2016

Transcriptional and Post-Transcriptional Regulation of Nodule-Specific Gene Expression in Soybean

Sajag Adhikari
South Dakota State University

TRANSCRIPTIONAL AND POST-TRANSCRIPTIONAL REGULATION OF NODULE-SPECIFIC GENE EXPRESSION IN SOYBEAN

BY

SAJAG ADHIKARI

A dissertation submitted in partial fulfillment of the requirements for the

Doctor of Philosophy

Major in Plant Science

South Dakota State University

2016

TRANSCRIPTIONAL AND POST-TRANSCRIPTIONAL REGULATION OF NODULE-SPECIFIC GENE EXPRESSION IN SOYBEAN

This dissertation is approved as a creditable and independent investigation by a candidate for the Doctor of Philosophy in Plant Science degree and is acceptable for meeting the dissertation requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Senthil Subramanian, PhD
Dissertation Advisor Date

David Wright, PhD
Head, Department of Plant Science Date

Dean, Graduate School Date

Dedicated to my parents:

Kumar Narsingh Adhikari and Bal Kumari Adhikari

ACKNOWLEDGEMENTS

I would like to acknowledge my advisor Dr. Senthil Subramanian for providing necessary guidance and support throughout my PhD studies and research. His vast knowledge and experience on the subject helped me to achieve my research goals.

I would like to thank members of my graduate student committee, Dr. Anne Fennell, Dr. Xijin Ge, and Dr. Cristina Lammers for their valuable suggestions and candid discussions for the improvement of my research.

I am thankful to collaborators Dr. Blake C. Meyers and Dr. Siwaret Arikit, functional genomics core facility of SDSU, for allowing me to use necessary equipment to conduct my research and funding agencies US Department of Agriculture and South Dakota Soybean Research & Promotion Council.

I am grateful to Brian Moore and Ms. Padmapriya Swaminathan for bioinformatics support.

I would like to express my thanks to my colleagues and lab members Dr. Marie turner, Suresh Damodaran, Yoaan Lo Monaco, Dr. Narashima Rao Nizampatnam, Laura White, Spencer Schreier, Dr. Carl Fellbaum, Sunita Pathak, Haque Md Ehsanul, Shuchi Smita, Paul Gaillard, and Krishna Acharya for their direct and indirect help, co-operation, and company.

My sincere gratitude to my parents, siblings, and all other family members. Their continuous love and support throughout the years is highly remarkable.

I would like to acknowledge my little boy Sujay whose smile kept me motivated. It’s hard to put into words how much I love you.

Finally, but most importantly, I thank my husband Laxman Karki. Without his unending support, encouragement and love, I could have never completed the PhD program.

TABLE OF CONTENTS

LIST OF FIGURES xiii

LIST OF TABLES xix

ABBREVIATIONS xxi

ABSTRACT xxii

Chapter 1 1

1. Literature review 1

1.1. Biological Nitrogen Fixation 1

1.2. Symbiotic Nitrogen Fixation 2

1.2.1. Development of a Root Nodule 2

1.2.2. Molecular Mechanisms Involved in Root Nodule Development 5

1.2.3. Epidermal Responses 5

1.2.3.1. Infection Thread Formation and Growth 6

1.2.3.2. Cortical Responses and Primordium Formation 7

1.2.3.3. Primordium Formation and Nodule Development 8

1.2.4. Regulation of Gene Expression during Nodule Development 11

1.2.4.1. Transcriptional Regulation of Nodule Development 11

1.2.4.2. Post-transcriptional Regulation of nodule development 13

1.2.5. Evolution of nodule development 17

1.3. Objectives 21

References 25

Chapter 2 48

2. Comparative transcriptomics reveals distinct transcription factors and hormone action pathways in nodules and lateral roots of soybean 48

2.1. Introduction 48

2.2. Results 52

2.2.1. Overview of the transcriptome libraries 52

2.2.2. Predominant mapping to coding sequences and consistency among replicates in our libraries 59

2.2.3. Novel Transcribed Regions (NTRs) in soybean 64

2.2.4. Mature nodules had the largest difference in global gene expression patterns 69

2.2.5. The most common biological process between EN and ELR was cell division 69

2.2.6. Transcriptional regulators in root lateral organs 74

2.2.7. Hormone biosynthesis and signaling patterns during lateral root and nodule development 77

2.2.7.1. Local auxin biosynthesis occurs in both nodules and LRs 77

2.2.7.2. TIR/AFB-mediated auxin signaling is predominant in LRs compared to nodules 78

2.2.7.3. Cytokinin biosynthesis and signaling are activated in nodules, but not in LRs 81

2.2.7.4. A coordinated increase in active gibberellin levels might occur during nodule development 84

2.2.7.5. Brassinolide activity is enhanced during lateral organ formation 86

2.2.7.6. Abscisic acid, ethylene, salicylic acid and jasmonic acid activities are reduced during LR and nodule development 88

2.2.7.7. Strigolactone activity is high in emerging nodules 97

2.2.8. A number of shoot axillary meristem markers are enriched in nodule tissues 99

2.3. Discussion 102

2.3.1. An accurate set of nodule enriched genes 102

2.3.2. Stage-specific action of nod factor signaling components 104

2.3.3. Transcription factor families associated with nodule development 105

2.3.4. Is the soybean nodule a modified shoot axillary organ? 106

2.4. Experimental Procedures 108

2.4.1. Plant material and RNA isolation 108

2.4.2. Transcriptome library preparation and sequencing 109

2.4.3. Quality control of Raw Reads 109

2.4.4. Read Alignment and Assembly 110

2.4.5. Identification and functional annotation of Novel Transcribed Regions 111

2.4.6. Singular Enrichment (SEA) analysis 112

2.4.7. Transcription Factors Analysis 112

2.4.8. Assessment of hormone biosynthesis and signaling 113

References 114

Chapter 3 139

3. Small RNAs in root nodules of soybean 139

3.1. Introduction 139

3.2. Materials and Methods 147

3.2.1. Plant growth and *B. japonicum* treatment 147

3.2.2. Tissue harvest/ RNA isolation 147

3.2.3. Small RNA library preparation and sequencing 150

3.2.4. PARE library preparation and sequencing 152

3.2.5. Processing of the reads 154

3.2.6. Identification of miRNA targets in the PARE data 154

3.3. Results 154

3.3.1. Small RNA library reads characterization 154

3.3.2. Processing of small RNA reads for miRNA prediction 158

3.3.3. Identification of novel miRNA sequences with potential hairpin-forming precursors 160

3.3.4. Differential expression of mature sequences in ABMN and MN libraries 167

3.3.5. Differential expression of known soybean miRNAs in ABMN and MN libraries 169

3.3.6. PARE library reads distribution 172

3.3.7. Identification of the gene targets of the potential mature miR sequences 176

3.3.8. Identification of the targets of known soybean miRNAs 177

3.4. Discussion 188

3.5. Conclusion 189

References 191

Chapter 4 199

4. Method Development for microRNA quantification 199

4.1. Introduction 201

4.2. Results and Discussion 206

4.2.1. Hairpin and polyA cDNA qPCR assays differ in their efficiencies in detecting 2'-OMe and 2'OH RNA molecules 206

4.2.2. polyA cDNA qPCR might result in erroneous abundance measurements of plant miRNAs 212

4.2.3. Use of U6 and miR1515 as normalization controls in hairpin cDNA qPCR 218

4.2.4. Multiplexing cDNA synthesis of miRNA and normalization controls for accurate and efficient qPCR assays 223

4.3. Material and methods 226

4.3.1. Plant materials, growth conditions and treatments 226

4.3.2. RNA isolation and synthesis 226

4.3.3. cDNA Synthesis and RT-qPCR 227

4.3.4. Northern hybridization 228

4.3.5. U6 and miR1515 as normalization control 228

Acknowledgements 228

References 230

Chapter 5 236

5. Role of microRNAs in root nodule development 236

5.1. Materials and Methods 241

5.1.1. DNA vectors 241

5.1.2. TOPO-TA cloning 241

5.1.3. LR clonase 242

5.1.4. Agrobacterium rhizogenes-mediated gene transfer 243

5.1.5. Hairy root transformation 245

5.1.6. *Bradyrhizobium* Inoculation and Nodulation Assays 245

5.1.7. Tissue Sampling and RNA isolation 248

5.1.8. cDNA Synthesis 248

5.1.9. Gene Expression Analysis 249

5.1.10. Bradyrhizobium japonicum GUS assay 250

5.1.11. Microscopy 251

5.1.12. Statistical Analysis 251

5.2. Results 252

5.2.1. Validation of overexpression 252

5.2.2. Validation of misexpression lines 255

5.2.3. miR169c role in controlling root nodule number 258

5.2.4. miR169c role in nodule maturation 264

5.2.5. miR169c role in early nodulation event 266

5.2.6. miR169c regulation of downstream genes 271

5.2.7. Spatial regulation of miR169c during root nodule formation 279

5.3. Discussion 283

5.4. Conclusion 287

References **289**

Appendix A **292**

LIST OF FIGURES

Figure 1.1 Stages of root nodule development in determinate nodule.. 4

Figure 1.2 Nod factor perception and signal transduction leading to epidermal response and cortical response.. 10

Figure 2.1 Representative images of root lateral organs and corresponding control tissues harvested for transcriptome library construction. 54

Figure 2.2 Summary of read alignment, distribution of novel transcribed regions and annotated transcripts in nodule and root transcriptome. 62

Figure 2.3 Distribution of FPKM (log transformed) in the transcriptome libraries. 63

Figure 2.4 Linear regression plot between FPKM vs Coverage. 66

Figure 2.5 Numbers of significantly enriched NTRs in each tissue. 67

Figure 2.6 Expression pattern of annotated genes and NTRs.. 68

Figure 2.7 Venn diagram showing the number of transcripts specifically enriched in each tissue and in more than one tissue..... 72

Figure 2.8 Co-expression clusters with different patterns of gene expression identified using K-means cluster analysis. 73

Figure 2.9 Number of family members with enriched expression in nodule or lateral root tissues in different transcription factor (TF) families. 76

Figure 2.10 Heat maps indicating organ-specific enrichment (log_2 significant fold change – organ vs. control) of different genes overlaid on (a) auxin biosynthesis and (b) signaling pathways. 80

Figure 2.11 Heat maps indicating organ-specific enrichment (log_2 significant fold change – organ vs. control) of different genes overlaid on (a) cytokinin biosynthesis and (b) signaling pathways. 83

Figure 2.12 Heat maps indicating organ-specific enrichment (log_2 significant fold change – organ vs. control) of different genes overlaid on (a) gibberellin biosynthesis and (b) signaling pathways. 85

Figure 2.13 Heat maps indicating organ-specific enrichment (log_2 significant fold change – organ vs. control) of different genes overlaid on (a) brassinolide biosynthesis and (b) signaling pathways. 87

Figure 2.14 Heat maps indicating organ-specific enrichment (log_2 significant fold change – organ vs. control) of different genes overlaid on (a) ABA biosynthesis and (b) signaling pathways. 89

Figure 2.15 Heat maps indicating organ-specific enrichment (log_2 significant fold change – organ vs. control) of different genes overlaid on (a) ethylene biosynthesis and (b) signaling pathways. 91

Figure 2.16 Heat maps indicating organ-specific enrichment (log_2 significant fold change – organ vs. control) of different genes overlaid on (a) jasmonic acid biosynthesis and (b) signaling pathways. 93

Figure 2.17 Heat maps indicating organ-specific enrichment (log_2 significant fold change – organ vs. control) of different genes overlaid on (a) salicylic acid biosynthesis and (b) signaling pathways.. 96

Figure 2.18 Heat maps indicating organ-specific enrichment (log_2 significant fold change – organ vs. control) of different genes overlaid on (a) strigolactone biosynthesis and (b) signaling pathways. 98

Figure 2.19 Heat map indicating organ-specific enrichment (log_2 significant fold change – organ vs. control) of different genes involved in shoot axillary meristem development. 101

Figure 3.1 Schematic representation of root-enriched miRNAs that play roles in nodule development. 146

Figure 3.2 Tissue harvest for small RNA and PARE libraries. 149

Figure 3.3 Schematic representation of small RNA library preparation. 151

Figure 3.4 Schematic representation of PARE library preparation. 153

Figure 3.5 Total reads distribution in the ABMN library. 156

Figure 3.6 Total reads distribution in the MN library. 157

Figure 3.7 Processing of the ABMN and MN small RNA libraries for miRNA prediction. 159

Figure 3.8 Processing of the miRNA precursor sequences. 161

Figure 3.9 : Schematic representation of miRNA precursors falling into three different categories. 163

Figure 3.10 Distributions of sequence lengths and first nucleotides of the mature miRNA sequences. 166

Figure 3.11 Differential expression of nodule-enriched vs root-enriched mature miRNA sequences. 168

Figure 3.12 Distribution of sequence length and first nucleotide of known soybean miRNAs. 170

Figure 3.13 Differential expression of known soybean miRNAs. 171

Figure 3.14 Distribution of 20-nt signature in ABMN library. 174

Figure 3.15 Distribution of 20-nt signature in MN library. 175

Figure 4.1 Schematic showing the principle of two methods in this study to generate cDNAs from mature miRNAs. 204

Figure 4.2 Comparison of hairpin cDNA qPCR and polyA cDNA qPCR to assay synthetic RNA molecules with and without 2'-O methylation. 209

Figure 4.3 Amplification plots from hairpin and polyA cDNA qPCR assays. 211

Figure 4.4 Comparison of hairpin cDNA qPCR and polyA cDNA qPCR to assay plant miRNAs.. 215

Figure 4.5 Dissociation curves from polyA-cDNA-qpCR assays. 216

Figure 4.6 Relative abundance of miR164. 217

Figure 4.7 U6 primer design. 219

Figure 4.8 Validation of the novel U6 fusion primer for cDNA synthesis on multiple plant species. 220

Figure 4.9 Expression of miR1515 in different soybean tissues 222

Figure 4.10 Multiplexing during hairpin cDNA synthesis did not affect assay efficiency. 225

Figure 5.1 Schematic representation of nodule-excluded miRNAs, the absence of which allows target expression in the nodule. 240

Figure 5.2 Representative image of emerging and mature root nodules of soybean. 247

Figure 5.3 Relative expression of miR2218, miR169c and miR1513c in corresponding overexpression lines. 253

Figure 5.4 Relative expression of target genes of miR2218, miR169c and 1513c in the corresponding overexpression lines.. 254

Figure 5.5 Relative expression of miR2218, miR169c, and miR1513c in the corresponding misexpression lines. 256

Figure 5.6 Relative expression levels of the target genes of miR2218, miR169c, and miR1513c in the corresponding misexpression lines..... 257

Figure 5.7 Effect of misexpression of miRNAs on nodule formation. 260

Figure 5.8 miR169c plays a role in nodule maturation..... 265

Figure 5.9 Effect of ENOD169 in root hair deformation..... 269

Figure 5.10 Effect of ENOD169 in nodule primordium formation. 269

Figure 5.11 Effect of overexpression of miR169 in early nodulation events. 270

Figure 5.12 Relative expression of miR169c in OX169c and ENOD169c lines compared to vector control. 273

Figure 5.13 Relative expression of targets of miR169c in OX169c and ENOD169c lines compared to vector control. 274

Figure 5.14 miR169c regulation of GA20ox gene..... 276

Figure 5.15 miR169c regulation of CKX genes. 278

Figure 5.16 Spatial regulation of miR169c in root tip of soybean plants. 280

Figure 5.17 Spatial regulation of miR169c in emerging nodules of soybean plants.. 281

Figure 5.18 Spatial regulation of miR169c in mature nodules of soybean plants. 282

Figure 5.19 Schematic representation of miR169 action in nodule development. 288

LIST OF TABLES

Table 1.1 Studies that identified or evaluated miRNA regulation during nodule development in legumes. 14

Table 1.2 Similarities and differences between lateral roots and nodules. 20

Table 2.1 Expression and enrichment of selected organ-specific genes.. 57

Table 2.2 Average number of reads. 61

Table 3.1 Read distribution in ABMN and MN small RNA libraries.. 155

Table 3.2 Summary result of categorization of precursors.. 165

Table 3.3 Read distribution in ABMN and MN PARE libraries. 173

Table 3.4 List of previously unreported mature sequence/target pairs.. 179

Table 3.5 List of potential miRNA/target pairs that show an inverse expression in small RNA vs transcriptome libraries. 181

Table 3.6 Annotation of nodule-enriched and root-enriched targets. 182

Table 3.7 Known soybean miRNAs and targets validated by PARE data.. 183

Table 3.8 Known soybean miRNA sequences that showed an inverse expression patterns.. 186

Table 3.9 Annotation of the targets of known soybean miRNAs (Table 3.8). 187

Table 4.1 Major advantages and disadvantages of methods used to assay mature miRNA levels in plants..... 203

Table 4.2 Ct value from qPCR assays using cDNA synthesized from different amounts of 2’-OH and 2’-OMe RNA molecules. 210

Table 5.1 List of primers used for cloning and qPCR of miR2218, miR169c and miR1513c. 244

Table 5.2 Summary of emerging nodules count (average ± SE) in misexpression lines of miRNAs vs VC. 261

Table 5.3 Summary of mature nodule count (average ± SE) in misexpression lines of miRNAs vs VC.. 262

Table 5.4 Summary of total nodule count (average ± SE) in misexpression lines of miRNAs vs VC.. 263

Table 5.6 Role of miR169 in nodule primordium (NP) formation.. 268

ABBREVIATIONS

°C = Centrigrade

µl = microliter

Cat no: = catalogue number

cDNA = complementary DNA

DNA = Deoxyribonucleic Acid

dpi = days post inoculation

h = hours

min = minutes

miRNAs = microRNAs

nt = nucleotide

PARE = Parallel Analysis of RNA Ends

qPCR = quantitative Polymerase Chase Reaction

RACE = RNA ligase mediated 5’ Rapid amplification of cDNA ends

RNA = Ribonucleic Acid

rpm = rotation per minute

s = second

x g = relative centrifugal force

ABSTRACT

TRANSCRIPTIONAL AND POST-TRANSCRIPTIONAL REGULATION OF NODULE-SPECIFIC GENE EXPRESSION IN SOYBEAN

SAJAG ADHIKARI

2016

Lateral roots and nodules are two important nutrient acquisition organs in soybean. The evolutionary origin of nodules from lateral roots has been highly hypothesized based on morphological similarities and genetic studies, but gene expression profiles during the formation of these organs have not been compared. In addition, the role of post-transcriptional gene regulation during nodule development has not been thoroughly explored. Bridging these knowledge gaps is crucial to develop genetic/biotechnological strategies to optimize nutrient acquisition and sustainable production of crops. To answer some of the outstanding questions about regulation of gene expression during nodule development, (i) global transcriptome analyses of lateral root and nodule were compared to identify organ-specific enrichment patterns of transcription factors and hormone signaling elements; (ii) small RNA and degradome/Parallel Analysis of RNA Ends (PARE) libraries were generated to identify miRNAs and their cleavage products respectively in nodule tissues; (iii) miRNA qPCR quantification methods were optimized; and (iv) the effect of misexpression of selected miRNAs on nodule development were evaluated.

Analysis of transcriptome data showed very little overlap in transcription factor expression profiles between emerging nodules and emerging lateral roots. The expression

profiles of certain key hormone biosynthesis and signaling genes were distinct between nodules and lateral roots. Interestingly, members of gene families associated with shoot axillary meristem formation were enriched in nodules, but not in lateral roots.

Analysis of small RNA and PARE libraries resulted in the identification of 497 previously unknown miRNA precursors and validated 353 miRNA-target pairs. These and additional results suggested that inverse expression of miRNA and target is likely to be one of the mechanisms that direct nodule-specific gene expression. In addition, methods for miRNA quantification by qPCR were optimized, and a potential role for miR169 in regulating hormone homeostasis during nodule development was identified through functional assays.

In summary, nodules might have adopted not only the developmental pathways of lateral roots, but also shoot axillary meristems. Furthermore, the inverse expression of miRNAs and their targets between nodules and adjacent root tissues might be a mechanism that spatially limits target gene expression to the nodule and/or root tissues.

Chapter 1

1. Literature review

1.1. Biological Nitrogen Fixation

Nitrogen is a key component of a number of essential biomolecules, such as DNA, RNA, and protein (Howard and Rees 1996). Plants are the direct or indirect sources of nitrogen in all animals. In plants, nitrogen is an important component of key biomolecules such as chlorophyll, amino acids, Adenosine Triphosphate (ATP), and nucleic acids, making it one of the limiting elements for plant growth and development (Vance 2001, Wagner 2011). Nitrogen is available in three forms: 1) dinitrogen or N_2; 2) bound to carbon (organic form); and 3) nitrogen nutrients (inorganic forms) (Socolow 1999). Dinitrogen is the most available form of nitrogen in the atmosphere; organic forms of nitrogen include biomolecules such as amino acids; and inorganic forms of nitrogen include nitrogen ions [ammonium ($NH4^+$) and nitrate ($NO3^-$)] and nitrogen gases [ammonia (NH_3), nitric oxide, (NO), nitrogen dioxide (NO_2)] (Socolow 1999). Plants obtain nitrogen either by assimilation of nitrate and ammonium from soil (primary forms of artificial fertilizer), by decomposition of organic matter, or by biological nitrogen fixation (BNF) (Stougaard 2000, Vance 2001). BNF is a mechanism by which atmospheric dinitrogen is converted into ammonia by a group of prokaryotes aided by the enzyme nitrogenase (Howard and Rees 1996, Halbleib and Ludden 2000). The reaction during BNF (Hoffman et al. 2014) occurs as follows:

$$N_2 + 8e^- + 8H^+ + 16ATP \rightarrow 2\ NH_3 + H_2 + 16ADP + 16\ Pi$$

The prokaryotes that carry out BNF could be symbiotic, associative, or free-living in soils and water (Dixon and Kahn 2004). *Azotobacter, Clostridium, Rhodospirilum* and various cyanobacteria are free-living nitrogen-fixing (NF) prokaryotes; *Azosprillium, Klebsiella, Bacillus, Acetobacter,* and *Pseudomonas*-like genera represent associative NFing prokaryotes; while *Rhizobium* spp. are symbiotic form (Brill 1980, Elmerich et al. 1992).

1.2. Symbiotic Nitrogen Fixation

Bacteria that form a symbiotic association with plants belonging to the Leguminosae family are collectively known as "rhizobia" (singular "rhizobium") (Wang et al. 2012). Rhizobia include the genera *Allorhizobium, Bradyrhizobium, Mesorhizobium, Rhizobium,* and *Sinorhizobium* (Ferguson et al. 2010). The symbiotic association results in the formation of unique structures in plants known as root nodules. Exceptions to the legume-rhizobia symbiotic relationship include *Parasponia*-rhizobia symbiosis and *Frankia*-non-legume symbiosis. *Parasponia,* a medium-sized tropical tree, is the only non-legume host plant that is nodulated by rhizobia (Trinick 1973, Akkermans et al. 1978). *Frankia* spp. induce formation of nodules in eight dicotyledonous families commonly known as actinorhizal plants; nodules formed in these plants are known as actinorhizal nodules (reviewed by (Perrine-Walker et al. 2011, Pawlowski and Demchenko 2012, Santi et al. 2013)). Bacteria inside the root nodule convert atmospheric nitrogen to ammonia, which is readily available to the plants; bacteria in return get carbohydrate from the plant.

1.2.1. Development of a Root Nodule

Rhizobium infects the root tissue of the legume plant, and the interaction leads to the formation of root nodule. During this process, compatible rhizobia bacteria perceive a

flavonoid compound produced by the host plants, which leads to the activation of *nod* genes and secretion of nod factors (NFs) by the bacteria. The bacteria attach to the root hair, causing deformation and elongation of the root hair tip and encapsulation of bacteria (Figure 1.1). At the point of attachment, the cell wall is partially dissolved and the bacteria enter inside the root hair cell by the invagination of the plasma membrane known as infection thread which is surrounded by cell wall like material. At the meantime, pericycle cells get activated, and a nodule primordium is formed by division of cortical cells. In temperate legumes like *Medicago truncatula*, *Medicago sativa,* and *Pisum sativum*, the initial cell division occurs at the inner cortical cell followed by continuous cell division at the apical region of the nodule, resulting in an indeterminate nodule. In tropical legumes such as *Glycine max* and *Lotus japonicus*, the initial cell division occurs in the outer cortical cells. Subsequently, the mitotic activity stops and nodule growth occurs through expansion (Nap and Bisseling 1990). Pericycle cells continue to divide, and the masses of dividing cells from the pericycle and cortex fuse to form a single clump, or globular nodule, in plants like *G. max* and *L. japonicus*, or elongated nodules with persistent meristem in plants like *M. truncatula* and *M. sativa*. The infection thread grows into the primordia cells. The branched infection threads release the bacteria into the host cytoplasm by an endocytotic process. These bacteria are surrounded by a host-derived peribacteroid membrane, and the resulting structures are known as symbiosomes. Inside the symbiosome, the bacterium multiplies and differentiates into bacteroids that fix the atmospheric nitrogen (reviewed by (Nap and Bisseling 1990, Mylona et al. 1995, Gage 2004, Ferguson et al. 2010)) (Figure 1.1).

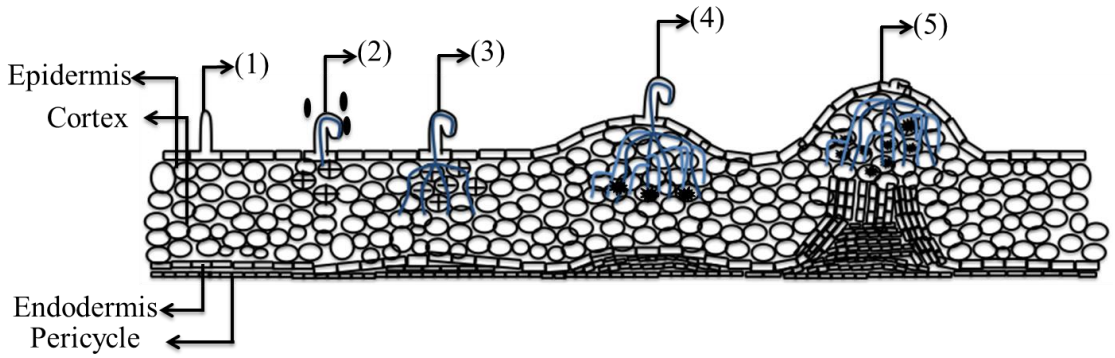


Figure 1.1 Stages of root nodule development in determinate nodule. Secretion of flavonoid compounds by the root hair. (2) Perception of flavonoids by rhizobia bacteria; bacterial attachment to root hair; formation of infection thread. In response, outer cortex cells divide. (3) Growth and branching of infection thread into the dividing cortex cells. Pericycle cells also divide. (4) Cell division continues in outer cortex cells and pericycle cells. Bacteria are released into the primordia cells. (5) Dividing cortex and pericycle cells fuse, yielding a globular nodule. Nodule continues to grow by elongation and differentiation. This figure is adopted from Ferguson et al. (Ferguson et al. 2010).

1.2.2. Molecular Mechanisms Involved in Root Nodule Development

The perception of NFs by NF receptors in legume trigger simultaneous transcriptional cascades in the epidermal and cortical cells, resulting in infection and root nodule organogenesis, respectively (Desbrosses and Stougaard 2011). The molecular responses will be summarized in the following sections.

1.2.3. Epidermal Responses

At the epidermal cells, membrane-localized Lysine Motif (LysM) receptor-like kinases (RLKs) perceive the NFs, and have been described in several legumes: *L. japonicus* NFR1 and NFR5 (for Nod Factor Receptor 1 and 5), *P. sativum* PsSYM2A and PsSYM10, *M. truncatula* MtLYK3/MtLYK4 (for LysM domain-containing receptor-like Kinases (LYKs)) and MtNFP (for Nod factor perception), and *G. max* GmNFR1α/β and GmNFR5α/β (for LysM-type receptor kinase genes) (Limpens et al. 2003, Radutoiu et al. 2003, Indrasumunar and Gresshoff 2010, Indrasumunar et al. 2010, Moling et al. 2014). The perception of NFs causes calcium spiking in the nucleus that is mediated by potassium channel proteins encoded by LjCASTOR/POLLUX, MtDMI1 (for Does not Make Infections 1), MtDMI2 (for Does Not Make Infections 2), MtHMGR1 (for 3-Hydroxy-3-MethylGlutaryl coenzyme A Reductase 1) (Ane et al. 2004, Imaizumi-Anraku et al. 2005, Peiter et al. 2007, Charpentier et al. 2008, Venkateshwaran et al. 2015) and nucleoporins encoded by LjNUP133 and LjNUP85 (Kanamori et al. 2006, Saito et al. 2007). Ca^{2+} ions bind to calcium and a calmodulin-dependent protein kinase (CCaMK) encoded by MtDMI3 (for Does Not Make Infections 3) /PsSYM9, leading to auto phosphorylation of the CCaMK and of the associated protein CYCLOPS (Mitra et al. 2004, Yano et al. 2008). The activation of CCaMK and CYCLOPS activates

transcription factors (TFs) like NSP1, NSP2 (for Nodulation Signaling Pathway1 and 2), ERN (for (Ets2 Repressor Factor (ERF) Required for Nodulation) and NIN (for Nodule Inception) (Schauser et al. 1999, Catoira et al. 2000, Oldroyd and Long 2003, Smit et al. 2005, Marsh et al. 2007, Middleton et al. 2007, Vernie et al. 2015). These TFs work together to regulate expression of *ENOD* genes (for early nodulation) in the epidermis, followed by initiation and infection of root hair by the bacteria. Various *ENOD* genes have been described in *P. sativum* (*PsENOD5* and *PsENOD12), M. truncatula (MtENOD11*and *Mt* ENOD12), and *G. max (GmENOD40)* (Scheres et al. 1990a, Yang et al. 1993, Journet et al. 1994) (Figure 1.2).

1.2.3.1. Infection Thread Formation and Growth

Infection through the root hair is one of the advanced modes by which rhizobia gain entry to plant root (Madsen et al. 2010). The infection thread is formed by inward tubular growth of the root hair cell plasma membrane at the site of infection by the bacteria and grows into the dividing cortical cells. The *Nap1* (for *Nck-associated protein 1*) and *Pir1* (for *121F-specific p53 inducible RNA*) genes in *L. japonicus* as well as the *CERBERUS* and *Vapyrin (VPY*) genes in *M. truncatula* play important, individual roles in infection thread growth. These roles include actin rearrangement, elongation, and penetration of the cortical cells for rhizobia release (Yano et al. 2009, Yokota et al. 2009, Murray et al. 2011). The genetic dissection of the organogenesis and infection pathways in *L. japonicus* showed that nod receptors (NFR1 and NFR5), nucleoporins (NUP133 and NUP85) and potassium channel proteins (CASTOR/POLLUX) are required for both infection thread formation and infection of the nodule primordia, whereas *Nap1*, *Pir1*,

and *CERBERUS* are required for the proper progression of infection thread growth (Madsen et al. 2010).

1.2.3.2. Cortical Responses and Primordium Formation

Following the perception of NFs in the epidermis, a rapid response also occurs in the cortical cells of the root, as suggested by expression of *ENOD40* and by cytoskeleton rearrangements (Yang et al. 1993, Timmers et al. 1999, Mathesius et al. 2000a). NIN restricts expression of *ENOD11* in epidermis but promotes activity of *CRE1* (for Cytokinin Response 1) in cortex (Vernie et al. 2015). The authors (Vernie et al. 2015) proposed that NIN or its product might play a role in the communication between the epidermis and cortex. However, the signaling mechanism from the epidermal cells to cortical cells required for nodule morphogenesis is unknown, although evidence suggests that the signaling molecule might be cytokinin (Downie 2014). For instance, the application of cytokinin alone is able to induce formation of nodule primordia followed by expression of nodulin genes (Heckmann et al. 2011). Moreover, a gain-of-function mutation of *LHK1* (for Lotus Histidine Kinase 1) forms spontaneous nodules without NF perception (Tirichine et al. 2007). Finally, cytokinin is perceived by a cytokinin receptor in the root cortex encoded by MtCRE1/LjLHK1 (Gonzalez-Rizzo et al. 2006b, Tirichine et al. 2007). The perception of cytokinin leads to the activation of NSP1-, NSP2-, and NIN-mediated organogenesis (Hayashi et al. 2010, Madsen et al. 2010) (Figure 1.2). Nuclear factor Y (NF-Y), a heterotrimeric CCAAT box-binding protein complex consisting of three subunits A, B and C, NF-YA1 and NF-YB1 are the targets of NIN and are responsible for initial cell division during nodule organogenesis (Soyano et al. 2013).

In *M. truncatula* NF-YA1 (HAP2-1 (for Heme Activator Protein2-1)) is involved in meristem persistence (Combier et al. 2006).

1.2.3.3. Primordium Formation and Nodule Development

In response to infection and entry of the rhizobia, the pericycle cell near to the protoxylem poles gets activated, followed by cell division of inner and outer cortical cells, and eventual development of indeterminate nodules. For determinate nodules, cell division occurs first in outer cortical cells, and followed by division in the pericycle cells. The activation of the pericycle cell opposite to protoxylem cells in both kinds of nodules might be induced by uridine, also named as a stele factor, while ACC oxidase prevents the activation of cells in the protophloem poles (reviewed by (Cohn et al. 1998, Gage 2004)). In response to infection and organogenesis, two major classes of genes are induced in host plants: *ENOD* and late nodulin genes. *ENOD* genes are expressed during initiation of nodule development and are NF stimulated, while late nodulin genes are primarily expressed during onset of nitrogen fixation (Nap and Bisseling 1990, Horvath et al. 1993). In response to rhizobia infection or nod factor application, *ENOD12* and *RIP1* are expressed in the root epidermis, while *ENOD40* is expressed in the nodule primordia (reviewed by (Oldroyd 2001)). *ENOD40* is continually expressed during nodule development in uninfected cells of the central tissue and nodule vascular bundles in determinate nodules (Yang et al. 1993). In *M. sativa*, it is expressed in nodule meristem, cells at the periphery of the central region, and in the vascular bundle (Fang and Hirsch 1998). *ENOD2* is expressed in nodule parenchyma of both indeterminate and determinate nodules independent of the infection by the symbiotic bacteria, suggesting its role in nodule morphogenesis (Franssen et al. 1987, van de Wiel et al. 1990, Zhao et al.

2012). What determines the development of nodule primordia cells into various cell types in nodules, such as infected cell, non-infected cortical cells and vascular bundles, is not completely known.

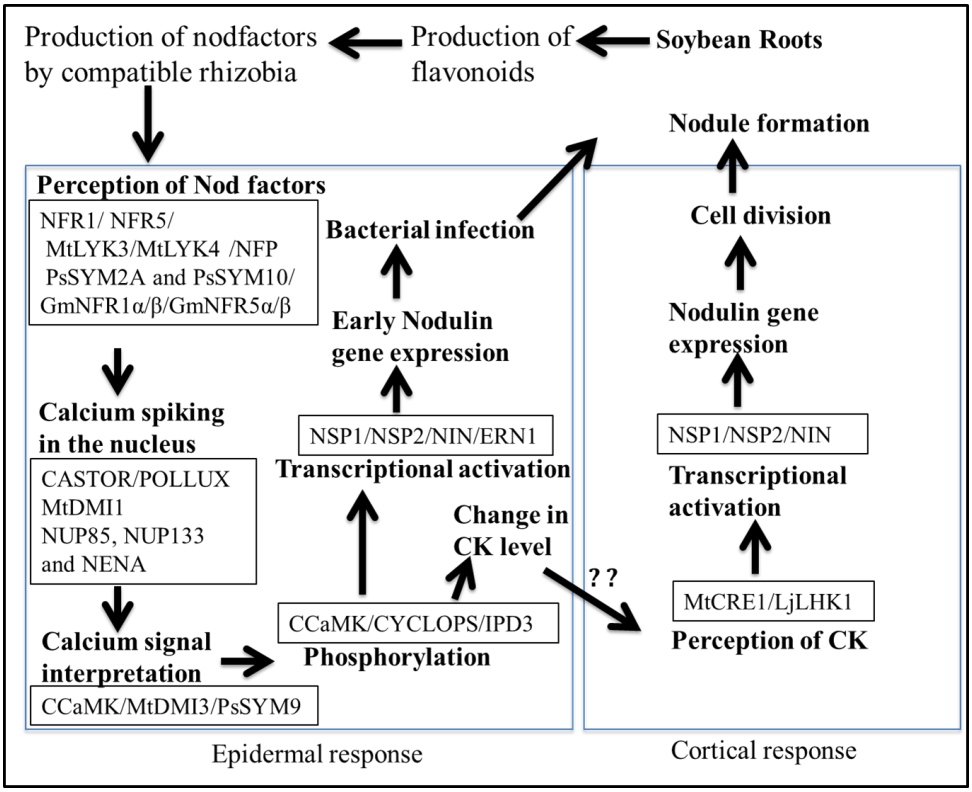


Figure 1.2 Nod factor perception and signal transduction leading to epidermal response and cortical response. The pathway is modified from (Ferguson et al. 2010).

1.2.4. Regulation of Gene Expression during Nodule Development

The phenotype of a cell or of an organism as a whole is a result of highly regulated gene expression (Mata et al. 2005), and understanding its expression profile provides insights into biological function (DeRisi et al. 1997). Two successive events happen to make phenotype possible, first, the transcription of DNA into mRNA and, second, translation of mRNA into protein. The amount of mRNA made through transcription of a gene is controlled by the combined effects of the structural properties of DNA and its interaction with TFs, known as transcriptional regulation (Phillips 2008). TFs have an ability to interpret the regulatory information encoded in the DNA, such as enhancers and promoters (Kadonaga 2004). After the successful transcription of mRNA, the mRNA goes through several other levels of regulation, such as processing, export, and translation, known as post-transcriptional regulation. Post-transcriptional control of mRNA is also mediated through RNA binding proteins (RBPs) and small RNA-mediated silencing pathways (Mata et al. 2005). The development and functioning of nitrogen fixing root nodules require the precise spatial regulation of gene expression, some of which is described at the transcriptional and post-transcriptional levels in the following sections.

1.2.4.1. Transcriptional Regulation of Nodule Development

Expression of the *ENOD* and late nodulin genes during nodule development has been reported in several legumes, including *G. max* (Gloudemans et al. 1987, Kouchi and Hata 1993), *P. sativum* (Bisseling et al. 1983, Scheres et al. 1990b), and *M. sativa* (Lang-Unnasch and Ausubel 1985, Vance et al. 1985, Norris et al. 1988). To unravel the mechanism of nodule-specific gene expression, the study of promoter regions of the late

nodulin genes N23, leghemoglobin (*lbc3*) in soybean and *Srglb3* gene in *Sesbania rostrata* has revealed four regulatory regions: a strong positive element (enhancer), a weak positive element (WPE), an organ-specific element (OSE) with conserved AAAGAT-taTTGT-CTCTT box, and a negative element (NE) (Stougaard 2000). However, the nodule-specific expression of *Srglb3* was driven by using a root-specific promoter and, in addition to the nodule-specific expression, its activity was observed in root and phloem cells of stem and petiole when transferred into tobacco plants (Szabados et al. 1990, Szczyglowski et al. 1994, Szczyglowski et al. 1996). Very little information on regulatory proteins in the promoter regions of nodulin genes has been identified. *NAT1*, *NAT2*, CPP1 have been identified as a trans-acting proteins in the promoter of soybean *lbc3* gene (Jensen et al. 1988, Jacobsen et al. 1990, Cvitanich et al. 2000). Similarly, NMH7 (Heard and Dunn 1995) and HAP2-1 (Combier et al. 2006) represent very few TFs known to play a role in nodule development.

Recent genomic techniques have been used to identify global gene expression changes during various stages of nodule development in legumes such as *M. truncatula*, *L. japonicus*, and *G. max*. This includes gene expression changes during establishment of the symbiosis (Lohar et al. 2006, Libault et al. 2010a, Breakspear et al. 2014) and at various stages (7 dpi (days post inoculation), 14 dpi, 21 dpi) of nodule growth (Franssen et al. 1987, Kouchi et al. 2004, Lee et al. 2004, Brechenmacher et al. 2008, Høgslund et al. 2009, Libault et al. 2010b, Moreau et al. 2011, Limpens et al. 2013). These studies have identified potential TFs involved in nodule development, but their functional role is yet to be discovered.

1.2.4.2. Post-transcriptional Regulation of nodule development

Non-coding RNAs, such as siRNAs and miRNAs, regulate gene expression at the post-transcriptional level (Phillips 2008). miRNAs are 20- to 24-nucleotide long RNAs transcribed from MIR genes by RNA polymerase II. During biogenesis of plant miRNAs, a single-stranded primary transcript that forms a stem-loop structure is cleaved by DICER-LIKE 1 (DCL1) into premiRNA and miRNA/miRNA* duplex, sequentially. The majority of the known plant miRNAs cleave target mRNA, and very few of them block translation of the target gene (reviewed by (Meng et al. 2011). miRNAs are reported to be present at various stages of root nodule development and play a regulatory role. For example, members of conserved and novel miRNA families were expressed during early stages of infection and after nodules were formed in various legumes, such as soybean (Subramanian et al. 2008, Wang et al. 2009, Joshi et al. 2010, Yan et al. 2015a), *M. truncatula* (Lelandais-Briere et al. 2009), and *L. japonicus* (De Luis et al. 2012, Breakspear et al. 2014). Targets of many plant miRNAs are TFs that play regulatory roles (Rhoades et al. 2002, Jones-Rhoades et al. 2006); several TFs in nodules are under miRNA-mediated regulation (Table 1.1).

Table 1.1 Studies that identified or evaluated miRNA regulation during nodule development in legumes. This table lists studies by various authors on miRNA-mediated regulation of nodule development in various legumes. Legume species, miRNA with their target and function identified in the study is provided.

Authors	**Species**	**miRNA/target/role**
(Li et al. 2010)	*G. max*	miR482/ miR1512/miR1515/ nodule numbers
(Turner et al. 2013)	*G. max*	miR160/ ARFs/ inhibits nodule primordia development
(Yan et al. 2013)	*G. max*	miR156, gma-miR172/ AP2 TFs/non symbiotic hemoglobin
(Wang et al. 2014)	*G. max*	miR172/ Nodule Number Control1/represses expression of the early nodulin gene, ENOD40
(Wang et al. 2015)	*G. max*	miR167c/ GmARF8a and GmARF8b/increase nodule and lateral root number
(Yan et al. 2015b)	*G. max*	miR393j-3p/ ENOD93/reduce nodule number
(Yan et al. 2016)	*G. max*	gma-miR4416/hizobium-induced peroxidase 1 (RIP1)-like peroxidase

(Yan et al. 2016)	*G. max*	gma-miR2606b/Mannosyl-oligosaccharide 1, 2-alpha-mannosidase gene/nodule number
(De Luis et al. 2012)	*L. japonicus*	miR171c/NSP2/bacterial infection in nodules
(De Luis et al. 2012)	*L. japonicus*	miR397/member of laccase copper protein family/nitrogen fixation related copper family
(Combier et al. 2006)	*M. truncatula*	miR169/ MtHAP2-1/ meristem maintenance
(Boualem et al. 2008)	*M. truncatula*	miR166/HD-ZIPIII/vascular bundle pattern, lateral root and nodule number
(Lelandais-Briere et al. 2009)	*M. truncatula*	miR2586/miR107/miR396 tissue specific expression nodule meristem/ root tip/might play a role in stem
(D'Haeseleer et al. 2011)	*M. truncatula*	miR164/MtNAC1/nodule number
(Bustos-Sanmamed et al. 2013)	*M. truncatula*	miR160/root growth and nodule number

(Nova-Franco et al. 2015)	*Phaseolus vulgaris*	miR172/rhizobial infeciton, nodulation and nitrogen fixation

1.2.5. Evolution of nodule development

Plants develop lateral root and shoot organs throughout their lifetime, albeit in different ways. A group of cells in each shoot apical meristem (SAM) remain meristematic, giving rise to new cells that form above-ground, lateral organs like leaves, branches and flowers. Lateral roots are formed from root founder cells at the pericycle (Malamy and Benfey 1997). Roots of legume plants form additional root lateral organ known as root nodules, which develop by the dedifferentiation of pericycle and cortex cells in the root (Libbenga and Harkes 1973, Dudley et al. 1987, Yang et al. 1994, Timmers et al. 1999). Unlike lateral roots, nodules are formed as a result of symbiotic interaction between rhizobia and host legume plants, making it an interesting organ to study in relation to its origin and development (Hirsch et al. 1997).

Over the years, the legume nodule was hypothesized to originate from lateral root, stem or carbon storage organs (reviewed by (Hirsch et al. 1997)), with a root origin most widely studied. This is first based on the fact the nodule formation occurs close to the site of primary and secondary lateral roots (Nutman 1948), and further morphological resemblance of actinorhizal and *Parasponia* nodules to lateral root, including a central vasculature and pericycle as the site of origin (reviewed by (Hirsch et al. 1997), and similarities in organ formation and gene expression in cortical cells (Mathesius et al. 2000b). Support for the root origin of nodule is also based on identification of mutants such as *LATD* that affects the development of both the nodule and lateral root (Bright et al. 2005), mutant with affected in their ability to synthesize hormones gibberellins or brassinosteroids in roots of *P. sativum* (Ferguson et al. 2005), and loss of nodule identity and formation of root in *noot* and *coch mutants* of *M. truncatula* and *P. sativum*

respectively (Couzigou et al. 2012). Expression of the root meristem gene WOX5 (for WUSCHEl-related hemeobox 5) in wild type as well as the supernodulating mutants of *M. truncatula* and *P. sativum* (Osipova et al. 2012, Roux et al. 2014) and expression of the root meristem-specific *PLETHORA* gene in nodules (Franssen et al. 2015) also supports the root origin of the legume nodule.

An indeterminate nodule more resembles a lateral root, with its persistent meristem, than does a determinate nodule, which lacks a persistent meristem (Table 1.2). The hypothesized root origin of legume nodules primarily represent an indeterminate nodule, and we yet do not know if a similar pattern can be expected in a determinate nodule like that of soybean. Contrary to the observed similarities between a lateral root and a nodule as mentioned in the above studies, the auxin and cytokinin requirements for these organs are completely opposite. Auxin plays a positive role in lateral root formation (Casimiro et al. 2001, Swarup et al. 2008), but very low auxin activity is required for determinate nodule formation. Indeed, increased activity of auxin inhibits formation of the determinate nodule (Suzaki et al. 2012, Turner et al. 2013). Moreover, cytokinin promotes nodule formation (Gonzalez-Rizzo et al. 2006a, Murray et al. 2007, Plet et al. 2011), but inhibits lateral root formation (Gonzalez-Rizzo et al. 2006a, Li et al. 2006, Laplaze et al. 2007, Bielach et al. 2012).

Although a nodule is considered similar to a lateral root, it is distinct from a lateral root based on such characteristics as site of origin, initial cell division, vasculature structure, and response to hormones (Hirsch et al. 1997). There is a continuous interest to investigate how a nodule is originated and how its origin is associated with other plant organs, such as shoots and carbon storage organs (Hirsch et al. 1997). Despite active

research in rhizobia-legume biology, origin and development of nodule has yet to be addressed.

Table 1.2 Similarities and differences between lateral roots and nodules. The table is modified from Hirsch et al. (Hirsch et al. 1997)

Property	**Lateral Root**	**Indeterminate nodule**	**Determinate nodule**
Site of origin	Pericycle	Inner cortex	Outer cortex
Radial Position	Adjacent of xylem poles	Most adjacent to xylem poles	Most adjacent to xylem poles
Meristem	Subapical	Apical	Short lived
Vasculature	Central	Peripheral	Peripheral
Organ growth	Cell division and expansion	Primary cell division and expansion	Primarily cell expansion
Plant species	All higher plants	Temperate legumes (*P. sativum, M. truncatula*)	Tropical legumes (*G. max, L. japonicus*)

1.3. Objectives

Lateral roots and nodules are the major lateral organs in the root. The evolutionary origin of nodules from the lateral roots has been highly hypothesized based on morphological similarities and loss-of-function mutation studies that affect both lateral root and nodule formation. However, none of the studies have compared the transcriptome of a lateral root with that of a nodule in a way that would identify the similarities and differences in the development of these organs. Likewise, the interaction of rhizobia with the host plant leads to expression of *ENOD* and late nodulin genes, but very little information is available on the transcriptional regulators that cause these nodule-specific gene expression responses.

Recent work demonstrates that miRNAs play crucial roles during nodule development through the post-transcriptional regulation of transcriptional regulators (Table 1.1). This is a list of some miRNAs that play regulatory roles in nodule development, but there could be many more not yet identified. There are also many reported miRNAs that might play a role in nodule development, but require functional characterization to identify their role. In the following chapters, I will address the above-mentioned gaps in our understanding of root nodule development by using three main objectives.

Objective 1: Compare the transcriptomes of lateral roots and of symbiotic nodules to identify organ-specific transcription factor and hormone activities

The goal here is to evaluate the evolutionary relationships between lateral roots and nodules by comparing the activities of key transcription factors, and hormone biosynthesis and signaling pathways. To meet this objective, transcriptomes of lateral

roots and nodules at both emerging and mature stages of development would be compared using RNA sequencing (RNA-seq). Genes enriched in emerging nodules, mature nodules, emerging lateral roots, and young lateral roots in soybean would be identified by comparing global gene expression profiles between each of these organs and adjacent root segments. Transcriptomes of soybean root nodules have been reported previously (Lee et al. 2004, Brechenmacher et al. 2008, Libault et al. 2010b) but the limitations of these studies is that nodule-enriched genes were identified using uninoculated root as control and the nodule tissue alone was used to identify nodule-enriched genes, over-representing their nodule enrichment. While a lateral root transcriptome has been reported for *A. thaliana* (Okushima et al. 2005, Paponov et al. 2008, Swarup et al. 2008, Lewis et al. 2013), no lateral root transcriptome of soybean or any other legume has been reported. Unlike in *Arabidopsis*, lateral root development in *M. truncatula* involves pericycle as well as endodermis and inner cortex cells (Herrbach et al. 2014). By comparing the lateral root and nodule transcriptomes, this study would help to tease out the commonalities and differences in the developmental processes of these two organs and would help to answer if the nodule has adopted the developmental pathways of the lateral root or not. The nodule and lateral root transcriptome alone would help to identify developmental regulators during the formation of these organs that have not been reported before.

Objective 2: Identify miRNA-target pairs with inverse expression between nodules and adjacent root tissues

This objective would help to obtain support for the hypothesis that miRNAs enriched in root tissues adjacent to nodules might contribute to nodule-specific

expression of their targets. The approach is global identification and quantification of known and novel miRNAs and their targets by using small RNA-seq and Parallel Analysis of RNA ends (PARE) analysis, respectively. Studies have been conducted to identify miRNAs at various stages of nodule development in soybean root 3 h post-inoculation (Subramanian et al. 2008), nodules 7, 4, and 21 days after inoculation (Joshi et al. 2010), root hair 12, 18, 24 and 48 h post-inoculation (Yan et al. 2015a), and 28-day-old nodules (Wang et al 2008). However, the limitation of these studies is that non-inoculated roots were used to identify nodule-specific expression of miRNAs. To address this, small RNA and PARE libraries would be generated by harvesting only nodules (MN) from inoculated root and root sections above and below the nodule but devoid of any lateral organ (ABMN).

Objective 3: Evaluate the role of selected miRNA-target pairs with inverse expression between nodules and adjacent root tissues

The first specific aim is to optimize techniques to quantify miRNA abundance using quantitative PCR (qPCR). miRNA can be quantified by various methods such as cloning, northern hybridization and microarray analyses, but each of these methods have some limitations. For example, cloning is laborious and time-consuming and is not able to identify poorly expressed miRNAs, northern hybridization requires a large quantity of starting RNA, and microarray lacks the sensitivity and specificity required to reliably detect and quantify miRNAs. These limitation can be overcome by a specially designed RT-qPCR (for Reverse Transcriptase quantitative Polymerase Chain Reaction) assay (Varkonyi-Gasic et al. 2007). But this protocol (Varkonyi-Gasic et al. 2007) is limited to quantification of only one miRNA at a time and lacks an appropriate reference gene. To

address this limitation, an RT-qPCR technique will be optimized to quantify more than one miRNA and a miRNA reference gene will be identified.

The second specific aim is to overexpress or misexpress nodule-excluded miRNAs in nodule tissues and then evaluate nodule development in order to identify the roles of miRNA/targets in nodule development. To achieve the second aim, miRNAs of interest will be overexpressed or misexpressed in nodules using CsVMV and nodule-specific ENOD40 promoter respectively, and the effect on nodule development will be evaluated.

Overall, achieving these objectives would shed light on transcriptional and post-transcriptional regulation of root nodule-specific gene expression in soybean.

References

Akkermans, A.D.L., Abdulkadir, S. and Trinick, M.J. (1978). N2-fixing root nodules in Ulmaceae:Parasponia or (and) Trema spp.? *Plant Soil*, 49, 711-715.

Ane, J.M., Kiss, G.B., Riely, B.K., Penmetsa, R.V., Oldroyd, G.E., Ayax, C., Levy, J., Debelle, F., Baek, J.M., Kalo, P., Rosenberg, C., Roe, B.A., Long, S.R., Denarie, J. and Cook, D.R. (2004). Medicago truncatula DMI1 required for bacterial and fungal symbioses in legumes. *Science (New York, N.Y.)*, 303, 1364-1367.

Bielach, A., Podlešáková, K., Marhavý, P., Duclercq, J., Cuesta, C., Müller, B., Grunewald, W., Tarkowski, P. and Benková, E. (2012). Spatiotemporal Regulation of Lateral Root Organogenesis in Arabidopsis by Cytokinin. *The Plant Cell Online*, 24, 3967-3981.

Bisseling, T., Been, C., Klugkist, J., Kammen, A.v. and Nadler, K. (1983) Nodule-specific host proteins in effective and ineffective root nodules of Pisum sativum. *The EMBO journal*, 2, 961-966.

Boualem, A., Laporte, P., Jovanovic, M., Laffont, C., Plet, J., Combier, J.P., Niebel, A., Crespi, M. and Frugier, F. (2008). MicroRNA166 controls root and nodule development in Medicago truncatula. *The Plant journal : for cell and molecular biology*, 54, 876-887.

Breakspear, A., Liu, C., Roy, S., Stacey, N., Rogers, C., Trick, M., Morieri, G., Mysore, K.S., Wen, J., Oldroyd, G.E.D., Downie, J.A. and Murray, J.D. (2014). The Root Hair “Infectome” of Medicago truncatula Uncovers Changes in Cell Cycle Genes

and Reveals a Requirement for Auxin Signaling in Rhizobial Infection. *The Plant Cell*, 26, 4680-4701.

Brechenmacher, L., Kim, M.Y., Benitez, M., Li, M., Joshi, T., Calla, B., Lee, M.P., Libault, M., Vodkin, L.O., Xu, D., Lee, S.H., Clough, S.J. and Stacey, G. (2008). Transcription profiling of soybean nodulation by Bradyrhizobium japonicum. *Molecular plant-microbe interactions : MPMI*, 21, 631-645.

Bright, L.J., Liang, Y., Mitchell, D.M. and Harris, J.M. (2005). The LATD gene of Medicago truncatula is required for both nodule and root development. *Molecular plant-microbe interactions : MPMI*, 18, 521-532.

Brill, W.J. (1980). Biochemical genetics of nitrogen fixation. *Microbiological reviews*, 44, 449-467.

Bustos-Sanmamed, P., Mao, G., Deng, Y., Elouet, M., Khan, G.A., Bazin, J., Turner, M., Subramanian, S., Yu, O., Crespi, M. and Lelandais-Brière, C. (2013). Overexpression of miR160 affects root growth and nitrogen-fixing nodule number in Medicago truncatula. *Functional Plant Biology*, 40, 1208-1220.

Casimiro, I., Marchant, A., Bhalerao, R.P., Beeckman, T., Dhooge, S., Swarup, R., Graham, N., Inze, D., Sandberg, G., Casero, P.J. and Bennett, M. (2001). Auxin transport promotes Arabidopsis lateral root initiation. *Plant Cell*, 13, 843-852.

Catoira, R., Galera, C., de Billy, F., Penmetsa, R.V., Journet, E.P., Maillet, F., Rosenberg, C., Cook, D., Gough, C. and Denarie, J. (2000). Four genes of Medicago truncatula controlling components of a nod factor transduction pathway. *Plant Cell*, 12, 1647-1666.

Charpentier, M., Bredemeier, R., Wanner, G., Takeda, N., Schleiff, E. and Parniske, M. (2008). Lotus japonicus CASTOR and POLLUX are ion channels essential for perinuclear calcium spiking in legume root endosymbiosis. *Plant Cell*, 20, 3467-3479.

Cohn, J., Day, R.B. and Stacey, G. (1998). Legume nodule organogenesis. *Trends in Plant Science*, 3, 105-110.

Combier, J.P., Frugier, F., de Billy, F., Boualem, A., El-Yahyaoui, F., Moreau, S., Vernie, T., Ott, T., Gamas, P., Crespi, M. and Niebel, A. (2006). MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in Medicago truncatula. *Genes & development*, 20, 3084-3088.

Couzigou, J.M., Zhukov, V., Mondy, S., Abu el Heba, G., Cosson, V., Ellis, T.H., Ambrose, M., Wen, J., Tadege, M., Tikhonovich, I., Mysore, K.S., Putterill, J., Hofer, J., Borisov, A.Y. and Ratet, P. (2012). NODULE ROOT and COCHLEATA maintain nodule development and are legume orthologs of Arabidopsis BLADE-ON-PETIOLE genes. *Plant Cell*, 24, 4498-4510.

Cvitanich, C., Pallisgaard, N., Nielsen, K.A., Hansen, A.C., Larsen, K., Pihakaski-Maunsbach, K., Marcker, K.A. and Jensen, E.Ø. (2000). CPP1, a DNA-binding protein involved in the expression of a soybean leghemoglobin c3 gene. *Proceedings of the National Academy of Sciences*, 97, 8163-8168.

D'Haeseleer, K., Den Herder, G., Laffont, C., Plet, J., Mortier, V., Lelandais-Briere, C., De Bodt, S., De Keyser, A., Crespi, M., Holsters, M., Frugier, F. and Goormachtig, S. (2011). Transcriptional and post-transcriptional regulation of a

NAC1 transcription factor in Medicago truncatula roots. *The New phytologist*, 191, 647-661.

De Luis, A., Markmann, K., Cognat, V., Holt, D.B., Charpentier, M., Parniske, M., Stougaard, J. and Voinnet, O. (2012). Two microRNAs linked to nodule infection and nitrogen-fixing ability in the legume Lotus japonicus. *Plant Physiol*, 160, 2137-2154.

DeRisi, J.L., Iyer, V.R. and Brown, P.O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science (New York, N.Y.)*, 278, 680-686.

Desbrosses, Guilhem J. and Stougaard, J. (2011). Root Nodulation: A Paradigm for How Plant-Microbe Symbiosis Influences Host Developmental Pathways. *Cell Host & Microbe*, 10, 348-358.

Dixon, R. and Kahn, D. (2004). Genetic regulation of biological nitrogen fixation. *Nat Rev Micro*, 2, 621-631.

Downie, J.A. (2014). Legume nodulation. *Current Biology*, 24, R184-R190.

Dudley, M.E., Jacobs, T.W. and Long, S.R. (1987). Microscopic studies of cell divisions induced in alfalfa roots by Rhizobium meliloti. *Planta*, 171, 289-301.

Elmerich, C., Zimmer, W. and Vieille, C. (1992). Associative nitrogen-fixing bacteria. In *Biological nitrogen fixation*: Chapman and Hall New York, London, pp. 213-258.

Fang, Y. and Hirsch, A.M. (1998). Studying Early Nodulin Gene ENOD40 Expression and Induction by Nodulation Factor and Cytokinin in Transgenic Alfalfa. *Plant Physiology*, 116, 53-68.

Ferguson, B.J., Indrasumunar, A., Hayashi, S., Lin, M.H., Lin, Y.H., Reid, D.E. and Gresshoff, P.M. (2010). Molecular analysis of legume nodule development and autoregulation. *Journal of integrative plant biology*, 52, 61-76.

Ferguson, B.J., Ross, J.J. and Reid, J.B. (2005). Nodulation phenotypes of gibberellin and brassinosteroid mutants of pea. *Plant Physiol*, 138, 2396-2405.

Franssen, H.J., Nap, J.-P., Gloudemans, T., Stiekema, W., Van Dam, H., Govers, F., Louwerse, J., Van Kammen, A. and Bisseling, T. (1987). Characterization of cDNA for nodulin-75 of soybean: A gene product involved in early stages of root nodule development. *Proceedings of the National Academy of Sciences*, 84, 4495-4499.

Franssen, H.J., Xiao, T.T., Kulikova, O., Wan, X., Bisseling, T., Scheres, B. and Heidstra, R. (2015). Root developmental programs shape the Medicago truncatula nodule meristem. *Development*, 142, 2941-2950.

Gage, D.J. (2004) Infection and Invasion of Roots by Symbiotic, Nitrogen-Fixing Rhizobia during Nodulation of Temperate Legumes. *Microbiology and Molecular Biology Reviews*, 68, 280-300.

Gloudemans, T., de Vries, S., Bussink, H.J., Malik, N.S., Franssen, H.J., Louwerse, J. and Bisseling, T. (1987). Nodulin gene expression during soybean (Glycine max) nodule development. *Plant molecular biology*, 8, 395-403.

Gonzalez-Rizzo, S., Crespi, M. and Frugier, F. (2006a). The Medicago truncatula CRE1 Cytokinin Receptor Regulates Lateral Root Development and Early Symbiotic Interaction with Sinorhizobium meliloti. *The Plant Cell Online*, 18, 2680-2693.

Gonzalez-Rizzo, S., Crespi, M. and Frugier, F. (2006b) The Medicago truncatula CRE1 cytokinin receptor regulates lateral root development and early symbiotic interaction with Sinorhizobium meliloti. *Plant Cell*, 18, 2680-2693.

Halbleib, C.M. and Ludden, P.W. (2000). Regulation of Biological Nitrogen Fixation. *The Journal of Nutrition*, 130, 1081-1084.

Hayashi, T., Banba, M., Shimoda, Y., Kouchi, H., Hayashi, M. and Imaizumi-Anraku, H. (2010). A dominant function of CCaMK in intracellular accommodation of bacterial and fungal endosymbionts. *The Plant journal : for cell and molecular biology*, 63, 141-154.

Heard, J. and Dunn, K. (1995). Symbiotic induction of a MADS-box gene during development of alfalfa root nodules. *Proceedings of the National Academy of Sciences*, 92, 5273-5277.

Heckmann, A.B., Sandal, N., Bek, A.S., Madsen, L.H., Jurkiewicz, A., Nielsen, M.W., Tirichine, L. and Stougaard, J. (2011). Cytokinin induction of root nodule primordia in Lotus japonicus is regulated by a mechanism operating in the root cortex. *Molecular plant-microbe interactions : MPMI*, 24, 1385-1395.

Herrbach, V., Rembliere, C., Gough, C. and Bensmihen, S. (2014). Lateral root formation and patterning in Medicago truncatula. *Journal of plant physiology*, 171, 301-310.

Hirsch, A.M., Larue, T.A. and Doyle, J. (1997). Is the legume nodule a modified root or stem or an organ sui generis? *Critical Reviews in Plant Sciences*, 16, 361-392.

Hoffman, B.M., Lukoyanov, D., Yang, Z.-Y., Dean, D.R. and Seefeldt, L.C. (2014). Mechanism of Nitrogen Fixation by Nitrogenase: The Next Stage. *Chemical reviews*, 114, 4041-4062.

Høgslund, N., Radutoiu, S., Krusell, L., Voroshilova, V., Hannah, M.A., Goffard, N., Sanchez, D.H., Lippold, F., Ott, T., Sato, S., Tabata, S., Liboriussen, P., Lohmann, G.V., Schauser, L., Weiller, G.F., Udvardi, M.K. and Stougaard, J. (2009). Dissection of Symbiosis and Organ Development by Integrated Transcriptome Analysis of Lotus japonicus Mutant and Wild-Type Plants. *PLoS ONE*, 4, e6556.

Horvath, B., Heidstra, R., Lados, M., Moerman, M., Spaink, H.P., Prome, J.C., van Kammen, A. and Bisseling, T. (1993). Lipo-oligosaccharides of Rhizobium induce infection-related early nodulin gene expression in pea root hairs. *The Plant journal : for cell and molecular biology*, 4, 727-733.

Howard, J.B. and Rees, D.C. (1996). Structural Basis of Biological Nitrogen Fixation. *Chemical reviews*, 96, 2965-2982.

Imaizumi-Anraku, H., Takeda, N., Charpentier, M., Perry, J., Miwa, H., Umehara, Y., Kouchi, H., Murakami, Y., Mulder, L., Vickers, K., Pike, J., Downie, J.A., Wang, T., Sato, S., Asamizu, E., Tabata, S., Yoshikawa, M., Murooka, Y., Wu, G.J., Kawaguchi, M., Kawasaki, S., Parniske, M. and Hayashi, M. (2005). Plastid proteins crucial for symbiotic fungal and bacterial entry into plant roots. *Nature*, 433, 527-531.

Indrasumunar, A. and Gresshoff, P.M. (2010). Duplicated Nod-Factor receptor 5 (NFR5) genes are mutated in soybean. *Plant signaling & behavior*, 5, 535-536.

Indrasumunar, A., Kereszt, A., Searle, I., Miyagi, M., Li, D., Nguyen, C.D., Men, A., Carroll, B.J. and Gresshoff, P.M. (2010). Inactivation of duplicated nod factor receptor 5 (NFR5) genes in recessive loss-of-function non-nodulation mutants of allotetraploid soybean (Glycine max L. Merr.). *Plant & cell physiology*, 51, 201-214.

Jacobsen, K., Laursen, N.B., Jensen, E.O., Marcker, A., Poulsen, C. and Marcker, K.A. (1990). HMG I-like proteins from leaf and nodule nuclei interact with different AT motifs in soybean nodulin promoters. *Plant Cell*, 2, 85-94.

Jensen, E.Ø., Marcker, K.A., Schell, J. and Bruijn, F.J.d. (1988). Interaction of a nodule specific, trans-acting factor with distinct DNA elements in the soybean leghaemoglobin Ibc(3) 5' upstream region. *The EMBO journal*, 7, 1265-1271.

Jones-Rhoades, M.W., Bartel, D.P. and Bartel, B. (2006). MicroRNAS and their regulatory roles in plants. *Annual review of plant biology*, 57, 19-53.

Joshi, T., Yan, Z., Libault, M., Jeong, D.H., Park, S., Green, P.J., Sherrier, D.J., Farmer, A., May, G., Meyers, B.C., Xu, D. and Stacey, G. (2010). Prediction of novel miRNAs and associated target genes in Glycine max. *BMC bioinformatics*, 11 Suppl 1, S14.

Journet, E.P., Pichon, M., Dedieu, A., de Billy, F., Truchet, G. and Barker, D.G. (1994). Rhizobium meliloti Nod factors elicit cell-specific transcription of the ENOD12

gene in transgenic alfalfa. *The Plant journal : for cell and molecular biology*, 6, 241-249.

Kadonaga, J.T. (2004). Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell*, 116, 247-257.

Kanamori, N., Madsen, L.H., Radutoiu, S., Frantescu, M., Quistgaard, E.M., Miwa, H., Downie, J.A., James, E.K., Felle, H.H., Haaning, L.L., Jensen, T.H., Sato, S., Nakamura, Y., Tabata, S., Sandal, N. and Stougaard, J. (2006). A nucleoporin is required for induction of Ca2+ spiking in legume nodule development and essential for rhizobial and fungal symbiosis. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 359-364.

Kouchi, H. and Hata, S. (1993). Isolation and characterization of novel nodulin cDNAs representing genes expressed at early stages of soybean nodule development. *Molecular & general genetics : MGG*, 238, 106-119.

Kouchi, H., Shimomura, K., Hata, S., Hirota, A., Wu, G.J., Kumagai, H., Tajima, S., Suganuma, N., Suzuki, A., Aoki, T., Hayashi, M., Yokoyama, T., Ohyama, T., Asamizu, E., Kuwata, C., Shibata, D. and Tabata, S. (2004). Large-scale analysis of gene expression profiles during early stages of root nodule formation in a model legume, Lotus japonicus. *DNA research : an international journal for rapid publication of reports on genes and genomes*, 11, 263-274.

Lang-Unnasch, N. and Ausubel, F.M. (1985). Nodule-Specific Polypeptides from Effective Alfalfa Root Nodules and from Ineffective Nodules Lacking Nitrogenase. *Plant Physiology*, 77, 833-839.

Laplaze, L., Benkova, E., Casimiro, I., Maes, L., Vanneste, S., Swarup, R., Weijers, D., Calvo, V., Parizot, B., Herrera-Rodriguez, M.B., Offringa, R., Graham, N., Doumas, P., Friml, J., Bogusz, D., Beeckman, T. and Bennett, M. (2007). Cytokinins Act Directly on Lateral Root Founder Cells to Inhibit Root Initiation. *The Plant Cell Online*, 19, 3889-3900.

Lee, H., Hur, C.G., Oh, C.J., Kim, H.B., Pakr, S.Y. and An, C.S. (2004). Analysis of the root nodule-enhanced transcriptome in soybean. *Molecules and cells*, 18, 53-62.

Lelandais-Briere, C., Naya, L., Sallet, E., Calenge, F., Frugier, F., Hartmann, C., Gouzy, J. and Crespi, M. (2009). Genome-wide Medicago truncatula small RNA analysis revealed novel microRNAs and isoforms differentially regulated in roots and nodules. *Plant Cell*, 21, 2780-2796.

Lewis, D.R., Olex, A.L., Lundy, S.R., Turkett, W.H., Fetrow, J.S. and Muday, G.K. (2013). A Kinetic Analysis of the Auxin Transcriptome Reveals Cell Wall Remodeling Proteins That Modulate Lateral Root Development in Arabidopsis. *The Plant Cell*, 25, 3329-3346.

Li, H., Deng, Y., Wu, T., Subramanian, S. and Yu, O. (2010). Misexpression of miR482, miR1512, and miR1515 increases soybean nodulation. *Plant Physiol*, 153, 1759-1770.

Li, X., Mo, X., Shou, H. and Wu, P. (2006). Cytokinin-mediated cell cycling arrest of pericycle founder cells in lateral root initiation of Arabidopsis. *Plant & cell physiology*, 47, 1112-1123.

Libault, M., Farmer, A., Brechenmacher, L., Drnevich, J., Langley, R.J., Bilgin, D.D., Radwan, O., Neece, D.J., Clough, S.J., May, G.D. and Stacey, G. (2010a) Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to Bradyrhizobium japonicum infection. *Plant Physiol*, 152, 541-552.

Libault, M., Farmer, A., Joshi, T., Takahashi, K., Langley, R.J., Franklin, L.D., He, J., Xu, D., May, G. and Stacey, G. (2010b). An integrated transcriptome atlas of the crop model Glycine max, and its use in comparative analyses in plants. *The Plant journal : for cell and molecular biology*, 63, 86-99.

Libbenga, K.R. and Harkes, P.A. (1973). Initial proliferation of cortical cells in the formation of root nodules in Pisum sativum L. *Planta*, 114, 17-28.

Limpens, E., Franken, C., Smit, P., Willemse, J., Bisseling, T. and Geurts, R. (2003). LysM domain receptor kinases regulating rhizobial Nod factor-induced infection. *Science (New York, N.Y.)*, 302, 630-633.

Limpens, E., Moling, S., Hooiveld, G., Pereira, P.A., Bisseling, T., Becker, J.D. and Kuster, H. (2013). cell- and tissue-specific transcriptome analyses of Medicago truncatula root nodules. *PLoS One*, 8, e64377.

Lohar, D.P., Sharopova, N., Endre, G., Penuela, S., Samac, D., Town, C., Silverstein, K.A. and VandenBosch, K.A. (2006). Transcript analysis of early nodulation events in Medicago truncatula. *Plant Physiol*, 140, 221-234.

Madsen, L.H., Tirichine, L., Jurkiewicz, A., Sullivan, J.T., Heckmann, A.B., Bek, A.S., Ronson, C.W., James, E.K. and Stougaard, J. (2010). The molecular network

governing nodule organogenesis and infection in the model legume Lotus japonicus. *Nat Commun*, 1, 10.

Malamy, J.E. and Benfey, P.N. (1997). Organization and cell differentiation in lateral roots of Arabidopsis thaliana. *Development*, 124, 33-44.

Marsh, J.F., Rakocevic, A., Mitra, R.M., Brocard, L., Sun, J., Eschstruth, A., Long, S.R., Schultze, M., Ratet, P. and Oldroyd, G.E.D. (2007). Medicago truncatula NIN Is Essential for Rhizobial-Independent Nodule Organogenesis Induced by Autoactive Calcium/Calmodulin-Dependent Protein Kinase. *Plant Physiology*, 144, 324-335.

Mata, J., Marguerat, S. and Bähler, J. (2005). Post-transcriptional control of gene expression: a genome-wide perspective. *Trends in Biochemical Sciences*, 30, 506-514.

Mathesius, U., Charon, C., Rolfe, B.G., Kondorosi, A. and Crespi, M. (2000a). Temporal and spatial order of events during the induction of cortical cell divisions in white clover by Rhizobium leguminosarum bv. trifolii inoculation or localized cytokinin addition. *Molecular plant-microbe interactions : MPMI*, 13, 617-628.

Mathesius, U., Weinman, J.J., Rolfe, B.G. and Djordjevic, M.A. (2000b). Rhizobia can induce nodules in white clover by "hijacking" mature cortical cells activated during lateral root development. *Molecular plant-microbe interactions : MPMI*, 13, 170-182.

Meng, Y., Shao, C., Wang, H. and Chen, M. (2011). The Regulatory Activities of Plant MicroRNAs: A More Dynamic Perspective. *Plant Physiology*, 157, 1583-1595.

Middleton, P.H., Jakab, J., Penmetsa, R.V., Starker, C.G., Doll, J., Kaló, P., Prabhu, R., Marsh, J.F., Mitra, R.M., Kereszt, A., Dudas, B., VandenBosch, K., Long, S.R., Cook, D.R., Kiss, G.B. and Oldroyd, G.E.D. (2007). An ERF Transcription Factor in Medicago truncatula That Is Essential for Nod Factor Signal Transduction. *The Plant Cell*, 19, 1221-1234.

Mitra, R.M., Gleason, C.A., Edwards, A., Hadfield, J., Downie, J.A., Oldroyd, G.E.D. and Long, S.R. (2004). A Ca(2+)/calmodulin-dependent protein kinase required for symbiotic nodule development: Gene identification by transcript-based cloning. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 4701-4705.

Moling, S., Pietraszewska-Bogiel, A., Postma, M., Fedorova, E., Hink, M.A., Limpens, E., Gadella, T.W. and Bisseling, T. (2014). Nod factor receptors form heteromeric complexes and are essential for intracellular infection in medicago nodules. *Plant Cell*, 26, 4188-4199.

Moreau, S., Verdenaud, M., Ott, T., Letort, S., de Billy, F., Niebel, A., Gouzy, J., de Carvalho-Niebel, F. and Gamas, P. (2011). Transcription Reprogramming during Root Nodule Development in Medicago truncatula. *PLoS ONE*, 6, e16463.

Murray, J.D., Karas, B.J., Sato, S., Tabata, S., Amyot, L. and Szczyglowski, K. (2007). A cytokinin perception mutant colonized by Rhizobium in the absence of nodule organogenesis. *Science (New York, N.Y.)*, 315, 101-104.

Murray, J.D., Muni, R.R., Torres-Jerez, I., Tang, Y., Allen, S., Andriankaja, M., Li, G., Laxmi, A., Cheng, X., Wen, J., Vaughan, D., Schultze, M., Sun, J., Charpentier,

M., Oldroyd, G., Tadege, M., Ratet, P., Mysore, K.S., Chen, R. and Udvardi, M.K. (2011). Vapyrin, a gene essential for intracellular progression of arbuscular mycorrhizal symbiosis, is also essential for infection by rhizobia in the nodule symbiosis of Medicago truncatula. *The Plant journal : for cell and molecular biology*, 65, 244-252.

Mylona, P., Pawlowski, K. and Bisseling, T. (1995). Symbiotic Nitrogen Fixation. *The Plant Cell*, 7, 869-885.

Nap, J.P. and Bisseling, T. (1990). Developmental biology of a plant-prokaryote symbiosis: the legume root nodule. *Science (New York, N.Y.)*, 250, 948-954.

Norris, J.H., Macol, L.A. and Hirsch, A.M. (1988). Nodulin gene expression in effective alfalfa nodules and in nodules arrested at three different stages of development. *Plant Physiol*, 88, 321-328.

Nova-Franco, B., Iniguez, L.P., Valdes-Lopez, O. and Alvarado-Affantranger, X. (2015). The micro-RNA72c-APETALA2-1 node as a key regulator of the common bean-Rhizobium etli nitrogen fixation symbiosis. 168, 273-291.

Nutman, P. (1948). Physiological studies on nodule formation: I. The relation between nodulation and lateral root formation in red clover. *Annals of Botany*, 12, 81-96.

Okushima, Y., Overvoorde, P.J., Arima, K., Alonso, J.M., Chan, A., Chang, C., Ecker, J.R., Hughes, B., Lui, A., Nguyen, D., Onodera, C., Quach, H., Smith, A., Yu, G. and Theologis, A. (2005). Functional genomic analysis of the AUXIN RESPONSE FACTOR gene family members in Arabidopsis thaliana: unique and overlapping functions of ARF7 and ARF19. *Plant Cell*, 17, 444-463.

Oldroyd, G.E. and Long, S.R. (2003). Identification and characterization of nodulation-signaling pathway 2, a gene of Medicago truncatula involved in Nod actor signaling. *Plant Physiol*, 131, 1027-1032.

Oldroyd, G.E.D. (2001). Dissecting Symbiosis: Developments in Nod Factor Signal Transduction. *Annals of Botany*, 87, 709-718.

Osipova, M.A., Mortier, V., Demchenko, K.N., Tsyganov, V.E., Tikhonovich, I.A., Lutova, L.A., Dolgikh, E.A. and Goormachtig, S. (2012). Wuschel-related homeobox5 gene expression and interaction of CLE peptides with components of the systemic control add two pieces to the puzzle of autoregulation of nodulation. *Plant Physiology*, 158, 1329-1341.

Paponov, I.A., Paponov, M., Teale, W., Menges, M., Chakrabortee, S., Murray, J.A. and Palme, K. (2008). Comprehensive transcriptome analysis of auxin responses in Arabidopsis. *Molecular plant*, 1, 321-337.

Pawlowski, K. and Demchenko, K.N. (2012). The diversity of actinorhizal symbiosis. *Protoplasma*, 249, 967-979.

Peiter, E., Sun, J., Heckmann, A.B., Venkateshwaran, M., Riely, B.K., Otegui, M.S., Edwards, A., Freshour, G., Hahn, M.G., Cook, D.R., Sanders, D., Oldroyd, G.E.D., Downie, J.A. and Ané, J.-M. (2007). The Medicago truncatula DMI1 Protein Modulates Cytosolic Calcium Signaling. *Plant Physiology*, 145, 192-203.

Perrine-Walker, F., Gherbi, H., Imanishi, L., Hocher, V., Ghodhbane-Gtari, F., Lavenus, J., Benabdoun, F.M., Nambiar-Veeti, M., Svistoonoff, S. and Laplaze, L. (2011).

Symbiotic signaling in actinorhizal symbioses. *Current protein & peptide science*, 12, 156-164.

Phillips, T. (2008). Regulation of transcription and gene expression in eukaryotes. *Nature Education*, 1, 199.

Plet, J., Wasson, A., Ariel, F., Le Signor, C., Baker, D., Mathesius, U., Crespi, M. and Frugier, F. (2011). MtCRE1-dependent cytokinin signaling integrates bacterial and plant cues to coordinate symbiotic nodule organogenesis in Medicago truncatula. *The Plant Journal*, 65, 622-633.

Radutoiu, S., Madsen, L.H., Madsen, E.B., Felle, H.H., Umehara, Y., Gronlund, M., Sato, S., Nakamura, Y., Tabata, S., Sandal, N. and Stougaard, J. (2003). Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature*, 425, 585-592.

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

Roux, B., Rodde, N., Jardinaud, M.F., Timmers, T., Sauviac, L., Cottret, L., Carrere, S., Sallet, E., Courcelle, E., Moreau, S., Debelle, F., Capela, D., de Carvalho-Niebel, F., Gouzy, J., Bruand, C. and Gamas, P. (2014). An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *The Plant journal : for cell and molecular biology*, 77, 817-837.

Saito, K., Yoshikawa, M., Yano, K., Miwa, H., Uchida, H., Asamizu, E., Sato, S., Tabata, S., Imaizumi-Anraku, H., Umehara, Y., Kouchi, H., Murooka, Y., Szczyglowski,

K., Downie, J.A., Parniske, M., Hayashi, M. and Kawaguchi, M. (2007). NUCLEOPORIN85 is required for calcium spiking, fungal and bacterial symbioses, and seed production in Lotus japonicus. *Plant Cell*, 19, 610-624.

Santi, C., Bogusz, D. and Franche, C. (2013). Biological nitrogen fixation in non-legume plants. *Annals of Botany*.

Schauser, L., Roussis, A., Stiller, J. and Stougaard, J. (1999). A plant regulator controlling development of symbiotic root nodules. *Nature*, 402, 191-195.

Scheres, B., Van De Wiel, C., Zalensky, A., Horvath, B., Spaink, H., Van Eck, H., Zwartkruis, F., Wolters, A.-M., Gloudemans, T., Van Kammen, A. and Bisseling, T. (1990a) .The ENOD12 gene product is involved in the infection process during the pea-rhizobium interaction. *Cell*, 60, 281-294.

Scheres, B., van Engelen, F., van der Knaap, E., van de Wiel, C., van Kammen, A. and Bisseling, T. (1990b). Sequential induction of nodulin gene expression in the developing pea nodule. *The Plant Cell*, 2, 687-700.

Smit, P., Raedts, J., Portyanko, V., Debelle, F., Gough, C., Bisseling, T. and Geurts, R. (2005). NSP1 of the GRAS protein family is essential for rhizobial Nod factor-induced transcription. *Science (New York, N.Y.)*, 308, 1789-1791.

Socolow, R.H. (1999). Nitrogen management and the future of food: Lessons from the management of energy and carbon. *Proceedings of the National Academy of Sciences*, 96, 6001-6008.

Soyano, T., Kouchi, H., Hirota, A. and Hayashi, M. (2013). NODULE INCEPTION Directly Targets NF-Y Subunit Genes to Regulate Essential Processes of Root Nodule Development in Lotus japonicus. *PLoS Genetics*, 9, e1003352.

Stougaard, J. (2000). Regulators and Regulation of Legume Root Nodule Development. *Plant Physiology*, 124, 531-540.

Subramanian, S., Fu, Y., Sunkar, R., Barbazuk, W.B., Zhu, J.K. and Yu, O. (2008). Novel and nodulation-regulated microRNAs in soybean roots. *BMC genomics*, 9, 160.

Suzaki, T., Yano, K., Ito, M., Umehara, Y., Suganuma, N. and Kawaguchi, M. (2012). Positive and negative regulation of cortical cell division during root nodule development in Lotus japonicus is accompanied by auxin response. *Development (Cambridge, England)*, 139, 3997-4006.

Swarup, K., Benkova, E., Swarup, R., Casimiro, I., Peret, B., Yang, Y., Parry, G., Nielsen, E., De Smet, I., Vanneste, S., Levesque, M.P., Carrier, D., James, N., Calvo, V., Ljung, K., Kramer, E., Roberts, R., Graham, N., Marillonnet, S., Patel, K., Jones, J.D.G., Taylor, C.G., Schachtman, D.P., May, S., Sandberg, G., Benfey, P., Friml, J., Kerr, I., Beeckman, T., Laplaze, L. and Bennett, M.J. (2008). The auxin influx carrier LAX3 promotes lateral root emergence. *Nat Cell Biol*, 10, 946-954.

Szabados, L., Ratet, P., Grunenberg, B. and de Bruijn, F.J. (1990). Functional analysis of the Sesbania rostrata leghemoglobin glb3 gene 5'-upstream region in transgenic Lotus corniculatus and Nicotiana tabacum plants. *Plant Cell*, 2, 973-986.

Szczyglowski, K., Potter, T., Stoltzfus, J., Fujimoto, S.Y. and de Bruijn, F.J. (1996). Differential expression of the Sesbania rostrata leghemoglobin glb3 gene promoter in transgenic legume and non-legume plants. *Plant molecular biology*, 31, 931-935.

Szczyglowski, K., Szabados, L., Fujimoto, S.Y., Silver, D. and de Bruijn, F.J. (1994). Site-specific mutagenesis of the nodule-infected cell expression (NICE) element and the AT-rich element ATRE-BS2* of the Sesbania rostrata leghemoglobin glb3 promoter. *The Plant Cell*, 6, 317-332.

Timmers, A.C., Auriac, M.C. and Truchet, G. (1999). Refined analysis of early symbiotic steps of the Rhizobium-Medicago interaction in relationship with microtubular cytoskeleton rearrangements. *Development*, 126, 3617-3628.

Tirichine, L., Sandal, N., Madsen, L.H., Radutoiu, S., Albrektsen, A.S., Sato, S., Asamizu, E., Tabata, S. and Stougaard, J. (2007). A gain-of-function mutation in a cytokinin receptor triggers spontaneous root nodule organogenesis. *Science (New York, N.Y.)*, 315, 104-107.

Trinick, M.J. (1973). Symbiosis between Rhizobium and the Non-legume, Trema aspera. *Nature*, 244, 459-460.

Turner, M., Nizampatnam, N.R., Baron, M., Coppin, S., Damodaran, S., Adhikari, S., Arunachalam, S.P., Yu, O. and Subramanian, S. (2013). Ectopic Expression of miR160 Results in Auxin Hypersensitivity, Cytokinin Hyposensitivity, and Inhibition of Symbiotic Nodule Development in Soybean. *Plant Physiology*, 162, 2042-2055.

van de Wiel, C., Scheres, B., Franssen, H., van Lierop, M.-J., van Lammeren, A., Van Kammen, A. and Bisseling, T. (1990). The early nodulin transcript ENOD2 is located in the nodule parenchyma (inner cortex) of pea and soybean root nodules. *The EMBO journal*, 9, 1.

Vance, C., Boylan, K., Stade, S. and Somers, D. (1985). Nodule specific proteins in alfalfa (Medicago sativa L.). *Symbiosis (USA).*

Vance, C.P. (2001). Symbiotic Nitrogen Fixation and Phosphorus Acquisition. Plant Nutrition in a World of Declining Renewable Resources. *Plant Physiology*, 127, 390-397.

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

Venkateshwaran, M., Jayaraman, D., Chabaud, M., Genre, A., Balloon, A.J., Maeda, J., Forshey, K., den Os, D., Kwiecien, N.W., Coon, J.J., Barker, D.G. and Ane, J.M. (2015). A role for the mevalonate pathway in early plant symbiotic signaling. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 9781-9786.

Vernie, T., Kim, J., Frances, L., Ding, Y., Sun, J., Guan, D. and Niebel, A. (2015). The NIN Transcription Factor Coordinates Diverse Nodulation Programs in Different Tissues of the Medicago truncatula Root. 27, 3410-3424.

Wagner, S.C. (2011). Biological Nitrogen Fixation. *Nature*, 3, 15.

Wang, D., Yang, S., Tang, F. and Zhu, H. (2012) Symbiosis specificity in the legume – rhizobial mutualism. *Cellular Microbiology*, 14, 334-342.

Wang, Y., Li, K., Chen, L., Zou, Y. and Liu, H. (2015). MicroRNA167-Directed Regulation of the Auxin Response Factors GmARF8a and GmARF8b Is Required for Soybean Nodulation and Lateral Root Development. 168, 984-999.

Wang, Y., Li, P., Cao, X., Wang, X., Zhang, A. and Li, X. (2009). Identification and expression analysis of miRNAs from nitrogen-fixing soybean nodules. *Biochemical and biophysical research communications*, 378, 799-803.

Wang, Y., Wang, L., Zou, Y., Chen, L., Cai, Z., Zhang, S., Zhao, F., Tian, Y., Jiang, Q., Ferguson, B.J., Gresshoff, P.M. and Li, X. (2014). Soybean miR172c targets the repressive AP2 transcription factor NNC1 to activate ENOD40 expression and regulate nodule initiation. *Plant Cell*, 26, 4782-4801.

Yan, Z., Hossain, M.S., Arikit, S., Valdes-Lopez, O., Zhai, J., Wang, J., Libault, M., Ji, T., Qiu, L., Meyers, B.C. and Stacey, G. (2015a). Identification of microRNAs and their mRNA targets during soybean nodule development: functional analysis of the role of miR393j-3p in soybean nodulation. *The New phytologist*.

Yan, Z., Hossain, M.S., Arikit, S., Valdes-Lopez, O., Zhai, J., Wang, J., Libault, M., Ji, T., Qiu, L., Meyers, B.C. and Stacey, G. (2015b) Identification of microRNAs and their mRNA targets during soybean nodule development: functional analysis of the role of miR393j-3p in soybean nodulation. *The New phytologist*, 207, 748-759.

Yan, Z., Hossain, M.S., Valdes-Lopez, O., Hoang, N.T., Zhai, J., Wang, J., Libault, M., Brechenmacher, L., Findley, S., Joshi, T., Qiu, L., Sherrier, D.J., Ji, T., Meyers, B.C., Xu, D. and Stacey, G. (2016). Identification and functional characterization of soybean root hair microRNAs expressed in response to Bradyrhizobium japonicum infection. *Plant biotechnology journal*, 14, 332-341.

Yan, Z., Hossain, M.S., Wang, J., Valdes-Lopez, O., Liang, Y., Libault, M., Qiu, L. and Stacey, G. (2013). miR172 regulates soybean nodulation. *Molecular plant-microbe interactions : MPMI*, 26, 1371-1377.

Yang, W.C., de Blank, C., Meskiene, I., Hirt, H., Bakker, J., van Kammen, A., Franssen, H. and Bisseling, T. (1994). Rhizobium nod factors reactivate the cell cycle during infection and nodule primordium formation, but the cycle is only completed in primordium formation. *Plant Cell*, 6, 1415-1426.

Yang, W.C., Katinakis, P., Hendriks, P., Smolders, A., de Vries, F., Spee, J., van Kammen, A., Bisseling, T. and Franssen, H. (1993). Characterization of GmENOD40, a gene showing novel patterns of cell-specific expression during soybean nodule development. *The Plant journal : for cell and molecular biology*, 3, 573-585.

Yano, K., Shibata, S., Chen, W.L., Sato, S., Kaneko, T., Jurkiewicz, A., Sandal, N., Banba, M., Imaizumi-Anraku, H., Kojima, T., Ohtomo, R., Szczyglowski, K., Stougaard, J., Tabata, S., Hayashi, M., Kouchi, H. and Umehara, Y. (2009). CERBERUS, a novel U-box protein containing WD-40 repeats, is required for formation of the infection thread and nodule development in the legume-

Rhizobium symbiosis. *The Plant journal : for cell and molecular biology*, 60, 168-180.

Yano, K., Yoshida, S., Muller, J., Singh, S., Banba, M., Vickers, K., Markmann, K., White, C., Schuller, B., Sato, S., Asamizu, E., Tabata, S., Murooka, Y., Perry, J., Wang, T.L., Kawaguchi, M., Imaizumi-Anraku, H., Hayashi, M. and Parniske, M. (2008). CYCLOPS, a mediator of symbiotic intracellular accommodation. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 20540-20545.

Yokota, K., Fukai, E., Madsen, L.H., Jurkiewicz, A., Rueda, P., Radutoiu, S., Held, M., Hossain, M.S., Szczyglowski, K., Morieri, G., Oldroyd, G.E., Downie, J.A., Nielsen, M.W., Rusek, A.M., Sato, S., Tabata, S., James, E.K., Oyaizu, H., Sandal, N. and Stougaard, J. (2009). Rearrangement of actin cytoskeleton mediates invasion of Lotus japonicus roots by Mesorhizobium loti. *Plant Cell*, 21, 267-284.

Zhao, Y., Yu, Y., Zhai, J., Ramachandran, V., Dinh, T.T., Meyers, B.C., Mo, B. and Chen, X. (2012). The Arabidopsis nucleotidyl transferase HESO1 uridylates unmethylated small RNAs to trigger their degradation. *Curr Biol*, 22, 689-694.

Chapter 2

2. Comparative transcriptomics reveals distinct transcription factors and hormone action pathways in nodules and lateral roots of soybean

2.1. Introduction

Root lateral organs such as lateral roots (LRs) and symbiotic nodules are unique in that they arise from differentiated cells whereas shoot lateral organs arise from founder cells in the meristem. Throughout plant development, the shoot apical meristem (SAM) produces leaf primordia and axillary meristems that develop into a lateral branch or flower meristem (Barton 2010). On the other hand, root lateral organs are not formed directly from the root apical meristem (RAM), but through *de novo* differentiation (De Smet 2012, Sozzani and Iyer-Pascuzzi 2014). LRs are present in all higher plants and are initiated in response to both developmental and primarily abiotic environmental cues (e.g. nutrients, mechanical stimuli). Along with LR, roots of some higher plants are capable of forming nodules in association with nitrogen-fixing bacteria termed rhizobia and *Frankia*. The majority of legumes and the non-legume *Parasponia* initiate symbiotic nodule development in response to rhizobia-derived lipo-chitooligosaccharide nodulation (*nod*) signals specifically under nitrogen-deprived conditions (Trinick 1973, Marvel et al. 1987, Hirsch 1992). *Frankia* interact with eight different families of dicotyledon plants commonly known as actinorhizal plants [reviewed by (Benson and Silvester 1993, Wall 2000, Pawlowski and Demchenko 2012)].

Many evidences suggest that nodule and LRs are related. First, the correlation between the number of LRs and nodules has been observed in many legume species (Nutman 1948, Carroll et al. 1985, Nishimura et al. 2002). In addition, actinorhizal and

Parasponia nodules share several similarities to roots including a central vasculature (Pawlowski and Bisseling 1996, Gualtieri and Bisseling 2000, Pawlowski and Demchenko 2012). Nod factors also affect root architecture in addition to nodule formation. These observations indicated that nodules might share developmental pathways with LRs and suggested a root origin in nodule formation. Indeed, root identity of nodules appears to be suppressed by orthologs of the *Arabidopsis thaliana BLADE-ON-PETIOLE* orthologs in *Medicago truncatula* and *Pisum sativum*. Loss of function mutations in these genes led to the development of roots from nodule vascular initials (Couzigou et al. 2012). However, a legume nodule differs from the root in both development and morphology (Hirsch et al. 1997, Mathesius 2003). LRs arise from a few initial pericycle cells adjacent to a xylem pole and undergo a defined program of cell division and expansion. For example, in *A. thaliana*, the formation of LRs occurs by the division of the pericycle cells and can be divided into 8 stages based on anatomical characteristics and cell divisions (Malamy and Benfey 1997). In *M. truncatula*, LR formation involves cell division in pericycle as well as endodermis (Herrbach et al. 2014). Root nodule initiation also involves a defined program of cell division and expansion, but the site of initiation of legume nodules is cortex cells (with a few exceptions (Allen and Allen 1940, Bond 1948)). Furthermore, after initiation, the two lateral organs have clear differences in their development. A conspicuous example is the presence of central vasculatures in LRs as opposed to peripheral vasculatures in nodules. Dissection of genetic pathways associated with the initiation of the nodule and lateral root development indicates a conservation between the primary root and LRs but not in

nodules which show some cellular structures similarities with the shoot (Desbrosses and Stougaard 2011).

There are two major types of nodules formed in legume roots: indeterminate and determinate [reviewed in (Hirsch 1992, Mathesius 2003)]. Indeterminate nodules are oblong and characterized by the presence of a persistent nodule meristem analogous to LRs. Examples of plants that form indeterminate nodules include temperate legumes viz. *P. sativum* (pea), *M. truncatula* (Barrel Medic) and *Trifolium* species (clover). In contrast, determinate nodules are spherical and lack a persistent nodule meristem. There is no sustained cell division during determinate nodule development; nodule growth is more a result of cell expansion rather than cell division. Tropical/subtropical legumes viz. *G. max* (soybean), *Vicia faba* (common bean), and *L. japonicus* represent determinate nodules. Additionally, indeterminate nodules arise from inner cortical cell layers whereas determinate nodules arise from outer cortical cell layers.

Therefore nodules, especially determinate nodules differ from LRs on various properties including the site of origin, type of initial cell divisions, meristem persistency, position relative to parent cortex and vascular position (Hirsch et al. 1997). The hormone requirements for development also appear to be different between these organs. For example, auxin is known to promote LRs (Casimiro et al. 2001, Swarup et al. 2008) whereas there is very low auxin activity during nodule initiation and increased auxin activity inhibits nodule formation, especially in determinate nodule forming legumes (Suzaki et al. 2012, Turner et al. 2013). On the other hand, cytokinin promotes nodule formation (Gonzalez-Rizzo et al. 2006, Murray et al. 2007, Plet et al. 2011), but inhibits LR formation (Gonzalez-Rizzo et al. 2006, Li et al. 2006, Laplaze et al. 2007, Bielach et

al. 2012). It is possible that initiation of LRs and nodules might be dictated by distinct auxin-cytokinin ratios. So far, information on morphological characteristics of root and legume nodule and the expression patterns of organ identity genes (“nodulins”) or regulatory gene [e.g. MADS-box gene (*NMH*)], are not able to definitively answer if the nodule is indeed a modified lateral root. Similarly, not enough evidence is present to suggest if it originated from stem or carbon storage organs (Hirsch et al. 1997). Therefore, whether nodules evolved by adopting developmental signaling pathways of LR formation is an open question.

To address some of the outstanding questions on the similarities and differences in signaling and developmental pathways associated with the development of these lateral organs, we compared transcriptomes of lateral root and nodule tissues at two different stages of development in soybean. While a number of studies have evaluated gene expression during nodule formation (Kouchi et al. 2004, Benedito et al. 2008, Brechenmacher et al. 2008, Libault et al. 2010c, Takanashi et al. 2012, Limpens et al. 2013), global gene expression profiles during LR formation has not been evaluated in legumes. In addition, the LR initiation process in legumes might be slightly different from that in *Arabidopsis* (Herrbach et al. 2014). RNA-seq was used to obtain transcriptome profiles, and used for comparative analysis to identify organ- and developmental stage-specific transcripts enriched in LRs and nodule at two developmental stages of soybean. Comparison of LR and nodule transcriptome suggest that just a few biological processes primarily associated with cell division, and a small set of transcripts are common between these organs and that the development of these organs involves distinct transcription factor (TF) families and hormonal activities. Evaluation of

the expression and enrichment of key marker genes associated with shoot apical and axillary meristems in *Arabidopsis* to LR and nodule transcriptomes suggested that soybean nodules are more similar to a shoot axillary organ than a lateral root. These data strongly suggest that symbiotic nodules might have co-opted developmental and hormonal signaling pathways active in shoot axillary meristems in addition to those in lateral roots during evolution. The data generated in this study will serve as an excellent resource to the community for comparative analysis of molecular and cellular processes active during lateral organ development in plants. Examples include but are not limited to carbon and nitrogen metabolism (Wang et al. 2000, Palenchar et al. 2004, Gutiérrez et al. 2007), specialized secondary metabolites transporter activity, plant defense responses, and specific gene families of significance (Ehlting et al. 2008, Cvrčková et al. 2010, Ghorbani et al. 2015). In addition, the dataset can be used as a key resource for generation of gene regulatory networks using co-expression models (Canales et al. 2014) or graphical Gaussian models (Ma et al. 2007), as well as systems biology studies when combined with proteomic and metabolomic analyses (Lan et al. 2013, Hossain et al. 2015).

2.2. Results

2.2.1. Overview of the transcriptome libraries

To compare the transcriptome of LRs and nodules, RNA was obtained from emerging nodules (EN), mature nodules (MN), emerging lateral roots (ELR) and young lateral roots (YLR), and strand-specific RNA-seq libraries were constructed by using polyA-enriched RNA preparations. Root sections above and below these organs devoid of any lateral organs (designated ABEN, ABMN, ABELR, and ABYLR respectively)

were used to construct respective control tissue libraries (Figure 2.1). “Empty” root segments above and below the respective organs was used as age-appropriate controls to identify organ-specific/enriched genes. At an early stage, both organs (EN and ELR) had minimal differentiation and had not completely emerged out of the primary root. The comparison of transcriptomes at this stage would identify common developmental pathways if any that exist between initiation/formation of LRs and nodules. The mature stages were expected to identify distinct pathways that characterize specific functions of these organs. EN and ABEN from soybean (Williams 82) seedlings were harvested 8 days post inoculation (dpi) with *Bradyrhizobium japonicum*. MN and ABMN were harvested at 14-16 dpi. EN was distinguished as a slight bump in the root surface while mature MN were completely protruded and pink in color (Figure 2.1 a and b). ELR and YLR and their controls were harvested from 3-5 days old soybean seedlings not inoculated with *B. japonicum*. ELR were selected based on a sharp protrusion with no tissue separation on the epidermis, which is also different from the bump seen in the cortical region of an emerging nodule. YLR were completely protruded out and about 1-2 mm in length (Figure 2.1 c and d). Three replicate samples were harvested for each of the eight tissue types for the preparation of twenty-four RNA-seq libraries.

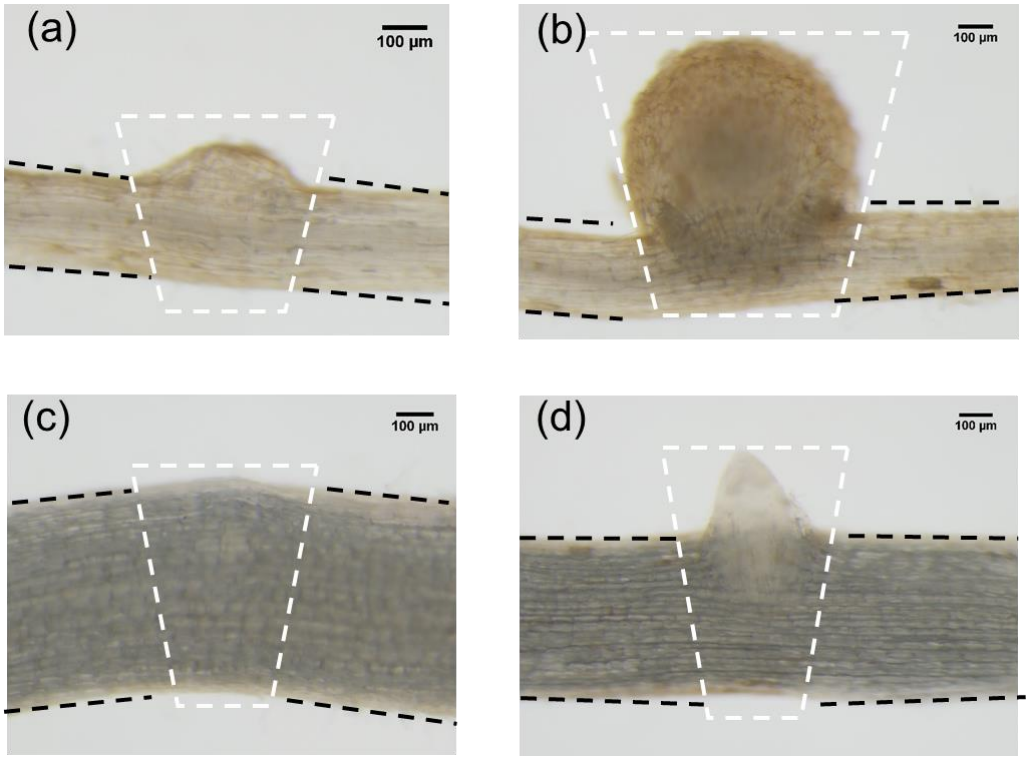


Figure 2.1 Representative images of root lateral organs and corresponding control tissues harvested for transcriptome library construction. The portions shown enclosed in white dotted lines show emerging nodules (EN) (a), mature nodules (MN) (b), emerging lateral roots (ELR) (c), and young lateral roots (YLR) (d). The portions shown in black dotted lines show adjacent root tissues that served as controls for the respective organs.

The libraries were sequenced using an Illumina HiSeq 2000 and RNA-seq analysis was performed to identify organ-specific/enriched genes and pathways. Read mapping, transcript assembly and differential expression analyses were performed using the "Tuxedo" pipeline (Trapnell et al. 2012), (See *experimental procedures* for parameters). Previously reported nodule-specific genes were used as controls to verify the authenticity of the tissue-specific RNA-seq libraries (Table 2.1). For example, *FRUIT WEIGHT 2.2 -LIKE 1* (*FWL1)* (Glyma09g31910) is induced early in the root hair and subsequently highly enriched in nodule tissues (higher FPKM in MN compared to ABMN) in response to rhizobium inoculation (Libault et al. 2010a, Libault et al. 2010c). Consistently, *FWL1* expression was observed in both EN and MN; however, a significant enrichment was not observed in EN and this could likely be due to its expression in adjacent control tissues colonized by rhizobia (Table 2.1). In contrast, high expression and significant enrichment of *FWL*1 was observed in MN.*GLYCINE MAX EARLY NODULIN GENE* (*GmENOD40)* was enriched in both EN and MN as expected. Similarly, all four symbiotic leghemoglobin (Ott et al. 2005) tested were very highly expressed in MN, but barely detected in EN; among these three leghemoglobin were specifically enriched in MN (Table 2.1). *NODULE INCEPTION* (*NIN)* and *NODULATION SIGNALING PATHWAY 1* (*NSP1)* also showed expected patterns of expression (Heckmann et al. 2006, Heckmann et al. 2011). Importantly, the expression of these nodule marker genes was not detected in ELR, YLR and their controls except *ENOD40* which is known to be expressed at low levels in the roots (Papadopoulou et al. 1996). For lateral root markers, potential soybean orthologs of *Arabidopsis* lateral root primordium marker genes (Smith and Fedoroff 1995, Parizot et al. 2010) were obtained

from LegumeIP (http://plantgrn.noble.org/LegumeIP) (Li et al. 2012). At least one ortholog of each family of lateral root primordium marker genes tested such as *AUXIN RESPONSE FACTOR* (*ARF5), CYTOKININ RESPONSE FACTOR* (*CRF2), and LATERAL ROOT PRIMORDIUM 1* (*LRP1)* except *GATA TRANSCRIPTION FACTOR* (*GATA23)* was significantly enriched in ELR and/or LR as expected. Consistent with their role in LR formation, orthologs of *PIN-FORMED1* (*PIN1)* and *TARGET OF MONOPTEROS 7* (*TMO7)* were specifically enriched in ELR and were highly expressed in LR tissues vs. nodule tissues. Importantly, none of the LR primordium markers were enriched in nodule tissues. Finally, a set of housekeeping genes, *Actin11* and *cons4* previously identified to be expressed uniformly in multiple soybean tissues (Libault et al. 2010c) showed no difference in expression between the different organs and their control tissues (Table 2.1). Together these results confirmed that tissue harvests were of very high quality and specificity.

Table 2.1 Expression and enrichment of selected organ-specific genes. Expression levels are FPKM values observed in each lateral organ tissue. Enrichment refers to log_2 fold value calculated by using the expression values in each tissue vs respective control tissue. Numbers in parenthesis indicate enrichment (log_2 fold change in expression level compared to the respective control tissues). Fold change values showing significant enrichment are highlighted in boldface with noise corrected p-value (q-value) <0.05.*

Transcript ID	Annotation	Expression levels (FPKM) and enrichment *			
		EN	MN	ELR	YLR
Nodule Marker genes					
Glyma09g31910.	*FWL1*	152.0 (0.0)	**1267.6 (8.0)**	0.0 (0.0)	0.0 (0.0)
Glyma02g04180.	*Enod40*	**1482.2 (2.9)**	**1799.5 (4.7)**	12.4 (-1.0)	13.2 (-2.3)
Glyma10g34290.	*LBC_A*	0.2 (0.0)	**9191.6 (8.8)**	0.0 (0.0)	0.0 (0.0)
Glyma10g34280.	*LBC_C1*	0.7 (0.0)	**9839.5 (9.0)**	0.0 (0.0)	0.0 (0.0)
Glyma20g33290.	*LBC_C2*	**27.9 (3.2)**	**6968.7 (9.0)**	0.0 (0.0)	0.0 (0.0)
Glyma10g34260.	*LBC_C3*	19.3 (0.0)	**12680.1 (9.0)**	0.0 (0.0)	0.0 (0.0)
Glyma04g00210.	*NIN1*	16.7 (0.0)	84.8 (0.0)	0.0 (0.0)	0.2 (0.0)
Glyma16g01020.	*NSP1*	**9.8 (3.5)**	**19.7 (5.5)**	0.4 (0.0)	0.7 (0.0)
Lateral root marker genes					
Glyma14g40540.	*ARF5*	6.7 (0.0)	0.9 (-1.5)	2.9 (0.0)	2.6 (0.0)
Glyma17g37580.	*ARF5*	12.6 (0.0)	2.5 (0.0)	**6.3 (1.0)**	6.3 (0.0)

Glyma05g37120.	*CRF2*	1.0 (0.0)	2.2 (0.0)	**3.7 (2.0)**	**1.6 (3.7)**
Glyma08g02460.	*CRF2*	4.1 (0.0)	7.2 (0.0)	**9.0 (1.9)**	**6.1 (2.8)**
Glyma03g39220.	*GATA23*	0.0 (0.0)	0.1 (0.0)	0.1 (0.0)	0.0 (0.0)
Glyma19g41780.	*GATA23*	4.5 (0.0)	6.3 (0.0)	6.9 (0.0)	5.0 (0.0)
Glyma02g44860.	*LRP1*	19.5 (0.0)	8.6 (0.0)	11.3 (0.0)	6.1 (0.0)
Glyma07g35780.	*LRP1*	11.3 (0.0)	6.0 (0.0)	5.0 (0.0)	3.3 (0.0)
Glyma14g03900.	*LRP1*	24.8 (0.0)	9.8 (0.0)	**21.4 (0.7)**	9.4 (0.0)
Glyma07g11550.	*PIN1*	7.7 (0.0)	2.2 (-2.6)	**24.7 (1.0)**	22.6 (0.0)
Glyma08g05900.	*PIN1*	8.2 (0.0)	2.2 (0.0)	**19.6 (1.5)**	9.6 (0.0)
Glyma09g30700.	*PIN1*	7.8 (-1.0)	1.7 (-3.2)	15.5 (0.0)	12.6 (0.0)
Glyma04g34080.	*TMO7*	0.0 (0.0)	0.0 (0.0)	**5.2 (2.7)**	**11.5 (5.0)**
Glyma06g20400.	*TMO7*	0.2 (0.0)	0.0 (0.0)	**23.9 (2.3)**	26.7 (0.0)
Housekeeping genes					
Glyma02g10170.	*Actin11*	149.7 (0.0)	102.5 (0.0)	204.5 (0.0)	223.8 (0.0)
Glyma12g02310.	*Cons4*	14.1 (0.0)	20.4 (0.0)	17.9 (0.0)	28.5 (0.0)
Glyma12g05510.	*Cons6*	24.2 (0.0)	28.3 (0.0)	36.6 (0.0)	19.4 (0.0)

2.2.2. Predominant mapping to coding sequences and consistency among replicates in our libraries

To improve the read quality, trimming (PHRED quality score >=20) and filtering (minimum read length = 25-nt) of the adapter trimmed reads were conducted. This resulted in 28 million reads per library on average, which is expected to provide sufficient depth of coverage to reliably quantify gene expression using RNA-seq (Liu et al. 2014). The majority of the libraries retained ~85% reads post quality trimming and filtering; this indicated that these were high-quality libraries. It should be noted that YLR and ABYLR libraries had to be sequenced twice due to technical reasons and they retained only 65% reads, but 33 million reads on average (Table 2.2). Reads were aligned against the soybean reference genome (Gmax_v1.1_1.89; 54,175 gene models; 73,269 transcripts) allowing no mismatches or indels. On average, 84% of the high quality reads successfully mapped to the genome. Among these, ~18% were junction reads (that spanned two different exons; Table 2.2). Detailed examination of alignment positions in the genome demonstrated that most of the bases (76%) mapped to coding sequences followed by the UTRs (18%). A smaller percentage of reads mapped to intronic (4%) and intergenic regions (2%) (Figure 2.2 a). Next, the distribution of FPKM (Fragments Per Kilobase of transcripts per Million mapped reads) values in each library (obtained from *isoforms.read-group-tracking* output of *cuffdiff*) was compared and identified similar median values and FPKM distribution among the libraries (Figure 2.3) indicating that they can be reliably used to compare gene expression. Finally, hierarchical clustering was performed based on correlation distances and the result indicated excellent consistency among the three different replicates of the same tissue types (Figure 2.2 b). Together,

these data indicated that nodule and LR RNA-seq libraries were of very high quality and very well suited for global gene expression analysis.

Table 2.2 Average number of reads. Read quality and mapping summary of the 24 RNA-seq libraries analysed in this study is provided.

Samples	**Initial reads**	**Good reads**	**Mapped reads**	**%**	**Junctions**
EN	27,316,270	23,085,065	19,677,401	85.3	3,486,503
ABEN	31,217,551	26,366,781	22,393,649	84.9	2,418,097
MN	32,160,389	26,628,397	22,783,171	85.6	4,303,688
ABMN	31,981,548	26,449,908	22,386,232	84.7	3,676,281
ELR	33,457,163	28,959,046	24,769,480	85.8	4,915,012
ABELR	32,532,005	27,323,610	23,073,851	84.4	3,948,228
YLR	57,215,586	37,217,266	29,650,094	79.7	6,889,720
ABYLR	44,304,260	29,487,948	23,305,692	79.0	4,014,888

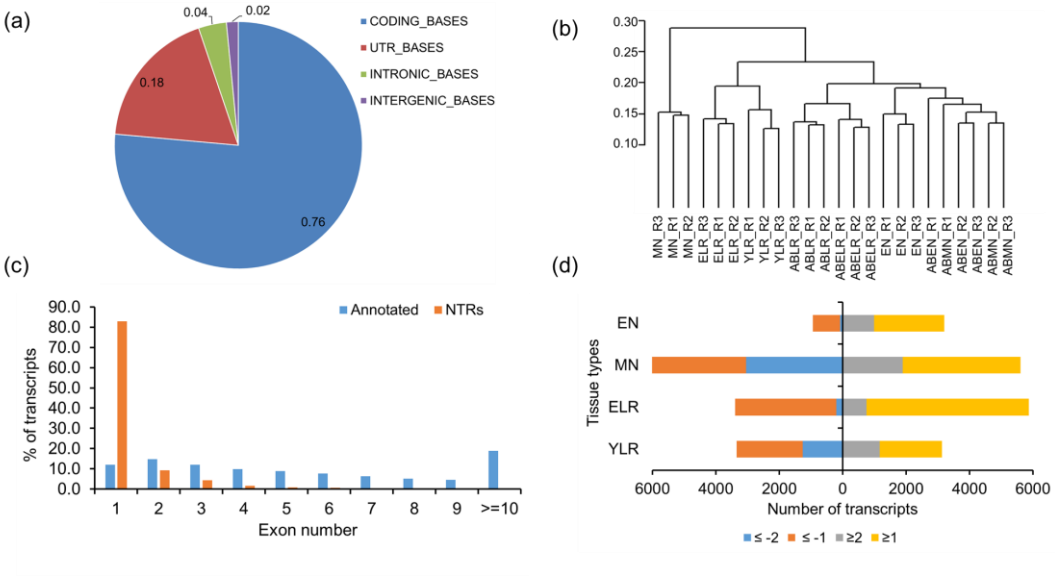


Figure 2.2 Summary of read alignment, distribution of novel transcribed regions and annotated transcripts in nodule and root transcriptome. (a) Post alignment summary of reads mapped to the soybean reference genome from all 24 libraries. The proportion of nucleotide bases mapping to coding regions (blue), UTRs (red), introns (green) and intergenic regions (purple) are indicated. (b) A dendrogram resulting from hierarchical clustering of transcript FPKM values using correlation distances between the different libraries. (c) Exon distribution in annotated transcripts (blue) and novel transcribed regions (orange) obtained from this study. (d) Numbers of significantly differentially expressed transcripts in each tissue. Only those with FPKM ≥1 in at least one tissue type are shown. The proportion of transcripts with $\log_2$ fold change of ≤-1 (orange), ≤-2 (blue), ≥1 (grey), and ≥2 (yellow) in EN, MN, ELR or YLR vs. the respective control tissues are indicated.

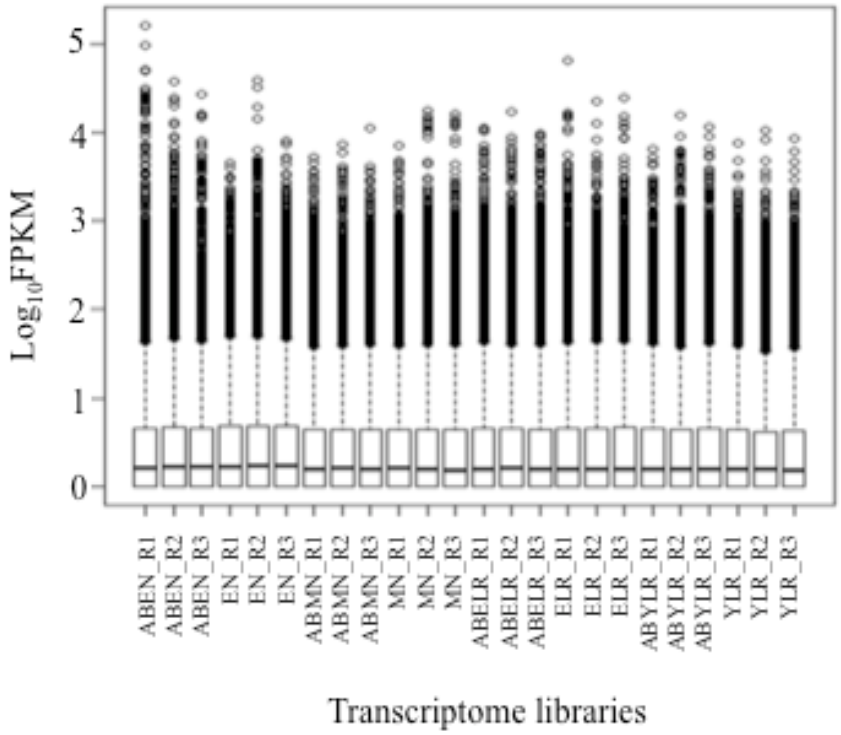


Figure 2.3 Distribution of FPKM (Log base10 transformed) in the transcriptome libraries. FPKM values belonging to all 24 libraries were used to generate the box plot.

2.2.3. Novel Transcribed Regions (NTRs) in soybean

The higher depth and unbiased coverage of RNA-seq data of nodule and LR libraries (based on Table 2.2 and Figure 2.2 a) were used to identify novel transcribed regions (NTRs) in the soybean genome. Transcripts that were assembled using reads that mapped "outside" of annotated transcripts in the genome (Gmax_v1.1_1.89) were considered as potential NTRs. A total of 3,334 potential NTRs with various gene structures and expression levels were identified. However, it is possible that part of these NTRs might be due to erroneous transcription and potential DNA contamination in the libraries. The depth of coverage was used as an indication of the authenticity of these NTRs. First, the FPKM value that will yield 100% coverage of a gene was estimated by using a regression analysis between FPKM vs. coverage depth for the known transcripts (annotated in Gmax_v1.1_1.89; with non-zero FPKM and coverage values). The resulting linear regression model (Coverage =4.138 + 26.332FPKM) indicated that an FPKM value of 3.64 with upper and lower confidence limit of 1.68 and 5.3 ($1.68< X=3.64<5.3$) corresponds to 100% coverage in our dataset (Figure 2.4). Using FPKM $\geq$ 3.64 as a filter, a list of 496 NTRs that are likely to be authentic transcripts was obtained. Unique GO assignments were obtained for 96 of the 496 NTRs suggesting that these might indeed have a functional role.

Further, the number of exons and transcript lengths of these NTRs to those of annotated genes were compared. The vast majority of NTRs appeared to be single exon genes (Figure 2.1c) with a median gene length of ~514 nucleotides (nt) while nearly 88% of the annotated genes had more than one exon and a median gene length of ~3384-nt. Indeed, about 19% of annotated genes had 10 or more exons (Figure 2.1 c). Finally,

differential expression of the 496 NTRs in the different root lateral organ libraries was evaluated. MN had the highest number of both up-regulated and down-regulated NTRs with a dominance of down-regulated NTRs (Figure 2.5). The differential expression profiles (i.e. distribution of $\log_2$ fold change values) of annotated genes vs. NTRs were comparable in EN, MN, and YLR. Interestingly, a larger number of NTRs were down-regulated in ELR compared to annotated genes (Figure 2.6). These NTRs might have a functional role in the development of LRs and nodules.

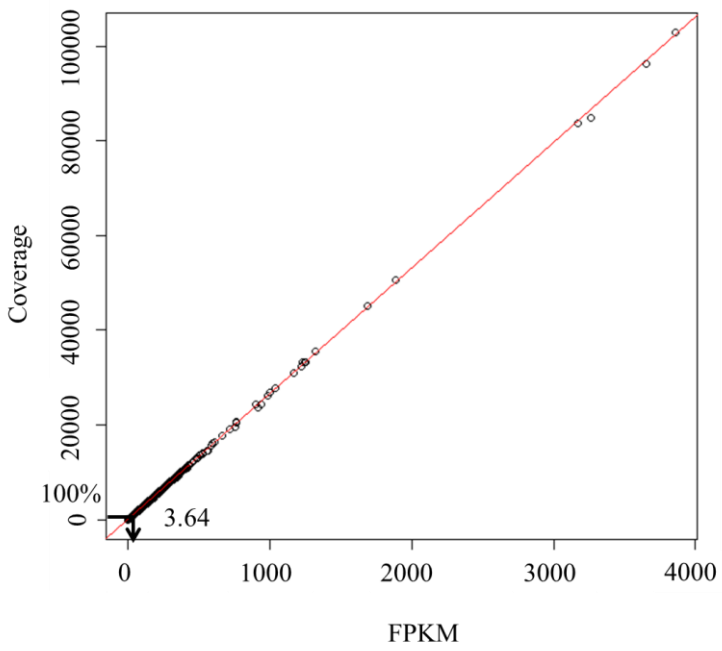


Figure 2.4 Linear regression plot between FPKM vs. Coverage. The plot was generated by using FPKM and coverage values for all annotated genes in the soybean genome (version 1.189) obtained by mapping and subsequent analysis of all 24 RNA-seq libraries together. The resultant model (Coverage =4.138 + 26.332FPKM) was used to back-predict the FPKM value required for 100% coverage.

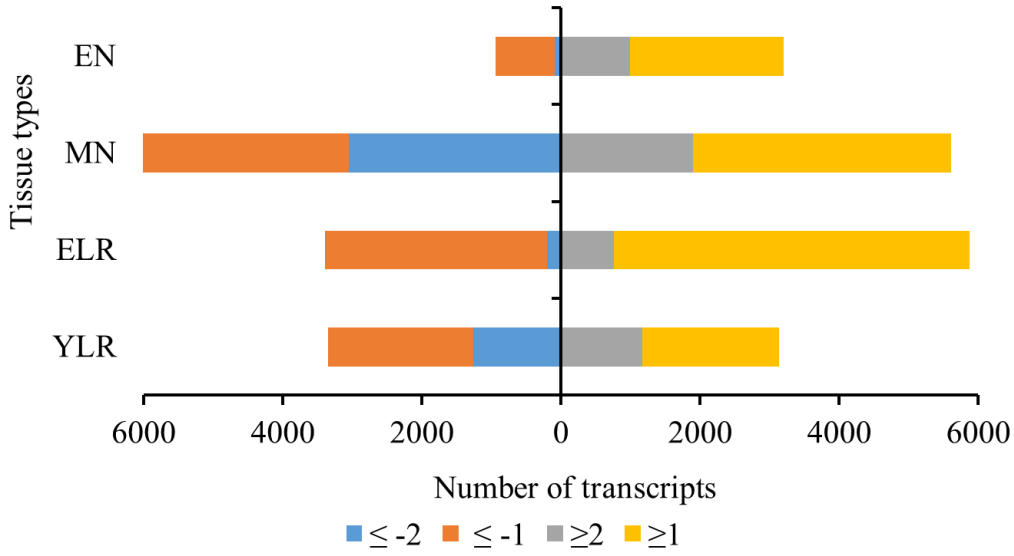


Figure 2.5 Numbers of significantly enriched NTRs in each tissue. The number of transcripts with log_2 significant fold change of ≤-1 (orange), ≤-2 (blue), ≥1 (grey), and ≥2 (yellow) in emerging nodule (EN), mature nodule (MN), emerging lateral root (ELR), and young lateral root (YLR) vs. the respective control tissues are indicated.

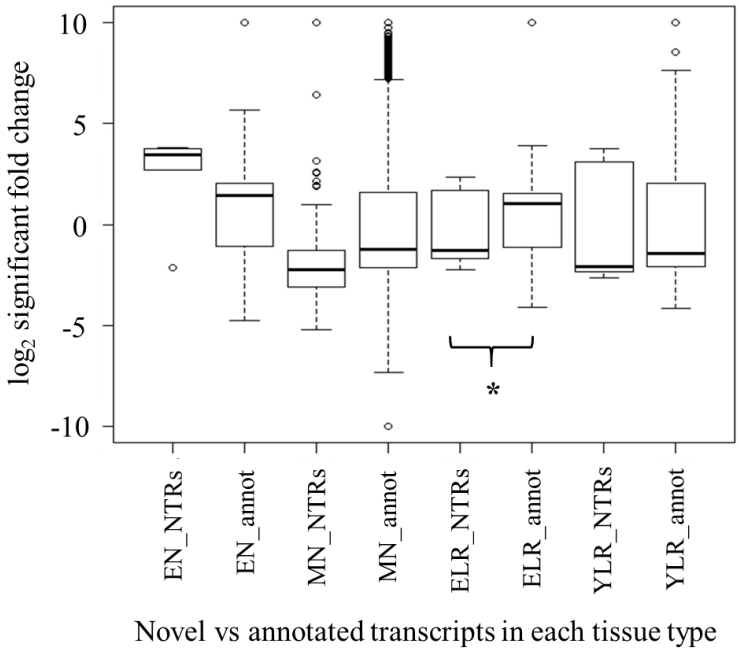


Figure 2.6 Expression pattern of annotated genes and NTRs. Both NTRs and annotated transcripts in emerging nodules (EN), mature nodules (MN), emerging lateral roots (ELR) and young lateral root (YLR) were filtered using the criteria FPKM ≥3.64 (corresponding to 100% coverage; See Figure 2.4) for consistency. Among these, those meeting the criteria of test status =OK and q-value <0.05 were used to generate boxplots. Asterisk indicates a significant difference in mean fold change value between NTRs vs. annotated genes (Welch t-test, P<0.05). Novel transcribed regions (NTRs) and annotated (annot).

2.2.4. Mature nodules had the largest difference in global gene expression patterns

To identify organ-enriched transcripts, global transcriptome in each lateral organ to its corresponding control was compared. A minimum expression threshold of FPKM ≥1 was used to reliably identify transcripts differentially expressed in respective organs vs. its control tissue (EN vs. ABEN, MN vs. ABMN, ELR vs. ABELR and YLR vs. ABYLR). Those that were differentially expressed were further classified based on the extent of differential expression i.e. $\log_2$significant fold change of ≥ 1, ≥ 2, ≤ -1 and ≤ -2. Based on the fold change, it is further categorized as "enriched" if the $\log_2$significant fold change is >0 and "reduced" if $\log_2$significant fold change is <0 compared to its control tissue. MN and ELR had the largest number of transcripts differentially expressed relative to their control root segments (Figure 2.1 d). Overall, MN had the largest number of transcripts differentially expressed vs. its control tissue. Among all four lateral organ tissues, MN also had the highest number of reduced transcripts compared to its control. This can be attributed to the specialization of MN in both metabolic and developmental pathways relative to root tissues. At younger stages of both these lateral organs (EN and ELR), there were more transcripts that were enriched in EN and ELR compared to ABEN and ABELR (Figure 2.1 d) suggesting that a number of developmental pathways are being activated in these tissues.

2.2.5. The most common biological process between EN and ELR was cell division

The list of transcripts enriched in each lateral organ tissue (≥1 $\log_2$ fold-change) was further compared to identify unique transcripts and those that are common between different lateral organ tissues using VENNY (Oliveros 2007). EN, MN, ELR, and YLR had 337, 2298, 2579 and 215 transcripts that were specifically enriched in the respective

tissue type (Figure 2.7). EN, ELR, and YLR shared 593 transcripts, and EN and ELR shared 335 transcripts. The biological processes these transcripts are involved was analyzed by singular enrichment analysis (SEA) using agriGO (Du et al. 2010) and showed that the majority of the transcripts were “down-stream” genes such as those involved in DNA replication, DNA metabolic process, microtubule-based movement, cell cycle, microtubule-based process, and DNA recombination based on. The biological processes specific to EN and ELR were consistent with the expectation that cell division and related biological processes are active in these tissues. This also suggested that very few upstream signaling or developmental pathways if any is shared between LRs and nodules. Only 32 transcripts were common between EN, MN, and YLR. In contrast, a higher number of transcripts were common between EN and ELR (335), and MN and ELR (364) (Figure 2.7). It is likely that EN or MN share specific developmental or signaling pathways with LRs at a younger stage of lateral root development, and once the root is mature they share very few. Indeed, global comparative analysis of enriched biological processes indicated that these two root lateral organs might share cell division and associated gene sets during early stages. However, these organs distinguish themselves in general at the transcriptome level consistent with the distinct hormone requirements for their initiation and development.

Among transcripts that were enriched in more than one root and nodule organs, ELR and YLR shared the most number of transcripts (563). In contrast, MN shared a relatively smaller number of enriched genes specifically with EN (299) (Figure 2.7). This might reflect the non-persistent nature of the nodule meristem compared to the persistent nature of the LR meristem in soybean. Such an observation is supported by global

correlation analysis of transcriptomes of all tissues used in the analysis (Figure 2.1 b) where ELR and YLR are clustered together in a branch as opposed to EN and MN with distinct well-separated branches. SEA revealed unique biological processes enriched in the different lateral organs e.g. oxidative stress in EN; oxidation-reduction and transmembrane transport in MN; and cell wall modification, organization or biogenesis in YLR.

To identify specific patterns of gene expression during the development of these two root lateral organs, cluster analysis was performed by using transcripts with significant enrichment in at least one tissue but a non-zero variance when enrichment values were compared between the tissues. In other words, the transcript must be specifically enriched in one or more, but not all lateral organ tissues at the same level. The number of clusters was determined by using a Fit Of Merit (FOM) analysis and subjected the resulting 17,229 transcripts to k-means cluster analysis using MeV (v4.9), in TM4 suite (Saeed et al. 2003). The resulting clusters were closely examined for patterns of fold change among tissue types. This pattern search detected organ-specific (each of EN, MN, ELR and LR), lateral root-specific (both in ELR and LR), emerging organ-specific (both in EN and ELR), and mature organ-specific (both in MN and LR) clusters. The largest clusters included the MN-enriched (3854) and MN-reduced (3383) followed by ELR-enriched (2876) and ELR-reduced (2191) (Figure 2.8).

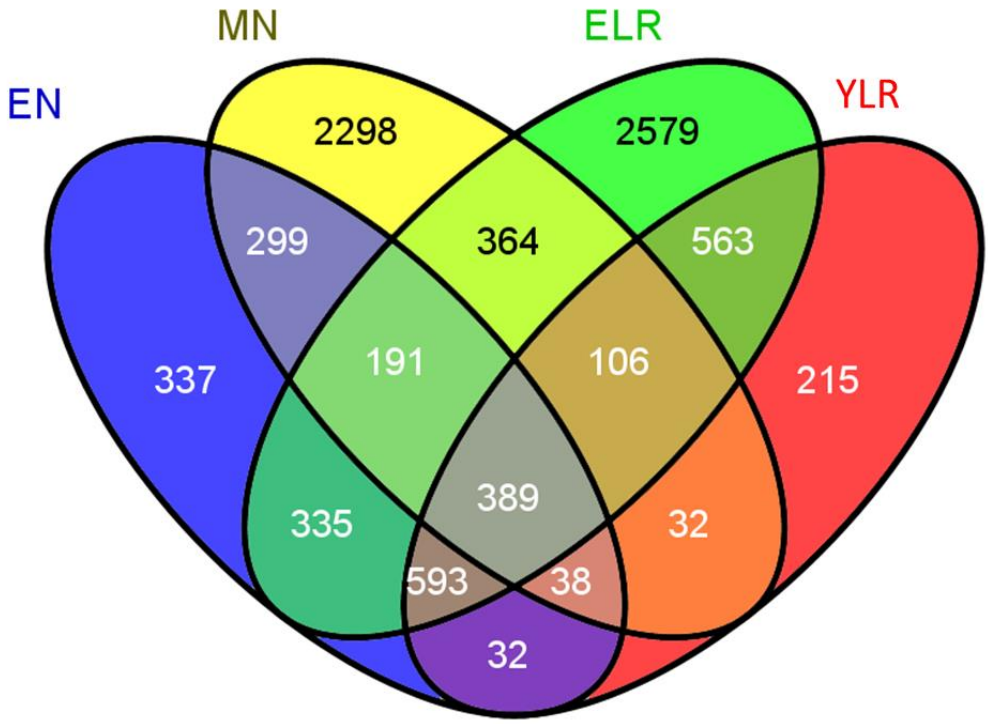


Figure 2.7 Venn diagram showing the number of transcripts specifically enriched in each tissue and in more than one tissue. Transcripts were selected based on FPKM ≥1 and log_2 *significant fold change ≥1 compared to the respective controls in emerging nodules (EN), mature nodules (MN), emerging lateral roots (ELR), and young lateral roots (YLR). The list was then supplied to http://bioinfogp.cnb.csic.es/tools/venny/ to obtain tissue-specific transcripts. Transcripts specific to EN (337), MN (2298), ELR (2579) and YLR (215) are displayed as dark blue, yellow, green, and red respectively. Based on the overlap of transcripts among EN, MN, ELR, and YLR colors and number of transcripts common between tissue types are displayed. Raw data can be obtained by contacting Subramanian lab: senthil.subramanian@sdstate.edu.*

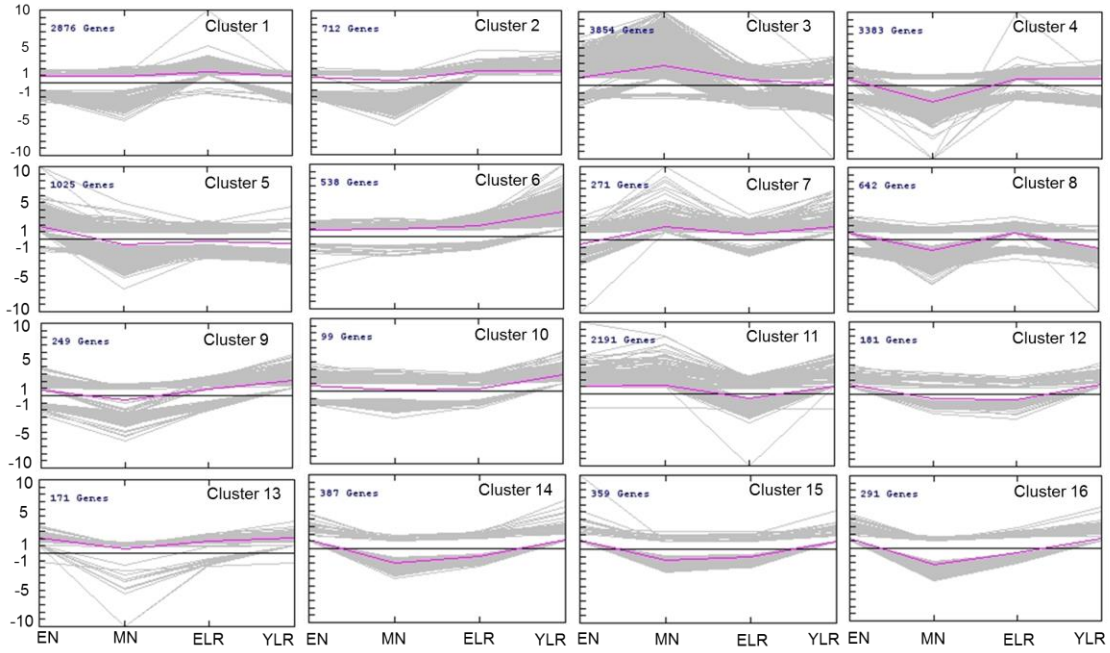


Figure 2.8 Co-expression clusters with different patterns of gene expression identified using K-means cluster analysis. The clusters were generated by using MeV (v4.9; http://www.tm4.org/). Sixteen clusters were generated guided by Fit Of Merit (FOM) analysis. Raw data can be obtained by contacting Subramanian lab: senthil.subramanian@sdstate.edu.

2.2.6. Transcriptional regulators in root lateral organs

Soybean transcription factor annotations from the Plant transcription factor database (PlantTFDB v3.0; http://planttfdb.cbi.pku.edu.cn/) (Jin et al. 2014) was used to identify those differentially expressed and enriched (≥1 $\log_2$ fold-change) in nodule and LR tissues vs. their controls. Among 58 TFs annotated in soybean 48 TF families had at least one member differentially expressed in one of the four organ tissue types. For each TF family, the unique transcripts that were enriched in EN and/or MN were summed to calculate a total number of family members enriched in nodule tissues. Similarly, the number of TFs enriched in lateral root tissues was calculated. By comparing the number of family members enriched in nodule vs. lateral root tissues, nodule-specific or -enriched, lateral root-specific or -enriched, and lateral organ non-specific (equal number of transcripts in lateral root and nodules) TF families were identified (Figure 2.9). Statistical analysis (Fisher's Exact test, $P<0.05$) of nodule- vs. lateral root-specific enrichment showed that TALE, MYB-related, MIKC, C2H2, bZIP, G2-like, WRKY, and NFY-B were either nodule-specific or significantly enriched in nodules (Figure 2.9). Overall, very distinct families of TFs appear to be active in nodule and lateral roots although morphological similarities are present between these organs.

The organ specificity of individual TF family members was also analyzed by using VENNY (Oliveros 2007). EN, MN, ELR, and YLR had 37, 189, 96 and 20 members respectively belonging to 18, 32, 34 and 11 TF families respectively that were specifically enriched in these tissues. Among some of these organ-specific TFs, MN-

specific members were represented by TFs with well-known roles in nodule development e.g. NIN-like (Castaings et al. 2009, Konishi and Yanagisawa 2013), NFY-A, NFY-B (Combier et al. 2006). Similarly, ELR showed the enrichment of GATA family members known to be associated with LR identity (De Rybel et al. 2010). Only 8 members belonging to 7 TF families were enriched in all the four tissues. EN and ELR shared 23 family members that belonged to 16 TF families including bHLH, ERF, GRAS, HD-ZIP, and MYB. However, only a single member of two TF family members (bHLH and ERF) was shared between EN and YLR. MN and ELR shared 14 family members belonging to B3, GATA, GRAS, MYB, NAC, and trihelix; but none with YLR. 23 family members were common between EN, ELR and YLR and belonged to AP2, C2H2, and GRF families. However, MN had only single members of bHLH, C2H2, and HRT-like that were common with ELR and YLR. This supported our observation above that MN and LR tissues have very few commonalities compared to ELR. EN and MN shared 41 family members that were dominated by TF families belonging to B3, bHLH, bZIP, C2H2, ERF, G2-like, MYB, and NFY. Similarly, ELR and YLR shared 38 family members belonging to ARF, bHLH, GATA, GRF, LBD and NAC.

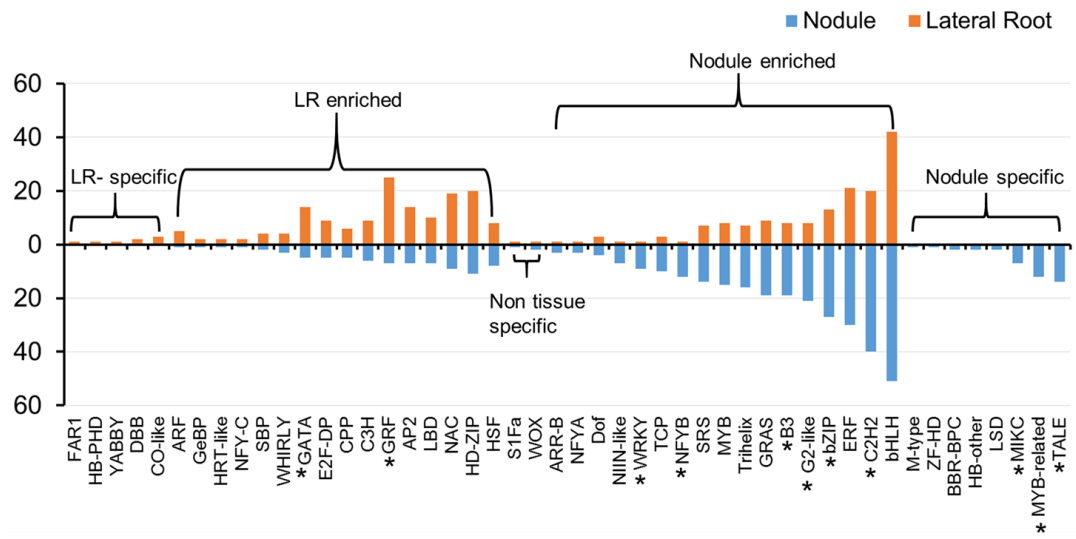


Figure 2.9 Number of family members with enriched expression in nodule or lateral root tissues in different transcription factor (TF) families. Transcripts with log_2 significant fold change of ≥1 associated with each TF family was used to identify the number of transcripts enrichment in nodule (blue) and lateral root (orange). TF annotations are based on Plant Transcription Factor Databases (PlantTFDB). Asterisks indicate TF families that were significantly nodule specific, nodule enriched and root enriched (Fisher's Exact test, P<0.05). Raw data can be obtained by contacting Subramanian lab: senthil.subramanian@sdstate.edu.

2.2.7. Hormone biosynthesis and signaling patterns during lateral root and nodule development

To gain a deeper understanding of hormone regulation, which is a key difference between LRs and nodules formation, using known Arabidopsis genes as the query, these components in soybean was identified and annotated (See experimental procedures for details). The expression patterns of genes encoding hormone biosynthesis and signaling components was examined in greater detail.

2.2.7.1. Local auxin biosynthesis occurs in both nodules and LRs

The majority of auxin biosynthesis in plants occurs via the indole pyruvate pathway composed of TRYPTOPHAN AMINOTRANSFERASE (TAA)-YUCCA (YUC) enzymes (Mashiguchi et al. 2011, Mano and Nemoto 2012). Twelve and 24 soybean orthologs for *TAA* and *YUC* was identified, respectively. Only one of the *TAA* genes was significantly differentially expressed in YLR vs. its control (Figure 2.10 a). In contrast, a number of *YUC* genes (that encode the rate-limiting step) were differentially expressed in EN, MN or ELR (Figure 2.10 a). Interestingly, the majority of them were specifically induced in either LRs or nodules; i.e. those induced in nodules were either unaffected or reduced in LRs and *vice versa*. The *GRETCHEN HAGEN3* (*GH3*) family encoding auxin conjugate enzymes also included organ-specific, and emerging stage-specific family members (Figure 2.10 a). Interestingly, the majority of *CYTOCHROME P450* (*CYP79B1* and *CYP83B1*) family members known to compete with the IPA pathway for tryptophan substrate were reduced in both nodules and LRs except a single *CYP83B1* gene that was enriched specifically in the MN (Figure 2.10 a). These data suggested that local biosynthesis of auxin might occur during nodule and LR

development in soybean, and that specific gene family members might have acquired nodule or LR specific roles. Auxin activity has been detected during nodule initiation in both determinate and indeterminate nodule forming legumes (Mathesius et al. 1998, Boot et al. 1999, Wasson et al. 2006, Suzaki et al. 2012, Turner et al. 2013). However, the source of auxin for nodule initial cell divisions and subsequent nodule development is unclear. While auxin accumulation resulting from inhibition of polar auxin transport is a likely source of auxin during indeterminate nodule formation, such inhibition of auxin transport is not crucial for determinate nodule formation (Subramanian et al. 2007). Therefore, local auxin biosynthesis might be a source of auxin during determinate nodule formation. Consistent with this notion, *YUC* orthologs in the indeterminate nodule-forming legume *M. truncatula* do not appear to have a nodule-specific expression (Benedito et al. 2008). In mature nodules, auxin-responsive gene expression has been reported in proximity to the vasculature (Takanashi et al. 2011). Sustained expression of *YUC*s in mature nodules is likely to contribute to nodule vascular development.

2.2.7.2. TIR/AFB-mediated auxin signaling is predominant in LRs compared to nodules

The majority of auxin-responsive gene expression occurs through the *TRANSPORTER INHIBITOR RESPONSE 1/AUXIN RELATED F-BOX- AUXIN/ INDOLE ACETIC ACID- AUXIN RESPONSE FACTOR* (*TIR1/AFB-Aux/IAA-ARF)* signaling pathway (Teale et al. 2006, Lau et al. 2008). *TIR1/AFB* family members were specifically induced in ELR and MN. Among the *Aux/IAA* family members, just a single-family member was specifically enriched in each of EN, MN, or LR (both ELR and YLR) tissues. Most others were induced in EN as well as in ELR or LR (and one of them

in all four tissues types). This suggested that *Aux/IAA*s might not have diversified to function specifically in LRs or nodules like *YUC*s and *GH3*s. None of *ARFs* genes were specifically induced in EN (Figure 2.10 b) which is consistent with the observation that auxin signaling/activity is perhaps suppressed during soybean nodule formation (Turner et al. 2013). There was a single *ARF3* ortholog that was specifically induced in MN. A number of *ARF*s were enriched in ELR as expected (Okushima et al. 2007, Goh et al. 2012) (Figure 2.10 b). These observations indicated that the classical *TIR/AFB-Aux/IAA-ARF* signaling cascade plays a larger role in LRs compared to nodule development even though local auxin biosynthesis appears to occur in both EN and MN.

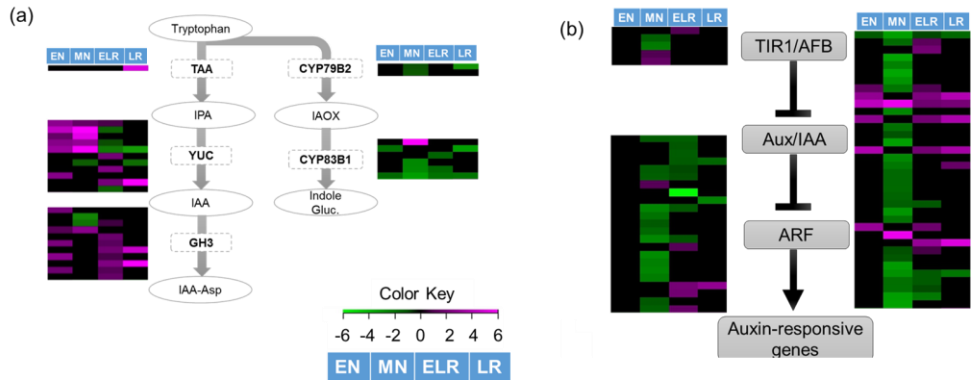


Figure 2.10 Heat maps indicating organ-specific enrichment (log_2 significant fold change – organ vs. control) of different genes overlaid on (a) auxin biosynthesis and (b) signaling pathways. Each row in a heat map corresponds to a gene family member. The four columns (as represented in the legend) in the heat map correspond to enrichment in emerging nodules (EN), mature nodules (MN), emerging lateral roots (ELR) and young lateral roots (YLR) respectively. The color key represents enrichment of transcript expression (purple), reduction of transcript expression (green) and no change in transcript expression (black). TRYPTOPHAN AMINO TRANSFERASES (TAA); Indole Pyruvic Acid (IPA); Flavin monooxygenases YUCCAs (YUC); Indole-3-acetic acid (IAA); GRETCHEN HAGEN 3 (GH3); INDOLE-3-ACETALDOXIME (IAOX); glucosinolate (Gluc); TRANSPORTER INHIBITOR RESPONSE 1/AUXIN RELATED F-BOX (TIR1/AFB); AUXIN/ INDOLE ACETIC ACID (AUX/IAA); AUXIN RESPONSE FACTOR (ARF). Raw data can be obtained by contacting Subramanian lab: senthil.subramanian@sdstate.edu.

2.2.7.3. Cytokinin biosynthesis and signaling are activated in nodules, but not in LRs

ISOPENTENYL TRANSFERASE (IPT) and LONELY GUY (LOG) are key enzymes that promote active cytokinin (CK) levels (Takei et al. 2001, Kurakawa et al. 2007) while CYTOKININ OXIDASE (CKX) irreversibly deactivates cytokinin (Hare and van Staden 1994, Jones and Schreiber 1997). Members of *IPT* and *LOG* gene families were enriched in nodules relative to LRs. In fact, not even a single *IPT* or *LOG* family member was enriched in ELR (Figure 2.11 a). Enrichment of transcripts of CK biosynthesis genes in nodules and reduction in LRs is consistent with the known positive and negative roles of cytokinin in the formation of these lateral organs (Lohar et al. 2004, Gonzalez-Rizzo et al. 2006, Laplaze et al. 2007, Bielach et al. 2012). A number of cytokinin catabolizing *CKX* genes were specifically enriched in nodules (Figure 2.11 a) indicating that active regulation of cytokinin levels occurs during nodule development. Indeed, a number of *CKX* genes are induced by cytokinin, acting in a negative feedback loop (Jones and Schreiber 1997, Kamínek et al. 1997, Motyka et al. 2003), and their enrichment might suggest high cytokinin activity.

Cytokinin perception occurs via a phosphorylation cascade involving HISTIDINE KINASE (HK), HISTIDINE PHOSPHOTRANSFER PROTEIN (HP) and RESPONSE REGULATOR (RR)-B proteins (Hwang and Sheen 2001, Inoue et al. 2001, Hutchison and Kieber 2002). This cascade transcriptionally activates *RR-A* genes whose protein products generally act as negative regulators of cytokinin signaling. *RR-B* genes were enriched only in MN (Figure 2.11 b). However, the majority of *RR-A* genes were enriched in both EN and MN; and a smaller number in LRs. Together, these data indicate

that there is high cytokinin activity during nodule vs. LR development. Enrichment of *HKs* and *HPs* were observed in the ELR and LR (Figure 2.11 b) but it is important to note that HKs, HPs, and type-B RRs are activated by phosphorylation and transcript levels might not be a true indicator of their activities. Higher expression of *RR* genes specifically in nodule tissues indicated a larger role for CK signaling during nodule development. These observations are consistent with the known roles of these hormones in nodule and LR development (Lohar et al. 2004, Gonzalez-Rizzo et al. 2006, Murray et al. 2007, Madsen et al. 2010). In addition, these observations have led to the identification of specific family members of cytokinin biosynthesis and signaling components likely to be crucial for nodule development in soybean.

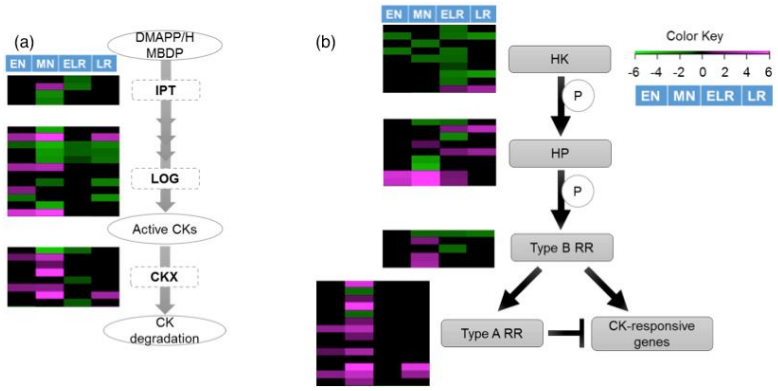


Figure 2.11 Heat maps indicating organ-specific enrichment (log_2 significant fold change – organ vs. control) of different genes overlaid on (a) cytokinin biosynthesis and (b) signaling pathways. Each row in a heat map corresponds to a gene family member. The four columns (as represented in the legend) in the heat map correspond to enrichment in emerging nodule (EN), mature nodule (MN), emerging lateral root (ELR) and young lateral root (YLR) respectively. The color key represents enrichment of transcript expression (purple), reduction of transcript expression (green) and no change in transcript expression (black). Cytokinin (CK); ISOPENTYL TRANSFERASE (IPT); LONELY GUY (LOG); CYTOKININ DEHYDROGENASE (CKX); HISTIDINE KINASE (HK); HISTIDINE PHOSPHOTRANSFER (HP), RESPONSE REGULATOR (RR). Raw data can be obtained by contacting Subramanian lab: senthil.subramanian@sdstate.edu.

2.2.7.4. A coordinated increase in active gibberellin levels might occur during nodule development

Gibberellic acid (GA) 20-oxidase (*GA20ox*) and GA3 oxidase (*GA3ox*) are involved in GA biosynthesis while GA2 oxidase (*GA2ox*) is involved in GA catabolism (reviewed by (Yamaguchi 2008)). Genes encoding GA biosynthesis (*GA20ox* and *GA3ox*) were induced in both EN and MN whereas those involved in GA inactivation (*GA2ox*) were reduced in MN (Figure 2.12 a). This suggested that active GA levels are promoted during nodule development. Such consistent change in GA levels was not observed in LR tissues. Gene encoding components of GA perception (*GIBBERELLIN INSENSITIVE DWARF1* (*GID1*) and *DELLA*) (Sun 2010) were primarily induced in MN and mostly reduced in ELR (Figure 2.12 b). This observation seemingly contradicts previous reports on the negative effect of GA on determinate nodule development. For example, the exogenous application of the GA reduces the number of infection thread and nodules in *L. japonicus* (Maekawa et al. 2009). However, GA appears to inhibit the infection process, but not the later stages of nodule development such as primordium initiation or nodule maturation in soybean (Williams and Sicardi de Mallorca 1984). Indeed, up-regulation of GA biosynthesis genes *GA20ox* and *GA30x* was observed at very early stage of nodule development in soybean (Hayashi et al. 2012). Therefore, local biosynthesis of active GAs in nodule tissues might play a key role in soybean nodule development. The low activity of GA biosynthesis and signaling genes in ELR and LR could be related to the known negative role of GA in lateral root development (Gou et al. 2010) and root biomass (Berova and Zlatev 2000).

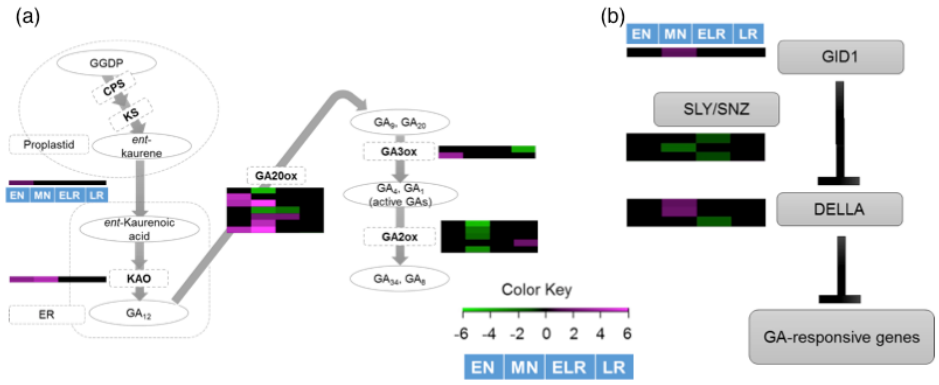


Figure 2.12 Heat maps indicating organ-specific enrichment (log_2 significant fold change – organ vs. control) of different genes overlaid on (a) gibberellin biosynthesis and (b) signaling pathways. Each row in a heat map corresponds to a gene family member. The four columns (as represented in the legend) in heat map correspond to enrichment in emerging nodule (EN), mature nodule (MN), emerging lateral root (ELR) and young lateral root (YLR) respectively. The color key represents enrichment of transcript expression (purple), reduction of transcript expression (green) and no change in transcript expression (black). ENT-COPALYL DIPHOSPHATE SYNTHASE (CPS); ENT-KAURENE SYNTHASE (KS); ENT-KAURENE OXIDASE (KO); ENT-KAURENOIC ACID OXIDASE (KAO); GA_{20}-OXIDASES (GA20ox); GA_3-OXIDASES (GA3ox); GA_2-OXIDASE (GA2ox); GIBBERELLIN INSENSITIVE DWARF1 (GID1); THE DELLA GROWTH INHIBITORS (DELLA); SLEEPY/SNEEZY (SLY/SNZ). Raw data can be obtained by contacting Subramanian lab: senthil.subramanian@sdstate.edu.

2.2.7.5. Brassinolide activity is enhanced during lateral organ formation

Evaluation of brassinosteroid (BR) biosynthesis and signaling genes (Shimada et al. 2003, Ye et al. 2011) indicated that a number of key biosynthesis genes (*DWARF3* (*DWF3)* and *CYP83A* were enriched in both nodule and LR tissues (Figure 2.13 a). Similarly, one of the transcripts encoding *BRASSINOSTEROID INSENSITIVE 1* (*BRI1*) was enriched in all lateral organ tissue types. The general theme was that BR signaling was induced in both LRs and nodules, but to a relatively higher level in LRs (Figure 2.13 b). Previous work suggested that BRs might play both negative (Hunter 2001, Terakado et al. 2005) and positive (Upreti and Murti 2004, Ferguson et al. 2005) roles in nodule formation (reviewed by (Ferguson and Mathesius 2014)). In soybean, foliar application or direct injection of BR in root base of supernodulating mutant (EN6500) reduced the nodule number but not in the parent line (Terakado et al. 2005). In contrast, reduction in the numbers of nodules as well as LRs was observed in BR mutants of Pea (Ferguson et al. 2005). Increase in nodule number was observed in *Phaseolus vulgaris* (L. cv. Arka Suvidha) by pretreating with BR and induction of stress (Upreti and Murti 2004). BR is known to play a positive role in lateral root formation by a synergetic effect with auxin (Bao et al. 2004, Vercruyssen et al. 2011). However, the overall pattern of differential expression of BR biosynthesis and signaling genes from our study shows that it might have a positive role in nodule and LR development.

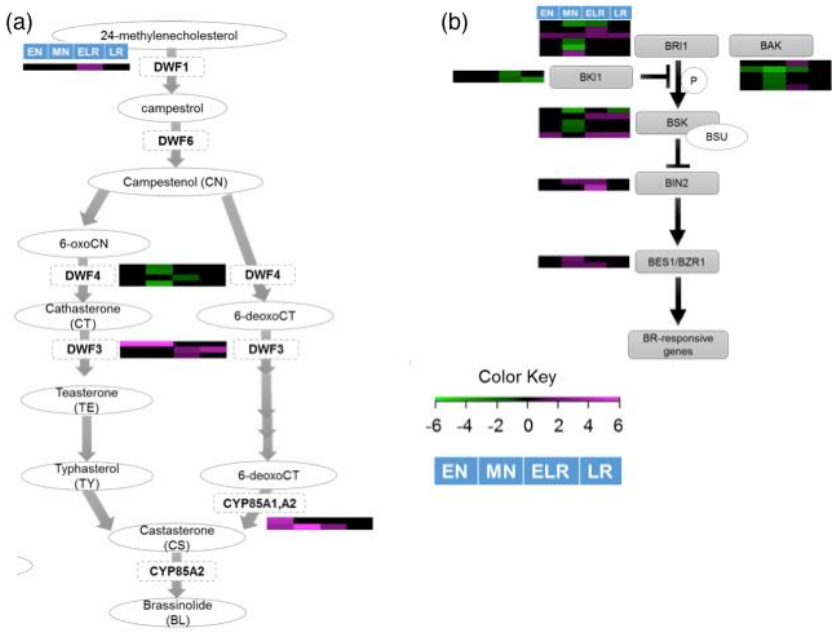


Figure 2.13 Heat maps indicating organ-specific enrichment (log_2 significant fold change – organ vs. control) of different genes overlaid on (a) brassinolide biosynthesis and (b) signaling pathways. Each row in a heat map corresponds to a gene family member. The four columns (as represented in the legend) in heat map correspond to enrichment in emerging nodule (EN), mature nodule (MN), emerging lateral root (ELR) and young lateral root (YLR) respectively. The color key represents enrichment of transcript expression (purple), reduction of transcript expression (green) and no change in transcript expression (black). DWARF1/DIMINUTO (DWF1/DIM); DWARF6/DEETIOLATED2 (DWF6/DET2); DWARF3 (DWF3); DWARF4 (DWF4); BRASSINOSTEROID INSENSITIVE 1 (BRI1); BRI1 KINASE INHIBITOR 1 (BKI1); BRI1- ASSOCIATED RECEPTOR KINASE 1 (BAK1); BR SIGNALING KINASE 1 (BSK1); BRI1-SUPPRESSOR 1 (BSU1); BRASSINOSTEROID INSENSITIVE 2 (BIN2); BRI1-EMS SUPPRESSOR 1 (BES1); BRASSINAZOLE RESISTANT 1 (BZR1). Raw data can be obtained by contacting Subramanian lab: senthil.subramanian@sdstate.edu.

2.2.7.6. Abscisic acid, ethylene, salicylic acid and jasmonic acid activities are reduced during LR and nodule development

Enrichment of transcripts encoding specific abscisic acid (ABA) biosynthetic genes (*SHORT CHAIN ALCOHOL DEHYDROGENASE REDUCTASE* (*SDR/ABA2)* and *ABSCISIC ALDEHYDE OXIDASE* (*AAO*) in nodules coupled with a reduction in the ABA catabolic gene *CYP707* suggested that ABA activity might be reduced in nodule tissues (Figure 2.14 a). However, transcripts encoding *9-CIS-EPOXYCAROTENOID DIOXYGENASES* (*NCED*), the classical rate-limiting enzyme (Xiong and Zhu 2003) did not appear to be enriched in nodules. A number of gene family members were regulated similarly in nodule and LR tissues suggesting that no functional diversification might have occurred. Except the enrichment of specific *PYRABACTIN RESISTANCE*/*PYR RELATED* (*PYR/PYL)* ABA receptors and the downstream *SUCROSE NON-FERMENTING-1-RELATED PROTEIN KINASES* (*SnRK2*s), most of the transcripts encoding components of ABA signaling modules (Hubbard et al. 2010) were reduced in MN and ELR (Figure 2.14 b). Given the known inhibitory role of ABA in nodule development (Suzuki et al. 2004, Ding et al. 2008, Tominaga et al. 2009), it was not surprising to observe minimal enrichment of ABA biosynthesis or signaling genes in nodule tissues.

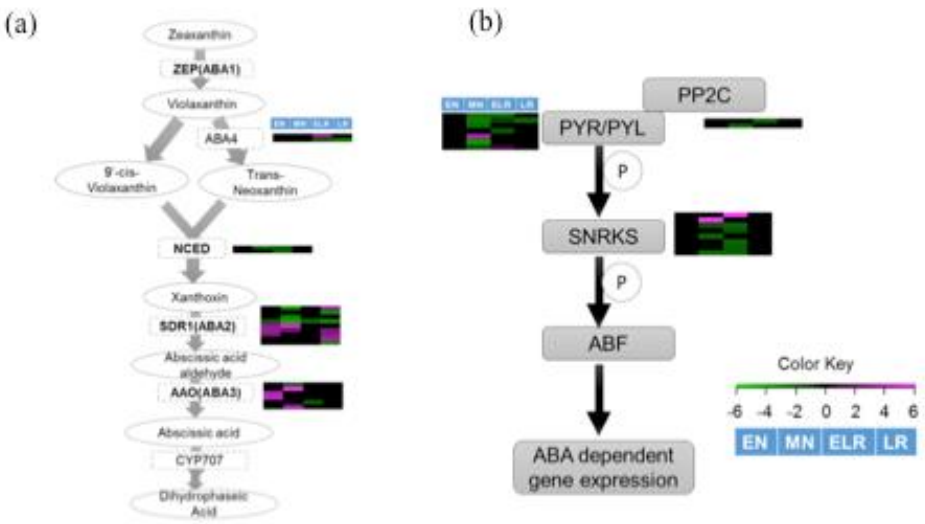


Figure 2.14 Heat maps indicating organ-specific enrichment (log_2 significant fold change – organ vs. control) of different genes overlaid on (a) ABA biosynthesis and (b) signaling pathways. Each row in a heat map corresponds to a gene family member. The four columns (as represented in the legend) in heat map correspond to enrichment in emerging nodule (EN), mature nodule (MN), emerging lateral root (ELR) and young lateral root (YLR) respectively. The color key represents enrichment of transcript expression (purple), reduction of transcript expression (green) and no change in transcript expression (black). ZEAXANTHIN EPOXIDASE (ZEP); 9-CIS-EPOXYCAROTENOID DIOXYGENASES (NCED); SHORT CHAIN ALCOHOL DEHYDROGENASE REDUCTASE (SDR1); ABSCISIC ALDEHYDE OXIDASE (AAO); PYRBACTIN RESISTANCE (PYR); PYR RELATED (PYL); SERINE/THREONINE PHOSPHATASES TYPE 2C (PP2C); SUCROSE NON-FERMENTING-1-RELATED PROTEIN KINASES (SNRKS). Raw data can be obtained by contacting Subramanian lab: senthil.subramanian@sdstate.edu.

The majority of the genes encoding *ACC SYNTHASE* (*ACS*) and *ACC OXIDASE* (*ACO*) that catalyze rate limiting steps of ethylene biosynthesis (Wang et al. 2002) were reduced in all the lateral organs (Figure 2.15 a). Ethylene signaling genes were mostly reduced or unaltered except one member of *ETHYLENE RECEPTOR 1* (*ETR1*) and *ETHYLENE INSENSITIVE3* (*EIN3*) transcript enriched in ELR (Figure 2.15 b). The expression pattern of the majority of these transcripts is consistent with known negative role of ethylene in lateral root formation (Negi et al. 2008, Negi et al. 2010, Lewis et al. 2011) and nodule formation in *P. sativum* (L. cv sparkle), *M. sativa*, and *L. japonicus* (Peters and Crist-Estes 1989, Ligero et al. 1991, Lee and LaRue 1992, Nukui et al. 2000) although soybean nodule formation might be independent of the ethylene signaling (Hunter 1993, Suganuma et al. 1995, Schmidt et al. 1999).

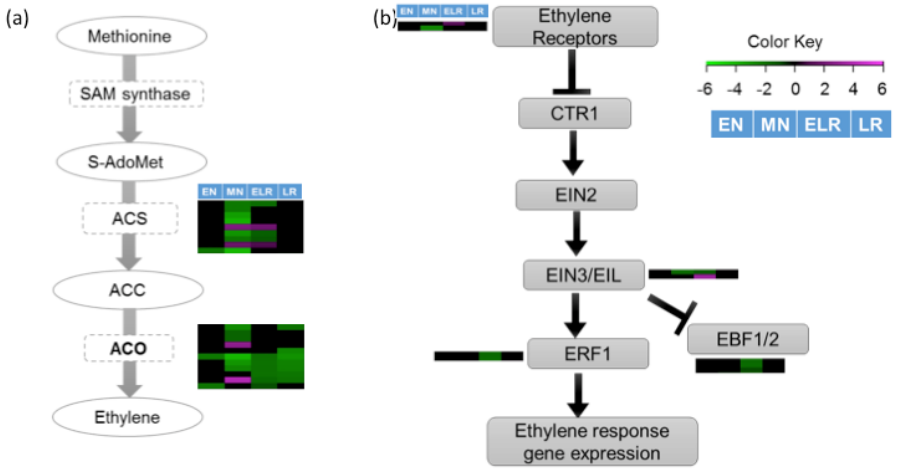


Figure 2.15 Heat maps indicating organ-specific enrichment (log_2 significant fold change – organ vs. control) of different genes overlaid on (a) ethylene biosynthesis and (b) signaling pathways. Each row in a heat map corresponds to a gene family member. The four columns (as represented in the legend) in heat map correspond to enrichment in emerging nodule (EN), mature nodule (MN), emerging lateral root (ELR) and young lateral root (YLR) respectively. The color key represents enrichment of transcript expression (purple), reduction of transcript expression (green) and no change in transcript expression (black). S –adenosyl methionine (S-AdoMet); ACC SYNTHASE (ACS); 1-aminocyclopropane-1-carboxylic acid (ACC); ACC OXIDASE (ACO); CONSTITUTIVE TRIPLE RESPONSE 1 (CTR1); ETHYLENE INSENSITIVE3/EIN3-LIKE (EIN/EIL); ETHYLENE RESPONSE FACTOR 1 (ERF1); EIN3-BINDING-F-BOX PROTEIN 1 OR 2 (EBF1/2). Raw data can be obtained by contacting Subramanian lab: senthil.subramanian@sdstate.edu.

The majority of jasmonic acid (JA) biosynthesis genes (e.g. *ALLENEOXIDE SYNTHASE* (*AOS)*, *ALLENE OXIDE CYCLASE* (*AOC),* and *12-OXOPHYTODIENOATE REDUCTASE3* (*OPR3)*(Zdyb et al. 2011) were unchanged or reduced in both nodule and LR tissues (Figure 2.16 a). Similarly, the majority of JA signaling components was also reduced. However, enrichment of *MYC2* transcripts was observed specifically in nodule tissues (Figure 2.16 b).

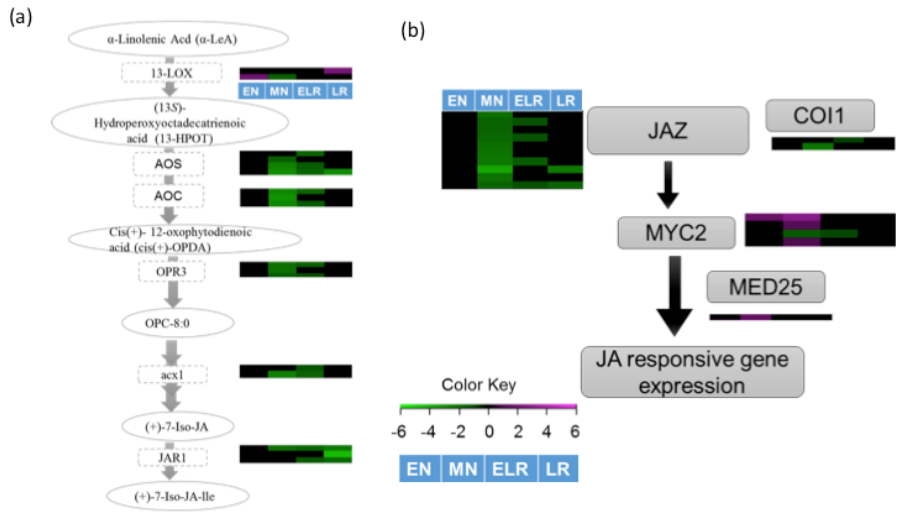


Figure 2.16 Heat maps indicating organ-specific enrichment (log_2 significant fold change – organ vs. control) of different genes overlaid on (a) jasmonic acid biosynthesis and (b) signaling pathways. Each row in a heat map corresponds to a gene family member. The four columns (as represented in the legend) in heat map correspond to enrichment in emerging nodule (EN), mature nodule (MN), emerging lateral root (ELR) and young lateral root (YLR) respectively. The color key represents enrichment of transcript expression (purple), reduction of transcript expression (green) and no change in transcript expression (black). 13-LIPOXYGENASE (13-LOX); ALLENEOXIDE SYNTHASE (AOS); ALLENE OXIDE CYCLASE (AOC); 12-OXOPHYTODIENOATE REDUCTASE3 (OPR3); ACYL COA OXIDASE (ACX); JASMONOYL ISOLEUCINE CONJUGATE SYNTHASE 1 (JAR1); jasmonoyl- L-isoleucine (JA-Ile); CORONATINE INSENSITIVE 1 (COI1); JAZMONATE ZIM DOMAIN (JAZ); MEDIATOR25 (MED25).

Raw data can be obtained by contacting Subramanian lab:

senthil.subramanian@sdstate.edu.

No difference in transcript activity of the salicylic acid (SA) biosynthesis gene (*ISOCHORISMATE SYNTHASE* (*ICS*) and *ISOCHORISMATE PYRUVATE LYASE* (*IPL*) was observed in all the tissue types compared to their control (Figure 2.17 a). An enrichment of a transcript encoding *SAGT* (*SA-glucosyltransferase*) involved in the conversion of SA to SGE (Salicylated Glucose Ester) was observed both in ELR and YLR (Figure 2.17 a). Other SA deactivation and activation enzymes were reduced in all tissues. Except a single *TGA* gene enriched in EN, the majority of genes encoding components of SA signaling were reduced in MN, ELR, and YLR (Figure 2.17 b). Suppression of JA biosynthesis did not affect nodule formation in *M. truncatula* roots indicating that JA does not play a key role in determinate nodule (Zdyb et al. 2011). However, a potential negative role of JA has been reported in *L. japonicus* where MeJA application reduced the number of nodules and lateral roots (Nakagawa and Kawaguchi 2006). However, SA appears to play a negative role in nodule formation (Sato et al. 2002, Stacey et al. 2006). Consistently, we observed reduction of the majority of JA and SA biosynthesis and signaling genes suggesting a neutral or negative role in soybean nodule development.

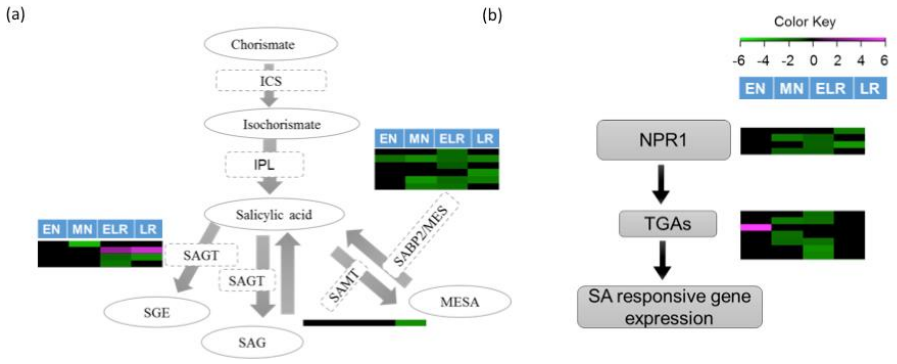


Figure 2.17 Heat maps indicating organ-specific enrichment (log_2 significant fold change – organ vs. control) of different genes overlaid on (a) salicylic acid biosynthesis and (b) signaling pathways. Each row in a heat map corresponds to a gene family member. The four columns (as represented in the legend) in heat map correspond to enrichment in emerging nodule (EN), mature nodule (MN), emerging lateral root (ELR) and young lateral root (YLR) respectively The color key represents enrichment of transcript expression (purple), reduction of transcript expression (green) and no change in transcript expression (black). ISOCHORISMATE SYNTHASE (ICS); ISOCHORISMATE PYRUVATE LYASE (IPL); SA GLUCOSYLTRANSFERASE (SAGT); SA METHYLTRANSFERASE (SAMT); SA BINDING PROTEIN (SABP2); METHYL ESTERASE (MES); NON-EXPRESSOR OF PATHOGENESIS RELATED GENES 1 (NPR1); TGA transcription factors (TGAs). Raw data can be obtained by contacting Subramanian lab: senthil.subramanian@sdstate.edu.

2.2.7.7. Strigolactone activity is high in emerging nodules

DWARF27 (D27), CAROTENOID CLEAVAGE DIOXYGENASES 7 and 8 (CCD7 and CCD8) and the cytochrome P450 monooxygenase, MORE AXILLARY GROWTH1 (MAX1) are the enzymes involved in strigolactone (SL) biosynthesis. In EN, *CCD8* was reduced and *MAX1* was enriched compared to ABEN but *D27* and *CCD7* were unchanged. Only *D27* was enriched in MN and reduced in ELR. However, no differential expression pattern was observed in YLR for any of the transcripts involved in SL biosynthesis (Figure 2.18 a). MORE AXILLARY GROWTH2 (MAX2) and DWARF14 (D14) are enzymes involved in SL signaling (Bennett and Leyser 2014, Koltai 2014). Differential expression of *MAX2* was observed in EN, MN, and ELR but *D14* only in MN. Similar to biosynthesis, no differential expression of transcripts involved in signaling was observed in YLR (Figure 2.18 b). SL is shown to play a positive role in the formation of indeterminate nodules e.g. pea (Foo and Davies 2011), and alfalfa (Soto et al. 2010) as well as determinate nodules e.g. *L. japonicus* (Liu et al. 2013). It is possible that SL play a positive role in nodule formation but not in nodule maturation and lateral root formation.

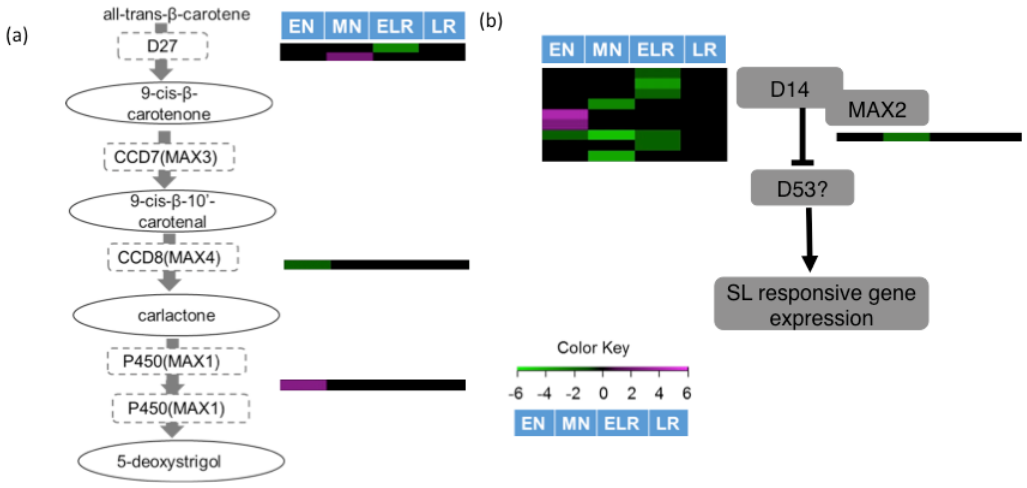


Figure 2.18 Heat maps indicating organ-specific enrichment (log_2 significant fold change – organ vs. control) of different genes overlaid on (a) strigolactone biosynthesis and (b) signaling pathways. Each row in a heat map corresponds to a gene family member. The four columns (as represented in the legend) in heat map correspond to enrichment in emerging nodule (EN), mature nodule (MN), emerging lateral root (ELR) and young lateral root (YLR) respectively. The color key represents enrichment of transcript expression (purple), reduction of transcript expression (green) and no change in transcript expression (black). DWARF27 (D27), CAROTENOID CLEAVAGE DIOXYGENASE 7 (CCD7), CAROTENOID CLEAVAGE DIOXYGENASE 8 (CCD8), CYTOCHROME P450 (P450), MORE AXILLARY GROWTH (MAX). Raw data can be obtained by contacting Subramanian lab: senthil.subramanian@sdstate.edu.

2.2.8. A number of shoot axillary meristem markers are enriched in nodule tissues

Comparative analysis of this large transcriptome data set suggested that a number of TFs and hormone biosynthesis and signaling genes had clearly diversified in their function between LRs and nodules. Interestingly, suppression of auxin activity and induction of cytokinin activity plays a key role in the formation of shoot lateral meristems (Wang et al. 2014a, Wang et al. 2014b) similar to nodule formation (Turner et al. 2013, Nizampatnam et al. 2015). This is in contrast to LR initiation that requires a suppression of cytokinin activity and promotion of auxin activity (Werner et al. 2003, Laplaze et al. 2007, Peret et al. 2009, Lavenus et al. 2013). The expression of orthologs and gene family members of shoot axillary meristem-specific genes in nodule and LR tissues was examined. Shoot axillary buds develop from boundary cells between the apical meristem and leaves through a sequence of axillary meristem initiation, bud differentiation, and shoot elongation. A number of genes that are specifically expressed in organ boundaries and also play a role in bud formation is known in Arabidopsis and other plants. Members of the *BLADE ON PETIOLE* (*BOP*), *CUP-SHAPED COTYLEDON* (*CUC*), *LATERAL ORGAN BOUNDARIES* (*LOB*), *LATERAL SUPPRESSOR* (*LAS*), *LATERAL ORGAN FUSION* (*LOF*), *ORGAN BOUNDARY1* (*OBO*), and *REGULATORS OF AXILLARY MERISTEMS* (*RAX*) families show boundary-specific expression patterns in *Arabidopsis* and other plant species (Aida et al. 1997, Shuai et al. 2002, Greb et al. 2003, Ha et al. 2003, Schmitz and Theres 2005, Aida and Tasaka 2006, Keller et al. 2006, Müller et al. 2006, Borghi et al. 2007). The meristem is initiated through the induction of the KNOX gene *SHOOT MERISTEMLESS* (*STM*). Subsequent suppression of *STM* expression leads to the formation of leaf/floral primordia on the axillary meristem transitioning the bud to

the differentiation stage (Long and Barton 2000). Other genes that act in axillary buds to regulate bud formation and differentiation include the TCP transcription factor *BRANCHED1* (Koyama et al. 2007) and *PENNYWISE* (*PNY*) and *POUND-FOOLISH* (*PNF*), which encode KNOX-interacting BELL homeodomain proteins (Kanrar et al. 2006). Potential orthologs and other family members of these gene families in soybean through ortho and tribe annotations (Li et al. 2012) were identified and their expression patterns was examined in the nodule and LR dataset. Interestingly, the expression of a number of genes belonging to these shoot axillary bud-specific gene families was significantly induced in nodule tissues (Figure 2.19). Importantly, these genes were not induced or were reduced in LR tissues. For example, orthologs of *RAX, PNF/PNY, BOP, LBD, LOF, OBO,* and *MAX1* gene family members were significantly induced in nodule tissues, but not in LR tissues. However, members of *CUC* and *STM* gene families did not show a significant induction in either nodule or LR tissues relative to the control tissues. This might either be due to their low expression levels or transient expression patterns (e.g. *STM*) or domains of expression being larger (e.g. *CUC* genes) to include part of the control tissues above and below the lateral organ tissues (Figure 2.19). The strikingly nodule-specific induction of gene families known to play a role in formation and differentiation of shoot axillary buds strongly indicates that soybean nodules might be modified shoot organs.

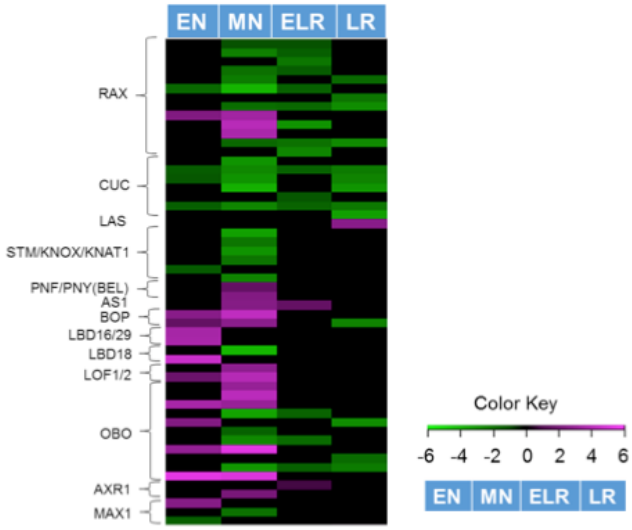


Figure 2.19 Heat map indicating organ-specific enrichment (log2 significant fold change – organ vs. control) of different genes involved in shoot axillary meristem development. Each row in the heat map corresponds to a gene family member. The four columns (as represented in the legend) in heat map correspond to enrichment in emerging nodule (EN), mature nodule (MN), emerging lateral root (ELR) and young lateral root (YLR) respectively. The color key represents enrichment of transcript expression (purple), reduction of transcript expression (green) and no change in transcript expression (black). REGULATORS OF AXILLARY MERISTEMS (RAX); CUP-SHAPED COTYLEDON (CUC); LATERAL SUPPRESSOR (LAS); SHOOT MERISTEMLESS/ KNOTTED-LIKE HOMEOBOX/HOMEOBOX PROTEIN KNOTTED-1-LIKE 1(STM/KNOX/KNAT1); POUND-FOOLISH/PENNYWISE (PNF/PNY); ASYMMETRIC LEAVES 1(AS1); BLADE ON PETIOLE (BOP); LATERAL ORGAN BOUNDARIES-DOMAIN (LBD); LATERAL ORGAN FUSION (LOF); ORGAN BOUNDARY1 (OBO); AUXIN RESISTANT 1 (AXR1); MORE AXILLARY GROWTH 1(MAX1). Raw data can be obtained by contacting Subramanian lab: senthil.subramanian@sdstate.edu.

In summary, analysis of LR and nodule transcriptome show that (i) several known signaling elements involved in the initiation of LRs (e.g. *ARF5*, *CRF2*, *LRP1*, *PIN1* and *TMO7*) were poorly expressed or not enriched in nodules; (ii) while a number of genes are enriched in both EN and ELR, the vast majority of them are associated with cell division and not upstream signaling; nodules; (iii) distinct sets of transcription factor families are enriched in LR vs. nodule tissues; (iv) local auxin biosynthesis occurs in both LRs and nodules, but with distinct family members active in each organ; (v) auxin signaling pathway is predominant in LRs, but down-regulated in nodules; (vi) cytokinin biosynthesis and signaling are active in nodules, but not in LRs; (vii) GA biosynthesis is specifically activated in nodules but suppressed in LRs; and (viii) orthologs of several transcription factors involved in shoot axillary bud formation are specifically enriched in nodules but not LRs. Comparative transcriptome results also suggest that soybean nodules are not similar to LRs in their gene expression profiles especially when comparing those that encode key TFs and hormone biosynthesis and signaling components, whereas they share similarities with shoot axillary buds not only in expression patterns of marker genes but also hormone requirements.

2.3. Discussion

2.3.1. An accurate set of nodule enriched genes

To determine key developmental and hormonal pathways active in LRs and nodules, the organ-specific/enriched transcriptomes of these tissues in soybean at two different developmental stages were compared. The use of adjacent root segments as control tissues helped clarify some outstanding questions on nodule-specific/enriched genes identified by comparing nodule transcriptomes to mock-inoculated roots. Indeed,

global correlation analysis of transcriptome libraries of LRs and nodules indicated that control tissue of each lateral organ had transcriptomes distinct enough to distinguish themselves from other tissues (Figure 2.1b); the only exception being ABEN and ABMN samples. Therefore, root tissues adjacent to nodules had distinct transcriptome profiles compared to similarly aged tissues in uninoculated roots. The use of adjacent root segments as controls identified that a number of genes previously annotated as being nodule-specific/enriched (Schmutz et al. 2010) were actually enriched also in inoculated root tissues adjacent to nodules (ABMN). Notable among these include Glyma15g20180.1 and Glyma09g08550.5 homeolog of Glyma13g17420 (*Nod100*), Glyma02g15370.1 a homeolog of Glyma07g33090 (*SAN1A*), Glyma08g08090.1 a paralog of Glyma05g25010.1 (*Nodulin 21*), Glyma14g23780.1, Glyma17g03340.1, Glyma15g00620.1, Glyma12g04390.1 (*GmNARK*), Glyma01g36080.2, and Glyma11g09330.1. These genes while highly expressed in nodules were not necessarily nodule-specific/enriched. Within nodules, stage-specific expression of several nodulin genes were also detected. For example Glyma08g12650.1 (*Nodulin26a*), Glyma19g22210.1 (*Nodulin26b*), and Glyma10g06810.1 (*Nodulin61*) had a very high expression in mature nodules but no or very low expression in emerging nodules. On the other hand, Glyma05g25010.1 (*Nodulin21*), Glyma18g02230.1, Glyma10g23790.1 (*Nod35*), Glyma02g36580.1 (*Nod55-1*), Glyma17g08110.1 (*Nod55-2*), Glyma17g08110.1 (*N315*), and Glyma02g43320.1 (*Nodulin16*) had high expression and enrichment in emerging and mature nodules compared to their controls. In addition, several genes previously annotated as nodule-specific were also found to be expressed at high levels in LRs. For example Glyma20g05530.1, Glyma14g01160.1 showed

significant enrichment in MN as well as the ELR and YLR. Likewise, Glyma01g36080.2 and Glyma11g09330.1 were not nodule-specific and were indeed enriched in ELR and YLR, but reduced in MN. Thus, use of the adjacent root tissue as control has identified an accurate set of nodule-specific genes in soybean.

2.3.2. Stage-specific action of nod factor signaling components

The expression patterns of genes encoding nod factor signaling components in soybean were evaluated. Known signaling components identified in *L. japonicus* and *M. truncatula* was used to identify potential orthologs in soybeans and evaluated their expression in EN and MN. The nod factor receptor Glyma02g43860 (*NFR1*) was enriched in EN and MN; however *NFR5* was not enriched in nodules relative to the control tissues and was reduced in YLR. Therefore, unlike *NFR1, NFR5* might act primarily in epidermal responses and not participate in continued nod factor recognition through nodule development. The expression pattern of *NFR1* and *NFR5* and their homoelogs is consistent with previous reports in soybean (Libault et al. 2010b, Libault et al. 2010c, Hayashi et al. 2012). As generally expected for receptor proteins, NFR1 and NFR5 had a lower expression and enrichment (perhaps due to their constitutive expression) compared to ENOD40a, ENOD40b, NSP2b, ERN1, and RIC1a as previously reported (Hayashi et al. 2012). Transcripts of *SYMRK/DMI2*, *CASTOR*, and *POLLUX/DMI1*were significantly enriched in MN but not in EN or LRs. The most likely reason for the lack of enrichment of these early signaling components in EN is their induction all over the root tissues during rhizobial infection resulting in high expression in both ABEN and EN. We determined the potential cell types where the MN-enriched genes might be active by comparing our data set to the nodule zone-specific expression

identified in *M. truncatula* (Limpens et al. 2013). It appears that DMI2 and DMI3 are enriched in the infected cells of fixation zone, while ERN1 and ERN2 are enriched early in infected cells. Therefore, continuous perception and signaling of nod factors must occur in these cells. Additional signaling components such as *NUP133* and *NUP85* were significantly enriched in MN, EN, and ELR. *CCaMK/DMI3* and *CYCLOPS* showed nodule-specific enrichment. Interestingly, *NSP2* (Glyma04g43090.1) did not show any significant enrichment in any of the lateral organs but its homeolog (Glyma13g02840.1) showed significant enrichment in YLR. We observed significant enrichment of *NSP1* (Glyma07g04430.1) in EN but not in MN, however, its homeolog (Glyma16g01020.1) showed significant enrichment in both EN and MN. This suggested that even though both *NSP1* and *NSP2* are activated during initial nodulation signaling, only *NSP1* is specific to nodule formation. In summary, specific paralogs of some nod factor signaling components appear to act in a stage-specific manner during soybean nodule formation.

2.3.3. Transcription factor families associated with nodule development

Analyses of TF expression profiles in different organs were not only consistent with results from previous studies (e.g. (Libault et al. 2009, Libault et al. 2010c, Boscari et al. 2013)), but also pointed to potential roles in specific stages of nodule development. For example, MYB-related TFs Glyma03g31980 and Glyma17g07330 are induced in soybean nodules at various stages of development in response to rhizobium inoculation (Libault et al. 2009). Our data indicated a reduction of Glyma03g31980 in MN, ELR, and LR while significant enrichment of Glyma17g07330 in MN. In agreement with nodule transcriptome atlas of soybean (Libault et al. 2010c), specific enrichment of *C2H2* family in the nodules was observed. In *M. truncatula* a member of *C2H2* family of TFs

described as *REGULATOR OF SYMBIOSOME DIFFERRENTIATION* (*RSD*) is shown to play a role in normal symbiosome differentiation during nodule development (Sinharoy et al. 2013). In our study two orthologs of *RSD*, Glyma07g16290 and Glyma18g40360 showed very high enrichment specifically in MN (>200 fold). The role of these two orthologs in soybean symbiosome differentiation is worth further understanding. We also detected nodule-enriched family members of DCL, B3, and PHD finger proteins, which have not been reported previously.

Very few TFs common between LRs and nodules were observed during this study, suggesting that developmental pathways associated with the formation of these organs are distinct from each other. The activity of TFs promoting the activities of plant hormones cytokinin and auxin was especially striking. While *ARR-B* was highly enriched in nodule tissues, *Aux/IAA*s, and *ARF*s were enriched in LRs. This is consistent with our observation that there is minimal auxin activity during nodule initiation in soybean and a higher cytokinin-lower auxin condition is required for nodule formation (Turner et al. 2013). It is interesting to note that shoot lateral organ formation and nodule formation require similar auxin-cytokinin balance in contrast to lateral root formation which is favored by a high auxin-low cytokinin condition (Wang et al. 2014b).

2.3.4. Is the soybean nodule a modified shoot axillary organ?

Enrichment of key signaling elements involved in shoot axillary bud formation in nodule tissues (Figure 2.19) strongly pointed towards co-option of shoot axillary meristem gene for nodule development. On the other hand, very little overlap between LRs and nodules in the expression of marker genes and transcription factors associated with LR formation was observed (Table 2.1 and Figure 2.9). As previously discussed by

others (Hirsch et al. 1997, Desbrosses and Stougaard 2011), conspicuous differences between LRs and nodules (especially determinate nodules) include the central vs. peripheral position of vascular bundles, persistent vs. non-persistent meristems, and the requirement of auxin vs. cytokinin action for initiation. Nodules and shoot axillary buds are similar with one exception that axillary buds are dormant meristem and determinate nodules have non-persistent meristem. One can imagine the determinate (and possibly indeterminate) nodule being an "arrested" axillary bud where the organs are fused together i.e. organ separation has been inhibited. For example, the peripheral vascular strands in nodules could be analogous to those from young leaves in a vegetative bud or floral organs in a flower bud. The morphological and gene expression similarities strongly suggest that such a possibility is likely to be beyond mere coincidence. Another somewhat distant similarity worth pointing out is the one between a pollen tube that transports foreign nuclei into the flower and releases them into the eggs vs. the infection thread that transport rhizobia into the infection zone and release them into infected cells. Interestingly, both fertilization and rhizobial colonization turn flowers and nodules respectively into sink tissues. To test the hypothesis that signaling and developmental pathways involved in nodule development could have evolved not only from lateral roots but also from shoot axillary bud formation, it would be crucial to explore these similarities through detailed phylogenetic analysis of genes expressed in axillary meristems, functional analysis of orthologs and gene family members. During evolution, nodules appear to have co-opted signaling and developmental pathways from both lateral roots and shoot axillary buds.

2.4. Experimental Procedures

2.4.1. Plant material and RNA isolation

Soybean (Williams82 cv.) seeds were surface-sterilized (Subramanian et al. 2008), sown in a mixture of autoclaved vermiculite and perlite (2:1 ratio). Seedlings were grown in a growth chamber (Conviron,Manitoba, Canada) at 25^0C and 16/8 h light cycle, and watered with nitrogen free plant nutrient solution (N^-PNS; (Subramanian et al. 2008) for both nodule and root tissue harvest. To harvest emerging (EN) and mature (MN) nodules and their controls, plants were inoculated with *B. japonicum* USDA 110 (grown in Vincent rich medium (Schwartz 1972) at 30 °C shaking at 200 rpm). For inoculation, the *B. japonicum* cells were resuspended in N^-PNS to a concentration of OD 0.08 OD at 600 nm (Bhuvaneswari et al. 1980) and applied to the roots. EN and control ABEN tissues were harvested at 5-7 dpi; MN and ABMN tissues were harvested at 14-16 dpi. ELR and YLR and their controls were harvested from uninoculated plants. All growth conditions including watering were identical between plants used for LR and nodule harvests except rhizobium inoculation. ELR and YLR tissues and controls were harvested 5-7 days after germination. For lateral organ harvests, plants were removed from vermiculite and perlite mixture, washed thoroughly in sterile water to remove the dirt, and organs were dissected using sterile scalpel blades under a dissection microscope. The dissected tissues were wipe dried on a sterile paper towel and collected in TRI reagent (Sigma-Aldrich, St.Louis, MO) in a pre-weighed 2 ml microcentrifuge tube on ice, and stored at -80 °C until RNA isolation. Plant growth, inoculations, and tissue harvests were performed in batches for each of the three biological replicates. Total RNA was isolated using the protocol recommended by the manufacturer with slight modifications

(Subramanian et al. 2008). The major exception was that a tissue lyser (QIAGEN, Hilden, Germany) was used for tissue homogenization rather than manual grinding with a pestle and mortar.

2.4.2. Transcriptome library preparation and sequencing

Directional RNA-seq libraries from each of the 24 RNA preparations was prepared using 5 µg of total RNA for each library (4 lateral organ tissues + 4 respective control tissues) x 3 biological replicates. RNA quality was checked using a Bioanalyzer (Agilent Technologies, Santa Clara, CA), and the library was prepared using ScriptSeq™ v2 RNA-Seq Library Preparation Kit (Epicentre, Madison, WI) according to the manufacturers recommendation. Briefly, PolyA RNA was isolated and reverse transcribed using random hexamer primers with a 5' tagging sequence. The 5' tagged cDNAs were then tagged at their 3'ends using terminal tagging oligo (TTO) with blocked 3' end. The resulting ditagged cDNA was linearly amplified using primers with adapter sequences (Pease and Sooknanan 2012). The amplified libraries were then sequenced on an Illumina HIseq2000 (single end, 50-nt read length) using 4 lanes and 6 samples per lane. The three replicates of each lateral organ tissue and its control were run on a single lane. At the end of the run, sequences were multiplexed into 24 different sequences files using the specific tag barcodes. All library construction, sequencing, and adapter trimming were performed at the Genomics Core Facility, University of Missouri, Columbia, MO.

2.4.3. Quality control of Raw Reads

The read quality was evaluated using fastqc -v0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The results suggested that a

portion of the reads might have few nt at the 5' end and ~7- to 10-nt at the 3' end with a poor phred score (<20). The sequences were filtered and trimmed for quality using prinseq-lite -v0.19.5 (Schmieder and Edwards 2011). We trimmed the reads at both the ends using the following parameters: minimum phred quality score = 20, a window size of 5, no Ns, and minimum post trimming length = 25. After trimming, the read quality was again checked using FastQC, which indicated a significant improvement in quality with the all nucleotides with phred quality score > 20.

2.4.4. Read Alignment and Assembly

RNA-seq read alignment and assembly were done based on the *tuxedo* pipeline (Trapnell et al. 2012). In brief, reads were aligned using *tophat* (v2.0.5) to the soybean reference genome (Gmax_v1.1_189, Phytozome v9.0). For differential expression analysis, each library was mapped individually, and for analysis of overall mapping statistics and coverage vs. FPKM analysis, all the libraries were mapped together. *tophat* was run with the following parameters: no transcriptome, genome read, no read and segment mismatches, no deletions and insertions, maximum intron length -5000, and library type- fr-second strand. The mapping was guided using gene models in Gmax_v1.1_189_gene.gff3 and was run without the –no-novel junctions option to identify the novel splice junctions. The mapped reads from each library were assembled into transcripts using *cufflinks* where transcriptome file was used as a guide and genome sequence file was used to enhance assembly. The maximum intron length was kept at 5000-nt, library type was fr -second strand and u-option were used to accurately weigh multiple mappings of the same read. The resulting assemblies were then merged using *cuffmerge* using both transcriptome and genome file as a guide to obtain a master gene

transfer format (gtf) reference which combines the novel splice junctions identified from this study with previously annotated ones. Differential expression analysis of the different lateral organs vs. respective controls was performed using *cuffdiff* (FDR<0.05). All of these three tools, *cufflinks*, *cuffmerge* and *cuffdiff* are available in the suite cufflinks (*cufflinks* v2.0.2). Instead of directly using the fold change output provided by *cuffdiff* we calculated "significant fold change", where the fold change was converted to zero if the stat test was not significant. This helped to filter out transcripts with a fold change, but the statistics test was either invalid or the result was not significant. Post Alignment Summaries were obtained by using *samtools*-v0.1.18 (Li et al. 2009) and *picard*-v1.98 (http://broadinstitute.github.io/picard/). *Samtools* was used to obtain the number of mapped reads, and unique reads for each library. CollectAlignmentSummaryMetrics tool in *picard* was used to obtain the mapping quality and genomic position annotation for each mapped base.

2.4.5. Identification and functional annotation of Novel Transcribed Regions

The protein coding sequences of all the 3334 transcripts designated as unannotated by *cufflinks* were extracted from the genome sequence using a custom Perl script. The sequences were used as a query in a BLASTX (*ncbi-blast-2.2.28+*) search against the non-redundant database of protein sequences (e-value of 0.00001, word size 3, gapopen 11, gapextend 3 and output in xml format). The xml output file was then fed into the *b2g4pipe_v2.5* (Conesa et al. 2005, Götz et al. 2008) to obtain gene ontology annotations. Out of the 3,334 query sequences, we were able to obtain GO annotations for 682. As described in the *results* subsection, we filtered the 3,334 novel transcribed region based on the FPKM≥3.64 for authenticity. The resulting 496 NTRs were evaluated

for differential gene expression in all lateral organ tissues and those with >2 fold change in at least one tissue were manually annotated for gene model. For this, we obtained genomic sequences 1000bp on either side of locus and manually determined and validated gene models using softberry (www.softberry.com).

2.4.6. Singular Enrichment (SEA) analysis

For global comparison of biological processes among lateral organs, transcripts with FPKM ≥1 and significant $\log_2$ fold change of ≥1 compared to their controls was used for SEA analysis using AgriGO-1.2 (Du et al. 2010). SEA analysis was also done for transcripts with tissue-specific enrichment, transcripts common between emerging organs (emerging nodule and lateral root) and transcripts enriched in organ-specific clusters. For SEA, soybean genome locus (Phytozome v1.1), Fisher statistical test method with Yekutieli (FDR under dependency), significance level =0.05, and complete GO was used.

2.4.7. Transcription Factors Analysis

Soybean transcription factor annotations from the Plant transcription factor database (PlantTFDB v3.0; http://planttfdb.cbi.pku.edu.cn/) (Jin et al. 2014) was used to obtain the TFs families and associated members. For each TF family, the number of members enriched in nodules tissues (EN and/or MN) or LR tissues (ELR and/or YLR) was counted. Only those transcripts with FPKM ≥1 and a $\log_2$ significant fold change of ≥1 compared to the corresponding control were counted. A number of transcripts for each TF family in EN and MN were combined (counted only once if present in both tissues) for enrichment in nodule tissue. Similarly, the number of transcripts enriched for each TF family in lateral root tissue was calculated by combining transcripts enriched in ELR and

YLR. The nodule and lateral root specific enrichment were compared by Fisher's Exact test ($P<0.05$).

2.4.8. Assessment of hormone biosynthesis and signaling

Peptide sequences of genes involved in biosynthesis and signaling of auxin, cytokinin, gibberellin, brassinolide, abscisic acid, ethylene, jasmonic acid, salicylic acid and strigolactone in *A. thaliana* were retrieved from The Arabidopsis Information Resource (TAIR,https://www.arabidopsis.org/) (Lamesch et al. 2012). These peptide sequences were used as a query in a TBLASTN search against the soybean genome in LegumeIP (Li et al. 2012). Soybean orthologs obtained in this way were checked for tissue-specific gene expression pattern in our transcriptome data. Heatmaps were generated for each biosynthesis and signaling ortholog using significant log_2fold change value, without dendrogram and data scaling using Heatmap.2 from gplots package (Warnes et al. 2014) in R (2.12.1).

References

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

Aida, M. and Tasaka, M. (2006) Morphogenesis and patterning at the organ boundaries in the higher plant shoot apex. *Plant Mol Biol*, 60, 915-928.

Allen, O.N. and Allen, E.K. (1940) Response of the Peanut Plant to Inoculation with Rhizobia, with Special Reference to Morphological Development of the Nodules. *Botanical Gazette* 102, 121.

Bao, F., Shen, J., Brady, S.R., Muday, G.K., Asami, T. and Yang, Z. (2004) Brassinosteroids Interact with Auxin to Promote Lateral Root Development in Arabidopsis. *Plant Physiology*, 134, 1624-1631.

Barton, M.K. (2010) Twenty years on: The inner workings of the shoot apical meristem, a developmental dynamo. *Developmental Biology*, 341, 95-113.

Benedito, V.A., Torres-Jerez, I., Murray, J.D., Andriankaja, A., Allen, S., Kakar, K., Wandrey, M., Verdier, J., Zuber, H., Ott, T., Moreau, S., Niebel, A., Frickey, T., Weiller, G., He, J., Dai, X., Zhao, P.X., Tang, Y. and Udvardi, M.K. (2008) A gene expression atlas of the model legume Medicago truncatula. *The Plant journal : for cell and molecular biology*, 55, 504-513.

Bennett, T. and Leyser, O. (2014) Strigolactone signalling: standing on the shoulders of DWARFs. *Current opinion in plant biology*, 22, 7-13.

Benson, D.R. and Silvester, W.B. (1993) Biology of Frankia strains, actinomycete symbionts of actinorhizal plants. *Microbiological reviews*, 57, 293-319.

Berova, M. and Zlatev, Z. (2000) Physiological response and yield of paclobutrazol treated tomato plants (Lycopersicon esculentum Mill.). *Plant Growth Regulation*, 30, 117-123.

Bhuvaneswari, T.V., Turgeon, B.G. and Bauer, W.D. (1980) Early Events in the Infection of Soybean (Glycine max L. Merr) by Rhizobium japonicum: I. LOCALIZATION OF INFECTIBLE ROOT CELLS. *Plant Physiology*, 66, 1027-1031.

Bielach, A., Podlešáková, K., Marhavý, P., Duclercq, J., Cuesta, C., Müller, B., Grunewald, W., Tarkowski, P. and Benková, E. (2012) Spatiotemporal Regulation of Lateral Root Organogenesis in Arabidopsis by Cytokinin. *The Plant Cell Online*, 24, 3967-3981.

Bond, L. (1948) Origin and Development Morphology of Root Nodules of Pisum sativum. *Botanical Gazette*, 109, 411.

Boot, K.J.M., van Brussel, A.A.N., Tak, T., Spaink, H.P. and Kijne, J.W. (1999) Lipochitin Oligosaccharides from Rhizobium leguminosarum bv. viciae Reduce Auxin Transport Capacity in Vicia sativa subsp. nigra Roots. *Molecular Plant-Microbe Interactions*, 12, 839-844.

Borghi, L., Bureau, M. and Simon, R. (2007) Arabidopsis JAGGED LATERAL ORGANS is expressed in boundaries and coordinates KNOX and PIN activity. *The Plant cell*, 19, 1795-1808.

Boscari, A., Del Giudice, J., Ferrarini, A., Venturini, L., Zaffini, A.L., Delledonne, M. and Puppo, A. (2013) Expression dynamics of the Medicago truncatula transcriptome during the symbiotic interaction with Sinorhizobium meliloti: which role for nitric oxide? *Plant Physiol*, 161, 425-439.

Brechenmacher, L., Kim, M.-Y., Benitez, M., Li, M., Joshi, T., Calla, B., Lee, M.P., Libault, M., Vodkin, L.O., Xu, D., Lee, S.-H., Clough, S.J. and Stacey, G. (2008) Transcription Profiling of Soybean Nodulation by Bradyrhizobium japonicum. *Molecular Plant-Microbe Interactions*, 21, 631-645.

Canales, J., Moyano, T.C., Villarroel, E. and Gutiérrez, R.A. (2014) Systems analysis of transcriptome data provides new hypotheses about Arabidopsis root response to nitrate treatments. *Frontiers in Plant Science*, 5, 22.

Carroll, B.J., McNeil, D.L. and Gresshoff, P.M. (1985) A Supernodulation and Nitrate-Tolerant Symbiotic (nts) Soybean Mutant. *Plant Physiol*, 78, 34-40.

Casimiro, I., Marchant, A., Bhalerao, R.P., Beeckman, T., Dhooge, S., Swarup, R., Graham, N., Inze, D., Sandberg, G., Casero, P.J. and Bennett, M. (2001) Auxin transport promotes Arabidopsis lateral root initiation. *The Plant cell*, 13, 843-852.

Castaings, L., Camargo, A., Pocholle, D., Gaudon, V., Texier, Y., Boutet-Mercey, S., Taconnat, L., Renou, J.P., Daniel-Vedele, F., Fernandez, E., Meyer, C. and Krapp, A. (2009) The nodule inception-like protein 7 modulates nitrate sensing and metabolism in Arabidopsis. *The Plant journal : for cell and molecular biology*, 57, 426-435.

Combier, J.-P., Frugier, F., de Billy, F., Boualem, A., El-Yahyaoui, F., Moreau, S., Vernié, T., Ott, T., Gamas, P., Crespi, M. and Niebel, A. (2006) MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in Medicago truncatula. *Genes & Development*, 20, 3084-3088.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M. and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics (Oxford, England)*, 21, 3674-3676.

Couzigou, J.M., Zhukov, V., Mondy, S., Abu el Heba, G., Cosson, V., Ellis, T.H., Ambrose, M., Wen, J., Tadege, M., Tikhonovich, I., Mysore, K.S., Putterill, J., Hofer, J., Borisov, A.Y. and Ratet, P. (2012) NODULE ROOT and COCHLEATA maintain nodule development and are legume orthologs of Arabidopsis BLADE-ON-PETIOLE genes. *Plant Cell*, 24, 4498-4510.

Cvrčková, F., Bezvoda, R. and Žárský, V. (2010) Computational identification of root hair-specific genes in Arabidopsis. *Plant Signaling & Behavior*, 5, 1407-1418.

De Rybel, B., Vassileva, V., Parizot, B., Demeulenaere, M., Grunewald, W., Audenaert, D., Van Campenhout, J., Overvoorde, P., Jansen, L., Vanneste, S., Möller, B., Wilson, M., Holman, T., Van Isterdael, G., Brunoud, G., Vuylsteke, M., Vernoux, T., De Veylder, L., Inzé, D., Weijers, D., Bennett, M.J. and Beeckman, T. (2010) A Novel Aux/IAA28 Signaling Cascade Activates GATA23-Dependent Specification of Lateral Root Founder Cell Identity. *Current Biology*, 20, 1697-1706.

De Smet, I. (2012) Lateral root initiation: one step at a time. *New Phytologist*, 193, 867-873.

Desbrosses, Guilhem J. and Stougaard, J. (2011) Root Nodulation: A Paradigm for How Plant-Microbe Symbiosis Influences Host Developmental Pathways. *Cell Host & Microbe*, 10, 348-358.

Ding, Y., Kalo, P., Yendrek, C., Sun, J., Liang, Y., Marsh, J.F., Harris, J.M. and Oldroyd, G.E.D. (2008) Abscisic Acid Coordinates Nod Factor and Cytokinin Signaling during the Regulation of Nodulation in Medicago truncatula. *The Plant Cell Online*, 20, 2681-2695.

Du, Z., Zhou, X., Ling, Y., Zhang, Z. and Su, Z. (2010) agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Research*, 38, W64-W70.

Ehlting, J., Chowrira, S.G., Mattheus, N., Aeschliman, D.S., Arimura, G. and Bohlmann, J. (2008) Comparative transcriptome analysis of Arabidopsis thaliana infested by diamond back moth (Plutella xylostella) larvae reveals signatures of stress response, secondary metabolism, and signalling. *BMC genomics*, 9, 154.

Ferguson, B.J. and Mathesius, U. (2014) Phytohormone regulation of legume-rhizobia interactions. *Journal of chemical ecology*, 40, 770-790.

Ferguson, B.J., Ross, J.J. and Reid, J.B. (2005) Nodulation Phenotypes of Gibberellin and Brassinosteroid Mutants of Pea. *Plant Physiology*, 138, 2396-2405.

Foo, E. and Davies, N.W. (2011) Strigolactones promote nodulation in pea. *Planta*, 234, 1073-1081.

Ghorbani, S., Lin, Y.-C., Parizot, B., Fernandez, A., Njo, M.F., Van de Peer, Y., Beeckman, T. and Hilson, P. (2015) Expanding the repertoire of secretory peptides controlling root development with comparative genome analysis and functional assays. *Journal of Experimental Botany*.

Goh, T., Kasahara, H., Mimura, T., Kamiya, Y. and Fukaki, H. (2012) Multiple AUX/IAA–ARF modules regulate lateral root formation: the role of Arabidopsis SHY2/IAA3-mediated auxin signalling. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 1461-1468.

Gonzalez-Rizzo, S., Crespi, M. and Frugier, F. (2006) The Medicago truncatula CRE1 Cytokinin Receptor Regulates Lateral Root Development and Early Symbiotic Interaction with Sinorhizobium meliloti. *The Plant Cell Online*, 18, 2680-2693.

Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talón, M., Dopazo, J. and Conesa, A. (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, 36, 3420-3435.

Gou, J., Strauss, S.H., Tsai, C.J., Fang, K., Chen, Y., Jiang, X. and Busov, V.B. (2010) Gibberellins Regulate Lateral Root Formation in Populus through Interactions with Auxin and Other Hormones. *The Plant Cell Online*, 22, 623-639.

Greb, T., Clarenz, O., Schafer, E., Muller, D., Herrero, R., Schmitz, G. and Theres, K. (2003) Molecular analysis of the LATERAL SUPPRESSOR gene in Arabidopsis reveals a conserved control mechanism for axillary meristem formation. *Genes Dev*, 17, 1175-1187.

Gualtieri, G. and Bisseling, T. (2000) The evolution of nodulation. *Plant Mol Biol*, 42, 181-194.

Gutiérrez, R.A., Lejay, L.V., Dean, A., Chiaromonte, F., Shasha, D.E. and Coruzzi, G.M. (2007) Qualitative network models and genome-wide expression data define carbon/nitrogen-responsive molecular machines in Arabidopsis. *Genome Biology*, 8, R7-R7.

Ha, C.M., Kim, G.-T., Kim, B.C., Jun, J.H., Soh, M.S., Ueno, Y., Machida, Y., Tsukaya, H. and Nam, H.G. (2003) The BLADE-ON-PETIOLE 1 gene controls leaf pattern formation through the modulation of meristematic activity in Arabidopsis. *Development (Cambridge, England)*, 130, 161-172.

Hare, P.D. and van Staden, J. (1994) Cytokinin oxidase: Biochemical features and physiological significance. *Physiologia Plantarum*, 91, 128-136.

Hayashi, S., Reid, D.E., Lorenc, M.T., Stiller, J., Edwards, D., Gresshoff, P.M. and Ferguson, B.J. (2012) Transient Nod factor-dependent gene expression in the nodulation-competent zone of soybean (Glycine max [L.] Merr.) roots. *Plant Biotechnology Journal*, 10, 995-1010.

Heckmann, A.B., Lombardo, F., Miwa, H., Perry, J.A., Bunnewell, S., Parniske, M., Wang, T.L. and Downie, J.A. (2006) Lotus japonicus Nodulation Requires Two GRAS Domain Regulators, One of Which Is Functionally Conserved in a Non-Legume. *Plant Physiology*, 142, 1739-1750.

Heckmann, A.B., Sandal, N., Bek, A.S., Madsen, L.H., Jurkiewicz, A., Nielsen, M.W., Tirichine, L. and Stougaard, J. (2011) Cytokinin induction of root nodule

primordia in Lotus japonicus is regulated by a mechanism operating in the root cortex. *Molecular plant-microbe interactions : MPMI*, 24, 1385-1395.

Herrbach, V., Remblière, C., Gough, C. and Bensmihen, S. (2014) Lateral root formation and patterning in Medicago truncatula. *Journal of Plant Physiology*, 171, 301-310.

Hirsch, A.M. (1992) Developmental Biology of Legume Nodulation. *New Phytologist*, 122, 211-237.

Hirsch, A.M., Larue, T.A. and Doyle, J. (1997) Is the Legume Nodule a Modified Root or Stem or an Organ sui generis? *Critical Reviews in Plant Sciences*, 16, 361-392.

Hossain, M.S., Joshi, T. and Stacey, G. (2015) System approaches to study root hairs as a single cell plant model: current status and future perspectives. *Frontiers in Plant Science*, 6, 363.

Hubbard, K.E., Nishimura, N., Hitomi, K., Getzoff, E.D. and Schroeder, J.I. (2010) Early abscisic acid signal transduction mechanisms: newly discovered components and newly emerging questions. *Genes Dev*, 24, 1695-1708.

Hunter, W.J. (1993) Ethylene Production by Root Nodules and Effect of Ethylene on Nodulation in Glycine max. *Applied and Environmental Microbiology*, 59, 1947-1950.

Hunter, W.J. (2001) Influence of Root-Applied Epibrassinolide and Carbenoxolone on the Nodulation and Growth of Soybean (Glycine max L.) Seedlings. *Journal of Agronomy and Crop Science*, 186, 217-221.

Hutchison, C.E. and Kieber, J.J. (2002) Cytokinin Signaling in Arabidopsis. *The Plant Cell Online*, 14, S47-S59.

Hwang, I. and Sheen, J. (2001) Two-component circuitry in Arabidopsis cytokinin signal transduction. *Nature*, 413, 383-389.

Inoue, T., Higuchi, M., Hashimoto, Y., Seki, M., Kobayashi, M., Kato, T., Tabata, S., Shinozaki, K. and Kakimoto, T. (2001) Identification of CRE1 as a cytokinin receptor from Arabidopsis. *Nature*, 409, 1060-1063.

Jin, J., Zhang, H., Kong, L., Gao, G. and Luo, J. (2014) PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res*, 42, D1182-1187.

Jones, R. and Schreiber, B.N. (1997) Role and function of cytokinin oxidase in plants. *Plant Growth Regulation*, 23, 123-134.

Kamínek, M., Motyka, V. and Vaňková, R. (1997) Regulation of cytokinin content in plant cells. *Physiologia Plantarum*, 101, 689-700.

Kanrar, S., Onguka, O. and Smith, H.M. (2006) Arabidopsis inflorescence architecture requires the activities of KNOX-BELL homeodomain heterodimers. *Planta*, 224, 1163-1173.

Keller, T., Abbott, J., Moritz, T. and Doerner, P. (2006) Arabidopsis REGULATOR OF AXILLARY MERISTEMS1 Controls a Leaf Axil Stem Cell Niche and Modulates Vegetative Development. *The Plant Cell Online*, 18, 598-611.

Koltai, H. (2014) Receptors, repressors, PINs: a playground for strigolactone signaling. *Trends in Plant Science*, 19, 727-733.

Konishi, M. and Yanagisawa, S. (2013) Arabidopsis NIN-like transcription factors have a central role in nitrate signalling. *Nat Commun*, 4, 1617.

Kouchi, H., Shimomura, K., Hata, S., Hirota, A., Wu, G.J., Kumagai, H., Tajima, S., Suganuma, N., Suzuki, A., Aoki, T., Hayashi, M., Yokoyama, T., Ohyama, T., Asamizu, E., Kuwata, C., Shibata, D. and Tabata, S. (2004) Large-scale analysis of gene expression profiles during early stages of root nodule formation in a model legume, Lotus japonicus. *DNA research : an international journal for rapid publication of reports on genes and genomes*, 11, 263-274.

Koyama, T., Furutani, M., Tasaka, M. and Ohme-Takagi, M. (2007) TCP transcription factors control the morphology of shoot lateral organs via negative regulation of the expression of boundary-specific genes in Arabidopsis. *The Plant cell*, 19, 473-484.

Kurakawa, T., Ueda, N., Maekawa, M., Kobayashi, K., Kojima, M., Nagato, Y., Sakakibara, H. and Kyozuka, J. (2007) Direct control of shoot meristem activity by a cytokinin-activating enzyme. *Nature*, 445, 652-655.

Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M., Karthikeyan, A.S., Lee, C.H., Nelson, W.D., Ploetz, L., Singh, S., Wensel, A. and Huala, E. (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res*, 40, D1202-1210.

Lan, P., Li, W., Lin, W.D., Santi, S. and Schmidt, W. (2013) Mapping gene activity of Arabidopsis root hairs. *Genome Biol*, 14, R67.

Laplaze, L., Benkova, E., Casimiro, I., Maes, L., Vanneste, S., Swarup, R., Weijers, D., Calvo, V., Parizot, B., Herrera-Rodriguez, M.B., Offringa, R., Graham, N., Doumas, P., Friml, J., Bogusz, D., Beeckman, T. and Bennett, M. (2007) Cytokinins Act Directly on Lateral Root Founder Cells to Inhibit Root Initiation. *The Plant Cell Online*, 19, 3889-3900.

Lau, S., Jürgens, G. and De Smet, I. (2008) The Evolving Complexity of the Auxin Pathway. *The Plant Cell Online*, 20, 1738-1746.

Lavenus, J., Goh, T., Roberts, I., Guyomarc'h, S., Lucas, M., De Smet, I., Fukaki, H., Beeckman, T., Bennett, M. and Laplaze, L. (2013) Lateral root development in Arabidopsis: fifty shades of auxin. *Trends in plant science*, 18, 450-458.

Lee, K.H. and LaRue, T.A. (1992) Exogenous Ethylene Inhibits Nodulation of Pisum sativum L. cv Sparkle. *Plant Physiology*, 100, 1759-1763.

Lewis, D.R., Negi, S., Sukumar, P. and Muday, G.K. (2011) Ethylene inhibits lateral root development, increases IAA transport and expression of PIN3 and PIN7 auxin efflux carriers. *Development (Cambridge, England)*, 138, 3485-3495.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25, 2078-2079.

Li, J., Dai, X., Liu, T. and Zhao, P.X. (2012) LegumeIP: an integrative database for comparative genomics and transcriptomics of model legumes. *Nucleic Acids Research*, 40, D1221-D1229.

Li, X., Mo, X., Shou, H. and Wu, P. (2006) Cytokinin-mediated cell cycling arrest of pericycle founder cells in lateral root initiation of Arabidopsis. *Plant & cell physiology*, 47, 1112-1123.

Libault, M., Farmer, A., Brechenmacher, L., Drnevich, J., Langley, R.J., Bilgin, D.D., Radwan, O., Neece, D.J., Clough, S.J., May, G.D. and Stacey, G. (2010b) Complete Transcriptome of the Soybean Root Hair Cell, a Single-Cell Model, and Its Alteration in Response to Bradyrhizobium japonicum Infection. *Plant Physiology*, 152, 541-552.

Libault, M., Farmer, A., Joshi, T., Takahashi, K., Langley, R.J., Franklin, L.D., He, J., Xu, D., May, G. and Stacey, G. (2010c) An integrated transcriptome atlas of the crop model Glycine max, and its use in comparative analyses in plants. *The Plant Journal*, 63, 86-99.

Libault, M., Joshi, T., Takahashi, K., Hurley-Sommer, A., Puricelli, K., Blake, S., Finger, R.E., Taylor, C.G., Xu, D., Nguyen, H.T. and Stacey, G. (2009) Large-Scale Analysis of Putative Soybean Regulatory Gene Expression Identifies a Myb Gene Involved in Soybean Nodule Development. *Plant Physiology*, 151, 1207-1220.

Libault, M., Zhang, X.-C., Govindarajulu, M., Qiu, J., Ong, Y.T., Brechenmacher, L., Berg, R.H., Hurley-Sommer, A., Taylor, C.G. and Stacey, G. (2010a) A member of the highly conserved FWL (tomato FW2.2-like) gene family is essential for soybean nodule organogenesis. *The Plant Journal*, 62, 852-864.

Ligero, F., Caba, J.M., Lluch, C. and Olivares, J. (1991) Nitrate Inhibition of Nodulation Can Be Overcome by the Ethylene Inhibitor Aminoethoxyvinylglycine. *Plant Physiology*, 97, 1221-1225.

Limpens, E., Moling, S., Hooiveld, G., Pereira, P.A., Bisseling, T., Becker, J.D. and Kuster, H. (2013) cell- and tissue-specific transcriptome analyses of Medicago truncatula root nodules. *PLoS One*, 8, e64377.

Liu, J., Novero, M., Charnikhova, T., Ferrandino, A., Schubert, A., Ruyter-Spira, C., Bonfante, P., Lovisolo, C., Bouwmeester, H.J. and Cardinale, F. (2013) CAROTENOID CLEAVAGE DIOXYGENASE 7 modulates plant growth, reproduction, senescence, and determinate nodulation in the model legume Lotus japonicus. *Journal of Experimental Botany*, 64, 1967-1981.

Liu, Y., Zhou, J. and White, K.P. (2014) RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics (Oxford, England)*, 30, 301-304.

Lohar, D.P., Schaff, J.E., Laskey, J.G., Kieber, J.J., Bilyeu, K.D. and Bird, D.M. (2004) Cytokinins play opposite roles in lateral root formation, and nematode and Rhizobial symbioses. *The Plant Journal*, 38, 203-214.

Long, J. and Barton, M.K. (2000) Initiation of Axillary and Floral Meristems in Arabidopsis. *Developmental Biology*, 218, 341-353.

Ma, S., Gong, Q. and Bohnert, H.J. (2007) An Arabidopsis gene network based on the graphical Gaussian model. *Genome Research*, 17, 1614-1625.

Madsen, L.H., Tirichine, L., Jurkiewicz, A., Sullivan, J.T., Heckmann, A.B., Bek, A.S., Ronson, C.W., James, E.K. and Stougaard, J. (2010) The molecular network

governing nodule organogenesis and infection in the model legume Lotus japonicus. *Nat Commun*, 1, 1-12.

Maekawa, T., Maekawa-Yoshikawa, M., Takeda, N., Imaizumi-Anraku, H., Murooka, Y. and Hayashi, M. (2009) Gibberellin controls the nodulation signaling pathway in Lotus japonicus. *The Plant Journal*, 58, 183-194.

Malamy, J.E. and Benfey, P.N. (1997) Organization and cell differentiation in lateral roots of Arabidopsis thaliana. *Development*, 124, 33-44.

Mano, Y. and Nemoto, K. (2012) The pathway of auxin biosynthesis in plants. *J Exp Bot*, 63, 2853-2872.

Marvel, D.J., Torrey, J.G. and Ausubel, F.M. (1987) Rhizobium symbiotic genes required for nodulation of legume and nonlegume hosts. *Proceedings of the National Academy of Sciences of the United States of America*, 84, 1319-1323.

Mashiguchi, K., Tanaka, K., Sakai, T., Sugawara, S., Kawaide, H., Natsume, M., Hanada, A., Yaeno, T., Shirasu, K., Yao, H., McSteen, P., Zhao, Y., Hayashi, K.-i., Kamiya, Y. and Kasahara, H. (2011) The main auxin biosynthesis pathway in Arabidopsis. *Proceedings of the National Academy of Sciences*, 108, 18512-18517.

Mathesius, U. (2003) Conservation and divergence of signalling pathways between roots and soil microbes – the Rhizobium-legume symbiosis compared to the development of lateral roots, mycorrhizal interactions and nematode-induced galls. *Plant and Soil*, 255, 105-119.

Mathesius, U., Schlaman, H.R., Spaink, H.P., Of Sautter, C., Rolfe, B.G. and Djordjevic, M.A. (1998) Auxin transport inhibition precedes root nodule formation in white clover roots and is regulated by flavonoids and derivatives of chitin oligosaccharides. *The Plant journal : for cell and molecular biology*, 14, 23-34.

Motyka, V., Vaňková, R., Čapková, V., Petrášek, J., Kamínek, M. and Schmülling, T. (2003) Cytokinin-induced upregulation of cytokinin oxidase activity in tobacco includes changes in enzyme glycosylation and secretion. *Physiologia Plantarum*, 117, 11-21.

Müller, D., Schmitz, G. and Theres, K. (2006) Blind Homologous R2R3 Myb Genes Control the Pattern of Lateral Meristem Initiation in Arabidopsis. *The Plant Cell Online*, 18, 586-597.

Murray, J.D., Karas, B.J., Sato, S., Tabata, S., Amyot, L. and Szczyglowski, K. (2007) A cytokinin perception mutant colonized by Rhizobium in the absence of nodule organogenesis. *Science (New York, N.Y.)*, 315, 101-104.

Nakagawa, T. and Kawaguchi, M. (2006) Shoot-applied MeJA Suppresses Root Nodulation in Lotus japonicus. *Plant and Cell Physiology*, 47, 176-180.

Negi, S., Ivanchenko, M.G. and Muday, G.K. (2008) Ethylene regulates lateral root formation and auxin transport in Arabidopsis thaliana. *The Plant Journal*, 55, 175-187.

Negi, S., Sukumar, P., Liu, X., Cohen, J.D. and Muday, G.K. (2010) Genetic dissection of the role of ethylene in regulating auxin-dependent lateral and adventitious root formation in tomato. *The Plant journal : for cell and molecular biology*, 61, 3-15.

Nishimura, R., Hayashi, M., Wu, G.J., Kouchi, H., Imaizumi-Anraku, H., Murakami, Y., Kawasaki, S., Akao, S., Ohmori, M., Nagasawa, M., Harada, K. and Kawaguchi, M. (2002) HAR1 mediates systemic regulation of symbiotic organ development. *Nature*, 420, 426-429.

Nizampatnam, N.R., Schreier, S.J., Damodaran, S., Adhikari, S. and Subramanian, S. (2015) microRNA160 dictates stage-specific auxin and cytokinin sensitivities and directs soybean nodule development. *The Plant Journal*, 84, 140-153.

Nukui, N., Ezura, H., Yuhashi, K.-I., Yasuta, T. and Minamisawa, K. (2000) Effects of Ethylene Precursor and Inhibitors for Ethylene Biosynthesis and Perception on Nodulation in Lotus japonicus and Macroptilium atropurpureum. *Plant and Cell Physiology*, 41, 893-897.

Nutman, P.S. (1948) Physiological Studies on Nodule Formation. *Annals of Botany*, 12, 81-96.

Okushima, Y., Fukaki, H., Onoda, M., Theologis, A. and Tasaka, M. (2007) ARF7 and ARF19 regulate lateral root formation via direct activation of LBD/ASL genes in Arabidopsis. *The Plant cell*, 19, 118-130.

Oliveros, J.C. (2007) VENNY. An interactive tool for comparing lists with Venn Diagrams. http://bioinfogp.cnb.csic.es/tools/venny/index.html.

Ott, T., van Dongen, J.T., Gu¨nther, C., Krusell, L., Desbrosses, G., Vigeolas, H., Bock, V., Czechowski, T., Geigenberger, P. and Udvardi, M.K. (2005) Symbiotic Leghemoglobins Are Crucial for Nitrogen Fixation in Legume Root Nodules but Not for General Plant Growth and Development. *Current Biology*, 15, 531-535.

Palenchar, P.M., Kouranov, A., Lejay, L.V. and Coruzzi, G.M. (2004) Genome-wide patterns of carbon and nitrogen regulation of gene expression validate the combined carbon and nitrogen (CN)-signaling hypothesis in plants. *Genome Biology*, 5, R91-R91.

Papadopoulou, K., Roussis, A. and Katinakis, P. (1996) Phaseolus ENOD40 is involved in symbiotic and non-symbiotic organogenetic processes: expression during nodule and lateral root development. *Plant Mol Biol*, 30, 403-417.

Parizot, B., De Rybel, B. and Beeckman, T. (2010) VisuaLRTC: A New View on Lateral Root Initiation by Combining Specific Transcriptome Data Sets. *Plant Physiology*, 153, 34-40.

Pawlowski, K. and Bisseling, T. (1996) Rhizobial and Actinorhizal Symbioses: What Are the Shared Features? *The Plant cell*, 8, 1899-1913.

Pawlowski, K. and Demchenko, K.N. (2012) The diversity of actinorhizal symbiosis. *Protoplasma*, 249, 967-979.

Pease, J. and Sooknanan, R. (2012) A rapid, directional RNA-seq library preparation workflow for Illumina[reg] sequencing. *Nat Meth*, 9.

Peret, B., De Rybel, B., Casimiro, I., Benkova, E., Swarup, R., Laplaze, L., Beeckman, T. and Bennett, M.J. (2009) Arabidopsis lateral root development: an emerging story. *Trends in plant science*, 14, 399-408.

Peters, N.K. and Crist-Estes, D.K. (1989) Nodule Formation Is Stimulated by the Ethylene Inhibitor Aminoethoxyvinylglycine. *Plant Physiology*, 91, 690-693.

Plet, J., Wasson, A., Ariel, F., Le Signor, C., Baker, D., Mathesius, U., Crespi, M. and Frugier, F. (2011) MtCRE1-dependent cytokinin signaling integrates bacterial and plant cues to coordinate symbiotic nodule organogenesis in Medicago truncatula. *The Plant Journal*, 65, 622-633.

Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V. and Quackenbush, J. (2003) TM4: a free, open-source system for microarray data management and analysis. *BioTechniques*, 34, 374-378.

Sato, T., Fujikake, H., Ohtake, N., Sueyoshi, K., Takahashi, T., Sato, A. and Ohyama, T. (2002) Effect of exogenous salicylic acid supply on nodule formation of hypernodulating mutant and wild type of soybean. *Soil Science and Plant Nutrition*, 48, 413-420.

Schmidt, J.S., Harper, J.E., Hoffman, T.K. and Bent, A.F. (1999) Regulation of Soybean Nodulation Independent of Ethylene Signaling. *Plant Physiology*, 119, 951-960.

Schmieder, R. and Edwards, R. (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics (Oxford, England)*, 27, 863-864.

Schmitz, G. and Theres, K. (2005) Shoot and inflorescence branching. *Curr Opin Plant Biol*, 8, 506-511.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., Xu, D., Hellsten, U., May, G.D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M.K., Sandhu, D., Valliyodan, B.,

Lindquist, E., Peto, M., Grant, D., Shu, S., Goodstein, D., Barry, K., Futrell-Griggs, M., Abernathy, B., Du, J., Tian, Z., Zhu, L., Gill, N., Joshi, T., Libault, M., Sethuraman, A., Zhang, X.-C., Shinozaki, K., Nguyen, H.T., Wing, R.A., Cregan, P., Specht, J., Grimwood, J., Rokhsar, D., Stacey, G., Shoemaker, R.C. and Jackson, S.A. (2010) Genome sequence of the palaeopolyploid soybean. *Nature*, 463, 178-183.

Schwartz, W. (1972) J. M. Vincent, A Manual for the Practical Study of the Root-Nodule Bacteria (IBP Handbuch No. 15 des International Biology Program, London). XI u. 164 S., 10 Abb., 17 Tab., 7 Taf. Oxford-Edinburgh 1970: Blackwell Scientific Publ., 45 s. *Zeitschrift für allgemeine Mikrobiologie*, 12, 440-440.

Shimada, Y., Goda, H., Nakamura, A., Takatsuto, S., Fujioka, S. and Yoshida, S. (2003) Organ-specific expression of brassinosteroid-biosynthetic genes and distribution of endogenous brassinosteroids in Arabidopsis. *Plant Physiol*, 131, 287-297.

Shuai, B., Reynaga-Pena, C.G. and Springer, P.S. (2002) The lateral organ boundaries gene defines a novel, plant-specific gene family. *Plant Physiol*, 129, 747-761.

Sinharoy, S., Torres-Jerez, I., Bandyopadhyay, K., Kereszt, A., Pislariu, C.I., Nakashima, J., Benedito, V.A., Kondorosi, E. and Udvardi, M.K. (2013) The C2H2 Transcription Factor REGULATOR OF SYMBIOSOME DIFFERENTIATION Represses Transcription of the Secretory Pathway Gene VAMP721a and Promotes Symbiosome Development in Medicago truncatula. *The Plant Cell Online*, 25, 3584-3601.

Smith, D.L. and Fedoroff, N.V. (1995) LRP1, a gene expressed in lateral and adventitious root primordia of arabidopsis. *The Plant Cell Online*, 7, 735-745.

Soto, M.J., Fernández-Aparicio, M., Castellanos-Morales, V., García-Garrido, J.M., Ocampo, J.A., Delgado, M.J. and Vierheilig, H. (2010) First indications for the involvement of strigolactones on nodule formation in alfalfa (Medicago sativa). *Soil Biology and Biochemistry*, 42, 383-385.

Sozzani, R. and Iyer-Pascuzzi, A. (2014) Postembryonic control of root meristem growth and development. *Current Opinion in Plant Biology*, 17, 7-12.

Stacey, G., McAlvin, C.B., Kim, S.Y., Olivares, J. and Soto, M.J. (2006) Effects of endogenous salicylic acid on nodulation in the model legumes Lotus japonicus and Medicago truncatula. *Plant Physiol*, 141, 1473-1481.

Subramanian, S., Fu, Y., Sunkar, R., Barbazuk, W.B., Zhu, J.-K. and Yu, O. (2008) Novel and nodulation-regulated microRNAs in soybean roots. *BMC genomics*, 9, 160.

Subramanian, S., Stacey, G. and Yu, O. (2007) Distinct, crucial roles of flavonoids during legume nodulation. *Trends in plant science*, 12, 282-285.

Suganuma, N., Yamauchi, H. and Yamamoto, K. (1995) Enhanced production of ethylene by soybean roots after inoculation with Bradyrhizobium japonicum. *Plant Science*, 111, 163-168.

Sun, T. (2010) Gibberellin-GID1-DELLA: A Pivotal Regulatory Module for Plant Growth and Development. *Plant Physiology*, 154, 567-570.

Suzaki, T., Yano, K., Ito, M., Umehara, Y., Suganuma, N. and Kawaguchi, M. (2012) Positive and negative regulation of cortical cell division during root nodule development in Lotus japonicus is accompanied by auxin response. *Development (Cambridge, England)*, 139, 3997-4006.

Suzuki, A., Akune, M., Kogiso, M., Imagama, Y., Osuki, K.-i., Uchiumi, T., Higashi, S., Han, S.-Y., Yoshida, S., Asami, T. and Abe, M. (2004) Control of Nodule Number by the Phytohormone Abscisic Acid in the Roots of Two Leguminous Species. *Plant and Cell Physiology*, 45, 914-922.

Swarup, K., Benkova, E., Swarup, R., Casimiro, I., Peret, B., Yang, Y., Parry, G., Nielsen, E., De Smet, I., Vanneste, S., Levesque, M.P., Carrier, D., James, N., Calvo, V., Ljung, K., Kramer, E., Roberts, R., Graham, N., Marillonnet, S., Patel, K., Jones, J.D.G., Taylor, C.G., Schachtman, D.P., May, S., Sandberg, G., Benfey, P., Friml, J., Kerr, I., Beeckman, T., Laplaze, L. and Bennett, M.J. (2008) The auxin influx carrier LAX3 promotes lateral root emergence. *Nat Cell Biol*, 10, 946-954.

Takanashi, K., Sugiyama, A. and Yazaki, K. (2011) Involvement of auxin distribution in root nodule development of Lotus japonicus. *Planta*, 234, 73-81.

Takanashi, K., Takahashi, H., Sakurai, N., Sugiyama, A., Suzuki, H., Shibata, D., Nakazono, M. and Yazaki, K. (2012) Tissue-specific transcriptome analysis in nodules of Lotus japonicus. *Molecular plant-microbe interactions : MPMI*, 25, 869-876.

Takei, K., Sakakibara, H. and Sugiyama, T. (2001) Identification of Genes Encoding Adenylate Isopentenyltransferase, a Cytokinin Biosynthesis Enzyme, inArabidopsis thaliana. *Journal of Biological Chemistry*, 276, 26405-26410.

Teale, W.D., Paponov, I.A. and Palme, K. (2006) Auxin in action: signalling, transport and the control of plant growth and development. *Nat Rev Mol Cell Biol*, 7, 847-859.

Terakado, J., Fujihara, S., Goto, S., Kuratani, R., Suzuki, Y., Yoshida, S. and Yoneyama, T. (2005) Systemic Effect of a Brassinosteroid on Root Nodule Formation in Soybean as Revealed by the Application of Brassinolide and Brassinazole. *Soil Science & Plant Nutrition*, 51, 389-395.

Tominaga, A., Nagata, M., Futsuki, K., Abe, H., Uchiumi, T., Abe, M., Kucho, K.-i., Hashiguchi, M., Akashi, R., Hirsch, A.M., Arima, S. and Suzuki, A. (2009) Enhanced Nodulation and Nitrogen Fixation in the Abscisic Acid Low-Sensitive Mutant enhanced nitrogen fixation1 of Lotus japonicus. *Plant Physiology*, 151, 1965-1976.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L. and Pachter, L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protocols*, 7, 562-578.

Trinick, M.J. (1973) Symbiosis between Rhizobium and the Non-legume, Trema aspera. *Nature*, 244, 459-460.

Turner, M., Nizampatnam, N.R., Baron, M., Coppin, S., Damodaran, S., Adhikari, S., Arunachalam, S.P., Yu, O. and Subramanian, S. (2013) Ectopic Expression of miR160 Results in Auxin Hypersensitivity, Cytokinin Hyposensitivity, and Inhibition of Symbiotic Nodule Development in Soybean. *Plant Physiology*, 162, 2042-2055.

Upreti, K.K. and Murti, G.S.R. (2004) Effects of Brassmosteroids on Growth, Nodulation, Phytohormone Content and Nitrogenase Activity in French Bean Under Water Stress. *Biologia Plantarum*, 48, 407-411.

Vercruyssen, L., Gonzalez, N., Werner, T., Schmülling, T. and Inzé, D. (2011) Combining Enhanced Root and Shoot Growth Reveals Cross Talk between Pathways That Control Plant Organ Size in Arabidopsis. *Plant Physiology*, 155, 1339-1352.

Wall, L.G. (2000) The Actinorhizal Symbiosis. *Journal of plant growth regulation*, 19, 167-182.

Wang, K.L., Li, H. and Ecker, J.R. (2002) Ethylene biosynthesis and signaling networks. *The Plant cell*, 14 Suppl, S131-151.

Wang, Q., Kohlen, W., Rossmann, S., Vernoux, T. and Theres, K. (2014a) Auxin Depletion from the Leaf Axil Conditions Competence for Axillary Meristem Formation in Arabidopsis and Tomato. *The Plant cell*, 26, 2068-2079.

Wang, R., Guegler, K., LaBrie, S.T. and Crawford, N.M. (2000) Genomic Analysis of a Nutrient Response in Arabidopsis Reveals Diverse Expression Patterns and Novel

Metabolic and Potential Regulatory Genes Induced by Nitrate. *The Plant cell*, 12, 1491-1510.

Wang, Y., Wang, J., Shi, B., Yu, T., Qi, J., Meyerowitz, E.M. and Jiao, Y. (2014b) The Stem Cell Niche in Leaf Axils Is Established by Auxin and Cytokinin in Arabidopsis. 26, 2055-2067.

Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W.H.A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M. and Venables, B. (2014) gplots: Various R programming tools for plotting data. R package version 2.14.1. http://CRAN.R-project.org/package=gplots.

Wasson, A.P., Pellerone, F.I. and Mathesius, U. (2006) Silencing the Flavonoid Pathway in Medicago truncatula Inhibits Root Nodule Formation and Prevents Auxin Transport Regulation by Rhizobia. *The Plant Cell Online*, 18, 1617-1629.

Werner, T., Motyka, V., Laucou, V., Smets, R., Van Onckelen, H. and Schmulling, T. (2003) Cytokinin-deficient transgenic Arabidopsis plants show multiple developmental alterations indicating opposite functions of cytokinins in the regulation of shoot and root meristem activity. *The Plant cell*, 15, 2532-2550.

Williams, P.M. and Sicardi de Mallorca, M. (1984) Effect of gibberellins and the growth retardant CCC on the nodulation of soya. *Plant and Soil*, 77, 53-60.

Xiong, L. and Zhu, J.K. (2003) Regulation of abscisic acid biosynthesis. *Plant Physiol*, 133, 29-36.

Yamaguchi, S. (2008) Gibberellin Metabolism and its Regulation. *Annual Review of Plant Biology*, 59, 225-251.

Ye, H., Li, L. and Yin, Y. (2011) Recent Advances in the Regulation of Brassinosteroid Signaling and Biosynthesis PathwaysF. *Journal of Integrative Plant Biology*, 53, 455-468.

Zdyb, A., Demchenko, K., Heumann, J., Mrosk, C., Grzeganek, P., Göbel, C., Feussner, I., Pawlowski, K. and Hause, B. (2011) Jasmonate biosynthesis in legume and actinorhizal nodules. *New Phytologist*, 189, 568-579.

Chapter 3

3. Small RNAs in root nodules of soybean

3.1. Introduction

Small RNAs are 20 to 24-nt, non-coding RNAs that regulate gene expression in a transcriptional or post-transcriptional manner (Chen 2009, Chen 2010, Katiyar-Agarwal and Jin 2010, Axtell 2013). miRNAs (microRNAs) are a small RNA that belong to the hairpin RNAs (hpRNAs) class, i.e., they are derived from single-stranded precursors that form a hairpin structure during their biogenesis (Axtell 2013). Although miRNAs regulate diverse developmental processes in both plants and animals, differences exist in the genomic arrangement of the miRNA loci, the method of excision of miRNA precursors, the involvement of cellular compartments during processing of the miRNAs, and the major mode of target repression in the organisms belonging to the two kingdoms (Axtell et al. 2011).

In plants, most of the miRNA-encoding genes are located at intergenic regions. A pri-miRNA of 70-300-nt with a 5’ cap and 3’-polyadenylation is transcribed by DNA-dependent RNA Polymerase II (Pol II). The pri-miRNA is processed into a stem-loop pre-miRNA by the activity of DICER Like 1 (DCL1) along with HYPONASTIC LEAVES (HYL1), SERRATE (SE), and nuclear cap binding protein complex (CBC). In animals, this cleavage is mediated by two RNA III endonucleases, DROSHA and DICER. Among the four different types of DCLs present in *Arabidopsis*, DCL1 is primarily involved in miRNA biogenesis, and DCL4 is involved in the cleavage of double-stranded substrates. The pre-miRNA is further processed by DCL1, yielding a duplex with the miRNA and its cognate miRNA* with characteristic 2-nt overhangs at

the 3’ ends. In plants, biogenesis of the miRNA/miRNA* duplex is accomplished in the nucleus. The 3’ ends of the duplex are 2’-*O*-methylated by HEN1, which helps protect the miRNAs from 3’-end truncation and oligouridylation. In animals, pre-miRNAs are transported into cytoplasm for further processing by Exportin -5 and the 2’-*O*-methylation step is absent. The methylated duplex is exported to the cytoplasm, where the miRNA (guide strand) is associated with the RISC (RNA-Induced Silencing Complex) for cleavage or repression of mRNA, and the miRNA* (passenger strand) is degraded (reviewed by (Chen 2005, Chen 2009, Rogers and Chen 2013, Xie et al. 2014)) (Figure 3.1). The miRNA strand is generally thought to become an active RNA molecule (Bartel 2004) and incorporated into RISC, however, new evidence shows that miRNA* is also involved in the regulation of the mRNA (Devers et al. 2011).

lin-4 was the first miRNA discovered in *Caenorhabditis elegans* (Lee et al. 1993), followed by *let-7*. Stage-specific expression of *lin-4* and *let-7* regulates the developmental timing from the larval to the adult stages of *C. elegans* (Reinhart et al. 2000). After the discovery of *lin-4* and *let-7* using forward genetic techniques, cloning of small RNAs in drosophila, human and worm (Lagos-Quintana et al. 2001, Lau et al. 2001, Lee and Ambros 2001) resulted in an extensive numbers of miRNAs. The primary approach of miRNA gene identification in these studies was isolation of size-fractionated small RNA, adapter ligation, and reverse transcription followed by cloning and sequencing. An informatics approach that used an RNA folding program along with this classical method was later used in the discovery of numerous miRNA genes in *C. elegans* (Lee and Ambros 2001). In plants, additional approaches, such as computational prediction based on conserved sequence and secondary structure, cloning of small RNA

libraries, and high throughput sequencing of small RNA libraries, have been widely used for miRNA identification (Sun et al. 2014).

The above-described methods have their own advantages and disadvantages. For example, the miRNAs *lin-4* and *let-7* were discovered by a forward genetics technique. This method is time- consuming, requires an enormous effort to identify the gene responsible for the mutant phenotype, and depends on a visible or detectable phenotype. Computational modeling allows prediction of miRNAs independent of expression pattern, but requires downstream analyses like cloning or northern hybridization (Zhang et al. 2006). A cDNA cloning method is unable to find miRNAs that are expressed at low levels or in condition- or cell type-specific manners. However, a method combining cDNA cloning and next generation sequencing (NGS) allows comparison of miRNA abundance between different samples. NGS techniques are more reliable and sensitive in the identification of miRNAs, but require computational tools for identification and functional tools for characterization of the predicted miRNAs.

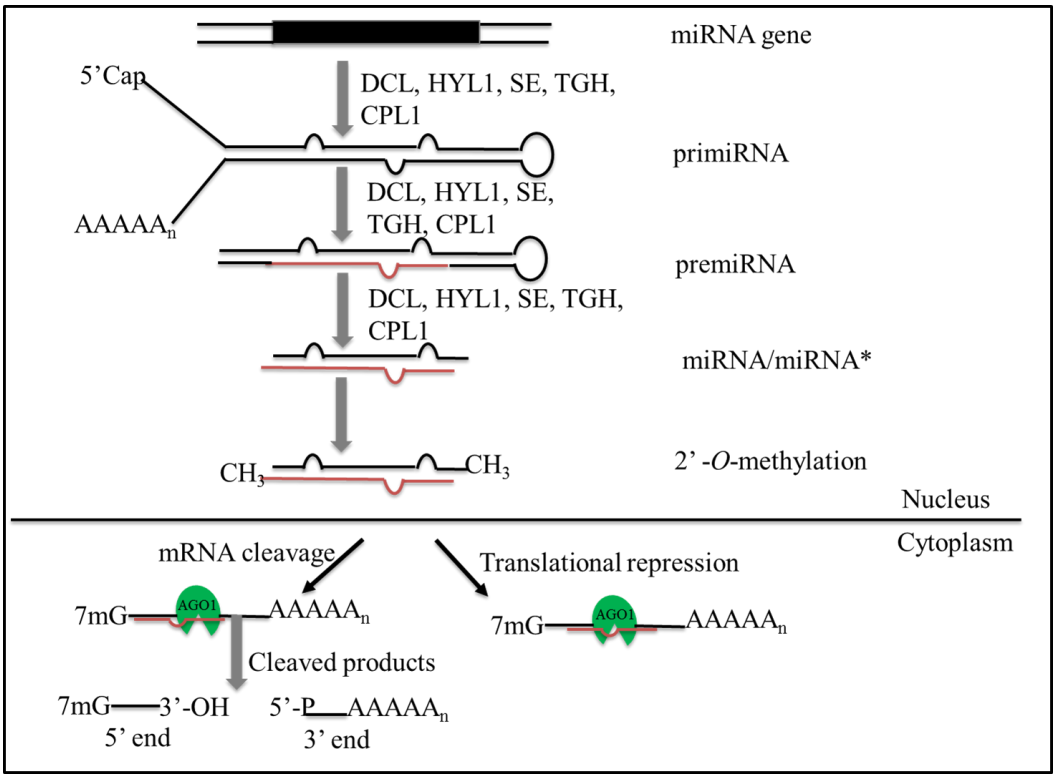


Figure 3.1. Schematic representation of plant miRNA biogenesis (adapted from (Rogers and Chen 2013)).

With the discovery of an immense number of miRNAs in different species, there is an interest as well as a necessity to functionally characterize the identified miRNAs. miRNAs in plants regulate the target mRNA with near perfect complementarity, and this has made target prediction relatively easier compared to in animals (Rhoades et al. 2002, Jones-Rhoades and Bartel 2004). Use of a modified RNA ligase-mediated 5' rapid amplification of cDNA ends (5' RLM-RACE), northern blot, and western blot have become common experimental approaches for the identification of the cleaved target fragment (Sun et al. 2014). Based on a principle similar to that of RLM-RACE, PARE (Parallel Analysis of RNA Ends) provides high throughput information of the cleaved targets (German et al. 2008).

miRNAs play important roles during organ development in plants, by modulating the levels of several key transcription factors (TFs). For example, miR164 regulates NAC-like transcription factors CUC1 and CUC2 (for CUP SHAPED COTYLEDON 1 and 2), which are required for shoot axillary meristem initiation and floral organ development (Mallory et al. 2004a, Raman et al. 2008). Another example is miR166, which regulates HDZIP-III genes crucial for proper leaf development (Juarez et al. 2004, Mallory et al. 2004b). Root nodules are a class of root lateral organs formed in response to a symbiotic interaction between a legume and a class of nitrogen-fixing bacteria commonly known as rhizobia. Bacteria reside in the nodules, fix the atmospheric nitrogen to ammonia ($NH_{3)}$, and exchange it with the legume in return for carbon. This mutually beneficial relationship helps to provide significant amounts of nitrogen to the host plant. Nitrogen is one of the elements that limits crop productivity, and therefore is highly

important to agriculture. In most cases, legume crop residues provide enough residual nitrogen to benefit subsequent crops (Frankow-Lindberg and Dahlin 2013).

Various miRNAs are expressed at different stages of root nodule development in a number of legumes, such as *Glycine max*, *Medicago truncatula*, and *Lotus japonicus*, suggesting key roles for these molecules in the processes involved in development of these unique structures (Subramanian et al. 2008, Lelandais-Briere et al. 2009, Wang et al. 2009, Joshi et al. 2010, Turner et al. 2012, Yan et al. 2016). The miRNAs miR169, miR166, miR172, miR482, miR1515, miR1512, miR160, miR171, and miR397 are known to play crucial roles in proper nodule development in legumes (Combier et al. 2006, Boualem et al. 2008, Li et al. 2010, D'Haeseleer et al. 2011, De Luis et al. 2012, Turner et al. 2013) (see Table 1.1, Chapter 1). However, this list is not exhaustive and there are sure to be additional miRNAs that are yet to be discovered and evaluated for their role in nodule development.

The objective of this study is the global identification of spatially regulated miRNAs in soybean root nodules. Soybean is not only an important food and feed crop, but also a significant producer of fixed nitrogen (Graham and Vance 2003). Earlier studies that identified miRNAs in functional soybean root nodules (Wang et al. 2009, Joshi et al. 2010) used uninoculated root as a control. Comparison of differential gene expression between nodule and uninoculated roots might overrepresent the "nodule-enriched" expression of miRNAs and would miss miRNAs that were involved in spatial regulation of nodule development. To address these limitations, we prepared small RNA libraries by using only nodules (Mature Nodule, MN) from inoculated root and root sections above and below the nodule and devoid of any lateral organ (adjacent root

Above and Below Nodule, ABMN) as the control to identify “root enriched” and “nodule enriched” miRNAs. To qualify as “root enriched” miRNAs the negative change value was greater than $\log_2$fold when differential expression in MN is compared to that in ABMN and as “nodule enriched” the positive change value was $\log_2$fold when differential expression in MN is compared to that in ABMN. We hypothesize that root-enriched miRNAs modulate target gene levels for proper nodule development (Figure 3.2). To identify miRNA-regulated targets in root nodules, we generated Parallel Analysis of RNA Ends (PARE) libraries of the MN and ABMN tissues.

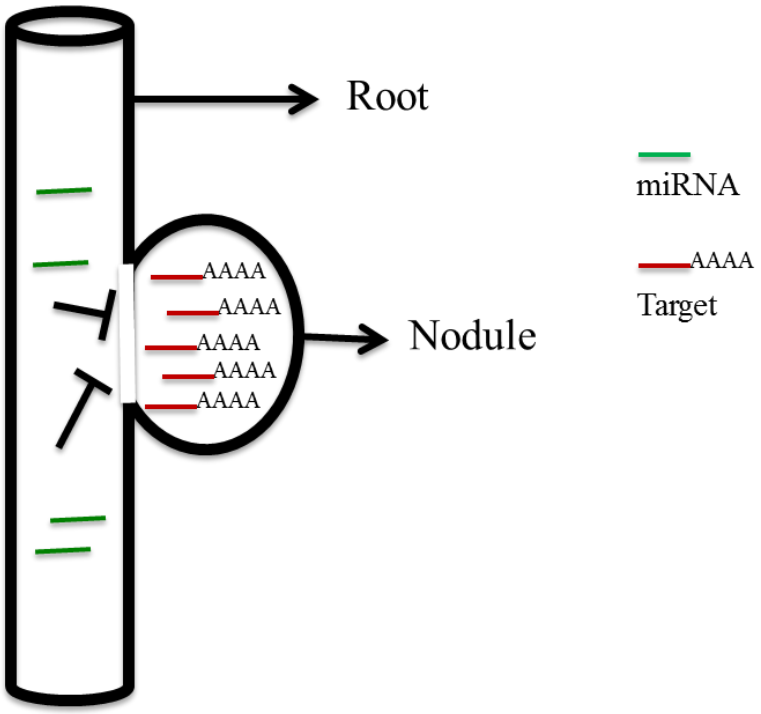


Figure 3.1 Schematic representation of root-enriched miRNAs that play roles in nodule development. Root-enriched miRNAs (shown as green lines) cleave the target gene (shown as a red line with polyA end) in the root, causing nodule-enriched target expression.

3.2. Materials and Methods

3.2.1. Plant growth and *B. japonicum* treatment

Glycine max (cv. Williams 82) seeds were surface-sterilized (Subramanian et al. 2008) and sown in a mixture of autoclaved vermiculite and perlite (2:1 ratio). Seedlings were grown in a growth chamber (Conviron, Manitoba, Canada) at 25 ºC and under a 16-h/8-h light cycle and were watered with nitrogen-free plant nutrient solution (N^-PNS, (Subramanian et al. 2008). To harvest MN and ABMN tissues, plants were inoculated with *Bradyrhizobium japonicum* (USDA 110) grown in Vincent rich medium (Schwartz 1972) at 30 ºC (shaking at 200 rpm). *B. japonicum* cells were centrifuged (3000 x g, 8 min @ 4 ºC) using an Eppendorf 5804 R centrifuge (rotor F-34-6-38) (Eppendorf, NY). The pellet was resuspended in N^-PNS to an OD of 0.08 at 600 nm (Bhuvaneswari et al. 1980) and was applied at 25 ml/pot to the roots.

3.2.2. Tissue harvest/ RNA isolation

MN and ABMN tissues were harvested at 14-16 dpi (days post inoculation). For tissue harvest, plants were removed from the vermiculite and perlite mixture and washed thoroughly in sterile water to remove the vermiculite/perlite particles attached to the root surface. The MN portion was harvested using a sterile forceps under a dissection microscope, and the corresponding ABMN tissue was dissected out using a sterile scalpel blade (Figure 3.2). The dissected tissues were quickly wipe dried on a sterile paper towel, placed into TRI reagent (Sigma-Aldrich, St. Louis, MO) in a pre-weighed 2 ml microcentrifuge tube on ice, and stored at -80 ºC until RNA isolation. Plant growth, inoculation and tissue harvesting were performed in batches. Total RNA was isolated using the protocol recommended by the manufacturer with slight modifications

(Subramanian et al. 2008a). The major exception was that tissues were ground with a bead beater rather than manual grinding with a pestle and mortar.

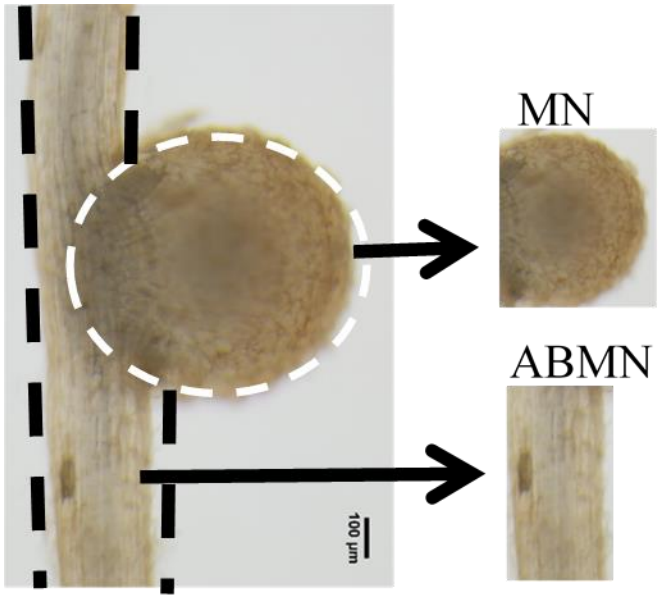


Figure 3.2 Tissue harvest for small RNA and PARE libraries. Nodule (MN) and corresponding root sections above and below nodule (ABMN) were harvested from root at 14 dpi with. B. japonicum.

3.2.3. Small RNA library preparation and sequencing

Libraries were prepared using the Illumina TruSeq RNA preparation kit and sequenced on an Illumina HiSeq 2000 at the Delaware Biotechnology Institute, University of Delaware. Briefly, 3’ and 5’ adapters were ligated at the 3’ end and 5’ end of total RNA sequentially. Sequences with adapters ligated at both ends were reverse transcribed to create single stranded cDNA. The cDNA was amplified by using a common primer and a primer with 6-base indexes to multiplex the samples and then sequenced (Figure 3.3).

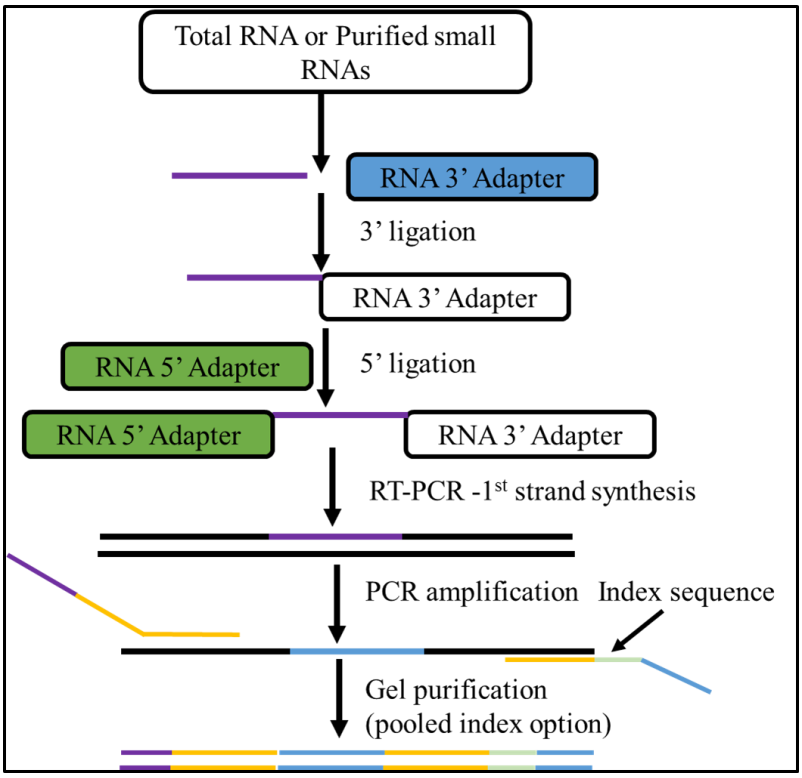


Figure 3.3 Schematic representation of small RNA library preparation. The libraries were prepared using Illumina TrueSeq Small RNA library preparation Kit.

3.2.4. PARE library preparation and sequencing

PARE library construction is based on an in-house PARE protocol (German et al. 2008) developed at the Delaware Biotechnology Institute, University of Delaware. This method uses modified RACE to capture the cleaved mRNA fragments followed by high throughput sequencing (German et al 2008). Briefly, polyadenylated small RNA that could either be a miRNA-mediated cleaved small mRNA or some random mRNA decay product were separated from the total RNA. The presence of a 5'-phosphate in the cleaved fragments and the absence of a 5'-cap makes the fragments compatible for RNA adapter ligation. First, adapters were ligated at the 5'-end of the cleaved fragments, and cDNA was synthesized with a DNA oligonucleotide consisting of oligo (dT) and 3' adapter sequences. The cDNA was then PCR amplified. The PCR products were digested with the *Mme*I enzyme, recognition sites of which are located in the 5' and 3' adapters. The *Mme*I cleaves the PCR products into 20-nt signature sequences. Finally, the *Mme*I digested products were ligated with a 3' adapter, PCR amplified, gel purified, and sequenced (German et al. 2008) (Figure 3.4).

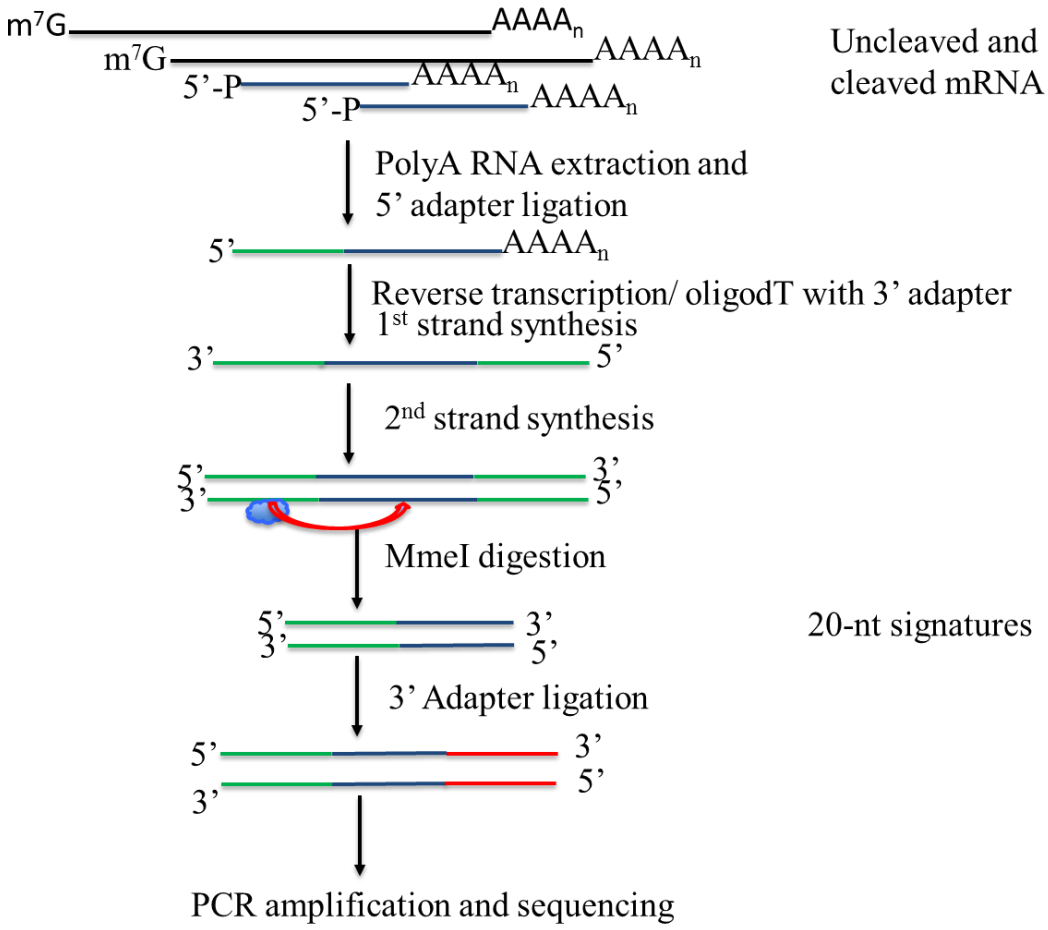


Figure 3.4 Schematic representation of PARE library preparation. This method is based on modification of 5'-rapid amplification of cDNA ends (RACE) to capture the miRNA-mediated cleaved mRNA fragments followed by high throughput sequencing (German et al. 2008, Zhai et al. 2014).

3.2.5. Processing of the reads

Raw reads from both the small RNA and PARE libraries were processed for initial quality control, such as 5’ and 3’ adapter-match, absence of adapter match, and sequencing errors, by using an in-house Perl script. The reads were adapter-trimmed and filtered to obtain sequences with size ranges from 18- to 34-nt. For PARE libraries, reads were adapter-trimmed and filtered to get sequences with size ranges from 18 to 20-nt.

3.2.6. Identification of miRNA targets in the PARE data

miRNA targets were identified and validated by using small RNA-PARE target analyzer (sPARTA) software (Kakrana et al. 2014) with the following parameters: -genomeFile -gffFile -genomeFeature 0 -miRNAFile -libs -tarPred, -tarScore, --tag2FASTA, --map2DD --validate --repeats.

3.3. Results

3.3.1. Small RNA library reads characterization

After the initial trimming and the filtering for adapter removal and quality control, MN and ABMN small RNA libraries had 29 and 55 million reads, respectively, which were then used for miRNA prediction (Table 3.1). Total reads both before and after filtering in the ABMN library were dominated by reads 21- and 24-nt in length (Table 3.1, Figure 3.5), but distinct reads were dominated by 24-nt long sequences (Table 1). In MN, the total reads were dominated by 21- and 24-nt sequences, but distinct reads were dominated by 24-nt sequences (Table 3.1, Figure 3.6).

Table 3.1 Read distribution in ABMN and MN small RNA libraries. The total numbers of reads before and after adapter trimming and filtering in ABMN and MN libraries were counted. The number of reads (total and distinct) for the two most abundant read lengths (nt) are reported after adapter trimming and filtering in the libraries.

Adapter trimming	Total reads in library		Most abundant reads by library type					
	ABMN	MN	ABMN			MN (Total)	21-nt	7,338,471
Before	56,782,201	29,773,886	Total	21-nt	8,703,876		24-nt	7,094,411
After	55,092,309	29,154,995	Distinct	24-nt	1,051,565	MN (Distinct)	24-nt	3,133,607

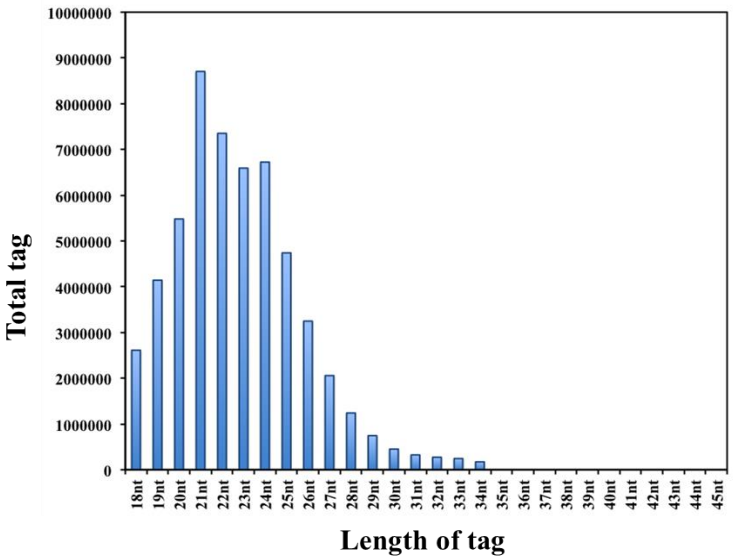


Figure 3.5 Total reads distribution in the ABMN library. Raw reads were trimmed for adapter sequences and filtered to obtain sequences in the size range of 18- to 34-nt.

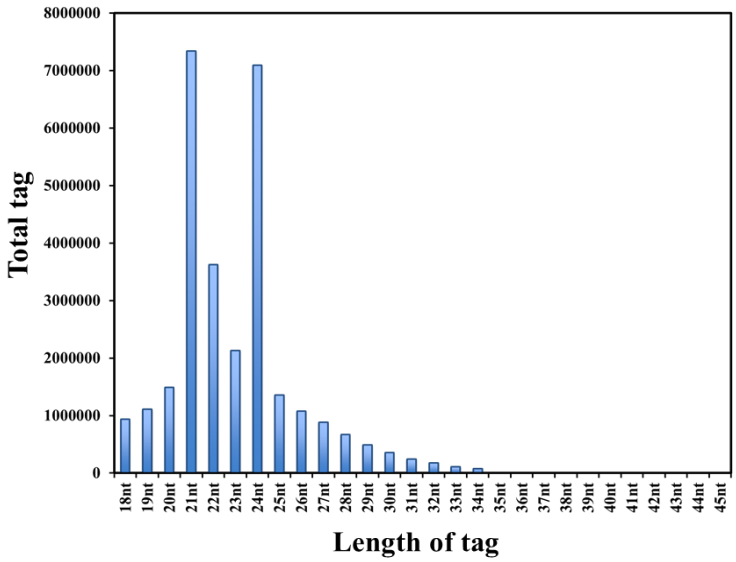


Figure 3.6 Total reads distribution in the MN library. Raw reads were trimmed for adapter sequences and filtered to obtain sequences in the size range of 18- to 34-nt.

3.3.2. Processing of small RNA reads for miRNA prediction

To process the reads for miRNA prediction, AB-BLAST was used to generate a non-redundant dataset from the ABMN and MN libraries by preserving the read count information from each library. Then, reads were filtered through various steps by mapping to: a) chloroplast, mitochondrial DNA, ribosomal DNA and repeats of soybean (phytozome_v6); b) tRNAs of *M. truncatula* (http://plantrna.ibmp.cnrs.fr/) and soybean snoRNAs (http://bioinf.scri.sari.ac.uk/cgi-bin/plant_snorna/); c) known soybean microRNAs (miRBase V20); and d) presence of at least 8 reads in both libraries combined. One limitations of filter d is that reads which were unique in one library but had less than 8 reads were removed. Reads that passed the above filtering steps were mapped to the soybean genome (Gmax_v1.1_189) using bowtie (Bowtie 2). For the mapped reads, miRNA precursor sequences were obtained and subjected to secondary structure prediction.

A higher percentage of the reads was retained in from the MN (~80%) library compared to the ABMN (~20%) library after mapping to chloroplast, mitochondrial, protoplast, rDNA and repeats of soybean (filter a). Read loss in the ABMN library was more but couldn't be explained. About 1% of the reads from both ABMN and MN libraries were mapped to tRNAs and snoRNAs (filter b), and ~ 4% and ~15% of reads were mapped to known soybean miRNAs (filter c) in ABMN and MN respectively. About 7% (ABMN) and 24% (MN) of the original reads had at least 8 reads from one or both of the libraries. In the end, ~ 3.6% (ABMN) and 13.9% (MN) of the original reads were mapped to the soybean genome (Figure 3.7) for a total of non-redundant sequences of potential interest. Overall, more reads were retained from MN compared to ABMN.

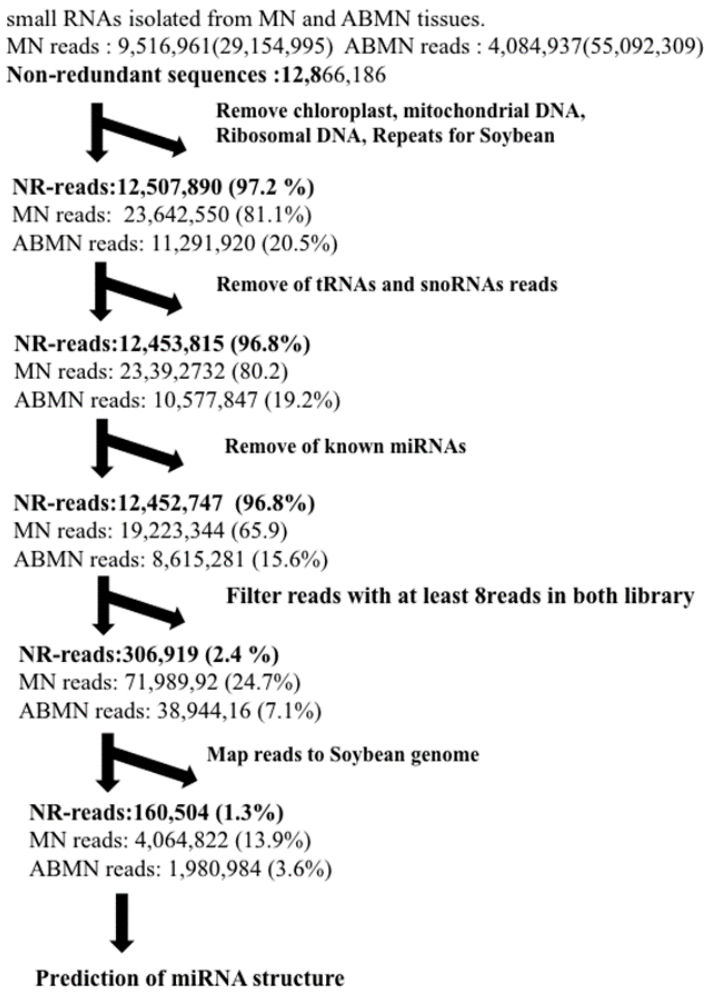


Figure 3.7 Processing of the ABMN and MN small RNA libraries for miRNA prediction. To increase computing efficiency, non-redundant reads from both libraries, keeping the read count information, were generated and used for filtering. The number of non-redundant unique reads and corresponding total reads from ABMN and MN libraries and their percentage retained compared to original reads after each filtering step is provided.

3.3.3. Identification of novel miRNA sequences with potential hairpin-forming precursors

For each mature sequence that had a perfect match to the soybean genome, the sequence was tested to see if it was potentially derived from a hairpin precursor. For this, potential precursors were generated by using a custom Perl script that extracted a 600-nt region spanning the genomic match position (on either side). Using sequence, secondary structures were predicted and validated by using the miRcheck ($MAX_UNPAIR = 4, $MAX_STAR_UNPAIR = 6, $MAX_SIZEDIFFERENCE = 3, $MAX_MIR_GAP = 2, $MAX_STAR_GAP = 2, $MIN_FBACK_SIZE = 54, $MAX_MIR_AS_BULGE = 2, $MIN_UNPAIR = 1, $BP_EXTENSION = 2). This resulted in 15, 616 mappings and 6879 precursors. The precursors were filtered through various steps: a) if they were from repeat region of the genome b) reads mapped ≥ 35 times and c) manual annotation based on Meyers et al. (Meyers et al. 2008) (Figure 3.8). The avoidance of miRNA predicted from repeat regions (filter a) would insure prediction of miRNAs rather than siRNAs (Rajagopalan et al. 2006, Kasschau et al. 2007), and the cut-off of reads ≤ 35 was chosen based on genome duplication of soybean and would ensure prediction of genuine miRNAs (Turner et al. 2012). Among 6879 predicted precursors, 500 unique precursors passed the filtering criteria.

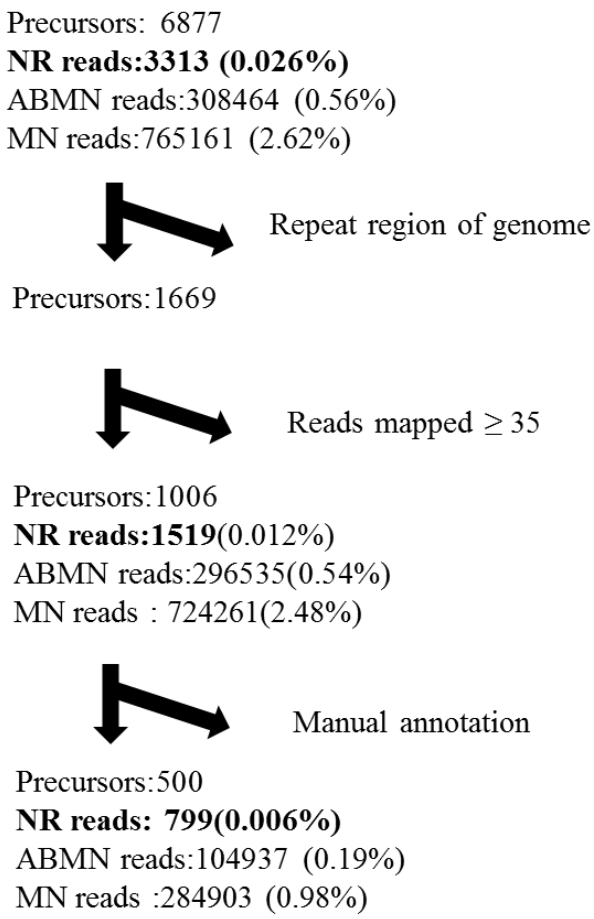


Figure 3.8 Processing of the miRNA precursor sequences. Precursors represent total number of unique precursor. For reads associated with precursors, the number of non-redundant, unique reads and corresponding total reads from the ABMN and MN libraries and their percentage retained compared to original reads after each filtering step is provided.

The potential precursors were further grouped by using three criteria: 1. presence of miRNA (miR)/miR* duplex with 2-nt overhangs; 2. presence of miR/miR* duplex plus other sequences; and 3. presence of potential miR without miR* (Figure 3.9). Criteria 2 and 3 included many precursors that had many reads mapping to the same precursors and an additional filter was used to find the best potential miRNA mature sequence: 1. for precursors in category 2, proportion of the reads mapped to each precursor was calculated and, with the exclusion of the miR/miR* duplex mature sequence, those that had at least ≥25% reads were kept; 2. for precursors in category 3, mature sequence that had a least ≥25% reads were kept. A total of 10, 17 and 470 precursors that fit to criteria 1, 2 and 3, respectively, were found (Table 3.2).

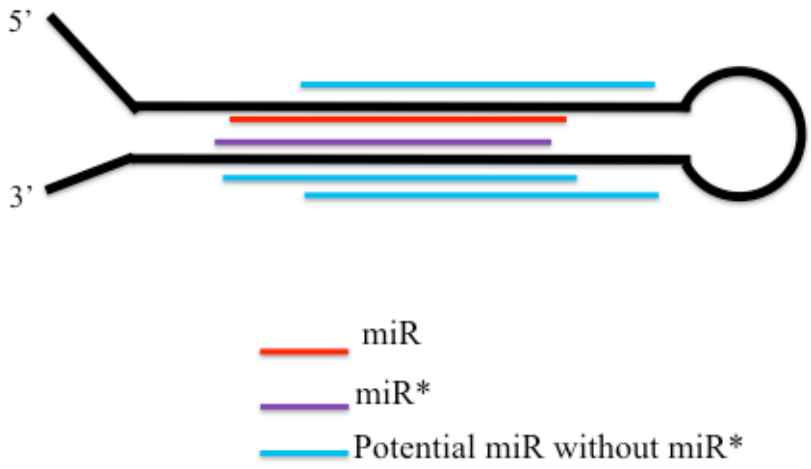


Figure 3.9 Schematic representation of miRNA precursors falling into three different categories. Criteria 1 represents precursors with a duplex (miR/miR), criteria 2 represents precursors with miR/miR* plus other sequences, and criteria 3 represents precursors with potential miR but without miR*. The miR/miR* duplex are color-coded as red and purple, and sequences other than duplex are color-coded as blue.*

To identify if these precursors were “previously reported” or “not reported”, 455 mature sequences associated with these precursors were BLAST searched (blastn) against all plant miRNAs available in miRBase (v21). The blastn parameter was set to report only 1 best alignment. Based on length of alignment and bit score, the candidates were divided into two groups. If the aligned reads had a longer match (>15-nt) and or bit score ≥30, they were classified as “conserved”, but otherwise were considered “novel”. Using these criteria, 330 precursors were found to be previously not reported, 149 precursors were previously reported, and 12 precursors were previously reported by mature sequences were not reported. (Table 3.2). An additional 7 previously unreported mature sequences were also identified from previously reported soybean precursors that were present in miRbase (v21) and added to the 455 mature sequences for further downstream analysis. Analyses of distribution of sequence length and first nucleotide of the 462 unique mature sequences associated with these precursors showed that 21-nt mature sequences were most abundant and that uridine (U) was the most prevalent first nt. This result is consistent with reported sequence size and first nt of the miRNAs (Cuperus et al. 2011) (Figure 3.10).

Table 3.2 Summary result of categorization of precursors. Sequences in each precursor were mapped to miRbase (v21) and categorized as previously reported or not reported.

Categories	Feature of categories	Number of Precursors			
		Each category	Not previously reported	Previously reported but mature sequence not reported	Precursor previously
Category 1	Presence of miR/ miR*	10	4	2	4
Category 2	Presence of miR/miR* duplex and other sequences	17	1	7	9
Category 3	Potential miR without miR*	470	325	9	136
	Total	497	330	18	149

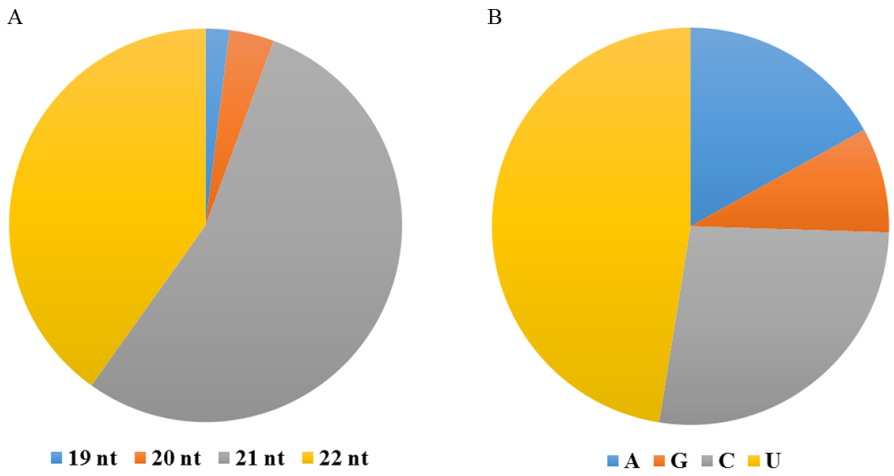


Figure 3.10 Distributions of sequence lengths and first nucleotides of the mature miRNA sequences. These mature sequences were associated with 497 precursors identified in this study. A. Nucleotide length -the mature sequences were dominated by 19-nt (blue), 20-nt (orange), 21-nt (gray) and 22-nt (blue) reads. B. first nucleotide of the mature sequences – Adenine (A), Guanine (G), Cytosine (C), and Uracil (U) were represented with blue, red, gray and orange colors respectively.

3.3.4. Differential expression of mature sequences in ABMN and MN libraries

Differential expression of the mature sequences was described as "nodule-enriched" if the log_2fold change in MN vs ABMN was ≥1, and was described as "root-enriched" if the log_2fold was ≤ -1. Analysis of differential expression of 462 potential miRNA mature sequences in MN compared to ABMN showed that 193 of these were nodule-enriched (log_2fold ≥1). Among these mature miRNA sequences, 48 were previously reported and 145 were not previously reported. Interestingly, 117 mature sequences were nodule-reduced (log_2fold ≤ -1). Among these mature miRNA sequences, 54 were previously reported and 62 were not previously reported (Figure 3.11). Many of the not "previously reported" mature sequences were up-regulated compared to "previously reported" mature sequences, suggesting they might have an as-yet-undefined functional role in nodule development.

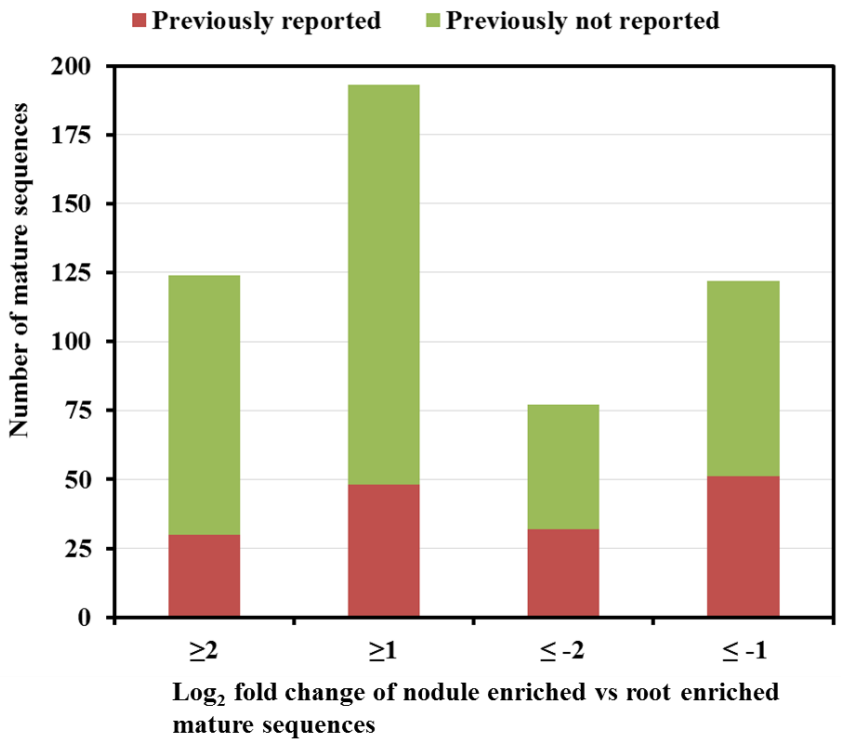


Figure 3.11 Differential expression of nodule-enriched vs. root-enriched mature miRNA sequences. Log_2 fold change of ≥1 and ≥2 represent nodule enrichment, and log_2 fold change of ≤-1 and ≤-2 represent root enrichment. For each read in the ABMN and MN libraries, the proportion was calculated by dividing with total reads mapped to the genome from each library and normalized to million reads. This proportion was used to calculate MN vs. ABMN log_2 base differential expression. The differential expression was calculated for previously reported (red) and previously not reported (green) mature sequences.

3.3.5. Differential expression of known soybean miRNAs in ABMN and MN libraries

Non-redundant reads from the ABMN and MN libraries that did not map to chloroplast, mitochondrial, protoplast, rDNA, repeats, tRNAs, and snoRNAs of soybean but mapped to soybean miRNAs (miRbase v20, 0 mismatch) were considered as known soybean miRNAs. A total of 554 known soybean miRNAs belonging to 219 families were recorded. To increase the confidence of known soybean mature sequence, the mature sequences were filtered using the following criteria: a) if the proportion of the reads that mapped to the soybean genome was greater than or equal to 25% and b) if read numbers from both libraries were at least 8. After these two filters, 228 known soybean mature sequences belonging to 92 families passed the criteria. Distribution of read length showed that sequence length varied from 18-24 nt, but 21 nt was the most abundant sequence length. As expected, uridine (U) was the most prevalent first nucleotide (Figure 3.12). Analyses of differential expression of these 235 known mature sequences showed that 70 known soybean miRNAs were enriched ($\log2fold \geq 1$), while 84 were reduced ($\log2fold \leq -1$) (Figure 3.13).

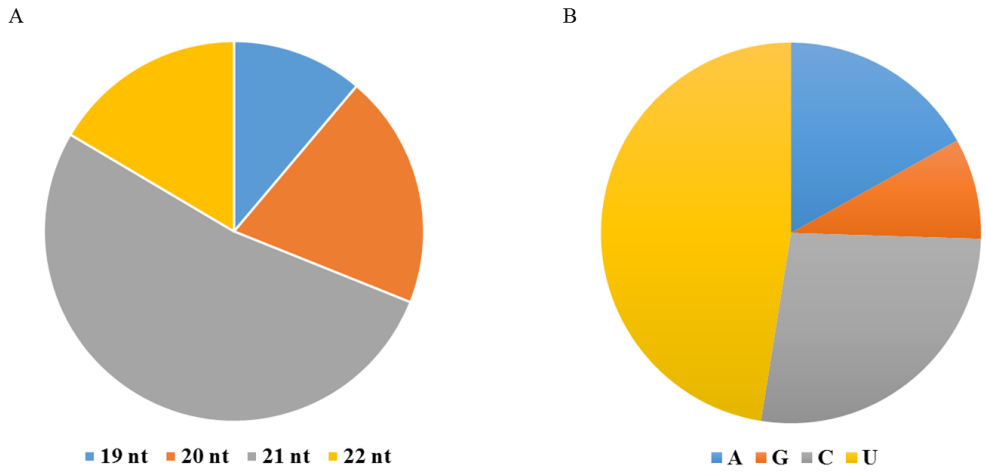


Figure 3.12 Distribution of sequence length and first nucleotide of known soybean miRNAs. A. Nucleotide length -the mature sequences were dominated by 19-nt (blue), 20-nt (orange), 21-nt (gray) and 22-nt (blue) reads. B. first nucleotide of the mature sequences – Adenine (A), Guanine (G), Cytosine (C), and Uracil (U) are represented with blue, red, gray and orange colors respectively.

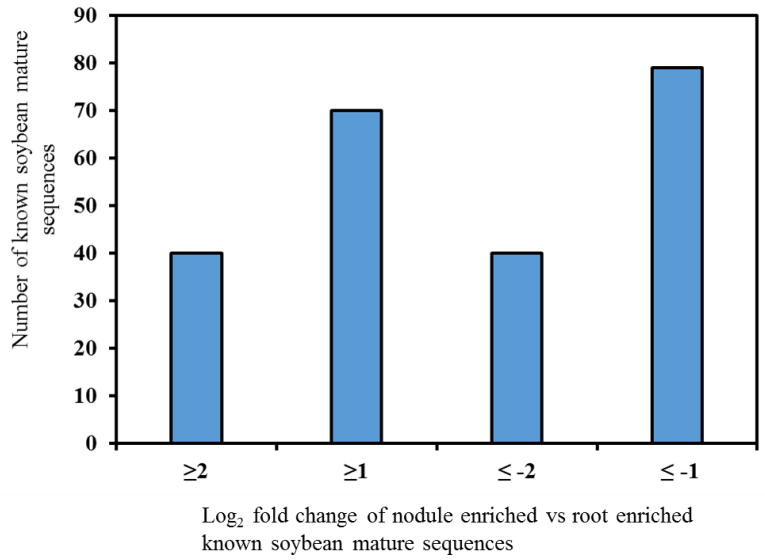


Figure 3.13 Differential expression of known soybean miRNAs. Log_2 fold change of ≥1 and ≥2 represent nodule enrichment, and log_2 fold change of ≤-1 and ≤-2 represent root enrichment. For each sequence, read proportion was calculated by dividing number of reads for each sequence with total read numbers that mapped to genome for each library and normalized it to million reads. This proportion was used to calculate MN vs ABMN log_2 base differential expression.

3.3.6. PARE library reads distribution

After trimming the adapters, ABMN and MN PARE libraries had about 21 and 34 million total sequenced reads. MmeI restriction site, present at the 5' adapter (German et al. 2008), was used to cleave the PCR product into 20-nt fragments; therefore reads >20-nt were trimmed to obtain 20-nt sequences. The total number of 20-nt long reads increased from about 10 million to 20 million in ABMN and from 17 to 31 million in MN after trimming and filtering (Table 3.3, Figure 3.14 and 3.15).

Table 3.3 Read distribution in ABMN and MN PARE libraries. Number of total and distinct 20-nt signatures is provided.

Trimming	Total		Total 20-nt signatures		Distinct 20-nt signatures	
	MN	ABMN	MN	ABMN	MN	ABMN
Before	34,625,442	21,986,317	17,739,349	10,436,765	6,235,096	4,476,047
After	34,484,872	21,755,376	31,952,932	20,499,892	9,800,848	7,355,167

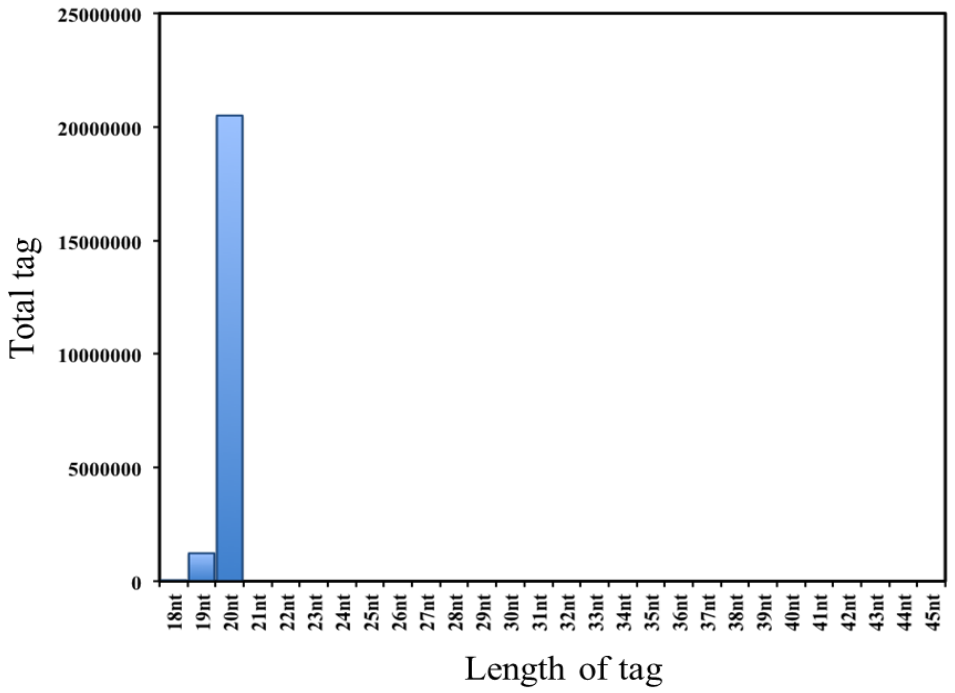


Figure 3.14 Distribution of 20-nt signature in ABMN library. The post-trimmed reads were filtered to get 18- to 20-nt signatures and used for target prediction and validation.

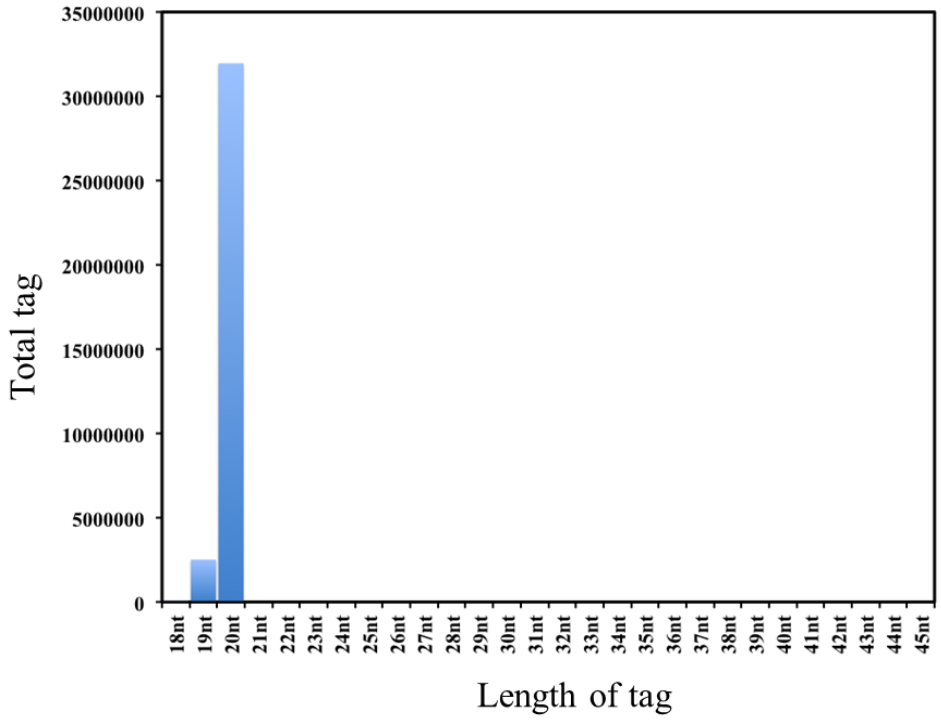


Figure 3.15 Distribution of 20-nt signature in MN library. The trimmed reads were filtered to get 18- to 20-nt signatures and used for target prediction and validation.

3.3.7. Identification of the gene targets of the potential mature miR sequences

Analysis of PARE data by sPARTA validated targets for 248 of the 462 potential miRNA mature sequences. The mature sequence/target pairs were then filtered based on various criteria: a) score ≤5; b) noise corrected p-value <0.05; and c) targets belonging to category 0 and 1. The score penalizes the predicted miRNA and target alignment for a mismatch at the 10^{th} and 11^{th} position between the miRNA and target, for a wobble or mismatch with a single flanking mismatch or for a mismatch on both sides. The category provides information about how reliable a predicted target is. Category 0/1 represents a PARE signal with the most promising site of miRNA-directed cleavage. This resulted in 128 mature sequences/target pairs. These targets were annotated as related to transcription factor families like TCP, MYB, NAC, GRAS and WRKY. Among the 128 mature sequences/target pairs, 63 belonged to previously unreported mature sequences; a further filter based on PARE reads ≥ 5 resulted into 20 previously unreported mature sequence/ target pairs (Table 3.4).

The expression patterns of the 248 mature sequence/target pairs from the sPARTA validation were examined in small RNA and transcriptome libraries of MN and ABMN. This analysis revealed 97 mature sequence/ target pairs with inverse differential expression between the two libraries (i.e., if the miRNA was enriched in ABMN, the target was enriched in MN or the other way round). To increase the confidence of the predicted spatial regulation of the target by miRNAs, targets with significant $\log_2$ fold change (noise corrected p-value (q-value) <0.05) was used as an criteria to filter differentially expressed mature sequence/target pairs with an inverse expression pattern in MN vs ABMN tissues. Based on this criteria, 10 mature sequence/target pairs were

identified, of which 6 were root-enriched mature sequence/nodule-enriched target and 4 were nodule-enriched mature sequence/root-enriched target (Table 3.5). Annotation of targets associated with these root-enriched and nodule-enriched mature sequences showed that Glyma.16g050500 was related to auxin signaling and that Glyma.08g173400 and Glyma08g173400 were related to a NAC domain-containing protein (Table 3.6)

3.3.8. Identification of the targets of known soybean miRNAs

PARE analyses validated targets for 105 of the 228 known soybean miRNAs (identified based on section 3.3.5, proportion of reads ≥ 25, sum of number of reads from both libraries ≥8). These soybean miRNA/target pairs were filtered based on a) score <5, b) PARE reads ≥5 and c) noise corrected p-value <0.05, resulting in 41 of these passing the criteria (Table 3.7). Many of these known soybean miRNA/target pairs were represented by previously predicted and RACE-verified miRNA/target pairs, such as miR160/auxin response factor 10, miR166/Homeobox-leucine zipper family protein, miR169/nuclear factor Y, and miR171/GRAS family transcription factor (Llave et al. 2002, Rhoades et al. 2002, McHale and Koning 2004, Kim et al. 2005, Combier et al. 2006). The presence of RACE-verified targets in our data supports the PARE/degradome validation of targets of previously unreported mature miRNA sequences.

Among the 105 validated known miRNA/target pairs, 47 had inverse differential expression patterns between the MN and ABMN small RNA and transcriptome libraries. These 47 known miRNA/target pairs were also filtered out if they had a target with a significant $\log_2$ fold change in the MN vs. ABMN transcriptome data. There were 8 known miRNA/target pairs that passed the filter, of which 6 were root-enriched miRNA/nodule-enriched target pairs and 2 were nodule-enriched miRNA/root-enriched

target pairs (Table 3.8 and 3.9). One of root-enriched miRNA/nodule-enriched target pairs represented gma-miR393. Previously, overexpression of miR393 in soybean resulted in reduced sensitivity to auxin, but didn't inhibit nodule formation (Turner et al. 2013). It will be interesting to see how miR393 affects spatial regulation through auxin signaling.

Table 3.4 List of previously unreported mature sequence/target pairs. For mature sequences, the numbers of reads in the ABMN and MN small RNA libraries are provided. For the targets, the Fragments per Kilo base of transcript per Million Mapped reads (FPKM) values of ABMN and MN from transcriptome libraries and annotation is provided (G. max 1.89).

Sequence	Reads		Target	FPKM		Annotation
	ABMN	MN		ABMN	MN	
GUGGUAUUGCUUCUGUUCAUCU	50	576	Glyma.03g206400	45.2	61.4	eukaryotic translation initiation factor
GUGGUAUUGCUUCUGUUCAUC	93	853	Glyma.06g142200	8.7	26.3	0
ACAAUCCUCACCUUACAACCCA	2	8	Glyma.20g173100	0.0	0.0	transducin family protein / WD-40 repeat family protein
AAUUUACAGUGGAACAUUCUC	1	31	Glyma.16g124500	0.0	0.0	RADIATION SENSITIVE 17
UUCGAUGUGGGACUUCAACAU	4	9	Glyma.06g217200	1.3	0.0	Peptidase C13 family
CGGAAUGCCAAAGGACAUGUU	251	182	#N/A	#N/A	#N/A	#N/A
UCGGAAUGCCAAAGGACAUGUU	338	350	#N/A	#N/A	#N/A	#N/A
AGGACAACUCUUCUUUGGCGC	4	13	Glyma.11g133800	9.4	53.2	dicarboxylate transporter 1

UCCAUGUCUGAACUUUGUCCA	5	16	Glyma.11g243000	19.8	30.2	beta-1,4-N-
AUAUAUUCGGAUAUUCACAUU	15	3	Glyma.10g078400	76.4	56.4	2-cysteine peroxiredoxin B
UCAUUGGGCUCUACAAUGCACC	2	10	Glyma.02g085900	1.9	2.2	response regulator 2
CAUUGGGCUCUACAAUGCACC	2	9	Glyma.02g085900	1.9	2.2	response regulator 2
UUCAACUGAGAUCCGGUACCU		8	Glyma.05g246500	18.5	21.9	S-adenosyl-L-methionine-dependent
AUUUUCUUGAGAAUUUGGCCU	69	413	Glyma.15g135200	1.6	2.1	ent-kaurenoic acid hydroxylase 2
ACUCCUGUUUGCACUAAAGAUU	9	13	Glyma.14g098700	35.6	28.9	0
AACAGGAUGCCAAACUUAAGG	2	14	Glyma.20g168100	125.2	180.8	ribosomal protein 1
UUGAUACACCAGGUGCAAUGUC	14	2	Glyma.04g033900	14.4	20.4	0
CGACCGUGUUUCCUUGGUUAA	5	5	Glyma.11g073600	31.0	42.6	ubiquitin-conjugating enzyme 16
TAGCCAAGGATGGACTTGCCTA	10	11	Glyma.02g195000	0.0	0.0	nuclear factor Y, subunit A10
TTGGACTGAAGGGAGCTCCTTC	1418	185	Glyma.13g292500	0.0	0.0	TCP family transcription factor 4

Table 3.5 List of potential miRNA/target pairs that show an inverse expression in small RNA vs transcriptome libraries. miRNA mature sequence and target pairs were characterized as root-enriched mature sequence /nodule-enriched target pair or nodule-enriched mature sequence/root-enriched target pair based on log_2 significant fold change of target in transcriptome libraries between MN and ABMN tissues (described in chapter 2). Annotations of these PARE- validated targets are provided in Table 7.

Mature sequence	Number of reads		Log_2fold	Target	FPKM		Log_2fold
Root-enriched	ABMN	MN		Nodule-enriched	ABMN	MN	
UCCAAAGGGAUCGCAUUGAUCU	6236	5324	-1.29	Glyma.16g050500	20.5	39.1	0.93
CAUGUGCCCCCCUUCCCCAUC	438	29	-4.98	Glyma.18g036300	2.8	66.7	4.55
UGGAGAAGCAGGGCACAUGCU	13	21	-0.37	Glyma.08g173400	10.1	42.4	2.07
UGGCAAGCUUCCUCGGCUAUU	25	32	-0.71	Glyma.08g057900	90.8	165.5	0.87
AUUCUUCCCCUAUGUCUUGUCC	28	6	-3.28	Glyma.14g035100	0.0	38.8	9.84
AUGGCCAUGGUUCUUGCAGCUG	4	6	-0.48	Glyma.19g019200	3.1	61.7	4.31
Nodule-enriched				Root-enriched			
CCAUUGGUGUAAAUCUUCCUG	1	9	2.11	Glyma.15g019400	16.8	4.3	-1.95
UAAGAAUUUGUUUAUAUUGUU	7	16	0.13	Glyma.19g198000	8.1	1.5	-2.43
UUGUUUUACCUAUUCCACCCAU	550	1956	0.77	Glyma.12g212500	2.7	0.7	-1.89
UAUGGAUAACAUGCAGAAGCU	2	16	1.94	Glyma.10g142700	69.8	4.4	-3.99

Table 3.6 Annotation of nodule-enriched and root-enriched targets. PARE-validated targets annotation is based on G. max 1.89.

Target	Annotation
Nodule-enriched	
Glyma.16g050500	auxin signaling F-box 3
Glyma.19g019200	H(+)-ATPase 11
Glyma.18g036300	purine biosynthesis 4
Glyma.08g173400	NAC domain containing protein 1
Glyma.08g057900	NA
Glyma.14g035100	NA
Root enriched	
Glyma.15g019400	Homeodomain-like superfamily protein
Glyma.19g198000	HVA22-like protein J
Glyma.12g212500	disease resistance protein (TIR-NBS-LRR class), putative
Glyma.10g142700	Low temperature and salt responsive protein family

Table 3.7 Known soybean miRNAs and targets validated by PARE data. PARE-validated targets were annotated by using soybean gene annotation (G. max 1.89). miRNAs with previously RACE-validated targets are identified in our data in red font.

miRNA	ABMN	MN	Target	Annotation
gma-	106363	263909	Glyma.11G212800	LRR and NB-ARC domains-containing disease
gma-	4	4	Glyma.16G072400	Glutaredoxin family protein
gma-	415	563	Glyma.16G072400	Glutaredoxin family protein
gma-	2854	3881	Glyma.04G195800	#N/A
gma-	119	175	Glyma.04G195800	#N/A
gma-	10519	23647	Glyma.15G232600	Disease resistance protein (TIR-NBS-LRR class)
gma-	21	43	Glyma.07G048000	NAC transcription factor-like 9
gma-miR156r	2	9	Glyma.04G159600	Squamosa promoter-binding protein-like (SBP
gma-miR156s	352	425	Glyma.04G159600	Squamosa promoter-binding protein-like (SBP
gma-miR156u	794	3599	Glyma.04G159600	Squamosa promoter-binding protein-like (SBP
gma-miR159a-	52110	82632	Glyma.04G125700	myb domain protein 33
gma-miR159c	13	175	Glyma.04G125700	myb domain protein 33
gma-miR160a-	646	2509	Glyma.11G145500	auxin response factor 10

gma-miR166n	2086	9130	Glyma.05G166400	Homeobox-leucine zipper family protein / lipid-
gma-miR167g	956	4538	Glyma.18G046800	auxin response factor 8
gma-miR171b-	1200	22227	Glyma.04G251900	GRAS family transcription factor
gma-miR171k-	235	26	Glyma.01G177200	GRAS family transcription factor
gma-miR171o	174	60	Glyma.04G251900	GRAS family transcription factor
gma-miR171s	581	9550	Glyma.04G251900	GRAS family transcription factor
gma-miR172b-	16	10	Glyma.14G067600	#N/A
gma-miR172e	33	2352	Glyma.15G044400	related to AP2.7
gma-miR172h-	17	20	Glyma.15G044400	related to AP2.7
gma-miR172i-		31	Glyma.15G044400	related to AP2.7
gma-miR2109	23	21	Glyma.03G075300	Disease resistance protein (TIR-NBS-LRR class)
gma-	411	204	Glyma.13G303900	Calcium-binding EF-hand family protein
gma-miR2119	205	219	Glyma.14G121200	alcohol dehydrogenase 1
gma-miR319m	5834	259	Glyma.13G292500	TCP family transcription factor 4
gma-miR393e	140	90	Glyma.16G050500	auxin signaling F-box 3
gma-miR393g	832	415	Glyma.16G050500	auxin signaling F-box 3

gma-miR393j	4199	425	Glyma.16G050500	auxin signaling F-box 3
gma-miR395b	1	15	Glyma.18G168900	slufate transporter 2;1
gma-miR396i-	110008	104989	Glyma.09G112600	0
gma-miR398a	104	78	Glyma.08G209700	Rubredoxin-like superfamily protein
gma-miR398c	558	1741	Glyma.14G222700	Prolyl oligopeptidase family protein
gma-miR408b-	468	3734	Glyma.04G242300	plantacyanin
gma-miR408c-	70	395	Glyma.04G242300	plantacyanin
gma-	3	10	Glyma.13G227500	0
gma-miR482c-	172	681	Glyma.13G303900	Calcium-binding EF-hand family protein
gma-	190	1012	Glyma.04G200100	0
gma-	8		Glyma.10G295300	homogentisate phytyltransferase 1
gma-miR5767	48	27	Glyma.02G041600	F-box family protein

Table 3.8 Known soybean miRNA sequences that showed an inverse expression patterns. Known mature sequence and target pairs are characterized as root-enriched known miRNA /nodule-enriched target pair or nodule-enriched known miRNA/root-enriched target pair based on log_2 significant fold change of target in transcriptome libraries of MN vs ABMN tissues (described in chapter 2). Annotation of these targets are provided in Table 3.9.

Sequence	miRNA	Reads		Log_2fold	Target	FPKM		
		ABMN	MN			ABMN	MN	Log_2fold
	Root-enriched							
CGATGTTGGTGAGGTTCAATC	gma-miR171l	214	20	-4.48	Glyma.08G214000	1.48	4.89	1.72
TAATCTGCATCCTGAGGTTTA	gma-miR2111f	78	33	-2.30	Glyma.19G090600	3.42	10.02	1.55
TCCAAAGGGATCGCATTGATCC	gma-miR393e	140	90	-1.70	Glyma.16G050500	20.55	39.07	0.93
TCCAAAGGGATCGCATTGATC	gma-miR393g	832	415	-2.07	Glyma.16G050500	20.55	39.07	0.93
TTCCAAAGGGATCGCATTGATC	gma-miR393j	4199	425	-4.37	Glyma.16G050500	20.55	39.07	0.93
TGTGTTCTCAGGTCACCCCTT	gma-miR398a	104	78	-1.48	Glyma.08G209700	51.89	133.94	1.37
GGTACCCTTTCAGATAGTCTCA	gma-miR5034	69	9	-4.00	Glyma.10G081700	0.88	8.18	3.22
TGGAGGACCTTTGAAGGTGCA	gma-miR5767	48	27	-1.89	Glyma.02G041600	3.79	9.87	1.38
	Nodule-enriched							
CTGAAGTGTTTGGGGGAACTC	gma-miR395b	1	15	2.84	Glyma.18G168900	4.98	0.82	-2.61
TTGATTCTCATCACAACATGG	gma-miR4415a-3p	13	137	2.34	Glyma.08G243600	39.80	14.69	-1.44

Table 3.9 Annotation of the targets of known soybean miRNAs (Table 3.9). Targets of known soybean miRNAs that had an inverse expression patterns in nodule vs root tissue were annotated using G. max 1.89.

Target	Annotation
Root enriched	
Glyma.08G214000	Tetratricopeptide repeat (TPR)-like superfamily protein
Glyma.19G090600	Galactose oxidase/kelch repeat superfamily protein
Glyma.16G050500	auxin signaling F-box 3
Glyma.16G050500	auxin signaling F-box 3
Glyma.16G050500	auxin signaling F-box 3
Glyma.08G209700	Rubredoxin-like superfamily protein
Glyma.10G081700	BEL1-like homeodomain 1
Glyma.02G041600	F-box family protein
Nodule enriched	
Glyma.18G168900	sulfate transporter 2;1
Glyma.08G243600	cytochrome P450, family 716, subfamily A, polypeptide 1
Glyma.08G214000	Tetratricopeptide repeat (TPR)-like superfamily protein

3.4. Discussion

The objective of the study was global identification of miRNAs that might play roles in nodule development through spatial regulation of their target genes between nodule and adjacent root tissues. To meet this objective, small RNA and PARE libraries of MN and ABMN tissues were generated. Sequencing of ABMN and MN small RNA libraries generated 29 and 55 million reads, respectively, and PARE libraries generated 34 to 21 million reads, respectively. After stringent processing and filtering, we identified 497 precursors, 330 of which were not reported previously. Identification of a high number of previously unreported precursors, instead of previously reported miRNAs in soybean root nodules (Wang et al. 2009, Joshi et al. 2010), could be due to differences in depth of small RNA libraries used for the prediction of the miRNAs between these studies. Analysis of 462 potential miRNA mature sequences associated with these precursors showed that these sequences were dominated by 21-nt long reads and that uridine (U) was the most abundant first nt. Considering that AGO1 recruits miRNAs with 5' U, this provides strong evidence that these are miRNAs (Baumberger and Baulcombe 2005, Mi et al. 2008). Analysis of differential expression of the potential mature sequences showed that many of the previously unreported sequences were nodule-enriched. However, analysis of differential expression of known soybean miRNAs showed a higher number of root-enriched miRNAs. This suggests that previously unreported mature sequences might have important functional roles in nodule development.

Annotation of targets of potential miRNA mature sequences were validated by PARE/degradome (Addo-Quaye et al. 2008, German et al. 2008) and showed that many of these belong to TF families such as NAC, MYB, GRAS, WRKY, suggesting their possible role in nodule development. However, further experimentation is

required to pinpoint their specific function in nodule development. Differential expression of potential miRNA mature sequence/target pairs between the MN and ABMN tissues in the small RNA and transcriptome libraries were compared, resulting in the identification of 10 pairs with inverse expression patterns between them. Further categorization based on expression resulted into 6 root-enriched mature sequence/nodule-enriched target pairs and 4 nodule-enriched mature sequence/root-enriched target pairs. A search for inverse expression patterns of known soybean miRNA/target pairs in small RNA and transcriptome libraries of MN and ABMN tissues also revealed 8 of these pairs, 6 of which were root-enriched known miRNA/nodule-enriched target and 2 of which were nodule-enriched known miRNA/root-enriched target pairs. These inverse miRNA/target expression patterns identified in this study strongly suggests that spatial regulation of the miRNA, and thus the target, in nodule and surrounding tissues is one mechanism of nodule-specific gene expression. Prior to this study, only one miRNA was known to function through spatial regulation: miR169 control of MtHAP2-1 expression in meristematic tissue of root nodule of *M. truncatula* (Combier et al. 2006). Our study provides *in silico* evidence of the spatial regulation of miRNAs, and through them their targets, in nodule and nodule-adjacent root tissue.

3.5. Conclusion

Comparative analyses of small RNA libraries from nodule and nodule-adjacent root tissue identified 330 previously unreported miRNA precursors. PARE validation of the targets identified 20 novel miRNA/target pairs. Further analysis of expression patterns of the miRNAs in the small RNA libraries and their corresponding targets in the transcriptome libraries of nodule and adjacent tissues identified 10 previously unreported miRNA mature sequence/target pairs and 8

known soybean miRNA/target pairs with inverse expression patterns in the corresponding libraries. Functional analyses of the identified miRNA mature sequence/target pairs will help to shed light on their roles in nodule development.

References

Addo-Quaye, C., Eshoo, T.W., Bartel, D.P. and Axtell, M.J. (2008) Endogenous siRNA and miRNA Targets Identified by Sequencing of the Arabidopsis Degradome. *Current Biology*, 18, 758-762.

Axtell, M.J. (2013) Classification and comparison of small RNAs from plants. *Annual review of plant biology*, 64, 137-159.

Axtell, M.J., Westholm, J.O. and Lai, E.C. (2011) Vive la difference: biogenesis and evolution of microRNAs in plants and animals. *Genome Biol*, 12, 221.

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

Baumberger, N. and Baulcombe, D.C. (2005) Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 11928-11933.

Bhuvaneswari, T.V., Turgeon, B.G. and Bauer, W.D. (1980) Early Events in the Infection of Soybean (Glycine max L. Merr) by Rhizobium japonicum: I. LOCALIZATION OF INFECTIBLE ROOT CELLS. *Plant Physiology*, 66, 1027-1031.

Boualem, A., Laporte, P., Jovanovic, M., Laffont, C., Plet, J., Combier, J.P., Niebel, A., Crespi, M. and Frugier, F. (2008) MicroRNA166 controls root and nodule development in Medicago truncatula. *The Plant journal: for cell and molecular biology*, 54, 876-887.

Chen, X. (2005) microRNA biogenesis and function in plants. *FEBS Letters*, 579, 5923-5931.

Chen, X. (2009) Small RNAs and their roles in plant development. *Annual review of cell and developmental biology*, 25, 21-44.

Chen, X. (2010) Small RNAs - secrets and surprises of the genome. *The Plant journal: for cell and molecular biology*, 61, 941-958.

Combier, J.P., Frugier, F., de Billy, F., Boualem, A., El-Yahyaoui, F., Moreau, S., Vernie, T., Ott, T., Gamas, P., Crespi, M. and Niebel, A. (2006) MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in Medicago truncatula. *Genes & development*, 20, 3084-3088.

Cuperus, J.T., Fahlgren, N. and Carrington, J.C. (2011) Evolution and functional diversification of MIRNA genes. *Plant Cell*, 23, 431-442.

D'Haeseleer, K., Den Herder, G., Laffont, C., Plet, J., Mortier, V., Lelandais-Briere, C., De Bodt, S., De Keyser, A., Crespi, M., Holsters, M., Frugier, F. and Goormachtig, S. (2011) Transcriptional and post-transcriptional regulation of a NAC1 transcription factor in Medicago truncatula roots. *The New phytologist*, 191, 647-661.

De Luis, A., Markmann, K., Cognat, V., Holt, D.B., Charpentier, M., Parniske, M., Stougaard, J. and Voinnet, O. (2012) Two microRNAs linked to nodule infection and nitrogen-fixing ability in the legume Lotus japonicus. *Plant Physiol*, 160, 2137-2154.

Devers, E.A., Branscheid, A., May, P. and Krajinski, F. (2011) Stars and Symbiosis: MicroRNA- and MicroRNA*-Mediated Transcript Cleavage Involved in Arbuscular Mycorrhizal Symbiosis. *Plant Physiology*, 156, 1990-2010.

Frankow-Lindberg, B.E. and Dahlin, A.S. (2013) N2 fixation, N transfer, and yield in grassland communities including a deep-rooted legume or non-legume species. *Plant Soil*, 370, 567-581.

German, M.A., Pillay, M., Jeong, D.H., Hetawal, A., Luo, S., Janardhanan, P., Kannan, V., Rymarquis, L.A., Nobuta, K., German, R., De Paoli, E., Lu, C., Schroth, G., Meyers, B.C. and Green, P.J. (2008) Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nature biotechnology*, 26, 941-946.

Graham, P.H. and Vance, C.P. (2003) Legumes: Importance and Constraints to Greater Use. *Plant Physiology*, 131, 872-877.

Jones-Rhoades, M.W. and Bartel, D.P. (2004) Computational Identification of Plant MicroRNAs and Their Targets, Including a Stress-Induced miRNA. *Molecular Cell*, 14, 787-799.

Joshi, T., Yan, Z., Libault, M., Jeong, D.H., Park, S., Green, P.J., Sherrier, D.J., Farmer, A., May, G., Meyers, B.C., Xu, D. and Stacey, G. (2010) Prediction of novel miRNAs and associated target genes in Glycine max. *BMC bioinformatics*, 11 Suppl 1, S14.

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

Kakrana, A., Hammond, R., Patel, P., Nakano, M. and Meyers, B.C. (2014) sPARTA: a parallelized pipeline for integrated analysis of plant miRNA and cleaved mRNA data sets, including new miRNA target-identification software. 42, e139.

Kasschau, K.D., Fahlgren, N., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A. and Carrington, J.C. (2007) Genome-wide profiling and analysis of Arabidopsis siRNAs. *PLoS biology*, 5, e57.

Katiyar-Agarwal, S. and Jin, H. (2010) Role of small RNAs in host-microbe interactions. *Annual review of phytopathology*, 48, 225-246.

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

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

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

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

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

Lelandais-Briere, C., Naya, L., Sallet, E., Calenge, F., Frugier, F., Hartmann, C., Gouzy, J. and Crespi, M. (2009) Genome-wide Medicago truncatula small

RNA analysis revealed novel microRNAs and isoforms differentially regulated in roots and nodules. *Plant Cell*, 21, 2780-2796.

Li, H., Deng, Y., Wu, T., Subramanian, S. and Yu, O. (2010) Misexpression of miR482, miR1512, and miR1515 increases soybean nodulation. *Plant Physiol*, 153, 1759-1770.

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

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

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

McHale, N.A. and Koning, R.E. (2004) MicroRNA-Directed Cleavage of Nicotiana sylvestris PHAVOLUTA mRNA Regulates the Vascular Cambium and Structure of Apical Meristems. *The Plant Cell*, 16, 1730-1740.

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

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

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

Raman, S., Greb, T., Peaucelle, A., Blein, T., Laufs, P. and Theres, K. (2008) Interplay of miR164, CUP-SHAPED COTYLEDON genes and LATERAL SUPPRESSOR controls axillary meristem formation in Arabidopsis thaliana. *The Plant journal: for cell and molecular biology*, 55, 65-76.

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

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

Rogers, K. and Chen, X. (2013) Biogenesis, Turnover, and Mode of Action of Plant MicroRNAs. *The Plant Cell*, 25, 2383-2399.

Schwartz, W. (1972) J. M. Vincent, A Manual for the Practical Study of the Root-Nodule Bacteria (IBP Handbuch No. 15 des International Biology Program, London). XI u. 164 S., 10 Abb., 17 Tab., 7 Taf. Oxford-Edinburgh 1970: Blackwell Scientific Publ., 45 s. *Zeitschrift für allgemeine Mikrobiologie*, 12, 440-440.

Subramanian, S., Fu, Y., Sunkar, R., Barbazuk, W.B., Zhu, J.-K. and Yu, O. (2008) Novel and nodulation-regulated microRNAs in soybean roots. *BMC Genomics*, 9, 160.

Sun, X., Zhang, Y., Zhu, X., Korir, N.K., Tao, R., Wang, C. and Fang, J. (2014) Advances in identification and validation of plant microRNAs and their target genes. *Physiologia plantarum*, 152, 203-218.

Turner, M., Nizampatnam, N.R., Baron, M., Coppin, S., Damodaran, S., Adhikari, S., Arunachalam, S.P., Yu, O. and Subramanian, S. (2013) Ectopic Expression of miR160 Results in Auxin Hypersensitivity, Cytokinin Hyposensitivity, and Inhibition of Symbiotic Nodule Development in Soybean. *Plant Physiology*, 162, 2042-2055.

Turner, M., Yu, O. and Subramanian, S. (2012) Genome organization and characteristics of soybean microRNAs. *BMC genomics*, 13, 169.

Wang, Y., Li, P., Cao, X., Wang, X., Zhang, A. and Li, X. (2009) Identification and expression analysis of miRNAs from nitrogen-fixing soybean nodules. *Biochemical and biophysical research communications*, 378, 799-803.

Xie, M., Zhang, S. and Yu, B. (2014) microRNA biogenesis, degradation and activity in plants. *Cellular and Molecular Life Sciences*, 72, 87-99.

Yan, Z., Hossain, M.S., Valdes-Lopez, O., Hoang, N.T., Zhai, J., Wang, J., Libault, M., Brechenmacher, L., Findley, S., Joshi, T., Qiu, L., Sherrier, D.J., Ji, T., Meyers, B.C., Xu, D. and Stacey, G. (2016) Identification and functional characterization of soybean root hair microRNAs expressed in response to Bradyrhizobium japonicum infection. *Plant biotechnology journal*, 14, 332-341.

Zhai, J., Arikit, S., Simon, S.A., Kingham, B.F. and Meyers, B.C. (2014) Rapid construction of parallel analysis of RNA end (PARE) libraries for Illumina sequencing. *Methods (San Diego, Calif.)*, 67, 84-90.

Zhang, B., Pan, X., Cobb, G.P. and Anderson, T.A. (2006) Plant microRNA: A small regulatory molecule with big impact. *Developmental Biology*, 289, 3-16.

Chapter 4

4. Method Development for microRNA quantification

This chapter has been published in two journals:

Sajag Adhikari, Marie Turner, Senthil Subramanian (2013) Hairpin priming is better suited than *in vitro* polyadenylation to generate cDNA for plant miRNA qPCR. Mol Plant. 2013 Jan; 6(1):229-31. doi: 10.1093/mp/sss106. Epub 2012 Sep 30

Marie Turner, Sajag Adhikari & Senthil Subramanian (2013) Optimizing stem-loop qPCR assays through multiplexed cDNA synthesis of U6 and miRNAs, Plant Signaling & Behavior, 8:8, e24918, DOI: 10.4161/psb.24918

Abstract

microRNAs (miRNAs) are short, non-coding regulatory RNAs that play crucial roles in plant development and responses to environmental cues. Reliable and accurate measurements of miRNA abundance by Northern hybridization and qPCR have significantly contributed to our understanding of miRNA function. Two most commonly used methods to generate cDNAs for miRNA qPCR are (i) "polyA tailing", where miRNAs are polyadenylated *in vitro* followed by reverse transcription using oligo-dT adapters and (ii) "Hairpin priming" or "stem-loop RT", where a hairpin primer with a 3' overhang complementary to the 3' end of the miRNA is used in reverse transcription. We reasoned that inhibition of polyadenylation by 2'-O-methylation in 3' ends of plant miRNAs might render the polyA tailing method less-suited to generate cDNA for plant miRNA qPCR assays. We examined the ability of the above two methods to assay (i) Synthetic RNA molecules with and without 2'-O-Me modification, (ii) Selected miRNAs in wild-type *Arabidopsis* and the small RNA methylation-deficient mutant, *hen1*, (iii) Selected miRNAs in respective *Arabidopsis* over-expression lines and (iv) miR164 in soybean roots during the course of nodulation. Our results suggest that hairpin priming is better suited than polyA tailing to generate cDNA for plant miRNA qPCR. We also designed primers to multiplex hairpin cDNA synthesis for normalization controls, U6 (all plant species) and miR1515 (tropical legumes), together with other miRNAs.

4.1. Introduction

miRNAs were first identified in *Caenorhabditis elegans* (Lee et al. 1993, Wightman et al. 1993) and subsequently in a number of other eukaryotes. They are short non-coding RNAs that play crucial roles in regulating gene expression during developmental processes and in response to biotic and abiotic stresses, in both animals and plants (reviewed by (Chen 2009, Ruiz-Ferrer and Voinnet 2009, Subramanyam and Blelloch 2011, Alonso 2012, Khraiwesh et al. 2012, Mendell and Olson 2012, Sunkar et al. 2012). miRNA biogenesis starts with transcription of miRNA gene leading to the formation of pri-miRNA transcripts which are 5' capped and generally 3' poly-adenylated. These molecules are first cleaved into minimal hairpin structures called pre-miRNAs and subsequently into short duplex RNA molecules consisting of the mature miRNA (typically 20-24nt in length) and miRNA*. In animals, two different RNase III proteins, DROSHA and DICER, catalyze these cleavages respectively; in plants, DICER-LIKE 1 sequentially performs both these cleavages. In addition, plant miRNA duplexes (i.e. both the miRNA and miRNA*) are O-methylated at the 2' position of the 3' end nucleotide (See below). Typically, mature miRNAs are the active forms and are incorporated into different RNA-induced silencing complexes. They can regulate gene expression by inhibition of translation or cleavage of complementary mRNAs (reviewed by (Bartel 2004, Ghildiyal and Zamore 2009, Voinnet 2009).

The identification of conserved and species-specific miRNA families in a number of different species has raised an interest in the understanding of their function in growth, development and responses to external cues. The first step towards such an understanding is the reliable quantification of the mature and active forms of miRNAs in different tissues and in response to developmental and

environmental cues. Large-scale methods such as high throughput sequencing and miRNA microarrays are used primarily for miRNA discovery and global expression analyses. However, functional characterization necessitates assaying the expression of individual miRNAs in a more detailed manner. Northern hybridization has routinely been used for this purpose and is perhaps the gold standard in miRNA expression analysis. However, the technique requires large quantities of RNA, the use of radioactivity, and is generally less sensitive (Table 4.1). qPCR assays are more sensitive than Northern assays; but, mature miRNAs are too short to serve as templates in a classical RT-qPCR assay. Therefore, several modified methods for reverse transcription of mature miRNAs into cDNAs have been successfully developed and used for reliable quantification of miRNAs by qPCR (reviewed by (Benes and Castoldi 2010, Varkonyi-Gasic and Hellens 2010, Chen et al. 2011, Hurley et al. 2012). Two of the most commonly used methods for cDNA synthesis from mature miRNAs are (i) "polyA-tailing method" which involves *in vitro* polyadenylation followed by oligo dT-mediated reverse transcription (Figure 4.1; (Shi and Chiang 2005, Shi et al. 2012)) and (ii) "hairpin priming method" (or "stem-loop RT") which involves reverse transcription using a hairpin primer that is complementary to the 3' end of miRNA (Figure 4.1; (Chen et al. 2005, Fu et al. 2006)). cDNA generated by either method can be used in SYBR® green or Taqman® qPCR assays. Hairpin-priming method is more specific and requires individual cDNA synthesis reaction (or reverse transcription primer) for each miRNA. On the other hand, cDNA generated by the polyA tailing method can be used to assay the expression of multiple miRNAs or even their target genes (Table 4.1).

Table 4.1 Major advantages and disadvantages of methods used to assay mature miRNA levels in plants.

Method	**Northern hybridization**	**polyA cDNA qPCR**	**Hairpin cDNA qPCR**
Advantages	Well-established method Direct measurement of RNA by hybridization Can identify the size of mature miRNAs (e.g. 21 vs 22nt)	Relatively more sensitive A single cDNA synthesis reaction can be used to assay the expression of multiple miRNAs and even their targets Might be able to distinguish closely related family members	Relatively more sensitive Might be able to distinguish closely related family members Probably not influenced by the methylation status of miRNAs (*examined in this study*)
Disadvantages	Relatively less sensitive and requires large amounts of RNA Might not efficiently distinguish closely related family members Involves the use of radioactivity Difficult to multiplex	Requires extensive standardization (e.g. linearity assays) In vitro poly-adenylation might be less efficient on RNA molecules with 2'O-methylation at 3' end (Li et al. 2005) (*examined in this study*)	Requires extensive standardization (e.g. linearity assays) Separate cDNA synthesis for each miRNA to be assayed including normalization controls (*suitability of multiplexing examined in this study*)

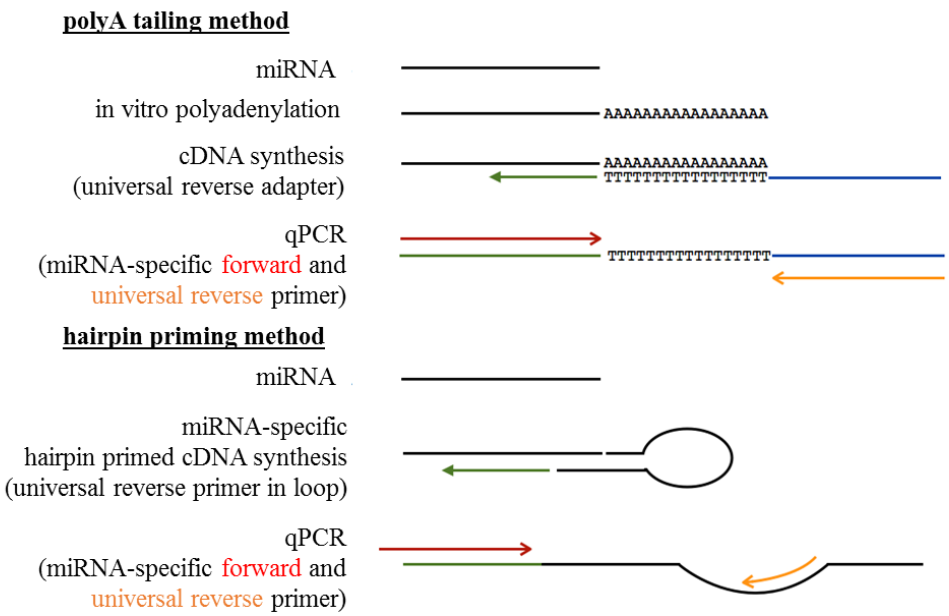


Figure 4.1 Schematic showing the principle of two methods in this study to generate cDNAs from mature miRNAs. In polyA cDNA qPCR, in vitro polyadenylation of mature miRNAs followed by oligo dT- primed cDNA synthesis allows subsequent detection by qPCR [Shi and Chiang (2005) BioTechniques 39:519]. In hairpin (or stem loop) cDNA qPCR, cDNAs are generated from mature miRNAs using specific hairpin primers with 3' overhang complementary to 3' end (6nt) of mature miRNAs and subsequently detected by qPCR [Chen et al (2005) NAR 33(20):179]

The hairpin priming method was initially developed for animal miRNAs, but later adapted and validated to assay plant miRNAs (Varkonyi-Gasic et al. 2007a). A number of studies have also used polyA tailing method to assay the levels of plant miRNAs. However, we reasoned that 2'-*O* methylation on the 3'-ends of plant miRNAs and the potential inhibition of polyadenylation by such methylation might warrant a systematic examination of the suitability of this method to generate cDNA for plant miRNA qPCR. Plant miRNAs possess a 2'-*O*-methyl modification on the ribose of their 3' termini catalyzed by HUA ENHANCER 1 (HEN1), a small RNA methyltransferase (first identified in *Arabidopsis* (Yu et al. 2005, Yang et al. 2006, Ji and Chen 2012)). This methyl modification protects mature miRNAs against polyuridylation by HEN1 SUPRESSOR1 (HESO1) and subsequent degradation (Ren et al. 2012, Zhao et al. 2012). This modification prevents the addition of nucleotides to the 3'-end of mature plant miRNAs both *in vitro* and *in vivo* (Li et al. 2005, Yu et al. 2010). We systematically examined the suitability of polyA-tailing method compared to hairpin priming method, to assay plant miRNAs. Our results indicate that while the polyA method is more sensitive, hairpin priming method is better suited to examine plant miRNA expression due to its ability to specifically detect methylated miRNAs.

Another crucial factor that determines the accuracy and reliability of qPCR assays is the "house-keeping" gene(s) used for normalization (Peltier and Latham 2008, D'Haene et al. 2012). In general, other small endogenous noncoding RNAs such as U6, U24 or U26 are used to normalize the expression of both plant and animal miRNAs (D'Haene et al. 2012). It is crucial that the reference genes and test miRNAs undergo identical reaction conditions during cDNA synthesis and qPCR. This would ensure equal efficiency of the reaction between the reference gene and the test

miRNA(s) leading to accurate and reliable measurements. The polyA-tailing method achieves this by using a single cDNA synthesis reaction for both reference genes and test miRNAs. On the other hand, the hairpin priming method typically involves independent cDNA synthesis for test and reference miRNAs and might suffer from inaccuracy due to differences in cDNA synthesis reaction efficiencies. Methods for large scale megaplex reactions to assay multiple miRNAs with small amounts of input RNA have been developed (Mestdagh et al. 2008). We designed primers to assay and utilize U6 as reference gene for hairpin cDNA qPCR and examined the suitability of small scale multiplexing (4 miRNAs + reference gene) in a single reaction to ensure equal amplification efficiency of test miRNAs and house-keeping genes. In addition, we have examined and developed the use of miR1515 as a suitable reference gene for miRNA qPCR assays, especially for soybean and potentially other leguminous plants.

4.2. Results and Discussion

4.2.1. Hairpin and polyA cDNA qPCR assays differ in their efficiencies in detecting 2'-OMe and 2'OH RNA molecules

We compared the ability of two most commonly used cDNA synthesis methods for miRNA qPCR to efficiently assay 2'-O methylated RNA molecules since a number of plant miRNAs are known to carry this modification (Yu et al. 2005, Yang et al. 2006, Yu et al. 2010). Synthetic RNA molecules identical in sequence to the mature form of gma-miR393a (the sequence of which is highly conserved among multiple plant species) with and without 2'-O methylation at the 3' ends nucleotide were synthesized. We will refer to these as "miR393-2'OMe" and "miR393-2'OH" respectively. cDNAs were synthesized from identical quantities of miR393-2'OH and miR393-2'OMe molecules using the polyA-tailing method ("polyA cDNA") or

hairpin priming method ("hairpin cDNA"). The efficiency of cDNA synthesis was measured using SYBR green-based qPCR assays (Figure 4.2).

For each cDNA synthesis method, we made pair-wise comparison of Ct values between reactions containing cDNA synthesized from identical amounts of miR393-2'OH or miR393-2'OMe RNA template (Figure 4.2, Table 4.2). In qPCR assays performed using hairpin cDNA, there was virtually no difference in Ct values (P =0.87; Student's t-test) between qPCR assays using miR393-2'OH or miR393-2'OMe (Compare Figure 4.2 A vs. C, Table 4.2). In contrast, in qPCR assays performed using polyA cDNA, there was a difference of at least 8 Ct values (P<0.0001; Student's t-test) between reactions containing the two types of RNA molecules (Compare Figure 4.2 B vs. D; Table 4.2). The high Ct values in assays using miR393-2'OMe compared to miR393-2'OH indicated that polyA tailing based cDNA synthesis was less efficient on 2'-OMe RNA molecules. This is consistent with the observation that 2'-OMe inhibits *in vitro* polyadenylation of miRNAs (Li et al. 2005).

We also used 90:10, 50:50 and 10:90 mixtures of 2'-OH and 2'-OMe miR393 molecules as templates in cDNA synthesis reactions using the two methods. Hairpin cDNA qPCR assays showed minimal or no difference in Ct values among these reactions (Figure 4.3 A; Variance =0.38). However, polyA cDNA qPCR assays containing more 2'-OH RNA molecules had significantly lower Ct values (and vice versa) compared to those containing equal mixtures of 2'-OH and 2'-OMe RNA molecules (Figure 4.3 B; Variance =3.04). Taken together, our results suggest that hairpin cDNA qPCR was equally efficient in assaying both 2'-OH and 2'-OMe RNA molecules, while polyA cDNA qPCR was less efficient in assaying 2'-OMe RNA molecules. However, it should be noted that polyA cDNA PCR appeared to be more

sensitive (2’-OH RNA molecules) than or as sensitive (2’-OMe RNA molecules) as hairpin cDNA method based on raw Ct values (Figure 4.2 and Table 4.2).

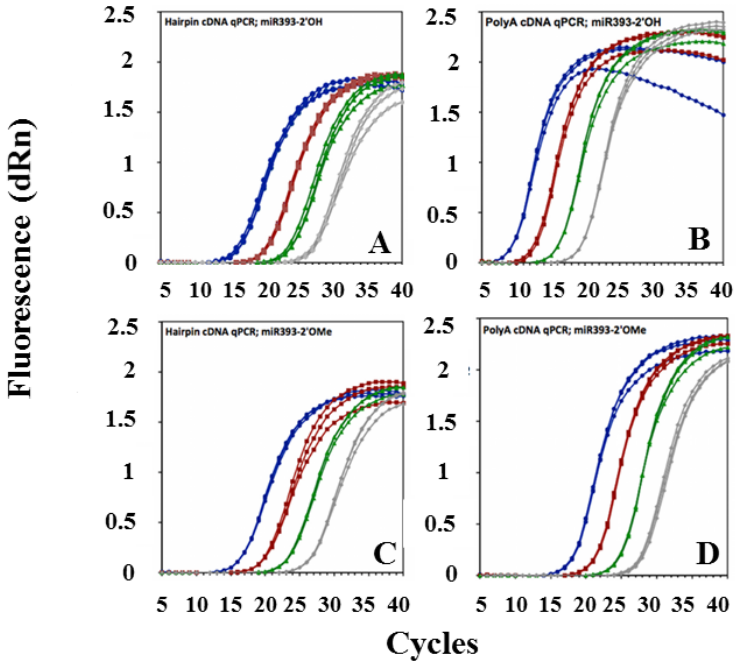


Figure 4.2 Comparison of hairpin cDNA qPCR and polyA cDNA qPCR to assay synthetic RNA molecules with and without 2'-O methylation. Amplification plots from hairpin (A & C) and polyA (B & D) cDNA qPCR assays to detect different amounts of 2'-OH and 2'-OMe RNA molecules (1 fmole, blue circles; 0.1 fmole, red squares; 0.01 fmole, green triangles and 0.001 fmole, grey diamonds).

Table 4.2 Ct value from qPCR assays using cDNA synthesized from different amounts of 2'-OH and 2'-OMe RNA molecules.

RNA equivalent in qPCR assay	*Hairpin cDNA qPCR*		*PolyA cDNA qPCR*	
	miR393-2'OH	*miR393-2'OMe*	*miR393-2'OH*	*miR393-2'OMe*
1 fmole	*16.8 ± 0.2*	*17.0 ± 0.0*	*9.6 ± 0.0*	*18.3 ± 0.1*
0.1 fmole	*20.9 ± 0.1*	*20.5 ± 0.1*	*12.9 ± 0.2*	*21.5 ± 0.1*
0.01 fmole	*24.2 ± 0.3*	*23.8 ± 0.1*	*16.4 ± 0.0*	*24.9 ± 0.1*
0.001 fmole	*27.6 ± 0.2*	*27.1 ± 0.1*	*19.7 ± 0.0*	*28.0 ± 0.2*
Linearity	*0.996*	*0.998*	*0.993*	*0.981*

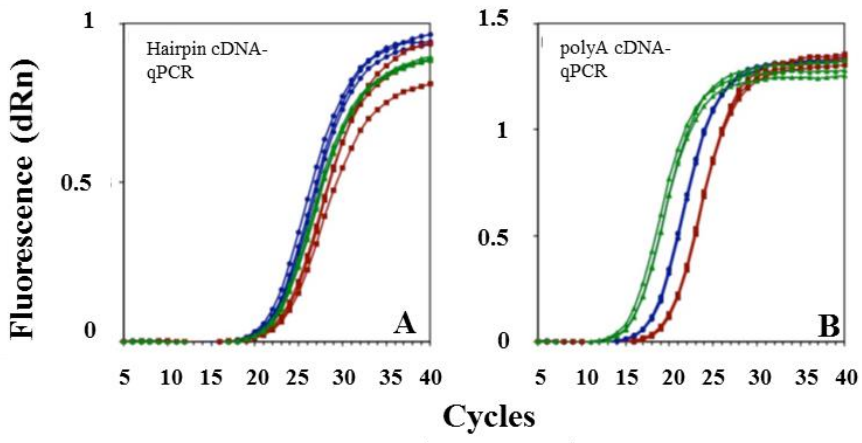


Figure 4.3 Amplification plots from hairpin (A) and polyA (B) cDNA qPCR assays. qPCR assays were conducted using mixtures of different ratios of 2'-OH and 2'-OMe RNA molecules (9 parts2'-OH: 1 part 2'-OMe, green triangles; 1 part 2'-OH: 1 part 2'-OMe, blue Circles; 1 part 2'-OH: 9 parts 2'-OMe, red squares). The total RNA in all cDNA synthesis reactions were identical (100pmoles) and 0.1fmole equivalent was present in each qPCR assay.

4.2.2. polyA cDNA qPCR might result in erroneous abundance measurements of plant miRNAs

We examined the efficiency of these methods in assaying miRNAs in plant RNA preparations from wild-type and the 2'O-methylation deficient *hen1*(Chen et al. 2002, Park et al. 2002, Yu et al. 2005) Arabidopsis seedlings. The *hen1* mutant accumulates very low levels of mature miRNAs (Park et al. 2002, Yu et al. 2005) and those that accumulate have additional nucleotides of varying length at the 3' end (Li et al. 2005) presumably rendering them inactive. We assayed the levels of miR160, 164, 166 and 393 in wild-type and *hen1* seedlings using hairpin and polyA cDNA qPCR. Results from hairpin cDNA qPCR indicated that the levels of all four miRNAs were lower in *hen1* seedlings compared to wild-type (Figure 4.4 A) as previously reported (Park et al. 2002). However, results from polyA cDNA qPCR indicated that miR160, 166 and 393 accumulated at higher levels in *hen1* compared to Col (Figure 4.4 A). We reasoned that the polyA tailing method perhaps did not distinguish between the canonical miRNA molecules and those with additional nucleotides in the 3'end resulting in artificially higher abundance in *hen1*. Indeed, when we examined the dissociation curves obtained from these qPCR assays, we observed broader peaks (Figure 4.5) indicating the presence of amplicons from templates of different lengths or nucleotide composition. These observations suggested that the use of cDNA generated through polyA tailing method in qPCR assays, might lead to erroneous miRNA abundance measurements. On the other hand, since hairpin priming depends on the last 6 nts at the 3' end of mature miRNAs, one can conclude that the method could distinguish canonical miRNA molecules from the modified molecules in *hen1* RNA preparations.

We also examined miRNA expression in three *Arabidopsis* transgenic lines over-expressing miR160 (Wang et al. 2005), 164 (Mallory et al. 2004) or 393 (Parry et al. 2009) using polyA cDNA qPCR and hairpin cDNA qPCR assays. Results from hairpin cDNA qPCR indicated that the respective miRNAs were expressed at significantly higher levels in all three over-expression (OX) lines compared to Col (Figure 4.4 B). However, polyA cDNA qPCR indicated significantly lower levels of miR393 in the OX line compared to Col (Figure 4.4 B) while being able to detect over-expression of miR160 (levels comparable to hairpin cDNA qPCR) and 164 (levels lower than hairpin cDNA qPCR). It is tempting to speculate that the miRNA methylation machinery might have kept up with miR393 in the miR393OX line which had the least overexpression (3-5 fold), but not with miR160 and 164 in their respective OX lines that had ~25-60 old over-expression. The presence of sufficient amounts of unmethylated miRNA molecules might have enabled polyA cDNA qPCR to detect over-expression of the respective miRNAs in miR160 and 164 over-expression lines.

We also compared the accuracy and sensitivity of hairpin and polyA cDNA qPCR assays to Northern hybridization analysis. We used all three methods to assay the expression of miR164 along a time-course of *B. japonicum* inoculation in soybean roots (Figure 4.6). Due to the difference in ability of these methods to distinguish miRNA family members with different sequences, we chose miR164 for these assays. All known miR164 precursors produce identical mature miRNAs in soybean (Zhai et al. 2011, Turner et al. 2012). We examined agreement between each qPCR method and Northern analysis by calculating correlation coefficients between expression levels from these methods. The time course of miR164 expression assayed by hairpin cDNA qPCR was largely in agreement with Northern hybridization in both mock (R^2

=0.91) and *B. japonicum*-inoculated (R^2 =0.68) samples. However, expression pattern assayed by polyA cDNA qPCR was not in agreement with that of Northern hybridization (R^2 values of -0.53 and -0.29 for mock and *B.japonicum*-inoculated samples respectively). Therefore, qPCR using hairpin cDNA can detect miRNA expression as reliably as Northern hybridization using much lower amounts of total RNA consistent with results obtained by Varkonyi-Gasic et al (Varkonyi-Gasic et al. 2007a).

Together, these results suggested that hairpin cDNA qPCR is better suited to quantify canonical plant miRNAs and that the use of polyA cDNA qPCR might lead to erroneous abundance measurements, perhaps depending on the methylation status of mature miRNAs. It is emerging that miRNA methylation might act to protect them from degradation (Ren et al. 2012, Zhao et al. 2012) and presumably demethylation would mark them for polyuridylation and subsequent degradation. Therefore, (de)methylation might be a mechanism by which miRNA activity is regulated in plants. This makes it crucial to be able to specifically assay methylated miRNA molecules as they are the active forms. Therefore we concluded that hairpin cDNA qPCR is better suited to assay plant miRNAs compared to polyA cDNA-qPCR despite the latter's apparent higher sensitivity.

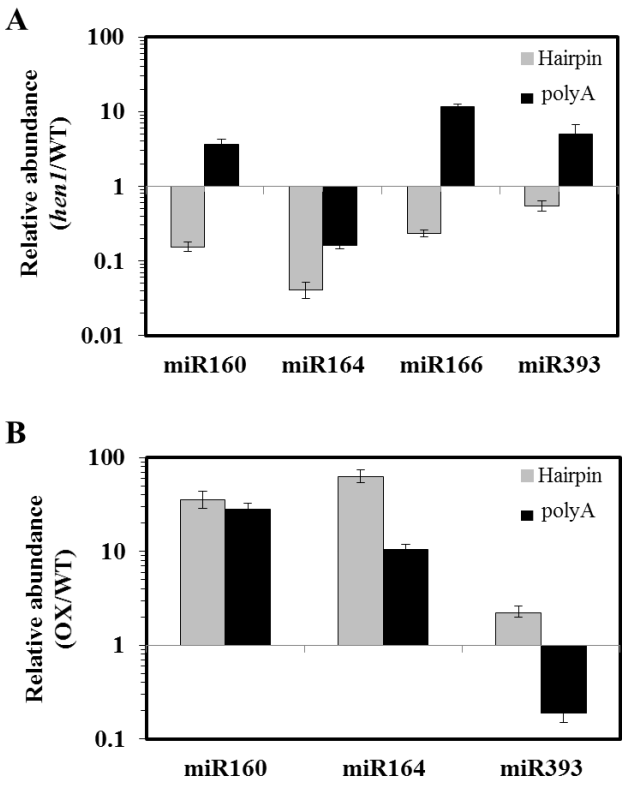


Figure 4.4 Comparison of hairpin cDNA qPCR and polyA cDNA qPCR to assay plant miRNAs. (A) Expression of miR160, 164, 166 and 393 in wild-type and hen1 Arabidopsis seedlings assayed by hairpin cDNA qPCR (shaded bars) and polyA cDNA qPCR (black bars). Data shown are the ratio of normalized miRNA abundance in hen1 over wild-type and representative of independent experiments. In all cases, expression in hen1 was significantly different from that in WT (P<0.01; Student's t-test). (B) Expression of miR160, 164 and 393 in wild-type and the respective miRNA overexpressing lines assayed by hairpin cDNA qPCR (shaded bars) and polyA cDNA qPCR (black bars). Data shown are the ratio of miRNA abundance in over-expressing lines over wild-type and representative of independent experiments. In all cases, expression in OX lines was significantly different from that in WT (P<0.01; Student's t-test).

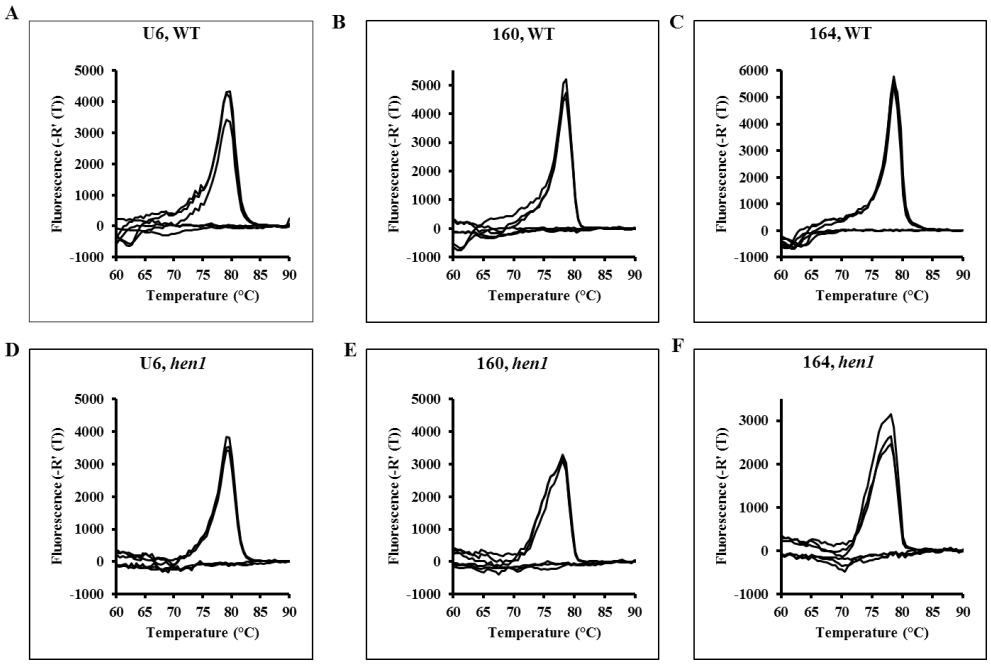


Figure 4.5 Dissociation curves from polyA-cDNA-qpCR assays. U6 (A and D), miR160 (B and E), miR164 (C and F), in wild type (A-C), and henl1 (D-F) mutants. The presence of leading shoulders in miRNA qPCR assays of hen1 samples (but not WT) suggested that amplicons of different lengths or nucleotide composition, probably due to uridylation and other modifications of non-methylated miRNA molecules in hen1. Note the dissociation curve of U6 assays were near identical between WT and hen1.

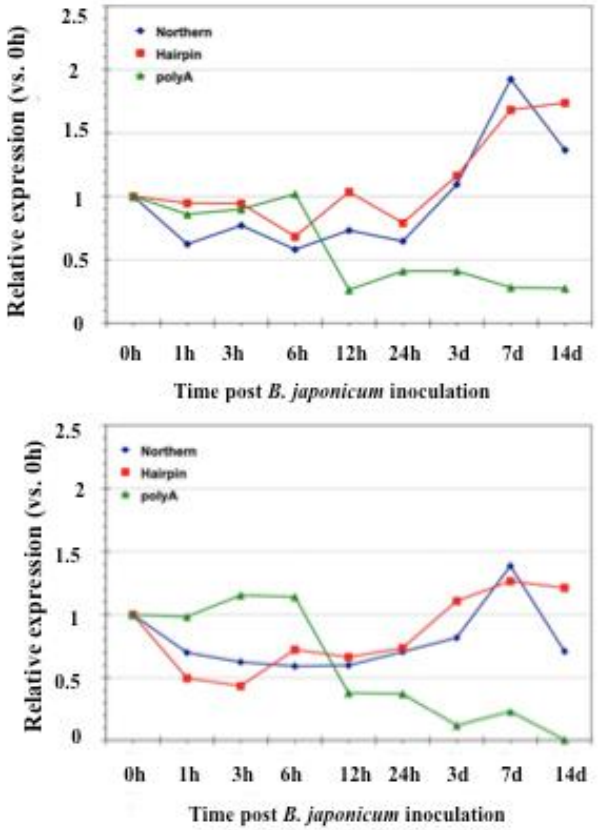


Figure 4.6 Relative abundance of miR164 (vs. 0h time-point) assayed by Northern hybridization (blue diamonds), hairpin cDNA qPCR (red squares) and polyA cDNA qPCR (green triangles) along a time-course of 14 days in (A) mock- and (B) B. japonicum-inoculated soybean roots.

4.2.3. Use of U6 and miR1515 as normalization controls in hairpin cDNA qPCR

In Northern hybridization, the small nucleolar RNA, U6 has been successfully used as a normalization control for a wide range of experiments (e.g. (Lelandais-Briere et al. 2009, Zhang et al. 2011). However, perhaps due to the specific nature of RT primers used in hairpin cDNA synthesis, protein-coding genes are often used as normalization controls in this method (e.g. (Varkonyi-Gasic et al. 2007a, Devers et al. 2011) rather than U6. Usually, cDNA synthesis for these normalization controls and the test miRNAs are performed in separate reactions. We designed RT primers to multiplex cDNA synthesis of U6 together with other miRNAs. We designed qPCR primers for U6 from a highly conserved region of the gene and added the 'universal' reverse primer used in hairpin cDNA qPCR (Figure 4.7) to the 5' end of U6-specific reverse primer (Figure 4.7). When this modified U6 reverse primer was used for cDNA synthesis, U6 can also be assayed using hairpin cDNA qPCR (U6-specific forward primer and universal reverse primer) along with other miRNA genes. The primers we designed were able to successfully detect U6 expression in 7 different plant species covering a range of monocots and dicots (both legumes and non-legumes) with reliable amplification efficiencies and linearity (Figure 4.8).

A

```
CLUSTAL 2.1 multiple sequence alignment


Oryza_sativa_cv_japonica      GTCTCTTCGGAGACATCCGATAAAATTGGAACGATACAGAGAAGATTAGC 50
Zea_mays                      GTCTCTTCGGAGACATCCGATAAAATTGGAACGATACAGAGAAGATTAGC 50
Triticum_aestivum             GCCCCTTCGGGGACATCCGATAAAATTGGAACGATACAGAGAAGATTAGC 50
Broad_bean                    ----CTTCGGGGACATCCGATAAAATTGGAACGACACAGAGAAGATTAGC 46
Medicago_truncatula           GTCCCTTTGGGGACATCTGATAAAATTGGAACGATACAGAGAAGATTAGC 50
Arabidopsis_thaliana          GTCCCTTCGGGGACATCTGATAAAATTGGAACGATACAGAGAAGATTAGC 50
                                  *** ** ****** ********** ********** **********

Oryza_sativa_cv_japonica      ATGGCCCCTGCGCAAGGATGACACGCACAAATCGAGAAATGGTCCAAATT 100
Zea_mays                      ATGGCCCCTGCGCAAGGATGACACGCACAAATCGAGAAATGGTCCAAATT 100
Triticum_aestivum             ATGGCCCCTGCGCAAGGATGACACGCACAAATCGAGAAATGGTCCAAATT 100
Broad_bean                    ATGGCCCCTGCGCAAGGATGACACGCACAAATCGAGAAATGGTCCAAATT 96
Medicago_truncatula           ATGGCCCCTGCGCAAGGATGACACGCACAAATCGAGAAATGGTCCAATTT 100
Arabidopsis_thaliana          ATGGCCCCTGCGCAAGGATGACACGCATAAATCGAGAAATGGTCCAAATT 100
                              ************************** ****************** **

Oryza_sativa_cv_japonica      TTT-------- 103
Zea_mays                      TTTTT------ 105
Triticum_aestivum             TTTTTGAGATT 111
Broad_bean                    TT--------- 98
Medicago_truncatula           TTT-------- 103
Arabidopsis_thaliana          TTT-------- 103
                              **
```

B

Primer Sequences:

U6 cDNA synthesis primer:	5'-GTGCAGGGTCCGAGGT**TTTGGACCATTTCTCGAT-3'**
U6 qPCR forward primer:	5'-**GGAACGATACAGAGAAGATTAGCA**-3'
Universal qPCR reverse primer:	5'-GTGCAGGGTCCGAGGT-3'

Figure 4.7 U6 primer design. A. Sequence alignment of U6 sequences from multiple plant species showing a conserved region from which qPCR primers were designed. U6 sequences from multiple plant species were obtained from non-coding RNA database (http://www.ncrna.org/frnadb/), NCBI and ncRNAdb. Sequences used to design "Forward" and "Reverse" primers are underlined. B. Novel U6 primer sequences designed in this study for multiplexed cDNA synthesis and qPCR. The universal qPCR primer used in this study is from Varkonyi-Gasic et al (2007).

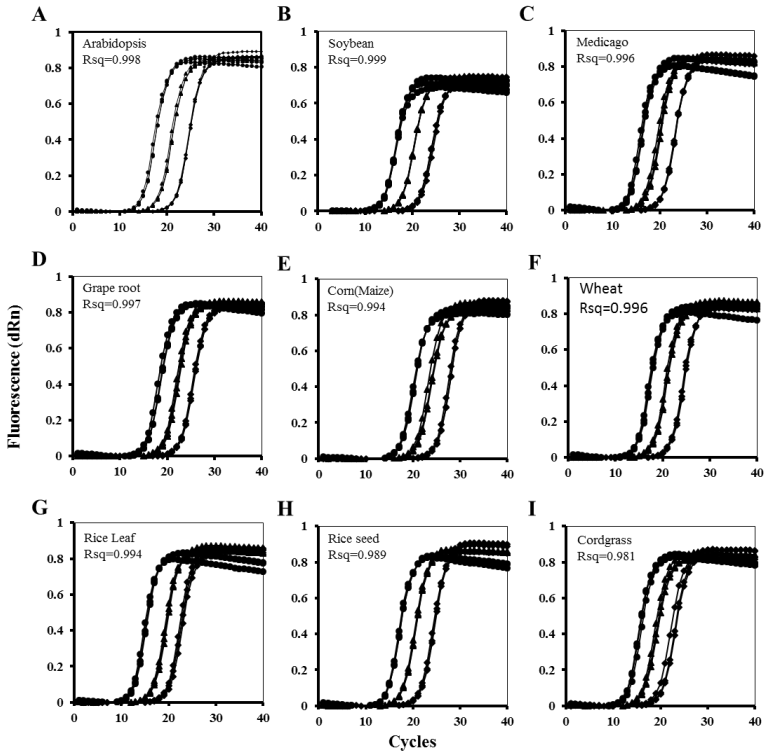


Figure 4.8 Validation of the novel U6 fusion primer for cDNA synthesis on multiple plant species. (A) Arabidopsis, (B) Soybean, (C) Medicago, (D) Grape root, (E) Corn, (F) Wheat, (G) Rice leaf, (H) Rice seed and (I) Prairie Cordgrass. Amplification plots of U6 in different plant species were obtained using primers designed in this study to adapt the use of U6 as normalization control in hairpin cDNA qPCR assays for miRNAs. Linearity (Rsq) values obtained by examining expression in different dilutions (circle: 1/50, triangle: 1/500 and diamond: 1/5000) of cDNA are indicated in each panel.

We also designed hairpin primers for the legume-specific miRNA, miR1515 (Zhai et al. 2011, Turner et al. 2012) that is uniformly expressed in different soybean tissues (Figure 4.9A) and in the roots along the course of nodulation (Figure (4.9B) as previously identified through Northern analysis (Li et al. 2010). We propose the use of miR1515 as normalization control for miRNA expression assays in soybean and other tropical legumes (e.g. common bean, cowpea, peanut). The use of a miRNA as normalization control for miRNA expression assays has some advantages over using protein coding genes. Since miRNAs biogenesis pathways are distinct from other RNAs, alterations in biosynthesis machinery (e.g. due to virus infection) might introduce artificial abundance difference between miRNAs and non-miRNA normalization controls. Using a miRNA as normalization control can help overcome such issues.

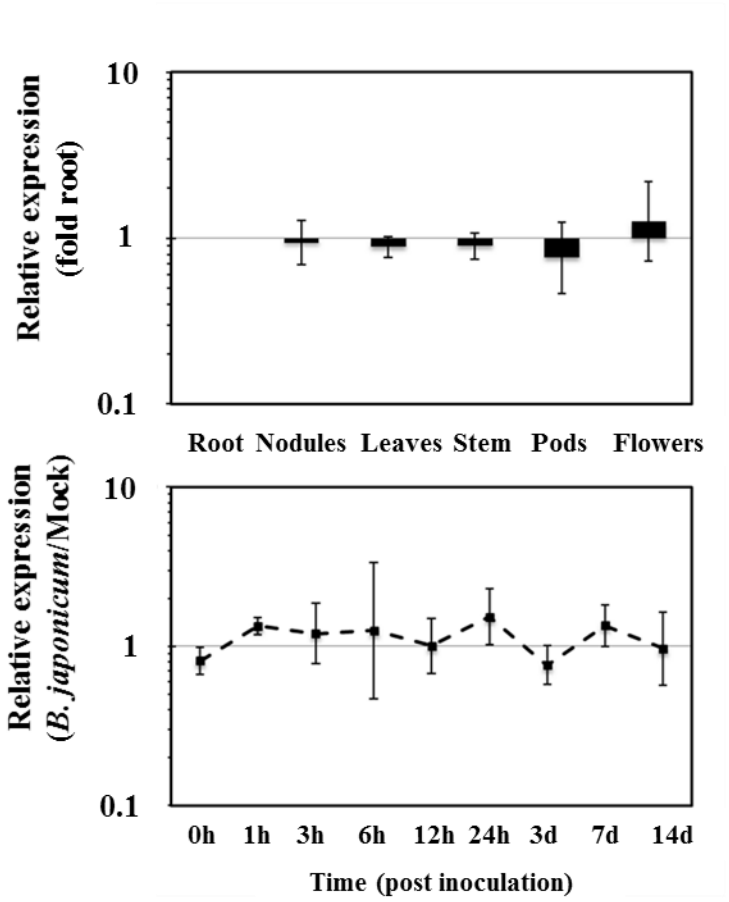


Figure 4.9 Expression of miR1515 in different soybean tissues and along the course of nodulation in soybean assayed by hairpin cDNA qPCR. (A) Relative expression compared to roots were calculated using average Ct values from four independent replicates (B) Relative expression at each time point post B. japonicum inoculation compared to mock inoculated roots. Data shown are average of four independent biological replicates. Error bars represent SD.

4.2.4. Multiplexing cDNA synthesis of miRNA and normalization controls for accurate and efficient qPCR assays

One disadvantage of independent cDNA synthesis concerning test miRNAs and normalization controls is possible differences in efficiency that contribute to differences in Ct values. We examined the suitability of U6 and miR1515 for multiplexing together with test miRNAs. We were able to multiplex up to five genes such as four miRNAs + miR1515 (Figure 4.10) or U6 (data not shown) without loss of efficiency. There was no significant difference in slope, intercept or the regression coefficients of standard curves (ANCOVA; $P>0.05$) between qPCR assays performed using independent or multiplexed cDNA synthesis reactions (Figure 4.10). Therefore, multiplexing during cDNA synthesis would not only enable the efficient use of minimal amounts of RNA for qPCR, but also enhance the accuracy of the assays by ensuring equal cDNA synthesis efficiency between tests and normalization controls.

There are other situations where multiplexing in hairpin cDNA qPCR might be handy. The ability of hairpin cDNA qPCR to distinguish family members or mature sequence variants can sometimes be a limitation e.g. when attempting to assay miRNA abundance of multiple/all family members. This could be addressed by multiplexing cDNA synthesis for multiple family members/variants in a single reaction using specific hairpin primers for each variant/family member. This limitation can also be addressed by the use of a recently developed method (miQPCR, reviewed in (Benes and Castoldi 2010)) where an RNA adapter is ligated to the 3' end of mature miRNAs and the ligated molecules are reverse transcribed using a primer complementary to the adapter sequence. However, this approach might as well suffer from the inability to distinguish methylated vs. non-methylated RNA molecules, especially in RNA preparations from plants.

We have conclusively demonstrated that hairpin cDNA qPCR is better suited for plant miRNA expression assays compared to polyA cDNA qPCR and developed a Microsoft Excel spreadsheet to easily design primers for this assay. In addition, we have designed RT primers to multiplex cDNA synthesis of normalization controls, U6 (all plant species) and miR1515 (tropical legumes), together with test miRNAs.

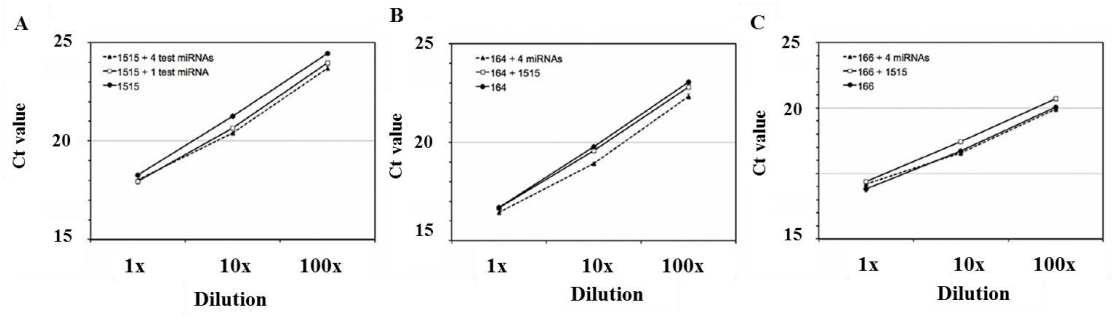


Figure 4.10 Multiplexing during hairpin cDNA synthesis did not affect assay efficiency. Ct values from hairpin cDNA qPCR assays to quantify (A) miR1515 (B) miR164 and (C) miR166 plotted using different dilutions of cDNAs obtained through independent (filled circles), duplex (open squares) or multiplex (filled triangles) synthesis reactions. Error bars indicate SD. In all cases, the slope of the curves or Ct values (at each dilution) did not differ significantly between different cDNAs.

4.3. Material and methods

4.3.1. Plant materials, growth conditions and treatments

Arabidopsis wild-type, *hen1-1* (obtained from ABRC, Columbus,OH: Stock# CS6583), miR160OX (a kind gift from Dr. X-Y Chen, Chinese Academy of Sciences, Shanghai, PRC), miR164OX (a kind gift from Dr. B Bartel, Rice University, Houston, TX) and miR393OX (a kind gift from Dr. M Estelle, University of California, San Diego, CA) seeds were surface-sterilized and grown in Sunshine mix #1 (Tessman Company, Sioux Falls, SD) under 16h light at 25°C. Soybean (*Glycine max* cv. Williams 82) seeds were surface-sterilized and grown essentially as described previously(Subramanian et al. 2008), using the following growth conditions (16h light; 25°C; 30% relative humidity). Pots were regularly watered with ¼ x (concentrated) nitrogen free plant nutrient solution (N^- PNS). *B. japonicum* (USDA110) was grown in Vincent's rich medium in the presence of chloramphenicol (20 mg/l) at 30°C and 200rpm. After ~4 days, cells (OD_{600}<1.0) were centrifuged (3500 × g for10 min @ 4 °C) and the pellet was resuspended in ¼ x N^- PNS to a final concentration of 1 x 10^8 cells/ml (OD_{600} =0.08). Three day-old soybean seedlings were inoculated (25 ml per pot) with this suspension (*B. japonicum*-inoculated) or N-PNS (mock-inoculated).

4.3.2. RNA isolation and synthesis

Total RNA extraction from plant samples was performed using TRI reagent (Sigma-Aldrich, St Louis, MO) essentially as described before (Schwab et al. 2005, Subramanian et al. 2008). RNA preparations were quantified using a Nanodrop spectrophotometer (Thermo Scientific, Waltham, MA) and checked for integrity by running an aliquot on an agarose gel before cDNA synthesis.

MiR393-2'OMe and miR393-2'OH molecules were synthesized by IDT Inc. (Ames, IA) and purity assessed by mass spectrometry (data not shown). Synthetic RNA molecules were resuspended in RNase free deionized water, quantified using a Nanodrop spectrophotometer and aliquots stored at -70 °C.

Tissue materials or RNA samples for detection of U6 in different plant species were obtained from collaborators and colleagues at SDSU. RNA preparations were obtained from Rice (leaf and seed), Wheat (leaf), Arabidopsis (seedling), Medicago (leaf), Corn (leaf meristematic tissue), Grape (root), Soybean (root), and Prairie cord grass (rhizome).

4.3.3. cDNA Synthesis and RT-qPCR

cDNA synthesis reactions were performed with 2 µg total RNA (plant samples) or 100 pmoles synthetic RNA (miR393-2'OMe/2'-OH) as starting template. polyA cDNA synthesis was performed using Mir-X first-strand synthesis kit (Product#638313 ; Clontech Mountain View, CA) according to the manufacturer's instructions (www.clontech.com). Hairpin cDNA synthesis (including RT primer design) was performed essentially as described previously (Varkonyi-Gasic et al. 2007b) using M-MuLV reverse transcriptase (NEB, Ipswich, MA). All cDNA synthesis reactions were performed on an ABI thermocycler (GeneAmp 9700).

qPCR assays were performed using a MX3000P thermocycler (Stratagene/Agilent Technologies, Santa Clara, CA) and SYBR Advantage qPCR premix (Clontech, Mountain view, CA, Cat no: 639676). Fold difference in miRNA expression between samples or between miRNAs and respective house-keeping genes were calculated using the ddCt method (Livak and Schmittgen 2001). Statistical

analyses for pair-wise comparison of Ct values was done using Student's t-test on Microsoft Excel. All other analyses were performed using SAS (www.sas.com).

4.3.4. Northern hybridization

RNA extracted from mock- and *B. japonicum*-inoculated roots at different time points (0 h, 1 h, 3 h, 6 h, 16 h, 24 h, 3 d, 7 d, and 14 d) was subjected to Northern hybridization (20 μg per lane). For detailed methods see (Schwab et al. 2005, Subramanian et al. 2008) using a radioactive probe against miR164. The same RNA samples (2 μg per reaction) were subject to hairpin or polyA cDNA qPCR to assay miR164. Expression levels were determined by normalizing band intensities (quantified using ImageJ; http://developer.imagej.net/) or Ct values for each time point against the initial time point (0h) for both mock- and *B. japonicum*-inoculated samples.

4.3.5. U6 and miR1515 as normalization control

Hairpin primed cDNAs were generated using U6-specific reverse primer (Figure 4.7) from RNA samples obtained from different plant species. Three different dilutions (1/5, 1/50, and 1/500) of cDNA from each plant species were examined for U6 expression by qPCR. Suitability of these primers to synthesize cDNA and assay U6 expression were determined by examining the linearity of amplification, quality of the amplification curves and dissociation curves. Multiplexing assays examining miR1515, miR164 and miR166 amplification efficiency were performed on soybean root cDNA preparations using primers described previously (Turner et al. 2012).

Acknowledgements

The authors thank ABRC (Columbus, OH), Drs Xiao-Ya Chen (Chinese Academy of Sciences, Shanghai, PRC), Bonnie Bartel (Rice University, Houston,

TX) and Mark Estelle (University of California, San Diego, CA) for different Arabidopsis genotypes/transgenic lines; Drs Bin Yu (University of Nebraska, Lincoln, NE), Jerome Verdier and Bikram Pant (Noble Foundation, Ardmore, OK) for comments on the manuscript; Drs Sharon Clay, Anne Fennel, Jose Gonzalez, Xing-You Gu, Jai Rohila and Yajun Wu (South Dakota State University, Brookings, SD) for plant materials or RNA preparations to test U6 primers and Drs Oliver Yu and Ying Deng (Donald Danforth Plant Science Center, St Louis, MO) for help with Northern assays. Data for gene expression assays were obtained using instruments available at the South Dakota State University's Functional Genomics Core Facility supported in part by the National Science Foundation/EPSCoR Grant No. 0091948 and by the State of South Dakota. Research in the authors' laboratory is supported by funds from USDA-AFRI, DOE Sun Grant Initiative, South Dakota Soybean Research and Promotion Council and South Dakota State University and Agricultural Experiment station.

References

Alonso, C.R. (2012) A complex 'mRNA degradation code' controls gene expression during animal development. *Trends Genet*, 28, 78-88.

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

Benes, V. and Castoldi, M. (2010) Expression profiling of microRNA using real-time quantitative PCR, how to use it and what is available. *Methods (San Diego, Calif.)*, 50, 244-249.

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

Chen, C., Tan, R., Wong, L., Fekete, R. and Halsey, J. (2011) Quantitation of microRNAs by real-time RT-qPCR. *Methods in molecular biology (Clifton, N.J.)*, 687, 113-134.

Chen, X. (2009) Small RNAs and their roles in plant development. *Annu Rev Cell Dev Biol*, 25, 21-44.

Chen, X., Liu, J., Cheng, Y. and Jia, D. (2002) HEN1 functions pleiotropically in Arabidopsis development and acts in C function in the flower. *Development*, 129, 1085-1094.

D'Haene, B., Mestdagh, P., Hellemans, J. and Vandesompele, J. (2012) miRNA expression profiling: from reference genes to global mean normalization. *Methods in molecular biology (Clifton, N.J.)*, 822, 261-272.

Devers, E.A., Branscheid, A., May, P. and Krajinski, F. (2011) Stars and symbiosis: microRNA- and microRNA*-mediated transcript cleavage involved in arbuscular mycorrhizal symbiosis. *Plant Physiol*, 156, 1990-2010.

Fu, H.J., Zhu, J., Yang, M., Zhang, Z.Y., Tie, Y., Jiang, H., Sun, Z.X. and Zheng, X.F. (2006) A novel method to monitor the expression of microRNAs. *Molecular biotechnology*, 32, 197-204.

Ghildiyal, M. and Zamore, P.D. (2009) Small silencing RNAs: an expanding universe. *Nature reviews. Genetics*, 10, 94-108.

Hurley, J., Roberts, D., Bond, A., Keys, D. and Chen, C. (2012) Stem-loop RT-qPCR for microRNA expression profiling. *Methods in molecular biology*, 822, 33-52.

Ji, L. and Chen, X. (2012) Regulation of small RNA stability: methylation and beyond. *Cell research*, 22, 624-636.

Khraiwesh, B., Zhu, J.K. and Zhu, J. (2012) Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochim Biophys Acta*, 1819, 137-148.

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

Lelandais-Briere, C., Naya, L., Sallet, E., Calenge, F., Frugier, F., Hartmann, C., Gouzy, J. and Crespi, M. (2009) Genome-wide Medicago truncatula small RNA analysis revealed novel microRNAs and isoforms differentially regulated in roots and nodules. *Plant Cell*, 21, 2780-2796.

Li, H., Deng, Y., Wu, T., Subramanian, S. and Yu, O. (2010) Misexpression of miR482, miR1512, and miR1515 increases soybean nodulation. *Plant Physiol*, 153, 1759-1770.

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

Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif.)*, 25, 402-408.

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

Mendell, J.T. and Olson, E.N. (2012) MicroRNAs in stress signaling and human disease. *Cell*, 148, 1172-1187.

Mestdagh, P., Feys, T., Bernard, N., Guenther, S., Chen, C., Speleman, F. and Vandesompele, J. (2008) High-throughput stem-loop RT-qPCR miRNA expression profiling using minute amounts of input RNA. *Nucleic acids research*, 36, e143.

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

Parry, G., Calderon-Villalobos, L.I., Prigge, M., Peret, B., Dharmasiri, S., Itoh, H., Lechner, E., Gray, W.M., Bennett, M. and Estelle, M. (2009) Complex regulation of the TIR1/AFB family of auxin receptors. *Proceedings of the*

National Academy of Sciences of the United States of America, 106, 22540-22545.

Peltier, H.J. and Latham, G.J. (2008) Normalization of microRNA expression levels in quantitative RT-PCR assays: identification of suitable reference RNA targets in normal and cancerous human solid tissues. *RNA (New York, N.Y.)*, 14, 844-852.

Ren, G., Chen, X. and Yu, B. (2012) Uridylation of miRNAs by hen1 suppressor1 in Arabidopsis. *Curr Biol*, 22, 695-700.

Ruiz-Ferrer, V. and Voinnet, O. (2009) Roles of plant small RNAs in biotic stress responses. *Annual review of plant biology*, 60, 485-510.

Schwab, R., Palatnik, J.F., Riester, M., Schommer, C., Schmid, M. and Weigel, D. (2005) Specific effects of microRNAs on the plant transcriptome. *Developmental cell*, 8, 517-527.

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

Shi, R., Sun, Y.H., Zhang, X.H. and Chiang, V.L. (2012) Poly(T) adaptor RT-PCR. *Methods in molecular biology*, 822, 53-66.

Subramanian, S., Fu, Y., Sunkar, R., Barbazuk, W.B., Zhu, J.K. and Yu, O. (2008) Novel and nodulation-regulated microRNAs in soybean roots. *BMC genomics*, 9, 160.

Subramanyam, D. and Blelloch, R. (2011) From microRNAs to targets: pathway discovery in cell fate transitions. *Curr Opin Genet Dev*, 21, 498-503.

Sunkar, R., Li, Y.F. and Jagadeeswaran, G. (2012) Functions of microRNAs in plant stress responses. *Trends in plant science*, 17, 196-203.

Turner, M., Yu, O. and Subramanian, S. (2012) Genome organization and characteristics of soybean microRNAs. *BMC Genomics*, 13, 169.

Varkonyi-Gasic, E. and Hellens, R.P. (2010) qRT-PCR of Small RNAs. *Methods in molecular biology*, 631, 109-122.

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

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

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

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

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

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

Yu, B., Bi, L., Zhai, J., Agarwal, M., Li, S., Wu, Q., Ding, S.W., Meyers, B.C., Vaucheret, H. and Chen, X. (2010) siRNAs compete with miRNAs for methylation by HEN1 in Arabidopsis. *Nucleic acids research*, 38, 5844-5850.

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R. and Chen, X. (2005) Methylation as a crucial step in plant microRNA biogenesis. *Science (New York, N.Y.)*, 307, 932-935.

Zhai, J., Jeong, D.H., De Paoli, E., Park, S., Rosen, B.D., Li, Y., Gonzalez, A.J., Yan, Z., Kitto, S.L., Grusak, M.A., Jackson, S.A., Stacey, G., Cook, D.R., Green, P.J., Sherrier, D.J. and Meyers, B.C. (2011) MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes & development*, 25, 2540-2553.

Zhang, L., Zheng, Y., Jagadeeswaran, G., Li, Y., Gowdu, K. and Sunkar, R. (2011) Identification and temporal expression analysis of conserved and novel microRNAs in Sorghum. *Genomics*, 98, 460-468.

Zhao, Y., Yu, Y., Zhai, J., Ramachandran, V., Dinh, T.T., Meyers, B.C., Mo, B. and Chen, X. (2012) The Arabidopsis nucleotidyl transferase HESO1 uridylates unmethylated small RNAs to trigger their degradation. *Curr Biol*, **22**, 689-694.

Chapter 5

5. Role of microRNAs in root nodule development

microRNAs (miRNAs) are 20- to 24-nt, small RNAs and are derived from a single-stranded precursor with a hairpin structure (Axtell 2013). During biogenesis, RNA polymerase II transcribes the miRNA genes into pri-miRNAs that are 70- to 300-nt in length. In plants, the Dicer homolog DCL1 (Dicer-like 1) cleaves the pri-miRNAs into ~70-nt pre-miRNAs which are further cleaved into a double-stranded miRNA/miRNA* (pronounced as "miRNA star") duplex with 2-nt overhangs at the 3' ends. The methylated duplex is transported into the cytoplasm, where, the miRNA associates with Argonaute (AGO) proteins to repress gene expression through cleaving the mRNA and/or repressing its translation (reviewed by (Chen 2005, Rogers and Chen 2013) (Figure 3.1, Chapter 3). Plant miRNAs regulate many of the transcription factors (TFs) that control various stages of plant growth and development (reviewed (Jones-Rhoades et al. 2006), defense against pathogens (Li et al. 2012, Shivaprasad et al. 2012), and response to abiotic stresses (reviewed (Sunkar et al. 2012).

miRNAs play a crucial role during root nodule development. In *Medicago truncatula*, the MtHAP2-1 transcription factor (TF) is strongly up-regulated at a very early stage of nodule development (El Yahyaoui et al. 2004). This gene was mostly abundant in the meristematic zone, but less present in the cells adjacent to the nodule infection zone. Conversely, the miRNA miR169, which targets MtHAP2-1, showed an inverse expression pattern, with high abundance in the infection zone proximal to the meristem, but not in the meristematic region (Combier et al. 2006). RNA interference (RNAi) of MtHAP2-1 causes a delay in nodulation and an arrest in nodule growth. Overexpression of miR169 led to a similar nodule developmental

defect as observed in MtHAP2-1 RNAi plants. However, a miR169-resistant version of mtHAP2-1 (mutated in the miR169 recognition site) developed nodules with no defect in tissue differentiation or nodule zonation, but with reduced elongation. This observation demonstrated that miR169 plays a key role in dictating cell differentiation and nodule development by the regulation of spatial expression patterns of MtHAP2-1 (Combier et al 2006). Likewise, miR166 regulates the expression of class III homeodomain leucine zipper (HD-ZIP III) genes that regulate nodule development in *M. truncatula* (Boualem et al. 2008). miR166 and its targets (MtCNA2, MtCNA1, and MtHB8) showed an overlapping pattern of expression in the meristematic region and vascular bundle of lateral roots and nodules. Overexpression of miR166 reduced lateral root and nodule number and affected root vascular organization (Boualem et al. 2008).

miRNAs also regulate nodule development by modulating auxin activity in determinate nodules (Turner et al. 2013). Hairy root composite plants were generated to overexpress miR393 (OX393) in order to silence a set of auxin receptors. Transgenic roots with OX393 were hyposensitive to auxin, but had the same number of nodules, suggesting that there is a minimal auxin requirement or signaling during the development of a determinate nodule. Similarly, the activity of repressor auxin response factor (repressor ARF 10/16/17) was suppressed by overexpression of the miR160 (OX160). OX160 roots induced hypersensitivity to auxin, resulting in a decrease formation of nodule primordium and suggesting that auxin hypersensitivity plays a role in the reduced nodule number (Turner et al. 2013). Further suppression of miR160 levels by using short tandem target mimic (STTM160) suggested that spatiotemporal regulation of auxin and cytokinin by miR160 is crucial for nodule development in soybean. Overexpression of STTM160 increased the number of

emerging nodules and reduced the number of mature nodules, indicating that low miR160 levels promote nodule formation but high miR160 level are important for nodule maturation. The suppression of miR160 also resulted in reduced sensitivity to auxin and enhanced sensitivity to cytokinin and provided the evidence that miR160 plays a role for proper nodule development in *Glycine max* by modulating auxin and cytokinin activity (Nizampatman et al 2015). Other examples of miRNAs that play roles in the formation of nodules include miR156 and miR172. Overexpression of miR156 reduced nodule number in *G.max* and *Lotus japonicus* (Yan et al. 2013, Wang et al. 2015). Overexpression of miR172 increased nodule number in *G. max* and *Phaseolus vulgaris* (Yan et al. 2013, Nova-Franco et al. 2015). Chapter 1, Table 1.1 includes a comprehensive list of miRNAs that have known roles in nodule development.

miRNAs are involved very early in root nodule development, as supported by the presence of several conserved and novel miRNAs in soybean roots 3 h post-inoculated with *B. japonicum* (Subramanian et al. 2008). Among these novel miRNAs, misexpression of miR482, miR1512 and miR1515 showed an increase in nodule numbers (Li et al. 2010). An updated annotation of novel miRNAs and their targets in soybean was performed in a genomic context (Turner et al. 2012). In the study, targets of miR169c, miR169g, and miR2218 were experimentally validated using a modified RNA ligase-mediated 5' Rapid amplification of cDNA ends (5' RLM- RACE) and assayed to see if they had nodule-specific expression (Turner et al. 2012). miR169, miR2218 and a few other miRNAs showed higher expression in root, stem, leaf and flower, but not in nodule. However, targets of the respective miRNAs showed higher expression in the organs with low miRNA expression, i.e. higher expression in nodule (Turner et al 2012). These miRNAs/target pairs with opposite

expression patterns in the root and nodule prompted us to hypothesize that nodules excluded miRNAs play roles in the nodule-specific expression of their targets (Figure 5. 1). For example, in *M. truncatula* miR169 plays a crucial role in nodule development by restricting the expression of the Nuclear Transcription Factor Y subunit alpha (NFY-A) also known as MtHAP-1 in the meristem region.

Inverse expression patterns of the miRNAs miR2218, miR169c, and miR1513c and their targets in the root and nodule were observed (Turner et al. 2012), but their mechanism of action during soybean nodule development is unknown. In this study, the roles of miR2218, miR169c, and miR1513c during soybean nodule development was investigated. To understand the roles of these miRNAs, the precursors of miR2218, miR169c, and miR1513c were overexpressed and misexpressed with CsVMV and ENOD40 promoters, respectively.

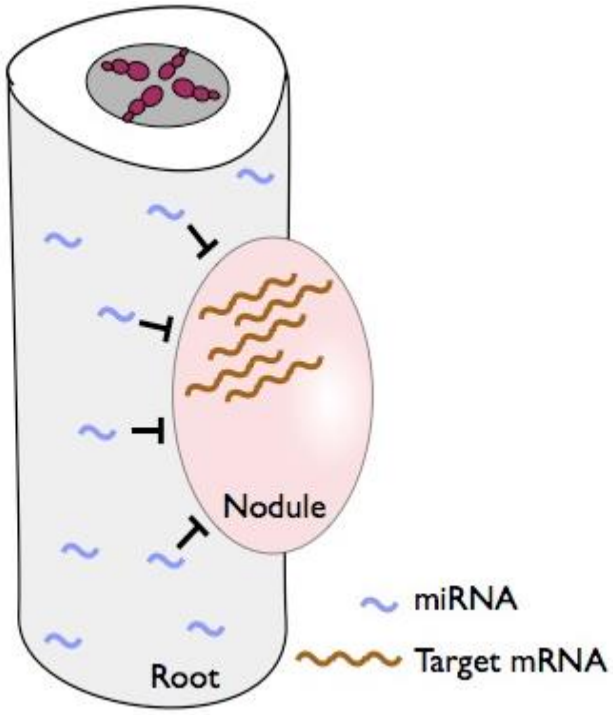


Figure 5.1 Schematic representation of nodule-excluded miRNAs, the absence of which allows target expression in the nodule. miRNA (purple curly line) is restricted to the root region while target mRNA (orange curly line) is restricted to the nodule.

5.1. Materials and Methods

5.1.1. DNA vectors

Binary vectors were generated by cloning precursors of miR169c, miR2218 and miR1513c into Gateway (GW) sites of pCAMGFP-CsVMV: GW (vector map, Appendix A, Figure 1) and pCAMGFP-ENOD40: GW (vector map, Appendix A, Figure 2). In pCAMGFP-CsVMV: GW vector, the constitutive CsVMV promoter is expected to overexpress the miRNAs. The ENOD40 promoter in pCAMGFP-ENOD40: GW is expected to misexpress the miRNAs (Turner et al 2013, Subramanian et al 2005). ENOD40 is expressed in different tissues at various stages of nodule development. During the early stage of soybean nodule development, it is expressed in dividing root cortical cells, nodule primordium, and the pericycle of the root vascular bundle. In mature nodules, it is expressed in uninfected cells of the central tissue and in pericycle cells of the nodule vascular bundles (Yang et al. 1993). With the help of the ENOD40 promoter, miRNAs that are primarily expressed in root are also expressed in the nodule. Both vectors have an additional super ubiquitin promoter driving the green fluorescent protein (GFP). GFP helps to distinguish the transgenic roots from non-transgenic roots when they are transformed into the binary vector *Agrobacterium rhizogenes* (strain AK599), during hairy root transformation.

5.1.2. TOPO-TA cloning

Precursors of miR2218, miR169c, and miR1513c (miRbase, Release 18) were PCR-amplified by using primer sequences (Table 5.1). The PCR reaction consisted of 2 µl (10 µm) of forward and reverse primers, 1 µl (40 ng) of genomic DNA and 45 µl of PCR supermix high fidelity (Invitrogen, 10790-020). The PCR conditions were: denaturation at 94 ºC, primer annealing at 55 ºC and extension at 68 ºC (PCR cycle:

94 ºC -2 min, 25 cycles of 94 ºC-30 s, 55 ºC -30 s, 68 ºC-60 s and a final extension at 68 ºC-5 min). Each PCR product was cloned into PCR8/GW (Invitrogen, cat no: K250020SC) vector (vector map, Appendix A, Figure 3). The reaction mix for the TOPO cloning consisted of TOPO vector =0.5 µl, PCR product =1.5 µl, and salt solution =1 µl. The reaction was kept at room temperature for 30 min. For the transformation, TOP10 *Escherichia coli* cells were thawed on ice (5-10 min), then 2 µl of the reaction mixture was added to the thawed cells for 20 min followed by heat shock at 42 ºC for 45 s and back to the ice for 2 min. 250 µl of the Super Optimal Broth (SOC) medium was added to the transformed cells. The cells were shaken at 37 ºC for 1 h (220 rpm), and plated onto Luria Broth (LB) plates with spectinomycin (100 µg/ml) as the selection antibiotic. The insertion of the pre-miRNA sequences was verified by EcoRI digestion (The 10 µL reaction consisted of Buffer NEB# 4 =1 µl, EcoRI-HF enzyme =0.25, DNA =2 µl - 4µl ~ 10 - 20 ng) followed by sequence verification (Beckman Coulter Genomics, Danvers MA) of the colonies verified by restriction enzyme digestion.

5.1.3. LR clonase

Sequence-verified TOPO-TA clones were cloned using LR clonase (Invitrogen, cat no: 11791-020) into the pCAMGFP-CsVMV: GW and pCAMGFP-ENOD40: GW destination vectors with PCR8/GW/miR2218, PCR8/GW/miR169c, and PCR8/GW/miR1513c as entry vectors. The LR clonase reaction consisted of 1.5 µl destination vector (~150 ng), 1.5 µl entry vector (~150 ng), TE buffer =1 µl, and proteinase K solution =1 µl. The reaction was incubated for 2 h, after which 2 µl of the LR reaction was transformed into the TOP10 *E. coli* cells. For the transformation, 2 µl of the LR reaction mix was added to thawed TOP10 *E. coli* cells (~50 µl) and incubated in ice for 2 min. The cells were heat shocked at 42 ºC for 45 s in a heat bath

and back to the ice for 2 min. 150 µl of the SOC medium was added to the cells and was shaken at 37 ºC for 1 h at 220 rpm and plated onto the LB plates with kanamycin (50 µg/ml) as the selection antibiotic. The plates were incubated overnight at 37 ºC to allow colony growth. The colonies were verified by restriction digestion and sequencing of the positive clones as determined by restriction digestion.

5.1.4. *Agrobacterium rhizogenes*-mediated gene transfer

Restriction digestion- and sequence-verified plasmids were transformed into *A. rhizogenes* (strain AK599) cells for hairy root transformation (see hairy root transformation, section 5.1.5). For the electroporation, 1 µl of the plasmid (~150 ng) was added to the AK599 competent cells (50 µl), mixed gently and incubated on ice for 20 min. The incubated culture (25 µl) was transferred into a 0.1 cm-gap electroporation cuvette (Eppendorf, MA) and electroporated at an electric charge of 25 µF, 400 ohm resistance, and 1.8 Kvolt using a Bio-Rad Gene Pulser Xcell Electroporation System (Bio-Rad Laboratories, CA). After electroporation, 1 ml of the LB was added to the cells in the cuvette and mixed gently. The cells were then transferred to a 2 ml microcentrifuge tube and shaken horizontally at 200 rpm and 30 ºC for 2 hr. The cells were plated onto LB plates with kanamycin (50 µg/ml) as the selection antibiotic and incubated at 30 ºC for 36-48 hr. Individual colonies were grown in liquid LB and glycerol stocks were prepared. Vector plasmids (pCAMGFP-CsVMV: GW and pCAMGFP-ENOD40: GW) without miRNA precursor sequences were also transformed into AK599 cells to generate "empty" vector control (VC).

Table 5.1 List of primers used for cloning and qPCR of miR2218, miR169c and miR1513c. Provided in the list are forward and reverse primers used in cloning precursor sequences of miR2218, miR169c, and miR1513c (miRbase, Release 18) and forward and reverse primers used to quantify expression of mature sequences by RT-qPCR.

MiRNA	Sequence (F)	Sequence (R)	Purpose
miR2218	GAGCTTGAGGAAGTG	GCGGAGAGGAGAGTGGA	Cloning
miR169c	TCATGGAGAGGTTGA	GAGGAAGAAAGCCAAAT	Cloning
miR1513	GTGGTGCATGATCTGA	AAAATGAGATATTGCAAA	Cloning
miR2218	TCGCTttgccgattccaccca	GTCGTATCCAGTGCAGGG	qPCR
miR169c	TCGCTcagccaaggatga	GTCGTATCCAGTGCAGGG	qPCR
miR1513	TCGCTtgagagaaagccatg	GTCGTATCCAGTGCAGGG	qPCR

5.1.5. Hairy root transformation

Hairy root transgenic composite plants were generated as described previously (Collier et al. 2005). Briefly, soybean seeds (Williams 82) were grown in 4 inch pots (Hummert, cat no: 1433561) containing a mixture of 1:2 vermiculite and perlite (Hummert, cat no: 10-2200-1, 10-1123-1) for 14 days in growth chambers (16-/8-h light/dark cycle). *A. rhizogenes* was grown for 16 -18 h (200 rpm, 30 ºC) in a shaker in LB medium (Fisher Scientific, cat no: BP1426500) with kanamycin (50 µg/ml) as a selection antibiotic. The culture was centrifuged (3000 x g, 8 min) and the pellet was resuspended using nutrient solution without nitrogen (N^-PNS) (Subramanian et al. 2008) . Rock wool plugs (Hummert, cat no: 1424841) were prepared, autoclaved, and punctured in the middle to add about 5-8 ml of the bacterial suspension with the help of a pipet (VWR, cat no: 89130-900). The tip of the soybean seedlings, with emerging trifoliate leaves and the first two leaves above, was cut at a slant with a razor blade and placed immediately in the rock wool plug with the culture. The transformed plants were kept under a 16-/8-h light cycle at, 25 ºC with a lid and watered once the rock wool plug started to dry out.

5.1.6. *Bradyrhizobium* Inoculation and Nodulation Assays

For nodulation assays, roots of transgenic composite plants (three weeks post-transformation) were screened for GFP. GFP-positive plants with respective miRNA overexpression and misexpression were transferred into 4 inch pots containing a mixture of vermiculite and perlite (1:2 ratio). They were allowed to grow for 1 week (25 ºC, 16/8 light/dark and 50% humidity) and inoculated with *B. japonicum* USDA 110 (OD at 600 nm = 0.08). *B. japonicum* was grown in Vincent's rich (VR) medium (30 ºC, 200 rpm) with chloramphenicol (20 µg/ml) as selection antibiotic. The culture was centrifuged (3500 x g, 8 min) and the supernatant was removed. The pellet was

resuspended in an equal volume of N^-PNS. Plants were grown for another 2 weeks by alternatively applying deionized water and N^-PNS solution. GFP-positive roots were separated and nodules at two developmental stages (emerging and mature) were counted under a dissection microscope (Olympus SZX12) (Olympus Corporation, PA). Nodules were counted at 14 days post inoculation (dpi). The nodules were categorized as emerging if they had a slight bump on the surface but had not protruded out completely or as mature if completely protruding out and pink- in color (Figure 5.2).

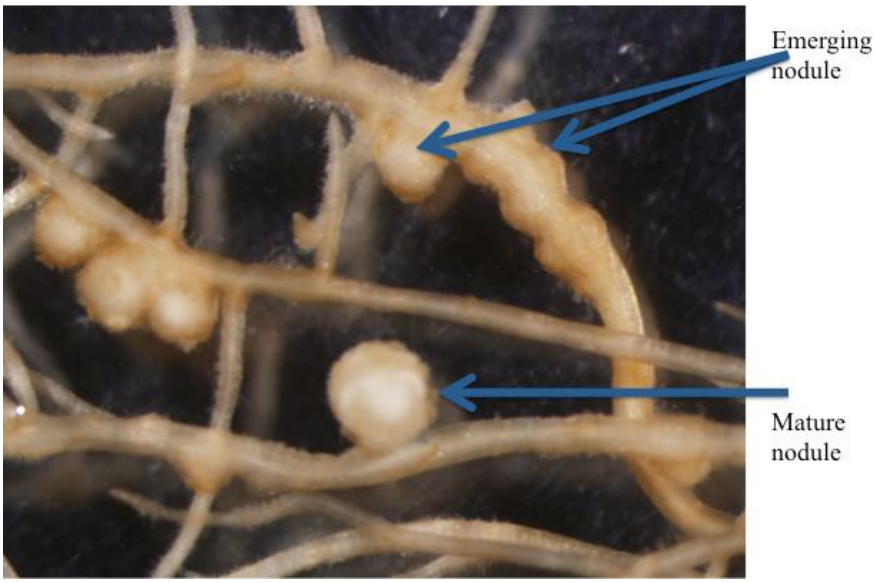


Figure 5.2 Representative image of emerging and mature root nodules of soybean. Composite soybean plant roots were inoculated with B. japonicum and were harvested at 14 dpi for nodule count.

5.1.7. Tissue Sampling and RNA isolation

GFP-positive roots of transgenic composite plants with overexpression (OX) and misexpression (ENOD) of miR2218, miR169c, and miR1513c were used for RNA isolation. The samples were harvested in liquid nitrogen and stored at -80 ºC until RNA isolation. GFP- positive roots without inoculation (only root) and with inoculation (14 dpi and 28 dpi roots with nodules) were harvested based on the objectives of the study.

RNA was isolated as described previously (Subramanian et al. 2008). Briefly, roots stored at -80 ºC were ground in liquid nitrogen using mortar and pestle for total RNA isolation using Tri Reagent (Sigma-Aldrich, cat no: T9424). The RNA was treated with DNase (New England Biolabs, cat no: M0303S) to remove any DNA contamination. The cDNA was synthesized using M-MuLv Reverse transcriptase (New England Biolabs, cat no: M0253S).

5.1.8. cDNA Synthesis

For messenger RNA (mRNA) cDNA synthesis, 2 µg of DNase-treated total RNA, 1 µl of 10 mM dNTP mix, 1 µl of 10 mM oligo (dT) primer and DEPC-treated water to a final volume of 16.5 µl was added to a PCR tube. The sample was incubated in a thermal cycler (ABI VERITI 96 well Thermal Cycler) at 75 °C for 5 min and snap-cooled in ice for 5 min. To this reaction mix, 2 µl of 10X Reverse transcriptase reaction buffer, 0.5 µl of MuMLV Reverse transcriptase (200,000 U/ml) (New England Biolabs, cat no: M0253S), and 1 µl of RNase inhibitor (40,000 U/ml) (New England Biolabs, cat no: M0314S) was added. The sample was mixed gently and incubated in the thermal cycler at 42 °C for 60 min and 95 °C for 5 min for heat inactivation of reverse transcriptase. After the completion of cDNA synthesis, the

samples were stored at -20 °C until further use.

For miRNA cDNA synthesis, the hairpin-priming method (Varkonyi-Gasic et al. 2007) was used. Initially, 1 µg of total RNA, 0.5 µl of 10 mM dNTP mix and DEPC-treated water to a final volume of 12.5 µl was added in a PCR tube. The sample was incubated at 65 °C for 5 min in a thermal cycler and snap-cooled on ice for 5 min. To this reaction mix, 2 µl of 10X Reverse transcriptase reaction buffer, 0.5 µl of MuMLV Reverse transcriptase (NEB, MA), 1 µl each 1 mM miRNA hairpin loop primer (designed for miR2218, miR169c, miR1513c, miR1515 and U6 as mentioned in the primer table list; Table 5.1) were added. The mixture was incubated in a thermal cycler at 16 °C for 30 min, then 60 cycles of 30 °C for 30 s, 42 °C for 30 s and 50 °C for 1 s, and then 85 °C for 5 min to heat-inactivate the reverse transcriptase. The samples were stored at -20 °C for future experiments.

5.1.9. Gene Expression Analysis

The expression levels of each miRNA and its target mRNA were quantified by using reverse transcription-quantitative polymerase chain reaction (RT-qPCR). To quantify the expression of miRNA, a 20 µl reaction was set up with 10 µl of 2X SYBR premix, 1 µl of cDNA, 0.4 µl of forward primer, 0.4 µl common reverse primer, 0.4 µl of 50x ROX, and 7.8 µl of DEPC-treated water. For each gene, three technical replicate reactions were prepared in a 96-well plate. The thermal cycle of 95 °C for 10 s, then 45 cycles of 95 °C for 5 s, 60 °C – 64 °C for 20 s, and 72 °C for 1 s was used. At the end of the thermal cycling, dissociation curve analysis was performed by using a thermal cycle of 95 °C for 1 min, 55 °C for 30 s, and 95 °C for 30 s. Expression of miRNAs was normalized to reference genes miR1515 and U6. Expression level was calculated by using the delta delta threshold cycle (ddCt) method (Livak and Schmittgen 2001).

To determine the expression level of each mRNA, a 20 µl reaction was set up with 10 µl of 2x SYBR premix (Clontech, cat no: 639676), 1.6 µl of cDNA as template, 0.4 µl of 10 µM forward primer, 0.4 µl of 10 µM reverse primer, 0.4 µl of 50x ROX reference dye, and 7.2 µl of DEPC-treated water. For each gene, three technical replicate reactions were prepared in a 96-well plate. The thermal cycle of 95 °C for 10 s, then 40 cycles of 95 °C for 5 s, 58 °C - 64° C for 20 s was used. At the end of the thermal cycle, a dissociation curve analysis was performed. For this, the reactions were cooled to 55 °C, and the fluorescence emission was collected continuously by heating the reaction @ 0.1 °C/s to 95 °C for 1 min. Gene expression was normalized to GmActin and GmCons4 using the delta delta threshold cycle (ddCt) method (Livak and Schmittgen 2001).

5.1.10. *Bradyrhizobium japonicum* GUS assay

To examine infection thread formation and the early stage of nodule development, composite plants were inoculated with *B. japonicum* (USDA110) transformed with an nptII: GUS (Bj-gus) construct (a kind gift from Dr. Gary Stacey, University of Missouri, Columbia, MO). The Bj-gus cells were cultured in VR medium similar to the method for *B. japonicum*, with chloramphenicol (20 µg/ml) and tetracycline (10 µg/ml) as selection antibiotics. The suspension for inoculation was prepared similar to *B. japonicum* (see *Bradyrhizobium* inoculation and nodulation assays). For GUS staining, the staining solution was prepared by mixing the following solutions: 1 M sodium phosphate buffer (pH 7.0), 0.1 M $K_3[Fe(CN)_6]$, 0.1 M $K_4[Fe(CN)_6]$, 0.1 M EDTA(pH7.0), 10% triton X-100 and 100% DMSO + X-Gluc (Gold Biotechnology, cat no: G1281C2). The sodium phosphate buffer was prepared by combining 1 M Na_2HPO_4 and 1 M NaH_2PO_4. X-Gluc was prepared by dissolving in DMSO to the final concentration of 1 mg/ml. The staining solution was filter

sterilized by using stericup-GV 45MM (Fisher Scientific, cat no: SCGVT05RE). The GFP-positive Bj-gus-inoculated roots were submerged in the staining solution and vacuum infiltrated for 5 min. The samples were incubated at 37 ºC for at least 24 h. The root samples were then fixed in fixing solution. The fixing solution was prepared by combining 3% v/v glutaraldehyde, 0.03% triton X-100, 100 mM Tris (pH 7.2). Samples were vacuum infiltrated for 5 min and kept for 24-48 h in dark at 4 ºC. The fixed tissues were mounted in a glass slide with 10% glycerol (Sigma-Aldrich, cat no: G5516) and observed under a microscope (Olympus AX70).

5.1.11. Microscopy

Images for miR169c sensor and VC (see section 5.2.7 for the details about the construct) were obtained by using a laser scanning microscope (Olympus FV300). Root tip, emerging nodules at 7 dpi, and mature nodules at 14 dpi were scanned with the following settings: Channel 1: GFP (488- nm excitation/685-nm emission), Channel 2 : tdTomato (700-nm emission), 5 μm sections, 6% gain and 4% offset, Kalman acquisition 3, objective 10x.

5.1.12. Statistical Analysis

Root nodules number (emerging, mature and total), root hair curl (rhc), and nodule primordia (np) in miRNA overexpression and misexpression lines was compared to the VC with the Poisson regression ($P < 0.05$) by using R software (version 3.0.2). Expression of mRNA (target of miRNA) and miRNA in overexpression and misexpression lines was compared to expression in the VC with Two-Sample *t*-test ($P < 0.05$) in MS Excel.

5.2. Results

5.2.1. Validation of overexpression

To understand the roles of miR2218, miR169c, and miR1513c in soybean nodule development, the precursors of these miRNAs were overexpressed using the CsVMV promoter. RNA isolated from GFP-positive transgenic hairy roots without inoculation was used to verify the expression of respective miRNAs by RT-qPCR (see materials and methods). The relative expression levels of miR2218, miR169c and miR1513c were significantly increased in all overexpression lines compared to those expressing the VC (*t*-test, $P < 0.05$) (Figure 5.3).

One of the common methods of mRNA regulation in plants is cleavage of their cognate target genes. An increase in miRNA levels usually causes a decrease in the expression levels of their target genes. The expression levels of the targets of miR2218, miR169c and miR1513c were also determined using the same tissue that was used for examining the expression levels of miRNAs. Turner et al (Turner et al. 2012) predicted the targets of miR2218, miR169c, and miR1513c. Among the predicted targets, the targets of miR2118 (Glyma12g03040) and miR169c (Glyma10g10240, Glyma17g05929, and Glyma15g18970) were verified by using a modified 5' - Rapid Amplification of cDNA Ends (RACE) assay. A significant reduction in the target of miR2218 was observed (Figure 5.4 A). Two of the targets of miR169c (Glyma10g10240 and Glyma17g05929) were significantly reduced, but not Glyma15g18970 (Figure 5.4 B, C, and D). This could be due to tissue-specific expression of the target mRNAs and their regulation by miR169c. Surprisingly, the target level of miR1513c (Glyma08g27820) showed a significant increase in expression level compared to VC (Figure 5.4 E).

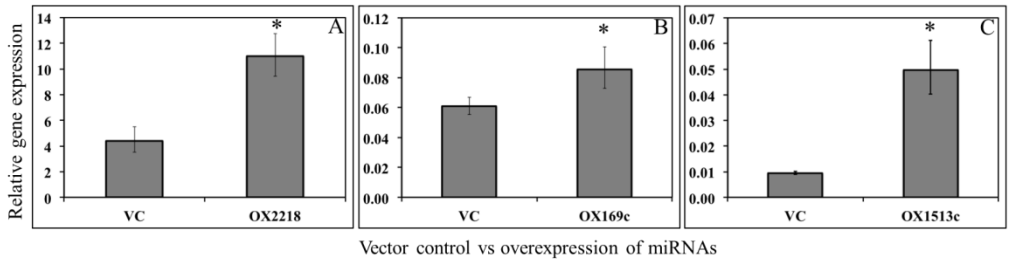


Figure 5.3 Relative expression of miR2218, miR169c and miR1513c in corresponding overexpression lines. Transgenic hairy roots of composite soybean plants grown for 1 wk in 1:2 vermiculite to perlite were harvested for RNA isolation and gene expression analysis. ***A.*** *miR2218* ***B.*** *miR169c and* ***C.*** *miR1513c. Relative expression refers to gene expression based on threshold cycle (ct) normalized with housekeeping gene miR1515 using three technical replicates. * Indicates significant expression (t-test, P <0.05).*

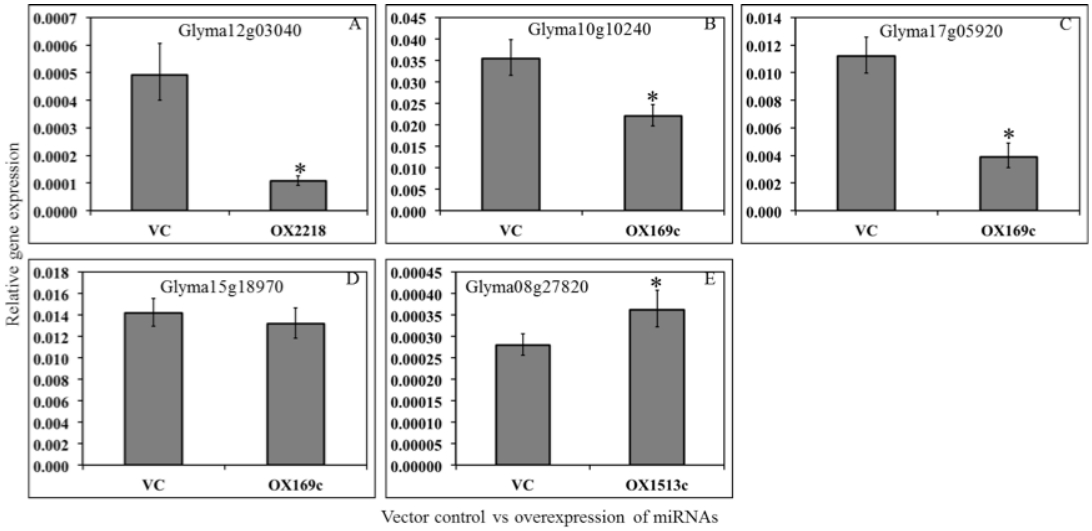


Figure 5.4 Relative expression of target genes of miR2218, miR169c and 1513c in the corresponding overexpression lines. Transgenic hairy roots of composite soybean plants grown for 1 wk in 1:2 vermiculite to perlite were harvested for expression analysis. ***A.*** *Glyma12g03040 target of miR2218,* ***B.*** *Glyma10g10240* ***C.*** *Glyma17g05920 and* ***D.*** *Glyma15g18970 targets of miR169c and* ***E****. Glyma08g27820 target of miR1513c. Relative expression refers to gene expression based on threshold cycle (ct) normalized with housekeeping gene GmActin using three technical replicates. * indicates significant expression (t-test, P <0.05).*

5.2.2. Validation of misexpression lines

To test the hypothesis that nodule-excluded miRNAs play roles in nodule development, precursors of miR2218, miR169c and miR1513c were misexpressed by using the ENOD40 promoter. The misexpression of the miRNAs was validated by RNA harvested from 14 dpi mature nodules. No increases in the miRNA levels in the misexpression lines were observed in ENOD2218, ENOD169c and ENOD1513c (*t*-test, $P < 0.05$) (Figure 5.5). This could be due to a relatively small increase in the expression of these miRNAs or to additional levels of regulation that keep the levels of mature miRNAs at a lower level. However, expression levels of Glyma12g03040 (target of miR2218), Glyma10g10240 (target of miR169c) and Glyma08g27820 (target of miR1513c) (Figure 5.6 A, B, and E) were significantly decreased (*t*-test, $P < 0.05$), suggesting cleavage of the target mRNAs. The expression levels of the remaining two targets of miR169c, excluding Glyma10g10240, were not significantly decreased, possibly due to tissue-specific expression of these target and their regulation by miRNAs.

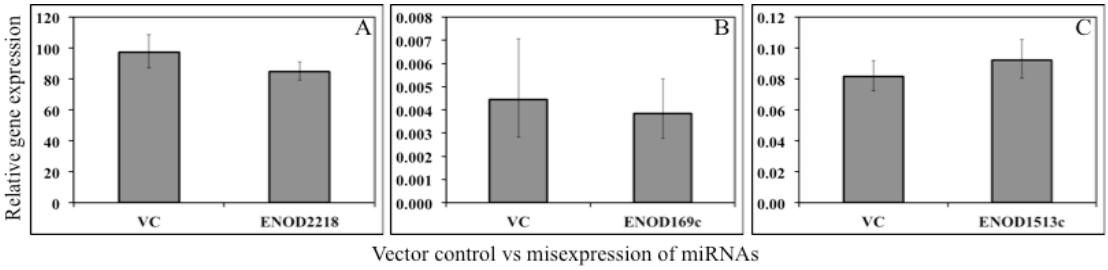


Figure 5.5 Relative expression of miR2218, miR169c, and miR1513c in the corresponding misexpression lines. Composite soybean plants grown for 1 wk in 1:2 vermiculite to perlite were inoculated with B. japonicum (USDA110). Nodules from transgenic hairy root plants were harvested for expression analysis at 14 dpi. ***A.*** *miR2218* ***B.*** *miR169c and* ***C.*** *miR1513c. Relative expression refers to gene expression based on threshold cycle (ct) normalized with housekeeping gene miR1515 using three technical replicates. * Indicates significant expression (t-test, P <0.05).*

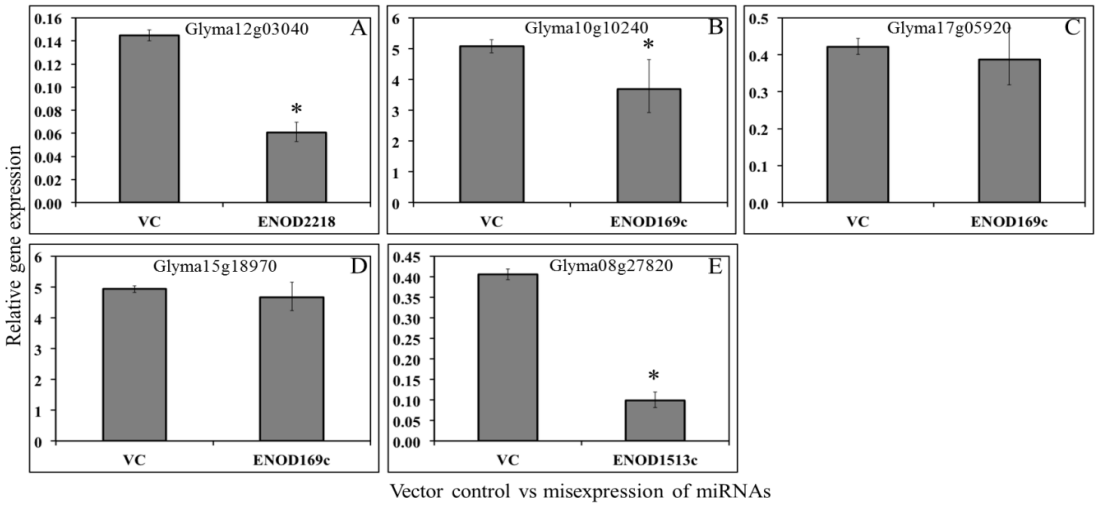
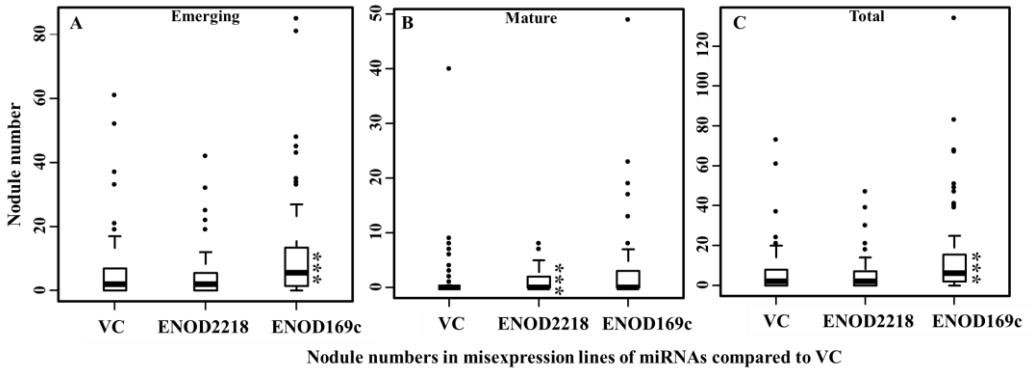


Figure 5.6 Relative expression levels of the target genes of miR2218, miR169c, and miR1513c in the corresponding misexpression lines. Composite soybean plants grown for 1 wk in 1:2 vermiculite to perlite were inoculated with B. japonicum (USDA110). Nodules from transgenic hairy root plants were harvested for the expression analysis at 14 dpi. ***A.*** *Glyma12g03040 target of miR2218,* ***B.*** *Glyma10g10240* ***C.*** *Glyma17g05920 and* ***D.*** *Glyma15g18970 targets of miR169c and* ***E****. Glyma08g27820 target of miR1513c. Relative expression refers to gene expression based on threshold cycle (ct) normalized with housekeeping gene GmActin using three technical replicates. * indicates significant expression (t-test, P <0.05).*

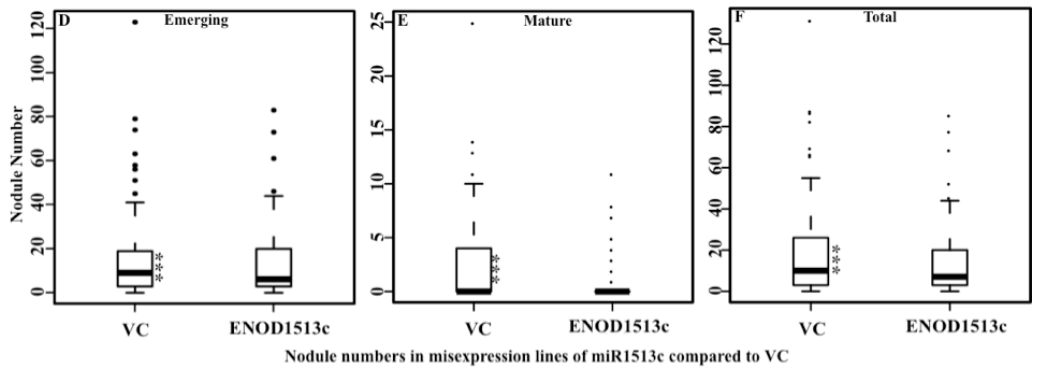
5.2.3. miR169c role in controlling root nodule number

To test our hypothesis that nodule-excluded miRNA plays a role in nodule development, we inoculated roots of the three ENOD40 misexpression lines with *B. japonicum* and harvested for nodule count at 14 dpi. Nodule numbers at emerging (em) and mature (mt) stages of development were counted in ENOD2218, ENOD169c, ENOD1513c, and VC. The experiments were repeated three times. For ENOD2218, significant decreases in both emerging and mature nodules were observed only in replicate 3 (Poisson regression, $P<0.001$) (Figure 5.7 J and K, Table 5. 2, Table 5.3). For ENOD169c, a significant increase in emerging nodules was observed in all three replicate experiments (Poisson regression, $P<0.01$) (Figure 5.7 A, G and J, Table 5.2) and mature nodules were significantly decreased in replicates 2 and 3 (Poisson regression, $P<0.01$) (Figure 5.7 H and K, Table 5.3). A significant increase ($P<0.001$) in total nodule number was observed in replicates 1 and 2 in ENOD169c compared to VC (Figure 5.7 C and I, Table 5.4). For ENOD1513c, a significant decrease in emerging, mature and total nodule number was observed in replicate 1; however, this trend was not seen in replicates 2 and 3. (Figure 5.7 D, E and F, Table 5.2, 5.3 and 5.4). These results strongly suggest that miR169c might play a role in controlling nodule number.

Replicate 1

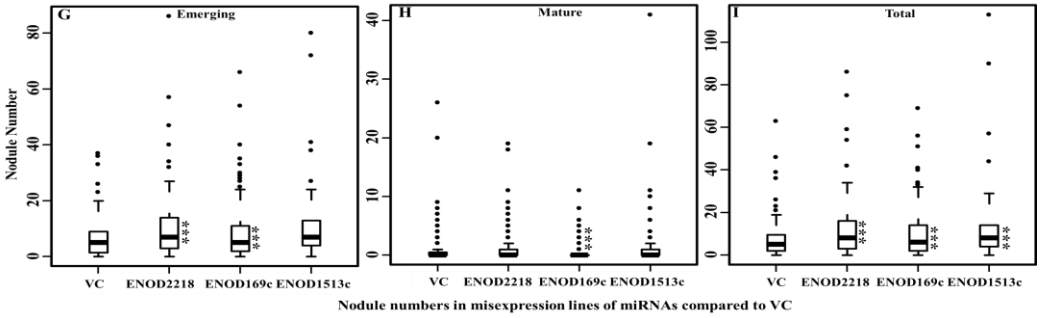


Nodule numbers in misexpression lines of miRNAs compared to VC



Nodule numbers in misexpression lines of miR1513c compared to VC

Replicate 2



Nodule numbers in misexpression lines of miRNAs compared to VC

Replicate 3

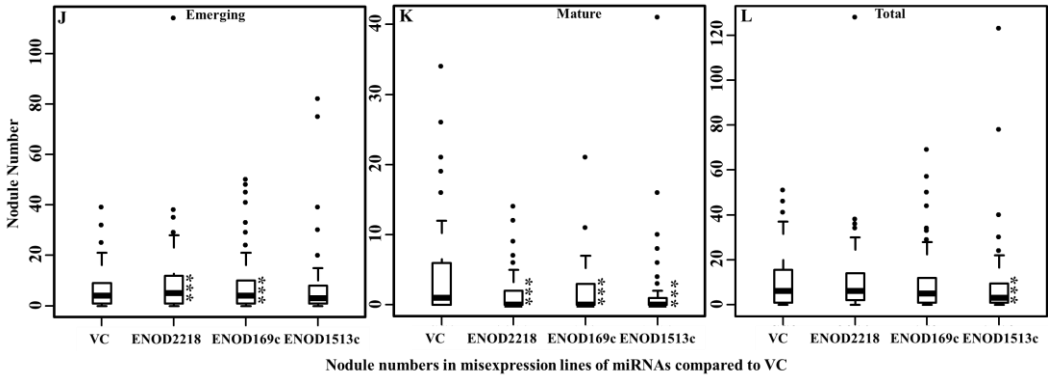


Figure 5.7 Effect of misexpression of miRNAs on nodule formation. Hairy roots of ENOD2218, ENOD169c, ENOD1513c, and VC were inoculated with B. japonicum. Nodules categorized as emerging and mature were counted at 14 dpi. Total nodule numbers were calculated by summing emerging and mature nodule numbers. The nodule count experiment was repeated three times. ***A, B, C*** *first replicate counts for emerging, mature and total nodule numbers, respectively, compared to VC in ENOD2218 and ENOD169.* ***D, E, F*** *first replicate counts for emerging, mature and total nodule numbers, respectively, in ENOD1513c compared to VC.* ***G, H, I*** *second replicate counts for emerging, mature and total nodule numbers respectively.* ***J, K, L*** *third replicate counts for emerging, mature and total nodule numbers, respectively, in ENOD2218, ENOD169, and ENOD1513c compared to VC. The significant difference in nodule number was tested by using Poisson regression (P <0.05). * indicates significant expression.*

*Table 5.2 Summary of emerging nodules count (average ± SE) in misexpression lines of miRNAs vs VC. Hairy roots of ENOD2218, ENOD169c, ENOD1513c, and VC were inoculated with B. japonicum and harvested for nodule count at 14 dpi. The number in parenthesis represents the number of roots used for nodule count. The significant difference in nodule number was tested by using Poisson regression (*** P <0.001, ** P <0.01, and * P <0.05).*

Replicate	VC	ENOD2218	ENOD169c	ENOD1513c
1	(73) 6.3 ± 1.29	(52) 5.17 ± 1.19	(68) 11.6 ± 2.12***	17.7 ± 2.64(VC) (73) 14.2 ± 2.03***
2	(75) 7.17 ± 0.95	(76) 11.7 ± 1.65***	(89) 9.8 ± 1.32***	(87) 10.7 ± 1.36***
3	(55) 6.8 ± 1.1	(81) 8.8 ± 1.61***	(69) 8.2 ± 1.4**	(83) 7.04 ± 1.4

*Table 5.3 Summary of mature nodule count (average ± SE) in misexpression lines of miRNAs vs VC. Hairy roots of ENOD2218, ENOD169c, ENOD1513c, and VC were inoculated with B. japonicum and harvested for nodule count at 14 dpi. The number in parenthesis represents the number of roots used for nodule count. The significant difference in nodule number was tested by using Poisson regression (*** P <0.001, ** P <0.01, and * P <0.05).*

Replicate	VC	ENOD2218	ENOD169c	ENOD1513c
1	1.2 ± 0.22	1.19 ± 0.30	3.0 ± 0.88	2.8 ± 0.61(VC)
2	1.4 ± 0.48	1.8 ± 0.43	0.8 ± 0.2**	1.6 ± 0.55
3	4.5 ± 1.0	1.8 ± 0.35***	1.9 ± 0.4***	1.43 ± 0.6***

*Table 5.4 Summary of total nodule count (average ± SE) in misexpression lines of miRNAs vs VC. Hairy roots of ENOD2218, ENOD169c, ENOD1513c, and VC were inoculated with B. japonicum and harvested for nodule count at 14 dpi. The number in parenthesis represents the number of roots used for nodule count. The significant difference in nodule number was tested by using Poisson regression (*** P <0.001, ** P <0.01, and * P <0.05).*

Replicate	VC	ENOD2218	ENOD169c	ENOD1513c
1	7.5 ± 1.39	6.37 ± 1.41	14.5 ± 2.82***	20.5 ± 3.04(VC)
2	8.57 ± 1.30	13.5 ± 1.89***	10.6 ± 1.41***	12.4 ± 1.77
3	11.3 ± 1.8	10.6 ± 1.8	10.1 ± 1.7	8.47 ± 1.8***

5.2.4. miR169c role in nodule maturation

Since a significant increase in emerging nodules and a significant decrease of mature nodules were observed in ENOD169c compared to VC (Table 5.2 and 5.3), any role for miR169c in nodule maturation was further tested. Nodule growth time was increased from 2 to 4 weeks in both overexpression and misexpression lines of miR169. Significant increases in emerging, mature, and total nodule numbers were observed in ENOD169c compared to VC (Poisson regression, $P < 0.001$) (Figure 5.8). A significant decrease in mature nodule number was observed in OX169c compared to VC (Poisson regression, $P < 0.05$) (Figure 5.8). This suggested that miR169 might play a role in nodule maturation, but additional experiments are necessary to verify this result.

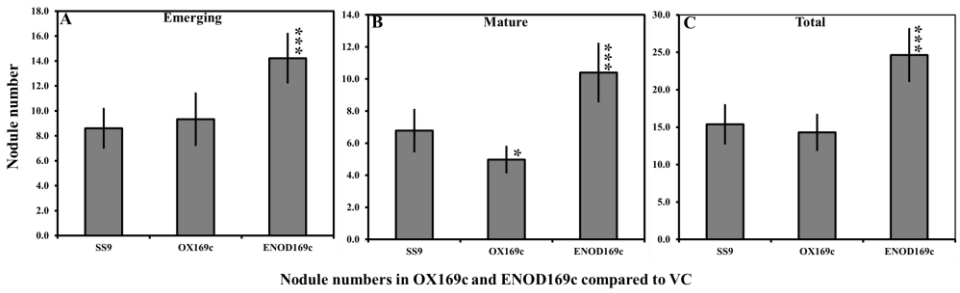


*Figure 5.8 miR169c plays a role in nodule maturation. Hairy roots of OX169c, ENOD169c, and VC were inoculated with B. japonicum and harvested for nodule count at 24 dpi. The number of roots used for nodule count were 39, 62 and 50, respectively. The significant difference in nodule number was tested by using Poisson regression (*** P <0.001, ** P <0.01, and * P <0.05).*

5.2.5. miR169c role in early nodulation event

To elucidate whether the increase in emerging nodule numbers observed in ENOD169c was due to an effect in early nodulation events, two early nodulation events (i) number of root hair deformations counted as number of root hair curl (RHC) and (ii) nodule primordium (NP) were counted in OX169c and ENOD169c after inoculation with *B. japonicum* at 7 dpi. *B. japonicum* expressing nptII: GUS was used to identify NP and RHC. A significant increase in RHC was observed in replicate 1 of ENOD169c (Poisson regression, $P < 0.05$); however, in two other replicates no effect in RHC was observed (Table 5.5, Figure 5.9). The "RHC phenotype" is in agreement with expression of ENOD40 in nodule tissues but not in infection threads. No difference in NP was observed in ENOD169c compared to VC in replicate 1, however a significant decrease in NP was observed in replicate 2 and 3 (Table 5.6, Figure 5.10). The number of nodules overtime in ENOD169c seems contradictory: a significant decrease in NP at 7dpi; a significant increase in emerging nodules at 14 dpi; a significant decrease in mature nodules at 14 dpi; and a significant increase in both at 24 dpi. Likewise, a significant increase in both RHC and NP was observed in OX169 compared to VC at 7 dpi (Table 5.5 and 5.6, Figure 5.11) (Poisson regression, $P < 0.05$), however, a significant decrease in mature nodules were observed at 24 dpi and no significant increase in emerging and total nodules were observed. These results suggest that miR169c might spatially and /or temporally control nodule development.

Table 5.5. Role of miR169 in root hair deformation. Number of root hair curl (RHC) (average ± SE) is shown. Roots inoculated with B. japonicum *(expressing nptII: GUS) were harvested at 7 dpi for RHC count. Three replicate experiments were performed for comparison of ENOD169c to vector control (VC) and 1 replicate experiment was performed for comparison of OX169 to VC. The number in the parenthesis represents the number of roots used for RHC count. The significant differences in RHC were tested by using Poisson regression (*** P <0.001, ** P <0.01, and * P <0.05). NA represents no experiment.*

	ENOD169c compared to VC		OX169c comparison to VC	
Replicates	VC	ENOD169c	VC	OX169c
Rep1	13.2 ± 3.65 (18)	17.0 ± 4.30* (22)	10.7 ± 3.5 (15)	20.4 ± 13.9* (10)
Rep2	4.2 ± 1.04 (22)	2.6 ± 1.16 (21)	NA	NA
Rep3	61.8 ± 11.06(20)	46.6 ± 9.13 (20)	NA	NA

Table 5.5 Role of miR169 in nodule primordium (NP) formation. Number of NP (average ± SE) is shown. Roots inoculated with B. japonicum *(expressing nptII: GUS) were harvested at 7 dpi for NP count. Three replicate experiments were performed for comparison of ENOD169c to vector control (VC) and 1 replicate experiment was performed for comparison of OX169 to VC. Roots used for RHC (Table 5.5) were also used for NP count. The significant difference in NP was tested by using Poisson regression (*** P <0.001, ** P <0.01, and * P <0.05). NA represents no experiment.*

	ENOD169c compared to VC		OX169c compared to VC	
Replicates	SS9	ENOD169c	SS9	OX169c
Rep1	4.4 ± 1.48	4.0 ± 1.50	3.6 ± 1.4	7.1 ± 2.7*
Rep2	3.6 ± 0.97	1.6 ± 0.46***	NA	NA
Rep3	10.2 ± 2.39	7.6 ± 1.40***	NA	NA

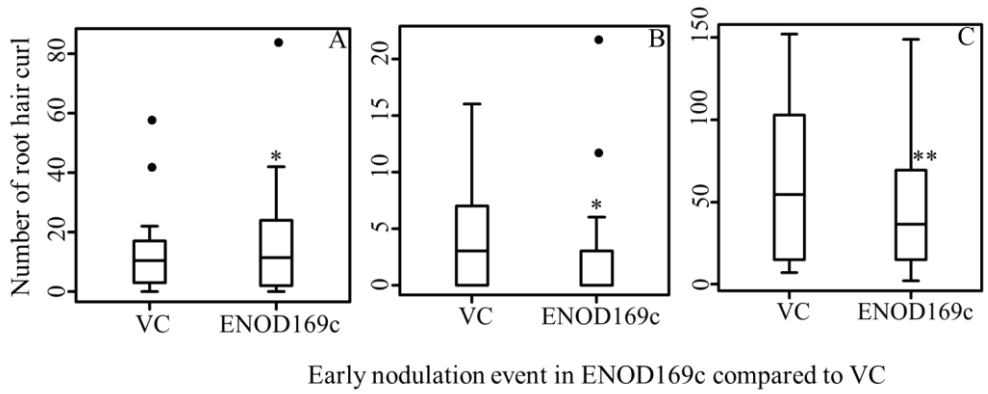


Figure 5.9 Effect of ENOD169 in root hair deformation. Number of root hair curl (RHC) was counted 7 dpi after B. japonicum inoculation. ***A.*** *First replicate* ***B.*** *Second replicate* ***C.*** *Third replicate. The significant difference in nodule number was tested by using Poisson regression (*** P <0.001, ** P <0.01, and * P <0.05).*

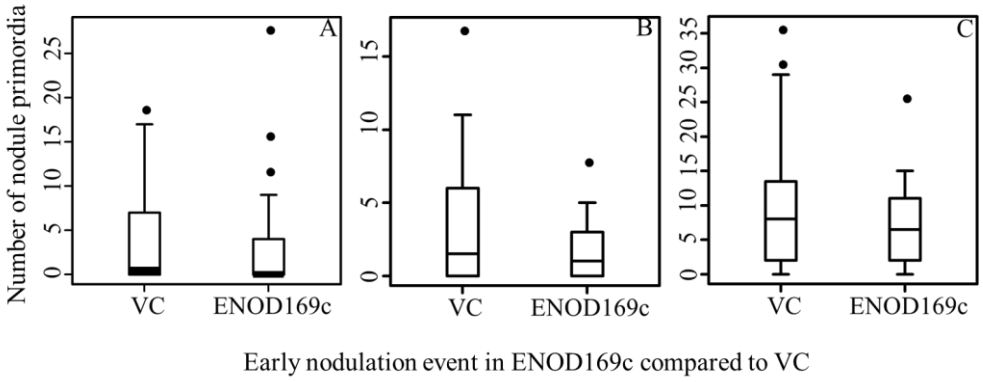


Figure 5.10 Effect of ENOD169 in nodule primordium formation. Number of nodule primordia was counted 7 dpi after B.japonicum inoculation. ***A.*** *First replicate* ***B.*** *Second replicate* ***C.*** *Third replicate. The significant difference in nodule number was tested by using Poisson regression (P <0.05).*

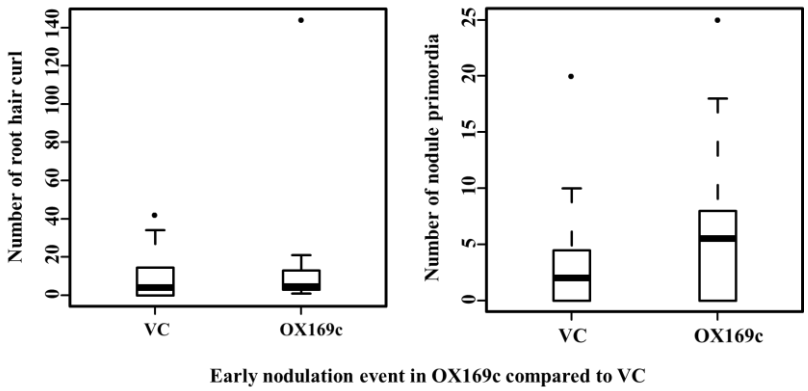


Figure 5.11 Effect of overexpression of miR169 in early nodulation events. ***A.*** *root hair curl (RHC)* ***B.*** *nodule primordium (NP). Number of RHC and NP was counted 7 dpi after B. japonicum inoculation. The significant difference in nodule number was tested by using Poisson regression (*** P <0.001, ** P <0.01, and * P <0.05).*

5.2.6. miR169c regulation of downstream genes

A network analysis of the root nodule transcriptome (Chapter 2) identified a cytokinin oxidase (CKX) and Gibberellin 20-oxidase (GA20ox) module that acts downstream of the Nuclear Transcription Factor Y subunit alpha (NFY-A) transcription factor, NFY-As are cleaved by miR169c. To understand the miR169c-NFY-A mediated regulation of downstream genes, expression levels of GA20ox and CKX was determined. RNA was isolated from uninoculated roots (without nodules) and inoculated roots (with nodules harvested at 14 dpi and 28 dpi) of OX169c and ENOD169c lines.

A significant increase in miR169c expression level was observed in root of OX169c lines but not in ENOD169c lines (*t*- test, $P < 0.05$) (Figure 5.12, A). This could be due to differences in the promoter in the two lines. ENOD40 primarily expresses at emerging and mature stages of nodule but not in the root, and root samples in ENOD169c were devoid of nodules. At 14dpi, a significant increase in miR169c expression level was observed only in OX169c (*t*- test, $P < 0.05$) (Figure 5.12, B). At 24 dpi, significant increases in miR169c levels were observed in both OX169c and ENOD169c (*t*- test, $P < 0.05$) (Figure 5.12, C).

Among the three targets of miR169c, the expression levels of Glyma10g10240 and Glyma15g18970 were significantly reduced in the uninoculated root tissues of OX169c, consistent with the significant increase in miR169c expression in these roots (Figure 5.13 A, D). Moreover, expression of Glyma10g10240 and Glyma17g05920 were significantly increased in ENOD169c roots, though no significant difference in miR169c level was observed in these roots (*t*- test, $P < 0.05$) (Figure 5.13, A, G). At 14 dpi, expression of Glyma10g10240 was significantly decreased in the ENOD169c line, but none of the targets were significantly reduced in the OX169c line (Figure

5.13, B, E, and H). At 24 dpi, the expression of Glyma10g10240 and Glyma17g05920 was significantly reduced in ENOD169c roots, but none of the targets of miR169c were significantly reduced in OX169c (Figure 5.13 C, F, and I). These inconsistencies in expression of miR169c and target genes could be attributed to sample-to-sample expression level variations, to dilution of expression levels due to harvest of whole root, and/or to temporal regulation of miR169c.

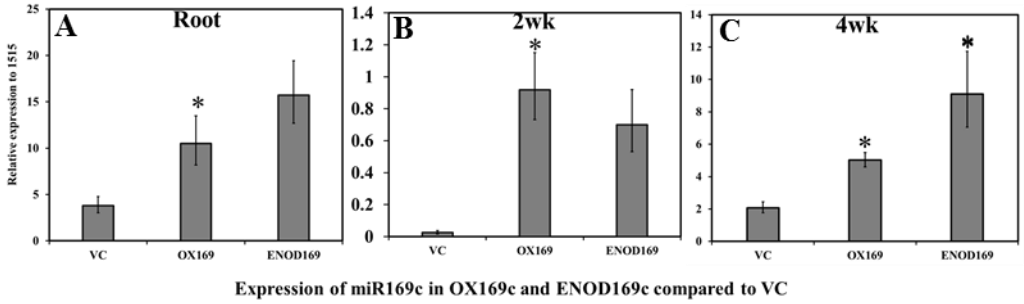


Figure 5.12 Relative expression of miR169c in OX169c and ENOD169c lines compared to vector control (VC). Transgenic hairy roots of composite soybean plants were harvested before and after inoculation for gene expression analysis. A. unincoculated root (root from petri dish, without nodules) B. 14 dpi root with nodules and C. 28 dpi root with nodules. Relative expression refers to gene expression based on threshold cycle (ct) normalized with housekeeping gene miR1515 using three technical replicates. indicates significant expression (t-test, P <0.05).*

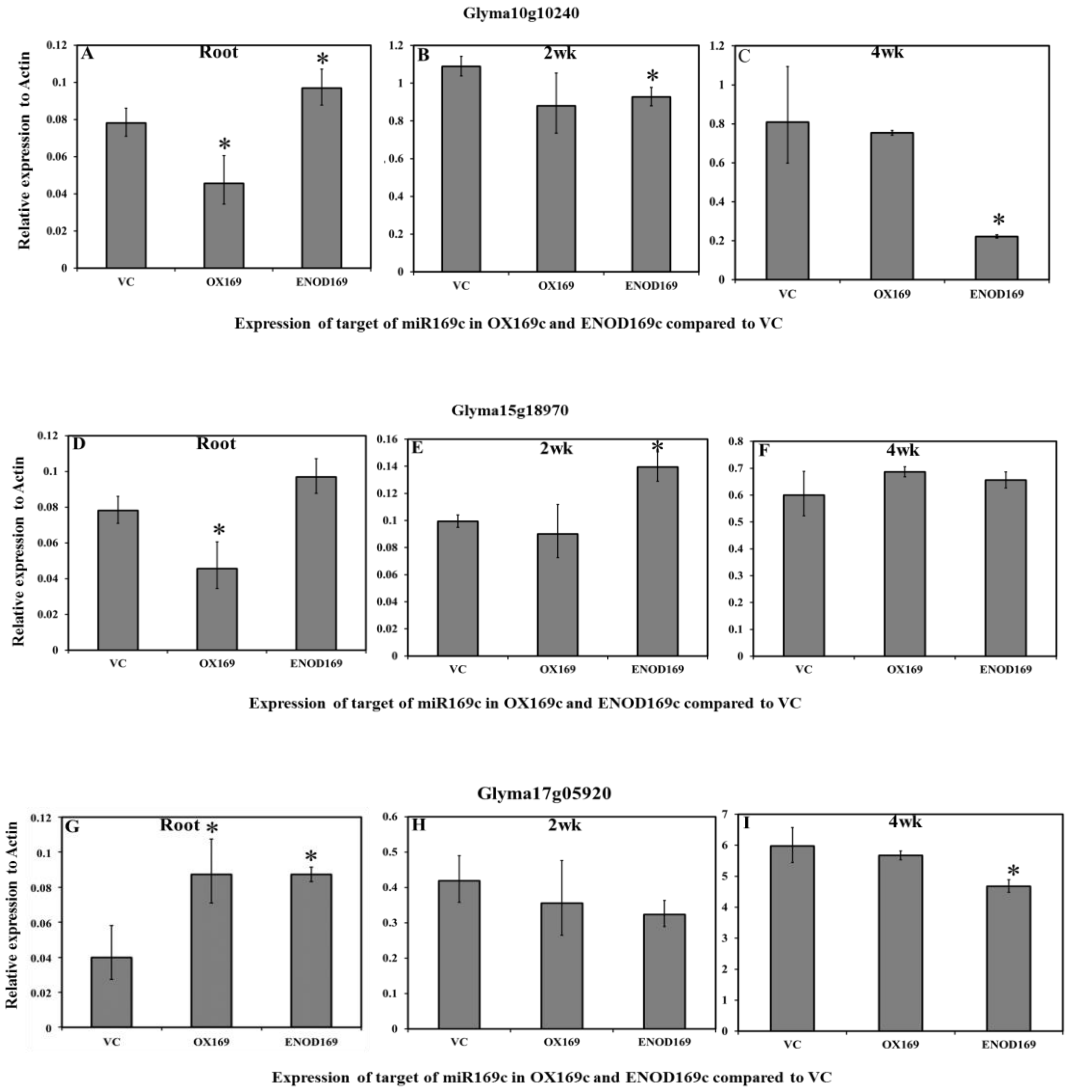


Figure 5.13 Relative expression of targets of miR169c in OX169c and ENOD169c lines compared to vector control (VC). Transgenic hairy roots of composite soybean plants were harvested before (uninoculated roots) and after inoculation (root with nodules at 14 dpi and 24 dpi) for gene expression analysis. ***A, B, C,*** *Glyma10g10240 in root, 14 dpi root with nodule and 24 dpi root with nodule, respectively.* ***D, E, F*** *Glyma15g18970 in root, 14 dpi root with nodule and 24 dpi root with nodule, respectively.* ***G, H, I*** *Glyma17g05920 in root, 14 dpi root with nodule and 24 dpi root with nodule, respectively. Relative expression refers to gene expression based on threshold cycle (ct) normalized with housekeeping gene GmActin using three technical replicates. * indicates significant expression (t-test, P <0.05).*

In the root tissue, all the GA20ox gene family members used in this study were significantly reduced in the OX169c line compared to the VC (*t*-test, $P<0.05$) (Figure 5.14 A and D). This pattern is consistent with the observed significant increase in expression for miR169c and decrease for two of the three targets in the OX169c line (Figure 5.12 A, Figure 5.13 A and D). However, the GA20ox genes were significantly increased in the ENOD169c roots compared to the VC (*t*-test, $P<0.05$) (Figure 5.14 A, D). No significant reduction of GA20x expression level was observed at 14 dpi in either OX169 or ENOD169 lines (Figure 5.14 B, E). However, at 24 dpi, significant reduction of both GA20ox gene family members was observed in both the OX169c and ENOD169c lines (*t*-test, $P<0.05$) (Figure 5.14 C and F).

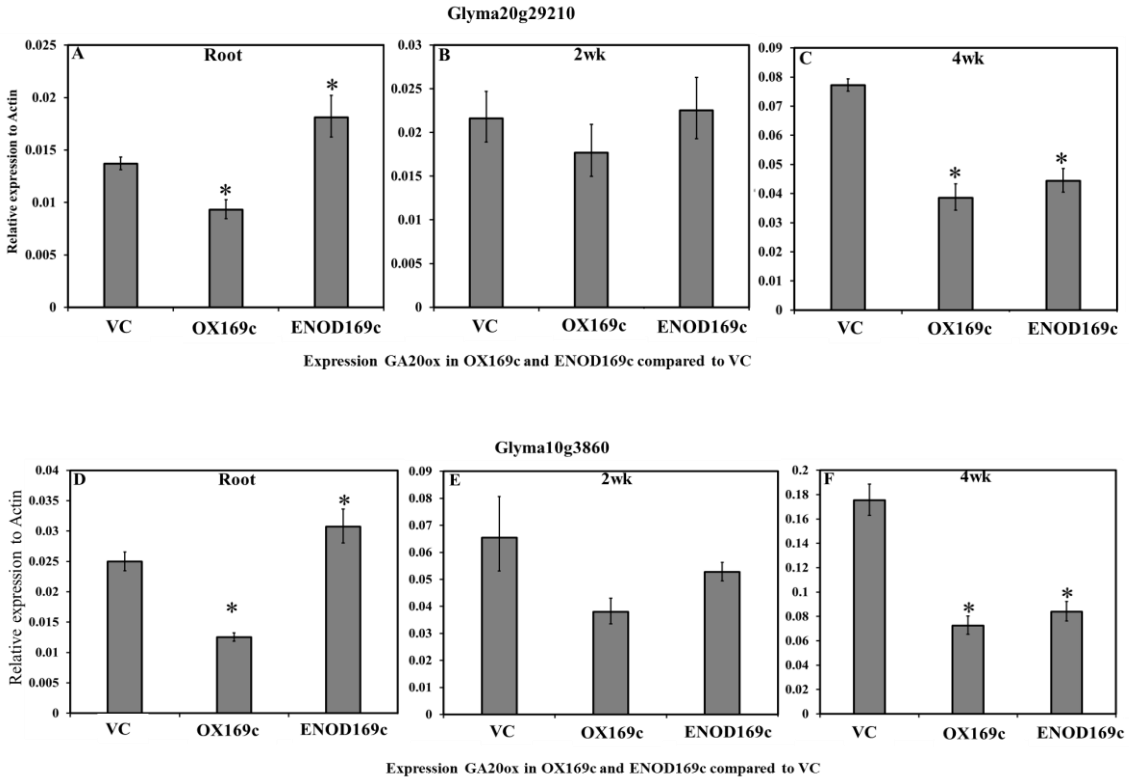


Figure 5.14 miR169c regulation of GA20ox gene. Transgenic hairy roots of composite soybean plants were harvested before (uninoculated roots) and after inoculation (root with nodules at 14 dpi and 24 dpi) for gene expression analysis ***A, B, C*** *Glyma20g29210 (a GA20ox) in root, 14-dpi root with nodule, and 24-dpi root with nodule, respectively.* ***D, E, F*** *Glyma10g38600 (a GA20ox) in root, 14-dpi root with nodule, and 24-dpi root with nodule, respectively. Relative expression refers to gene expression based on threshold cycle (ct) normalized with housekeeping gene GmActin using three technical replicates. * Indicates significant expression (t-test, P <0.05).*

Similar to GA20ox genes, all the CKX gene family members used in this study were significantly reduced in the OX169c line compared to the VC in uninoculated roots (*t*-test, $P<0.05$) (Figure 5.15 A, D, and G). However, among the three CKX members, Glyma13g16430 and Glyma17g06220 showed significant increases in ENOD169c uninoculated roots (Figure 5.15 A, D). At 14 dpi, Glyma13g16420 and Glyma17g06220 were significantly reduced in the ENOD169c line (Figure 5.15 B and E). At 24 dpi, Glyma17g06220 was significantly reduced in ENOD169c, and Glyma13g16420 was significantly reduced in the OX169c line (*t*-test, $P<0.05$) (Figure 5.15 F and I). Overall, a consistent pattern of an increase in miR169c expression followed by the down regulation of the target gene Glyma10g10240 and of members of the GA20x and CKX gene families was observed only in in the uninoculated roots of the OX169c and ENOD169c lines, but not in root with nodules at 14 dpi or 24 dpi.

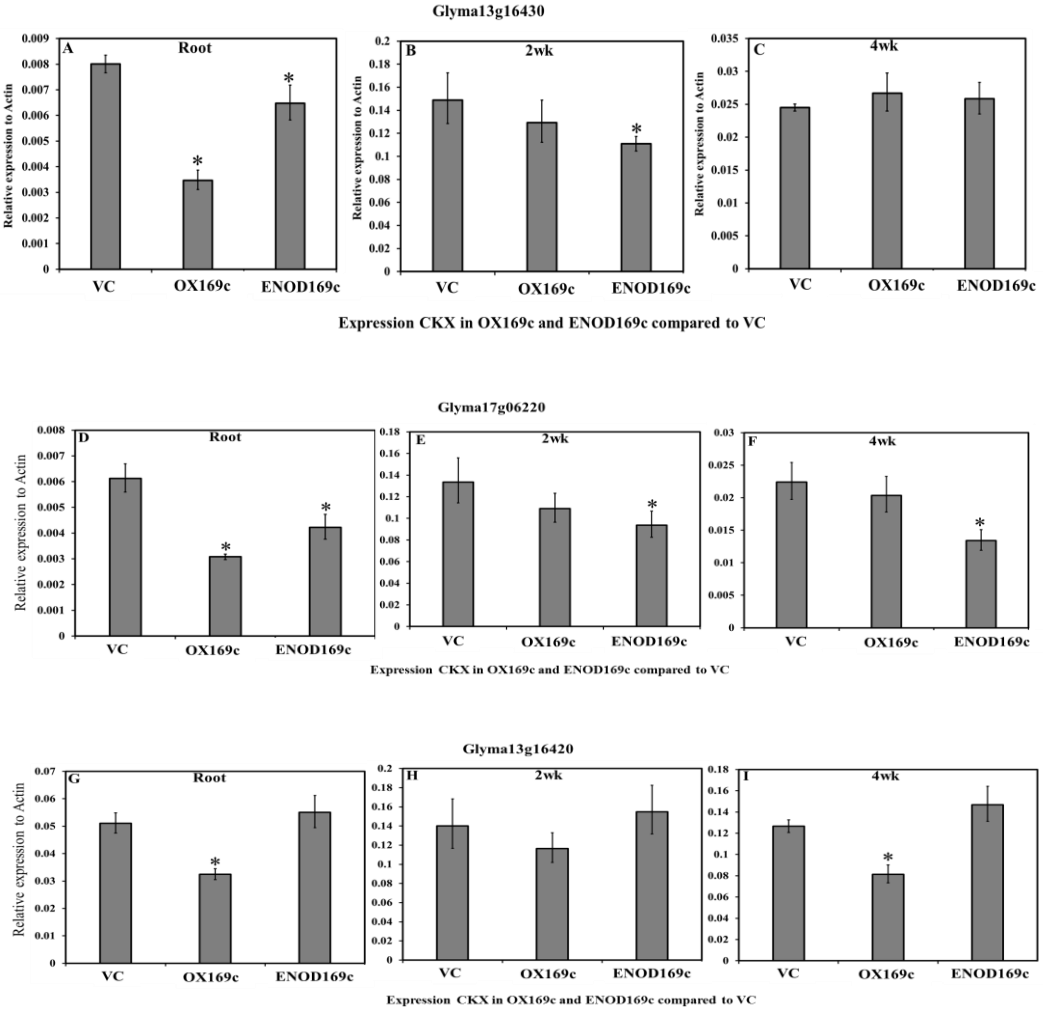


Figure 5.15 miR169c regulation of CKX genes. Transgenic hairy roots of composite soybean plants were harvested before (uninoculated roots) and after inoculation (root with nodules at 14 dpi and 24 dpi) for gene expression analysis. ***A, B, C*** *Glyma13g16430 (CKX) in root, 14-dpi root with nodule, and 24-dpi root with nodule, respectively.* ***D, E, F*** *Glyma17g06220 (CKX) in root, 14-dpi root with nodule and 24-dpi root with nodule, respectively.* ***G, H, I,*** *Glyma13g16420 (CKX) in root, 14-dpi root with nodule and 24-dpi root with nodule, respectively. Relative expression refers to gene expression based on threshold cycle (ct) normalized with housekeeping gene GmActin using three technical replicates. * Indicates significant expression (t-test, P <0.05).*

5.2.7. Spatial regulation of miR169c during root nodule formation

To understand the spatial expression of miR169c during root nodule development, a miR16c sensor was constructed by creating a construct that had the miR169c binding site at the 3'UTR of the Tdtomato (tdt) fluorescent marker gene driven by CsVMV promoter. The construct also included a GFP marker driven by super Ubiquitin GFP promoter that would produce a green signal to help identify the transgenic roots from the non-transgenic ones. The hypothesis is that endogenous miR169c would bind and cleave the tdt marker at the 3'UTR, and this cleaved construct would result in only green signal. If the miR169c doesn't cleave the mRNA, there would be expression of both the red signal generated by the tdt marker and the green signal generated by the GFP, a combination that would result in a yellow signal. A VC was generated without the miRNA-binding site.

Scanning of the entire root tips of the plant did not reveal any difference in the root tips of the VC and miR169c-sensor plants (Figure 5.16). Similarly, no differences in emerging nodules were observed between the VC and the miR169c-sensor lines (Figure 5.17). Hand-sectioned mature nodules were obtained using confocal optical sections (3 μm). No clear pattern of expression of miR169c was observed in the miR169c-sensor line compared to the VC line (Figure 5.18).

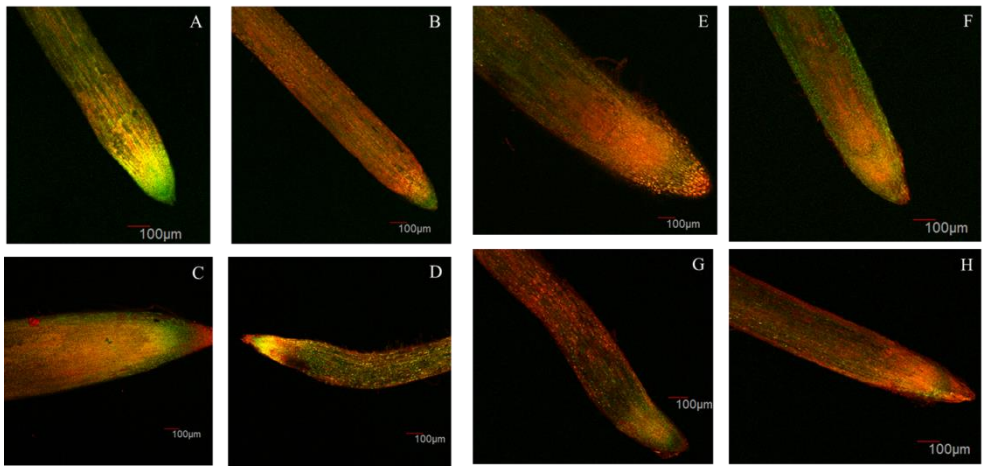


Figure 5.16 Spatial regulation of miR169c in root tip of soybean plants. Composite hairy root plants carrying a miR169c binding site (miR169c sensor plants) and vector control (without miR169c binding site) were inoculated with B. japonicum. Transgenic roots were harvested at 7 dpi after inoculation for the microscopic study. Entire root tips were used for laser dissection microscopy. Vector control (VC): ***A, B, C,*** *and* ***D****. miR169-sensor:* ***E, F, G,*** *and* ***H****. Scale bars =100 μm.*

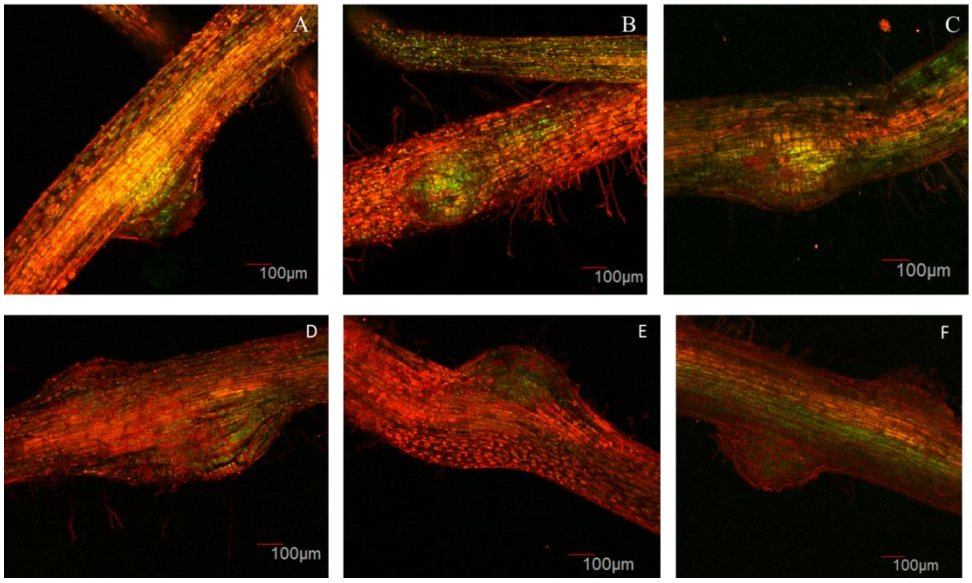


Figure 5.17 Spatial regulation of miR169c in emerging nodules of soybean plants. Composite hairy root plants carrying a miR169c binding site (miR169c-sensor plants) and vector control (without miR169c binding site) were inoculated with B. japonicum. Transgenic roots were harvested at 7 dpi for the microscopic study. Entire nodules were used for laser dissection microscopy. Vector control: ***A, B, C,*** *and* ***D****. miR169-sensor:* ***E, F, G,*** *and* ***H****. Scale bars =100 µm.*

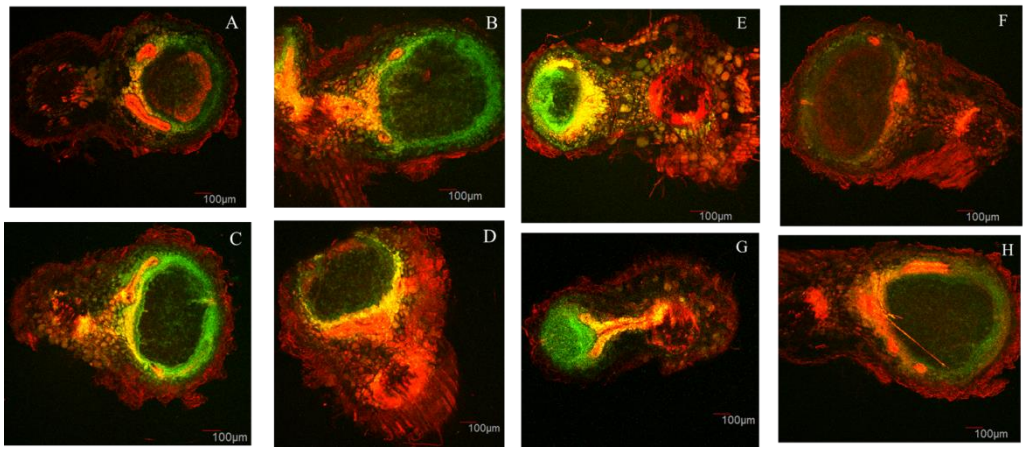


Figure 5.18 Spatial regulation of miR169c in mature nodules of soybean plants. Composite hairy root plants carrying a miR169c binding site (miR169c-sensor plants) and vector control (without miR169c binding site) were inoculated with B. japonicum. Nodules in transgenic roots were harvested at 14 dpi and sectioned for the microscopic study. Vector control: ***A, B, C,*** *and* ***D****. miR169sensor:* ***E, F, G,*** *and* ***H****. Scale bars =100 µm.*

5.3. Discussion

Recent studies have identified miRNAs that play regulatory roles in root nodule development (See list Chapter 1, Table 1.1). However, very few miRNAs that play spatial regulation roles during nodule development are known. A notable example of miRNA-mediated spatial regulation of nodule development is miR169c. miR169c regulates nodule development in *M. truncatula* by spatial regulation of the NFY-A TF. Analysis of nodule-specific miRNAs in soybean identified miR2218, miR169c, and miR1513c. These miRNAs were highly expressed in the roots compared to nodule and targets of these miRNAs were highly expressed in the nodules (Turner et al. 2012). To test the hypothesis that these nodule-excluded miRNAs play a spatial regulation in soybean nodule development, hairy root composite soybean plants were generated with overexpression and misexpression of miR2218, miR169c, and miR1513c.

Verification of overexpression by RT-qPCR showed that all overexpression lines resulted in a significant increase in miR2218, miR169c and miR1513c expression, respective to the line analyzed. Further support for overexpression was seen as decreases in expression of genes targeted by miR2218 and miR169c. Surprisingly, a significant increase in the target gene of miR1513c was observed. Verification of misexpression of miR2218, miR169c and miR1513c was done by using nodule tissues harvested at 14 dpi. None of the miRNAs were significantly increased in misexpression lines compared to the VC line. However, the targets of miR2218 and miR1513c and one of the targets of miR169c were significantly reduced. These results suggest high turnover of nodule-excluded miRNAs from the nodule tissue. Among the three targets of miR169c, only expression of Glyma10g10240 was significantly reduced, possibly due to tissue-specific

expression of the miR169c targets inside the nodule tissues and their regulation by miR169c.

Nodule numbers at 14 dpi were quantified to understand how expression of miR2218, miR169c and miR1513c inside the nodule would affect nodule formation. Among the three studied miRNAs, miR169c misexpression showed a significant increase in the number of emerging nodules across the three replicates and a decrease in the number of mature nodules in at least 2/3 of the replicates. To understand the mechanism behind increased numbers of emerging nodules in miR169c, two approaches were taken. First, nodulation time was extended from 14 dpi to 24 dpi. Analyses of nodulation data at 24 dpi showed an increase in the emerging, mature and total nodule numbers, suggesting that miR169c might play a role in nodule maturation. Second, early nodulation events, namely root hair curls (RHC) and nodule primordia (NP), were counted at 7 dpi after inoculation with *B. japonicum* expressing nptII: GUS. No effect in RHC numbers was observed in at least 2/3 of the replicates in ENOD169c lines, although a significant decrease in NP was obtained in 2/3 of the replicates. Although the decrease in NP numbers at 7 dpi is contradictory to the increase in emerging nodules at 14 dpi, the number of mature nodules was decreased at 14 dpi, indicating that the increase in emerging nodules could be due to delay in the maturation of the nodules. It is also possible that miR169c-mediated regulation is independent of the initial nodulation event. In OX169c lines, the RHC and NP numbers showed a significant increase. However, nodule count at 24 dpi in the OX169 lines showed a significant decrease in number of mature nodules, but no effect on emerging or total number of nodules. Difference in nodule phenotype between OX169c and ENOd169c lines could be due to specificity of

ENOD40 promoter and significant silencing of the expression of miR169c inside the nodules compared to the constitutive promoter. Overall, nodule phenotype data suggests that miR169c plays an important role in nodule development, but further validation with additional experiments, such as engineering miR-resistant targets and overexpression or misexpression of miR169c in a super-nodulting mutant root, are necessary.

To determine the response of genes downstream of NFY-A, which is under the regulation of miR169c, gene expression analysis of GA20X and CKX gene family members were determined by RT-qPCR. RNA from uninoculated root and inoculated root with nodules (sampled at 14 and 24 dpi) were used for RT-qPCR analysis. First, the expression of miR169c and its targets in the OX169c and ENOD169c lines were determined in all root tissue types (uninoculated root, inoculated root at 14 dpi and 24 dpi). Expression of miR169c was significantly increased in the three tissue types in OX169, but was significantly increased in ENOD169c only at 24 dpi. The targets of miR169c also did not show consistent significant reductions across all tissue types or with time. This could be due to tissue-specific expression of the targets or temporal and spatial regulation of the targets by mR169c.

Consistent decreases in both target (NFY-A) levels and in the downstream GA20ox and CKX members were obtained only in uninoculated OX169c roots. At 14 dpi, no significant increase in miR169c expression was observed in ENOD169c compared to VC plants; however, Glyma10g10240, one of the targets of miR169c, was significantly reduced. This result is consistent with expression of miR169c and Glyma10g10240 in nodule-only samples from ENOD169c line. An opposite expression pattern was observed between miR169c and its target Glyma10g101240 in both the

OX169c and ENDO169c lines at 14 dpi. This suggests that miR169c might rapidly turnover inside the nodules and that the expression of miR169c is more effective when using nodule-specific promoter rather than a constitutive promoter. At 14 dpi, none of the GA20x members were significantly reduced in OX169c and ENOD169c lines. Nonetheless, two CKX members (Glyma13g16430 and Glyma17g06220) showed significant decrease in the ENOD169c line. At 24 dpi, a significant decrease in miR169c level was observed only in OX169c and ENOD169c but target levesl were significantly reduced only in ENOD169c roots. Significant decreases in both GA20x members (Glyma20g29210, Glyma10g38600) were observed in both the OX169c and ENOD169c lines. However, significant reductions were seen for different members of the CKX family depending on the lines: Glyma13g16420 was reduced in OX169c but Glyma17g06220 was reduced in ENOD169c. This suggests that there is spatiotemporal regulation of GA20ox and CKX genes by miR169c.

A consistent expression pattern of miR169 and its targets as well as the downstream signaling targets in the GA20x and CKX families was not observed in inoculated root tissues with nodules at 14 dpi or 24 dpi, as was observed in the OX169c uninoculated root tissue. This could be attributed to the different natures of the samples used for expression analyses in these tissues. Uninocualted roots were devoid of nodules and showed consistent, significant reductions in expression of Glyma10g10240 and members of the GA20X and CKX families, but roots at 14 dpi and 24 dpi had varied numbers and size of nodules. Some of these nodules might have been harvested during a period of low expression for miR169c. Harvesting size- or developmentally specific nodules at each time point could ameliorate this problem.

The spatial expression of miR169 was determined by generating a sensor construct with a miRNA cleavage site at the 3'UTR of the fluorescent tdt marker gene driven by the CsVMV promoter (miR169-sensor). No differences in spatial expression of miR169 between emerging nodules, lateral root tips and mature nodules sections were observed in roots expressing the miR169c-sensor construct vs. VC. In the plants harboring the VC, we would expect to see yellow signal due to the absence of the miRNA cleavage site, but in the miR169c-sensor line we would expect a green signal due to presence of miRNA cleavage site and elimination of the red signal. However, the vector control lines showed both red and green fluorescence (making yellow). This could be the possible reason that no clear pattern of miR169c activity was observed in our data.

5.4. Conclusion

The nodulation study showed that misexpression of miR169c leads to increase in number of emerging nodules, suggesting that miR169c might delay maturation of root nodules. Study of RHC and NP suggested miR169c regulation might be independent of early nodulation events. From our network analyses, we would expect to see decreases in expression of NFY-A members under the regulation of miR169c and subsequent reduction in expression levels of members of GA20x and CKX. The decrease in CKX and GA20ox would relate to the observed delay in maturation and increase in emerging nodules (Figure 5.19). But this expected pattern was seen only in the uninoculated roots and in the ENOD169c roots at 24 dpi. Based on these results, it is likely that miR169c affects nodule development by spatial and temporal regulation.

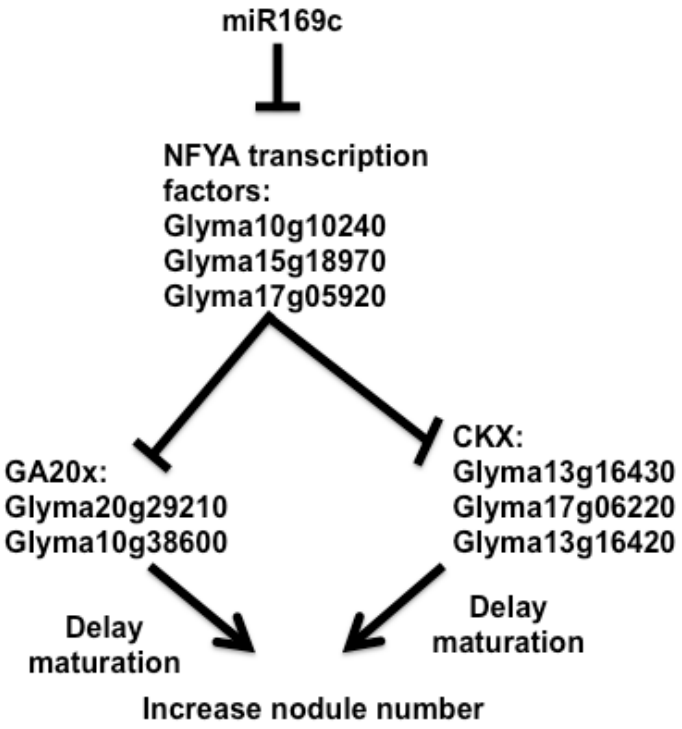


Figure 5.19 Schematic representation of miR169 action in nodule development. Increases in expression of miR169c causes decreased expression of NFY-A TFs, which is followed by decreased expression of GA20x and CKX genes. Decreases in GA20x and CKX genes might have an effect on nodule maturation, leading to an increase in emerging nodules (nodules with delayed maturation) and total nodules.

References

Axtell, M.J. (2013) Classification and comparison of small RNAs from plants. *Annual review of plant biology*, 64, 137-159.

Boualem, A., Laporte, P., Jovanovic, M., Laffont, C., Plet, J., Combier, J.P., Niebel, A., Crespi, M. and Frugier, F. (2008) MicroRNA166 controls root and nodule development in Medicago truncatula. *The Plant journal : for cell and molecular biology*, 54, 876-887.

Chen, X. (2005) microRNA biogenesis and function in plants. *FEBS Letters*, 579, 5923-5931.

Collier, R., Fuchs, B., Walter, N., Kevin Lutke, W. and Taylor, C.G. (2005) Ex vitro composite plants: an inexpensive, rapid method for root biology. *The Plant journal : for cell and molecular biology*, 43, 449-457.

Combier, J.P., Frugier, F., de Billy, F., Boualem, A., El-Yahyaoui, F., Moreau, S., Vernie, T., Ott, T., Gamas, P., Crespi, M. and Niebel, A. (2006) MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in Medicago truncatula. *Genes & development*, 20, 3084-3088.

El Yahyaoui, F., Küster, H., Amor, B.B., Hohnjec, N., Pühler, A., Becker, A., Gouzy, J., Vernié, T., Gough, C., Niebel, A., Godiard, L. and Gamas, P. (2004) Expression Profiling in Medicago truncatula Identifies More Than 750 Genes Differentially Expressed during Nodulation, Including Many Potential Regulators of the Symbiotic Program. *Plant Physiology*, 136, 3159-3176.

Jones-Rhoades, M.W., Bartel, D.P. and Bartel, B. (2006) MicroRNAS and their regulatory roles in plants. *Annual review of plant biology*, 57, 19-53.

Li, F., Pignatta, D., Bendix, C., Brunkard, J.O., Cohn, M.M., Tung, J., Sun, H., Kumar, P. and Baker, B. (2012) MicroRNA regulation of plant innate immune receptors. *Proceedings of the National Academy of Sciences*, 109, 1790-1795.

Li, H., Deng, Y., Wu, T., Subramanian, S. and Yu, O. (2010) Misexpression of miR482, miR1512, and miR1515 increases soybean nodulation. *Plant Physiol*, 153, 1759-1770.

Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif.)*, 25, 402-408.

Nova-Franco, B., Iniguez, L.P., Valdes-Lopez, O. and Alvarado-Affantranger, X. (2015) The micro-RNA72c-APETALA2-1 node as a key regulator of the common bean-Rhizobium etli nitrogen fixation symbiosis. 168, 273-291.

Rogers, K. and Chen, X. (2013) Biogenesis, Turnover, and Mode of Action of Plant MicroRNAs. *The Plant Cell*, 25, 2383-2399.

Shivaprasad, P.V., Chen, H.M., Patel, K., Bond, D.M., Santos, B.A. and Baulcombe, D.C. (2012) A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. *Plant Cell*, 24, 859-874.

Subramanian, S., Fu, Y., Sunkar, R., Barbazuk, W.B., Zhu, J.K. and Yu, O. (2008) Novel and nodulation-regulated microRNAs in soybean roots. *BMC genomics*, 9, 160.

Sunkar, R., Li, Y.-F. and Jagadeeswaran, G. (2012) Functions of microRNAs in plant stress responses. *Trends in Plant Science*, 17, 196-203.

Turner, M., Nizampatnam, N.R., Baron, M., Coppin, S., Damodaran, S., Adhikari, S., Arunachalam, S.P., Yu, O. and Subramanian, S. (2013) Ectopic Expression of miR160 Results in Auxin Hypersensitivity, Cytokinin Hyposensitivity, and Inhibition of Symbiotic Nodule Development in Soybean. *Plant Physiology*, 162, 2042-2055.

Turner, M., Yu, O. and Subramanian, S. (2012) Genome organization and characteristics of soybean microRNAs. *BMC genomics*, 13, 169.

Wang, Y., Wang, Z., Amyot, L., Tian, L., Xu, Z., Gruber, M.Y. and Hannoufa, A. (2015) Ectopic expression of miR156 represses nodulation and causes morphological and developmental changes in Lotus japonicus. *Molecular genetics and genomics : MGG*, 290, 471-484.

Yan, Z., Hossain, M.S., Wang, J., Valdes-Lopez, O., Liang, Y., Libault, M., Qiu, L. and Stacey, G. (2013) miR172 regulates soybean nodulation. *Molecular plant-microbe interactions : MPMI*, 26, 1371-1377.

Yang, W.C., Katinakis, P., Hendriks, P., Smolders, A., de Vries, F., Spee, J., van Kammen, A., Bisseling, T. and Franssen, H. (1993) Characterization of GmENOD40, a gene showing novel patterns of cell-specific expression during soybean nodule development. *The Plant journal: for cell and molecular biology*, 3, 573-585.

Appendix A

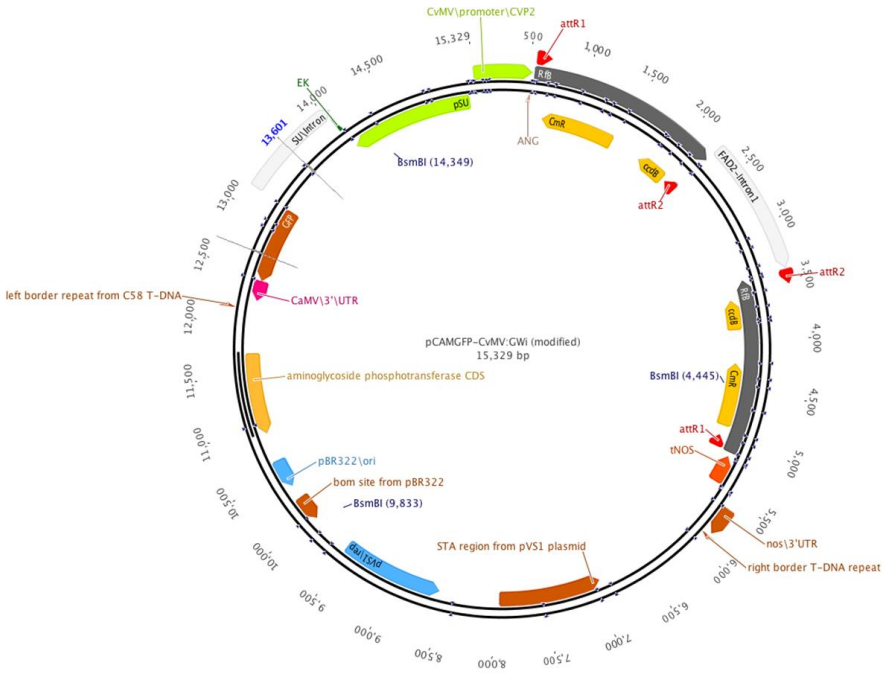


Figure 1 Vector map of pCAMGFP-CsVMV: GW

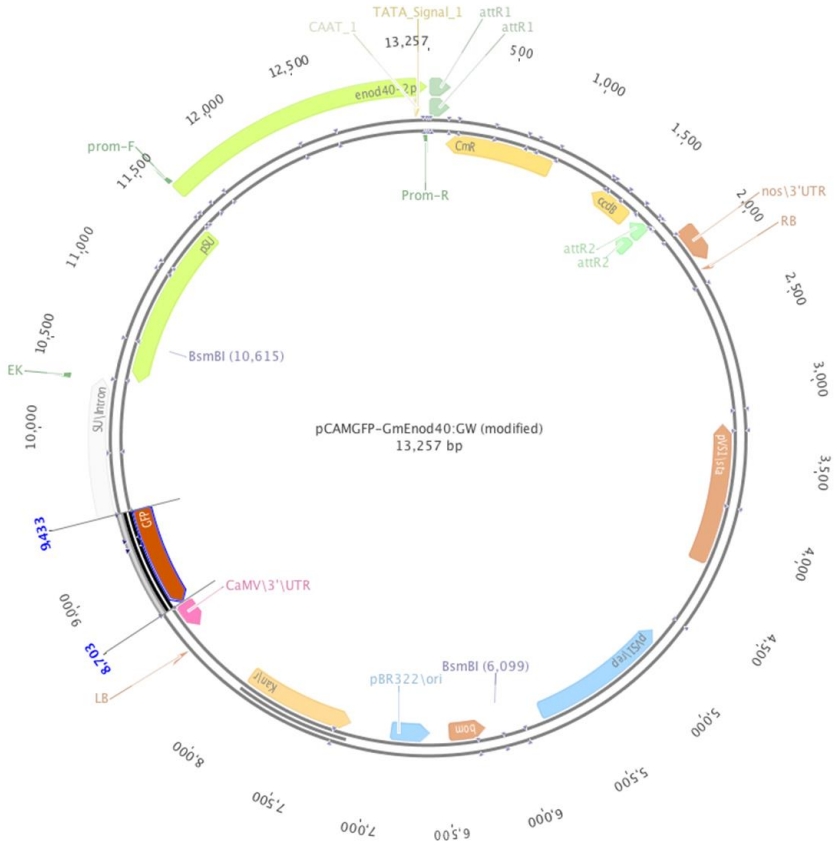


Figure 2 Vector map of pCAMGFP-ENOD40: GW

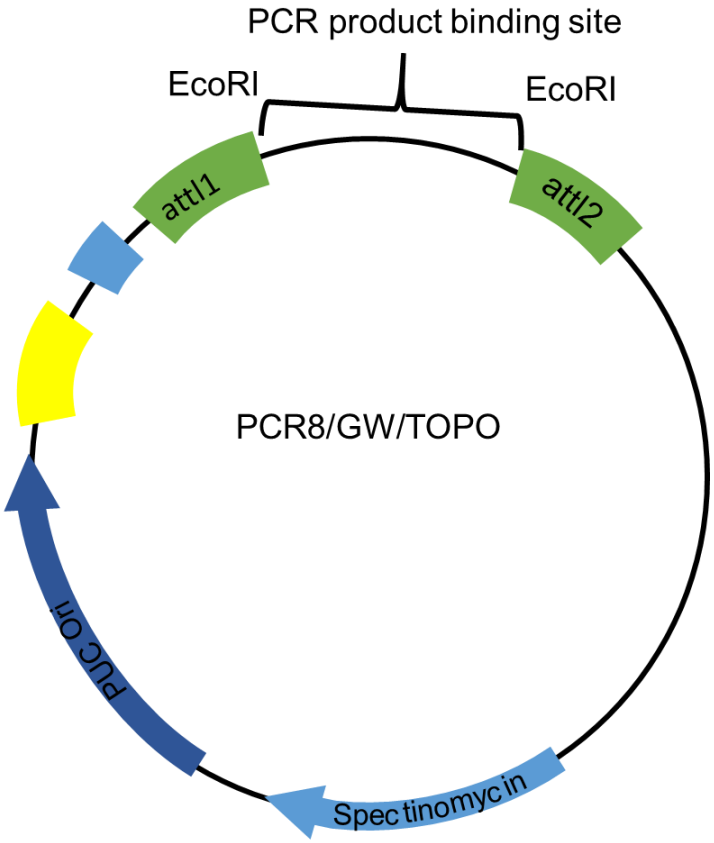


Figure 3 PCR8/GW/TOPO vector map (Adapted from ThermoFisher Scientific).