	UUCCACAGCUUUCUUGAACUU		LG_XVIII:12256356-12256336	5'
ptc-miR396f	UUCCACAGCUUUCUUGAACUU		LG_III:14406867-14406887	5'
ptc-miR396g	UUCCACGGCUUUCUUGAACUG		LG_VI:17347052-17347032	5'
ptc-miR396h	UUCCACGGCUUUCUUGAACUU		LG_III:14406951-14406931	5'

ptc-miR396i	UCCCACAGCUUUAUUGAACCG		LG_VI:4075758-4075737	5'
ptc-miR396j	UCCCACAGCUUUAUUGAACCG		LG_XVIII:12249447-12249427	5'
ptc-miR396k	UCCCACAGCUUUCUUGAGCUU		LG_XVIII:12256230-12256250	5'
	UGUGUUCUCAGGUCACCCCUG (4)	21		
ptc-miR398a	UGUGUUCUCAGGUCACCCCUU		LG_X:14819174-14819194	3'
ptc-miR398b	UGUGUUCUCAGGUCGCCCCUG		LG_I:19895303-19895283	3'
ptc-miR398c	UGUGUUCUCAGGUCGCCCCUG		LG_IX:6181890-6181910	3'
	AUGCACUGCCUCUUCCCUGGC (2)	21		
ptc-miR408	AUGCACUGCCUCUUCCCUGGC		LG_II:16909463-16909443	3'
	UUUUCCCUACUCCACCCAUCC (5)	22		
ptc-miR472a	UUUUCCCUACUCCACCCAUCC		scaffold_131:138364-138385	3'
	UUUUCCCAACUCCACCCAUCC (3)	22		
ptc-miR472b	UUUUCCCAACUCCACCCAUCC		LG_XIII:203728-203749	3'
	UUACAGUGCCCAUUGAUUAAG (2)	21		
ptc-miR475a	UUACAGUGCCCAUUGAUUAAG		LG_II:20949518-20949538	3'
ptc-miR475b	UUACAGUGCCCAUUGAUUAAG		LG_VIII:13883578-13883558	3'
ptc-miR475c	UUACAAUGUCCAUUGAUUAAG		LG_X:12045864-12045884	3'
ptc-miR475d	UUACAGAGUCCAUUGAUUAAG		LG_VIII:8026896-8026876	3'

[a] ptc-miRNA gene name as assigned by sequence similarity to known miRNAs identified in previous studies. In cases where the miRNA family contains multiple members, the gene name is kept the same followed by a letter to distinguish between members.

[b] Sequence of identified ptc-miRNAs given in 5' to 3' orientation. The cloned small RNA sequences are shown in bold, followed by the frequency of cloning as indicated in parentheses. Sequences in normal type were identified through computational analysis as described in the Materials and Methods section of this chapter.

[c] Length of the cloned small RNA in nucleotides (nt). Only the longest cloned sequence of each family is shown.

[d] Genome locations are presented as "genome ID (http://genome.jgi.psf.org/Poptc1/ Poptc1.home.html). Ptc-miR start site-stop site with the ID "for *Populus trichocarpa*.

[e] Location in the predicted hairpin structure as 5' or 3' depending on stem-loop arm position as shown in Figure 2.2.

Table 2.3. Putative targets and predicted gene functions for *Populus* microRNAs.

Common function[a]	ptc-miR family[b]	Protein annotation[c]	*Populus* target genes[d]
Transcription regulation			
Transcription factor	ptc-miR69*	Zn-finger (B-box) protein	fgenesh1_pg.C_LG_I002367(2), fgenesh1_pg.C_LG_I002595(2.5)
	ptc-miR159	MYB family	eugene3.00400064(2), eugene3.00011133(2), estExt_fgenesh1_pg_v1.C_LG_IX1359(2), grail3.0100003501(2.5)
	ptc-miR166	HD-ZIP family	fgenesh1_pm.C_LG_I000485(1.5), estExt_fgenesh1_pm_v1.C_LG_III0591(1.5), estExt_Genewise1_v1.C_660759(1.5), fgenesh1_pg.C_LG_IX001407(2), fgenesh1_pm.C_LG_XVIII000075(2), estExt_fgenesh1_pm_v1.C_LG_VI0133(2), eugene3.63260001(2.5), estExt_Genewise1_v1.C_LG
	ptc-miR396	Putative growth regulating factor	eugene3.00190480(2), eugene3.00190478(2), eugene3.00021070(2), eugene3.00010688(2), grail3.0126003701(2), eugene3.00280010(2), eugene3.01450005(2), eugene3.00150073(2), fgenesh1_pg.C_LG_XIV000039(2), fgenesh1_pg.C_scaffold_28000290(2), fgenesh1_pg.C_scaf
DNA binding	ptc-miR398	SNF2-domain containing protein	fgenesh1_pg.C_LG_XVI000242(3)
	ptc-miR69*	LIM-domain containing protein	estExt_fgenesh1_pg_v1.C_LG_X1752(3)[1]
RNA processing and modification	ptc-miR26*	RRM motif-containing protein	estExt_fgenesh1_pg_v1.C_LG_II0395(2.5), eugene3.00012603(2.5)
	ptc-miR69*	D111/G-patch domain-containing protein	eugene3.00181030(2.5), eugene3.133550001(3), fgenesh1_pg.C_LG_VII000496(3), fgenesh1_pm.C_LG_I000995(3),
	ptc-miR235*	Polypyrimidine tract-binding protein	estExt_fgenesh1_pg_v1.C_LG_XV0200(3)
	ptc-miR396	SWAP-domain containing protein	eugene3.00120492(2)
Floral development	ptc-miR156	Squamosa-promotor binding protein (SBP)	estExt_Genewise1_v1.C_1240187(1)[1], eugene3.01640028(1), eugene3.00160416(1), grail3.0010026801(1.5), fgenesh1_pg.C_LG_X001417(1.5), eugene3.00400033(2), grail3.0033026601(2), estExt_fgenesh1_pg_v1.C_LG_XII0333(2), estExt_Genewise1_v1.C_LG_XV2186(2), es
	ptc-miR396	MADS-box transcription factor	estExt_Genewise1_v1.C_LG_VII4000(2.5)

Meristem determinacy	ptc-miR81*	B-box transcription factor	estExt_fgenesh1_kg_v1.C_LG_XIII0078(3)[2]
Auxin signalling	ptc-miR167	Transcription factor B3 family	fgenesh1_pg.C_LG_IV000787(2), fgenesh1_pm.C_scaffold_44000055(2), eugene3.00051147(3), grail3.1006000101(3), fgenesh1_pm.C_LG_I000947(3), estExt_Genewise1_v1.C_LG_II0899(3), estExt_fgenesh1_pg_v1.C_LG_XI0597(3)

Metabolism and cellular processes

Fatty acid metabolism	ptc-miR39*	Acyl-CoA oxidase 1	fgenesh1_pm.C_scaffold_29000034(2.5)
Carbohydrate metabolism	ptc-miR170*	Glycosyl hydrolase protein family	fgenesh1_pg.C_LG_IX000743(1.5)
Amino acid metabolism	ptc-miR159	Asparagine synthase (1-4)-Beta-mannan endohydrolase	fgenesh1_pg.C_LG_X001429(2.5), eugene3.00161335(2.5)
Metabolism	ptc-miR39*	Short-chain dehydrogenase	fgenesh1_pg.C_scaffold_88000086(2.5)
Electron transport	ptc-miR39*	Thioredoxin family	eugene3.00080305(2.5)
	ptc-miR99*	Unknown	eugene3.00091595(3)
	ptc-miR398	Cytochrome C oxidase	eugene3.00011055(2)
	ptc-miR408	Plastocyanin-like	estExt_fgenesh1_pg_v1.C_400101(1)[1], eugene3.00002467(1.5), estExt_Genewise1_v1.C_LG_I9886(2)
Cell cycle control	ptc-miR396	Microtubule severing	eugene3.00110287(2.5)
Cellular homeostasis	ptc-miR56*	Ankyrin-domain containing protein	grail3.0456000301(1)
Organelle biogenesis	ptc-miR235*	PPR	eugene3.00660195(3)
	ptc-miR396	PPR	eugene3.00030883(2.5), fgenesh1_pg.C_LG_XIV000577(2.5)
	ptc-miR475	PPR	eugene3.00131192(0.5), eugene3.00131188(1), eugene3.00131190(1), eugene3.00160204(1), eugene3.03980005(1), fgenesh1_pm.C.LG_XVIII000087(1), eugene3.00061748(1.5), eugene3.00062011(1.5), eugene3.00190210(1.5), eugene3.15500003(1.5), fgenesh1_pg.C_LG_IV0008

Protein regulation and signalling

Protein transport	ptc-miR26*	Ras small GTPase	estExt_fgenesh1_pm_v1.C_LG_XIII0079(2.5)
	ptc-miR44*	ABC (ATP-binding cassette) transporter	eugene3.17000001(2), fgenesh1_pm.C_LG_IV000044(2), estExt_fgenesh1_pg_v1.C_LG_XI0590(3)
	ptc-miR156	High-affinity nitrate transporter	grail3.0015018701(2.5)
	ptc-miR170	Sexual differentiation protein ISP4	eugene3.01370001(2.5), eugene3.00160002(2.5)

Ubiquitination pathway	ptc-miR44*	Predicted E3 ubiquitin	fgenesh1_pg.C_scaffold_82000021(3)
	ptc-miR472	Cyclin-like F-box protein	eugene3.00140082 (2.5)
Molecular chaperone	ptc-miR396	Heat shock protein DNAJ	estExt_fgenesh1_pm_v1.C_LG_I0511(2.5)[2]
Protein phoshorylation	ptc-miR69*	Putative leucine rich repeat protein kinase	fgenesh1_pg.C_LG_I000446(3)
Protein dephosphorylation	ptc-miR81*	Protein phosphatase 2C	fgenesh1_pg.C_LG_V001686(3)
Signal transduction	ptc-miR396	Ceramidases	eugene3.00161045(3)
Endocytosis	ptc-miR26*	Dynamin family protein	fgenesh1_pm.C_LG_I000286(2.5)
Cell death	ptc-miR398	Copper superoxide dismutase (SOD1)	eugene3.00130294(2), fgenesh1 pg.C LG IX001505(2.5)
Environmental response			
Integral membrane protein	ptc-miR167	Hly-III related proteins	estExt_Genewise1_v1.C_LG_XVIII3084(2.5), grail3.0030022401(2.5)
Defense response	ptc-miR69*	Putative disease resistance protein	eugene3.00102261(1.5),eugene3.00102261(1.5), eugene3.00190077(2), eugene3.00190012(2.5), eugene3.00190139(2), grail3.0140004801(1.5), fgenesh1_pg.C_scaffold_546000002(1.5), fgenesh1_pg.C_scaffold_7992000001(1.5), fgenesh1_pg.C_LG_XIX000014(2),
	ptc-miR472	Putative disease resistance protein	eugene3.06450005(1.5), eugene3.00110125(1.5), fgenesh1_pg.C_LG_XI000133(1.5), eugene3.03110003(2), eugene3.00700173(2), eugene3.08360002(2), eugene3.113680001(2.5), eugene3.00110135(2.5),eugene3.00030137(2.5), eugene3.32720001(2.5), eugene3.00770103(2.
Unknown			
	ptc-miR26*	Unknown	fgenesh1_pg.C_LG_X000935(2.5), fgenesh1_pg.C_LG_IX001362(2.5), grail3.0109002101(2.5)
	ptc-miR39*	Choline transporter-like protein	eugene3.00081101(2)
	ptc-miR39*	Unknown	eugene3.117700001(0)[2], eugene3.173450001(0)[2], eugene3.116780001(0)[2]
	ptc-miR44*	Unknown	fgenesh1_pg.C_LG_V000523(2), fgenesh1_pm.C_LG_VI000309(3)
	ptc-miR56*	Unknown	grail3.13238000201(0), grail3.9008000101(0), grail3.21597000101(0), grail3.17468000201(0), grail3.1214000201(0)

ptc-miR69*	Retrotransposon	fgenesh1_pg.C_scaffold_817000001(3)
ptc-miR69*	Unknown	eugene3.00102261(1.5)
ptc-miR235*	DUF26 domain containing protein	eugene3.06280002(2)[1], fgenesh1_pm.C_LG_VIII000433(2.5)[1]
ptc-miR156	Unknown	grail3.0008008901(1), fgenesh1_pg.C_scaffold_9189000001(2), grail3.0047015801(2.5), eugene3.04930001(2.5), fgenesh1_pg.C_scaffold_464000004(2.5),
ptc-miR159	Unknown	grail3.0043007201(0), grail3.0016033301(1.5), eugene3.01340043(1.5), eugene3.00111000(1.5), grail3.0123001501(2.5),
ptc-miR167	Unknown	eugene3.02230021(0), fgenesh1_pm.C_LG_XII000036(2), fgenesh1_pm.C_LG_VI000047(2), estExt_Genewise1_v1.C_LG_XIV2145(2.5), eugene3.00660215(2.5), estExt_fgenesh1_pg_v1.C_LG_VIII1319(3), estExt_fgenesh1_pg_v1.C_LG_VI1517(3)
ptc-miR408	Early-responsive to dehydration-related	estExt_fgenesh1_pm_v1.C_LG_XI0014(3), estExt_fgenesh1_pm_v1.C_1660007(3), estExt_Genewise1_v1.C_14050002 (3)
ptc-miR408	Unknown	fgenesh1_pg.C_LG_II001930(0)[2]

[a] Putative targets are grouped according to a common functionality. Functions for target gene families were predicted using a protein domain search (http://www.ebi.ac.uk/interpro).

[b] Name of the ptc-miRNA family. The targets listed here apply to an entire small RNA family and not to individual family members. * Newly identified ptc-miRNAs.

[c] Target loci are given as putative gene models as shown from http://genome.jgi-psf.org/Poptr1/Poptr1.home.html).

[d] For each putative target locus, the number of mismatches between the miRNA and the mRNA is indicated in parentheses. Only putative targets with scores/mismatches of 3 or less are shown here. For targets of novel ptr-miRNAs, target sites are located in predicted open reading frames (ORF) unless specifically noted. [1] Target site located in 3' untranslated region. [2] Target site located in 5' untranslated region.

2.9 REFERENCES

Achard, P., Herr, A., Baulcombe, D.C., and Harberd, N.P. (2004). Modulation of floral development by a gibberellin-regulated microRNA. Development **131,** 3357-3365.

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.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25,** 3389-3402.

Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M., Matzke, M., Ruvkun, G., and Tuschl, T. (2003). A uniform system for microRNA annotation. RNA **9,** 277-279.

Aravind, L., and Koonin, E.V. (1999). G-patch: a new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins. Trends Biochem. Sci. **24,** 342-344.

Aukerman, M.J., and Sakai, H. (2003). Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. Plant Cell **15,** 2730-2741.

Bai, C., Sen, P., Hofmann, K., Ma, L., Goebl, M., Harper, J.W., and Elledge, S.J. (1996). SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. Cell **86,** 263-274.

Bandiziulis, R.J., Swanson, M.S., and Dreyfuss, G. (1989). RNA-binding proteins as developmental regulators. Genes Dev. **3,** 431-437.

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

Beguin, P. (1990). Molecular biology of cellulose degradation. Annu. Rev. Microbiol. **44,** 219-248.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature **409,** 363-366.

Birney, E., Kumar, S., and Krainer, A.R. (1993). Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. Nucleic Acids Res. **21,** 5803-5816.

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. USA **101,** 115511-111516.

Bork, P. (1993). Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally. Proteins **17,** 363-374.

Calderon-Villalobos, L.I., Kuhnle, C., Dohmann, E.M., Li, H., Bevan, M., and Schwechheimer, C. (2005). The evolutionarily conserved TOUGH protein is required for proper development of *Arabidopsis thaliana*. Plant Cell **17,** 2473-2485.

Capaldi, R.A., Malatesta, F., and Darley-Usmar, V.M. (1983). Structure of cytochrome c oxidase. Biochim. Biophys. Acta. **726,** 135-148.

Carles, C.C., Choffnes-Inada, D., Reville, K., Lertpiriyapong, K., and Fletcher, J.C. (2005). ULTRAPETALA1 encodes a SAND domain putative transcriptional regulator that controls shoot and floral meristem activity in *Arabidopsis*. Development **132,** 897-911.

Chang, S., Puryear, J., and Cairney, J. (1993). A simple and efficient method for isolating RNA from pine trees. Plant Mol. Biol. Rep. **11,** 113-116.

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

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., Desmond, G., and Thompson, J.D. (2003). Multiple sequence alignment with Clustal series of programs. Nucleic Acids Res. **31,** 3497-3500.

Chiou, T.-J., Aung, K., Lin, S.-I., Wu, C.-C., Chiang, S.-F., and Su, C.-l. (2005). Regulation of phosphate homeostasis by microRNA in *Arabidopsis*. Plant Cell **18,** 412-421.

Coen, E.S., and Meyerwitz, E.M. (1991). The war of the whorls: Genetic interactions controlling flower development. Nature **353,** 31-37.

Craig, K.L., and Tyers, M. (1999). The F-box: a new motif for ubiquitin dependent proteolysis in cell cycle regulation and signal transduction. Prog. Biophys. Mol. Biol. **72,** 299-328.

Davies, G., and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. Structure **3,** 853-859.

Dezulian, T., Remmert, M., Palatnik, J.F., Weigel, D., and Huson, D.H. (2006). Identification of plant microRNA homologs. Bioinformatics **22,** 359-360.

Doench, J.G., Petersen, C.P., and Sharp, P.A. (2003). siRNAs can function as miRNAs. Genes Dev. **17,** 438-442.

Dresios, J., Aschrafi, A., Owens, G.C., Vanderklish, P.W., Edelman, G.M., and Mauro, V.P. (2005). Cold stress-induced protein Rbm3 binds 60S ribosomal subunits, alters microRNA levels, and enhances global protein synthesis. Proc. Natl. Acad. Sci. USA **102,** 1865-1870.

Dsouza, M., Larsen, N., and Overbeek, R. (1997). Searching for patterns in genomic data. Trends Genet. **13,** 497-498.

Eckardt, N.A. (2005). MicroRNAs regulate auxin homeostasis and plant development. Plant Cell **17,** 1335-1338.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev. **15,** 188-200.

Fletcher, J.C. (2001). The ULTRAPETALA gene controls shoot and floral meristem size in *Arabidopsis*. Development **128,** 1323-1333.

Griffiths-Jones, S. (2004). The microRNA Registry. Nucleic Acids Res. **32,** D109-111.

Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., and Enright, A.J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. **34,** D140-144.

Guo, H.-S., Xie, Q., Fei, J.-F., and Chua, N.-H. (2005). MicroRNA164 directs NAC1 mRNA cleavage to downregulate auxin signals for lateral root development. Plant Cell **17,** 1376-1386.

Hancock, J.F. (2003). Ras proteins: different signals from different locations. Nat. Rev. Mol. Cell Biol. **4,** 373-385.

Hartmann, U., Hohmann, S., Nettesheim, K., Wisman, E., Saedler, H., and Huijser, P. (2000). Molecular cloning of SVP: a negative regulator of the floral transition in *Arabidopsis*. Plant J. **21,** 351-360.

Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science **293,** 834-838.

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.

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

Kim, J., Jung, J.H., Reyes, J.L., Kim, Y.S., Kim, S.Y., Chung, K.S., Kim, J.A., Lee, M., Lee, Y., Narry Kim, V., Chua, N.H., and Park, C.M. (2005). microRNA-directed cleavage of ATHB15 mRNA regulates vascular development in *Arabidopsis* inflorescence stems. Plant J **42,** 84-94.

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

Ko, J., Prassinos, C., and Han, K. (2006). Developmental and seasonal expression of *PtaHB1*, a *Populus* gene encoding a class III HD-Zip protein, is closely associated with secondary growth and inversely correlated with the level of microRNA (miR166). New Phytol. **169,** 469-478.

Koh, S., Wiles, A.M., Sharp, J.S., Naider, F.R., Becker, J.M., and Stacey, G. (2002). An oligopeptide transporter gene family in *Arabidopsis*. Plant Physiol. **128,** 21-29.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science **294,** 853-858.

Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., and Tuschl, T. (2002). Identification of tissue-specific microRNAs from mouse. Curr. Biol. **12,** 735-739.

Lai, E.C. (2003). microRNAs: runts of the genome assert themselves. Curr. Biol. **13,** R925-936.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science **294,** 858-862.

Lee, R.C., and Ambros, V. (2001). An extensive class of small RNAs in *Caenorhabditis elegans*. Science **294,** 862-864.

Lee, Y., Jeon, K., Lee, J.-T., Kim, S., and Kim, V.N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. EMBO J. **21,** 4663-4670.

Li, D., and Roberts, R. (2001). WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. Cell Mol. Life Sci. **58,** 2085-2097.

Li, Y., Li, W., and Jin, Y. (2005). Computational identification of novel family members of microRNA genes in *Arabidopsis thaliana* and *Oryza sativa.* Acta. Biochim. Biophys. Sin. (Shanghai) **37,** 75-87.

Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., and Bartel, D.P. (2003). The microRNAs of *Caenorhabditis elegans*. Genes Dev. **17,** 991-1008.

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.

Lu, S., Sun, Y.-H., Shi, R., Clark, C., Li, L., and Chiang, V.L. (2005). Novel and mechanical stress-responsive microRNAs in *Populus trichocarpa* that are absent from *Arabidopsis*. Plant Cell **17,** 2186-2203.

Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M.L., Mireau, H., Peeters, N., Renou, J.P., Szurek, B., Taconnat, L., and Small, I. (2004). Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell **16,** 2089-2103.

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

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

Mansfield, J., Harfe, B., Nissen, R., Obenauer, J., Srineel, J., Chaudhuri, A., Farzan-Kashani, R., Zuker, M., Pasquinelli, A., Ruvkun, G., Sharp, P., Tabin, C., and McManus, M. (2004). MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression. Nat. Genet. **36,** 1079-1083.

Mathews, D.H., Sabina, J., Zuker, M., and Turner, D.H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol. Biol. **288,** 911-940.

McConnell, J.R., Emery, J., Eshed, Y., Bao, N., Bowman, J., and Barton, M.K. (2001). Role of *PHABULOSA* and *PHAVOLUTA* in determining radial patterning in shoots. Nature **411,** 709-713.

Michelsen, J.W., Schmeichel, K.L., Beckerle, M.C., and Winge, D.R. (1993). The LIM motif defines a specific zinc-binding protein domain. Proc. Natl. Acad. Sci. USA **90,** 4404-4408.

Nakahara, K., and Carthew, R.W. (2004). Expanding roles for miRNAs and siRNAs in cell regulation. Curr. Opin. Cell Biol. **16,** 127-133.

Nakamura, T., Schuster, G., Sugiura, M., and Sugita, M. (2004). Chloroplast RNA-binding and pentatricopeptide repeat proteins. Biochem. Soc. Trans. **32,** 571-574.

Nersissian, A.M., Immoos, C., Hill, M.G., Hart, P.J., Williams, G., Herrmann, R.G., and Valentine, J.S. (1998). Uclacyanins, stellacyanins, and plantacyanins are distinct subfamillies of phytocyanins: Plant-specific mononuclear blue copper proteins. Protein Sci. **7,** 1915-1929.

Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C., and Weigel, D. (2003). Control of leaf morphogenesis by microRNAs. Nature **425,** 257-263.

Pao, S.S., Paulsen, I.T., and Saier, M.H., Jr. (1998). Major facilitator superfamily. Microbiol. Mol. Biol. Rev. **62,** 1-34.

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.

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

Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., and Bartel, D.P. (2002). Prediction of plant microRNA targets. Cell **110,** 513-520.

Riechmann, J.L., and Meyerwitz, E.M. (1997). MADS domain proteins in plant development. Biol. Chem.**, 378,** 1079-1101.

Saier, M.H., Jr., Beatty, J.T., Goffeau, A., Harley, K.T., Heijne, W.H., Huang, S.C., Jack, D.L., Jahn, P.S., Lew, K., Liu, J., Pao, S.S., Paulsen, I.T., Tseng, T.T., and Virk, P.S. (1999). The major facilitator superfamily. J. Mol. Microbiol. Biotechnol. **1,** 257-279.

Sever, S., Muhlberg, A.B., and Schmid, S.L. (1999). Impairment of dynamin's GAP domain stimulates receptor-mediated endocytosis. Nature **398,** 481-486.

Skowyra, D., Craig, K.L., Tyers, M., Elledge, S.J., and Harper, J.W. (1997). F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. Cell **91,** 209-219.

Smith, T.F., Gaitatzes, C., Saxena, K., and Neer, E.J. (1999). The WD repeat: a common architecture for diverse functions. Trends Biochem. Sci. **24,** 181-185.

Stacey, G., Koh, S., Granger, C., and Becker, J.M. (2002a). Peptide transport in plants. Trends Plant. Sci. **7,** 257-263.

Stacey, M.G., Koh, S., Becker, J., and Stacey, G. (2002b). AtOPT3, a member of the oligopeptide transporter family, is essential for embryo development in *Arabidopsis*. Plant Cell **14,** 2799-2811.

Stacey, M.G., Osawa, H., Patel, A., Gassmann, W., and Stacey, G. (2006). Expression analyses of *Arabidopsis* oligopeptide transporters during seed germination, vegetative growth and reproduction. Planta **223,** 291-305.

Stone, S.L., Hauksdottir, H., Troy, A., Herschleb, J., Kraft, E., and Callis, J. (2005). Functional analysis of the RING-type ubiquitin ligase family of *Arabidopsis*. Plant Physiol. **137,** 13-30.

Sunkar, R., and Zhu, J.K. (2004). Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. Plant Cell **16,** 2001-2019.

Sunkar, R., Girke, T., Jain, P.K., and Zhu, J.K. (2005). Cloning and characterization of microRNAs from rice. Plant Cell **17,** 1397-1411.

Tanaka, R., Hirashima, M., Satoh, S., and Tanaka, A. (2003). The *Arabidopsis*-accelerated cell death gene *ACD1* is involved in oxygenation of pheophorbide a:inhibition of the pheophorbide a oxygenase activity does not lead to the "stay-green" phenotype in *Arabidopsis*. Plant Cell Physiol. **44,** 1266-1274.

Tang, G. (2005). siRNA and miRNA: an insight into RISCs. Trends Biochem. Sci. **30,** 106-114.

Theissen, G., Becker, A., Di Rosa, A., Kanno, J., Kim, J.T., Munster, T., Winter, K.U., and Saedler, H. (2000). A short history of MADS box genes in plants. Plant Mol. Biol. **42,** 115-149.

Wang, J.-F., Zhou, H., Chen, Y.-Q., Luo, Q.-J., and Qu, L.-H. (2004). Identification of 20 microRNAs from *Oryza sativa*. Nucleic Acids Res. **32,** 1688-1695.

Weigel, D., and Meyerwitz, E.M. (1994). The ABCs of floral homeotic genes. Cell **78,** 203-209.

Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S.A., and Carrington, J.C. (2005). Expression of *Arabidopsis MIRNA* genes. Plant Physiol. **138,** 2145-2154.

Zeng, Y., and Cullen, B.R. (2003). Sequence requirements for microRNA processing and function in human cells. RNA **9,** 112-123.

Zeng, Y., Yi, R., and Cullen, B.R. (2005). Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. EMBO J. **24,** 138-148.

Zhong, R., and Ye, Z.H. (2004). *Amphivasal vascular bundle 1*, a gain-of-function mutation of the IFL1/REV gene, is associated with alterations in the polarity of leaves, stems and carpels. Plant Cell Physiol. **45,** 369-385.

Zhu-Shimoni, J.X., and Galili, G. (1998). Expression of an *Arabidopsis* aspartate Kinase/Homoserine dehydrogenase gene is metabolically regulated by photosynthesis-related signals but not by nitrogenous compounds. Plant Physiol. **116,** 1023-1028.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. **31,** 3406-3415.

2.10 SUPPLEMENTARY MATERIAL

Table S 2.1. Putative endogenous siRNAs cloned from *Populus*.

ID[a]	Sequence (5'-3')[b]	Length (nt)[c]	Cloning frequency[d]	Putative targets[e]	Predicted target function[f]
1	CCCCAAGAUGAGUGCUCUCC	20	1	eugene3.27810001(0)	Unknown
9	CUGCUAUUAAGGUGCUCAAG	20	2	fgenesh1_pm.C_scaffold_152000020(1)	Leucine rich repeat protein
13	UUUCUGGUGCCAUUAUUCCUA	21	2	eugene3.00050623(2)	Unknown
15	GGCUAAGCGAUCUGCCGAAG	20	1	eugene3.125570001(0)	Hypothetical protein
16	GUAUGCAGCUGAGGCAUCCUAAC	23	2	fgenesh1_pg.C_LG_II000237(1)	Amine oxidase
17	AUAACGGUCCUAAAGUAGCGA	21	1	eugene3.125570001(1), eugene3.150890001(1), eugene3.27810001(1)	Hypothetical protein
22	GCUCUUCCUAUCAUUGUGAA	20	1	eugene3.57820001(0), eugene3.81440001(0), eugene3.86230001(0), eugene3.05760001(0), eugene3.72670001(0), grail3.0005004301(0), grail3.0182000701(0), estExt_Genewise1_v1.C_91780006(0), eugene3.10290008(1), eugene3.01960031(1)	Unknown
23	UUAAGCCAUGCAUGUGUAAGUC	22	1	grail3.9190000201(0), eugene3.24490002(0), grail3.13238000201(0), grail3.8915000201(0)	Unknown
32	UAGCUAACCCAAGGAAAACAGC	22	1	fgenesh1_pg.C_LG_XIX000309(1)	Peptidase S26, signal peptidase I
				eugene3.00190264(1)	Unknown
				fgenesh1_pg.C_LG_XIX000308(1), fgenesh1_pg.C_scaffold_195000001(1), fgenesh1_pg.C_scaffold_2307000001(1), fgenesh1_pg.C_scaffold_2674000001(1), fgenesh1_pg.C_scaffold_365000003(1), grail3.9675000101(1)	Transposon-encoded proteins
34	UAGUUCACCUCCCUAUGGUUCGAC	24	1	fgenesh1_pg.C_LG_I003000(2)	Aquaporin

42	UGGCCAAGCCCUCGUGGUGAA	21	1	grail3.0005004301(1), eugene3.05760001(1), eugene3.57820001(1), eugene3.81440001(1), eugene3.86230001(1), eugene3.10290008(1), eugene3.01960031(1), grail3.0182000701(1), grail3.7269000201(1)	Unknown
46	AACGACUCUCGGCAACGGAUA	21	5	eugene3.08530002(0), grail3.1214000201(0), grail3.0196001501(0), grail3.9008000101(0), eugene3.60690001(0), eugene3.10290007(0)	Unknown
53	UCACCGGGUGUAAAUCAGCUUG	22	2	grail3.0001100701(2)	Predicted lipase
56	GGUGGUGCAUGGCCGUUCUU	20	1	grail3.13238000201(0), grail3.9008000101(0), grail3.21597000101(0), grail3.17468000201(0), grail3.1214000201(0), grail3.0456000301(1) eugene3.18420003(0)	Unknown Reverse transcriptase
70	AUAACGGUCCUAAGGUAGCGAA	22	2	eugene3.27810001(0), eugene3.150890001(0), eugene3.125570001(0)	Hypothetical protein
72	ACAAAGCAUUGCGAUGGUCCCU	22	2	estExt_Genewise1_v1.C_91780006(0), grail3.0182000901(0), grail3.0072005601(0), eugene3.72670001(1)	Unknown
73	AGCAUAUCAAUAAGCGGAGG	20	1	estExt_Genewise1_v1.C_97120003(0), eugene3.166320001(0), grail3.0214000401(0), eugene3.137670001(0), eugene3.106100001(0), eugene3.105860001(0), estExt_Genewise1_v1.C_94310004(0), eugene3.83720001(0), eugene3.01960004(1), grail3.19774000201(1)	Unknown
74	CAUGCCGCGAAUCCUCUUG	19	1	grail3.13773000101(0), estExt_Genewise1_v1.C_95310005(0), grail3.0903000301(0)	Unknown
75	AGCUUCAUAGGUGGUGGUCA	20	1	fgenesh1_pg.C_LG_II000780(1)	Unknown
79	CUGCCAGGACCGCUGGGCC	19	1	estExt_Genewise1_v1.C_94310004(0), grail3.9008000101(0), grail3.0196001501(0), eugene3.08530002(0), eugene3.83720001(0), eugene3.60690001(0), eugene3.10290007(0)	Unknown
80	UCCUGCCAGUAGUCAUAUGCUUG	23	3	eugene3.24490002(0)	Unknown
82	AGGCAGUGUGGUUAGCU	17	2	fgenesh1_pg.C_LG_I002125(1)	UDP-glycosyl transferase
83	AUGAUAUUGCUUCCUUGGAUGA	22	1	estExt_Genewise1_v1.C_LG_X3123(1), estExt_Genewise1_v1.C_LG_VIII0614(1)	Glycosyl transferase, family 14

87	UCGGGAGAAGGGGUGCCUAC	22	1	eugene3.125570001(0), eugene3.27810001(0), eugene3.150890001(0)	Hypothetical protein
88	UUCGAUUCCGGAGAGGGAGC	20	1	grail3.9190000201(0), grail3.8915000201(0), grail3.21481000201(0), grail3.0190000501(0), grail3.17468000201(0), grail3.1323800201(0)	Unknown
89	UCCAGCUCCAAUAGCGUAUA	20	1	grail3.17468000201(0), grail3.0190000501(0), grail3.21481000201(0), grail3.0033002201(0)	Unknown
90	GAAGGGCCGUCGCUCAACGG	20	1	grail3.5952000201(0), eugene3.150890001(0), eugene3.125570001(0), eugene3.27810001(1)	Unknown
				fgenesh1_pg.C_scaffold_8474000002(1)	Putative cell wall associated hydrolase
94	UCAAUAAGCGGAGGAAAAGA	20	1	grail3.0214000401(0), eugene3.106100001(0), grail3.19774000201(0), estExt_Genewise1_v1.C_94310004(0), estExt_Genewise1_v1.C_97120003(0), eugene3.137670001(0), eugene3.01960004(0), eugene3.83720001(0), eugene3.105860001(0), eugene3.166320001(0)	Unknown
95	UCCGGAAUGAUUGGGCGUAA	20	1	estExt_Genewise1_v1.C_95310005(0)	Unknown
103	ACAUCUGUGUGCACUAACUC	20	2	fgenesh1_pm.C_scaffold_228000001(0), fgenesh1_pm.C_LG_XIII000193(0)	Copia-like retrotransposon family
104	AAUGGGUUAAGCAGGUUCUC	20	1	fgenesh1_pg.C_LG_XVIII000301(1)	Unknown
106	CGAUCCACUGAGAUUCAGCCC	21	1	eugene3.57820001(0), eugene3.04560006(0), eugene3.05760001(0), eugene3.66100001(0), eugene3.81440001(0), grail3.0182000701(0), grail3.7269000201(0), eugene3.86230001(0), eugene3.01960031(1)	Unknown
				eugene3.100860001(0)	Hypothetical protein
				fgenesh1_pg.C_scaffold_10733000001(0)	SNF2 family DNA-dependent ATPase
107	CGUCUGCCUGGGUGUCACGC	20	1	eugene3.83720001(0), eugene3.08530002(0), eugene3.60690001(0), grail3.0196001501(0), eugene3.10290007(0), grail3.9008000101(0), estExt_Genewise1_v1.C_94310004(0)	
113	AUCCCGUGAACCUGUAGGCACCUG	24	1	fgenesh1_pm.C_LG_V000019(2)	Putative methyltransferase DUF248
129	UGACGCGCAUGAAUGGAUU	19	1	grail3.10696000101(0)	Zinc finger protein

				estExt_Genewise1_v1.C_91780006(0), eugene3.57820001(0), eugene3.72670001(0)	Unknown
131	UGGCUGUCGUCAGCUCGUGUCGUGA	25	1	grail3.0012000801(0), grail3.13773000101(1), estExt_Genewise1_v1.C_95310005(1)	Unknown
144	AAUGCCAAUGGGUCAAAUUUUGACA	25	1	fgenesh1_pm.C_LG_XII000258(2), grail3.0074008701(2)	Unknown
				estExt_fgenesh1_pg_v1.C_LG_IX0880(2)	Putative Zinc finger protein
146	CCAGUAGUCCUAGCCGUAAA	20	1	estExt_Genewise1_v1.C_95310005(0)	Unknown
155	GUAGUCAUAUGCUUGUCUCA	20	2	grail3.9190000201(0), eugene3.24490002(0)	Unknown
				grail3.13238000201(0), grail3.8915000201(0)	Putative succinate dehydrogenase
158	CGGUUCGAGUCCGAGUGGCGGCA	23	1	eugene3.00110764(2)	Calmodulin-binding family
160	CACAAAUCGAGAAAUGGUCC	20	1	eugene3.00012570(1)	Uncharacterized conserved protein
165	GGGAGCUUCGGCGAAGGUGA	20	1	eugene3.72670001(0), grail3.7269000301(0), estExt_Genewise1_v1.C_91780006(0), eugene3.57820001(1)	Unknown
167	AUAACGGUCCUAAGGUAGCG	20	1	eugene3.27810001(0), eugene3.150890001(0), eugene3.125570001(0)	Hypothetical protein, cell wall associated hydrolase
				eugene3.125570001(0)	Conserved hypothetical protein
171	GGGGGAGUAUGGUCGCAAGG	20	2	grail3.17468000201(0), grail3.13238000201(0)	Unknown
172	AUGGUUGGUGCAACUUCUUUUGA	23	3	estExt_fgenesh1_pg_v1.C_LG_I2049(1)	Protein involved in membrane traffic (YOP1/TB2/DP1/HVA22 family)
173	AAUCAGGUUAUGGACCGAUAAAAAA	25	1	fgenesh1_pg.C_LG_VI000069(2)	Cytochrome P450 CYP4/CYP19/CYP26 subfamilies
174	AUUAGUGACGCGCAUGAAUGGAUU	24	1	estExt_Genewise1_v1.C_91780006(0), eugene3.57820001(0), eugene3.72670001(0)	Unknown
				grail3.10696000101(0)	Zinc finger protein
				fgenesh1_pm.C_LG_I000669(2)	Plectin-related protein
182	UGUGUUGUGUAGUGAUGUGA	20	2	estExt_fgenesh1_pg_v1.C_LG_III0067(0)	MYB-related protein
183	UGAAAUGCCAACUGGGGGU	19	1	grail3.0091000801(1)	Acetylglucosaminyltransferase
186	GCUGUCGAAGUUCCAUCUACAAA	23	1	fgenesh1_pg.C_LG_II002119(1)	Major intrinsic protein family
189	AACCCGGUUCAAACUCUAACA	21	2	grail3.0093002701(2)	PPR repeat
195	AUGCAAAAACUCACGUGAUGGUUA	24	1	eugene3.00161333(2)	Ribosomal protein S6 kinase and related proteins
197	UGUGUCUGUAAUGGAUGAAGU	21	1	eugene3.01270009(2)	PPR repeat
				estExt_fgenesh1_pg_v1.C_LG_III0099(2)	Kinesin (KAR3 subfamily)

198	CACAAAUCGAGAAAUGGUCC	20	1	eugene3.00012570(1)	Uncharacterized conserved protein
205	GCUUCUUUGAAAACAAGUCCAUAAA	25	1	eugene3.01700038(1), eugene3.01700060(1)	Unknown
208	AGGAUGAUCAGCCACACUGGGAC	23	1	grail3.0286000901(0), grail3.0286000901(0), grail3.13773000101(0), estExt_Genewise1_v1.C_95310005(2)	Unknown
225	CCGUGAGAUUAGUCGAGGUGCGU	23	1	fgenesh1_pg.C_LG_II000444(1), fgenesh1_pm.C_LG_VI000336(1), fgenesh1_pg.C_scaffold_130000053(1), fgenesh1_pg.C_scaffold_66000036(1), fgenesh1_pg.C_LG_V0010899(1), fgenesh1_pg.C_LG_VII000951(1), grail3.17427000101(1), fgenesh1_pg.C_LG_XVIII000545(1)	Unknown
				grail3.0066014202(1)	Ca2+-binding protein
				estExt_fgenesh1_pg_v1.C_LG_IX0420(1)	Subtilisin kexin isozyme-1/site 1
				eugene3.00031423(1)	Acyl-CoA synthetase
226	CAUUUCUUCUUAGCUGCUUGGC	22	1	grail3.0030020501(2)	Unknown
				fgenesh1_pg.C_LG_I002327(2)	DNA binding WRKY protein
				fgenesh1_pg.C_LG_XVIII000611(2)	Nuclear pore complex, Nup98
229	CAUAAUCUCAUCGUUUCGUU	20	2	estExt_Genewise1_v1.C_LG_XII1015(0)	Unknown
230	GCAAAGGCAGAAGGGAGCUUGA	22	1	eugene3.125570001(0), eugene3.150890001(0), grail3.5952000201(0), eugene3.27810001(0)	Conserved hypothetical protein
				fgenesh1_pg.C_LG_II002220(1)	Karyopherin (importin) beta 3
232	UCGAUCCUUUAGACCUUCGGAA	22	3	fgenesh1_pg.C_scaffold_182000013(0), estExt_Genewise1_v1.C_91780006(0), estExt_Genewise1_v1.C_91780006(0), grail3.9892000201(0), grail3.0182000701(0), eugene3.05760001(0), eugene3.72670001(0), eugene3.86230001(0), eugene3.57820001(1), eugene3.81440001(1)	Unknown
233	GUAUCCAGAGCGUAGGCUUG	20	1	eugene3.148650001(0), eugene3.02190005(0), eugene3.131840001(0), eugene3.111770001(0), eugene3.102800001(0), grail3.10051000101(0), grail3.11871000101(0)	Unknown

grail3.12102000101(0), grail3.0196000201(0),	Unknown
grail3.9345000101(0), grail3.1043000101(1),	
eugene3.98720001(1), estExt_Genewise1_v1.C_62520001(1),	
grail3.0214000501(2)	
grail3.0196001101(0)	Hypothetical apoptosis regulator
grail3.9008000501(0)	Hypothetical protein

[a] The small RNA ID assigned according in the order of cloning.

[b] Sequence of putative siRNAs given in 5' to 3' orientation.

[c] Length of the cloned small RNA in nucleotides (nt).

[d] The frequency of cloning represents the number of times the specific sequence was cloned.

[e] Putative targets are given as the predicted gene model name obtained from the draft populus genome assembly at http://genome.jgi-psf.org/Poptr1/Poptr1.home.html. Number of mismatches of the putative siRNA and the complementary site on the target are given in parentheses. For siRNAs longer than 20 nucleotides only targets with up to 1 mismatch are given, for siRNAs longer than 20 nucleotides predicted targets containing up to 2 mismatches are listed.

[f] Predicted target functions were assigned based on preliminary annotations from the draft *populus* genome (http://genome.jgi-psf.org/Poptr1/Poptr1.home.html) and a protein domain search (http://www.ebi.ac.uk/interpro).

Table S2.2. MiRNA stem-loop locations and precursor sequences in *Populus*.

ptc-miRNA gene[a]	miR stem location[b]	Precursor Sequences (5' - 3')[c]
ptc-MIR26a	scaffold_628:11989-11808	CCTCCTATCAAATAATTTAAATCTAGTTATTAGAGACTTAAAGTGTTTTT ACATGGTATATTAGTTATTACCAAATGATCAGGGGTTAATGTGTTTCTG CATTTTAAAAATATAATGAAAAAAACTAAAATACTTTTAAAAGCAAAATC TTCAATGCTGGTGTTAAAAGGCAATTTGATAT
ptc-MIR26b	LG_XIII:4014178-4013847	GCCCAATTTGAGCTTAATTGAGGAGATTGAAATTTTAGAGGACCAAACAT TAAATTTTTACACAGTCGATTGGTTGAAATCAGGGATACAATTGCAAGAA AATCAAAGTTTAGGGTTAATTAGGGGTTAAATTGAAGAAATTCACAGCCA AAGATGTTTTTTGAAAAAGGCGCCGAACTCTGAGGACTCAATTGATTGAA ATCAAATGTGAAATTGAAGAGAATTGAAAGTTTAATGGTCAATTGAGAGT CAATTTGCATAAATCCAAGACCAAAGACCAAAATAAAAAAAGGCGCTGAA ATCTTCAATGCTGATGTTGAAGTTCGACAAG
ptc-MIR39a	LG_XVI:12106164-12106072	CACTAAGCATCCATGGTTGAATAGTTAAAGCACCCAACTCATAATTGGTG AATTCATAGTTTTAATTCCAACTAGATGAATGCCAATGTGACC
ptc-MIR39b	LG_XVI:12079317-12079260	TTAGTTATTCACCATGTATGAGGATCCCCGCCAAGCATCCATGGTTGAAT GGTTGAA
ptc-MIR44	LG_VIII:8093051-8093071	CTGGGAGTTACGGGTATGGGAGGATTGGACAGTACTGCTTGGTTTTAATT AACCAAAGTCTGTTCTTTCCAACGCCTCCCATACCGGTAATCTCCAGC
ptc-MIR56a	scaffold_21597:141-339	GAATTCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACG GAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGG GAAACCTCACCAGGCCCGGACACTGGAAGGATTGACAGATTGAGAGCTCT TTCTTGATTCAGTGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGC
ptc-MIR56b	scaffold_6643:1495-1780	GCACCCAAAGAAGAGGTGGCTATGAGACTCATCATCCACATGGCAGAAAA TACAAGTGGAATCAGTTTGGAGGAAGTGAAGCCTATCTCTTGTACGAAGC CTGCCCAAGATTGCCAACCACAGGATGAAACTGAACCGAGGCAGGGACCA GGGCTCCCAAATAACTCTTCTCCAGTGGACTTGGGAGCTGCGAAATCTTA AGTAGTCATAGGCATTGGATGAGAATGGGCAACTGAGAGCTCTTTCTTGA TTCTATGGGTGGTGGTGCATGGCCGTTCTTAGTTG
ptc-MIR56c	scaffold_520:40810-40965	AACAGGGAAGGACAAGAGCAAACAACGGCAGTAGATCAAACTCGAGGGAA AAGCTTCACAGAGAAGGAGAGCATCTCTCTCTAAATTCTCTCTCTGACTG AGAGCTCTTTCTTGATTCTATGGGTGGTGGTGCATGGCCGTTCTTAGTTG GTGGAG
ptc-MIR56d	scaffold_196:124123-123997	GGCTAATTACTGGAGAAAGGGACGGGTGCAGTAGGCAGCCATGATCGAAG GAGGAAACAGCTTCCGCTCTACTACTCTTTCTTGATTCTATGGGTGGTGG TGCATGGCCGTTCTTAGTTGGTGGAGC
ptc-MIR69a	LG_VIII:8092808-8092913	AGTCCTAGCAAGTCTTTGGAGATGGGAGAGTATGCAAGAAGGAAAAATTC ATGATTTAATATTCTTTCTTGCCTACTCCTCCCATTCCATCTGCTTTCTG CGACTC
ptc-MIR69b	LG_VIII:8092806-8092913	GAGAGTCCTAGCAAGTCTTTGGAGATGGGAGAGTATGCAAGAAGGAAAAA TTCATGATTTAATATTCTTTCTTGCCTACTCCTCCCATTCCATCTGCTTT CTGCGACTC
ptc-MIR81	scaffold_129:575576-575709	ATCATGCACAAATTGAGAAATGGTCTGTGTGCTTGGCAGAGCTGAACTAG GATGCATGAAATTGGTGATAGTTGACTGCATGTCTTGATAGCTTCAATTG CTATAATAGGGCAATTATTAATTTGTGGATGTAT
ptc-MIR99	LG_VII:120921-121011	TGCAACCATGAATATATTTGTTGATAGTCATCTAGTGAACATTGAAAGTC ATATTAGATGACCATCAACGAAAACATTCATGATATTGCAA
ptc-MIR170	LG_IV: 6840773-6840614	CTTTAATTCAAAGAAGGCCACTGTTTGTTTTGTTTGATGGCTGCTTGATG ATTCTCTTTTCTATGAGAGTGGAGGTTGTTTTATAATCTCTCTTTTTATA GAGGAAGAAGAGGGCTTTTGTCAAGCATTACAAACAAGTTTTTCTTTGG CTTAAACAAA

ptc-MIR235a   LG_XVIII:5423584-5423917

```
GAAGATCGATGGCTATGGAAGTTATGCATGTAGGTATTTGCATCTTTCAT
TTATTGATTTGTTGCTCGTTACATTTCTTTTAGTTTAGCAGCCATAGTAG
TGGCATTCTCAAAATTGAGCCATGCCTCTTTCTCTCGCTTAGAAGAGAAA
GAGAGGATTAATACCTGCTGGCAACGATTGCAGTTACTTTATGTTTAATC
TAATATAATATAATATGGTTAAATATCGTTGTTGGAAGTTGGCATCTCTG
TTGAAAGAGGAGACTGTCTAAACCCTTAAACCAAAATAAACTACCCCCAG
CAAAACCCAGAACCATCATGATTCATATGCATGG
```

ptc-MIR235b   LG_XVII:2785382-2785553

```
GATCGTGTAGCTCGAAGTTATGGAAGTTATTCAGGAACGTTTGGTGTGTC
TTGAGATTTTGATAGGGCAAGAGGAATAAGGTGAAGAATGCTCTATCATT
GATCGATTAAAGGAAGCTGTGGAAAGTGTTGAAAGGATTGAAAGTTTGTA
CATTTCTCTTGCGAAAAAATCG
```

ptc-MIR156d   LG_XVIII:1365495-1365685

```
GGGTTTTAAGGGTAAGTGAGTTGACAGAAGAGAGTGAGCACACAGGGTAC
TTTCTTGCATGACGTTCATGCTTGAAGCTTTGCGTGCCCACCCTCTATCT
GTCACCCCATCACCATCTCTCCTTCTCTTCCCCTTTCCTTGCCTGTCCTG
CGTACTTATGTTGTGTGTGAATGAATTCTTAGTGTTTGAAT
```

ptc-MIR156e   LG_XVIII:6180951-6181169

```
ATTTTATGCTCTCTCTTTCCTTTCCATTTTTGGGTATGTTTTATCTTCCA
CTCAGAGCTCAAAACTAGAACAGAAAAGGTTTTGCTTGCTTGTTTTCTTT
CCCGTTTTTTTCCAGCTTCTGGGACATTAGAAATTGACAGAAGAGAGTGAG
CACACAGAGGCATATTTGTATAAAATTATACCATTGCTTTTGCGTGCTCA
TTTCTCTTTCTGTCACTT
```

ptc-MIR156g   LG_II:21820430-21820436

```
TGTTGACAGAAGATAGAGAGCACAGATGATGAAATGCATGGAGCTTAATT
GCATCTCACTCCTTTGTGCTCTCTAGTCTTCTGTCATCACC
```

ptc-MIR156h   LG_IV:6006026-6006119

```
TGTTGACAGAAGATAGAGAGCACTGACGATGAAATGCATGGAGCTTAATT
GCATCTCACTCCTTTGTGCTCTCTAGTCTTCTGTCATCAC
```

ptc-MIR156i   LG_XII:4450915-4450822

```
TGTTGACAGAAGATAGAGAGCACAGATGATGTTTTGCAGTAGACTCTGGA
TCTCACTCCTTTGTGCTCTCTATGCTTCTGCCATCACCT
```

ptc-MIR156j   LG_XV:3242395-3242488

```
TGTTGACAGAAGATAGAGAGCACAGATGATGATATGCAATGGACTCTGCA
TCCCACTCCTTTGTGCTCTCTATGCTTCTGTCATCA
```

ptc-MIR159a   LG_XIII:4201528-4201717

```
GAGTGGAGCTCCTTGAAGTCCAATAGAGGTTCTTGCTGGGTAGATTAAGC
TGCTAAGCTATGGATCCACAGTCCTATCTATCAACTGAAGGATAGGTTTG
CGGCTTGCATATCTCAGGAGCTTTATTGCCTAATGTTAGATCCCTTTTTG
GATTGAAGGGAGCTCTAAA
```

ptc-MIR159b   LG_XV:3478465-3478654

```
GAGTGGAGCTCCTTGAAGTCCAATAGAAGCTCCTGCTGGGTAGATCGAGC
TGCTGAGCTATGAATCCCACAGCCCTATCACCATCAGTCATTTTGATGGG
CCTGCGGCTTGCATATCTCAGGAGCTTTATTACCTAATGTTAGATCTTTT
TTTGGATTGAAGGGAGCTCTAAACCTTTGATCTTTTCAGC
```

ptc-MIR159c   scaffold_1140:8347-8536

```
TAGGGAGTGGAGCTCCTTGAAGTCCAATAGAGGTTCTTGCTGGGTAGATT
AAGCTGCTAAGCTATGGATCCACAGTCCTATCTATCAACCGAAGGATAGG
TTTGCGGCTTGCATATCTCAGGAGCTTTATTGCCTAATGTTAGATCCCTT
TTTGGATTGAAGGGAGCTCTAAACCCATAACCCCTTCAAC
```

ptc-MIR166a   LG_I:6096929-6097080

```
ATAGAAATGGAAGCTATTTCTTTTGAGGGGAATGTTGTCTGGCTCGAGGT
CACTAATGGGATCTATGATTTTATCTCAATTGATTGATTTTCTTTTAAAT
TCTAGTAAATTGAATTGAGAGATATCATGATCAACTTATATTTAATGATG
TCGGACCAGGCTTCATTCCCCCCAATTGTTTCTTCCTTTC
```

ptc-MIR166b   scaffold_122:445042-445157

```
GGAAAGTGAGGATGTGGGGGAATGTCGTCTGACTCGAGACAACAGGTAAA
CTGAAGGAATGGCCGGAGTTTAAGAGTTCTCTCGGACCAGGCTTCATTCC
CCTCATCCACACCTTCC
```

ptc-MIR166c   LG_XV:460476-460361

```
TAGCTAGGAAGCTTAACTTTTGTGGTGAATGTTGTCTGGTTCAAGGCATG
GCCACCACATCTCTTGGTGAATATATGTCCTCGGACCAGGCTTCATTCCC
CTCAATTAATGCTTCC
```

ptc-MIR166d   LG_XIV:3909056-3909197

```
GGGGAATGTTGTCTGGCTCGAGGACTTTTTGTTCATCAATCTAATCGAAC
TTTCTACCTGTAGATCTAGTATCTTATTTAAGATTGATCACGTATTAGGG
TTGTCGGACCAGGCTTCATTCCC
```

ptc-MIR166e   LG_XIV:3895087-3895227

```
GGGGAATGTTGTCTGGCTCGAGGACTTTTTGTTCATCAATCTAATCGAAC
TTTCTACCTGTAGATCTAGTATCTTATTTAAGATTGATCACGTATTAGGG
TTGTCGGACCAGGCTTCATTCCCCCCAATCATTGCTTCCAT
```

ptc-MIR166f   LG_XIII:2546006-2546111

```
GGAAGCTTGTCTTTTGAGGGGAATGTTGTTTGGTTCAAGGCCTGGCCACC
CCATGTCTTGGAATTTAAAATATCATGTCCTCGGACCAGGCTTCATTCCC
CTCAATTACTGCTTCC
```

ptc-MIR166g	LG_X:8492275-8492368	AGTTGAGGGGAATGCTGTCTGATTCGAGACCATTCACTTTAAGCACACAT TCATCTTTCGAATGATCTCGGACCAGGCTTCATTCCCCCCAACT
ptc-MIR166h	LG_VIII:11091691-11091583	TGCTACACAGTTGAGGGGAATGCTGTCTGGTTCGAGACCATTCACCTGAA GAGCACGCATTCATCTTTTGAGTGATCTCGGACCAGGCTTCATTCCCCCC AACTCAAAA
ptc-MIR166i	LG_V:824495-824399	GTTGAGGGGAATGTTGGCTGGCTCGAAGCTTAAGCAAAGAGTTTTCTCTC AAGAAACAACTGTTAAGGCTTCGGACCAGGCTTCATTCCCCTCAACCAAT
ptc-MIR166j	LG_VII:3395337-3995234	GTGTGTTGAGGGGAATGTTGGCTGGCTCGAAGCTTAAGCAAAGAGTTTCC TAACATGAAACAACTGTTAAGGCTTCGGACCAGGCTTCATTCCCCTCAAA CATA
ptc-MIR166k	LG_V:15276296-15276176	TCTCGCGTATTAACTGTTGAGGGGATTGTCGTCTGGTTCGATGTCATTCA TGAGAAGCTCAAACATAAACGTAATATTGAATGATTTCGGACCAGGCTTC ATTCCCCCCAACAATGTTTGA
ptc-MIR166l	LG_II:2737618-2737737	GCATGTATTAACTGTTGAGGGGATTGCTGTCTGGTTCGATGTCATTCATG TGAAGCTTTAACATTAATGTAGTATTGAGTGATTTCGGACCAGGCTTCAT TCCCCCCAACTATATTTGTT
ptc-MIR166m	LG_II:13668142-13668286	AGCTTCTTTTTTTGAGGGGAATGTTGTCTGGCTCGAGGACTCTTTCTTGA TCAGTCTGATCAAGTGTTCTATCTTTAGATCTAATATCTTAGATCATGTG TTAGGGTCGTCGGACCAGGCTTCATTCCCCCCAATTATTGCTTCC
ptc-MIR166n	LG_XIX:5391607-5391268	GGGGTGTTTGGAATGAAGTTTGATCCAAGATCCTTGTCTCTCCCGTTAAC TTAGTCTCTGTTACTGTTAGGTTTTCATTACTGAGTGTATTTGCAGCCCC CTTGTCTGATTTTAGCATCATGAGAGTAGGAAGTGCTGGTGATCATAGGG TTTGGTTCAAGATCCATTTGACTCTTCTCTTTATATCTCTCCTGTGTCCT AGCTGGTAATCTGTAGTAGTCTTAATTATCCCTTACTTCTTGTTAATTTT TAGGTTTGATCTTGCAAGTAATTATATTTGGAAAATATCATAAAATGATC TCGGACCAGGCTTCATTCCTTACACCTTGCTTTTCTTTTA
ptc-MIR166o	LG_VII:3395022-3394878	GTTGAGAGGAATGTTGTCTGGCTCGAGGTCATTGAGGCCATGATTATACA GACATGGCATTACCTGATGACAGCCGAGAAAATTCAAAGTCTGTGTGTAT CTTGTACCTCGATGATTCTCGGACCAGGCTTCATTCCTTCCAACA
ptc-MIR166p	LG_V:824362-824206	GCGTAAGGTTGAGAGGAACGCTGTCTGGGTCGAGGTCATGGAGGCCATGA TTATACATAAATGGCATTATCTGATGACAGCCCAGATAATCGATGCACCT GTCTTGAACCTAAATGATTCTCGGACCAGGCTCCATTCCTTCCAACCATC ATTTGCT
ptc-MIR166q	LG_XIII:4057334-4057502	GTGAGCACTAGTGATAGAGTTTGATTCAAGATCCATCTCATGTGTGCTAG CTAGCTTTTAAATTTTTAATCAGATCCCTACTTCCTGTTAATTCTTAGGT TTTATCTTGCAAGTTATTATATTTAGTAAACATCATGAACTGATCTCGGA CCAGGCTTCATTCCTTACA
ptc-MIR167a	LG_II:3057682-3057565	GGGAAAAAGTGAAGCTGCCAGCATGATCTATCTTTGGTTAGAGAAAGAAA GGACTAACCCTAGCTAGGTCATGCTGTGACAGCCTCACTC
ptc-MIR167b	scaffold_14274:302-178	GTTCACAAGGGAAAAGGTGAAGCTGCCAGCATGATCTATCTTTGGTTAGA GAAGGATAGAAGCGAAAGAACTAACCCTAGCTAGGTCATGCTCTGACAGC CTCACTCCTTCCTATTTGGGGACCC
ptc-MIR167c	LG_V:14949516-14949625	TCGTGCACCACTAGCAGTTGAAGCTGCCAGCATGATCTAAATTAACCTCC TTCTTTATCAAGGATGGATTAGATCATGTGGTAGTTTCACCTGCTGATGG CATCACGAAA
ptc-MIR167d	LG_II:3055016-3054896	CATCGTGCACCACTAGCAGTTGAAGCTGCCAGCATGATCTAACTTCCTTG CTTCTTTATCAAGGATGGATTTAGATCATGTGGTGGTTTCACCTGTTGAT GGCACCACGGAAAAATCCCAG
ptc-MIR167e	LG_VIII:10761707-10761813	TCGTGCACCACTATCAGTTGAAGCTGCCAGCATGATCTTAACCTCCCTCC TTTGTCGAGGAAAGAACAGATCATGTGGCAGTTTCACCTGAAGTTGGTAT CACGAGA
ptc-MIR167f	LG_X:9042307-9042189	TCGTGCACCTCTATCAGTTGAAGCTGCCAGCATGATCTTAGCCTTCCTCC TTTGTTGAGGAAAGAAACAGATCATGTGGCAGTTTCACCTGTTGTTGGTA TCACGAGAAACCCTAATTC
ptc-MIR167g	scaffold_70:484778-484660	TCATGCACCACTAGTAGTTGAAGCTGCCAGCATGATCTGAACTTTCCTTA ATTTTCCTATACGGGAAAGACTAGATCATGTGGTAGTTTCATCTATTGAT GGTAGCATGGGAAACCCTA

ptc-MIR167h   LG_XIII:2740285-2740396

TCATGCACCGCTATTAGTTGAAGCTGCCAACATGATCTGAGCTTTCCTTA
ATTTTCCTATACAGGAAAGACTAGATCATGTGGCAGTTTCACCTATTGAT
GGTAGTATGGAA

ptc-MIR167i   scaffold_70:484780-484668

AATCATGCACCACTAGTAGTTGAAGCTGCCAGCATGATCTGAACTTTCCT
TAATTTTCCTATACGGGAAAGACTAGATCATGTGGTAGTTTCATCTATTG
ATGGTAGCATGGG

ptc-MIR396a   LG_VI:4075637-4075775

TGACCCTCTTGGTATTCTTCCACAGCTTTCTTGAACTGCACCTATTTGAA
TATTGTTGTTGATGTTGCCGTGCATGTACATATGACATTGTATTTTTGTT
GCGGTTCAATAAAGCTGTGGGAAGATACAAACAGGATCA

ptc-MIR396b   LG_XVIII:12249317-12249464

TGACCCTCTTTGTATTCTTCCACAGCTTTCTTGAACTGCACCTATTAGAT
TAATGTTGATGTTGTTGTGCGATATGCCATGACCATATGACATTGTATTC
ATTTTTGCTGCGGTTCAATAAAGCTGTGGGAAGATACAAACAGGATCA

ptc-MIR396c   LG_VII:6490285-6490100

TAAGGTCGTTGCATGCTTTTCCACGGCTTTCTTGAACTTGGCACTCAAGA
GACATGAGAGTAAAGGGCTAGGCTTTCTTTTCTATTCCTTTCGTTCTTTG
AAATTTTCCTGTTCAAAGAATCCATACATATCAGTCTCTTTAGTACAAAG
CTCAAGAAAGCCGTGGGAAAATATGAAAGGATTTTG

ptc-MIR396d   LG_VI:4082560-4082366

AAGTCCTGGTCATGCTTTTCCACAGCTTTCTTGAACTTCCTTGCCATGCT
TAACTTGTGTGTGTGTGTGTGTGTGTGTAGATCACTATATATCTGTAT
AGTTATACACATTTAGCTAGCTAGCACCATGGAAGCTCAAGAAAGCCGTG
GGAGAACATGGCAACTCAGGATTTTTGGGGCGATCTGTTTCTTG

ptc-MIR396e   LG_XVIII:12256370-12256209

AAGTCCTGGTCATGCTTTTCCACAGCTTTCTTGAACTTCTTTGCCTTGCT
TAATCTGTGTATATATAGATCACTACATGTACAGCTCCTATATATATAAA
TATATGTATGTATAGCGCCATGGAAGCTCAAGAAAGCTGTGGGAGAACAT
GGCAATTCAGGATTT

ptc-MIR396f   LG_III:14406861-14406958

CATGTTTTTCCACAGCTTTCTTGAACTTCCTAGAGCCTAGAGGTGCTGCT
AGCTATACATATAACTTAAGAAGTTCAAGAAAGCCGTGGAATAGCATG

ptc-MIR396g   LG_VI:17347070-17346926

ATGGGTCTCTTGGTGATTTTCCACGGCTTTCTTGAACTGTATATATTATC
AATGGCTTTTTACAAGACTGGAAGATGGTTTCCATGGAGAAGAATTGTCA
CAAAAACAGTTCAAGAAAGCCCTGAAAAATTATTTAGAGGATCTA

ptc-MIR396h   LG_III:14406968-14406847

AACACCTGGTCATGCTATTCCACGGCTTTCTTGAACTTCTTAAGTTATAT
GTATAGCTAGCAGCACCTCTAGGCTCTAGGAAGTTCAAGAAAGCTGTGGA
AAAACATGGCAGTTTCAGGGTT

ptc-MIR396i   LG_VI:4075775-4075637

TGATCCTGTTTGTATCTTCCCACAGCTTTATTGAACCGCAACAAAAATAC
AATGTCATATGTACATGCACGGCAACATCAACAACAATATTCAAATAGGT
GCAGTTCAAGAAAGCTGTGGAAGAATACCAAGAGGGTCA

ptc-MIR396j   LG_XVIII:12249461-12249317

TGATCCTGTTTGTATCTTCCCACAGCTTTATTGAACCGCAGCAAAAATGA
ATACAATGTCATATGGTCATGGCATATCGCACAACAACATCAACATTAAT
CTAATAGGTGCAGTTCAAGAAAGCTGTGGAAGAATACAAAGAGGGTCA

ptc-MIR396k   LG_XVIII:12256212-12256370

TCCTGAATTGCCATGTTCTCCCACAGCTTTCTTGAGCTTCCATGGCGCTA
TACATACATATATTTATATATATAGGAGCTGTACATGTAGTGATCTATAT
ATACACAGATTAAGCAAGGCAAAGAAGTTCAAGAAAGCTGTGGAAAAGCA
TGACCAGGA

ptc-MIR398a   LG_X:14819090-14819207

AGCCGGTACACCCCAGAGGAGTGGCTCCAGAGAACACAGGGGGTTGGTTT
TCTAGCTGTAAGCTACAAGATGGACAAAGCACTCTGTGTTCTCAGGTCAC
CCCTTTGGGGCACCCTGT

ptc-MIR398b   LG_I:19895317-19895221

GTAAAAGGGTTCCGACAGGAGCGGCCTTGGGTCACATGTGGCGGCATCCA
ATGTACTTCATGTGTTCTCAGGTCGCCCCTGCCGGGCTTCCTCTACA

ptc-MIR398c   LG_IX:6181820-6181925

TGGAGGATTCCTACAGGAGCGACCTGAAATCACATGTGGGCTGCACCCTC
CTGGGTTATCTTGAGCAACATGTGTTCTCAGGTCGCCCCTGCCGGGCTTT
CCTCCA

ptc-MIR408   LG_II: 16909613-16909423

TGGCAGTAGCTGGCATGGTATGGAAGCACAATGAAAGGTGAAGGTACGTA
AACAGGGATCGTGGGGAGAGGGAGCAGAGACAGATGAAGACGGGGAACAG
GCAGAGCATGGATGGAGCTACTAACAGAAGTACTTGTTTTGGCTCTACCC
ATGCACTGCCTCTTCCCTGGCTTGTGGCTCTTCCTTTTCTC

ptc-MIR472a   scaffold_131:138297-138404

ATTTTCGGAAGGTTATCGGGTGGGTGAGCGGGGAAGATAACTTTGGTTTT
TGAGATAGTACTTGTTATTTTCCCTACTCCACCCATCCCATAGGTTTCCG
ATCATTCC

ptc-MIR472b	LG_XIII:203670-203768	GGTTTTCGGAAGGTTACTGGATGGGTGAGTGGGGAAGATAACTAAGCTCT GTTTGTTATTTTCCCAACTCCACCCATCCCATAGGTTTCCGATCATTCC
ptc-MIR475a	LG_II:20949411-20949550	GTTATAACATCTTGATCAATGGCCATTGTAAGAGTAGAAGGATCCATGAA GCAATAACTCTCTTTGCTAAAATGTGTGATAAAACATTGACTCCTGGTAT TTCACTCTTACAGTGCCCATTGATTAAGATGGTATAATGT
ptc-MIR475b	LG_VIII:13883728-138834538	GTTAAAATATAATGGTCAACCAAGGTTGCGCTTCTGGCTTTCATAGTCAT AACATCTTGATCAATGGCCATTGTAAGAGTAGAAGGATCCATGAAGCAAT AACTCTCTTTGCTGAAATGTGCGACCAAATAGACTCTTGATATTTCATTC TTACAGTGCCCATTGATTAAGATGTTATAACATTGTTCTCA
ptc-MIR475c	LG_X:12045764-12045893	ATAACATCTTGAACAATGGCCATTGTAAGAGTAAGTGGTCCATGAGTCAA TAACTCTCTTTGCAGAAATGCATGACAAAGCATTGACTCCTATTTCACCC TTACAATGTCCATTGATTAAGAGGTTATAA
ptc-MIR475d	LG_VIII:8026975-8026865	GTCATAACATCCTGATCAATGGCCATTGTAAGAGTAGAAGGATCCATGAA GCAAAACATTGACTCCTGATATTCCACTCTTACAGAGTCCATTGATTAAG ATGTTATAACA

[a] *Populus* small RNA gene name as specified by ptc-MIR. Putative novel ptc-miRNA genes and sequences are listed first followed by known ptc-miRNA genes.

[b] Genome locations of the miRNA precursor presented as "genome ID (http://genome.jgi-psf. org/ Poptr1/ Poptr1.home.html): Ptc-MIR start-stop site with the ID".

[c] Sequence of identified small RNA precursors given in 5' to 3' orientation. Only the portion of the analyzed sequence capable of forming a stable fold-back structure in figure 2.1 and 2.2 is presented here as the ptr-miRNA precursor molecule. The sequences presented do not represent full length small RNA genes.

CHAPTER 3

ISOLATION AND EXPRESSION PROFILING OF MICRORNAS FROM THE VASCULAR TISSUES OF *EUCALYPTUS*

Michelle Victor[1], Ying-Hsuan Sun[2], Shanfa Lu[2], Erika Asamizu[3], Satoshi Tabata[3], Vincent L. Chiang[2], Alexander A. Myburg[1]

[1] *Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, 0002, South Africa*

[2] *Forest Biotechnology Group, Department of Forestry and Environmental Resources, College of Natural Resources, North Carolina State University, Raleigh, NC, 27695, USA*

[3] *Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan*

This chapter has been prepared in the format of a manuscript for a peer-reviewed research journal, as is the standard procedure of our research group. I performed the isolation, sequencing and expression profiling of *Eucalyptus* microRNAs, and also wrote the manuscript. Putative miRNA precursor gene sequences were identified by Dr. Erika Asamizu from the Kazusa DNA Research Institute, under Dr. Satoshi Tabata. Prof. Vincent Chiang and Dr. Shanfa Lu assisted with guidance and expertise in the isolation and identification of microRNAs. Dr. Ying-Hsuan Sun provided the computer script for automated mfold analysis and provided target predictions of the putative *Eucalyptus* miRNAs. I analyzed the mfold results and target predictions in order to identify the putative miRNAs described in this chapter. Prof. Alexander Myburg provided the concept of this study and assisted in the editing of the chapter.

3.1 ABSTRACT

Xylogenesis, or wood formation, involves the differentiation of a lateral meristem tissue (the vascular cambium) into mature xylem cell types. MicroRNAs (miRNAs) are known to play specific roles in the regulation of developmental processes and the specification of cell lineages. In plants, miRNAs are involved in the regulation of developmental processes such as the differentiation of meristematic tissues into mature organs, leaf morphogenesis, flowering time, floral organ identity, and apical meristem identity. Since miRNAs are involved in the differentiation of other meristems, we hypothesized that miRNAs are involved in regulating aspects of cambial differentiation. In order to test this hypothesis we isolated and characterized small RNAs from three differentiating stem tissues of *Eucalyptus grandis*: mature xylem, immature xylem and phloem. We identified 48 predicted *Eucalyptus* miRNAs grouping into thirteen putative miRNA gene families. Twenty of these belong to five families previously identified in other plant species, whereas the remaining 28 miRNAs group into eight putative novel miRNA families. Target predictions performed using *Populus* and *Arabidopsis* genome sequences revealed 110 predicted target genes for these thirteen families. The target genes include transcription factor families involved in specific aspects of development, auxin response factors and signalling proteins. The expression patterns of these miRNAs were profiled across a number of xylogenic and non-xylogenic tissues of *Eucalyptus* and these results suggest that miRNAs play a functional role during vascular development in eucalypt trees.

3.2 INTRODUCTION

The genus *Eucalyptus* includes some of the most widely used tree species in industrial plantation forestry worldwide, and is amongst the most prolific producers of wood known, for pulp and paper production as well as solid wood products (Eldridge et al., 1993; Potts and Dungey, 2004). *Eucalyptus* tree species are renowned for fast-growing hybrids and are some

125

of the most efficient producers of wood fiber and cellulose on earth (Eldridge et al., 1993; Potts and Dungey, 2004; Poke et al., 2005). Clonal plantations of these hybrid genotypes can produce up to 40 m^3 of wood per hectare per year in countries such as Brazil and South Africa (FAO, http://apps.fao.org). The high rates of biomass production observed in these trees are largely due to very efficient cell proliferation and differentiation of the lateral meristem. The morphology and chemical composition of mature xylem cells are important factors affecting pulp processing and the quality of solid wood products derived from these trees (Rudie, 1998; Evans and Ilic, 2001; Barnett and Bonham, 2004).

Xylogenesis, or the formation of wood, is a complex developmental process known involving the differentiation of a meristematic tissue into the specialized cell types found in wood. This process involves the periclinal differentiation of a secondary lateral meristem tissue, the vascular cambium, into the vascular tissues of xylem and phloem. These cells continue to differentiate and finally give rise to secondary xylem (wood) and mature phloem tissues (reviewed by Burton et al., 2000; Plomion et al., 2001). Since these two main vascular tissues are functionally and chemically distinct, each requires the expression of a specific set of genes to ensure the maintenance of their specific functionality. Xylogenesis requires the coordinate expression of a large number of genes, some of which have been shown to be differentially expressed in the developing tissues surrounding the vascular cambium (Hertzberg et al., 2001; Israelsson et al., 2003; Yang et al., 2004; Ranik et al., 2006). Although the physiological processes and many of the genes involved in xylogenesis are relatively well documented, the regulation of key developmental genes such as hormone responsive transcription factors, and the regulation of differential gene expression in woody tissues is not well understood.

MicroRNAs (miRNAs) are a group of small non-coding RNAs involved in the post-transcriptional regulation of gene expression. In plants miRNAs have been shown to be

involved in the regulation of complex developmental processes including the specification of meristem identity (Jacobsen et al., 1999; Llave et al., 2002a; Kidner and Martienssen, 2005; Wu and Poethig, 2006; Wurschum et al., 2006). MiRNAs are further known to be involved in tissue-specific gene expression patterns in plants and animals. The temporal and spatial regulation of miRNA expression allows the controlled timing of developmental events through the regulation of asymmetric gene expression (Pasquinelli et al., 2000; Banerjee and Slack, 2002; Lagos-Quintana et al., 2002; Johnson et al., 2003; Ko et al., 2006a). In other studies, miRNAs have been shown to be essential in the differentiation of meristematic tissues into mature cell types, where the miRNAs themselves exhibit preferential expression in a temporal or tissue-specific manner (Babak et al., 2004; Chen et al., 2004; Rogelj and Giese, 2004; Bari et al., 2006; Sood et al., 2006).

The knowledge that miRNAs are key regulators in numerous developmental pathways (including hormone signalling), and are often expressed in a tissue-specific fashion, suggests a possible role for miRNAs during vascular development and xylogenesis. With the number of new miRNAs being discovered associated with diverse developmental pathways increasing (Sunkar and Zhu, 2004; Lu et al., 2005; Sunkar et al., 2005b; Bari et al., 2006; Mica et al., 2006), the differentiating woody tissues of *Eucalyptus* are a likely candidate for miRNA discovery. The role of miRNAs during wood formation and vascular development to date has not been studied in depth. Preliminary studies have indicated that miRNAs have a predicted role during reaction wood formation in poplar (Lu et al., 2005), and have been correlated with the seasonal and developmental regulation of a class III HD-Zip protein involved in vascular patterning in hybrid aspen (Ko et al., 2006a).

In this study we aimed to isolate miRNAs from three vascular tissues of *Eucalyptus* and determine whether these miRNAs exhibited tissue-specific expression patterns, which would indicate a possible role for these molecules during xylogenesis. We identified forty-

eight putative miRNAs, including hypothetical *Eucalyptus*-specific miRNAs. The functions of the predicted miRNA targets include transcriptional regulation, signal transduction and cellular metabolism. Expression profiling revealed that a number of the putative novel miRNAs are preferentially expressed in vascular tissues. These results indicate that certain identified putative egr-miRNAs could play a role during the tree-specific developmental process of the vascular tissue differentiation in the *Eucalyptus* stem.

3.3 MATERIALS AND METHODS

3.3.1 Plant materials

Three differentiating stem tissues (mature xylem, immature xylem, and immature phloem) and young leaves were collected from a four-year-old *Eucalyptus grandis* tree and were immediately frozen in liquid nitrogen for RNA isolation. Immature xylem, comprising xylem mother cells and early developing xylem tissue was harvested from the outer ca. 1-2 mm glutinous layer of the stem after the removal of the bark. Mature xylem, encompassing xylem cells in advanced stages of maturation, was recovered from 3-5 mm deep planings following the complete removal of the immature xylem layer. Immature phloem, which contained the majority of the vascular cambium and developing phloem cells, was sampled as a 1-2 mm layer from the inner surface of the bark. Young unfolded leaf tissues were taken from the top of the crown where the leaves were still light green in colour.

3.3.2 Small RNA isolation and identification

Total RNA was isolated from all tissues using a modified CTAB RNA extraction method (Chang et al., 1993). Small RNAs were isolated from each sample as previously described in Chapter 2 according to Elbashir et al. (2001) with certain modifications. Small RNA molecules (<200 bp) were separated from total RNA using the *mir*Vana™ miRNA Isolation Kit (Ambion, Austin, TX) as per the manufacturer's instructions. The small RNA fraction

was resolved on a denaturing (6 M Urea) 12% polyacrylamide gel and RNAs in the size range of ~ 17 – 27 nt were excised from the gel using a 10 bp DNA ladder (Promega, Madison, WI) as a reference. The small RNAs were eluted overnight in 0.3 M NaCl at 4°C and recovered by ethanol precipitation at -80°C for more than two hours. Small RNAs were processed as described in Chapter 2, with regards to the ligation of modified 5' and 3' adaptors, cDNA synthesis and PCR amplification. However, from this point, PCR products were resolved on a 12% polyacrylamide gel and size selection in the range of ~ 50 – 70 nt was performed as described above. Size-selected PCR products were blunt-end ligated in a 20 μl reaction, at 22°C for two hours, using 5% w/v PEG, 40 mM Tris-HCl, 10 mM MgCl$_2$, 10 mM DTT, 0.5 mM ATP, 15 U T4 DNA ligase (Fermentas), followed by heat inactivation of the enzyme for 10 min at 65°C. The sample was enriched for ligated products in a 50 μl PCR reaction using adaptor-specific primers overlapping the ligation site with a 5'-adaptor forward primer (5'-TAGGCACCATTCATCGTCGT–3') and a 3'-adaptor reverse primer (5'–CAGGTGCCTCA CGACGATGA–3'). The thermal cycling conditions were as follows: initial denaturation at 95°C for 2 min; followed by 31 cycles of 94°C for 30 sec, 50°C for 30 sec, 72°C for 30 sec, and a final elongation step of 7 min at 72°C. PCR products were resolved on a 2% agarose gel, concatamers ranging from ~ 400 – 800 nt were excised and recovered from the gel slice using the QIAquick® gel extraction kit (QIAGEN, Valencia, CA). The purified concatamers were directly cloned into the pCR2.1-TOPO vector using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA) as per the manufacturer's protocol. Plasmid DNA was extracted from positive colonies using R.E.A.L prep 96-plasmid isolation kit (QIAGEN) as per the manufacturer's instructions. DNA sequencing reactions were performed using the M13 forward primer (5'-CACGACGTTGTAAAACGAC-3') and sequenced in a BigDye cycle sequencing reaction (Version 3.1, Applied Biosystems, Foster City, CA) using an ABI3100 automated DNA sequencer (Applied Biosystems).

Small RNA sequences were identified from the concatamer sequences by manual identification of the known flanking adaptor sequences and parsing of the intervening small RNA sequences. Sequences shorter than 17 nt and longer than 25 nt were excluded from the data set. The remaining small RNA sequences were compared to the GenBank database using BLASTn (http://www.ncbi.nlm.nih.gov, (Altschul et al., 1997)) and the Sanger Institute miRBase miRNA database (http://www.microrna.sanger.ac.uk, Ambros et al., 2003; Griffiths-Jones, 2004). Sequences that matched known rRNA, tRNA, snRNA, and retrotransposons were excluded from further analysis. The remaining putative small RNAs were named with the prefix egr-miR to indicate species, namely *Eucalyptus grandis*, and were given arbitrarily designated numerical suffixes (e.g. egr-miR1). Existing miRNA designations were used in cases when the sequence was identical to known miRNAs.

3.3.3 Prediction of stem-loop structures

The remaining small RNA sequences were used in a search against all available BAC clone shotgun sequences of the *Eucalyptus camaldulensis* genome being generated at the Kazusa DNA Research Institute (Kisarazu, Japan). Small RNA sequences were identified in the genomic sequences using the EMBOSS nucleotide pattern search pattern Fuzznuc (http://bioinfo.hku.hk/EMBOSS) allowing up to three mismatches from the provided small RNA sequence. Identified genomic sequences were then surveyed within 500 nt (when available) upstream and downstream of the small RNA site for sequences with up to six mismatches from the small RNA sequence. These genomic sequences were then used for fold-back secondary structure prediction. This was performed using the mfold program (Zuker, 2003) with default parameters (http://www.bioinfo.rpi.edu/applications/mfold/rna/form1.cgi, version 3.2). Via manual inspection, only those sequences that could form thermodynamically stable stem-loops, and adhered to the criteria for miRNA precursors as described by Jones-Rhoades et al. (2006), were denoted as putative *Eucalyptus* miRNA

precursor sequences. Briefly, the criteria used specify that the 25 nucleotides centered on the mature miRNA in the stem-loop should not contain more than seven unpaired nucleotides, of which no more than three can be consecutive, and no more than two are without a corresponding unpaired nucleotide in the miRNA*. Only the stem-loop structures which adhered to these criteria were considered as candidate miRNAs and used for further analysis (Reinhart et al., 2002).

3.3.4 Quantitative real-time PCR of isolated miRNAs

Sample preparation for real-time PCR of miRNAs was performed using the protocol described by Rui et al (2005). Briefly, total RNA was extracted from xylem, immature xylem, phloem, and leaf tissues as described above. Contaminating DNA was removed by DNaseI digestion in a 20 μl reaction at 37°C for 30 min using 30 U RNase-free DNaseI, Roche, Indianapolis, IN, USA, 80 U RNasin® RNase inhibitor, Promega, 10 mM Tris, 50 mM KCl, pH 8.3. Approximately 1 μg treated total RNA from each tissue was polyadenylated with Poly(A) polymerase (PAP) and ATP using the Poly(A) Tailing Kit (Ambion) at 37°C for 1 hour in a 20 μl reaction as per the manufacturer's protocol and purified with phenol/chloroform extraction followed by ethanol precipitation. The polyadenylated total RNA from each tissue was reverse transcribed into first strand cDNA using 0.5 μg of a polyT-VN anchored adaptor oligonucleotide (Table S3.1) as a primer and 200 U SuperScript™ III reverse transcriptase (Invitrogen) according to the manufacturer's instructions.

In order to validate miRNA expression and to determine expression differences between the four *Eucalyptus* tissues, real-time PCR reactions were performed using the identified miRNA sequence as a forward primer and a poly-T adaptor-specific antisense reverse primer (Table S3.1). Forward primers containing more than 3 G/C base-pairs within the last five bases on the 3' end, were extended with one or two A's to ensure correct binding

and amplification during PCR reactions. The endogenous reference used for normalization of miRNA expression levels was the *Eucalyptus* 5.8S rRNA gene. To obtain an appropriate reference, a poly-T adaptor reverse primer and a gene-specific forward primer were used to generate a PCR product similar in length to the miRNA amplicon. All primers used in real-time PCR are listed in Table S3.1 and were synthesised by Inqaba Biotech (Pretoria, South Africa).

Real-time PCR was performed with the LightCycler 480 instrument (Roche Diagnostics, GmbH) using the LightCycler® 480 SYBR Green I Master kit (Roche) as per the manufacturer's instructions. Briefly, optimal template amount was determined by performing PCR reactions using serial dilutions of concentrated cDNA. Each 10 μl PCR reaction contained 1 μl diluted cDNA, 5 μl 2 X LightCycler® SYBR Green I PCR master mix (Roche), and 0.5 μM each of the forward and reverse primers. The following thermal cycling conditions were used for real-time PCR reactions: enzyme activation and denaturation at 95°C for 10 min followed by 42 cycles of 95°C for 5 s, 60°C for 10 s and 72°C for 5 s. The amplification was followed by melting curve analysis and agarose gel electrophoresis of the RT-PCR products for each miRNA primer, to confirm that they corresponded to single amplification products. All PCR reactions were performed in triplicate and included a no template water control.

For quantification of miRNAs, internal standard curves were created for each miRNA primer using a serial dilution of cDNA, which was PCR amplified, gel extracted and purified using the QIAquick® gel extraction kit (QIAGEN) and the concentration determined using the ND-1000 NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE). Standard curves were generated for each miRNA primer with the LightCycler 480 software (version 1), using the second derivative maximum method. Expression levels were normalized to the 5.8S rRNA reference gene and, for each miRNA, standardized to the tissue

showing the highest expression level (Figure 3.3). Finally, to allow a direct comparison of expression levels of different egr-miRNAs, a measure of the expression level of the 5.8S rRNA gene relative to the expression of the miRNA in the tissue where it is most abundant was calculated (i.e. where 45X indicates that the miRNA is expressed 45 times less than the 5.8S rRNA gene), as shown in parenthesis is Figure 3.3.

3.3.5 Prediction of targets for *Eucalyptus* miRNAs

Since the full *Eucalyptus* genome sequence is not available yet, and limited EST data is available for *Eucalyptus*, all target searches were performed using the *Populus trichocarpa* (http://genome.jgi-psf.org/Poptr1/Poptr1) and *Arabidopsis thaliana* (http://www.arabidopsis. org) annotated gene models. Target predictions for were performed using PatScan (Dsouza et al., 1997) to search for regions of complementarity between the identified miRNA and expressed transcripts within the two genomes analyzed. The number of mismatches allowed between the miRNA and a potential target was determined as described by Jones-Rhodes and Bartel (2004). Target gene function was taken as that described for each gene in the respective database. For genes which were not annotated or with unknown gene function, predicted functions were determined using searches of domains present in the target gene against the EMBL-EBI database of the European Bioinformatics Institute InterPro database (http://www.ebi.ac.uk/interpro) and the *Arabidopsis* transcription factor database (http://arabtfdb .bio.uni-potsdam.de/v1.1).

3.4 RESULTS

3.4.1 Isolation and sequence analysis of small RNAs from *Eucalyptus grandis* stem tissues

In order to identify miRNAs that may be specific to *Eucalyptus* trees, or those with a specialized role in wood formation, three independent small RNA cDNA libraries were constructed from xylem, immature xylem and phloem tissues of a fast-growing eucalypt tree.

Approximately 960 clones were sequenced (320 per tissue), resulting in a total of 1621 individual small RNA sequences. Only unique sequences in the size range of 19 to 24 nt were analyzed further. These sequences were then used in a plant specific (viridiplantae) BLASTn search against the GenBank database (http://www.ncbi.nlm.nih.gov/ BLAST) to identify cloned sequences which were rRNA, tRNA, snRNA and retrotransposon breakdown products, and therefore not true miRNAs. This analysis revealed that 38% of the identified sequences were not true small RNA sequences and they were therefore excluded from further analyses. This resulted in a remainder of 504 small RNA sequences from all three tissues, which represents approximately 31% of the total sequences cloned.

BLASTn analysis was used to determine whether any miRNAs cloned in previous studies were also identified here. A search against the Sanger Institute miRBase miRNA registry release 9.0 at http://www.microrna.sanger.ac.uk (Griffiths-Jones, 2004; Griffiths-Jones et al., 2006), revealed that our small RNA sequences showed similarity to seven known previously published families of miRNAs from *Arabidopsis*, poplar and other plant species (Reinhart et al., 2002; Rhoades et al., 2002; Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004; Lu et al., 2005). These sequences are referred to as known miRNAs in the remainder of the chapter, and have been grouped into their corresponding miRNA families as annotated in other species. These are the miR159, miR160, miR167, miR168, miR172, miR397, and miR472 families.

3.4.2 Identification of putative *Eucalyptus* miRNAs

One of the distinguishing features of true miRNAs is the presence of a miRNA precursor molecule with the ability to form a stable stem-loop structure (pre-miRNAs) in the context of the surrounding genomic sequence (Lee et al., 2002; Yan et al., 2003). Thus, stem-loop structure prediction is an important step during the identification of true miRNAs in the multitude of sequences obtained during small RNA isolation. In order to identify pre-miRNA

molecules, knowledge of the genomic sequence surrounding the mature miRNA is necessary. At the time of this study only a limited amount of DNA sequence for *Eucalyptus* was publicly available, and the majority of these sequences represented ESTs and known gene sequences. The Kazusa DNA Research Institute is in the process of generating a draft genome sequence of the *E. camaldulensis* genome. To identify putative *E. grandis* miRNA precursors, we searched all of the available shotgun sequence data for sequence similarity to the remaining 504 cloned small RNAs. Up to three nucleotide mismatches were allowed in the search as this is the largest number of mismatches observed between miRNA family members to date (reviewed by Jones-Rhoades et al., 2006). Of the 504 sequences searched, 358 did not show similarity to any of the shotgun sequences, and the remaining 146 sequences showed similarity to 5932 genomic shotgun sequences which were considered to be potential stem-loop candidates. For 98 of the 146 sequences, no stem-loop structures adhering to the correct criteria could be identified, and these were therefore excluded from further analysis. These sequences may indeed be true miRNAs, but without complete genomic sequence they cannot be characterized further. The remaining 48 sequences were capable of forming stable stem-loop structures in the context of the surrounding genomic sequences. These 48 sequences were considered to be putative *Eucalyptus* miRNAs (denoted as egr-miRNAs), and were grouped into thirteen different families, as shown in Tables 3.1 and 3.2.

Twenty-eight of the putative miRNA sequences group into eight distinct families and were considered to be putative novel miRNAs, as they did not exhibit sequence similarity to miRNAs in other plant species. These were the egr-miR31, egr-miR90, egr-miR140, egr-miR200, egr-miR293, egr-miR320, egr-miR359 and egr-miR362 families (Table 3.1). The naming of the new egr-miRNAs was based on the order of cloning. The remaining 20 miRNAs belonged to the known miRNA families as described above. However, for the miR167 and miR397 families, no *E. camaldulensis* genomic sequence could be identified and

these families could not be analyzed further. These twenty small RNAs therefore belong to five known miRNA families as represented in Table 3.2, and were subsequently denoted as egr-miR159, egr-miR160, egr-miR168, egr-miR172, and egr-miR472. For the egr-miRNAs belonging to known miRNA families, we adhered to the naming convention assigned by the original authors and miRBase. Cloning frequencies and the *E. camaldulensis* genome reference for each of the egr-miRNAs are included in Tables 3.1 and 3.2, with the small RNA sequences identified through cloning shown in bold. As shown in Tables 3.1 and 3.2, five of the putative egr-miRNA sequences had single copies in the *Eucalyptus* genome, whereas the rest corresponded to multiple (2 to 8) genomic loci.

Precursor sequences and *E. camaldulensis* genome references for all the identified putative miRNAs are summarized in Supplemental Table S3.2, where the mature miRNA sequence is depicted in bold and the miRNA* in italics. The precursor (pre-miRNA) stem-loop structures for all putative novel and known egr-miRNAs are shown in Figures 3.1 and 3.2 respectively, with the mature miRNAs in bold. The pre-miRNA stem-loop structures presented in Figures 3.1 and 3.2 represent only the mature miRNA/loop/miRNA* portion of each precursor, starting at the miRNA sequence itself. Thus, this does not represent the full-length transcribed miRNA precursor molecule, the extent of which is currently unknown.

The number of putative egr-miRNA members per family varied greatly between families. For instance, the egr-miR31 family had only one family member, whereas the egr-miR359 and egr-miR172 families contained 9 and 7 members respectively. Although the family identity and miRNA sequence is highly conserved across species, the number of miRNAs in each family seems to differ between species (miRBase). However, due to the fact that the entire *Eucalyptus* genome sequence was not available, inferences about family size could not be made as there may have been family members that could not be identified yet.

3.4.3 Prediction of putative targets for *Eucalyptus* miRNAs

In contrast to other species, plant miRNAs exhibit near-perfect complementarity to their target genes, which allows reliable computational prediction of miRNA targets (Rhoades et al., 2002; Bonnet et al., 2004; Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004; Wang et al., 2004a; Adai et al., 2005). We used all available *Eucalyptus* genomic and EST sequences for target predictions however could not identify viable matches to the miRNA sequences at this time. It is known that plant miRNAs, and a large proportion of their target genes, are conserved across species, even among a herbaceous plant such as *Arabidopsis* and a woody plant such as poplar (Rhoades et al., 2002; Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004; Wang et al., 2004b; Lu et al., 2005; Jiang et al., 2006; Zhang et al., 2006). We therefore assumed that at least some targets would be conserved between *Eucalyptus*, poplar and *Arabidopsis*. For this reason, hypothetical targets for the putative egr-miRNAs were predicted using the *Populus* and *Arabidopsis* annotated genome sequences. Manual inspection of the mismatch pattern between the miRNA: mRNA duplex of each putative target gene was performed in order to eliminate non-authentic targets. The scoring system for mismatch patterns is described in Chapter 2.

We identified a total of 110 putative target loci from poplar and *Arabidopsis* for the thirteen egr-miRNA families (Table 3.3 and 3.4). For the eight novel miRNA families we identified 45 candidate target genes (Table 3.3), and for the five known egr-miRNA families we identified 65 putative target loci (Table 4), which correlated well with the targets already predicted for these families in other species (Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004; Axtell and Bartel, 2005; Li and Zhang, 2005; Lu et al., 2005). Each miRNA family identified had a number of putative targets, ranging from one to twenty (Tables 3.3 and 3.4). Since it is not possible to predict if a specific miRNA family member regulates a

specific target gene Tables 3.3 and 3.4 provide a list of all possible target genes of an entire miRNA family.

Of the predicted targets, ~ 40% belonged to transcription factor families or function in transcriptional activation or regulation. A further ~ 15% of the targets had functions in defence response against pathogen infection. Approximately 11% of the targets were predicted to function in the metabolism of carbohydrates, lipids and proteins, and a further ~ 23% were of unknown function. The remainder of the targets had various functions, including signal transduction, calcium binding and electron transport. The predicted targets of the putative novel miRNAs appear to cover a more diverse range of targets, and fewer members of the same gene families, than the known miRNAs. This could be due to the fact that perhaps the targets of these miRNAs in *Eucalyptus*, are not targeted by miRNAs in these species, and the target sites of the miRNAs are not conserved. Thus, either *Eucalyptus* could have gained novel miRNAs, or they were ancestral but subsequently lost in *Arabidopsis* and poplar. The predicted targets of the known egr-miRNAs have been validated as true targets of these families in poplar and *Arabidopsis*, and thus it is not surprising that no new targets were identified for these families. Once the *Eucalyptus* genome sequence is available, target predictions for these miRNAs within the *Eucalyptus* genome may yield targets specific to this tree species. For the purpose of this study inferences about egr-miRNA function can only be made from the poplar and *Arabidopsis* sequences available, and further experimental validation of these targets within *Eucalyptus* is necessary for verification of the egr-miRNAs, as well as for establishing the precise functional role of the putative miRNAs in eucalypt trees.

3.4.4 Expression profiles of putative egr-miRNAs in *Eucalyptus*

To determine whether the identified putative egr-miRNAs exhibit tissue-specific expression patterns, and to deduce a possible functional role for these miRNAs during vascular

development, expression profiling was performed on all identified egr-miRNA families. Many miRNAs have been found to be differentially expressed at different developmental stages and in different tissues in a range of organisms (Horvath et al., 2003; Kidner and Martienssen, 2004; Robinson et al., 2004; Lu et al., 2005; Ko et al., 2006a). Previous studies have indicated that although the use of RNA gel blots to determine miRNA expression is a useful tool, many miRNAs expressed at low levels cannot be detected with this method (Sunkar and Zhu, 2004; Lu et al., 2005). Thus, we employed qRT-PCR in order to detect miRNA expression levels using the Roche LightCycler 480 instrument, which allows simultaneous analysis of up to 384 samples. This method has been shown to be a more powerful tool to quantitatively assess miRNA expression, and we used the technique developed by Shi et al. (2005) in order to obtain reliable expression patterns.

The expression levels of thirteen putative egr-miRNA families were determined by qRT-PCR in three secondary tissues and one primary tissue of *Eucalyptus grandis* (Figure 3.3). We designed primers based on the cloned egr-miRNA sequence for each family. However, we tested if different members of the same egr-miRNA family produce similar expression patterns across the four tissues. In cases were family members exhibited sequence differences from each other, two primers per family were used to determine if these sequence differences could alter PCR amplification efficiency and subsequently expression patterns. As shown for egr-miR200a and egr-miR200e (Figure 3.3), these slight sequence changes did not alter the overall expression pattern of this miRNA.

We verified the expression of all thirteen miRNAs tested in this study and found expression levels ranging from one to three orders of magnitude lower than that of the 5.8S rRNA gene. The expression levels presented in Figure 3.3 represent the contribution of an entire miRNA family and although specific family members could exhibit tissue-specific expression patterns, these could not be quantified using our approach. All of the putative egr-

miRNAs have at least a base level of expression in all of the tissues analyzed, with the exception of egr-miR359, which appeared to be specific to phloem. Several of the egr-miRNAs were strongly expressed in leaf tissues, namely egr-miR31, egr-miR90, egr-miR200, egr-miR320 and egr-miR168. Egr-miR31, egr-miR90, and egr-miR200 (a and e) exhibited similar expression patterns, showing highest expression in leaves and relatively similar levels in the vascular tissues, although egr-miR31 was expressed approximately 41 to 78 times less than these miRNAs. Egr-miR140 was expressed at similar levels in xylem, phloem and leaf tissues, and exhibited a ~ 5 fold lower expression in the immature xylem. Egr-miR293 and egr-miR160 were expressed at similar levels in all the vascular tissues, but occurred at concentrations of ~ 10 fold lower in the leaf tissue. Egr-miR320 was expressed at similar levels in the leaf and xylem tissues, but lowest in the immature xylem. Egr-miR359 was expressed very highly in phloem tissues and at almost undetectable levels in the other tissues. Egr-miR362, egr-miR168 and egr-miR172 appeared to be ubiquitously expressed in all tissues analyzed. Egr-miR159 was expressed at similar levels in xylem and phloem, and at low levels in the immature xylem and leaf. Egr-miR472 was expressed at the highest level in the phloem and leaf tissues, and at 5-fold lower levels in the xylem and immature xylem tissues. The expression levels of the known miRNAs (Figure 3.3B) correlated well with the expression patterns observed in vascular tissues of poplar (Lu et al., 2005). Furthermore, we observed a vast difference in expression levels between different egr-miRNA families, where expression ranged from one to four orders of magnitude less than that of the 5.8S rRNA gene.

3.5 DISCUSSION

An ever increasing understanding of the diverse functions of miRNAs has allowed novel insights into the regulation of complex developmental processes and the specification of specialized cell types. Xylogenesis is an important developmental process in trees that leads to the formation of wood, which is one of the most important natural resources with a

multitude of applications. Understanding the molecular mechanisms involved in the control of wood formation would allow the manipulation of wood properties. However, little is know about the molecular mechanisms regulating wood formation genes. It has been suggested that miRNAs could play a functional role in wood formation and vascular development (McHale and Koning, 2004; Lu et al., 2005; Ko et al., 2006a). However these studies were focused on reaction wood in response to mechanical stress and to a single miRNA. Thus, the identification of miRNAs regulating normal wood formation and vascular differentiation, as well as isolation of possibly species-specific miRNAs, is of great importance.

By sequencing libraries of small RNAs from three differentiating vascular tissues of *Eucalyptus*, we were able to identify 48 putative miRNAs, of which 28 have not yet been identified in other plant species. Target predictions revealed 45 putative targets for these miRNAs in the poplar and *Arabidopsis* genomes. Furthermore, expression profiling of the putative novel miRNAs across a number of xylogenic and non-xylogenic tissues indicated that certain of these miRNAs could possibly have specialized functions during the control of vascular differentiation, and subsequently wood formation, in eucalypt trees. We used a combination of small RNA library sequencing and computational analysis to identify putative *Eucalyptus* miRNAs. The cloning of small RNAs is useful in the identification of novel miRNAs that are species- or tissue-specific, and those that are expressed in response to external stimuli. However, a large number of sequences need to be generated in order to identify a large number of miRNA for a specific study. Computational identification of miRNAs through extensive genome searches allows the identification of further family members of a specific miRNA family, but does not allow the identification of novel miRNAs. Deeper sequencing of libraries from *Eucalyptus* vascular and other non-vascular tissues will reveal further miRNAs not identified here. Once the *Eucalyptus* genome is fully sequenced and annotated, the remainder of the small RNAs cloned in this study (which could not be

identified in the available genomic sequences to date) could be identified as further miRNAs. Furthermore, this would allow experimental validation of the putative egr-miRNAs, and their predicted target genes, which could not be performed in this study due to lack of *Eucalyptus* sequence data.

3.5.1 Characteristics of putative *Eucalyptus* miRNAs

The putative egr-miRNAs identified in this study conform to certain characteristics found to be common in other plant species, including similarities with regards to sequence composition, length, and location of the miRNA on the pre-miRNA molecule. Many plant miRNAs possess a uracil at their 5' end, and although this is not a requirement for a true miRNA, it is an observed trend. Of the putative novel miRNAs, approximately 61% begin with a 5' uracil. The known egr-miRNA sequences are highly conserved among all other plant species, with some possessing identical sequences to their orthologues, and others showing a maximum of up to three nucleotide differences from evolutionary distant monocot species, such as rice (miRBase). All of the putative egr-miRNAs identified here were between 20 to 23 nt in length, another trend seen in all know plant miRNAs (reviewed by Reinhart et al., 2002; Jones-Rhoades et al., 2006; Meyers et al., 2006). As discussed in Chapter 1, the mature miRNA molecule can be derived from either the 5' or 3' arm of the precursor molecule. Studies of plant miRNAs in other species have shown that all of the members within a single miRNA family derive their mature miRNAs from the same arm of the stem-loop (Reinhart et al., 2002; Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004; Wang et al., 2004a; Lu et al., 2005). Our results are consistent with this finding (Table 3.1 and 3.2). Furthermore, for the *Eucalyptus* miRNAs belonging to previously identified families, the mature miRNAs is located on the same arm as identified in other species, suggesting that these loci share a common ancestry (Reinhart et al., 2002; Lu et al., 2005).

It may seem that the known miRNAs identified in this study (Table 3.2) are overrepresented in our library when compared to results found in other studies. For example, egr-miR160 was represented 13 times in our small RNA libraries. The fact that the known egr-miRNAs seem to be overrepresented in our libraries cannot be used to make any inferences about their expression patterns over the three tissues. It could simply reflect that these miRNAs are more abundantly expressed than the novel miRNAs in *Eucalyptus*. Since these miRNAs were some of the first ever to be identified, suggesting a higher overall expression level (Llave et al., 2002b; Mette et al., 2002; Park et al., 2002; Reinhart et al., 2002), this seems a likely conclusion. For the putative novel egr-miRNAs, the cloning frequencies were more similar to those obtained in other studies, where all sequences were cloned once with the exception of egr-miR90, which is represented twice. In general, it seems that the putative novel miRNAs have lower cloning frequencies in small RNA libraries as these could be either species- or tissue-specific, or because they are expressed at very low levels, and thus have not been identified before this study (Bonnet et al., 2004; Lu et al., 2005; Sunkar et al., 2005b; Sunkar et al., 2005a).

3.5.2 Predicted targets of egr-miRNAs have diverse biological functions

A large proportion of previously identified miRNA targets in *Arabidopsis*, poplar and rice are transcription factors (Rhoades et al., 2002; Bartel and Bartel, 2003; Jones-Rhoades and Bartel, 2004; Mica et al., 2006). However, as more miRNAs were discovered, especially those from specialized tissues, or those expressed in response to various stimuli, it became clear that miRNAs regulate more diverse targets than initially expected (Bonnet et al., 2004; Sunkar and Zhu, 2004; Wang et al., 2004a; Adai et al., 2005; Lu et al., 2005). We were able to predict 45 genes (Table 3.3) that are likely targets of the putative novel egr-miRNAs. The predicted targets appear to be involved in a diverse range of biological processes. Of the predicted targets, ~ 16% are involved in transcriptional regulation, ~ 18 % are involved in signal

transduction, ~ 27% are involved in cellular metabolic processes and a further ~ 40% have no defined function (Table 3.3). These unknown targets are prime candidates for further analysis as they may represent *Eucalyptus*-specific genes, and could be part of novel gene regulatory pathways specific to this woody plant. For the known egr-miRNAs, the targets have been characterized in *Arabidopsis* and poplar in previous studies, and it is not surprising that no novel targets were identified. These known target genes have been discussed in detail in Chapters 1 and 2 and will not be discussed to great length here. The putative novel egr-miRNAs are predicted to target a large number of genes, with seemingly unrelated functions. This is due to the fact that these targets were identified purely through computational prediction and, until these targets are experimentally validated, we can not deduce whether these are true targets. Thus, we will discuss all putative targets for the novel egr-miRNAs.

Although miRNAs have diverse targets, additional targets with roles in transcriptional regulation are still being discovered. A CCAAT-binding factor (CBF) subunit C transcription factor (Li et al., 1992) is a predicted target of egr-miR31. A related CCAAT-binding factor is a confirmed miRNA target in rice and *Arabidopsis* (Rhoades et al., 2002; Mica et al., 2006). In *Arabidopsis*, many of the CCAAT-binding transcription factors are involved in embryogenesis initiation and development (Li et al., 1992; West et al., 1994; Parcy et al., 1997; Lotan et al., 1998), and certain members exhibit tissue-specific expression patterns (Gusmaroli et al., 2002). A putative target of egr-miR140 is a LIM-domain containing protein, where proteins containing this domain participate in important cellular processes in eukaryotes. They are thought to be involved in protein: protein interactions, and could be important for the assembly or activity of multicomponent complexes involved in transcriptional activation or repression (Baltz et al., 1992; Michelsen et al., 1993; Eliasson et al., 2000). Of the known egr-miRNAs, egr-miR159 and egr-miR172 target multiple members of the MYB and APETALA2 transcription factor families. The coordinate regulation of gene

expression is required during a large number of plant developmental processes, and transcriptional regulation by miRNAs is an essential part of plant development.

Signalling networks allow organisms to respond to external stimuli, such as environmental changes, hormones, organic and inorganic compound levels. A predicted target of egr-miR31 is a putative ethylene-responsive calmodulin-binding protein (CBP), which was identified as a target in this study in both poplar and *Arabidopsis*. *Arabidopsis* CBPs are predicted to function in multiple calcium-mediated signalling networks in plants (Reddy et al., 2002; Yang and Poovaiah, 2002). CBPs are rapidly and differentially induced by signals such as temperature extremes, UVB (light signal perception), salt, and wounding, hormones, and signal molecules (Yang and Poovaiah, 2002). Comparative transcriptome analysis of poplar identified a calmodulin-binding family protein in a xylem-specific gene set (Ko et al., 2006b), indicating that this target could in fact be involved in vascular development in woody plants.

Egr-miR362 is predicted to target a calcium-binding calmodulin, with a role in mediating plant calcium signalling, where in *Arabidopsis* these genes encode potential calcium sensors (McCormack et al., 2005). Egr-miR140 is targets three putative small transport proteins thought to be involved in signal transduction, including the regulation of phosphate transport (Spain et al., 1995; Lenburg and O'Shea, 1996; Lee et al., 2000). In the last few years, miRNAs have been shown to regulate hormone, sulphur and phosphate levels in other plant species (Fu and Harberd, 2003; Achard et al., 2004; Jones-Rhoades and Bartel, 2004; Aung et al., 2006; Bari et al., 2006). The newly identified putative egr-miRNAs seemingly play a similar role in *Eucalyptus*, especially in the major transport tissues of this tree - the xylem and phloem.

In *Arabidopsis*, egr-miR31 is predicted to target a component of the SCAR/ WAVE complex involved in the regulation trichome morphogenesis (Zhang et al., 2005). This

complex regulates actin polymerization/depolymerization and function in cellular morphogenesis pathways, where it plays an important role in regulating cell division, cell expansion, and cell shape (Szymanski, 2004; Zhang et al., 2005). A key step during xylogenesis is cell division followed by cell expansion and ultimately differentiation into mature secondary cell wall tissues, each of which has a specific cell shape, such as the greatly elongated fibre cells. Thus, this putative miRNA could play a role in differentiation and cell fate determination in *Eucalyptus*.

In this study we identified a number of predicted targets that are thought to function in normal cellular and metabolic processes. These targets include a pectinase (egr-miR31) involved in carbohydrate metabolism and the ripening of fruit (Huang and Schell, 1990; Ruttkowski et al., 1990; Davies and Henrissat, 1995), two proteolytic enzymes: a cysteine protease (egr-miR31) and a peptidase (egr-miR140), a patatin storage protein (egr-miR31) involved in lipid metabolism (Banfalvi et al., 1994), a protein kinase (egr-miR293), a predicted phosphate acyltransferase (egr-miR320), and a PAPS reductase/FAD synthetase (egr-miR293) involved in molybdopterin cofactor biosynthesis. Egr-miR200 has only one target with a predicted function, a pentatricopeptide repeat (PPR) protein, which is one of the largest plant protein families and is involved in plant organelle RNA metabolism and processing (Lurin et al., 2004; Nakamura et al., 2004). Egr-miR359 is predicted to target an iron/ascorbate family oxidoreductase with an isopenicillin N synthase (IPNS) domain. In plants, proteins with this domain are involved in the formation of flavonoids, catechins, proanthocyanidins, anthocyanidins and ethylene (Zhang et al., 1997; Lukacin et al., 2000). These predicted target genes of egr-miRNAs appear to function in diverse cellular processes, notably secondary metabolism, and these results could indicate that *Eucalyptus* miRNAs are involved in a more diverse range of processes than initially expected.

3.5.3 *Eucalyptus* miRNAs could have a putative role in vascular development

Expression analysis (Figure 3.3) indicated that the different putative egr-miRNA families analyzed in this study exhibit different tissue-specific expression patterns among the different tissues and differ in their expression levels. This could indicate that the egr-miRNA families are involved in different aspects of development in the vascular tissues of *Eucalyptus*. For example, egr-miR359 could only be detected in phloem tissues, whereas the remaining egr-miRNAs could be detected in all of the tissues, although at different levels.

Some of the putative egr-miRNAs could be associated with more specific aspects of vascular development. Egr-miR90 is predicted to target MADS-box transcription factor involved in various aspects of development (Rounsley et al., 1995; de Folter et al., 2005). MADS-box genes predominantly regulate different steps of flower development, including the specification of the identity of floral organ cell-types, floral meristems, and floral organs, and in the timing of flowering (Jack, 2001). However, the functions of these genes is not restricted to the control of flower development (Riechmann and Meyerowitz, 1997), and they have recently been shown to play a role in regulating primary and secondary xylem formation (Cseke et al., 2003). Egr-miR90 is expressed at high levels in leaf tissues, and relatively ubiquitously in the vascular tissues. This could indicate a role for this miRNA in suppressing genes involved specifically in vascular differentiation in other tissues, such as the leaf.

Egr-miR359 is to be predominantly expressed in phloem tissues. This miRNA is predicted to target two AT-rich interaction domain (ARID)-containing DNA-binding proteins involved in the regulation of developmental and/or tissue-specific gene expression (Kortschak et al., 2000; Riechmann et al., 2000; Wilsker et al., 2002). ARID-encoding genes are involved in a variety of biological processes including cell cycle control, embryonic development, and cell lineage gene regulation (Wilsker et al., 2002). Further analysis of this miRNA in *Eucalyptus* could reveal a role in the control of vascular cambium differentiation,

where it could specify organ boundaries, much like miR165/166 act in specifying leaf fate in *Arabidopsis* (Mallory et al., 2004; McHale and Koning, 2004; Timmermans et al., 2004).

Egr-mir359 is also predicted to target two multidrug/pheromone exporter proteins, belonging to the ATP-binding cassette (ABC) superfamily. These proteins are responsible for the controlled efflux and influx of substrates across cellular membranes, and are thought to function in cellular detoxification, cell-to-cell signalling, plant growth and development (reviewed by Martinoia et al., 2002). In *Arabidopsis*, members of the multidrug resistance P-glycoprotein (MDR/PGP) subfamily of ABC transporters function in cellular and long-distance auxin transport from the root tip of the plant (Terasaka et al., 2005; Geisler and Murphy, 2006). Each phase of wood formation is regulated by the interaction of differentiating cells with global impulses such as hormone signalling. Auxin is thought to be an important factor in the regulation of xylogenesis (reviewed by Friml, 2003) and Schrader et al. (2003) described genes involved in the polar transport of auxin in the developing vascular tissues, where the development of an of an auxin gradient seemed to be involved in governing stages of vascular cambium differentiation in the poplar stem. Auxin responsive transcription factors (ARFs) are also known to be required for correct vascular patterning in *Arabidopsis* (Hardtke and Berleth, 1998). Egr-miR160 is predicted to target 12 different ARF genes from poplar and *Arabidopsis* (Table 3.4), where it also exhibits higher expression levels in phloem compared to other tissues. This miRNA has been shown to be involved in whole plant development (Mallory et al., 2005) and root cap formation (Wang et al., 2005) in *Arabidopsis*. Furthermore, auxin signaling in plants is modulated by ubiquitination and degradation of the Aux/IAA transcriptional repressors through the action of an ubiquitin protein ligase, where this protein mediates Aux/IAA degradation and auxin-regulated transcription (Dharmasiri et al., 2005). Enzymes involved in this ubiquitination are miRNA targets in *Arabidopsis* (Sunkar and Zhu, 2004), and here we identified a ubiquitin-protein ligase involved in

ubiquitination and ubiquitin-dependant protein catabolism (Bates and Vierstra, 1999) as a putative target for egr-miR90. These results suggest that miRNAs described above could play a role in controlling auxin levels in the developing stem in *Eucalyptus*.

Egr-miR293 and Egr-miR472 together target 16 putative disease resistance proteins. Although their expression profiles were not identical (Figure 3.3), where egr-miR293 is expressed at similar levels in the vascular tissues, and egr-miR472 is highly expressed in the leaf and phloem tissues, together these two putative miRNAs could play a role in complex defence pathways in *Eucalyptus* in order to ensure plant survival under exposure to pathogens.

Egr-miR362 is predicted to target a Ras GTPase family protein, involved in intracellular signal transduction through cell membrane association (Hancock, 2003). Members of this family are thought to be involved in the regulation of haematopoietic cells, with roles in growth, survival, differentiation, cytokine production, chemotaxis, vesicle-trafficking, and phagocytosis (Reuther and Der, 2000; Ehrhardt et al., 2002). The expression pattern of this miRNA is interesting as it is expressed at higher levels in the differentiated tissues, xylem, phloem and leaf, but exhibits lower levels in the less differentiated immature xylem. These results indicate a role for this miRNA, and its target genes, in early developmental events in the vascular cambium of *Eucalyptus*.

Studies of miRNAs in *Arabidopsis* have been essential in understanding a number of plant developmental processes. However, in order to understand tree-specific processes, miRNAs need to be identified from these woody plants. This study has provided some insights into the possible role of miRNAs during the tree-specific developmental process of wood formation. A deeper sequencing of miRNAs from xylogenic and other non-xylogenic tissues from *Eucalyptus* will allow the identification of many more miRNAs from this tree species. Furthermore, once the *Eucalyptus* genome sequence is made publicly available, validation of these miRNAs and their targets by cleavage-assays and knock-out studies in

Arabidopsis can be performed in order to deduce the precise role for miRNAs during xylogenesis in *Eucalyptus*.

3.6 ACKNOWLEDGEMENTS

3.7 FIGURES

egr-miR31 UUACAAUCCAGUACAUCAGACU

```
              C                                            ACCA
5' AAUUACAAUC AGUACAUCAGACUCUACGCAAUCUUAGAGAUUUCAUAUG          \
   ||||||||||  |||||||||||||||||||||||||||||||||||||||          C
3' UUAAUGUUAG  UCAUGUAGUCUGAGAUGCGUUAGGAUCUCUAAAGUAUAC          /
              A                                            AAAA
```

----

egr-miR90 AGGUUGGUGCUUGGCUUUGG

```
                UU       A    CA                        C         GC    A-      G
5' AGGUUGGUGCUUGGC  UGGCCUAGU AGGC  AGCCUCGCCGGCCGUUGGG GAGCUCG  CUC   CCG    A
   ||||||||||||||||  ||||||||| ||||  |||||||||||||||||||| ||||||||  |||   |||
3' UCCAACCACGAACCG  ACCGGAUCA UCCG  UCGGAGCGGCCGGCGACCC CUCGAGU  GAG   GGU    U
                CC       C    AC                        A         UA    CG    C
```

----

egr-miR140 AUUGGCACAAGUGCAAUAGA

```
                                           _____
       CAC          A            /          \
5' AUUGG    AAGUGCAAUAG UUUAAGAAUUUUUUUUUGG       82 nt
   |||||    ||||||||||| ||||||||||||||||||||    COMPLEX
3' UAACC    UUUAUGUUAUC AAA-UCUUGAAAAAAAUU        LOOP
     ACA          C                       _____/
```

----

egr-miR200a CGCACCCUCGCCGGCCACAAGCGA

```
                                    C    CAGA
                              -CGGCCC CGCC    \
                            A |||||| ||||      |
                            / GCCGGG GCGG      /
                           G  /       A   ACCU
       G    CA                /     A
5' CGCACCCUCGCCC GCCA AGCGAGGGC        A            CGUU
   ||||||||||||| |||| ||||||||||                    A-CCCUCGCC      \
3' GCGUGGGAGCGG CGGU  UCGCUCCCG                     ||||||||      |
       G     AC          \        C-GGGAGCGG      /
                          UCGGCG              CCGG
```

egr-miR200b CGCACCCUCGUUGGCCACAAGCGA

```
   C  A                   A      CAA  C    CAA                    AAA       CGU
5' GC CCCUCGUUGGCCACA GCGAGGGCGU   CC CCGCC  AUCCAGGCGAGGGC-C-GA   CCUUGCC    U
   || |||||||||||||||| ||||||||||   || |||||  |||||||||||||| | ||  ||||||||   |
3' CG GGGAGCAGCUGGUGU CGCUCCCGCA   GG GGCGG  UAGGUCCGCUCCCGUGUCU   GGAGCGG    G
   A  C                 C          ACC  U    AUC                   CCG       CCG
```

151

egr-miR200c CGCACCCUCGCCGGCCACAAGCGA

```
                                  30 nt
                                  LOOP
     C           CG        C      \  /   C  CAU          CCU
5' GCACCCUCGC  GCCACAAG GAGGGC-GUC  ACCCU GCU  UGGCCGGCAAGGG    C
   ||||||||||  |||||||| |||||| |||  ||||| |||  |||||||||||||    |
3' CGUGGGAGCG  UGGUGUUC CUCCCGACGGCGUGGGA CGG  ACCGGCUGUUCCC    U
    A           UG       A                A  AC-            ACG
```

egr-miR200d CACGCCCUCAUCGGCCACAAGCGA

```
      A       A  G    A     GG  GA-  GC  CC-   AGA
5' C CGCCCUC UC GCCAC AGCGA GU  CG   CCC  CU     \
   | |||||||| || ||||| |||||  ||  ||  |||   ||     U
3' G GUGGGAG AG CGGUG UCGUU  CG  GC   GGG  GA     /
     C       C  G     C     AA  ACG  AA   AGC  GUC
```

egr-miR200e UGCACCCUCAUCGACCACAAGCGA

```
           AU  G     A            GCC
5' UGCACCCUC  C ACCAC AGCGAGGGC      \
   |||||||||  | ||||| ||||||||||     |
3' GCGUGGGAG  G UGGUG UCGCUCCCG      /
           CG G     C          ACG
```

----

egr-miR293a GGUCGGCCUCGCCAUGGCUGG

```
      C                              U
5' GGU GGCCUCGCCAUGGCUGGGUGAGC    U
   ||| |||||||||||||||||||||||
3' CCA CCGGAGCGGUACCGACCCGCUCG    G
     C                        A
```

egr-miR293b GGUCGGCCUCGCCAUGGCUGG

```
                           GCU
5' GGUCGGCCUCGCCAUGGCUGGGCGA    C
   ||||| ||||||||||||||||||||
3' CCGGC-GGAGCGGUACCGACCCGCU    G
                           GGA
```

egr-miR293c GGUCGGCCUCGCCAUGGCUGG

```
     GG      C                       U
5'   UCGGCCU GCCAUGGCUGGGCGAGC    U
     |||||||| |||||||||||||||||
3'   AGCUGGA CGGUACCGACCCGCUCG    G
   CA       A                     A
```

egr-miR293d GGUCGGCCUCGCCAUGGCUGG

```
     C                          A  U
5' GGU GGCCUCGCCAUGGCUGGGCG GC    C
   ||| ||||||||||||||||||||| ||
3' CCA CCGGAGCGGUACCGGCCCGC CG    G
     A                       C  A
```

152

egr-miR293e GGUCGGCCUCGCCAUGGCUAG

```
        C              A        U
5' GGU GGCCUCGCCAUGGCU GGCGAGC   C
   |||  |||||||||||||||  |||||||
3' CCA CCGGAGCGGUACCGA CCGCUCG   G
     U              C        A
```

----

egr-miR320a UAAGUCGCGAAUUGGUAUGGGCU

```
        C                                        G    A      A  C    /‾‾‾‾‾‾‾\
5'UAAGU GCGAAUUGGUAUGGGCUUGGGAGAGCUCGAGUUAAGU UAAA ACUUAGU GU CAUUAAC     56 nt
  |||||  |||||||||||||||||||||||||||||||| ||||||||| ||||  ||||||| || |||||||      |
3'AUUCA CGUUUAACCAUACCCGAACCCUUUCGA-CUCGAUUUA AUUU UGAAUCA CA GUAAUUG    LOOP
     A                                      A    C      C  A    _____/
```

egr-miR320b UAAGUCGCGAAUUGGUAUGGGCU

```
        G          A            C    A      A   /‾‾‾‾‾‾‾\
5' UAAGUCGCGAAUU GUAUGGGCUUG GAGAGCUCGAGCUAAG UUAAA ACUUAGUA UCCAUUA     174 nt
   ||||||||||||| ||||||||||| ||||||||||||||| ||||| ||||||||| ||||||| |
3' AUUCAGCGCUUAA CAUACCCGAAC CUCUCGGGCUUGAUUC AAUUU UGAAUCAU AGGUAAU    LOOP
        A          C            A    C      C   _____/
```

egr-miR320c UAAGUCUCGAAUUGGUAUGGGCU

```
     UC            A   C       UAAA          /‾‾‾‾‾‾‾\
5' UAAGUC GAAUUGGUAUGGGCUUGGGA AGCC GAGCUAAGUU   CUUAGUAGUCCAUUAGCC   191 nt
   ||||||  |||||||||||||||||| |||| |||||||||   |||||||||||||||||||| |
3' AUUCAG CUUAACCAUACCCGAACCCU UCGG CUUGAUUCAA   GAAUCAUUAGGUAAUCGG   LOOP
     CA            C   A       CUCC          _____/
```

egr-miR320d UAAGUCGUGAAUUGGUAUGGGCU

```
        GU  A    A        GA
5' UAAGUC  GA UUGGU UGGGCUU   \
   ||||||  || ||||| |||||||    |
3' AUUCAG  CU AAUCG ACCCGAG   /
        AU  A    A        AG
```

egr-miR320e UAAGUCGCGAAUUAGUAUGGGCU

```
        GC  A    A        G
5' UAAGUC  GA UUAGU UGGGCUU   G
   ||||||  || ||||| |||||||    |
3' AUUCGG  CU AAUUA ACCCGAG   G
        AU  A    A        A
```

egr-miR320f UAAGUCGAGAAUUGGUACGGGCU

```
        G   A    A        U
5' UAAGUC AGA UUGGU CGGGCUU   G
   ||||||  ||| ||||| |||||||    |
3' AUUCAG UUU AAUCA GCUCGAG   G
        A   G    A        A
```

153

egr-miR359a UCUGUUUAAAAAUUGUUCCAGG

```
     UC                   AG              AAA
5'   UGUUUAAAAAUUGUUCC   G GAAUAGAAAUA       A
     ||||||||||||||||     ||||||||||||
3'   ACAAAUUUUUAACAAGG   CCUUGUCUUUAU         A
   AA                  GG            AAA
```

egr-miR359b UCUGUUUAAAAAUUGUUCCAGG

```
     UC                 A       A     GAA
5'   UGUUUAAAAAUUGUUCC GG GAACA AAAUA      A
     ||||||||||||||||   ||||||| |||||
3'   ACAAGUUUUUAACAAGG CCCUUGU UUUGU       A
   AA                A      C    GAA
```

egr-miR359c UCUGUUUAAAAAUUGUUCCAGG

```
     UC                              A
5'   UGUUUAAAAAUUGUUCCAGG GAACAGAA     U
     |||||||||||||||||||||||||||||||
3'   ACAAGUUUUUAACAAGGUCCCUUGUCUU      A
   AA                           C
```

egr-miR359d UCUGUUCAAAAAUUGUUCCAGG

```
       U                 AG    C     GAA
5' UCUG UCAAAAAUUG-UUCC  G GAA AGAAAUA     A
   |||| ||||||||||| ||||   |||| |||||||
3' AGAC AGUUUUUAACAAAGG   UCUU UCUUUGU     A
     C                G-    A     GAA
```

egr-miR359e UCUGUUCAAAAAUUGUUCCAGG

```
       U                 AG              AAAA
5' UCUG UCAAAAAUUGUUCC   G GAACAGAAAUA         \
   |||| ||||||||||||||     ||||||||||||         A
3' AGAC AGUUUUUAACAAGG   CCUUGUCUUUGU          /
     C                  AA          GUAA
```

egr-miR359f UCUGUUCAAAAAUUGUUCCAGG

```
     UC   U               AG              AAA
5'   UG UCAAAAAUUGUUCC   G GAACAGAAAUA         \
     || ||||||||||||||     ||||||||||||         A
3'   AC AGUUUUUAACAAGG   CCUUGUUUUUGU          /
   AA   C                GG            AAA
```

egr-miR359g UCUGUUCAAAAAUUGUUCCAGG

```
     UC                   AG      G       AA
5'   UGUUCAAAAAUUGUUCC   G GAACA AAAUAG      A
     |||||||||||||||||     |||||| ||||||
3'   ACAAGUUUUUGACAAGG   CCUUGU UUUAUC       A
   AA                    AG     A      AA
```

egr-miR359h UCUGUUCAAAAAUUGUUCCAGG

```
     UC                   AG              AAA
5'   UGUUCAAAAAUUGUUCC   GGAACAGAAAUAG       \
     ||||||||||||||||||   ||||||||||||        A
3'   ACAAGUUUUUGACAAGG   CCUUGUCUUUAUU        /
     AA                  GG              AAA
```

----

egr-miR362 UCUUGGAACUCAACAUCCUGAA

```
                                                                   _____
             A  A           A         U  U       U      /          \
5' UCUUGGAACUC AC UCCUGAAGUAACUU AGUUCAUAAUCCUU AC ACAGUUGCA GGAAU       43 nt
   |||||||||||| || |||||||||||||| |||||||||||||| || ||||||||| |||||        |
3' AGAACCUUGAG UG AGGAUUUCAUUGAA UCAAGUAUUAGGAA UG UGUCA-CGU CCUUG       LOOP
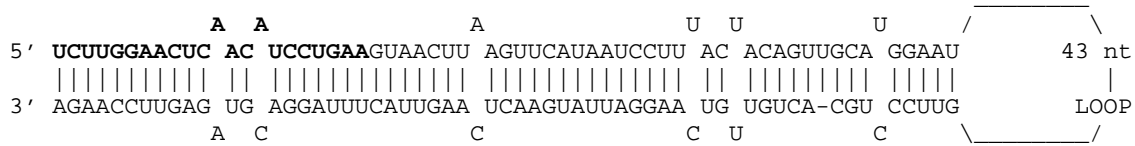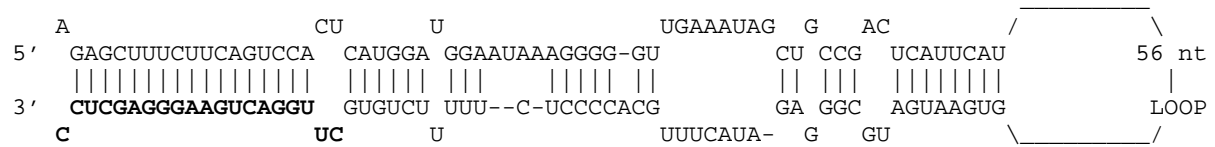             A  C           C         C  U       C      _____/
```

Figure 3.1. Predicted precursor stem-loop structures of putative novel *Eucalyptus grandis* egr-miRNAs. Precursor structures are presented as determined by computational analysis using the mfold program. Mature cloned miRNA sequences are shown in bold. MiRNA families are separated by dashed lines. The sequences shown do not represent only the shortest possible stem-loop structure, where only that portion of the stem-loop starting at the cloned sequence and ending at the complementary miRNA* sequence is shown.

egr-miR159a CUUGGACUGAAGGGAGCUCC

```
                                                                            _____
     A                   CU     U                UGAAAUAG  G   AC         /            \
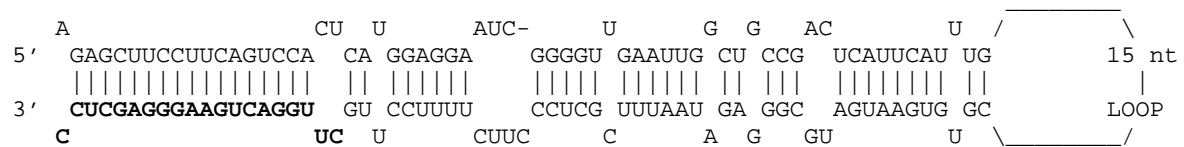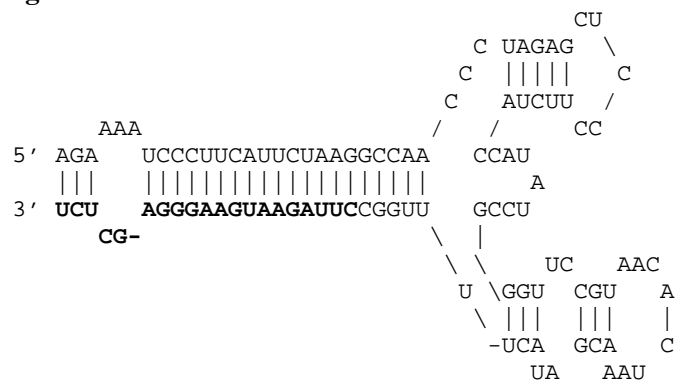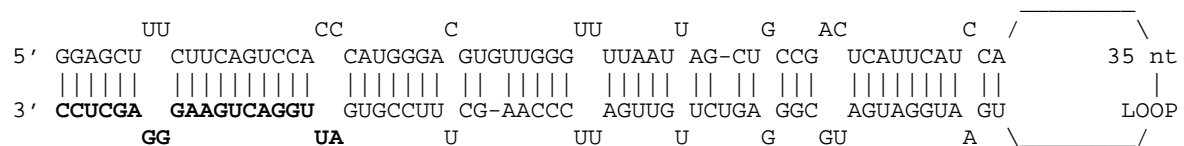5'  GAGCUUUCUUCAGUCCA  CAUGGA  GGAAUAAAGGGG-GU            CU CCG  UCAUUCAU       56 nt
    |||||||||||||||||  ||||||  |||         |||||| ||              || |||  ||||||||||       |
3'  CUCGAGGGAAGUCAGGU  GUGUCU  UUU--C-UCCCACG            GA GGC  AGUAAGUG       LOOP
     C                   UC     U                UUUCAUA-  G   GU         _____/
```

egr-miR159b CUUGGACUGAAGGGAGCUCC

```
                                                                            _____
     A    UU            CAC     U                 U   G   AC  U   UC      /            \
5'  GAGCU  CUUCGGUCCA  ACGGG  GGAUUUAGGGUUCAAU  GG-CU CCG  UCA UCA  CA         68 nt
    |||||  ||||||||||  |||||  ||||||||||||||||  || || |||  ||| |||  ||          |
3'  CUCGA  GAAGUCAGGU  UGUCC  UCU-AAUUCUAGGUUG  UCAGA GGC  AGU AGU  GU         LOOP
     C    GG           UCC    U                 U   G   GU  U   GA      _____/
```

egr-miR159c CUUGGACUGAAGGGAGCUCC

```
                                                                            _____
     A                   CU  U     AUC-       U    G  G  AC       U   /            \
5'  GAGCUUCCUUCAGUCCA  CA GGAGGA       GGGGU GAAUUG CU CCG  UCAUUCAU UG       15 nt
    |||||||||||||||||  ||  ||||||      ||||| |||||| || |||  |||||||| ||          |
3'  CUCGAGGGAAGUCAGGU  GU CCUUUU       CCUCG UUUAAU GA GGC  AGUAAGUG GC       LOOP
     C                   UC  U     CUUC       C    A  G  GU       U   _____/
```

egr-miR159d CUUAGAAUGAAGGGAGCUCU

```
                                     CU
                                   C UAGAG  \
                                  C  |||||     C
                                  C  AUCUU  /
         AAA                    /    /     CC
5' AGA      UCCCUUCAUUCUAAGGCCAA    CCAU
   |||      ||||||||||||||||||||           A
3' UCU      AGGGAAGUAAGAUUCCGGUU    GCCU
      CG-                         \   |
                                   \  \    UC  AAC
                                   U \GGU  CGU   A
                                    \ |||  |||   |
                                    -UCA  GCA   C
                                      UA  AAU
```

egr-miR159e AUUGGACUGAAGGGAGCUCC

```
                                                                            _____
       UU              CC      C            UU    U   G   AC         C   /            \
5' GGAGCU  CUUCAGUCCA  CAUGGGA GUGUUGGG  UUAAU AG-CU CCG  UCAUUCAU CA       35 nt
   ||||||  ||||||||||  ||||||| |||||||   ||||| || || |||  |||||||| ||          |
3' CCUCGA  GAAGUCAGGU  GUGCCUU CG-AACCC  AGUUG UCUGA GGC  AGUAGGUA GU       LOOP
       GG              UA      U            UU    U   G   GU         A   _____/
```

egr-miR159f AUUGGACUGAAGGGAGCUCC

```
                                                                            _____
                      A    U   -AC-     CC   ---   G    UC         /            \
5' GGAGCUCCUUUCAGUCCAAUACAG GGGC GAG       GCGGA  GA  GCU CCAU  CAUGCACC       82 nt
   |||||||||||||||||||||||| |||  |||       |||||  ||  ||| ||||  ||||||||          |
3' CCUCGAGGGAAGUCAGGUUA-GUU CUCG CUC       UGCCU  CU  CGA GGUG  GUACGUGG       LOOP
                      A    C   AUAU     UC   CAA   G    UG         _____/
```

156

egr-miR160a UGCCUGGCUCCCUGUAUGCCA

```
      UAU U  C      CU          A  A  UU--        UUC
5'    GC UGGCUCC  GUAUGCCAC AAC UA     CCAACCU      A
      || |||||||  ||||||||| ||| ||     |||||||      |
3'    CG ACCGAGG  CAUGCGGUG UUG AU     GGUUGGA      U
    AUC    A      AC        G   C  UACC        UUC
```

egr-miR160b UGCCUGGCUCCCUGUAUGCCA

```
         C      CU       U     C   C     A
5' UGUGC UGGCUCC  GUAUGCCAU UGCGUAG CCA CGG     G
   ||||| |||||||  ||||||||| ||||||| ||| |||     |
3' AUACG ACCGAGG  UAUGCGGUA GUGUGUC GGU GCC     A
         U      AG        U        C   A   A
```

egr-miR160c UGCCUGGCUCCCUGAAUGCCA

```
                                CAGUU
    U  C        GA       U        \ /      AGA
5'  GC UGGCUCCCU  AUGCCA CUAA-GAAG UUGCCAA    C
    || ||||||||||  |||||| |||| |||| |||||||    |
3'  CG ACCGAGGGG  UACGGU GAUUCCUUC-AACGGUU    A
   A  U        AG       U               GAG
```

egr-miR160d UGCCUGGCUCCCUGAAUGCCA

```
    U  C                 U       UU   CC    GUUGA AAA
5'  GC UGGCUCC-CUGAAUGCCA CUAAGGAG  GGUU  AACA    CC  G
    || |||||||| |||| ||||| |||||||||  ||||  ||||        ||
    CG ACUGAGGAGACU-ACGGU GAUUCUUC  UCAG  UUGU    GG  A
   A  U                 U      U-    U-   AGGA-   AAG
```

----

egr-miR168a UCGCUUGGUGCAGGUCGGGA

```
   GGU            C        U     A UG      ---GA    AUUGCCAGAUGGCUCGACAUGACU
5'    CUCUGAUUCG UUGGUGCAGG CGGGA C AUUCGGC   UUUG                          U
      |||||||||| |||||||||| ||||| |    ||||        ||||
3'    GAGGCUAAGU AACUACGUUC GCCCU G UAAGC-G    AAAC                          G
    --G         C        C      A GU      AUAAG    AAAAAAAAGGACAAAGGAAGGAAA
```

egr-miR168b UCGCUUGGUGCAGGUCGGGA

```
   GG   UA    C        U     A UG    ---GG -GAUU        AG        AC
5' UCUC  AUUCG UUGGUGCAGG CGGGA C AUUC     C      UGAUU-GCC AUGGCUAA  A
   ||||  |||||  |||||||||| ||||| |  ||||        |     ||||| ||| |||||||||
3' AGAG  UAAGU AACUACGUUC GCCCU G UAAG     G      AUUAA CGG UGUCGGUU   C
   --     GC    C        C      A GU    AGAAA AAAAU         AG        AG
```

----

egr-miR172a AGAAUCUUGAUGAUGCUGCAU

```
                          _____
                  A      /             \
5' GUGUAGCAUCAUCAAGAUUC CAUAU           94 nt
   |||||||||||||||||||| |||||           COMPLEX
3' UACGUCGUAGUAGUUCUAAG GUGUG           LOOP
                  A      _____/
```

157

egr-miR172b AGAAUCUUGAUGAUGCUGCAU

```
                          A         --C-                    UGUGU     UCG
5' GUGCAGCAUCAUUAAGAUUC CAUCGAUC     AAUUGCCACGGGGCGGCU      AGA      U
   |||||||||||||||||||| |||| |||     ||||||| |||||||||| ||         |||
3' UACGUCGUAGUAGUUCUAAG GUGG-UAG     UUAACG-UGUUUCGC-GG      UCU      U
                       A          UCGU                      --U--    UGU
```

egr-miR172c AGAAUCUUGAUGAUGCUGCAU

```
                                         UGAAAA
                                        G      A
    -C                     A   GUGAAAAA          \   / AU
5'    GUAGCAUCAUCAAGAUUC CAA         GGAC----GGC   GG    \
      |||||||||||||||||| |||         ||||    |||   ||     C
3'    CGUCGUAGUAGUUCUAAG GUU         CCUGAAAGCCGUUCC       /
    UA                    A   AGAGUCUA                   CG
```

egr-miR172d AGAAUCUUGAUGAUGCUGCA

```
    C                     A   GUGAAA              AG
5'    GUAGCAUCAUCAAGAUUC CAA       AUGGAC-GGCGAU    U
      |||||||||||||||||| |||       |||||| ||||||    |
3'    CGUCGUAGUAGUUCUAAG GUU       UACCUG CCGCUA    U
    A                     A   AGAGUC  /    \    GG
                                     A      C
                                     A      U
                                      AGCCGUU
```

egr-miR172e ACAAUCUUGAUGAUGCUGCA

```
     C   U              CA   GUGACA     -GG-     UU   A
5'   GCA CAUCAUCAAGAUU CAA        AUGGAC    AGAUGA  GG   U
     ||| ||||||||||||| |||        ||||||    || |||  ||
3'   CGU GUAGUAGUUCUAA GUU        UACCUG    UC-GUU  CC   C
    A   C              CA   AGAGUC      AAAA      UC   G
```

egr-miR172f GGAAUCUUGAUGAUGCUGCAG

```
    U           G        AAAGAACUU  UUUU   A                 _____
5'   UGCAGCAUCAU AAGAUUCCCA        UC    UUG AGAGAG-CGUCUUU  /          \
     |||||||||||| ||||||||||       ||    ||| ||||||| || ||||||   46 nt
3'   ACGUCGUAGUA UUCUAAGGGU        AG    AAC UCUCUCUGC-GAAAA  COMPLEX
    G           G        ACCAUUUU- UAU-   C                _____/   LOOP
```

egr-miR172g GGAAUCUUGAUGAUGCUGCAG

```
              36 nt
              LOOP
     U           A / \  U    AAG   ----        UUAAUG   CUUU
5' CUGCA CAUCAUCAAGAUUC CA    AUUAG GGGGUU  AUU     GGGGGA     GCC    \
   ||||| ||||||||||||||| ||    ||||| ||||||  |||     ||||||     |||     U
3' GACGU GUAGUAGUUCUAAG GUUAAGUAGUU CUCCGA  UGA     UCCCCU     CGG    /
          C             G      U   AAA  ACUC    CACUUU   AACA
```

158

egr-miR472 UUUCCCAAGGCCGCCCAUUCC

```
                           U    -----     A   GA             G
5' GGAAUGGGCGG-UUUGGGA AAAC        CGGAG AAC  CGU--UGUCGA    A
   ||||||||||| ||||||| ||||        ||||| |||  |||  ||||||
3' CCUUACCCGCCGGAACCCU UUUG        GCUUU UUG  GUACCACGGCU    U
                      U    AAUUA       C   A-             C
```

Figure 3.2. Predicted precursor stem-loop structures of *Eucalyptus grandis* miRNAs from previously identified families. Precursor structures are presented as determined by computational analysis using the mfold program. Mature cloned miRNA sequences are shown in bold. MiRNA families are separated by dashed lines. The sequences represent full-length only the shortest possible stem-loop structure, where only that portion of the stem-loop starting at the cloned sequence and ending at the complementary miRNA* sequence being shown.

A

B

Figure 3.3. Expression profiles of *Eucalyptus* miRNAs. Expression levels as determined by quantitative real-time PCR. The expression data were normalized to the 5.8S rRNA reference gene. These values are standardized to the tissue in which the miRNA exhibited the highest expression level and are thus depicted as percentages. The value in parentheses is a measure

160

of the expression level of the 5.8S rRNA gene relative to the expression of the miRNA in the tissue where it is most abundant (i.e. where 45X indicates that the miRNA is expressed 45 times less than the 5.8S rRNA gene). X, P, IX, and L: mature xylem, phloem, immature xylem and leaf. The specific miRNA family is indicated above the graph. The data is represented as means of three measurements per tissue ± SE. Egr-miR200a and Egr-miR200e depict expression profiles for these two family members, which contain sequence variation in the mature miRNA molecule.

(A) Relative expression levels of putative novel *Eucalyptus* miRNAs.

(B) Relative expression levels of known *Eucalyptus* miRNAs.

3.8 TABLES

Table 3.1. Putative novel microRNAs cloned from vascular tissues of *Eucalyptus* and gene family members identified in the *E. camaldulensis* genome sequence

miRNA[a]	miRNA Sequence (5' - 3')[b]	Length (nt)[c]	miRNA genome reference[d]	Arm[e]
	UUACAAUUCAGUACAUCAGACU(1)	22		
egr-miR31	UUACAAUCCAGUACAUCAGACU		E0152813f.ab1.255.1234_501	5'
	AGCUGGGUGCUUGACUUUGG (2)	20		
egr-miR90	AGGUUGGUGCUUGGCUUUGG		E0121074f.ab1.273.1267_378	5'
	AUUAGAACAAGUGCAAUAGA(1)	20		
egr-miR140	AUUGGCACAAGUGCAAUAGA		E1045769f.ab1.205.1149_343	5'
	CGCACCCUCAUCGGCCACAAGC(1)	22		
egr-miR200a	CGCACCCUCGCCGGCCACAAGC		E0224981r.ab1.1.816_293	5'
egr-miR200b	CGCACCCUCGUUGGCCACAAGC		E0709307f.ab1.1.748_225	5'
egr-miR200c	CGCACCCUCGCCGGCCACAAGC		E0162293r.ab1.1.811_159	5'
egr-miR200d	CACGCCCUCAUCGGCCACAAGC		E0130629f.ab1.1.858_335	5'
egr-miR200e	UGCACCCUCAUCGACCACAAGC		E1516675f.ab1.1.624_101	5'
	GGUCGGCCUCGCCAUGGCUGG(1)	21		
egr-miR293a	GGUCGGCCUCGCCAUGGCUGG		E1403267r.ab1.1.605_48	3'
egr-miR293b	GGUCGGCCUCGCCAUGGCUGG		E1231035r_051.ab1.1.616_59	3'
egr-miR293c	GGUCGGCCUCGCCAUGGCUGG		E1564604r.ab1.295.1288_464	3'
egr-miR293d	GGUCGGCCUCGCCAUGGCUGG		E1244745r.ab1.11.1031_464	3'
egr-miR293e	GGUCGGCCUCGCCAUGGCUAG		E1573050r.ab1.1.627_107	3'
	UAAGUCGCGAAUUGGUAUGGGCU(1)	23		
egr-miR320a	UAAGUCGCGAAUUGGUAUGGGCU		E1683802r.ab1.1.909_220	5'
egr-miR320b	UAAGUCGCGAAUUGGUAUGGGCU		E0467787f.ab1.1.975_453	5'
egr-miR320c	UAAGUCUCGAAUUGGUAUGGGCU		E1248077r.ab1.1.687_165	5'
egr-miR320d	UAAGUCGUGAAUUGGUAUGGGCU		E1626235f.ab1.10.1032_472	5'
egr-miR320e	UAAGUCGCGAAUUAGUAUGGGCU		E0404437r_053.ab1.1.746_224	5'
egr-miR320f	UAAGUCGAGAAUUGGUACGGGCU		E0246027r_019.ab1.47.946_501	5'
	UCUGUUUAAAAAUUGUUCCAGG(1)	22		
egr-miR359a	UCUGUUUAAAAAUUGUUCCAGG		E1013068f_091.ab1.1.693_170	5'
egr-miR359b	UCUGUUUAAAAAUUGUUCCAGG		E0532804f.ab1.192.1215_451	5'
egr-miR359c	UCUGUUUAAAAAUUGUUCCAGG		E0273663f_040.ab1.1.955_401	5'
egr-miR359d	UCUGUUCAAAAAUUGUUCCAGG		E1447690f.ab1.197.1204_451	5'
egr-miR359e	UCUGUUCAAAAAUUGUUCCAGG		E1381605r_022.ab1.1.836_262	5'
egr-miR359f	UCUGUUCAAAAAUUGUUCCAGG		E1196861r.ab1.1.581_58	5'
egr-miR359g	UCUGUUCAAAAAUUGUUCCAGG		E0570168r.ab1.106.1129_501	5'
egr-miR359h	UCUGUUCAAAAAUUGUUCCAGG		E0553137r.ab1.1.945_371	5'
	UCUUGGAACUCAACAUCCUG(1)	20		
egr-miR362	UCUUGGAACUCAACAUCCUG		E0121191r.ab1.1.831_163.331	5'

[a] egr-miRNA gene name assigned in the order of cloning. In cases where the miRNA family contains multiple members, the gene name is kept the same followed by a letter to distinguish between members.

b Sequence of identified egr-miRNAs given in 5' to 3' orientation. The cloned small RNA sequences are shown in bold, followed by the frequency of cloning as indicated in parentheses. Sequences in normal type were identified through computational analysis as described in the Materials and Methods section of this chapter.

c Length of the cloned small RNA in nucleotides (nt). Only the longest cloned sequence of each family is shown.

d *E. camaldulensis* genome reference for each small RNA location as determined by shotgun sequencing. Names were designated by the sequence file name followed by the 5' position of the small RNA within the transcript (after the underscore).

e Location (5' or 3' arm) of the small RNA within the predicted precursor stem-loop structure as shown in Figure 3.1.

Table 3.2. *Eucalyptus* miRNAs conserved in other species and gene family members

identified in the *E. camaldulensis* genome sequence

miRNA[a]	miRNA Sequence (5' - 3')[b]	Length (nt)[c]	miRNA genome reference[d]	Arm[e]
	UUUGGAUUGAAGGGAGCUCU(5)	20		
egr-miR159a	CUUGGACUGAAGGGAGCUCC		E0164506f.ab1.229.1248_337	3'
egr-miR159b	CUUGGACUGAAGGGAGCUCC		E0085684f.ab1.157.1176_322	3'
egr-miR159c	CUUGGACUGAAGGGAGCUCC		E0505209r.ab1.1.920_402	3'
egr-miR159d	CUUAGAAUGAAGGGAGCUCU		E1498862r.ab1.1.976_375	3'
egr-miR159e	AUUGGACUGAAGGGAGCUCC		E0435958r.ab1.360.1252_355	3'
egr-miR159f	AUUGGACUGAAGGGAGCUCC		E0406035r_051.ab1.1.894_376	3'
	UGCCUGGCUGCCCUGUAUGCCA(13)	21		
egr-miR160a	UGCCUGGCUCCCUGUAUGCCA		E0246270f_094.ab1.194.953_504	5'
egr-miR160b	UGCCUGGCUCCCUGUAUGCCA		E0443862f.ab1.360.1295_441	5'
egr-miR160c	UGCCUGGCUCCCUGAAUGCCA		E0238876r.ab1.1.778_191	5'
egr-miR160d	UGCCUGGCUCCCUGAAUGCCA		E0634921r.ab1.1.942_422	5'
	UCGCUUGGUGCAGGUCGGGA(1)	20		
egr-miR168a	UCGCUUGGUGCAGGUCGGGA		E0593603f.ab1.1.689_179	5'
egr-miR168b	UCGCUUGGUGCAGGUCGGGA		E0264641f_065.ab1.255.944_314	5'
	AGAAUCUUGAUGAUGCUGCAU(5)	21		
egr-miR172a	AGAAUCUUGAUGAUGCUGCAU		E0220437r.ab1.1.589_69	3'
egr-miR172b	AGAAUCUUGAUGAUGCUGCAU		E0162604f.ab1.1.733_213	3'
egr-miR172c	AGAAUCUUGAUGAUGCUGCAU		E0058639r_039.ab1.1.946_390	3'
egr-miR172d	AGAAUCUUGAUGAUGCUGCAU		E0182002r.ab1.1.913_315	3'
egr-miR172e	ACAAUCUUGAUGAUGCUGCAU		E0525459f.ab1.56.1076_423	3'
	GGAAUCUUGAUGAUGCUGCAG(4)	21		
egr-miR172f	GGAAUCUUGAUGAUGCUGCAG		E0381021f.ab1.1.970_450	3'
egr-miR172g	GGAAUCUUGAUGAUGCUGCAG		E0173163f.ab1.1.916_396	3'
	UUUCCCAAGGCCGCCCAUUCC(2)	21		
egr-miR472a	UUUCCCAAGGCCGCCCAUUCC		E0171896f.ab1.1.871_351	3'

[a] egr-miRNA gene name as assigned by sequence similarity to known miRNAs identified in previous studies.

[b] Sequence of identified egr-miRNAs given in 5' to 3' orientation. The cloned small RNA sequences are shown in bold, followed by the frequency of cloning as indicated in parentheses. Sequences in normal type were identified through computational analysis.

[c] Length of the cloned small RNA in nucleotides (nt). Only the longest cloned sequence of each family is shown.

[d] *E. camaldulensis* genome reference for each small RNA location as determined by shotgun sequencing. Names were designated by the sequence file name followed by the small RNA start site within the transcript (after the underscore).

[e] Location (5' or 3' arm) of the small RNA within the predicted precursor stem-loop structure as shown in Figure 3.2.

Table 3.3. Putative targets and predicted gene functions for novel egr-miRNAs in *Populus*

and *Arabidopsis*

miRNA[a]	Protein annotation[b]	*P. trichocarpa* and *Arabidopsis* target genes[c]	Predicted function[d]
egr-miR31	Putative ethylene-responsive CBP	fgenesh1_pg.C_LG_V000143(2.5), At2g22300/AtSR1 (2)	Signal transduction
	SCAR complex component	At2g35110(2)	Trichome morphogenesis
	Pectinase	At4g23820(2.5)	Carbohydrate metabolism
	Patatin	At2g39220(2.5)	Lipid metabolism
	Cysteine protease	At1g50670(2.5)	Proteolysis
	CCAAT-binding factor	eugene3.00130940(2.5), At1g08970	Transcription factor
egr-miR90	MADS-box family protein	At3g18650(1.5)	Transcription factor
	Ubiquitin protein ligase	At1g55860(2.5)	Protein ubiquitination
	Unknown	eugene3.00061771(2.5), eugene3.00040592(2.5), fgenesh1_pm.C_LG_XII000266(2.5)	Unknown
egr-miR140	Peptidase	At5g43600(2.5)	Proteolysis
	LIM domain containing protein	At5g17890(2.5)	Protein interactions
	Putative small molecule transporter	fgenesh1_pm.C_LG_X000537 (2.5), eugene3.20760001(2.5), At1g14040 (2.5)	Signal transduction
	Unknown	estExt_fgenesh1_pg_v1.C_LG_VIII0033(2.5)	Unknown
egr-miR200	PPR	fgenesh1_pm.C_LG_XVI000226(2.5)	Organelle biogenesis
	Unknown	At5g44170(2.5), At3g06930(2.5), eugene3.00131142(2.5), fgenesh1_pg.C_scaffold_12963000001(2.5)	Unknown
egr-miR293	Putative disease resistance protein	fgenesh1_pg.C_LG_VII000915(1)	Defense response
	PAPS reductase	grail3.8743000201(2.5)	Biosynthesis
	Protein kinase	fgenesh1_pg.C_LG_VII001044(2.5)	Phosphorylation
	Unknown	estExt_Genewise1_v1.C_LG_VIII2152 (2.5), eugene3.00180993(2.5)	Unknown
egr-miR320	Predicted phosphate acyltransferase	eugene3.00130312(2.5), estExt_fgenesh1_pg_v1.C_700114(2.5)	Lipid metabolism
	Unknown	fgenesh1_pg.C_scaffold_766000001(2.5), eugene3.00102297(2.5)	Unknown
egr-miR359	Oxidoreductase	estExt_fgenesh1_pg_v1.C_LG_X2093(2)	Electron transport
	AT-rich domain containing protein	eugene3.00180022(2.5), At3g43240 (2.5)	Transcription regulation
	ABC transporter	eugene3.00140576(2.5), eugene3.00011077(2)	Signal transduction
	Unknown	grail3.0143000601(2.5), fgenesh1_pg.C_LG_I000711(2.5)	Unknown

165

egr-miR362	Ras GTPase protein	eugene3.00150925(2.5)	Signal transduction
	Putative calmodulin	eugene3.00051226(2.5)	Signal transduction
	Unknown	fgenesh1_pg.C_LG_V000073(2.5), eugene3.00190215(2.5), At1g29630(2.5)	Unknown

[a] Name of the putative novel *Eucalyptus* miRNA family. The targets identified here apply to an entire small RNA family and not to individual family members.

[b] Poplar target loci are given as putative gene models as shown at http://genome.jgi-psf.org /Poptr1/Poptr1.home.html. *Arabidopsis* target loci are given as gene models from The Arabidopsis Information Resource database (http://www.arabidopsis.org).

[c] Genome location of the predicted targets were identified by searching the Egr-miR sequence patterns using PatScan. Searches performed using the draft *Populus* genome assembly (http://genome.jgi-psf.org/Poptr1/Poptr1.home.html) and The Arabidopsis Information Resource database (http://www.arabidopsis.org). For each putative target locus, the number of mismatches between the miRNA and the mRNA is indicated in parentheses. Only putative targets with scores/mismatches of 3 or less are shown here.

[d] Functions for target gene families were predicted using a protein domain search (http://www.ebi.ac.uk/interpro).

Table 3.4. Putative targets and predicted gene functions for known egr-miRNAs in *Populus*

and *Arabidopsis*

miRNA[a]	Protein annotation[b]	*P. trichocarpa* and *Arabidopsis* target genes[c]	Predicted function[d]
egr-miR159	MYB	eugene3.00400064(2), estExt_fgenesh1_pg_v1.C_LG_IX1359(2), eugene3.00011133(2), grail3.0100003501(2.5), At2g32460/AtMYB101(1.5), At3g60460(2), At2g26950/AtMYB104(2), At5g55020(2.5), At5g06100/AtMYB33(2.5), At3g11440/AtMYB65(2.5), At5g06100(2.5), At4g26930	Transcription factor activity
	Asparagine synthase	fgenesh1_pg.C_LG_X001429(1.5)	Amino acid metabolism
	Unknown	grail3.0043007201(0), grail3.0016033301(1.5), eugene3.00860143(2), grail3.0123001501(2.5), eugene3.00050584(2.5), fgenesh1_pg.C_LG_I002929(2.5), At1g29010 (2.5)	Unknown
egr-miR160	Auxin responsive factor	fgenesh1_pg.C_LG_V000905(0.5), eugene3.00020832(0.5), eugene3.00280173(0.5), eugene3.00280175(0.5), estExt_Genewise1_v1.C_LG_XVI2818(0.5), fgenesh1_pm.C_LG_IX000578(1), eugene3.00660262(1), fgenesh1_pm.C_LG_VIII000107(1), estExt_fgenesh1_pm_v1.C_LG_X0730(	Transcription factor activity
egr-miR168	ARGONAUTE1	grail3.0122002801(4)3, At1g48410/AGO (3)	MicroRNA biogenesis
	Vesicle coat protein complex COP1	eugene3.00120688(2.5), estExt Genewise1 v1.C LG XV1316(2.5)	Intracellular trafficking and protein transport
	Unknown	grail3.0110002401(0)	Unknown
egr-miR172	Homeotic protein APETALA2	fgenesh1_pg.C_LG_X001967(0.5), eugene3.00050501(0.5), grail3.0019003502(0.5), eugene3.00160775(0.5), fgenesh1_pg.C_scaffold_28000114(0.5), At4g36920(0.5), At5g60120(0.5), At2g28550(1.5), At5g67180(1.5), At2g39250(2)	Transcription factor activity
	Unknown	estExt_fgenesh1_pg_v1.C_LG_VIII1617(0.5), eugene3.00150782(2.5), eugene3.00091546(1.5)	Unknown
egr-miR472	Putative disease resistance protein	eugene3.05370005(2.5), eugene3.00440205(2.5), eugene3.00440201(2.5), eugene3.07650002(2.5), eugene3.00440206(2.5), eugene3.00440198(2.5), eugene3.00150937(2.5), eugene3.00440160(2.5), eugene3.07650001(2.5), eugene3.00440159(2.5), fgenesh1_pg.C_scaffold_63	Defense response

^a Name of the identified putative known *Eucalyptus* miRNA family. The targets identified here apply to an entire small RNA family and not to individual family members.

^b Poplar target loci are given as putative gene models as shown at http://genome.jgi-psf.org /Poptr1/Poptr1.home.html. *Arabidopsis* target loci are given as gene models from The Arabidopsis Information Resource database (http://www.arabidopsis.org).

^c Genome location of the predicted targets were identified by searching the egr-miR sequence patterns using PatScan. Searches performed using the draft *Populus* genome assembly (http://genome.jgi-psf.org/Poptr1/Poptr1.home.html) and The Arabidopsis Information Resource database (http://www.arabidopsis.org). For each putative target locus, the number of mismatches between the miRNA and the mRNA is indicated in parentheses. Only putative targets with scores/mismatches of 3 or less are shown here.

^d Functions for target gene families were predicted using a protein domain search (http://www.ebi.ac.uk/interpro).

3.9 REFERENCES

Achard, P., Herr, A., Baulcombe, D.C., and Harberd, N.P. (2004). Modulation of floral development by a gibberellin-regulated microRNA. Development **131,** 3357-3365.

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.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25,** 3389-3402.

Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M., Matzke, M., Ruvkun, G., and Tuschl, T. (2003). A uniform system for microRNA annotation. RNA **9,** 277-279.

Aung, K., Lin, S.-I., Wu, C.-C., Huang, Y.-T., Su, C.-l., and Chiou, T.-J. (2006). *pho2*, a phosphate overaccumulator, is caused by a nonsense mutation in a microRNA399 target gene. Plant Physiol. **141,** 1000-1011.

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

Babak, T., Zhang, W., Morris, Q., Blencowe, B.J., and Hughes, T.R. (2004). Probing microRNAs with microarrays: Tissue specificity and functional inference. RNA **10,** 1813-1819.

Baltz, R., Evrard, J.L., Domon, C., and Steinmetz, A. (1992). A LIM motif is present in a pollen-specific protein. Plant Cell **4,** 1465-1466.

Banerjee, D., and Slack, F. (2002). Control of developmental timing by small temporal RNAs: a paradigm for RNA-mediated regulation of gene expression. Bioessays **24,** 119-129.

Banfalvi, Z., Kostyal, Z., and Barta, E. (1994). *Solanum brevidens* possesses a non-sucrose-inducible patatin gene. Mol. Gen. Genet. **245,** 517-522.

Bari, R., Datt Pant, B., Stitt, M., and Scheible, W.-R. (2006). PHO2, microRNA399, and PHR1 define a phosphate-signaling pathway in plants. Plant Physiol. **141,** 988-999.

Barnett, J., and Bonham, V. (2004). Cellulose microfibril angle in the cell wall of wood fibres. Biol. Rev. Camb. Philos. Soc. **79,** 461-472.

Bartel, B., and Bartel, D.P. (2003). MicroRNAs: at the root of plant development. Plant Physiol. **132,** 709-717.

Bates, P.W., and Vierstra, R.D. (1999). UPL1 and 2, two 405 kDa ubiquitin-protein ligases from *Arabidopsis thaliana* related to the HECT-domain protein family. Plant J. **20,** 183-195.

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. USA **101,** 115511-111516.

Burton, R.A., Gibeaut, D.M., Bacic, A., Findlay, K., Roberts, K., Hamilton, A., Baulcombe, D.C., and Fincher, G.B. (2000). Virus-induced silencing of a plant cellulose synthase gene. Plant Cell **12,** 691-706.

Chang, S., Puryear, J., and Cairney, J. (1993). A simple and efficient method for isolating RNA from pine trees. Plant Mol. Biol. Rep. **11,** 113-116.

Chen, C.Z., Li, L., Lodish, H.F., and Bartel, D.P. (2004). MicroRNAs modulate hematopoietic lineage differentiation. Science **303,** 83-86.

Cseke, L.J., Zheng, J., and Podila, G.K. (2003). Characterization of PTM5 in aspen trees: a MADS-box gene expressed during woody vascular development. Gene **318,** 55-67.

Davies, G., and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. Structure **3,** 853-859.

de Folter, S., Immink, R.G.H., Kieffer, M., Parenicova, L., Henz, S.R., Weigel, D., Busscher, M., Kooiker, M., Colombo, L., Kater, M.M., Davies, B., and Angenent, G.C. (2005). Comprehensive interaction map of the *Arabidopsis* MADS box transcription factors. Plant Cell **17,** 1424-1433.

Dharmasiri, N., Dharmasiri, S., and Estelle, M. (2005). The F-box protein TIR1 is an auxin receptor. Nature **435,** 441-445.

Dsouza, M., Larsen, N., and Overbeek, R. (1997). Searching for patterns in genomic data. Trends Genet. **13,** 497-498.

Ehrhardt, A., Ehrhardt, G.R., Guo, X., and Schrader, J.W. (2002). Ras and relatives - job sharing and networking keep an old family together. Exp. Hematol. **30,** 1089-1106.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev. **15,** 188-200.

Eldridge, K., Davidson, J., Harwood, C., and Van Wyk, G. (1993). Eucalypt domestication and breeding. (Oxford: Oxford University Press).

Eliasson, A., Gass, N., Mundel, C., Baltz, R., Krauter, R., Evrard, J.L., and Steinmetz, A. (2000). Molecular and expression analysis of a LIM protein gene family from flowering plants. Mol. Gen. Genet. **264,** 257-267.

Evans, R., and Ilic, J. (2001). Rapid prediction of wood stiffness from microfibril, angle and density. Forest Prod. J. **51,** 53-57.

Friml, J. (2003). Auxin transport - shaping the plant. Curr. Opin. Plant Biol. **6,** 7-12.

Fu, X., and Harberd, N.P. (2003). Auxin promotes *Arabidopsis* root growth by modulating gibberellin response. Nature **421,** 740-743.

Geisler, M., and Murphy, A.S. (2006). The ABC of auxin transport: the role of p-glycoproteins in plant development. FEBS Lett **580,** 1094-1102.

Griffiths-Jones, S. (2004). The microRNA Registry. Nucleic Acids Res. **32,** D109-111.

Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., and Enright, A.J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. **34,** D140-144.

Gusmaroli, G., Tonelli, C., and Mantovani, R. (2002). Regulation of novel members of the *Arabidopsis thaliana* CCAAT-binding nuclear factor Y subunits. Gene **283,** 41-48.

Hancock, J.F. (2003). Ras proteins: different signals from different locations. Nat. Rev. Mol. Cell Biol. **4,** 373-385.

Hardtke, C.S., and Berleth, T. (1998). The *Arabidopsis* gene MONOPTEROS encodes a transcription factor mediating embryo axis formation and vascular development. EMBO J. **17,** 1405-1411.

Hertzberg, M., Aspeborg, H., Schrader, J., Andersson, A., Erlandsson, R., Blomqvist, K., Bhalerao, R., Uhlen, M., Teeri, T.T., Lundeberg, J., Sundberg, B., Nilsson, P., and Sandberg, G. (2001). A transcriptional roadmap to wood formation. Proc. Natl. Acad. Sci. USA **98,** 14732-14737.

Horvath, D.P., Schaffer, R., West, M., and Wisman, E. (2003). *Arabidopsis* microarrays identify conserved and differentially expressed genes involved in shoot growth and development from distantly related plant species. Plant J. **34,** 125-134.

Huang, J.H., and Schell, M.A. (1990). DNA sequence analysis of *pglA* and mechanism of export of its polygalacturonase product from *Pseudomonas solanacearum.* J. Bacteriol. **172,** 3879-3887.

Israelsson, M., Eriksson, M.E., Hertzberg, M., Aspeborg, H., Nilsson, P., and Moritz, T. (2003). Changes in gene expression in the wood-forming tissue of transgenic hybrid aspen with increased secondary growth. Plant Mol. Biol. **52,** 893-903.

Jack, T. (2001). Plant development going MADS. Plant Mol Biol **46,** 515-520.

Jacobsen, S.E., Running, M.P., and Meyerowitz, E.M. (1999). Disruption of an RNA helicase/RNAse III gene in *Arabidopsis* causes unregulated cell division in floral meristems. Development **126,** 5231-5243.

Jiang, D., Yin, C., Yu, A., Zhou, X., Liang, W., Yuan, Z., Xu, Y., Yu, Q., Wen, T., and Zhang, D. (2006). Duplication and expression analysis of multicopy miRNA gene family members in *Arabidopsis* and rice. Cell Res. **16,** 507-518.

Johnson, S.M., Lin, S.-Y., and Slack, F.J. (2003). The time of appearance of the *C. elegans let-7* microRNA is transcriptionally controlled utilizing a temporal regulatory element in its promoter. Dev. Bio. **259,** 364-379.

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.

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

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

Kidner, C.A., and Martienssen, R.A. (2005). The role of ARGONAUTE1 (AGO1) in meristem formation and identity. Dev. Biol. **280,** 504-517.

Ko, J., Prassinos, C., and Han, K. (2006a). Developmental and seasonal expression of *PtaHB1*, a *Populus* gene encoding a class III HD-Zip protein, is closely associated with secondary growth and inversely correlated with the level of microRNA (miR166). New Phytol. **169,** 469-478.

Ko, J.H., Beers, E.P., and Han, K.H. (2006b). Global comparative transcriptome analysis identifies gene network regulating secondary xylem development in *Arabidopsis thaliana*. Mol. Genet. Genomics. **276,** 517-531.

Kortschak, R.D., Tucker, P.W., and Saint, R. (2000). ARID proteins come in from the desert. Trends Biochem. Sci. **25,** 294-299.

Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., and Tuschl, T. (2002). Identification of tissue-specific microRNAs from mouse. Curr. Biol. **12,** 735-739.

Lee, M., O'Regan, S., Moreau, J.L., Johnson, A.L., Johnston, L.H., and Goding, C.R. (2000). Regulation of the Pcl7-Pho85 cyclin-cdk complex by PHO81. Mol. Microbiol **38,** 411-422.

Lee, Y., Jeon, K., Lee, J.-T., Kim, S., and Kim, V.N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. EMBO J. **21,** 4663-4670.

Lenburg, M.E., and O'Shea, E.K. (1996). Signaling phosphate starvation. Trends Biochem. Sci. **21,** 383-387.

Li, X., and Zhang, Y. (2005). Computational detection of microRNAs targeting transcription factor genes in *Arabidopsis thaliana*. Comput. Biol. Chem. **29,** 360-367.

Li, X.Y., Mantovani, R., Hooft van Huijsduijnen, R., Andre, I., Benoist, C., and Mathis, D. (1992). Evolutionary variation of the CCAAT-binding transcription factor NF-Y. Nucleic Acids Res. **20,** 1087-1091.

Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C. (2002a). Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. Science **297,** 2053-2056.

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

Lotan, T., Ohto, M., Yee, K.M., West, M.A., Lo, R., Kwong, R.W., Yamagishi, K., Fischer, R.L., Goldberg, R.B., and Harada, J.J. (1998). *Arabidopsis* LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. Cell **93,** 1195-1205.

Lu, S., Sun, Y.-H., Shi, R., Clark, C., Li, L., and Chiang, V.L. (2005). Novel and mechanical stress-responsive microRNAs in *Populus trichocarpa* that are absent from *Arabidopsis*. Plant Cell **17,** 2186-2203.

Lukacin, R., Groning, I., Pieper, U., and Matern, U. (2000). Site-directed mutagenesis of the active site serine290 in flavanone 3beta-hydroxylase from *Petunia hybrida*. Eur. J. Biochem. **267,** 853-860.

Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M.L., Mireau, H., Peeters, N., Renou, J.P., Szurek, B., Taconnat, L., and Small, I. (2004). Genome-wide

analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell **16,** 2089-2103.

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

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

Martinoia, E., Klein, M., Geisler, M., Bovet, L., Forestier, C., Kolukisaoglu, U., Muller-Rober, B., and Schulz, B. (2002). Multifunctionality of plant ABC transporters - more than just detoxifiers. Planta **214,** 345-355.

McCormack, E., Tsai, Y.C., and Braam, J. (2005). Handling calcium signaling: *Arabidopsis* CaMs and CMLs. Trends Plant Sci. **10,** 383-389.

McHale, N.A., and Koning, R.E. (2004). MicroRNA-directed cleavage of *Nicotiana sylvestris* PHAVOLUTA mRNA regulates the vascular cambium and structure of apical meristems. Plant Cell **16,** 1730-1740.

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

Meyers, B., Souret, F., Lu, C., and Green, P. (2006). Sweating the small stuff: microRNA discovery in plants. Curr. Opin. Biotechnol. **17,** 1-8.

Mica, E., Gianfranceschi, L., and Pe, M.E. (2006). Characterization of five microRNA families in maize. J. Exp. Bot. **57,** 2601-2612.

Michelsen, J.W., Schmeichel, K.L., Beckerle, M.C., and Winge, D.R. (1993). The LIM motif defines a specific zinc-binding protein domain. Proc. Natl. Acad. Sci. USA **90,** 4404-4408.

Nakamura, T., Schuster, G., Sugiura, M., and Sugita, M. (2004). Chloroplast RNA-binding and pentatricopeptide repeat proteins. Biochem. Soc. Trans. **32,** 571-574.

Parcy, F., Valon, C., Kohara, A., Misera, S., and Giraudat, J. (1997). The ABSCISIC ACID-INSENSITIVE3, FUSCA3, and LEAFY COTYLEDON1 loci act in concert to control multiple aspects of *Arabidopsis* seed development. Plant Cell **9,** 1265-1277.

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.

Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Muller, P., Spring, J., Srinivasan, A., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E., and Ruvkun, G. (2000). Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA. Nature **408,** 86-89.

Plomion, C., Leprovost, G., and Stokes, A. (2001). Wood formation in trees. Plant Physiol. **127,** 1513-1523.

Poke, F.S., Vaillancourt, R.E., Potts, B.M., and Reid, J.B. (2005). Genomic research in *Eucalyptus*. Genetica **125,** 79-101.

Potts, B.M., and Dungey, H.S. (2004). Interspecific hybrids of eucalypts: key issues for breeders and geneticists. New For. **27**.

Ranik, M., Creux, N.M., and Myburg, A.A. (2006). Within-tree transcriptome profiling in wood-forming tissues of a fast-growing *Eucalyptus* tree. Tree Physiol. **26,** 365-375.

Reddy, V.S., Ali, G.S., and Reddy, A.S.N. (2002). Genes encoding calmodulin-binding proteins in the *Arabidopsis* genome. J. Biol. Chem. **277,** 9840-9852.

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

Reuther, G.W., and Der, C.J. (2000). The Ras branch of small GTPases: Ras family members don't fall far from the tree. Curr. Opin. Cell Biol. **12,** 157-165.

Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., and Bartel, D.P. (2002). Prediction of plant microRNA targets. Cell **110,** 513-520.

Riechmann, J.L., and Meyerowitz, E.M. (1997). MADS domain proteins in plant development. Biol Chem **378,** 1079-1101.

Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Z, C., Jiang, Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J.Z., Ghandehari, D., Sherman, B.K., and L. Yu, G. (2000). *Arabidopsis* transcription factors: Genome-wide comparative analysis among eukaryotes. Science **290,** 2105-2110.

Robinson, S.J., Cram, D.J., Lewis, C.T., and Parkin, I.A.P. (2004). Maximizing the efficacy of SAGE analysis identifies novel transcripts in *Arabidopsis*. Plant Physiol. **136,** 3223-3233.

Rogelj, B., and Giese, K. (2004). Expression and function of brain specific small RNAs. Rev. Neurosci. **15,** 185-198.

Rounsley, S.D., Ditta, G.S., and Yanofsky, M.F. (1995). Diverse roles for MADS box genes in *Arabidopsis* development. Plant Cell **7,** 1259-1269.

Rudie, A.W. (1998). Wood and how it relates to paper products. Tappi J. **81,** 223-228.

Ruttkowski, E., Labitzke, R., Khanh, N.Q., Loffler, F., Gottschalk, M., and Jany, K.D. (1990). Cloning and DNA sequence analysis of a polygalacturonase cDNA from *Aspergillus niger* RH5344. Biochim. Biophys. Acta. **1087,** 104-106.

Schrader, J., Baba, K., May, S.T., Palme, K., Bennett, M., Bhalerao, R.P., and Sandberg, G. (2003). Polar auxin transport in the wood-forming tissues of hybrid aspen is under simultaneous control of developmental and environmental signals. Proc. Natl. Acad. Sci. USA **100,** 10096-10101.

Shi, R., and Chiang, V.L. (2005). Facile means for quantifying microRNA expression by real-time PCR. Biotechniques **39,** 519-525.

Sood, P., Krek, A., Zavolan, M., Macino, G., and Rajewsky, N. (2006). Cell-type-specific signatures of microRNAs on target mRNA expression. Proc. Natl. Acad. Sci. USA **103,** 2746-2751.

Spain, B.H., Koo, D., Ramakrishnan, M., Dzudzor, B., and Colicelli, J. (1995). Truncated forms of a novel yeast protein suppress the lethality of a G protein alpha subunit deficiency by interacting with the beta subunit. J. Biol. Chem. **270,** 25435-25444.

Sunkar, R., and Zhu, J.K. (2004). Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. Plant Cell **16,** 2001-2019.

Sunkar, R., Girke, T., and Zhu, J.-K. (2005a). Identification and characterization of endogenous small interfering RNAs from rice. Nucl. Acids Res. **33,** 4443-4454.

Sunkar, R., Girke, T., Jain, P.K., and Zhu, J.K. (2005b). Cloning and characterization of microRNAs from rice. Plant Cell **17,** 1397-1411.

Szymanski, D.B. (2004). Breaking the WAVE complex: the point of *Arabidopsis* trichomes. Curr. Opin. Plant Biol. **8,** 103-112.

Terasaka, K., Blakeslee, J.J., Titapiwatanakun, B., Peer, W.A., Bandyopadhyay, A., Makam, S.N., Lee, O.R., Richards, E.L., Murphy, A.S., Sato, F., and Yazaki, K. (2005). PGP4, an ATP binding cassette P-Glycoprotein, catalyzes auxin transport in *Arabidopsis thaliana* roots. Plant Cell **17,** 2922-2939.

Timmermans, M., Juarez, M., and Phelps-Durr, T. (2004). A conserved microRNA signal specifies leaf polarity. Cold Spring Harb Symp. Quant. Biol. **69,** 409-417.

Wang, J.-F., Zhou, H., Chen, Y.-Q., Luo, Q.-J., and Qu, L.-H. (2004a). Identification of 20 microRNAs from *Oryza sativa*. Nucleic Acids Res. **32,** 1688-1695.

Wang, J.-W., Wang, L.-J., Mao, Y.-B., Cai, W.-J., Xue, H.-W., and Chen, X.-Y. (2005). Control of root cap formation by microRNA-targeted auxin response factors in *Arabidopsis*. Plant Cell **17,** 2204-2216.

Wang, X.J., Reyes, J.L., Chua, N.H., and Gaasterland, T. (2004b). Prediction and identification of *Arabidopsis thaliana* microRNAs and their mRNA targets. Genome Biol. **5,** R65.

West, M., Yee, K.M., Danao, J., Zimmerman, J.L., Fischer, R.L., Goldberg, R.B., and Harada, J.J. (1994). LEAFY COTYLEDON1 Is an essential regulator of late embryogenesis and cotyledon identity in *Arabidopsis*. Plant Cell **6,** 1731-1745.

Wilsker, D., Patsialou, A., Dallas, P.B., and Moran, E. (2002). ARID proteins: a diverse family of DNA binding proteins implicated in the control of cell growth, differentiation, and development. Cell. Growth Differ. **13,** 95-106.

Wu, G., and Poethig, R.S. (2006). Temporal regulation of shoot development in *Arabidopsis thaliana* by miR156 and its target SPL3. Development **133,** 3539-3547.

Wurschum, T., Gross-Hardt, R., and Laux, T. (2006). APETALA2 regulates the stem cell niche in the *Arabidopsis* shoot meristem. Plant Cell **18,** 295-307.

Yan, K.S., Yan, S., Farooq, A., Han, A., Zeng, L., and Zhou, M.M. (2003). Structure and conserved RNA binding of the PAZ domain. Nature **426,** 468-474.

Yang, S.H., van Zyl, L., No, E.G., and Loopstra, C.A. (2004). Microarray analysis of genes preferentially expressed in differentiating xylem of loblolly pine (*Pinus taeda*). Plant Sci. **166,** 1185-1195.

Yang, T., and Poovaiah, B.W. (2002). A calmodulin-binding/CGCG box DNA-binding protein family involved in multiple signaling pathways in plants. J. Biol. Chem. **277,** 45049-45058.

Zhang, B., Pan, X., Cannon, C.H., Cobb, G.P., and Anderson, T.A. (2006). Conservation and divergence of plant microRNA genes. Plant J. **46,** 243-259.

Zhang, X., Dyachok, J., Krishnakumar, S., Smith, L.G., and Oppenheimer, D.G. (2005). IRREGULAR TRICHOME BRANCH1 in *Arabidopsis* encodes a plant homolog of the actin-related protein2/3 complex activator Scar/WAVE that regulates actin and microtubule organization. Plant Cell **17,** 2314-2326.

Zhang, Z., Barlow, J.N., Baldwin, J.E., and Schofield, C.J. (1997). Metal-catalyzed oxidation and mutagenesis studies on the iron(II) binding site of 1-aminocyclopropane-1-carboxylate oxidase. Biochemistry **36,** 15999-16007.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. **31,** 3406-3415.

3.10 SUPPLEMENTARY MATERIAL

Table S3.1. Primers used for *Eucalyptus* miRNA expression analysis

Primer name	Sequence (5' - 3')
Egr-miR31	TTACAATTCAGTACATCAGACT
Egr-miR90	AGCTGGGTGCTTGACTTTGG
Egr-miR140	ATTAGAACAAGTGCAATAGA
Egr-miR159	TTGGATTGAAGGGAGCTCTA
Egr-miR160	TGCCTGGCTCCCTGTATGCCAA
Egr-miR168	TCGCTTGGTGCAGGTCGGGAA
Egr-miR172	GGAATCTTGATGATGCTGCAGA
Egr-miR200a	CGCACCCTCATCGGCCACAAGCA
Egr-miR200e	TGCACCCTCATCGACCACAAGCA
Egr-miR293	GGTCGGCCTCGCCATGGCTGGA
Egr-miR320	TAAGTCGCGAATTGGTATGGGCTA
Egr-miR359	TCTGTTTAAAAATTGTTCCAGGA
Egr-miR362	TCTTGGAACTCAACATCCTGAA
Egr-miR472	TTTCCCAAGGCCGCCCATTCC
Egr-5.8S rRNA_Forward	ACGTCTGCCTGGGTGTCACAA
Poly-T adaptor	GACCACGCGTATCGATGGCTCAT$_{16}$V*
Poly-T adaptor_Reverse	GACCACGCGTATCGATGGCTCA

*V=A,G,C

Table S3.2. Precursor sequences of putative novel and conserved miRNAs identified in the

E. camaldulensis genome sequences

Gene names[a]	miRNA genome reference[b]	Precursor Sequences (5' - 3')[c]
egr-MIR31	E0152813f.ab1.255.1234_501.609	**TTACAATCCAGTACATCAGACT**CTACGCAATCTTAGAGATTTCATATGACCACAAAACATATGAAATCTCTAGGATTGCGTAG*AGTCTGATGTACTAGATTGTAA*
egr-MIR90	E0121074f.ab1.273.1267_378.520	**AGGTTGGTGCTTGGCTTTGG**CCTAGTAAGGCCAAGCCTCGCCGGCCGTTGGGCGAGCTCGGCCTCACCGGATCTGGGCGAGATTGAGCTCACCCAGCGGCCGGCGAGGCTCAGCCTCACTAGG*CCACCGCCAAGCACCAACCT*
egr-MIR140	E1045769f.ab1.205.1149_343.520	**ATTGGCACAAGTGCAATAGA**TTTAAGAATTTTTTTTGGTAATTCTCCTATCTAGCACTAACTGAGCTGAGCTTATACAGTCCAAAACCTTTATCATATACTCCCTTGAAGTAGGTTTCACATGGTTCAACATCATTTTTTTTTAAAAAAAGTTCTAAAC*CTATTGTATTTACACCAAT*
egr-MIR200a	E0224981r.ab1.1.816_293.414	**CGCACCCTCGCCGGCCACAAGCGA**GGGCGACGGCCCCCGCCCAGATCCAGGCGAGGGCCGAAACCCTCGCCCGTTGGCCGGCGAGGGCGCGGCTGCCC*TCGCTCATGGCGGGCGAGGGTGCG*
egr-MIR200b	E0709307f.ab1.1.748_225.376	**CGCACCCTCGTTGGCCACAAGCGA**GGGCGTCAACCCCCGCCCAAATCCAGGCGAGGGCCGAAACCTTGCCCGTTGGCCGGCGAGGGCCTCTGTGCCCTCGCCTGGATCTAGGCGGTGGCCAACGCCC*TCGCCTGTGGTCGACGAGGGCGCA*
egr-MIR200c	E0162293r.ab1.1.811_159.311	**CGCACCCTCGCCGGCCACAAGCGA**GGGCGTCAGCCCTCGCCCAGATCCAGGCGAGGGCCGAAACCCTCGCTCATTGGCCGGCAAGGGCCTCTGCACCCTTGTCGGCCACAGGCAAGGGTGCGGCAGCCC*TCACTTGTGGTGTGCGAGGGTGCA*
egr-MIR200d	E0130629f.ab1.1.858_335.425	**CACGCCCTCATCGGCCACAAGCGA**GGGTGACGGCCCCCCCTAGATCTGAGCGAGGGAACGGCAGCAA*TTGCTCGTGGCGGACGAGGGTGCG*
egr-MIR200e	E1516675f.ab1.1.624_101.162	**TGCACCCTCATCGACCACAAGCGA**GGGCGCCGCAGCCC*TCGCTCGTGGTGGGCGAGGGTGCG*
egr-MIR293a	E1403267r.ab1.1.605_48.105	**GGTCGGCCTCGCCATGGCTGG**GTGAGCTTGAGCTCGC*CCAGCCATGGCGAGGCCCACC*
egr-MIR293b	E1231035r_051.ab1.1.616_59.116	**GGTCGGCCTCGCCATGGCTGG**GCGAGCTCGAGGTCGC*CCAGCCATGGCGAGGCGGCCG*
egr-MIR293c	E1564604r.ab1.295.1288_464.521	**GGTCGGCCTCGCCATGGCTGG**GCGAGCTTGAGCTCGC*CCAGCCATGGCAAGGTCGAAC*
egr-MIR293d	E1244745r.ab1.11.1031_464.521	**GGTCGGCCTCGCCATGGCTGG**GCGAGCTCGAGCCCGC*CCGGCCATGGCGAGGCCAACC*
egr-MIR293e	E1573050r.ab1.1.627_107.164	**GGTCGGCCTCGCCATGGCTAG**GCGAGCTCGAGCTCGC*CCAGCCATGGCGAGGCCTACC*
egr-MIR320a	E1683802r.ab1.1.909_220.409	**TAAGTCGCGAATTGGTATGGGCT**TGGGAGAGCTCGAGTTAAGTGTAAAAACTTAGTAGTCCATTAACCTAGTTTTAGGTGGTGCAGCCAAAAAATTAGGTCGCGACAGCATCATCTAAAACTAAGTTAATGAACCACTAAGTCTTTAAATTTAGCTCAGCTTTCCCA*AGCCCATACCAATTTGCAACTTA*

egr-MIR320b   E0467787f.ab1.1.975_453.757   **TAAGTCGCGAATTTGTATGGGCT**TGGGAGAGCCCGAACTAAGT
TTAAAGACTTAGTAGTCCATTAACCTAGTTTTAGGTGATGCAC
CCAAAAATTAGGTCGCGATAATCACTTAGGAGTTTTTATGCAC
GTTTGACTATATATTGAAAATTTGATGCTATTTTCTCTCGTCT
TCAAAGAACATCTTTATAACCAATCTGTCGCGACCTAATTTTC
GCGAGTTCACTACCTAAAACTAGGTTAATGGATTACTAAGTTT
TTAAGCTTAGCTCGAGCTCTCTCA*AGCCCATACCAATTCGCGA*
CTTA

egr-MIR320c   E1248077r.ab1.1.687_165.498   **TAAGTCTCGAATTGGTATGGGCT**TGGGAAAGCCCGAGCTAAGT
TTAAACTTAGTAGTCCATTAGCCAAATCCTAAATAGTGCACCC
GAGGTTAGGTTGCAATAGCTTGGCAATTCCGCTGGAGAATGAA
CCCAGAATCTCTAGTTTGAGCATCATTCCATCAATCAATCATT
GTCGGAGTCATTCCATCTATGAGAGTGTCTCTGGAACGAACCC
AACGGAGTCGTCAAGTTGTCGCGATCTAACCTTGGGTGCATCA
CCTAGGGTTTGGCTAATGGATTACTAAGCCTCAACTTAGTTCA
GGCTCTCCCA*AGCCCATACCAATTCACGACTTA*

egr-MIR320d   E1626235f.ab1.10.1032_472.523   **TAAGTCTAGATTTAGCTTGGGCT**CTCTCA*AGCCCATACCAATT*
CACGACTTA

egr-MIR320e   E0404437r_053.ab1.1.746_224.275   **TAAGTCGCGAATTAGTATGGGCT**TGGGAG*AGCCCAAATTAAAT*
CTAGGCTTA

egr-MIR320f   E0246027r_019.ab1.47.946_501.552   **TAAGTCGAGAATTGGTACGGGCT**TTGGAG*AGCTCGAACTAAGT*
TTAGACTTA

egr-MIR359a   E1013068f_091.ab1.1.693_170.243   **TCTGTTTAAAAATTGTTCCAGG**GAATAGAAATAAAAAAAGTA
TTTCTGTTC*CGGGGAACAATTTTTAAACAAA*

egr-MIR359b   E0532804f.ab1.192.1215_451.524   **TCTGTTTAAAAATTGTTCCAGG**GAACAAAAATAGAAAAAAGTG
TTTCTGTTC*CCAGGAACAATTTTTGAACAAA*

egr-MIR359c   E0273663f_040.ab1.1.955_401.464   **TCTGTTTAAAAATTGTTCCAGG**GAACAGAAATACTTCTGTTC*C*
CTGGAACAATTTTTGAACAAA

egr-MIR359d   E1447690f.ab1.197.1204_451.524   **TCTGTTCAAAAATTGTTCCAGG**GAACAGAAATAGAAAAAAGTG
TTTCTATTC*TGGGAAACAATTTTTGACCAGA*

egr-MIR359e   E1381605r_022.ab1.1.836_262.336   **TCTGTTCAAAAATTGTTCCAGG**GAACAGAAATAAAAAAAATGT
GTTTCTGTTCC*AAGGAACAATTTTTGACCAGA*

egr-MIR359f   E1196861r.ab1.1.581_58.130   **TCTGTTCAAAAATTGTTCCAGG**GAACAGAAATAAAAAAAATGT
TTTTGTTC*CGGGGAACAATTTTTGACCAAA*

egr-MIR359g   E0570168r.ab1.106.1129_501.574   **TCTGTTCAAAAATTGTTCCAGG**GAACAGAAATAGAAAAAACTA
TTTATGTTC*CGAGGAACAGTTTTTGAACAAA*

egr-MIR359h   E0553137r.ab1.1.945_371.445   **TCTGTTCAAAAATTGTTCCAGG**GAACAGAAATAGAAAAAAATT
ATTTCTGTTC*CGGGGAACAGTTTTTGAACAAA*

egr-MIR362   E0121191r.ab1.1.831_163.331   **TCTTGGAACTCAACATCCTGAA**GTAACTTAAGTTCATAATCCT
TTACTACAGTTGCATGGAATGTACTTCAACAAGACAGGTGAAT
ACAATGATCTTCTTGAAGTACGTTCCCTGCACTGTTGTCAAGG
ATTATGAACTCAAGTTAC*TTTAGGACGTAGAGTTCCAAGA*

egr-MIR159a   E0164506f.ab1.229.1248_337.520   *AGAGCTTTCTTCAGTCCACT* CATGGATGGAATAAAGGGGGTTG
AAATAGCTGCCGACTCATTCATTCAATCACACAGTAGAGAACA
TGGGGCAAATTAGTTTGTGCTTTTACTGGTGACTGTGTGAATG
ATGCGGGAGATACTTTGCACCCCTCTTTTTTCTGTG**CTTGGACT**
GAAGGGAGCTCC

egr-MIR159b   E0085684f.ab1.157.1176_322.520   *AGAGCTTTCTTCGGTCCACA* CACGGGTGGATTTAGGGTTCAAT
TGGCTGCCGACTCATTCATCCAAATACTGAGTTTTATTCATGG
CGGATGTCAAAGTGCAATGGGGCAAATTATTATTGATTGCTCA
GTAAATGAGTGATTGATGCGGGAGACTTGTTGGATCTTAATCT
TCCTGTC**CTTGGACTGAAGGGAGCTCC**

egr-MIR159c   E0505209r.ab1.1.920_402.566   *AGAGCTTCCTTCAGTCCACT* CATGGAGGAATCGGGGTTGAATT
GGCTGCCGACTCATTCATTTGATCAATCAGTAGAAATCATGTG
GATCTACTATTGACCGTGTGAATGATGCGGGAGATAATTTCGC
TCCCTTCTTTTCCTTG**CTTGGACTGAAGGGAGCTCC**

179

egr-MIR159d	E1498862r.ab1.1.976_375.476	*GAAAATCCCTTCATTCTAA*GGCCAACCCTAGAGCTCCCTTCTA CCATATCCCGGTTCCGTAACACTAAACGATACTTTTGGC**CTTA GAATGAAGGGAGCTCT**
egr-MIR159e	E0435958r.ab1.360.1252_355.520	*GGAGCTTTCTTCAGTCCACC*CATGGGACGTGTTGGGTTTTAAT TAGCTGCCGACTCATTCATCCAAATACCGAGCGAGAGCAAGTA ACAGAGCTCCGTAAATGAATGGATGATGCGGGAGTCTTGTTGA TTCCCAAGCTTTCCGTG**ATTGGACTGAAGGGAGCTCC**
egr-MIR159f	E0406035r_051.ab1.1.894_376.586	*GGAGCTCCTTTCAGTCCAATACAGA*GGGCTGAGACGCGGACCG AGCTGCCATCTCATGCACCAGGCTATGCTCAATCTTTTCGGCC ACTAGTTGGAGGGATCAATCGGGTTGGCTAACCAATCAATCAG GGGTTGAGTCTAGCTAGGTGCATGGTGTGGGAGCAACTCCTTC CGTTATACTCCGCTCATTG**ATTGGACTGAAGGGAGCTCC**
egr-MIR160a	E0246270f_094.ab1.194.953_501.590	TA**TGCCTGGCTCCCTGTATGCCA**CAAACATATTCCAACCTTTC ATCTTAGGTTGGCCATTACGTTGG*TGGCGTACCAGGAGCCAAG CCTA*
egr-MIR160b	E0443862f.ab1.360.1295_438.521	TG**TGCCTGGCTCCCTGTATGCCA**TTTGCGTAGCCCACCGGAGA ACCGATGGCCTGTGTGTA*TGGCGTATGAGGAGCCATGCATA*
egr-MIR160c	E0238876r.ab1.1.778_191.278	**TGCCTGGCTCCCTGAATGCCA**TCTAAGAAGCAGTTTTGCCAAA GACAGAGTTGGCAACTTCCTTAGT*TGGCATGAGGGGAGCCATG CA*
egr-MIR160d	E0634921r.ab1.1.942_422.524	**TGCCTGGCTCCCTGAATGCCA**TCTAAGGAGTTGGTTCCAACAG TTGACCAAAGAGAAGGAGGATGTTTGACTTCTTCTTAGT*TGGC ATCAGAGGAGTCATGCA*
egr-MIR168a	E0593603f.ab1.1.689_168.531	GGTCTCTGAT**TCGCTTGGTGCAGGTCGGGA**ACTGATTCGGCGA TTTGATTGCCAGATGGCTCGACATGACTTGAAAGGAAGAAACA GGAAAAAAACAAAGAATAGCGAATTGGA*TCCCGCCTTGCATC AACTGA*ATCGGAGG
egr-MIR168b	E0264641f_065.ab1.255.944_303.522	GGTCTCTAAT**TCGCTTGGTGCAGGTCGGGA**ACTGATTCGGCGA TTTGATTGCCAGATGGCTAAACACGATTGGCTGTGAGGCAATT ATAAAAGAAAGAGAATTGGA*TCCCGCCTTGCATCAACTGA*ATC GGAGA
egr-MIR172a	E0220437r.ab1.1.589_69.215	*GTGTAGCATCATCAAGATTC*ACATATAGTTTTGCATCTTCCTC TTGGGGTAAGATTCCTAAAGATTATACTTGAATTCATGATAAA ATCGTGGGCAAGTGCAAAAGAAAGCTCCGTAGGAAGTGTG**AGA ATCTTGATGATGCTGCAT**
egr-MIR172b	E0162604f.ab1.1.733_213.328	*GTGCAGCATCATTAAGATTCA*CATCGATCCAATTGCCACGGGG CGGCTTGTGTAGATCGTTTGTTCTTGGCGCTTTGTGCAATTTG CTGATGGTG**AGAATCTTGATGATGCTGCAT**
egr-MIR172c	E0058639r_039.ab1.1.946_390.489	*CGTAGCATCATCAAGATTCA*CAAGTGAAAAAGGACGGCTATGA TTGGATCGCCCTTTGCCGAAAGTCCATCTGAGATTG**AGAATCT TGATGATGCTGCAT**
egr-MIR172d	E0182002r.ab1.1.913_315.413	*CGTAGCATCATCAAGATTCA*CAAGTGAAAATGGACGGCGATGA TTGGATCGCCCTTTGCCGAAAGTCCATCTGAGATTG**AGAATCT TGATGATGCTGCA**
egr-MIR172e	E0525459f.ab1.56.1076_423.521	*CGCATCATCATCAAGATTCA*CAAGTGACAATGGACGGAGATGA TTGGATCGCCCTTTGCTAAAAGTCCATCTGAGATTG**ACAATCT TGATGATGCTGCA**
egr-MIR172f	E0381021f.ab1.1.970_450.606	*TTGCAGCATCATGAAGATTC*CCAAAAGAACTTTCTTTTTTGAA GAGAGCGTCTTTTTTAATAATTAACGGCGGTTTAGTTTATGTG GTCCCTAATTCCACCTTAAAAGCGTCTCTCTCCAATATGATTT TACCATG**GGAATCTTGATGATGCTGCAG**

egr-MIR172g	E0173163f.ab1.1.916_396.564	*CTGCATCATCATCAAGATTCA*CAGATCTCTCCTTTCTTTTGCT CTTTTTAGGAGAGGTAATTAGTGGGGTTAAGATTGGGGGATTA ATGGCCCTTTTACAAGGCACTTTTTCACTCCCCTCTCAAGTAA AAGCCTCTTTGATGAATTG**GGAATCTTGATGATGCTGCAG**
egr-MIR472	E0171896f.ab1.1.871_351.447	*GGAATGGGCGGTTTGGGATA*AACCGGAGAAACGACGTTGTCGA GATCTCGGCACCATGAGTTCTTTCGATTAAGTT**TTTCCCAAGG CCGCCCATTCC**

---

[a] The *Eucalyptus* small RNA gene name as specified by egr-MIR. Putative novel egr-miRNA genes and sequences are listed first followed by known egr-miRNA genes.

[b] *E. camaldulensis* genome reference for each small RNA precursor location as determined by shotgun sequencing. Names were designated by the sequence file name followed by the precursor 5' site within the transcript (after the underscore).

[c] Sequence of identified small RNA precursors given in 5' to 3' orientation as shown in Figures 3.1 and 3.2. The small RNA sequences are shown in bold and miRNA* sequences are shown in italics. The sequences presented do not represent full length miRNA precursor genes. Only the stem-loop starting at the cloned sequence and ending at the complementary miRNA* sequence is shown.

CONCLUDING REMARKS

This M.Sc. study has identified the first miRNAs from *Eucalyptus*, and has contributed to the overall number of plant miRNAs. Furthermore, we isolated and characterized miRNAs from different developmental stages of two different forest tree species, in order to identify novel miRNAs unique to these woody plants. We were able to identify over forty putative novel tree miRNAs from the vascular tissues of *Eucalyptus* and *Populus* tissue culture plantlets. The miRNA target genes identified in this study point to a number of plant developmental programmes that may be under miRNA control, including vascular development and plant homeostasis. The isolation of miRNAs from two different tree species, and from a number of different tissues, provides clues to the molecular function of these molecules in commercially important forest trees. These results provide insights into a deeper understanding of the control of forest tree development, in both primary and secondary tissues, and of the underlying molecular mechanisms involved in tree growth.

One of the main aims of this study was to investigate the role of miRNAs during wood formation in *Eucalyptus*. It is becoming clear that miRNAs have roles during tissue patterning of both the primary and secondary vascular bundles from their respective meristem tissues (McHale and Koning, 2004; Kim et al., 2005; Ko et al., 2006). These molecules already have documented roles during aspects of wood formation in *Populus*, where they regulate reaction wood formation (Lu et al., 2005) and control the expression of a class III HD-ZIP protein involved in xylogenesis (Ko et al., 2006). However, there may be other aspects of wood formation under the control of miRNAs that are still to be identified. In this study, the identification of miRNAs from another forest tree genus, *Eucalyptus*, has allowed comparative analysis of miRNAs in different wood forming plants. This has provided a more comprehensive set of miRNAs from trees, and has allowed us to identify a number of putative target genes under miRNA control. Furthermore, the identification of miRNAs from the primary tissues of young, developing whole tissue culture plants allows insights into early

development in trees, such as the development of primary vascular tissues which will ultimately differentiate into the secondary tissues in wood.

Analysis of the target genes identified in this study reveals that miRNAs most likely have a role in higher level regulation of cambial meristem differentiation, for example through the targeting of transcription factor genes (egr-miR31, egr-miR90, egr-miR159, egr-miR172 and egr-miR359) and auxin response factor genes (egr-miR140 and egr-miR160). Other identified miRNA target genes, especially those identified from primary tissues, appear to be involved in more diverse plant pathways. Since we isolated miRNAs from poplar whole plantlets it is difficult to assess which miRNAs are involved in primary vascular patterning specifically until target gene studies are done. However, there is a trend towards targets with roles in plant metabolism and biosynthesis pathways, which could ultimately be involved in plant growth and development, as discussed in Chapter 2.

The miRNA isolation method applied in this study allowed the identification of a number of putative novel miRNAs that have not been identified in the past. Although this approach does not allow the identification of the entire complement of miRNAs from one tissue or species, it is the common approach used in miRNA identification projects of species and tissues that have not been studied previously. From the results it seems that there is a core set of highly conserved miRNAs present in most studied species, including rice, ferns, *Arabidopsis* and trees (http://microrna.sanger.ac.uk). These include the majority of the founding miRNA families such as miR159, miR160, miR166 and miR172. Although these miRNA families, and their functions *in vivo*, are important to understanding plant development, it is the novel miRNAs absent from non-tree species that may be of greatest interest, as these may contribute to a set of tree-specific miRNAs involved in processes and pathways unique to woody plants. The novel miRNAs identified in this study from both *Populus* and *Eucalyptus* could indeed signify tree-specific miRNAs that are absent from

184

Arabidopsis and rice. Without deeper sequencing of miRNAs from these two species, the results obtained in this study cannot be compared directly, but together provide a more comprehensive picture of miRNAs present in these two tree species.

Expression profiling across a number of different tissues provides further information as to where miRNAs are actively regulating gene expression. This allows a more powerful deduction of the molecular function of these molecules. The expression analysis of miRNAs identified in *Eucalyptus* strongly indicates a role for some of these molecules in vascular development. Several of the identified miRNAs exhibit unique expression patterns across the vascular tissues, such as exclusive expression in phloem or immature xylem tissues. Although we have already discussed a possibility of miRNA regulation of higher level patterning events, it may be that these miRNAs are involved in initial meristem differentiation, and continual expression in certain tissues maintains the differentiated state of the tissue in which they are expressed.

It is becoming clear that miRNAs have essential roles during vascular patterning and wood development. However, the use of miRNAs for the direct genetic manipulation of wood properties is unsure. Since it seems that miRNAs have roles in gene regulation upstream of genes involved in the chemical properties of wood, such as cellulose and lignin composition, the use of miRNAs to manipulate these aspects may not be viable at this time. However, other physical properties of wood, such as growth rate, quantity of xylem, and fibre properties, are traits influenced by vascular patterning effects, and thus could be directly affected, and thus manipulated by miRNAs. Furthermore, it has been shown that miRNA precursor molecules can be used in transgene studies and can stably express synthetic miRNAs against any user chosen target gene (Tsuda et al., 2005). Thus, although at this time, endogenous miRNAs themselves may have restricted applications for the genetic manipulation of wood properties, the use of the mechanism of miRNA-mediated gene

regulation may be of significant use for altering other aspects of wood development in trees. It seems that miRNAs have very unique, distinct and specific expression patterns, indicating that these genes are under the regulation of tissue-specific promotors. Although little is known about the molecular mechanisms regulating miRNA expression, it is known that the *MIR* genes are transcribe

ed by RNA polymerase II and contain other motifs in their promotors (Xie et al., 2005). Thus, the identification and use of certain miRNA promotors in transformation experiments will allow gene-function studies using miRNA-promoter driven expression of any transgenes in a very specific niche within the plant.

In plants, over the next few years it is likely that more investments will be made to identify miRNAs from a wide range of tissues, species and conditions, to validate predicted miRNAs and their targets. With the increasing availability of plant genome sequences, comparative genomics approaches are likely to be successful for the discovery of conserved miRNAs. A combination of approaches will be necessary to identify the total number of conserved and non-conserved miRNAs in a genome and to characterize the biological targets of these molecules. With new technologies, the tedious and time-consuming process of sequencing a large number of small RNAs is no longer a limiting factor in the discovery of novel miRNAs. The use of new technologies, such as MPSS and 454 sequencing, allows the sequencing of hundreds of thousands of molecules in a single run (Brenner et al., 2000; Margulies et al., 2005). This will facilitate the identification of miRNAs from virtually any tissue, condition and developmental stage, and subsequently allow the identification of miRNAs with novel functions in plant development (Meyers et al., 2004; Wang et al., 2004; Lu et al., 2006; Maher et al., 2006; Ronemus et al., 2006).

This M.Sc. project represents a pilot study of the identity and role of miRNAs in *Eucalyptus* trees development. These results pave the way for functional characterization of

these molecules in a multitude of future studies. The next step would be to confirm the authenticity of the miRNAs identified here with the use of plants deficient in miRNA precursor processing (i.e.DCL1 mutants), where the accumulation of miRNA precursor molecules is correlated with a corresponding decrease in mature miRNAs (Kurihara and Watanabe, 2004; Liu et al., 2005). Furthermore, the target genes identified here were based on computational predictions, and thus, we cannot say with total confidence that these will all be true miRNA targets. Confirmatory analysis of these targets is necessary, including 5'-RACE cleavage assays, ectopic expression of the miRNAs, and an inverse correlation between target gene and miRNA expression patterns. Once the *Eucalyptus* genome sequencing effort is complete, the use of whole-genome microarrays, using a multitude of conditions and tissues, will allow the coordinate analysis of the expression of all known miRNAs and their target genes simultaneously. This will provide a broader, overall picture of how these molecules function at a whole-tree level and provide information regarding global gene regulatory networks in systems biology approach (Shasha et al., 2001).

In conclusion, this study has successfully provided future researchers with insights into the genetic regulation of developmental processes of trees, especially the molecular mechanisms involved in tissue pattern formation and the differentiation of early meristematic tissues and specialized cell types, such as secondary xylem. The use of the miRNAs, and their target genes, identified in this study in future gene-function studies may contribute to the further elucidation of the genetic regulation of wood development. The information obtained in this study will help to unravel the complex gene regulatory networks involved in the development of trees.

REFERENCES

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S.R., Moon, K., Burcham, T., Pallas, M., DuBridge, R.B., Kirchner, J., Fearon, K., Mao, J., and Corcoran, K. (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat. Biotechnol. **18,** 630-634.

Kim, J., Jung, J.H., Reyes, J.L., Kim, Y.S., Kim, S.Y., Chung, K.S., Kim, J.A., Lee, M., Lee, Y., Narry Kim, V., Chua, N.H., and Park, C.M. (2005). microRNA-directed cleavage of ATHB15 mRNA regulates vascular development in *Arabidopsis* inflorescence stems. Plant J. **42,** 84-94.

Ko, J., Prassinos, C., and Han, K. (2006). Developmental and seasonal expression of *PtaHB1*, a *Populus* gene encoding a class III HD-Zip protein, is closely associated with secondary growth and inversely correlated with the level of microRNA (miR166). New Phytol. **169,** 469-478.

Kurihara, Y., and Watanabe, Y. (2004). *Arabidopsis* microRNA biogenesis through Dicer-like 1 protein functions. Proc. Natl. Acad. Sci. USA **101,** 12753-12758.

Liu, B., Li, P., Li, X., Liu, C., Cao, Y., Chu, C., and Cao, X. (2005). Loss of function of OsDCL1 affects microRNA accumulation and causes developmental defects in rice. Plant Physiol. **139,** 296-305.

Lu, C., Kulkarni, K., Souret, F.F., MuthuValliappan, R., Tej, S.S., Poethig, R.S., Henderson, I.R., Jacobsen, S.E., Wang, W., Green, P.J., and Meyers, B.C. (2006). MicroRNAs and other small RNAs enriched in the *Arabidopsis* RNA-dependent RNA polymerase-2 mutant. Genome Res. **16,** 1276-1288.

Lu, S., Sun, Y.-H., Shi, R., Clark, C., Li, L., and Chiang, V.L. (2005). Novel and mechanical stress-responsive microRNAs in *Populus trichocarpa* that are absent from *Arabidopsis*. Plant Cell **17,** 2186-2203.

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

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., and Chen, Z. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature **437,** 376-380.

McHale, N.A., and Koning, R.E. (2004). MicroRNA-directed cleavage of *Nicotiana sylvestris* PHAVOLUTA mRNA regulates the vascular cambium and structure of apical meristems. Plant Cell **16,** 1730-1740.

Meyers, B.C., Lee, D.K., Vu, T.H., Tej, S.S., Edberg, S.B., Matvienko, M., and Tindell, L.D. (2004). *Arabidopsis* MPSS. An online resource for quantitative expression Aanalysis. Plant Physiol. **135,** 801-813.

Ronemus, M., Vaughn, M.W., and Martienssen, R.A. (2006). MicroRNA-targeted and small interfering RNA-mediated mRNA degradation is regulated by Argonaute, Dicer, and RNA-Dependent RNA polymerase in *Arabidopsis*. Plant Cell **18,** 1559-1574.

Shasha, D.E., Kouranov, A.Y., Lejay, L.V., Chou, M.F., and Coruzzi, G.M. (2001). Using combinatorial design to study regulation by multiple input signals. A tool for parsimony in the post-genomics era. Plant Physiol. **127,** 1590-1594.

Tsuda, N., Kawano, K., Efferson, C., and Ioannides, C. (2005). Synthetic microRNA and double-stranded RNA targeting the 3'-untranslated region of HER-2/neu mRNA inhibit HER-2 protein expression in ovarian cancer cells. Int. J. Oncol. **27,** 1299-1306.

Wang, X.J., Reyes, J.L., Chua, N.H., and Gaasterland, T. (2004). Prediction and identification of *Arabidopsis thaliana* microRNAs and their mRNA targets. Genome Biol. **5,** R65.

Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S.A., and Carrington, J.C. (2005). Expression of *Arabidopsis MIRNA* genes. Plant Physiol. **138,** 2145-2154.

DISSERTATION SUMMARY

MicroRNAs in differentiating tissues of *Populus* and *Eucalyptus* trees

Michelle Victor

*Supervised by **Dr. Alexander A. Myburg and Prof. Henk Huismans***

*Submitted in partial fulfilment of the requirements for the degree **Magister Scientiae***

Department of Genetics

University of Pretoria

Summary

Trees exhibit many unique aspects of plant biology, one of which is the formation of wood. Wood is one of the most important natural products with a multitude of applications. The formation of wood (xylogenesis) is a highly ordered developmental process involving the patterned division and differentiation of the vascular cambium into secondary xylem and phloem tissue types. The progression of xylogenesis developmental process requires differential gene expression across the different tissue types. The tight regulation of wood formation is mediated by genes that regulate cambial meristem differentiation and xylem cell fate. MicroRNAs (miRNAs) are a group of endogenous ~ 20 to 24 nt RNA molecules that down regulate gene expression at the post-transcriptional level. MicroRNAs have validated roles in developmental processes through the regulation of meristem cell differentiation and developmental patterning in plants. They have been shown to spatially regulate differential gene expression patterns at different developmental stages. Thus, the vascular cambium and its derivatives are excellent candidate tissues for miRNA discovery. The aim of this M.Sc. study was to isolate microRNAs from actively differentiating tissues of two tree species in

order to determine possible gene regulatory networks involved in early meristem differentiation, tissue patterning and secondary vascular development.

A small RNA library from two-month old *in vitro* Populus trichocarpa plantlets was constructed to identify putative miRNAs contributing to the early postembryonic development of trees. This library, in conjunction with computational prediction of poplar miRNA homologues and precursor secondary structures, was used to identify a total of 72 poplar miRNAs. Sixteen of these were putative novel miRNAs, belonging to nine new miRNA families. A genome-wide search identified 55 putative target genes for the newly identified miRNAs. The target genes had diverse biological roles in developmental events and maintenance of cellular homeostasis. A number of the predicted targets were involved in plant organ development such as leaf cell fate, floral organ development and meristem differentiation. Other targets were involved in response to hormones, such as growth regulating factors and signaling proteins. Additionally, several targets were related to cellular metabolic processes, such as protein modification and ubiquitination. By isolating miRNAs from developing poplar plantlets, we were able to suggest possible developmental programmes under the control of these molecules, possibly affecting early seedling development and growth.

A similar approach was used to identify miRNAs from three differentiating vascular tissues of *Eucalyptus grandis*. Isolated small RNA sequences were used in a search against all available bacterial artificial chromosome (BAC) shotgun genomic sequences from an ongoing *Eucalyptus camaldulensis* genome sequencing initiative at the Kazusa DNA Research Institute in Japan. We were able to characterize the first *Eucalyptus* miRNAs, and identified 48 putative miRNAs grouping into thirteen gene families. Twenty of the miRNAs belong to five families previously identified in other plant species, whereas the remaining 28 miRNAs grouped into eight putative novel miRNA families. Searches of the *Populus* and

Arabidopsis annotated genomes revealed 45 putative target genes for the new families. Targets of particular interest included transcription factors involved in cell fate determination, including a MADS-box transcription factor involved in xylem formation. Further targets included auxin signaling proteins and auxin response factors, which could play a significant role during auxin regulation of vascular development. Expression profiling of the putative miRNAs using quantitative RT-PCR revealed that a number of the miRNAs exhibited differential expression patterns across xylogenic and non-xylogenic tissues. One miRNA showed expression in a single vascular tissue, whereas others were expressed at varying levels across the vascular tissues. This observation indicates a possible role for these putative miRNAs during vascular development and differentiation in eucalypt trees.

In this study we used a combination of small RNA library construction and computational prediction to identify microRNAs from two tree species. We identified a total of 120 putative miRNAs grouping into 31 families. Of these, 44 group into 17 putative novel tree-specific miRNAs. This study has allowed the identification of novel miRNAs from a unique set of tissues, and has contributed to the ever-growing number of plant-specific miRNAs. The results of this study further contribute to our expanding knowledge of the unique developmental process of vascular tissue differentiation of perennial woody plants such as *Eucalyptus* and *Populus* species.