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

© University of Pretoria
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
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:


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 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).
• To the Sequencing Facility at the University of Pretoria, Renate Zipfel and Gladys Shabangu, for fast and efficient service during the progress of this project.

• To the Genetics Department of the University of Pretoria and the Forestry and Agricultural Biotechnology Institute (FABI) for providing facilities and a sound academic environment.

• To the National Research Foundation of South Africa (NRF), for the Prestige scholarship awarded throughout the duration of this project and for funding of the project through the WFMG programme.

• To The Human Resources and Technology for Industry Programme (THRIP) for financial support of the WFMG programme.
# TABLE OF CONTENTS

DECLARATION........................................................................................................................................ ii  
PREFACE............................................................................................................................................. iii  
ACKNOWLEDGEMENTS ........................................................................................................................ vii  
TABLE OF CONTENTS ........................................................................................................................ ix  
LIST OF FIGURES ................................................................................................................................ xii  
LIST OF TABLES ................................................................................................................................ xii  
LIST OF ABBREVIATIONS .................................................................................................................. xiv  

CHAPTER 1 ........................................................................................................................................... 1  
LITERATURE REVIEW - MICRORNAS AND THEIR ROLE IN PLANT DEVELOPMENT ................................................................................................................................. 1  

1.1 Introduction ..................................................................................................................................... 2  

1.2 The discovery of RNA Silencing ................................................................................................. 3  

1.3 Biogenesis of plant miRNAs........................................................................................................ 7  

1.3.1 Plant miRNA genes .............................................................................................................. 7  

1.3.2 Plant miRNA processing ..................................................................................................... 8  

1.3.3 The silencing complex ....................................................................................................... 12  

1.3.4 Plant miRNA expression ................................................................................................... 14  

1.4 Mechanism of action ................................................................................................................ 16  

1.4.1 mRNA cleavage................................................................................................................ 17  

1.4.2 Translational repression ................................................................................................... 18  

1.5 MicroRNAs vs siRNAs............................................................................................................ 19  

1.6 MicroRNAs in plant development .......................................................................................... 21  

1.6.1 Organ polarity and meristem identity ............................................................................ 24  

1.6.2 Floral organ identity, flowering time and transitions in the plant life cycle ............. 27  

1.6.3 MicroRNA Biogenesis .................................................................................................. 29  

1.6.4 Stress-response genes and non-transcription factors ..................................................... 30  

1.7 Recent developments................................................................................................................ 32  

1.8 Xylogenesis: A possible target? ............................................................................................ 34  

1.11 References ............................................................................................................................ 37
3.3.5 Prediction of targets for *Eucalyptus* miRNAs ....................................................... 133
3.4 Results .......................................................................................................................... 133
  3.4.1 Isolation and sequence analysis of small RNAs from *Eucalyptus grandis* stem tissues ............................................................................................................................. 133
  3.4.2 Identification of putative *Eucalyptus* miRNAs ..................................................... 134
  3.4.3 Prediction of putative targets for *Eucalyptus* miRNAs ....................................... 137
  3.4.4 Expression profiles of putative egr-miRNAs in *Eucalyptus* ................................. 138
3.5 Discussion .................................................................................................................... 140
  3.5.1 Characteristics of putative *Eucalyptus* miRNAs ................................................ 142
  3.5.2 Predicted targets of egr-miRNAs have diverse biological functions ..................... 143
  3.5.3 *Eucalyptus* miRNAs could have a putative role in vascular development............. 147
3.6 Acknowledgements .................................................................................................... 150
3.7 Figures .......................................................................................................................... 151
3.8 Tables ........................................................................................................................... 162
3.9 References .................................................................................................................. 169
3.10 Supplementary material ........................................................................................... 177

CONCLUDING REMARKS ................................................................................... 182

References ...................................................................................................................... 188

DISSERTATION SUMMARY ................................................................................ 190
LIST OF FIGURES

Figure 1.1. Model for the biogenesis of small RNAs in *Arabidopsis*…………………………9

Figure 1.2. Structure and Function of *Arabidopsis* miR165…………………………………….26

Figure 2.1. Predicted stem-loop structures of non-conserved *P. trichocarpa* ptc-miRNAs.................................................................................................................................82

Figure 2.2. Predicted stem-loop structures of conserved *P. trichocarpa* ptc-miRNAs.................................................................................................................................85

Figure 2.3. Precursor alignments of selected ptc-miRNA families .................................95

Figure 3.1. Predicted precursor stem-loop structures of putative novel *Eucalyptus grandis* egr-miRNAs..............................................................................................................151

Figure 3.2. Predicted precursor stem-loop structures of *Eucalyptus grandis* miRNAs from previously identified families.................................................................156

Figure 3.3. Expression profiles of *Eucalyptus* miRNAs......................................................160
LIST OF TABLES

Table 1.1. Certain *Arabidopsis* miRNAs and their putative targets………………………23

Table 2.1. Newly identified microRNAs and gene family members from *Populus* plantlets
........................................................................................................................................100

Table 2.2. *Populus* microRNAs and gene family members from known families ..........101

Table 2.3. Putative targets and predicted gene functions for *Populus* microRNAs ........103

Table S2.1. Putative endogenous siRNAs cloned from *Populus*..................................113

Table S2.2. miRNA stem-loop locations and precursor sequences in *Populus*..........119

Table 3.1. Putative novel microRNAs cloned from vascular tissues of *Eucalyptus* and gene
family members identified in the *E. camaldulensis* genome sequence .......................162

Table 3.2. *Eucalyptus* miRNAs conserved in other species and gene family members
identified in the *E. camaldulensis* genome sequence..................................................164

Table 3.3. Putative targets and predicted gene functions for putative novel egr-miRNAs in
*Populus* and *Arabidopsis*..........................................................................................165

Table 3.4. Putative targets and predicted gene functions for known egr-miRNAs in *Populus*
and *Arabidopsis*...........................................................................................................167

Table S3.1. Primers used for *Eucalyptus* miRNA expression analysis.......................177

Table S3.2. Precursor sequences of putative novel and conserved miRNAs identified in the *E.
camaldulensis* genome sequences ..............................................................................178
# LIST OF ABBREVIATIONS

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

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 promoters, 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, polyadenylated, 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
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
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
Literature Review

(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 \textit{HASTY (HST)}, which is a member of the nucleocytoplasmic transporter family of proteins (Bollman et al., 2003; Park et al., 2005). \textit{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 \textit{C. elegans}, and SDE3 of \textit{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). Agol 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 cloning 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
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
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., 2004b).
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 promoter 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*

<table>
<thead>
<tr>
<th>miRNA family</th>
<th>Predicted target gene family</th>
<th>Predicted Target Gene</th>
<th>Predicted target gene function</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR156</td>
<td>Squamosa-promoter Binding Protein (SBP)-like transcription factors</td>
<td>10 SPL genes</td>
<td>Floral meristem identity and flowering time</td>
</tr>
<tr>
<td>miR157</td>
<td>SBP-like proteins</td>
<td>9 SPL genes</td>
<td>Floral meristem identity and flowering time</td>
</tr>
<tr>
<td>miR159a</td>
<td>MYB transcription factors</td>
<td>5 MYB genes, <em>GAMYB</em></td>
<td>Plant growth, anther development and flowering time</td>
</tr>
<tr>
<td>miR159b</td>
<td>MYB transcription factors</td>
<td>3 MYB genes, <em>MYB33, MYB65</em></td>
<td>As above</td>
</tr>
<tr>
<td>miR-JAW</td>
<td>TCP transcription factors</td>
<td>TCP2, TCP3, TCP4, TCP10, TCP24</td>
<td>Cell division, leaf development, and embryonic patterning</td>
</tr>
<tr>
<td>miR160</td>
<td>Auxin response factors</td>
<td><em>ARF10, ARF16, ARF17</em></td>
<td>Plant response to phytohormones during development</td>
</tr>
<tr>
<td>miR161</td>
<td>Pentatricopeptide repeat proteins</td>
<td>9 genes</td>
<td>Organelle metabolism</td>
</tr>
<tr>
<td>miR162</td>
<td>DICER</td>
<td><em>DCL1</em></td>
<td>Biogenesis of miRNAs and control of floral meristem development</td>
</tr>
<tr>
<td>miR163</td>
<td>SAM-dependant methyltransferases</td>
<td>5 genes</td>
<td>Control of boundary size in meristems and formation and separation of embryonic, vegetative and floral organs</td>
</tr>
<tr>
<td>miR164</td>
<td>NAC domain proteins</td>
<td><em>CUC1, CUC2, NAC1, 2 others</em></td>
<td>As above</td>
</tr>
<tr>
<td>miR165</td>
<td>HD-ZIP transcription factors</td>
<td><em>PHB, PHF, REV, ATHB8</em></td>
<td>Embryo patterning, postembryonic meristem initiation, axial meristem initiation, vascular development, leaf polarity and meristem size</td>
</tr>
<tr>
<td>miR166</td>
<td>HD-ZIP transcription factors</td>
<td><em>ATHB15</em></td>
<td>As above</td>
</tr>
<tr>
<td>miR167</td>
<td>Auxin response factors</td>
<td><em>ARF6, ARF8</em></td>
<td>Auxin signalling</td>
</tr>
<tr>
<td>miR168</td>
<td>ARGONAUTE</td>
<td><em>AGO1</em></td>
<td>miRNA biogenesis and pathway, stem cell function and organ formation</td>
</tr>
<tr>
<td>miR169</td>
<td>CCAAT-binding factor (CBF)-HAP2-like proteins</td>
<td>5 genes</td>
<td>Promotor DNA binding protein in Eukaryotes</td>
</tr>
<tr>
<td>miR170</td>
<td>GRAS domain transcription factors (SCARECROW-like)</td>
<td><em>SCL6-II, SCL6-III, SCL6-IV</em></td>
<td>Radial patterning in roots</td>
</tr>
<tr>
<td>miR171</td>
<td>GRAS domain transcription factors (SCARECROW-like)</td>
<td><em>SCL1, SCL6-II, SCL6-III, SCL6-IV</em></td>
<td>Radial patterning in roots</td>
</tr>
<tr>
<td>miR172</td>
<td>APETALA2-like transcription factors</td>
<td><em>AP2, TOE1, TOE2, TOE3</em></td>
<td>Specification of flower organ identity and flowering time</td>
</tr>
<tr>
<td>miR393</td>
<td>F-box proteins</td>
<td><em>TIR1</em> and 3 others</td>
<td>Component of SCF ubiquitin-ligase complexes for ubiquitin-mediated proteolysis</td>
</tr>
<tr>
<td>miR395</td>
<td>bHLH transcription factor</td>
<td>1 gene</td>
<td>Circadian rhythm control by light</td>
</tr>
<tr>
<td>miR396</td>
<td>ATP sulfurylase (3)</td>
<td><em>APS4</em> and 2 others</td>
<td>Sulphur metabolism</td>
</tr>
<tr>
<td>miR397</td>
<td>Growth Regulating Factor (GRF) transcription factors</td>
<td><em>GRL1, GRL2, GRL3, GRL7, GRL8, GRL9</em></td>
<td>Growth regulation</td>
</tr>
<tr>
<td>miR398</td>
<td>Laccase</td>
<td>3 genes</td>
<td>Normal cell wall structure and integrity of xylem fibers</td>
</tr>
<tr>
<td>miR399</td>
<td>Beta-6 tubulin</td>
<td>1 gene</td>
<td>Involved in microtubule based processes and response to cold</td>
</tr>
<tr>
<td>miR398</td>
<td>Copper superoxide dismutases</td>
<td><em>CSD1, CDS2</em></td>
<td>Involved in programmed cell death and apoptosis</td>
</tr>
</tbody>
</table>


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 axillary 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...
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* inflorescence 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).

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, lycopods 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, \textit{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 \textit{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 \textit{Arabidopsis}, snapdragon (Weir et al., 2004) and petunia (Souer et al., 1996) which suggests that regulation by miR165 may be specific to flowering plants.

\subsection*{1.6.2 Floral organ identity, flowering time and transitions in the plant life cycle}

The \textit{Arabidopsis} \textit{APETELA2 (AP2)} gene is known to be regulated by miR172 (Aukerman and Sakai, 2003; Chen, 2004). The \textit{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 \textit{AP2} was the first example of miRNA control through translational repression in plants. Although miR172 binds with near perfect complementarity to the \textit{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 \textit{AP2}-like gene family are also targets of miR172. These genes, \textit{TOE1} and \textit{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 \textit{AP2}-like homologue, \textit{Glossy15}, has also been shown to be regulated by miR172 (Lauter et al., 2005). \textit{Glossy15} regulates the transition from juvenile to adult leaf identity, where \textit{Glossy15} keeps the leaf in juvenile form and miR172-directed degradation results in transition to adult leaf fate (Lauter et al., 2005). Thus, miR172...
Literature Review 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)
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, microRNAs 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.,
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
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


CHAPTER 2

NOVEL MICRONNAS WITH DIVERSE ROLES DURING EARLY DEVELOPMENT OF POPULUS TRICHOCARPA PLANTLETS

Michelle Victor1, Shanfa Lu2, Ying-Hsuan Sun2, Vincent L. Chiang2 and Alexander A. Myburg1

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.,
MicroRNAs from Populus (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 MgCl₂, 1 mM spermidine, 0.1 mM ZnCl₂, 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’ – CTGTAGCACTTCAATTCATC – 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 MgCl₂, 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 MgCl₂, 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’–AGTCGTGAggcacagtgaac– 3’, uppercase DNA, lowercase RNA, underlined bases indicate BanI restriction enzyme site; Dhharmacon, 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
RT primer (5’ – GATGAATGGTGCTAC – 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 MgCl₂, 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 GGCACCGAAA – 3’, underlined bases indicate BanI restriction enzyme site) and the 3’ PCR primer (5’ – GATGAATGGTGCTACAG – 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 purifed 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 MgCl₂, 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
MicroRNAs from Populus 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,
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).
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 prediction 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, ULTRAPETALAI (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
MicroRNAs from Populus 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 polymerization (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 dependent 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 stimulated 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

I would like to thank Dr. S. Lu, Dr. Y-H. Sun, and Prof. V. Chiang for providing invaluable training and facilities at NC State University for the successful completion of this study. I would also like to thank Mr. M. Ranik and Mr. S.A. Garcia for insight and critical review of the manuscript. This work was supported by funding provided by the National Research Foundation (NRF), the University of Pretoria, the Forest Biotechnology Group at NC State University, and the Forest Molecular Genetics Programme at the University of Pretoria.
2.7 FIGURES

ptc-miR26a  UCUUCAUGCUUGUGUUAAA

ptc-miR26b  UCUUCAUGCUUGUGUUAAA

ptc-miR39a  GCAUCCAUGGUUGAAUAGUUA

ptc-miR39b  GCAUCCAUGGUUGAAUAGUUA

ptc-miR44  UCUUUCACGCCUCUCCAUAC

ptc-miR56a  UGGUGGUGCAUGGCCGUUCU
ptc-miR56b UGGUGGUGCAUGGCGUUCUU

AGCGAGUAGCGAUGCUAGUGAAUCAGGGAG AGU 200 nt
LOOP

ptc-miR56c UGGUGGUGCAUGGCGUUCUU

ptc-miR56d UGGUGGUGCAUGGCGUUCUU

AGAA GGAGCUAGGAGAGUA GCAGAUAGAGG AAAAUUAUUAAUU
LOOP

ptc-miR69a UCUUUGCCUACUCCUCCCUAU

ptc-miR69b UUGCCUACUCUCCCAUUCC

ptc-miR81 UGCAACAAUUGGAGAAGUAAUGGU

G A A G GUCUC AUCUAG AAGA UGCUAGUAGAAUGGAGAGUA GCAGAUAGAGG AAAAUUAUUAAUU
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-miR156e  UUGACAGAAGAGAGUGAGCAC

ptc-miR156g  UUGACAGAAGAUAGAGAGCAGC

ptc-miR156h  UUGACAGAAGAUAGAGAGCAGC

ptc-miR156i  UUGACAGAAGAUAGAGAGCAGC

ptc-miR156j  UUGACAGAAGAUAGAGAGCAGC

ptc-miR159a  UUUGGAUUGAAGGGAGCUCU

ptc-miR159b  UUUGGAUUGAAGGGAGCUCU
ptc-miR166g UCGGACCAGGCUUCAUUCCCC

ptc-miR166h UCGGACCAGGCUUCAUUCCCC

ptc-miR166i UCGGACCAGGCUUCAUUCCCC

ptc-miR166j UCGGACCAGGCUUCAUUCCCC

ptc-miR166k UCGGACCAGGCUUCAUUCCCC

ptc-miR166l UCGGACCAGGCUUCAUUCCCC
ptc-miR166m  UCGGACCAGGCUUCAUCCCU

ptc-miR166n  UCGGACCAGGCUUCAUCCUU

ptc-miR166o  UCGGACCAGGCUUCAUCCUU

ptc-miR166p  UCGGACCAGGCUUCAUCCUU
ptc-miR167a  UGAAGCUGCCAGCAUGAUCUA

AA A -C- U U U AGAAA

ptc-miR167b  UGAAGCUGCCAGCAUGAUCUA

ptc-miR167c  UGAAGCUGCCAGCAUGAUCUA

ptc-miR167d  UGAAGCUGCCAGCAUGAUCUA

ptc-miR167e  UGAAGCUGCCAGCAUGAUCUA

ptc-miR167f  UGAAGCUGCCAGCAUGAUCUA
MicroRNAs from Populus

**ptc-miR167g** UGAAGCUGCCAGCAUGAUCUG

| 5' | UCAUGC-ACCA UAGUAG UGAAGCUGCCAGCAUGAUCUG | CUUuccu | U |
| 3' | GGUACGAUGGU GUUAUC ACUUUGAAGGU-GUACUAGAU GAAGGG |
|    | A       U       CA      CAUAUC |

**ptc-miR167h** UGAAGCUGCCAACAUCAUGAUCUG

| 5' | UCAUGC-ACCG UA UAG UGAAGCUGCCAACAUCAUGAUCUG | CUUuccu | U |
| 3' | GGUAGGAUGGU GUUAUC ACUUUGAAGGU-GUACUAGAU GAAGGG |
|    | A       U       CA      CAUAUC |

**ptc-miR167i** UGAAGCUGCCAGCAUGAUCUG

| 5' | UCAUGC-ACCA UAGUAG UGAAGCUGCCAGCAUGAUCUG | CUUuccu | U |
| 3' | GGUACGAUGGU GUUAUC ACUUUGAAGGU-GUACUAGAU GAAGGG |
|    | A       U       CA      CAUAUC |

**ptc-miR396a** UUCCACAGCUUUCUUGAACUG

| 5' | UGA CCU UU GUAAUUCCACAGCUUUCUUGAACUGCUAUUACU GUUUUG-AUGUUGCU U |
| 3' | ACU GGA AA CAUA AGGGUGUGCAAA ACGUUGCGGUG GUUAUG ACAGUAUACACUGA |

**ptc-miR396b** UUCCACGCGCUUUCUUGAACUG

| 5' | UGA CCU UU GUAAUUCCACGCGCUUUCUUGAACUGCUAUUACU GAGU-UAAGUGUGAUGGU UG U |
| 3' | ACU GGA AA CAUA AGGGUGUGCAAA ACGUUGCGGUG GUUAUG UACAGUACACGUCA |

**ptc-miR396c** UUCCACGCGCUUUCUUGAACU

| 5' | CAUG UU UUCCACGCGCUUUCUUGAACU |
| 3' | GUAC AGAGGGUGUGCGAAGAACU |

**ptc-miR396d** UUCCACGCGCUUUCUUGAACU

| 5' | CAUG UU UUCCACGCGCUUUCUUGAACU |
| 3' | GUAC AGAGGGUGUGCGAAGAACU |

---

90
ptc-miR472a  UUUUCCCUACUCACCCAUCC

UU G C C G UUU U UG
5' GGAU UCGGAAG UUAU GGGUGGGUG AG GGGAAGAUAA C GGU UU A
3' CUUA AGCCUUU GAUA CCAUCCCCACCC CCCCUUUAUUG UCA GA G
CU G C C AU -U- U UA

ptc-miR472b  UUUUCCCUACUCACCCAUCC

UU G GU G U
5' GGGU UCGGAAG UUA GGAUGGGUG AGGGAAGAUACUAAGC \ C
3' CUUA AGCCUUU GAU CCAUCCCCACCC CCCCUUUAUUG-UUUG /
CU G AC U U

ptc-miR475a  UUACAGUGCCCAUUGAUUAAG

UAAC
AGCAA U
/ |||| U
UCGUU C
A C GAU U / / UUCU
5' GUUUAU CAUCUUGAUCUACGc GUAUGUAAGAGUGAAG CCA GA A
/ |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||
**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.
MicroRNAs from Populus

ptc-MIR396d  -AAGTCCTG----GTCACTGCTTCTCCACGTCATTCTGCTGTTGTCCTTCTCTCTCTCTCTGCTT
ptc-MIR396e  -AAGTCCTG----GTCACTGCTTCTCCACGTCATTCTGCTGTTGTCCTTCTCTCTCTGCTT
ptc-MIR396f  -AAGTCCTG----GTCACTGCTTCTCCACGTCATTCTGCTGTTGTCCTTCTCTCTCTGCTT
ptc-MIR396g  -AAGTCCTG----GTCACTGCTTCTCCACGTCATTCTGCTGTTGTCCTTCTCTCTCTGCTT
ptc-MIR396h  -AAGTCCTG----GTCACTGCTTCTCCACGTCATTCTGCTGTTGTCCTTCTCTCTCTGCTT
ptc-MIR396i  -AAGTCCTG----GTCACTGCTTCTCCACGTCATTCTGCTGTTGTCCTTCTCTCTCTGCTT
ptc-MIR396j  -AAGTCCTG----GTCACTGCTTCTCCACGTCATTCTGCTGTTGTCCTTCTCTCTCTGCTT
ptc-MIR396k  -AAGTCCTG----GTCACTGCTTCTCCACGTCATTCTGCTGTTGTCCTTCTCTCTCTGCTT
ptc-MIR396l  -AAGTCCTG----GTCACTGCTTCTCCACGTCATTCTGCTGTTGTCCTTCTCTCTCTGCTT
ptc-MIR396m  -AAGTCCTG----GTCACTGCTTCTCCACGTCATTCTGCTGTTGTCCTTCTCTCTCTGCTT
ptc-MIR396n  -AAGTCCTG----GTCACTGCTTCTCCACGTCATTCTGCTGTTGTCCTTCTCTCTCTGCTT
ptc-MIR396o  -AAGTCCTG----GTCACTGCTTCTCCACGTCATTCTGCTGTTGTCCTTCTCTCTCTGCTT
ptc-MIR396p  -AAGTCCTG----GTCACTGCTTCTCCACGTCATTCTGCTGTTGTCCTTCTCTCTCTGCTT
ptc-MIR396q  -AAGTCCTG----GTCACTGCTTCTCCACGTCATTCTGCTGTTGTCCTTCTCTCTCTGCTT
ptc-MIR396r  -AAGTCCTG----GTCACTGCTTCTCCACGTCATTCTGCTGTTGTCCTTCTCTCTCTGCTT
ptc-MIR396s  -AAGTCCTG----GTCACTGCTTCTCCACGTCATTCTGCTGTTGTCCTTCTCTCTCTGCTT
ptc-MIR396t  -AAGTCCTG----GTCACTGCTTCTCCACGTCATTCTGCTGTTGTCCTTCTCTCTCTGCTT
ptc-MIR396u  -AAGTCCTG----GTCACTGCTTCTCCACGTCATTCTGCTGTTGTCCTTCTCTCTCTGCTT
ptc-MIR396v  -AAGTCCTG----GTCACTGCTTCTCCACGTCATTCTGCTGTTGTCCTTCTCTCTCTGCTT
ptc-MIR396w  -AAGTCCTG----GTCACTGCTTCTCCACGTCATTCTGCTGTTGTCCTTCTCTCTCTGCTT
ptc-MIR396x  -AAGTCCTG----GTCACTGCTTCTCCACGTCATTCTGCTGTTGTCCTTCTCTCTCTGCTT
ptc-MIR396y  -AAGTCCTG----GTCACTGCTTCTCCACGTCATTCTGCTGTTGTCCTTCTCTCTCTGCTT
ptc-MIR396z  -AAGTCCTG----GTCACTGCTTCTCCACGTCATTCTGCTGTTGTCCTTCTCTCTCTGCTT
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
<table>
<thead>
<tr>
<th>Ptc-miRNA&lt;sup&gt;a&lt;/sup&gt;</th>
<th>MiRNA Sequence (5’-3’)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Length (nt)&lt;sup&gt;c&lt;/sup&gt;</th>
<th>MiRNA Genome Location&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Arm&lt;sup&gt;e&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ptc-miR26a</td>
<td>UCUUCAUGCUUGGUUUAAAA</td>
<td>(1)</td>
<td>scaffold_628:11840-11822</td>
<td>3’</td>
</tr>
<tr>
<td>Ptc-miR26b</td>
<td>UCUUCAUGCUUGGUUUAAAA</td>
<td></td>
<td>LG_XIII:4013878-4013860</td>
<td>3’</td>
</tr>
<tr>
<td>Ptc-miR26c</td>
<td>GCAUCCAGGUUAUAGUAAA</td>
<td>(2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ptc-miR26d</td>
<td>GCAUCCAGGUUAUAGUAAA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ptc-miR39a</td>
<td>GCAUCCAGGUUAUAGUAAA</td>
<td>(2)</td>
<td>LG_XVI:12106649-12106126</td>
<td>5’</td>
</tr>
<tr>
<td>Ptc-miR39b</td>
<td>GCAUCCAGGUUAUAGUAAA</td>
<td></td>
<td>LG_XVI:1207926-1207926</td>
<td>5’</td>
</tr>
<tr>
<td>Ptc-miR44</td>
<td>UUUUUCAACGUUCCGCUUUA</td>
<td>(1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ptc-miR56a</td>
<td>UGGUGGUGCAUGCGCUUUCUU</td>
<td></td>
<td>scaffold_25197:306-307</td>
<td>3’</td>
</tr>
<tr>
<td>Ptc-miR56b</td>
<td>UGGUGGUGCAUGCGCUUUCUU</td>
<td></td>
<td>scaffold_6563:1755-1774</td>
<td>3’</td>
</tr>
<tr>
<td>Ptc-miR56c</td>
<td>UGGUGGUGCAUGCGCUUUCUU</td>
<td></td>
<td>scaffold_520:40935-40954</td>
<td>3’</td>
</tr>
<tr>
<td>Ptc-miR56d</td>
<td>UGGUGGUGCAUGCGCUUUCUU</td>
<td></td>
<td>scaffold_196:120430-120409</td>
<td>3’</td>
</tr>
<tr>
<td>Ptc-miR69a</td>
<td>UUUUCAAAGCGCUUCCUUC</td>
<td>(1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ptc-miR69b</td>
<td>UUUUCAAAGCGCUUCCUUC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ptc-miR81</td>
<td>UGGUGGUGCAUGCGCUUUCUU</td>
<td></td>
<td>scaffold_129:575580-575599</td>
<td>5’</td>
</tr>
<tr>
<td>Ptc-miR99</td>
<td>UGGUGGUGCAUGCGCUUUCUU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ptc-miR170</td>
<td>UUGAGUCCAGCCGCGAGGG</td>
<td>(3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ptc-miR235a</td>
<td>UUGAGUCCAGCCGCGAGGG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ptc-miR235b</td>
<td>UUGAGUCCAGCCGCGAGGG</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> 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.

<sup>b</sup> 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.

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

<sup>d</sup> 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.

<sup>e</sup> 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
<table>
<thead>
<tr>
<th>ptc-miRNA(^a)</th>
<th>miRNA Sequence (5' - 3')(^b)</th>
<th>Length(^c) (nt)</th>
<th>miRNA genome location(^d)</th>
<th>Arm(^e)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UUGACAGAAGAUGAGACAC (5)</td>
<td>21</td>
<td>LG_XVIII:1365515-1365535</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR156d</td>
<td>UUGACAGAAGAUGAGACAC</td>
<td>21</td>
<td>LG_XVIII:13618085-13618110</td>
<td>5'</td>
</tr>
<tr>
<td>ptc-miR156e</td>
<td>UUGACAGAAGAUGAGACAC</td>
<td>LG_II:21820227-2182040</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR156h</td>
<td>UUGACAGAAGAUGAGACAC</td>
<td>LG_II:21820427-21820447</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR156i</td>
<td>UUGACAGAAGAUGAGACAC</td>
<td>LG_II:21820448-21820468</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR156j</td>
<td>UUGACAGAAGAUGAGACAC</td>
<td>LG_II:21820469-21820489</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>UUGAAUGAAGGAGACUCUA (5)</td>
<td>20</td>
<td>LG_XII:4201679-4201678</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR159a</td>
<td>UUGAAUGAAGGAGACUCUA</td>
<td>LG_XV:3478615-3478634</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR159b</td>
<td>UUGAAUGAAGGAGACUCUA</td>
<td>scaffold_1140:8367-8386</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR159c</td>
<td>UUGAAUGAAGGAGACUCUA</td>
<td>LG_I:6097079-6097060</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR166a</td>
<td>UGGACAGCCUGCUAUCUCCC</td>
<td>21</td>
<td>LG_VIII:11091614-11091595</td>
<td>3'</td>
</tr>
<tr>
<td>ptc-miR166b</td>
<td>UGGACAGCCUCUCAUCCCC</td>
<td>scaffold_122:445122-445141</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR166c</td>
<td>UGGACAGCCUCUCAUCCCC</td>
<td>LG_II:3057594-3057574</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR166d</td>
<td>UGGACAGCCUCUCAUCCCC</td>
<td>LG_V:14949534-14949554</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR166e</td>
<td>UGGACAGCCUCUCAUCCCC</td>
<td>LG_II:3054996-3054976</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR166f</td>
<td>UGGACAGCCUCUCAUCCCC</td>
<td>LG_II:3057597-3057599</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR166g</td>
<td>UGGACAGCCUCUCAUCCCC</td>
<td>LG_X:8492342-8492361</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR166h</td>
<td>UGGACAGCCUCUCAUCCCC</td>
<td>LG_XIII:4057479-4057498</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>UGAACUGCCAGGAGACUCAUCUAUCU (1)</td>
<td>21</td>
<td>LG_VI:17347052-17347032</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR167a</td>
<td>UGAACUGCCAGGAGACUCAUCUAUCU</td>
<td>scaffold_14274:285-265</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR167b</td>
<td>UGAACUGCCAGGAGACUCAUCUAUCU</td>
<td>LG_V:14949534-14949554</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR167c</td>
<td>UGAACUGCCAGGAGACUCAUCUAUCU</td>
<td>LG_II:3054996-3054976</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR167d</td>
<td>UGAACUGCCAGGAGACUCAUCUAUCU</td>
<td>scaffold_10:484760-484740</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR167e</td>
<td>UGAACUGCCAGGAGACUCAUCUAUCU</td>
<td>LG_X:9042289-9042269</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR167f</td>
<td>UGAACUGCCAGGAGACUCAUCUAUCU</td>
<td>LG_XIII:2740303-2740323</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR167g</td>
<td>UGAACUGCCAGGAGACUCAUCUAUCU</td>
<td>scaffold_10:484760-484740</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>UUCCACAGCUUUCAUAGAUSCU (4)</td>
<td>21</td>
<td>LG_V:14075653-14075673</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR396a</td>
<td>UUCCACAGCUUUCAUAGAUSCU</td>
<td>LG_XVIII:12249333-12249353</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR396b</td>
<td>UUCCACAGCUUUCAUAGAUSCU</td>
<td>LG_VI:6490267-6490247</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR396c</td>
<td>UUCCACAGCUUUCAUAGAUSCU</td>
<td>LG_V:9042289-9042269</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR396d</td>
<td>UUCCACAGCUUUCAUAGAUSCU</td>
<td>LG_VIII:12256356-12256336</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR396e</td>
<td>UUCCACAGCUUUCAUAGAUSCU</td>
<td>LG_III:14406867-14406887</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR396f</td>
<td>UUCCACAGCUUUCAUAGAUSCU</td>
<td>LG_VI:14075653-14075673</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR396g</td>
<td>UUCCACAGCUUUCAUAGAUSCU</td>
<td>LG_VIII:12249333-12249353</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR396h</td>
<td>UUCCACAGCUUUCAUAGAUSCU</td>
<td>LG_III:14406867-14406887</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miRNA</td>
<td>Sequence</td>
<td>Genome Location</td>
<td>Orientation</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>--------------------------------------------------------------------------</td>
<td>----------------------------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>ptc-miR396i</td>
<td>UCCCAACGCUUUAAUGAACCG</td>
<td>LG_VI:4075737-4075737</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR396j</td>
<td>UCCCAACGCUUUAAUGAACCG</td>
<td>LG_XVIII:12249447-12249427</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR396k</td>
<td>UCCCAACGCUUUACUGACUU</td>
<td>LG_XVIII:12256230-12256250</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR398a</td>
<td>UGGUGUUCAGGGUCACCCCUG</td>
<td>LG_X:14819174-14819194</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR398b</td>
<td>UGGUGUUCAGGCACCCCG</td>
<td>LG_I:19895303-19895283</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR398c</td>
<td>UGGUGUUCAGGCACCCCG</td>
<td>LG_IX:6181890-6181910</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR408</td>
<td>AUGCAUCUCCUUUCUCCCCUG</td>
<td>LG_I:16909463-16909443</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR472a</td>
<td>UUUUCCCCUACUCCACCCAUCC</td>
<td>scaffold_131:138364-138385</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR472b</td>
<td>UUUUCCCCUACUCCACCCAUCC</td>
<td>LG_XIII:203728-203749</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR475a</td>
<td>UUACAGUGCCCAUUGAUUAG</td>
<td>LG_I:20949518-20949538</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR475b</td>
<td>UUACAGUGCCCAUUGAUUAG</td>
<td>LG_VIII:13883578-13883558</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR475c</td>
<td>UUACAAUGCCCAUUGAUUAG</td>
<td>LG_X:12045864-12045884</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>ptc-miR475d</td>
<td>UUACAGUGCCCAUUGAUUAG</td>
<td>LG_VIII:8026896-8026876</td>
<td>3'</td>
<td></td>
</tr>
</tbody>
</table>

- **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.

<table>
<thead>
<tr>
<th>Common function</th>
<th>ptc-miR family</th>
<th>Protein annotation</th>
<th><em>Populus</em> target genes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Transcription regulation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transcription factor</td>
<td>ptc-miR69*</td>
<td>Zinc-finger (B-box) protein</td>
<td>fgenesh1_pg.C_LG_I002367(2), fgenesh1_pg.C_LG_I002595(2.5)</td>
</tr>
<tr>
<td></td>
<td>ptc-miR159</td>
<td>MYB family</td>
<td>eugene3.00400064(2), eugene3.00011133(2), estExt_fgenesh1_pg_v1.C_LG_IX1359(2), grail3.01000350(2.5)</td>
</tr>
<tr>
<td></td>
<td>ptc-miR166</td>
<td>HD-ZIP family</td>
<td>fgenesh1_pm.C_LG_I000485(1.5), estExt_fgenesh1_pm_v1.C_LG_I00003591(1.5), estExt_Genewise1_v1.C_660759(1.5), fgenesh1_pg.C_LG_IX001407(2), fgenesh1_pm.C_LG_XVII000007(2), estExt_fgenesh1_pm_v1.C_LG_VI0133(2), eugene3.63260001(2.5), estExt_Genewise1_v1.C_LG</td>
</tr>
<tr>
<td></td>
<td>ptc-miR398</td>
<td>SNF2-domain containing protein</td>
<td>fgenesh1_pg.C_LG_XVI000242(3)</td>
</tr>
<tr>
<td>DNA binding</td>
<td>ptc-miR69*</td>
<td>LIM-domain containing protein</td>
<td>estExt_fgenesh1_pg_v1.C_LG_XI752(3)</td>
</tr>
<tr>
<td>RNA processing and modification</td>
<td>ptc-miR26*</td>
<td>RRM motif-containing protein</td>
<td>estExt_fgenesh1_pg_v1.C_LG_I00395(2.5), eugene3.00012603(2.5)</td>
</tr>
<tr>
<td></td>
<td>ptc-miR69*</td>
<td>D111/G-patch domain-containing protein</td>
<td>eugene3.00181030(2.5), eugene3.133550001(3), fgenesh1_pg.C_LG_VI000496(3), fgenesh1_pm.C_LG_I000995(3), estExt_fgenesh1_pm_v1.C_LG_XV0200(3)</td>
</tr>
<tr>
<td></td>
<td>ptc-miR235*</td>
<td>Polypyrimidine tract-binding protein</td>
<td>eugene3.00120492(2)</td>
</tr>
<tr>
<td></td>
<td>ptc-miR396</td>
<td>SWAP-domain containing protein</td>
<td>eugene3.00124092(2)</td>
</tr>
<tr>
<td></td>
<td>ptc-miR396</td>
<td>MADS-box transcription factor</td>
<td>estExt_Genewise1_v1.C_LG_VII4000(2.5)</td>
</tr>
<tr>
<td>Metabolism and cellular processes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
<td>---------------------------------</td>
<td>-----------------</td>
<td></td>
</tr>
<tr>
<td><strong>Fatty acid metabolism</strong></td>
<td>ptc-miR39*</td>
<td>Acyl-CoA oxidase 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ptc-miR170*</td>
<td>Glycosyl hydrolase protein family</td>
<td></td>
</tr>
<tr>
<td><strong>Carbohydrate metabolism</strong></td>
<td>ptc-miR159</td>
<td>Asparagine synthase (1-4)-Beta-mannan endohydrolase</td>
<td></td>
</tr>
<tr>
<td><strong>Amino acid metabolism</strong></td>
<td>ptc-miR39*</td>
<td>Short-chain dehydrogenase</td>
<td></td>
</tr>
<tr>
<td><strong>Metabolism</strong></td>
<td>ptc-miR439*</td>
<td>Thioredoxin family</td>
<td></td>
</tr>
<tr>
<td><strong>Electron transport</strong></td>
<td>ptc-miR99*</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ptc-miR398</td>
<td>Cytochrome C oxidase</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ptc-miR408</td>
<td>Plastocyanin-like</td>
<td></td>
</tr>
<tr>
<td><strong>Cell cycle control</strong></td>
<td>ptc-miR396</td>
<td>Microtubule severing</td>
<td></td>
</tr>
<tr>
<td><strong>Cellular homeostasis</strong></td>
<td>ptc-miR56*</td>
<td>Ankyrin-domain containing protein</td>
<td></td>
</tr>
<tr>
<td><strong>Organelle biogenesis</strong></td>
<td>ptc-miR235*</td>
<td>PPR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ptc-miR396</td>
<td>PPR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ptc-miR475</td>
<td>PPR</td>
<td></td>
</tr>
<tr>
<td><strong>Protein regulation and signalling</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Protein transport</strong></td>
<td>ptc-miR26*</td>
<td>Ras small GTPase</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ptc-miR44*</td>
<td>ABC (ATP-binding cassette) transporter</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ptc-miR156</td>
<td>High-affinity nitrate transporter</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ptc-miR170</td>
<td>Sexual differentiation protein ISP4</td>
<td></td>
</tr>
<tr>
<td>Pathway/Response</td>
<td>ptc-miRNA</td>
<td>Predicted Target</td>
<td>Genes/Proteins</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>-----------</td>
<td>-----------------</td>
<td>-------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Ubiquitination pathway</td>
<td>ptc-miR44*</td>
<td>Predicted E3 ubiquitin</td>
<td>fgenesh1_pg.C_scaffold_82000021(3), eugene3.00140082(2.5)</td>
</tr>
<tr>
<td></td>
<td>ptc-miR472</td>
<td>Cyclin-like F-box protein</td>
<td>estExt_fgenesh1_pm_v1.C_LG_I0511(2.5), fgenesh1_pg.C_LG_I000446(3)</td>
</tr>
<tr>
<td>Molecular chaperone</td>
<td>ptc-miR396</td>
<td>Heat shock protein DNAJ</td>
<td>estExt_fgenesh1_pm_v1.C_LG_I0511(2.5)</td>
</tr>
<tr>
<td>Protein phosphorylation</td>
<td>ptc-miR69*</td>
<td>Putative leucine rich repeat protein kinase</td>
<td>fgenesh1_pg.C_LG_V001686(3)</td>
</tr>
<tr>
<td>Protein dephosphorylation</td>
<td>ptc-miR81*</td>
<td>Protein phosphatase 2C</td>
<td>fgenesh1_pg.C_LG_V001686(3)</td>
</tr>
<tr>
<td>Signal transduction</td>
<td>ptc-miR396</td>
<td>Ceramidases</td>
<td>eugene3.00161045(3)</td>
</tr>
<tr>
<td>Endocytosis</td>
<td>ptc-miR26*</td>
<td>Dynamin family protein</td>
<td>fgenesh1_pm.C_LG_I000286(2.5)</td>
</tr>
<tr>
<td>Cell death</td>
<td>ptc-miR398</td>
<td>Copper superoxide dismutase (SOD1)</td>
<td>fgenesh1_pg.C_LG_I000286(2.5), eugene3.00130294(2), fgenesh1_pg.C_LG_I000286(2.5)</td>
</tr>
<tr>
<td><strong>Environmental response</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integral membrane protein</td>
<td>ptc-miR167</td>
<td>Hly-III related proteins</td>
<td>estExt_Genewise1_v1.C_LG_XVIII3084(2.5), grail3.0030022401(2.5)</td>
</tr>
<tr>
<td>Defense response</td>
<td>ptc-miR69*</td>
<td>Putative disease resistance protein</td>
<td>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_799200001(1.5), fgenesh1_pg.C_LG_XIX000014(2)</td>
</tr>
<tr>
<td></td>
<td>ptc-miR472</td>
<td>Putative disease resistance protein</td>
<td>eugene3.065450005(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)</td>
</tr>
<tr>
<td>Unknown</td>
<td>ptc-miR26*</td>
<td>Unknown</td>
<td>fgenesh1_pg.C_LG_X000935(2.5), fgenesh1_pg.C_LG_I0001362(2.5), grail3.0109002101(2.5)</td>
</tr>
<tr>
<td></td>
<td>ptc-miR39*</td>
<td>Choline transporter-like protein</td>
<td>eugene3.00081101(2)</td>
</tr>
<tr>
<td></td>
<td>ptc-miR39*</td>
<td>Unknown</td>
<td>eugene3.117700001(0), eugene3.173450001(0), eugene3.116780001(0)</td>
</tr>
<tr>
<td></td>
<td>ptc-miR44*</td>
<td>Unknown</td>
<td>fgenesh1_pg.C_LG_V000523(2), fgenesh1_pm.C_LG_V100309(3)</td>
</tr>
<tr>
<td></td>
<td>ptc-miR56*</td>
<td>Unknown</td>
<td>grail3.13238000201(0), grail3.9008000101(0), grail3.21597000101(0), grail3.17468000201(0), grail3.1214000201(0)</td>
</tr>
</tbody>
</table>
MicroRNAs from Populus

<table>
<thead>
<tr>
<th>ptc-miR69*</th>
<th>Retrotransposon</th>
<th>fgenesh1_pg.C_scaffold_817000001(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ptc-miR69*</td>
<td>Unknown</td>
<td>eugene3.00102261(1.5)</td>
</tr>
<tr>
<td>ptc-miR235*</td>
<td>DUF26 domain containing protein</td>
<td>eugene3.06280002(2)¹, fgenesh1_pm.C_LG_VIII000433(2.5)¹</td>
</tr>
<tr>
<td>ptc-miR156</td>
<td>Unknown</td>
<td>grail3.0008008901(1), fgenesh1Pg_C_scaffold_9189000001(2), grail3.0047015801(2.5), eugene3.04930001(2.5), fgenesh1_pg.C_scaffold_4640000004(2.5),</td>
</tr>
<tr>
<td>ptc-miR159</td>
<td>Unknown</td>
<td>grail3.0043007201(0), grail3.0016033301(1.5), eugene3.01340043(1.5), eugene3.00111000(1.5), grail3.0123001501(2.5),</td>
</tr>
<tr>
<td>ptc-miR167</td>
<td>Unknown</td>
<td>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_VII1517(3)</td>
</tr>
<tr>
<td>ptc-miR408</td>
<td>Early-responsive to dehydration-related</td>
<td>estExt_fgenesh1_pm_v1.C_LG_XI00014(3), estExt_fgenesh1_pm_v1.C_1660007(3), estExt_Genewise1_v1.C_14050002(3)</td>
</tr>
<tr>
<td>ptc-miR408</td>
<td>Unknown</td>
<td>fgenesh1_pg.C_LG_IL001930(0)²</td>
</tr>
</tbody>
</table>

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


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

<table>
<thead>
<tr>
<th>ID</th>
<th>Sequence (5’-3’)</th>
<th>Length (nt)</th>
<th>Cloning frequency</th>
<th>Putative targets</th>
<th>Predicted target function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CCCCCAAGAUGAGUGCUC</td>
<td>20</td>
<td>1</td>
<td>eugene3.27810001(0)</td>
<td>Unknown</td>
</tr>
<tr>
<td>9</td>
<td>UUUCUGGUGCCAUAUUCUCC</td>
<td>21</td>
<td>2</td>
<td>fgenesh1_pm.C_scaffold_15200020(1)</td>
<td>Leucine rich repeat protein</td>
</tr>
<tr>
<td>13</td>
<td>GGCUAAGGCAUUGCGAAG</td>
<td>20</td>
<td>1</td>
<td>eugene3.125570001(0)</td>
<td>Hypothetical protein</td>
</tr>
<tr>
<td>16</td>
<td>GUAUGCAGCGAGGCAUCCUAAC</td>
<td>23</td>
<td>2</td>
<td>fgenesh1_pg.C_LG_IL000237(1)</td>
<td>Amine oxidase</td>
</tr>
<tr>
<td>17</td>
<td>AUAACGGUCCUAAGUGCGA</td>
<td>21</td>
<td>1</td>
<td>eugene3.150890001(1), eugene3.27810001(1)</td>
<td>Hypothetical protein</td>
</tr>
<tr>
<td>22</td>
<td>GCUCUUCUCAUUGUGAA</td>
<td>20</td>
<td>1</td>
<td>eugene3.81440001(0), eugene3.05760001(0), eugene3.72670001(0), grail3.0005004301(0), grail3.0182000701(0), estExt_Genewise1_v1.C_91780006(0), eugene3.10290008(0), eugene3.01960031(1)</td>
<td>Unknown</td>
</tr>
<tr>
<td>32</td>
<td>UAGCUAACCAAGGAAACAGC</td>
<td>22</td>
<td>1</td>
<td>estExt_Genewise1_v1.C_91780006(0), eugene3.10290008(0), eugene3.01960031(1)</td>
<td>Transposon-encoded proteins</td>
</tr>
<tr>
<td>34</td>
<td>UAGUUCACCUCUJAUGUGAC</td>
<td>24</td>
<td>1</td>
<td>fgenesh1_pg.C_LG_10030002(1)</td>
<td>Aquaporin</td>
</tr>
</tbody>
</table>

113
<table>
<thead>
<tr>
<th>Gene Accession</th>
<th>Start</th>
<th>End</th>
<th>Description</th>
<th>Transcript Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>UGGCAAGGCCCUCUGGUGAA</td>
<td>42</td>
<td>21</td>
<td>UGCCAAGCCCUCGUGGUGAA</td>
<td>MicroRNA</td>
</tr>
<tr>
<td>AACGACUCUGGCAACGGAAU</td>
<td>46</td>
<td>21</td>
<td>AACGACUCUGGCAACGGAAU</td>
<td>MicroRNA</td>
</tr>
<tr>
<td>UCACCGGGGUAAACAGCUGU</td>
<td>53</td>
<td>22</td>
<td>UCACCGGGGUAAACAGCUGU</td>
<td>MicroRNA</td>
</tr>
<tr>
<td>GGUGUGCAUGGCGGUCUU</td>
<td>56</td>
<td>20</td>
<td>GGUGUGCAUGGCGGUCUU</td>
<td>MicroRNA</td>
</tr>
<tr>
<td>AUAACGGUCAAAGGUGCAGA</td>
<td>70</td>
<td>22</td>
<td>AUAACGGUCAAAGGUGCAGA</td>
<td>MicroRNA</td>
</tr>
<tr>
<td>ACAAGCAUUGGCAUGGUGCCU</td>
<td>72</td>
<td>22</td>
<td>ACAAGCAUUGGCAUGGUGCCU</td>
<td>MicroRNA</td>
</tr>
<tr>
<td>AGCAUAUCAAUAAGCGGGAG</td>
<td>73</td>
<td>20</td>
<td>AGCAUAUCAAUAAGCGGGAG</td>
<td>MicroRNA</td>
</tr>
<tr>
<td>CAUGCCGCAGAUCUUCUUG</td>
<td>74</td>
<td>19</td>
<td>CAUGCCGCAGAUCUUCUUG</td>
<td>MicroRNA</td>
</tr>
<tr>
<td>AGCUCUCAUAGGUGGUGGUA</td>
<td>75</td>
<td>20</td>
<td>AGCUCUCAUAGGUGGUGGUA</td>
<td>MicroRNA</td>
</tr>
<tr>
<td>CUGCCAGGACCGCUGGCCC</td>
<td>79</td>
<td>19</td>
<td>CUGCCAGGACCGCUGGCCC</td>
<td>MicroRNA</td>
</tr>
<tr>
<td>UCCUGCCAGUAAGCAUAGCU</td>
<td>80</td>
<td>23</td>
<td>UCCUGCCAGUAAGCAUAGCU</td>
<td>MicroRNA</td>
</tr>
<tr>
<td>AGCCAGUGGUUGGGACU</td>
<td>82</td>
<td>17</td>
<td>AGCCAGUGGUUGGGACU</td>
<td>MicroRNA</td>
</tr>
<tr>
<td>AUGAUAAUGCUUCCUUGGAUGA</td>
<td>83</td>
<td>22</td>
<td>AUGAUAAUGCUUCCUUGGAUGA</td>
<td>MicroRNA</td>
</tr>
</tbody>
</table>

114
<table>
<thead>
<tr>
<th></th>
<th>Sequence</th>
<th>Count</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>87</td>
<td>UCGGAAGGGUGGUGCCUAC</td>
<td>22</td>
<td>eugene3.12557001(0), eugene3.27810001(0), eugene3.15089001(0)</td>
<td>Hypothetical protein</td>
</tr>
<tr>
<td>88</td>
<td>UUGCAUUCCGGAGGGGAC</td>
<td>20</td>
<td>grail3.9190000201(0), grail3.8915000201(0), grail3.21481000201(0), eugene3.0190000501(0), eugene3.1746800201(0), eugene3.1323800201(0)</td>
<td>Unknown</td>
</tr>
<tr>
<td>89</td>
<td>UCCAGCUCCAUAGCGUAUA</td>
<td>20</td>
<td>grail3.1746800201(0), grail3.0190000501(0), grail3.21481000201(0), grail3.003302201(0)</td>
<td>Unknown</td>
</tr>
<tr>
<td>90</td>
<td>GAAGGGCGUCGCUCAACGG</td>
<td>20</td>
<td>grail3.5952000201(0), eugene3.15089001(0), eugene3.125570001(0), eugene3.27810001(1)</td>
<td>Unknown</td>
</tr>
<tr>
<td>91</td>
<td>fgenesh1_pg.C_scaffold_8474000002(1)</td>
<td></td>
<td></td>
<td>Putative cell wall associated hydrolase</td>
</tr>
<tr>
<td>92</td>
<td>UUGCAAGGGUGGAGGAAAGA</td>
<td>20</td>
<td>grail3.0214000401(0), eugene3.10610001(0), grail3.19774000201(0), estExt_Genewise1_v1.C_94310004(0), eugene3.137670001(0), eugene3.01960004(0), eugene3.83720001(0), eugene3.10586001(0), eugene3.166320001(0)</td>
<td>Unknown</td>
</tr>
<tr>
<td>93</td>
<td>UCGGAAUAGAUUGGCGUAA</td>
<td>20</td>
<td>estExt_Genewise1_v1.C_95310005(0)</td>
<td>Unknown</td>
</tr>
<tr>
<td>94</td>
<td>ACAUCUGUGUGCACUAACUC</td>
<td>20</td>
<td>fgenesh1_pm.C_scaffold_2280000001(0), fgenesh1_pm.C_LG_XIII000193(0)</td>
<td>Copia-like retrotransposon family</td>
</tr>
<tr>
<td>95</td>
<td>AAGGGUUAAGCGAGGUUCUC</td>
<td>20</td>
<td>fgenesh1_pg.C_LG_XVIII000301(1)</td>
<td>Unknown</td>
</tr>
<tr>
<td>96</td>
<td>CGAUCACUGAGAUUCAGCCC</td>
<td>21</td>
<td>eugene3.57820001(0), eugene3.04560006(0), eugene3.05760001(0), eugene3.6610001(0), eugene3.81440001(0), grail3.0182000701(0), grail3.7269000201(0), grail3.686320001(0), eugene3.01960031(1), eugene3.10086001(0), fgenesh1_pg.C_scaffold_10733000001(0)</td>
<td>Unknown</td>
</tr>
<tr>
<td>97</td>
<td>Hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>SNF2 family DNA-dependent ATPase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>UGUGAAGAGGUGGUGCCUAC</td>
<td>22</td>
<td>eugene3.0853002(0), eugene3.0196001501(0), eugene3.10290007(0), eugene3.9080001101(0), estExt_Genewise1_v1.C_94310004(0)</td>
<td>Unknown</td>
</tr>
<tr>
<td>100</td>
<td>AUCCCGUGAACCUGAAGGCACUG</td>
<td>24</td>
<td>fgenesh1_pm.C_LG_V000019(2)</td>
<td>Putative methyltransferase DUF248</td>
</tr>
<tr>
<td>101</td>
<td>UGACGCGCAUGAAUGGAGU</td>
<td>19</td>
<td>grail3.1069600101(0)</td>
<td>Zinc finger protein</td>
</tr>
</tbody>
</table>
131  UGGCUGUCGCACUGUCUGUGA  25  1  grail3.00122000801(0), grail3.13773000101(1), estExt_Genewise1_v1.C_95310005(1)  Unknown

144  AAUGCAUAUGGGUCAAAUUUGACAA  25  1  fgenesh1_pm.C_LG_XII000258(2), grail3.0074008701(2)  Unknown

146  CCAGUAGUCGUAGCGGUAA  20  1  estExt_Genewise1_v1.C_95310005(0)  Unknown

155  GUAGUCAUAUGCUUGUCUCA  20  2  grail3.9190000201(0), eugene3.2449000201(0)  Putative Zinc finger protein

158  CCGUUGAGAGGGAGGCGCA  23  1  eugene3.00110764(2)  Calmodulin-binding family

160  CACAAUUCGAGAAGUUC  20  1  eugene3.0012570(1)  Uncharacterized conserved protein

165  GGGAGCUUGCCGGAAGGGA  20  1  eugene3.72670001(0), grail3.7269000301(0), estExt_Genewise1_v1.C_91780006(0), eugene3.57820001(1)  Putative succinate dehydrogenase

167  AUAAACGUCCUAAGGUGACG  20  1  eugene3.27810001(0), eugene3.150890001(0), eugene3.125570001(0)  Hypothetical protein, cell wall associated hydrolase

171  GGGGGAUAGUGUCGCAAGG  20  2  grail3.17468000201(0), grail3.13238000201(0)  Conserved hypothetical protein

172  AUGGUUGGUGCAACUUCUUUUGA  23  3  estExt_fgenesh1_pg_v1.C_LG_12049(1)  Protein involved in membrane traffic (YOP1/TB2/DP1/HVA22 family)

173  AAUCAGGUUAUGGACCGAUA  25  1  fgenesh1_pg.C_LG_VI000069(2)  Cytochrome P450 CYP4/CYP19/CYP26 subfamilies

174  AUUAGUGACGCCGAUGAAUGGAU  24  1  estExt_Genewise1_v1.C_91780006(0), eugene3.57820001(0), eugene3.72670001(0)  Zinc finger protein

182  UUGUUGUGUGUAGUGUGUGA  20  2  estExt_fgenesh1_pg_v1.C_LG_III00667(0)  Pectin-related protein

183  UGAAAUUGCAACUGGGGGU  19  1  grail3.0091000801(1)  MYB-related protein

186  GCUGUCAAGUUGCCAAUCUACAAA  23  1  fgenesh1_pg.C_LG_II002119(1)  Acetylglucosaminyltransferase

189  AACCGGGUUGAAACUCAACACA  21  2  grail3.0093002701(2)  Major intrinsic protein family

195  AUGGCAAACACUCGGUAGGGUA  24  1  eugene3.00161333(2)  PPR repeat

197  UUGUGUGUUAUGGAGUAAGU  21  1  eugene3.01270009(2)  Ribosomal protein S6 kinase and related proteins

116
<table>
<thead>
<tr>
<th>Start</th>
<th>Sequence</th>
<th>Authority</th>
<th>Evidence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>198</td>
<td>CACAAAUCGAGAAUGGUCC</td>
<td>20</td>
<td>eugene3.00012570(1)</td>
<td>Uncharacterized conserved protein</td>
</tr>
<tr>
<td>205</td>
<td>GCUUCUUUGAGAAAACAGUCCAUAA</td>
<td>25</td>
<td>eugene3.01700038(1), eugene3.01700060(1)</td>
<td>Unknown</td>
</tr>
<tr>
<td>208</td>
<td>AGGAUGAUCCACACACUGGGAC</td>
<td>23</td>
<td>grail3.0286000901(0), grail3.0286000901(0), grail3.13773000101(0), estExt_Genewise1_v1.C_95310005(2)</td>
<td>Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ca2+-binding protein</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Subtilisin kexin isozyme-1/site 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Acyl-CoA synthetase</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DNA binding WRKY protein</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Nuclear pore complex, Nup98</td>
</tr>
<tr>
<td>226</td>
<td>CAUUCUCUUCUCUGCUUGCG</td>
<td>22</td>
<td>estExt_fgenesh1_pg_v1.C_LG_DX0420(1)</td>
<td>Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DNA binding WRKY protein</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Karyopherin (importin) beta 3</td>
</tr>
<tr>
<td>229</td>
<td>CAUAAUCAUCGUUUCGUU</td>
<td>20</td>
<td>eugene3.12557001(0), eugene3.150890001(0), eugene3.5952000201(0), eugene3.27810001(0)</td>
<td>Conserved hypothetical protein</td>
</tr>
<tr>
<td>230</td>
<td>GCAAAGCCAGAGGGAGCUUGA</td>
<td>22</td>
<td>fgenesh1_pg.C_LG_I0002327(2), fgenesh1_pg.C_LG_XVII000611(2)</td>
<td>Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DNA binding WRKY protein</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Nuclear pore complex, Nup98</td>
</tr>
<tr>
<td>232</td>
<td>UCGAUCCUUAGACCUCUGGAA</td>
<td>22</td>
<td>fgenesh1_pg.C_scaffold_182000013(0), estExt_Genewise1_v1.C_917800060(0), estExt_Genewise1_v1.C_917800060(0), grail3.9892000201(0), grail3.0182000201(0), eugene3.05760001(0), eugene3.72670001(0), eugene3.86230001(0), eugene3.57820001(0), eugene3.81440001(1)</td>
<td>Unknown</td>
</tr>
<tr>
<td>233</td>
<td>GUAUCCAGAGGUAGCGCUUG</td>
<td>20</td>
<td>eugene3.148650001(0), eugene3.02190005(0), eugene3.131840001(0), eugene3.111770001(0), eugene3.10280001(0), grail3.10051000101(0), grail3.11871000101(0)</td>
<td>Unknown</td>
</tr>
<tr>
<td>Small RNA ID</td>
<td>Description</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------------</td>
<td>------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>grail3.12102000101(0), grail3.019600201(0), grail3.9345000101(0),</td>
<td>Unknown</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>grail3.1043000101(1), eugene3.98720001(1), estExt_Genewise1_v1.C_62520001(1),</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>grail3.0214000501(2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>grail3.0196001101(0)</td>
<td>Hypothetical apoptosis regulator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>grail3.9008000501(0)</td>
<td>Hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^{a}\) 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>
<thead>
<tr>
<th>ptc-miRNA gene</th>
<th>miR stem location</th>
<th>Precursor Sequences (5' - 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>ptc-MIR26a</td>
<td>scaffold_628:11989-11808</td>
<td>CCTCCTATCAAAATAATTTAAATCTAGTTATTAGAGACCTTAAAGGTGTTTTTACATGTGTTATTAGTTATATTACCAAAATATCGGAGGGTTAATGGTTTCTG</td>
</tr>
<tr>
<td>ptc-MIR26b</td>
<td>LG_XIII:4014178-4013847</td>
<td>GCCCAATTGAGGCAATTGAGGAGAATAGTTAGTTATTACCAAAATATCGGAGGGTTAATGGTTTCTG</td>
</tr>
<tr>
<td>ptc-MIR39a</td>
<td>LG_XVI:12106164-12106072</td>
<td>CAATGAGGAAGGAGAAGGGAAATGGGAGAGTATGCAAGAAGGAAAAATTCATGATTTAATATTCTTTCTTGCAGTACCTTCTCTTGGTGGG</td>
</tr>
<tr>
<td>ptc-MIR56a</td>
<td>scaffold_6643:1495-1780</td>
<td>GCCCAATTTGAGCTTAATTGAGGAGATTGAAATTAGAGGACCAAACATGTTATTACCAAAATATCGGAGGGTTAATGGTTTCTG</td>
</tr>
<tr>
<td>ptc-MIR56b</td>
<td>scaffold_21597:141-339</td>
<td>GCCCAATTTGAGCTTAATTGAGGAGATTGAAATTAGAGGACCAAACATGTTATTACCAAAATATCGGAGGGTTAATGGTTTCTG</td>
</tr>
<tr>
<td>ptc-MIR56c</td>
<td>scaffold_520:40810-40965</td>
<td>GCCCAATTTGAGCTTAATTGAGGAGATTGAAATTAGAGGACCAAACATGTTATTACCAAAATATCGGAGGGTTAATGGTTTCTG</td>
</tr>
<tr>
<td>ptc-MIR56d</td>
<td>scaffold_196:124123-123997</td>
<td>GCCCAATTTGAGCTTAATTGAGGAGATTGAAATTAGAGGACCAAACATGTTATTACCAAAATATCGGAGGGTTAATGGTTTCTG</td>
</tr>
<tr>
<td>ptc-MIR69a</td>
<td>LG_VIII:8092808-8092913</td>
<td>GCCCAATTTGAGCTTAATTGAGGAGATTGAAATTAGAGGACCAAACATGTTATTACCAAAATATCGGAGGGTTAATGGTTTCTG</td>
</tr>
<tr>
<td>ptc-MIR69b</td>
<td>LG_VIII:8092806-8092913</td>
<td>GCCCAATTTGAGCTTAATTGAGGAGATTGAAATTAGAGGACCAAACATGTTATTACCAAAATATCGGAGGGTTAATGGTTTCTG</td>
</tr>
<tr>
<td>ptc-MIR81</td>
<td>scaffold_129:575576-575579</td>
<td>GCCCAATTTGAGCTTAATTGAGGAGATTGAAATTAGAGGACCAAACATGTTATTACCAAAATATCGGAGGGTTAATGGTTTCTG</td>
</tr>
<tr>
<td>ptc-MIR99</td>
<td>LG_VII:120921-121011</td>
<td>GCCCAATTTGAGCTTAATTGAGGAGATTGAAATTAGAGGACCAAACATGTTATTACCAAAATATCGGAGGGTTAATGGTTTCTG</td>
</tr>
<tr>
<td>ptc-MIR170</td>
<td>LG_IV:6840773-6840614</td>
<td>GCCCAATTTGAGCTTAATTGAGGAGATTGAAATTAGAGGACCAAACATGTTATTACCAAAATATCGGAGGGTTAATGGTTTCTG</td>
</tr>
</tbody>
</table>
MicroRNAs from *Populus*

- **ptc-MIR235a** LG_XVIII:5 423584-5423917
  - GAAGATCGATGGCTATGGAAGTTATGCATGTAGGTATTTGCATCTTTCAT
  - TTATTTAAATATATATATTTGCATCTTTCCTCTCTCTCGCTTAGAAGTAG
  - TGGCATTCTCAAAATTGAGCCATGCCTCTTTCTCTCTCCGCTTAGAAGTAG

- **ptc-MIR235b** LG_XVII:2785382-2785553
  - GAAGATCGATGGCTATGGAAGTTATGCATGTAGGTATTTGCATCTTTCAT
  - TTATTTAAATATATATATTTGCATCTTTCCTCTCTCTCGCTTAGAAGTAG

- **ptc-MIR156d** LG_XVIII:1 365495-1365685
  - GAAGATCGATGGCTATGGAAGTTATGCATGTAGGTATTTGCATCTTTCAT
  - TTATTTAAATATATATATTTGCATCTTTCCTCTCTCTCGCTTAGAAGTAG

- **ptc-MIR156e** LG_XVIII:6 180951-6181169
  - GAAGATCGATGGCTATGGAAGTTATGCATGTAGGTATTTGCATCTTTCAT
  - TTATTTAAATATATATATTTGCATCTTTCCTCTCTCTCGCTTAGAAGTAG

- **ptc-MIR156g** LG_II:21820430-21820436
  - GAAGATCGATGGCTATGGAAGTTATGCATGTAGGTATTTGCATCTTTCAT
  - TTATTTAAATATATATATTTGCATCTTTCCTCTCTCTCGCTTAGAAGTAG

- **ptc-MIR166a** LG_I:6096929-6097080
  - GAAGATCGATGGCTATGGAAGTTATGCATGTAGGTATTTGCATCTTTCAT
  - TTATTTAAATATATATATTTGCATCTTTCCTCTCTCTCGCTTAGAAGTAG

- **ptc-MIR166b** scaffold_122:445042-445157
  - GAAGATCGATGGCTATGGAAGTTATGCATGTAGGTATTTGCATCTTTCAT
  - TTATTTAAATATATATATTTGCATCTTTCCTCTCTCTCGCTTAGAAGTAG

- **ptc-MIR166c** LG_XIII:2546006-2546111
  - GAAGATCGATGGCTATGGAAGTTATGCATGTAGGTATTTGCATCTTTCAT
  - TTATTTAAATATATATATTTGCATCTTTCCTCTCTCTCGCTTAGAAGTAG
MicroRNAs from Populus

- ptc-MIR166g LG_X:8492275-8492368
  AGTGTAGGGGAAATGCGTATCAGGATCCAAGTTAAGGCAACACAT
  TCATCTTGGTCAAGTCTGGGACCAGCAGCTTCATTTTCCTCCTTCCACCACACT

- ptc-MIR166h LG_VIII:11091691-11091583
  TGCTCAACAGCTTGGGACCAGGACGACATGACACTCCCTGAA
  GAGACCGCATTCTACCTTTTGAGGTATCTGGGACCAGGCTTCATTTTCCTTCCACCACACT

- ptc-MIR166i LG_V:824495-8243999
  GTTGAGGGGAAATGCGTATCAGGATCCAAGTTAAGGCAACACAT
  AAAGAAACACATCTTGAAGGACCAGGCTTCATTTTCCTTCCACCACACT

- ptc-MIR166j LG_VII:15276296-15276176
  TGCTACACAGTTGAGGGGAATGCTGTCTGGTTCGAGACCATTCACCTGAA
  GAGACCGCATTCTACCTTTTGAGTGATCTGGGACCAGGCTTCATTCCCCCAACTCAAAA

- ptc-MIR166k LG_V:11091691-11091583
  TGCTACACAGTTGAGGGGAATGCTGTCTGGTTCGAGACCATTCACCTGAA
  GAGACCGCATTCTACCTTTTGAGTGATCTGGGACCAGGCTTCATTCCCCCAACTCAAAA

- ptc-MIR166l LG_II:2737618-2737737
  TGCTACACAGTTGAGGGGAATGCTGTCTGGTTCGAGACCATTCACCTGAA
  GAGACCGCATTCTACCTTTTGAGTGATCTGGGACCAGGCTTCATTCCCCCAACTCAAAA

- ptc-MIR166m LG_II:13668142-13668286
  GTGGGTTGGGAAATGCGTATCAGGATCCAAGTTAAGGCAACACAT
  AAAGAAACACATCTTGAAGGACCAGGCTTCATTTTCCTTCCACCACACT

- ptc-MIR166n LG_VIII:11091691-11091583
  TGCTACACAGTTGAGGGGAATGCTGTCTGGTTCGAGACCATTCACCTGAA
  GAGACCGCATTCTACCTTTTGAGTGATCTGGGACCAGGCTTCATTCCCCCAACTCAAAA

- ptc-MIR166o LG_V:824362-824206
  GTGGGTTGGGAAATGCGTATCAGGATCCAAGTTAAGGCAACACAT
  AAAGAAACACATCTTGAAGGACCAGGCTTCATTTTCCTTCCACCACACT

- ptc-MIR166p LG_V:11091691-11091583
  TGCTACACAGTTGAGGGGAATGCTGTCTGGTTCGAGACCATTCACCTGAA
  GAGACCGCATTCTACCTTTTGAGTGATCTGGGACCAGGCTTCATTCCCCCAACTCAAAA

- ptc-MIR166q LG_XIII:4057334-4057502
  GTGGGTTGGGAAATGCGTATCAGGATCCAAGTTAAGGCAACACAT
  AAAGAAACACATCTTGAAGGACCAGGCTTCATTTTCCTTCCACCACACT

- ptc-MIR166r LG_V:824495-8243999
  GTGGGTTGGGAAATGCGTATCAGGATCCAAGTTAAGGCAACACAT
  AAAGAAACACATCTTGAAGGACCAGGCTTCATTTTCCTTCCACCACACT

- ptc-MIR167a LG_II:3057682-3057565
  GTGGGTTGGGAAATGCGTATCAGGATCCAAGTTAAGGCAACACAT
  AAAGAAACACATCTTGAAGGACCAGGCTTCATTTTCCTTCCACCACACT

- ptc-MIR167b scaffold_14274:302-178
  GTGGGTTGGGAAATGCGTATCAGGATCCAAGTTAAGGCAACACAT
  AAAGAAACACATCTTGAAGGACCAGGCTTCATTTTCCTTCCACCACACT

- ptc-MIR167c scaffold_14274:302-178
  GTGGGTTGGGAAATGCGTATCAGGATCCAAGTTAAGGCAACACAT
  AAAGAAACACATCTTGAAGGACCAGGCTTCATTTTCCTTCCACCACACT

- ptc-MIR167d LG_II:3055016-3054896
  GTGGGTTGGGAAATGCGTATCAGGATCCAAGTTAAGGCAACACAT
  AAAGAAACACATCTTGAAGGACCAGGCTTCATTTTCCTTCCACCACACT

- ptc-MIR167e LG_VIII:10761707-10761813
  GTGGGTTGGGAAATGCGTATCAGGATCCAAGTTAAGGCAACACAT
  AAAGAAACACATCTTGAAGGACCAGGCTTCATTTTCCTTCCACCACACT

- ptc-MIR167f LG_VIII:10761707-10761813
  GTGGGTTGGGAAATGCGTATCAGGATCCAAGTTAAGGCAACACAT
  AAAGAAACACATCTTGAAGGACCAGGCTTCATTTTCCTTCCACCACACT

- ptc-MIR167g scaffold_70:484778-484660
  GTGGGTTGGGAAATGCGTATCAGGATCCAAGTTAAGGCAACACAT
  AAAGAAACACATCTTGAAGGACCAGGCTTCATTTTCCTTCCACCACACT

121
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 “genom 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 Victor¹, Ying-Hsuan Sun², Shanfa Lu³, Erika Asamizu³, Satoshi Tabata³, Vincent L. Chiang², Alexander A. Myburg¹

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
MicroRNAs from Eucalyptus 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 mirVana™ 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 MgCl₂, 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’-TAGGCACCATTCCATCGTCG–3’) and a 3’-adaptor reverse primer (5’–CAGGTGCCTCACGACGATGA–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
MicroRNAs from Eucalyptus 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
MicroRNAs from Eucalyptus have 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 known 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
MicroRNAs from Eucalyptus 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).
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 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 (Reu ther 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

I would like to thank Sappi Forestry for providing the plant materials, and the Kazusa DNA Research Institute for providing the E. camaldulensis sequences used in this study. I would also like to thank Prof. Henk Huismans and Dr. Jacques Theron for their contributions to the editing of the chapter. Saul Garcia and Martin Ranik provided assistance with technical aspects of this work and editing of the manuscript. I would also like to thank Dr. Y-H. Sun for writing the computer script and performing semi-automated mfold analysis and target predictions. This work was supported by funding provided by Mondi Business Paper South Africa and Sappi Forestry, through the Wood and Fibre Molecular Genetics Programme. Additional funding was provided by the Technology and Human Resources for Industry Programme and the National Research Foundation of South Africa.
3.7 FIGURES

egr-miR31 UUCAAAUCCAGUACAUCA

C

ACCA

5’ AAUUAACAAUC AGUAACAUCA

GUACGCAAUCUUAAGAAUUUAUUG G

3’ UUAUGGAAU UCAUGAUGUCAUGGAGUUGAAUCAUAC

/ A

---

egr-miR90 AGGUUGGUGC UGGCUUGGC

UU

A

CA

5’ AGGUUGGUGC C UGGCUAGU AGGC AGCCUCCGCGCGCUUGGG GAGCGUG CUC CCG A

3’ UCCAACCAAGAACC G ACCGAUCA UCC GUCGAGCGCGCGCCGGCAAC CUCGAU GAG GGU U

CC

A

U

A

3’ C

 Article

egr-miR140 AUUGGCAACAAGUGCAAUAG

C

A

5’ AUUGGCAACAAGUGCAUAG

UUAGAAAUUUUUUUUUUGG

82 nt

3’ UAACC UUAUGUUAUC AAA-UCDUUGAAAAAAAUU

LOOP

AC

C

egr-miR200a CGCACCCUCGCGGCCCAACAAGCGA

C

A

5’ CGCACCCUCGCC CGCA GAGCG A

GGGCU CC CCGCC AUCCAGGCGAGGGC-C-GA UGGCCCU CCUUGCC

3’ GCCUGGGAGCGG CGGU UCACUCCGG

C

AC

G

AC

3’ G-CCGGAGCGG /

UCGCGG CCGG

egr-miR200b CGCACCCUCGUUGGC

C

A

5’ CGCACCCUCGUUGGC CGGA A

CCGCGU CC CCGCC AUCCAGGCGAGGGC-C-GA CCUUGCC

3’ GGGAGCGACUGGCGU GCCUCCCGCA GA GGCAG GAGCGU GGGAGCGG

A

C

A

AC

U

A

A

C

CCG

CCG

151
egr-miR200c CGCACCCUCGCGCCACACAAGCGA

30 nt
LOOP
C CG C
5' GCACCCUCGC GCCAAG GAACC-GUC ACCCU GCU UGGCCGCAAGGG C
3' CGUGGGAGCG UGGUGUUC CUGCCACGCGGCGUAGGA CCG ACCCGCUGUCCUC U
A UG A A AC- ACG

egr-miR200d CACGCCCUAUCGGCCACACAAGCGA

A A G A
GG GA- GC CC- AGA
5' C CGCCUC UC GCCAC AGCGA
3' G GUGGGAG AG CGGUG UCGUU CG GCC GGG GA /
C C G C AA ACG AA ACG GUC

egr-miR200e UGCACCCUCAUCGACCACAAGCGA

AU G A
GG GA- GC CC- AGA
5' UGCACCCUC UC GCCAC AGCGA
3' GCGUGGGAG G UGGUGU CGCUCCCG /
CG G C ACG

---

egr-miR293a GGUCGGCCUCGCGCCAUGGCUGG

C U
5' GGU GGCUCGCCAUGGCUGGUGAGC G
3' CCA CGGACGGGUACCGACCACGCUCG G C

egr-miR293b GGUCGGCCUCGCGCCAUGGCUGG

GCU
5' GGUCGGCCUCGCCAUGGCUGGCGA C
3' CCGGC-GGAGCCGUACCGACCACCGGCU G GGA

egr-miR293c GGUCGGCCUCGCGCCAUGGCUGG

GG C U
5' UGGCCU GCCAUGGCGCUCCGAGC G
3' AGCUAGA CGGUACCGGCCGCU GC A

egr-miR293d GGUCGGCCUCGCGCCAUGGCUGG

C A U
5' GGU GGCUCGCCAUGGCUGGCG G C
3' CCA CGGACGGGUACCGGCCGCGC G A C A
egr-miR293e GGUCGGCCUCGCGCAUGGCUAG

<table>
<thead>
<tr>
<th>C</th>
<th>A</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’</td>
<td>GGU</td>
<td>GGCCUCGCCAUGGCU GGCAGGC C</td>
</tr>
<tr>
<td>3’</td>
<td>CCA</td>
<td>CCGGACCGGUACCGA CCGCUCCG G</td>
</tr>
</tbody>
</table>

egr-miR320a UAAGUCGCAGAUAUGGUAUGGGCU

<table>
<thead>
<tr>
<th>C</th>
<th>A</th>
<th>A</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’</td>
<td>UAAAU</td>
<td>GCGAUAUGGGUAUGGGCUUUGGAGAGCUCAGUGUAAGUU AAA ACUUAGU GU CAUAAC</td>
<td></td>
</tr>
<tr>
<td>3’</td>
<td>AUUCA</td>
<td>CGUGUACCAUAACCGAAGACCUCUUCUGCAGGAUGUU AAUU UGAAUCA CA GUAAUUG LOOP</td>
<td></td>
</tr>
</tbody>
</table>

egr-miR320b UAAGUCGCAGAUAUGGUAUGGGCU

<table>
<thead>
<tr>
<th>C</th>
<th>A</th>
<th>A</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’</td>
<td>UAAACU</td>
<td>GCGAUAUGGGUAUGGGCUUUGGAGAGCUCAGUGUAAGUU AAA ACUUAGU GU CAUAAC</td>
<td></td>
</tr>
<tr>
<td>3’</td>
<td>AUUCA</td>
<td>CGUGUACCAUAACCGAAGACCUCUUCUGCAGGAUGUU AAUU UGAAUCA CA GUAAUUG LOOP</td>
<td></td>
</tr>
</tbody>
</table>

egr-miR320c UAAGUCUCAGAUAUGGUAUGGGCU

<table>
<thead>
<tr>
<th>UC</th>
<th>A</th>
<th>A</th>
<th>C</th>
<th>UAAAA</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’</td>
<td>UAAAGU</td>
<td>CGAUAUGGGUAUGGGCUUUGGAGAGCUCAGUGUAAGUU AAA ACUUAGU GU CAUAAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3’</td>
<td>AUUCA</td>
<td>CGUGUACCAUAACCGAAGACCUCUUCUGCAGGAUGUU AAUU UGAAUCA CA GUAAUUG LOOP</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

egr-miR320d UAAGUCGUGAUAUGGUAUGGGCU

<table>
<thead>
<tr>
<th>GU</th>
<th>A</th>
<th>A</th>
<th>GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’</td>
<td>UAAAGU</td>
<td>GA UUGGU UGGGCUU U</td>
<td></td>
</tr>
<tr>
<td>3’</td>
<td>AUUCA</td>
<td>CU AAUCG ACCCGAG_AG</td>
<td></td>
</tr>
</tbody>
</table>

egr-miR320e UAAGUCGCAGAUAUGGUAUGGGCU

<table>
<thead>
<tr>
<th>GC</th>
<th>A</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’</td>
<td>UAAAGU</td>
<td>GA UUGGU UGGGCUU G</td>
<td></td>
</tr>
<tr>
<td>3’</td>
<td>AUUCA</td>
<td>CU AAUA ACCCGAG G</td>
<td></td>
</tr>
</tbody>
</table>

egr-miR320f UAAGUCGAGAUAUGGUAUGGGCU

<table>
<thead>
<tr>
<th>G</th>
<th>A</th>
<th>A</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’</td>
<td>UAAAGU</td>
<td>AGA UUGGU CGGGCUU G</td>
<td></td>
</tr>
<tr>
<td>3’</td>
<td>AUUCA</td>
<td>GCU ACUCA GCUGAG G</td>
<td></td>
</tr>
</tbody>
</table>

153
egrmR359a  USGUUUGAAAAAUUGUUCAGG

UC  AG  AAA
5’  USGUUUGAAAAAUUGUUC  GGAAUAGAAAUA  A
                      \              \  
3’  ACUAAUUUUACAGG  CCUCUGUCUUAU  A
    AA         GG          AAA

egrmR359b  USGUUUGAAAAAUUGUUCAGG

UC  A  A  GAA
5’  USGUUUGAAAAAUUGUUC  GGGAACA  AAAUA  A
                      \              \  
3’  ACAAGUUUUAACAGG  CCCUUGU  UUUGU  A
    AA         A       C   GAA

egrmR359c  USGUUUGAAAAAUUGUUCAGG

UC  A
5’  USGUUUGAAAAAUUGUUCAGG  GGAAACAGAA  U
                      \              \  
3’  ACAAGUUUUAACAGG  CCCUUGUCUUC  A
    AA         C

egrmR359d  USGUUUGAAAAAUUGUUCAGG

U  AG  C  GAA
5’  UCUGUCAAAAAUUGUUCAGG  GGAA  AGAAAUA  A
                      \              \  
3’  AGAC  AGUUUUUAACAGG  UCUUUCUUUGU  A
    C       G  A   GAA

egrmR359e  USGUUUGAAAAAUUGUUCAGG

U  AG  AAA
5’  UCUGUCAAAAAUUGUUCAGG  GGAAACAGAAAUA  \  
                      \              \  
3’  AGAC  AGUUUUUAACAGG  CCCUUGUCUUGU  /  
    C       AA  GUAA

egrmR359f  USGUUUGAAAAAUUGUUCAGG

UC  U  AG  AAA
5’  UCUGUCAAAAAUUGUUCAGG  GGAAACAGAAAUA  \  
                      \              \  
3’  AC  AGUUUUUAACAGG  CCCUUGUUUGU  /  
    AA  C       GG       AAA

egrmR359g  USGUUUGAAAAAUUGUUCAGG

UC  AG  G  AA
5’  UGUUUCAAAAAUUGUUCAGG  GGAAACA  AAAAG  A
                      \              \  
3’  ACAAGUUUUGACAAGG  CCCUGU  UUUAUC  A
    AA  AG  A  AA

154
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  CUUGGAUGAAGGGAGCUCC

egr-miR159b  CUUGGAUGAAGGGAGCUCC

egr-miR159c  CUUGGAUGAAGGGAGCUCC

egr-miR159d  CUUUGAAGAAGGGAGCUCC

egr-miR159e  AUUGGACUGAAGGGAGCUCC

egr-miR159f  AUUGGACUGAAGGGAGCUCC
egr-miR160a UGCCUGGCUCCCUGUAUGCCA

UCU C CC A A UU-- UUC
5’ GC UGGCUCC GUAGCCA C AAC UA CCAACC A
|| ||||| |||| || ||| ||
3’ CG ACGGAGG CAUGCUGGUA UUG AU GGUGGA U
AUC A AC G C UACC UUC

egr-miR160b UGCCUGGCUCCCUGUAUGCCA

C CU C C A
5’ UGGUCC UC AAGCCA UGGCUAG CCA CGG G
|| ||||| |||| || ||| ||
3’ AUACG ACGGAGG UAGCGGUA UUGUGUC GU GGC A
U AG U C A A

egr-miR160c UGCCUGGCUCCCUGAAUGCCA

U C GA U \ / AGA
5’ GC UGGCUCCU AAUGCCA UUGCCAA C
|| ||||| |||| || ||| ||
3’ CG ACGGAGGG UAGCGGU UAUCUUC UCAG UUGU GG A
A U AG U - U- AGGA- AAG

egr-miR160d UGCCUGGCUCCCUGAAUGCCA

U C U UU CC GUUGA AAA
5’ GC UGGCUCC-UAGCCA CUAGGAG GGUU AAACA CC GG
|| ||||| |||| || ||| ||
3’ CG ACUGAGGAGACU-AAGGU UAGUCUUC UCAG UUGU GG A
A U U U- U- AGGA- AAG

egr-miR168a UCGCUUGUGCACGGUCGGA

GGU C U A UG --GA AUUGCCAGAUGGCUCGACAGUCU
5’ CUCUGAU UCG UUGUGCAAG GGGGA C AUCUGGC UUG
|| ||||| |||| || ||| ||
3’ GAGGCUAA UACUACGUUC GCCCU G UAAAG- G AAAC G
--G C C A GU AUAG AAAAAAAGGACAAAGGAAAGAAA

egr-miR168b UCGCUUGUGCACGGUCGGA

GG IU C U A UG --GAUU AG AC
5’ UUCU AUUCG UUGUGCAAG GGGGA C AUUC C UGAU- GCC AUGCUCAA A
|| ||||| |||| || ||| ||
3’ AGAG UAGU UACUACGUUC GCCCU G UAAAG G UGUC GGGU C
-- GC C C A GU AGAA AAAAAU AG AG

egr-miR172a AGAAUCUUGAUGCAUGCUAGCAU

A
5’ GUGUAAGCAUCUAAGGUAUUC CAAIU 94 nt
|| ||| ||||| |||| COMPLEX
3’ UACGUGUAGUAGUUCUAAG GUGUG LOOP
**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.
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...
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

<table>
<thead>
<tr>
<th>miRNA&lt;sup&gt;a&lt;/sup&gt;</th>
<th>miRNA Sequence (5' - 3')&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Length (nt)&lt;sup&gt;c&lt;/sup&gt;</th>
<th>miRNA genome reference&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Arm&lt;sup&gt;e&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>egr-miR31</td>
<td>UUCAAAUCAUGAUAAUCAGACU (1)</td>
<td>22</td>
<td>E0152813f.ab1.1253.1234_501</td>
<td>5'</td>
</tr>
<tr>
<td>egr-miR90</td>
<td>AGGUUGGCCUUCCACGACU (2)</td>
<td>20</td>
<td>E0121074f.ab1.1273.1267_378</td>
<td>5'</td>
</tr>
<tr>
<td>egr-miR140</td>
<td>AUUACAAAGUGCAUGACUGA (1)</td>
<td>20</td>
<td>E1045769f.ab1.205.1149_343</td>
<td>5'</td>
</tr>
<tr>
<td>egr-miR200a</td>
<td>CGCACCCUCUGCCAUGACAGC (1)</td>
<td>22</td>
<td>E0224981r.ab1.1.816_293</td>
<td>5'</td>
</tr>
<tr>
<td>egr-miR200b</td>
<td>CGCACCCUCUGCCAUGACAGC</td>
<td>E0709307f.ab1.1.748_225</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>egr-miR200c</td>
<td>CGCACCCUCUGCCAUGACAGC</td>
<td>E1062193r.ab1.1.811_159</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>egr-miR200d</td>
<td>CGCACCCUCUGCCAUGACAGC</td>
<td>E0130629f.ab1.1.858_335</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>egr-miR200e</td>
<td>CGCACCCUCUGCCAUGACAGC</td>
<td>E0761665f.ab1.1.624_101</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>egr-miR293a</td>
<td>GGUCCCGCCUGCCGUAUGCG</td>
<td>E1403267r.ab1.1.605_48</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>egr-miR293b</td>
<td>GGUCCCGCCUGCCGUAUGCG</td>
<td>E1230135r.ab1.1.616_59</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>egr-miR293c</td>
<td>GGUCCCGCCUGCCGUAUGCG</td>
<td>E1564604r.ab1.295.1288_464</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>egr-miR293d</td>
<td>GGUCCCGCCUGCCGUAUGCG</td>
<td>E1244745r.ab1.1.1031_464</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>egr-miR293e</td>
<td>GGUCCCGCCUGCCGUAUGCG</td>
<td>E1573050r.ab1.1.627_107</td>
<td>3'</td>
<td></td>
</tr>
<tr>
<td>egr-miR293f</td>
<td>GGUCCCGCCUGCCGUAUGCG</td>
<td>E1683802r.ab1.1.909_220</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>egr-miR302a</td>
<td>UUAAAGUCCUGAUAUGUAGGCU</td>
<td>E0467787f.ab1.1.975_453</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>egr-miR302b</td>
<td>UUAAAGUCCUGAUAUGUAGGCU</td>
<td>E1248077r.ab1.1.687_165</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>egr-miR302c</td>
<td>UUAAAGUCCUGAUAUGUAGGCU</td>
<td>E1626235f.ab1.1.1032_472</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>egr-miR302d</td>
<td>UUAAAGUCCUGAUAUGUAGGCU</td>
<td>E4044437r.ab1.1.746_224</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>egr-miR302e</td>
<td>UUAAAGUCCUGAUAUGUAGGCU</td>
<td>E0246027r.ab1.4.946_501</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>egr-miR302f</td>
<td>UUAAAGUCCUGAUAUGUAGGCU</td>
<td>E0121191r.ab1.1.831_163.331</td>
<td>5'</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> 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.
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.

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

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).

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

<table>
<thead>
<tr>
<th>miRNAa</th>
<th>miRNA Sequence (5' - 3')b</th>
<th>Length (nt)c</th>
<th>miRNA genome referenced</th>
<th>Armse</th>
</tr>
</thead>
<tbody>
<tr>
<td>egr-miR159a</td>
<td>UUUGGAUUGAAGGGAGCUCU (5)</td>
<td>20</td>
<td>E0164506f.ab1.229.1248_337</td>
<td>3'</td>
</tr>
<tr>
<td>egr-miR159b</td>
<td>UUUGGAUUGAAGGGAGCUCU</td>
<td></td>
<td>E0085684f.ab1.157.1176_322</td>
<td>3'</td>
</tr>
<tr>
<td>egr-miR159c</td>
<td>UUUGGAUUGAAGGGAGCUCU</td>
<td></td>
<td>E0505209r.ab1.1.920_402</td>
<td>3'</td>
</tr>
<tr>
<td>egr-miR159d</td>
<td>CUGGAUUGAAGGGAGCUCU</td>
<td></td>
<td>E1498862r.ab1.1.976_375</td>
<td>3'</td>
</tr>
<tr>
<td>egr-miR159e</td>
<td>AUGCAGCAGAAGGGAGCUCU</td>
<td></td>
<td>E0435958r.ab1.360.1252_355</td>
<td>3'</td>
</tr>
<tr>
<td>egr-miR159f</td>
<td>AUGCAGCAGAAGGGAGCUCU</td>
<td></td>
<td>E0406035r_051.ab1.1.894_376</td>
<td>3'</td>
</tr>
<tr>
<td>egr-miR160a</td>
<td>UGCCCCUGGCUCCCUGUAGC (13)</td>
<td>21</td>
<td>E0246270f_094.ab1.1.194.953_504</td>
<td>5'</td>
</tr>
<tr>
<td>egr-miR160b</td>
<td>UGCCCCUGGCUCCCUGUAGC</td>
<td></td>
<td>E0443862f.ab1.360.1295_441</td>
<td>5'</td>
</tr>
<tr>
<td>egr-miR160c</td>
<td>UGCCCCUGGCUCCCUGUAGC</td>
<td></td>
<td>E0238876r.ab1.1.778_191</td>
<td>5'</td>
</tr>
<tr>
<td>egr-miR160d</td>
<td>UGCCCCUGGCUCCCUGUAGC</td>
<td></td>
<td>E0634921r.ab1.1.942_422</td>
<td>5'</td>
</tr>
<tr>
<td>egr-miR168a</td>
<td>UGCUGUGAGCGAGCUGGGA</td>
<td></td>
<td>E0593603f.ab1.1.689_179</td>
<td>5'</td>
</tr>
<tr>
<td>egr-miR168b</td>
<td>UGCUGUGAGCGAGCUGGGA</td>
<td></td>
<td>E0264641f_065.ab1.255.944_314</td>
<td>5'</td>
</tr>
<tr>
<td>egr-miR172a</td>
<td>AGAAUCUUAGAUGAUGCAGCUA (5)</td>
<td>21</td>
<td>E0220437r.ab1.1.589_69</td>
<td>3'</td>
</tr>
<tr>
<td>egr-miR172b</td>
<td>AGAAUCUUAGAUGAUGCAGCUA</td>
<td></td>
<td>E0162604f.ab1.1.733_213</td>
<td>3'</td>
</tr>
<tr>
<td>egr-miR172c</td>
<td>AGAAUCUUAGAUGAUGCAGCUA</td>
<td></td>
<td>E0058639r_039.ab1.1.946_390</td>
<td>3'</td>
</tr>
<tr>
<td>egr-miR172d</td>
<td>AGAAUCUUAGAUGAUGCAGCUA</td>
<td></td>
<td>E0182002r.ab1.1.913_315</td>
<td>3'</td>
</tr>
<tr>
<td>egr-miR172e</td>
<td>AGAAUCUUAGAUGAUGCAGCUA</td>
<td></td>
<td>E0525459f.ab1.56.1076_423</td>
<td>3'</td>
</tr>
<tr>
<td>egr-miR472a</td>
<td>UUUCCAAGGCGCGCCAUCC (2)</td>
<td>21</td>
<td>E0381021f.ab1.1.970_450</td>
<td>3'</td>
</tr>
</tbody>
</table>

---

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>
<thead>
<tr>
<th>miRNA&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Protein annotation&lt;sup&gt;b&lt;/sup&gt;</th>
<th>P. trichocarpa and Arabidopsis target genes&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Predicted function&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>egr-miR31</td>
<td>Putative ethylene-responsive CBP</td>
<td>fgenesh1_pg.C_LG_V000143(2.5), At2g22300/AtSR1(2)</td>
<td>Signal transduction</td>
</tr>
<tr>
<td></td>
<td>SCAR complex component</td>
<td>At2g35110(2)</td>
<td>Trichome morphogenesis</td>
</tr>
<tr>
<td></td>
<td>Pectinase</td>
<td>At4g23820(2.5)</td>
<td>Carbohydrate metabolism</td>
</tr>
<tr>
<td></td>
<td>Patatin</td>
<td>At2g39220(2.5)</td>
<td>Lipid metabolism</td>
</tr>
<tr>
<td></td>
<td>Cysteine protease</td>
<td>At1g50670(2.5)</td>
<td>Proteolysis</td>
</tr>
<tr>
<td></td>
<td>CCAAT-binding factor</td>
<td>eugene3.00130940(2.5), At1g08970</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>egr-miR90</td>
<td>MADS-box family protein</td>
<td>At3g18650(1.5)</td>
<td>Transcription factor</td>
</tr>
<tr>
<td></td>
<td>Ubiquitin protein ligase</td>
<td>At1g55860(2.5)</td>
<td>Protein ubiquitination</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>eugene3.00061771(2.5), eugene3.00040592(2.5), eugene3.00130940(2.5), eugene3.00040592(2.5), At1g08970</td>
<td>Unknown</td>
</tr>
<tr>
<td>egr-miR140</td>
<td>Peptidase</td>
<td>At5g43600(2.5)</td>
<td>Proteolysis</td>
</tr>
<tr>
<td></td>
<td>LIM domain containing protein</td>
<td>At5g17890(2.5)</td>
<td>Protein interactions</td>
</tr>
<tr>
<td></td>
<td>Putative small molecule transporter</td>
<td>fgenesh1_pm.C_LG_X000537(2.5), eugene3.20760001(2.5), At1g14040(2.5)</td>
<td>Signal transduction</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>estExt_fgenesh1_pg_v1.C_LG_VIII0033(2.5)</td>
<td>Unknown</td>
</tr>
<tr>
<td>egr-miR200</td>
<td>PPR</td>
<td>fgenesh1_pm.C_LG_XVI000226(2.5)</td>
<td>Organelle biogenesis</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>At5g44170(2.5), At3g06930(2.5), eugene3.00131142(2.5), eugene3.00131142(2.5), fgenesh1.pg.C_scaffold_12963000001(2.5)</td>
<td>Unknown</td>
</tr>
<tr>
<td>egr-miR293</td>
<td>Putative disease resistance protein</td>
<td>fgenesh1_pg.C_LG_VII000915(1)</td>
<td>Defense response</td>
</tr>
<tr>
<td></td>
<td>PAPS reductase</td>
<td>grail3.8743000201(2.5)</td>
<td>Biosynthesis</td>
</tr>
<tr>
<td></td>
<td>Protein kinase</td>
<td>fgenesh1_pg.C_LG_VII001044(2.5)</td>
<td>Phosphorylation</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>estExt_Genewise1_v1.C_LG_VIII2152(2.5), eugene3.00180993(2.5)</td>
<td>Unknown</td>
</tr>
<tr>
<td>egr-miR320</td>
<td>Predicted phosphate acyltransferase</td>
<td>eugene3.00130312(2.5), estExt_fgenesh1_pg_v1.C_LG_700114(2.5)</td>
<td>Lipid metabolism</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>fgenesh1.pg.C_scaffold_766000001(2.5), eugene3.00102297(2.5)</td>
<td>Unknown</td>
</tr>
<tr>
<td>egr-miR359</td>
<td>Oxidoreductase</td>
<td>estExt_fgenesh1_pg_v1.C_LG_X2093(2)</td>
<td>Electron transport</td>
</tr>
<tr>
<td></td>
<td>AT-rich domain containing protein</td>
<td>eugene3.00180022(2.5), At3g43240(2.5)</td>
<td>Transcription regulation</td>
</tr>
<tr>
<td></td>
<td>ABC transporter</td>
<td>eugene3.00140576(2.5), eugene3.00011077(2)</td>
<td>Signal transduction</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>grail3.0143000601(2.5), fgenesh1_pg.C_LG_1000711(2.5)</td>
<td>Unknown</td>
</tr>
</tbody>
</table>
MicroRNAs from Eucalyptus

<table>
<thead>
<tr>
<th>egr-miR362 Ras GTPase protein</th>
<th>eugene3.00150925(2.5)</th>
<th>Signal transduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Putative calmodulin</td>
<td>eugene3.00051226(2.5)</td>
<td>Signal transduction</td>
</tr>
<tr>
<td>Unknown</td>
<td>fgenesh1_pg.C_LG_V000073(2.5), eugene3.00190215(2.5), At1g29630(2.5)</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

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*

<table>
<thead>
<tr>
<th>miRNA&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Protein annotation&lt;sup&gt;b&lt;/sup&gt;</th>
<th><em>P. trichocarpa</em> and <em>Arabidopsis</em> target genes&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Predicted function&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>egr-miR159 MYB</td>
<td>eugene3.00400064(2), estExt_fgenesh1_pg_v1.C_LG_I1359(2), eugene3.00011133(2), grai3.0100003501(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</td>
<td>Asparagine synthase</td>
<td>Transcription factor activity</td>
</tr>
<tr>
<td>egr-miR160 Auxin responsive factor</td>
<td>fgenesh1_pg.C_LG_V000905(0.5), eugene3.0020832(0.5), eugene3.00280173(0.5), eugene3.00280175(0.5), estExt_Genewise1_v1.C_LG_XVI2818(0.5), fgenesh1_pm.C_LG_I0000578(1), eugene3.00660262(1), fgenesh1_pm.C_LG_VIII0000107(1), estExt_fgenesh1_pg_v1.C_LG_X0730(1)</td>
<td>Unknown</td>
<td>Transcription factor activity</td>
</tr>
<tr>
<td>egr-miR168 ARGONAUTE1 Vesicle coat protein complex COP1</td>
<td>grai3.012200281(4), At1g48410/AGO (3), eugene3.00102068(2.5), estExt_Genewise1_v1.C_LG_XV1316(2.5), grai3.0110002401(0)</td>
<td>Known</td>
<td>MicroRNA biogenesis</td>
</tr>
<tr>
<td>egr-miR172 Homeotic protein APETALA2</td>
<td>fgenesh1_pg.C_LG_X001967(0.5), eugene3.00050501(0.5), grai3.0019003502(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)</td>
<td>Unknown</td>
<td>Transcription factor activity</td>
</tr>
<tr>
<td>egr-miR472 Putative disease resistance protein</td>
<td>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</td>
<td>Unknown</td>
<td>Defense response</td>
</tr>
</tbody>
</table>
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


171
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.


### Table S3.1. Primers used for *Eucalyptus* miRNA expression analysis

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5' - 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Egr-miR31</td>
<td>TTACAATTCAGTACATCAGACT</td>
</tr>
<tr>
<td>Egr-miR90</td>
<td>AGCTGGGTGCTTGACTTTGG</td>
</tr>
<tr>
<td>Egr-miR140</td>
<td>ATTAGAACAAGTGCAATAGA</td>
</tr>
<tr>
<td>Egr-miR159</td>
<td>TTGGATTTGAAGGGAGCTCTA</td>
</tr>
<tr>
<td>Egr-miR160</td>
<td>TGCCCTGGCTCCCTGTATGCCAA</td>
</tr>
<tr>
<td>Egr-miR168</td>
<td>TCGCTTGGTGCAAGGTCGGAA</td>
</tr>
<tr>
<td>Egr-miR172</td>
<td>GGAATCTTGATGCTGCTGCAA</td>
</tr>
<tr>
<td>Egr-miR200a</td>
<td>CGCACCCTCATCGGCCACAAGCA</td>
</tr>
<tr>
<td>Egr-miR200e</td>
<td>TGCAACCCTCATCGACCACAAGCA</td>
</tr>
<tr>
<td>Egr-miR293</td>
<td>GTGGGCCCTGCACATGGCTGGA</td>
</tr>
<tr>
<td>Egr-miR320</td>
<td>TAAGTGGCGCAATTGTATGGGCTTA</td>
</tr>
<tr>
<td>Egr-miR359</td>
<td>TCTGTAAAATATGTTCCAGGA</td>
</tr>
<tr>
<td>Egr-miR362</td>
<td>TCCTGGAACATCAACATCTGAA</td>
</tr>
<tr>
<td>Egr-miR472</td>
<td>TTCCCAAGGCCGCCCAATTCC</td>
</tr>
<tr>
<td>Egr-5.8S rRNA_Forward</td>
<td>ACCTGTGCCTGGGTCACAA</td>
</tr>
<tr>
<td>Poly-T adaptor</td>
<td>GACCACGCGATCGATGGCTCAT16V*</td>
</tr>
<tr>
<td>Poly-T adaptor_Reverse</td>
<td>GACCACGCGATCGATGGCTCA</td>
</tr>
</tbody>
</table>

*V=A,G,C
### Table S3.2. Precursor sequences of putative novel and conserved miRNAs identified in the *E. camaldulensis* genome sequences

<table>
<thead>
<tr>
<th>Gene names&lt;sup&gt;a&lt;/sup&gt;</th>
<th>miRNA genome reference&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Precursor Sequences (5' - 3')&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>egr-MIR31</td>
<td>E0152813f.ab1.255.1234_501.609</td>
<td>TTACAATCCAGTACACACTGACCTCTGCGCAAATCTAGAGATTCTGATGATCAGCCTACCGCAATCTAGATTGAA</td>
</tr>
<tr>
<td>egr-MIR90</td>
<td>E0121074f.ab1.273.1267_378.520</td>
<td>AGGTGGTGCTTGTGCTTTGCGCTAGTTAGGAGACCTGAGCCGGCCAGGCTCCAACTAGCAGCAGGCGCACCCAACCT</td>
</tr>
</tbody>
</table>
| egr-MIR140             | E1045769f.ab1.205.1149_343.520 | ATTAGGGCACAAGTGGATAGATTAGTCTTCTACTGACACTGAGGAGCCGCTTACGCTACAGAGG |}

<sup>a</sup>Abbreviation in this column indicates the miRNA family.

<sup>b</sup>Abbreviation in this column indicates the gene name.

<sup>c</sup>For each miRNA sequence, the precursor sequence is given as the mature miRNA sequence followed by the full-length miRNA precursor sequence.
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
Concluding remarks

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 expression...
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 \textit{MIR} genes are transcribed 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 \textit{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


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.