

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

MicroRNA and mRNA analysis of two species of New Zealand
Pachycladon

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Science

in

Genetics

at Massey University, Manawatu, New Zealand.

Louise Michelle Carr

2014

Abstract

MicroRNAs (miRNAs) are small, non-coding RNAs important in post-transcriptional regulation. In this study, potential miRNAs from two New Zealand *Pachycladon* species, *P. cheesemanii* and *P. fastigiatum*, are identified and compared.

Sixteen miRNAs were differentially expressed between the species, most of which have roles in flower and leaf development. Potential targets for 15 miRNAs were located in expressed sequence tag (EST) libraries for *P. cheesemanii* and/or *P. fastigiatum*, including a new potential relationship in *P. cheesemanii* between miR825 and MYB29 (AT5G07690), a transcription factor involved in the synthesis of methionine-derived glucosinolates.

From the results of the differential expression analysis and target identification, 27 miRNAs from 21 miRNA families were chosen for pre-miRNA sequencing. Sequences of 15 *P. cheesemanii* miRNA hairpins and 13 *P. fastigiatum* miRNA hairpins were validated experimentally.

Additionally, mRNA-Seq data obtained at the same time as the miRNAs were analysed. A gene ontology analysis indicated enriched terms for defence responses and miRNAs in *P. fastigiatum*.

This study is the first investigation of the miRNAs present in *Pachycladon* and how their differential expression contributes to the adaptive divergence between the species.

Acknowledgements

Firstly, I thank my supervisor Claudia Voelckel for all your support and guidance over the past two years. Thanks to Peter Lockhart for being a wonderful officemate, and thanks to Trish McLenachan for assistance with lab work and proofreading. I could not have asked for a more caring group of people to work with, and I truly appreciate all of the help you have given me over the time we have shared.

Thanks to Lesley Collins for microRNA wisdom and helping to shape this project, and thanks to Nicole Gruenheit for helping with the bioinformatics. Thanks to everyone who assisted with sequencing and extraction of small RNA and mRNA data.

Thank you to my friends, and especially my flatmates, for all your companionship. And lastly, huge thanks to my parents, brother, and grandparents for your unerring love and support.

Table of Contents

Abstract.....	ii
Acknowledgements.....	iii
Table of Contents.....	iv
List of Figures.....	viii
List of Tables.....	ix
Supplementary Tables and Files.....	ix
Abbreviations.....	x
Nucleotide Ambiguity Code.....	xi
Chapter One: Literature Review	1
1.1 Introduction	1
1.2 MiRNA Biogenesis and Function	1
1.2.1 Biogenesis of miRNAs in plants	1
1.2.1.2 Processing in the nucleus.....	1
1.2.2.2 Processing in the cytosol.....	1
1.2.2 Functions of miRNAs.....	2
1.2.2.1 Roles of miRNAs in stress responses.....	2
1.2.2.2 MiRNAs and development	2
1.3 Approaches to Studying miRNAs.....	2
1.3.1 Identifying miRNA genes	2
1.3.2 Experimental and computational approaches to find miRNA targets	4
1.4 MiRNAs and Evolution	5
1.4.1 MiRNA families.....	5
1.4.2 MiRNA loss and gain.....	6
1.4.3 MiRNAs and polyploidy	6
1.5 <i>Pachycladon</i>	7
1.6 Project Overview	8
1.6.1 Background Information.....	8

1.6.2 Aims.....	9
Chapter Two: Identification and Differential Expression Analysis of miRNAs	10
2.1 Introduction	10
2.2 Methods.....	11
2.2.1 Small RNA sequencing.....	11
2.2.2 Sequence trimming and sorting.....	11
2.2.3 rRNA and tRNA filtering.....	11
2.2.4 Identification of miRNAs using small RNA data.....	11
2.2.5 Mining EST libraries of <i>Pachycladon</i> for miRNAs.....	12
2.2.6 Differential expression analysis	12
2.3 Results.....	13
2.3.1 Identification of miRNAs	13
2.3.1.1 Data Pre-processing.....	13
2.3.1.2 Identification of potential homeologs	14
2.3.2 Mining <i>Pachycladon</i> EST libraries for miRNAs.....	16
2.3.3 Differential expression analysis	18
2.3.3.1 Changing the number of mismatches	18
2.3.3.3 Variant Analysis	19
2.4 Discussion	20
2.4.1 Data Pre-processing	20
2.4.2 The effect of mismatches.....	20
2.4.3 Variant Analysis.....	21
2.4.4 Mining EST libraries for miRNAs	21
2.4.5 MiRNAs of interest from this section	22
2.4.5.1 MiRNAs up-regulated in <i>P. cheesemanii</i>	22
2.4.5.2 MiRNAs up-regulated in <i>P. fastigiatum</i>	22
Chapter Three: Identification of Potential miRNA Targets.....	25
3.1 Introduction	25
3.2 Methods.....	25

3.3 Results.....	27
3.3.1 miRNA-EST relationships that are the same in <i>A. thaliana</i> and <i>Pachycladon</i>	27
3.3.2 Potential loss of target in <i>Pachycladon</i>	29
3.3.3 Potential gain of target.....	31
3.4 Discussion	33
3.4.1 Bowtie vs psRNAtarget	33
3.4.2 Potential losses and gains of targets in <i>Pachycladon</i>	34
3.4.3 Differences between the species	35
3.4.4 MiRNAs with interesting target profiles.....	37
Chapter Four: Verification of miRNAs.....	38
4.1 Introduction	38
4.2 Methods.....	38
4.2.1 MiRNA candidate selection	38
4.2.2 Primer design	39
4.2.3 Experimental procedure	39
4.2.4 Hairpin prediction.....	40
4.3 Results and Discussion.....	41
4.3.1 PCR amplification	41
4.3.2 Validation of mature sequences by PCR	44
4.3.3 Hairpin prediction.....	45
4.3.4 Future Work	54
Chapter Five: mRNA-Seq Analysis	55
5.1 Introduction	55
5.2 Materials and Methods.....	55
5.2.1 mRNA sequencing and data pre-processing.....	55
5.2.2 Differential expression analysis	56
5.2.3 GO Analysis	56
5.2.3.1 Gene set enrichment analysis using R.....	56
5.2.3.2 AgriGO	57

5.2.4. MiRNA and mRNA expression comparison.....	57
5.3 Results.....	57
5.3.1 Differentially expressed genes	57
5.3.2 GO Analysis	60
5.3.2.1 Gene set enrichment analysis	60
5.3.2.2 AgriGO	61
5.3.2.3 GO terms shared between the R enrichment analysis and AgriGO results....	62
5.3.3 Is miRNA up-regulation predictive of target gene down-regulation?.....	63
5.4 Discussion	65
5.4.1 Approaches to identifying enriched GO terms	65
5.4.1.1 Averaging duplicate accession numbers	65
5.4.1.2 AgriGO versus gene set enrichment.....	65
5.4.2 Differentially expressed genes and GO terms	66
5.4.2.1 Genes up-regulated in <i>P. fastigiatum</i>	67
5.4.2.2 Genes up-regulated in <i>P. cheesemanii</i>	67
5.4.3 Comparing the expression of miRNAs and their targets	68
Chapter 6.0 Discussion	69
6.1 MiRNA Analysis by Small RNA-Seq: Pros & Cons.....	69
6.1.1 Reliability of small RNA sequencing.....	69
6.1.2 Limitations of EST libraries in miRNA analyses	69
6.1.3 Using another species as a reference	70
6.2 Biological Aspects.....	71
6.2.1 Differentially expressed miRNAs and target genes	71
6.2.2 MiRNAs of Interest.....	72
6.2.3 Comparing information obtained from miRNA and mRNA-Seq analyses.....	74
6.3 Conclusion.....	75
6.3.1 Future Work.....	76
References	77
Appendix.....	82

List of Figures

Figure 1: The pre-miRNA hairpin of miR157a in <i>A. thaliana</i> , a typical hairpin structure	3
Figure 2: MiRNA target identification using degradome sequencing.....	5
Figure 3: Phylogenetic tree of <i>Pachycladon</i>	9
Figure 4: Potential miR157 loci in <i>Pachycladon</i>	10
Figure 5: Comparison of the number of miRNAs identified in <i>P. cheesemanii</i> and <i>P. fastigiatum</i> with varying number of mismatches.....	14
Figure 6: Potential homeologs of miRNA genes	16
Figure 7: ClustalX alignment of miR414 (AT1G67195)	16
Figure 8: Hairpins of miR414 in <i>Pachycladon</i>	17
Figure 9: MiRNA-target relationships confirmed in <i>A. thaliana</i> that are present in <i>Pachycladon</i>	27
Figure 10: Predicted miRNA targets present in <i>Pachycladon</i> that do not map	30
Figure 11: ClustalX alignment of AT4G25210 in <i>A. thaliana</i> and <i>P. fastigiatum</i> , a potential loss of target in <i>Pachycladon</i>	31
Figure 14: ClustalX alignment of AT5G10480 in <i>P. fastigiatum</i> and <i>P. cheesemanii</i> , an example of a different miRNA-target relationship between the species.....	36
Figure 15: Example alignment for primer design.....	39
Figure 16: Amplification of miRNAs 160a, 396b, 398b, 472, 825, 848 and 852 in <i>Pachycladon</i>	42
Figure 17: <i>Pachycladon</i> miR825 mature sequences obtained by validation sequencing.....	45
Figure 18: Predicted <i>Pachycladon</i> miRNA hairpin structures.....	46
Figure 19: Alignment of miR472 in <i>Pachycladon</i> and <i>Arabidopsis</i>	52
Figure 20: Alignment of miR852 pre-miRNA regions in <i>Pachycladon</i> and <i>Arabidopsis</i>	53
Figure 21: Log fold change vs log p-value of 5685 mRNAs expressed in both <i>Pachycladon</i> species.	58
Figure 22: MiRNA expression versus target expression for <i>Pachycladon</i> miRNA-target relationships confirmed in <i>A. thaliana</i>	64
Figure 23: MiRNA expression vs target expression for new miRNA-target relationships in <i>Pachycladon</i>	64
Figure 24: MiRNA expression vs target gene expression for potential loss of targets in <i>Pachycladon</i>	65

List of Tables

Table 1: The effect of changing mismatches on the identification of differentially expressed miRNAs.....	18
Table 2: Differentially expressed sequence variants.	19
Table 3: Functions of targets predicted for only one <i>Pachycladon</i> species due to sequence difference between the species	37
Table 4: PCR and sequencing results of pre-miRNAs.	43
Table 5: Pairwise percent identity for miR852 of <i>Pachycladon</i> and <i>Arabidopsis</i>	54
Table 6: Top 10 differentially expressed genes up-regulated in <i>P. fastigiatum</i>	59
Table 7: Top 10 differentially expressed genes up-regulated in <i>P. cheesemanii</i>	59
Table 8: Top 10 GO terms enriched in <i>P. cheesemanii</i> using gene set enrichment.....	60
Table 9: Top 10 GO terms enriched in <i>P. fastigiatum</i> using gene set enrichment	61
Table 10: Top 10 GO terms up in <i>P. cheesemanii</i> using AgriGO	61
Table 11: Top 10 GO terms up in <i>P. fastigiatum</i> using AgriGO	62
Table 12: GO terms enriched in <i>P. cheesemanii</i> for both the R and AgriGO analyses.....	62
Table 13: GO terms enriched in <i>P. fastigiatum</i> for both the R and AgriGO analyses	63

Supplementary Tables and Files

Supplementary Table 1: Data pre-processing of small RNA reads	82
Supplementary Table 2: Sequences of miRNAs identified in <i>Pachycladon</i> species.....	82
Supplementary Table 3: Sequences of potential homeologs in <i>Pachycladon</i>	84
Supplementary Table 4: Primers designed for PCR of miRNA genes	85
Supplementary Table 5: Eighty-one genes up-regulated in <i>P. fastigiatum</i>	87
Supplementary Table 6: Genes up in <i>P. cheesemanii</i>	90
Supplementary File 1: Script for identification of differentially expressed miRNAs	95
Supplementary File 2: Genbank accession numbers for <i>Pachycladon</i> pre-miRNA sequences	97
Supplementary File 3: Pre-miRNA sequences of <i>P. cheesemanii</i> and <i>P. fastigiatum</i> in FASTA format	97
Supplementary File 4: Script for identification of differentially expressed mRNAs.....	100
Supplementary File 5: Script for identification of enriched GO terms	101

Abbreviations

Aly	<i>A. lyrata</i>
Ath	<i>A. thaliana</i>
BLAST	basic local alignment search tool
bp	base pair
EST	Expressed Sequence Tag
GO	Gene Ontology
kb	kilobase
logFC	log fold change
miRNA	microRNA
mRNA	messenger RNA
Mya	million years ago
nt	nucleotide
PC, Pch	<i>P. cheesemanii</i>
PCR	polymerase chain reaction
PF, Pfa	<i>P. fastigiatum</i>
RNA	ribonucleic acid
rRNA	ribosomal RNA
tRNA	transfer RNA

Nucleotide Ambiguity Code

Code	Represents	Complement
A	Adenine	T
G	Guanine	C
C	Cytosine	G
T	Thymine	A
Y	Pyrimidine (C or T)	R
R	Purine (A or G)	Y
W	weak (A or T)	W
S	strong (G or C)	S
K	keto (T or G)	M
M	amino (C or A)	K
D	A, G, T (not C)	H
V	A, C, G (not T)	B
H	A, C, T (not G)	D
B	C, G, T (not A)	V
X/N	any base	X/N