2019

Nodule Zone-Specific Gene Expression in Soybean

Sadikshya Aryal
South Dakota State University

NODULE ZONE-SPECIFIC GENE EXPRESSION IN SOYBEAN

BY

SADIKSHYA ARYAL

A thesis submitted in partial fulfillment of the requirements for the

Master of Science

Major in Plant Science

South Dakota State University

2019

THESIS ACCEPTANCE PAGE

Sadikshya

This thesis is approved as a creditable and independent investigation by a candidate for the master's degree and is acceptable for meeting the thesis 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
Advisor Date

David Wright
Department Head Date

Dean, Graduate School Date

I would like to dedicate my thesis to my parents, Hari Prasad Aryal and Shanta Aryal

ACKNOWLEDGEMENT

I would like to express my hearty thanks to my advisor Dr. Senthil Subramanian for the opportunity he provided me to be the part of his research group. His continuous guidance, support, and encouragement were invaluable. I would like to thank Dr. Jose Gonzalez and Dr. Yeyan Qiu for their contribution through the Genome Sequencing Core facility at SDSU. I would like to express my gratitude to members of my master's thesis committee, Dr. Jose Gonzalez and Dr. Hesham Fahmy, and former member Dr. Qin Ma, for their valuable feedbacks and suggestions.

I would like to express my vote of thanks to all my lab members, Dr. Bhanu Prakash Petla (for guiding and working with me during the initial stages of my research), Dr. Suresh Damodaran, Dr. Laura White, Spencer Schreier, Sunita Pathak, Paul Gaillard, Pratiksha KC, Jesus Loya, Dr. Marina Johnson and Mucahid Bozkus for helping and training me in various aspects of my research projects. I would like to thank all my course teachers for contributing to increasing my level of knowledge.

Reverence and hearty thanks to my family and friends for being with me and pushing me forward in every aspect of my life. I would like to take a moment to thank all my friends in Brookings for making Brookings a beautiful place to live.

TABLE OF CONTENTS

LIST OF FIGURES ix
LIST OF TABLES ix
ABSTRACT xi
1 INTRODUCTION 1
1.1 Nitrogen in agriculture 1
1.1.1 Importance of Nitrogen 1
1.1.2 Availability of nitrogen 1
1.2 Biological Nitrogen Fixation 2
1.2.1 Free-living nitrogen fixation 3
1.2.2 Associative nitrogen fixation 4
1.2.3 Symbiotic nitrogen fixation 5
1.3 Legume-Rhizobia Symbiosis 5
1.4 Nodule development 6
1.4.1 Indeterminate nodules 8
1.4.2 Determinate nodules 8
1.5 Nutrient exchange between host plant and nitrogen fixing bacteroids 10
1.6 Molecular mechanism of bacterial infection and nodule organogenesis 12
1.6.1 Rhizobial infection and nod factor (NF) perception 12
1.6.2 NF signaling cascade 13
1.6.3 The early stages of the nodulation 15
1.7 Determinate nodule- zones differentiation 16
1.8 Cell type-specific gene expression 19
1.8.1 Regulation of gene expression 19
1.8.2 Cell type-specific gene expression 20
1.8.3 INTACT 22
1.8.4 TRAP 22
2 MATERIAL AND METHODS 26
2.1 Gene cassettes preparation 26
2.1.1 Verification of destination vector 26
2.1.2 Verification of expression vector for Enod2 and Enod40 27
2.1.3 Construction of PMH40-CsVMV entry vector 28
2.1.4 Construction of expression vector for CsVMV 29

2.2 Growth and maintenance of soybean plants 30
2.3 Hairy Root Transformation..... 31
2.3.1 Preparation of competent cells of *Agrobacterium rhizogenes* K599 strain 31
2.3.2 Electroporation-mediated transfer of construct in *Agrobacterium rhizogenes* K599 strain..... 31
2.3.3 *Agrobacterium rhizogenes* mediated hairy root plant transformation. 32
2.4 Plant screening and transplanting 34
2.5 Inoculation of rhizobia..... 34
2.6 Microscopy 34
2.6.1 Fluorescence microscopy 34
2.6.2 Vibratome sectioning 35
2.6.3 Confocal microscopy..... 35
2.7 Translating ribosomes affinity purification (TRAP) 35
2.8 Western blotting..... 37
2.9 RNA isolation 38
2.10 DNase treatment 38
2.11 Library preparation 39
2.11.1 mRNA enrichment 39
2.11.2 Anneal the cDNA synthesis primer..... 41
2.11.3 Synthesize cDNA 41
2.11.4 Synthesize 3′-tagged DNA 42
2.11.5 Purify the cDNA..... 42
2.11.6 PCR amplify the library and add an index (barcode)..... 43
2.11.7 Purify the RNA-Seq library..... 43
2.11.8 Assess library quantity and quality 43
2.12 cDNA synthesis 43
2.13 DNA contamination test 44
2.14 Reverse Transcription – quantitative Polymerase Chain Reaction (RT-qPCR)44
2.15 Gene Annotation..... 45
3 RESULTS 46
3.1 Localization of the tissue-specific promoter activity..... 46

3.2 Affinity purification of tagged ribosomes 49
3.3 Library Preparation 50
3.4 mRNA Enrichment 51
3.5 Library preparation with rRNA-depleted RNA 52
3.6 Analysis of marker gene expression patterns at 7 and 10 dpi time points .. 53
3.6.1 Analysis of marker gene expression patterns at 7 dpi time point 54
3.6.2 Analysis of marker gene expression patterns at 10 dpi time point 55
3.7 Evaluation of precursor microRNA and non-coding RNA 57
3.8 Differential gene expression between the two nodule zones 58
3.8.1 Differential gene expression between two nodule zones at 7 dpi 58
3.8.2 Differential gene expression between two nodule zones at 10 dpi 60
3.9 Comparative enrichment analysis between nodule parenchyma and infection zone 63
4 DISCUSSION 65
4.1 Nodule zone-specific promoters 65
4.2 Tagged ribosomes affinity purification 66
4.3 mRNA enrichment 68
4.4 Library Synthesis 68
4.5 Expression of auxin-related genes 70
4.6 Expression of transcription factor families 71
4.7 Nitrate and phosphate related genes 72
5 CONCLUSION 74
REFERENCES 76
APPENDICES 84
Appendix A: Verification of destination vector 84
Appendix B: Verification of expression vector 85
Appendix C: Verification of expression vector 86
Appendix D: Verification of entry vector 87
Appendix E: Verification of expression vector 88
Appendix F: Composition of Hoagland solution 89
Appendix G: Composition of nitrogen free plant nutrient solution 90
Appendix H: Composition of Vincent - rich media 91
Appendix J: List of qPCR primers used in this study 93

Appendix K: Log_2 fold change (ENOD40p-/ENOD2p-TRAp derived samples) at 7dpi 94

Appendix L: Log_2 fold change (ENOD40p-/ENOD2p-TRAP derived samples) at 10dpi 95

Appendix M: List of cloning primers used in this study 96

Appendix N: List of candidate genes for gene expression analysis 97

Appendix O: Agarose gel image of RNA 99

LIST OF FIGURES

Figure 1.1 Stages of infection during nodule development.. 7
Figure 1.2 Two major types of root nodules.. 9
Figure 1.3 Early molecular signals in root nodule development.. 16
Figure 2.1 Hairy rot plant transformation in soybean.. 33
Figure 3.1 TRAP constructs used in the study 46
Figure 3.2 A CsVMVp-TRAP transgenic root containing nodules.. 47
Figure 3.3. A ENOD40p-TRAP transgenic root containing nodules.. 47
Figure 3.4 Spatio-temporal localization of GFP expression in transverse sections of nodules from a soybean root transformed with ENOD40p-TRAP cassette.. 48
Figure 3.5 Spatio-temporal localization of GFP expression in transverse sections of nodules from a soybean root transformed with ENOD2p-TRAP cassette. 49
Figure 3.6 Agarose gel image of different fractions obtained from affinity purification. 50
Figure 3.7 Detection of RPL18:GFP fusion protein. 50
Figure 3.8 Evaluation of RNA-Seq library quality using Bioanalyzer. 51
Figure 3.9 Evaluation of rRNA depletion and mRNA isolation approach.. 52
Figure 3.10 Evaluation of RNA-Seq library quality using Bioanalyzer.. 53
Figure 3.11 Marker genes expression pattern in ENOD2p-, ENOD40p- and CsVMVp-TRAP derived RNA at 7 dpi. 55
Figure 3.12 Marker genes expression pattern in ENOD2p-, ENOD40p- and CsVMVp-TRAP derived samples at 10 dpi.. 56
Figure 3.13 Abundance of miRNA 160 and miRNA 166 in TRAP samples.. 57
Figure 3.14 Gene expression pattern in ENOD2p-, ENOD40p- and CsVMVp-TRAP derived samples at 7 dpi.. 60
Figure 3.15 Gene expression pattern in ENOD2p-, ENOD40p- and CsVMVp-TRAP derived samples at 10 dpi. 62

LIST OF TABLES

Table 1 Composition of Hoagland solution 89
Table 2 Composition of nitrogen free plant nutrient solution. 90
Table 3 Composition of vincent - rich media 91
Table 4 qPCR primers used in this study 93
Table 5 Log_2 fold change (ENOD40p-/ENOD2p-TRAp derived samples) value at 7 dpi 94
Table 6 Log_2 fold change (ENOD40p-/ENOD2p-TRAp derived samples) value at 10 dpi 95
Table 7 List of cloning primers used in this study 96
Table 8 List of candidate genes and their expression patterns in INTACT samples 97

ABSTRACT

NODULE ZONE-SPECIFIC GENE EXPRESSION IN SOYBEAN

SADIKSHYA ARYAL

2019

Nitrogen is one of the most limiting nutrients for plant growth and yield. Leguminous plants such as soybean (*Glycine max*) have developed the ability to form a symbiotic association with nitrogen fixing rhizobia. This symbiotic association results in the formation of unique structures called nodules that originate from root cortex via *de novo* cell differentiation. During soybean nodule development, two major nodule zones, the Nodule primordium (Npr) in the center and the nodule parenchyma (Npa) in the periphery, are clearly distinguishable. Npr gives rise to infection zone (IZ), and the Npa holds vascular bundles. However, it is not clear what early signaling pathways drive the conspicuous development of these two nodule zones. To bridge this knowledge gap, we adapted TRAP (Translating ribosome affinity purification) technique for use in soybean hairy root composite plants and evaluated the enrichment of selected mRNAs in translating ribosomes of Npa (using the ENOD2 promoter) and Npr/IZ (using the ENOD40 promoter) in soybean nodules. Confocal images confirmed the expected tissue-specific expression of ENOD2 promoter-driven TRAP gene cassettes in the Npa region and ENOD40 promoter-driven TRAP gene cassettes in the Npr/IZ. Absence of non-coding RNAs in TRAP samples confirmed that TRAP derived RNAs are free of contamination from total or nuclear RNAs. Enrichment of nodule zone-specific translated mRNAs was validated by RT-qPCR assays on three different marker genes: *ENOD2*,

ENOD40, and *CYP83B1* whose nodule zone-enriched expression patterns are previously demonstrated. The expected expression pattern of tissue-specific marker genes at 7 and 10 dpi validated the suitability of our system and methods to evaluate nodule zone-specific translated mRNA profiles. Relative enrichment of selected genes in each nodule zone was evaluated using RT-qPCR. Abundance of mRNAs encoding transcription factors such as MyB-related transcription factor (GLYMA03G42260.1), bZIP transcription factor (GLYMA19G43420.1), and bHLH transcription factor (GLYMA08G04661.1) were significantly higher in nodule infection zone vs. parenchyma at 10 dpi suggesting that these genes might be involved in nitrogen fixation. Similarly, the abundance of mRNA encoding GmARF5 (GLYMA17G37580, potential ortholog of Arabidopsis ARF5) was significantly enriched in nodule parenchyma at 7 and infection zone at 10 dpi suggesting the tissue-specific roles for auxin during nodule development and maturation. The abundance of nitrate transporter (GLYMA11G04500.2) mRNA was significantly enriched in infection zone at 7 dpi and 10 dpi suggesting a possible role for this gene in nitrogen fixation. The abundance of phosphate transporter 1 (GLYMA10G00720) mRNA was significantly higher in nodule parenchyma and might be involved in transporting Pi from roots to nitrogen fixing bacteroids. These results helped identify potential roles of specific genes in processes associated with distinct nodule zones. Global transcriptomic analysis yield broader insight on other key determinants and/or signaling components involved in nodule zone differentiation. Ultimately, this knowledge can be used to devise biotechnological strategies to enhance nitrogen fixation or even potentially transfer N-fixation trait to

non-leguminous plants, and reduce environmental pollution caused by excessive use of chemical nitrogenous fertilizer.

1 INTRODUCTION

1.1 Nitrogen in agriculture

1.1.1 Importance of Nitrogen

Advances in agricultural innovations such as the genetic improvement of crops to better respond to fertilizers have allowed food production to keep in pace with increasing population growth. Nitrogen is one of the most essential nutrients for plants as it is a key component of many major biological compounds such as chlorophyll, nucleotides, and amino acids. Thus, sufficient nitrogen is essential to produce high-quality, protein-rich food (Vance 2001). Nearly fifty percentage of the global fertilizer supply is comprised of nitrogen fertilizer (Bumb and Baanante 1996).

1.1.2 Availability of nitrogen

Plants obtain nitrogen primarily from nitrogen containing minerals in the soil. Nitrogen in soil minerals becomes available as the mineral decomposes. However, mineral decomposition is a slow process and is unable to fulfill the nitrogen requirement of crop plants. Therefore, agricultural nitrogen requirement is mainly met using industrial nitrogen fertilizers. Nitrogen fertilizer production and application are expected to increase in the coming decades to feed an ever-growing world population (Follett and Hatfield 2001). However, production of nitrogen fertilizer alone accounts for 50% of fossil fuel used in agriculture (Canfield, Glazer, and Falkowski 2010). Moreover, carbon dioxide (CO_2) released during fossil fuel combustion, and nitrous oxide (N_2O) released during the decomposition of nitrogen fertilizer add to global greenhouse gas accumulation (Huntley et al. 2007). Less than

half of the nitrogen added is available to the plant, while the rest is lost to the environment (Westhoff 2009). This lost nitrogen results in higher cost for farmers, and threatens air, water, and soil quality, and biodiversity. A European study estimated the cost of nitrogen pollution caused by various sources of nitrogen such as fertilizer runoff from agriculture, fossil fuel burning, industry, and others to be between US $ 79 billion and $364 billion per year. This cost is more than double the value that nitrogen fertilizers added to farm income (Sutton et al. 2011). Excessive fertilizer use causes several environmental and ecological problems within and outside of farmlands, such as air pollution (Seniczak et al. 1998), soil acidification and degradation (McCauley, Jones, and Jacobsen 2009), water eutrophication, crop yield reduction, and undermining the sustainability of food and energy production from agricultural fields (Savci and Development 2012). Therefore, it is necessary to minimize the use of chemical nitrogen fertilizers and optimize other sustainable ways of meeting agricultural nitrogen demands.

1.2 Biological Nitrogen Fixation

Biological nitrogen fixation is an alternative to meet our agricultural nitrogen needs. Biological nitrogen fixation is the process in which atmospheric nitrogen is converted into ammonia (plant useable form of nitrogen) by a group of prokaryotes, termed diazotrophs (de Bruijn 2016). A wide range of diazotrophs has been studied, such as cyanobacteria, *Azotobacteraceae*, rhizobia, and *Frankia*. Several obligately anaerobic bacteria (e.g. *Clostridium*) also fix nitrogen. Similarly, some *Archaea* also contribute to nitrogen fixation. However, the biochemical machinery required for

nitrogen fixation is common in all diazotrophs, which is provided by the nitrogenase enzyme system. The overall reaction for dinitrogen reduction by nitrogenase is:

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

Nitrogenase also catalyzes the reduction of protons to hydrogen and the reduction of diverse alternate substrates such as acetylene, azide, or cyanide (Burgess, 1985). Promoting biological nitrogen fixation in agricultural systems could reduce the dependency on chemical fertilizers, resulting in economic benefits and ecological sustainability.

Plants have developed multiple solutions to associate with nitrogen-fixing bacteria. Based on the degree of intimacy and interdependence between the plant and the microbe, these associations are classified into three main types: free-living, associative, and symbiotic.

1.2.1 Free-living nitrogen fixation

Free-living diazotrophs are bacteria that live in soil and can survive without the direct influence of plant roots (Glick 1995). *Azotobacter*, *Clostridium*, and *Nostoc* are some of the examples of free-living bacteria. These bacteria respond to root exudates via chemotaxis and colonize the rhizosphere but do not penetrate the plant tissues. The energy required for nitrogen fixation is mainly obtained by oxidation of organic molecules released by other organisms or from decomposition. Some of the organisms also have chemolithotrophic capabilities and utilize inorganic compounds as a source of energy. As there are not enough carbon and energy sources for free-living organisms, they have less contribution to global nitrogen

fixation. These bacteria add up to 10-25 kg, of nitrogen/ha/annum. This association is the simplest form of nitrogen-fixing symbiosis (Shridhar 2012).

1.2.2 Associative nitrogen fixation

Associative bacteria, such as *Azospirillium*, live in rhizosphere environment and form a close association with the roots. The interaction of associative bacteria and host plant takes place in the rhizosphere and bacteria are activated by plant root exudates and are attracted by root mucilage (James 2000). Flavonoids are important plant signals for interaction with the bacteria. The first step in the root colonization is the migration of bacteria towards plants roots which is facilitated by the bacterial flagella. The bacteria adsorb to the roots as single cells. After adsorption, bacterial aggregates are formed which are firmly and irreversibly anchored to the root (Vanbleu and Vanderleyden 2007). It is predicted that anchorage of the bacteria depends on bacterial extracellular polysaccharide production. Some Associative bacteria such as *Azospirillum* strains penetrate the roots of their host and become established in the intercellular spaces between the epidermis and the cortex, and even in the vascular system (Patriquin, Döbereiner, and Jain 1983). Associative bacteria convert atmospheric nitrogen into ammonium through the action of the nitrogenase complex which takes place under microaerobic conditions at low nitrogen levels. Ammonium is assimilated mainly through the glutamine synthetase (GS)/ glutamate synthase (GltS) pathway. However, the mechanisms of the association process are still not well understood. This is mainly because of the absence of a clear plant phenotype to indicate the successful interaction as it makes a direct screening of large numbers of mutants not feasible. In addition to fixing

nitrogen, these bacteria also produce plant growth promoting hormones such as cytokinin and gibberellins. Associative nitrogen fixation can contribute to 20-25% of nitrogen requirement in rice and maize (Saikia and Jain 2007).

1.2.3 Symbiotic nitrogen fixation

Symbiotic nitrogen fixation occurs in plants that provide a niche and fixed carbon to nitrogen fixing bacteria in exchange for fixed nitrogen. In this relationship, rhizobia convert the atmospheric nitrogen into ammonia which is available to the plants and in return bacteria gets organic acids from plants. Association between leguminous plants and rhizobia species is a common example of symbiotic nitrogen fixation (Beringer et al. 1979). However, in some plants such as bayber and sweet fern, the symbiont is not rhizobia but *Frankia* (Tjepkema et al. 1986). Mechanism of nutrient exchange between symbiotic partners is described in section **1.5**

1.3 Legume-Rhizobia Symbiosis

Legumes are ranked as the second largest food and feed crops grown worldwide and contribute to more than 25% of world food production (European Association for Grain Legume Research, 2007). Legumes such as soybean have high seed oil content and can be used as a source of biofuel. The legume-rhizobia symbiosis itself provides around 200 million tons of nitrogen annually (Peoples et al. 2009). Therefore, leguminous crops have a special advantage in sustainably meeting agricultural nitrogen needs. Some leguminous plants such as soybean (*Glycine max*) have developed an ability to form a symbiotic association with nitrogen fixing bacteria, collectively termed as rhizobia (Wang et al. 2012). As a result of such symbiotic association, a unique structure known as root nodule is formed in the root

of leguminous crops. Bacteroids within these nodules provide reduced nitrogen to the host plant and obtain energy and carbon from the host plant in return. High amounts of ATP and oxygen reductant are required to meet the demands of the nitrogenase enzyme, but at the same time, nitrogenase is oxygen sensitive (Wang et al. 2012). Leghemoglobin which provides the pink color to effective nodules binds oxygen and transfers it to the bacterial electron transport chain, and ATP synthesis occurs. Thus, there is a reduced concentration of free oxygen in the nodule. Carbon demand and requirement of microaerobic conditions contribute to the high sensitivity of the nitrogen fixation process to environmental conditions (Mengel 1994). Therefore, it is necessary to have the tight regulation of nitrogen fixation rates and development of biotechnological and genetic improvements to biological nitrogen fixation requires a better understanding of nodule development.

1.4 Nodule development

Nodule development is a result of precisely coordinated interactions between legumes and rhizobia. Nodule development can be divided into two principal biological processes: bacterial infection and nodule organogenesis (Oldroyd and Downie 2008), and these processes are tightly regulated in a spatiotemporal fashion. Legumes release flavonoids into the rhizosphere, which induces the production of nod factors (NF) in the rhizobia (Abdel-Lateif et al. 2012; Lerouge et al. 1990). NF is a symbiosis-specific compound which activates nodule organogenesis and induces cellular changes associated with bacterial infection. When the host plant perceives NF, root hair curling and formation of infection thread (IT) occur (Oldroyd et al. 2011). Molecular signaling events occurring during NF perception are described in

sub-section **1.5.1**. IT carries the bacteria to the inner cortex. It causes the inner cortical cells and pericycle cells to divide subsequently forming the primordium. As the IT grows into and reaches the root cortex, bacteria are released into the host cytoplasm and are surrounded by a plant-derived membrane called the peri bacteroid membrane. The structure thus formed is called symbiosome. The bacteria continuously divide inside the symbiosome and differentiate into bacteroids (Roth and Stacey 1989) (Figure 1.1).

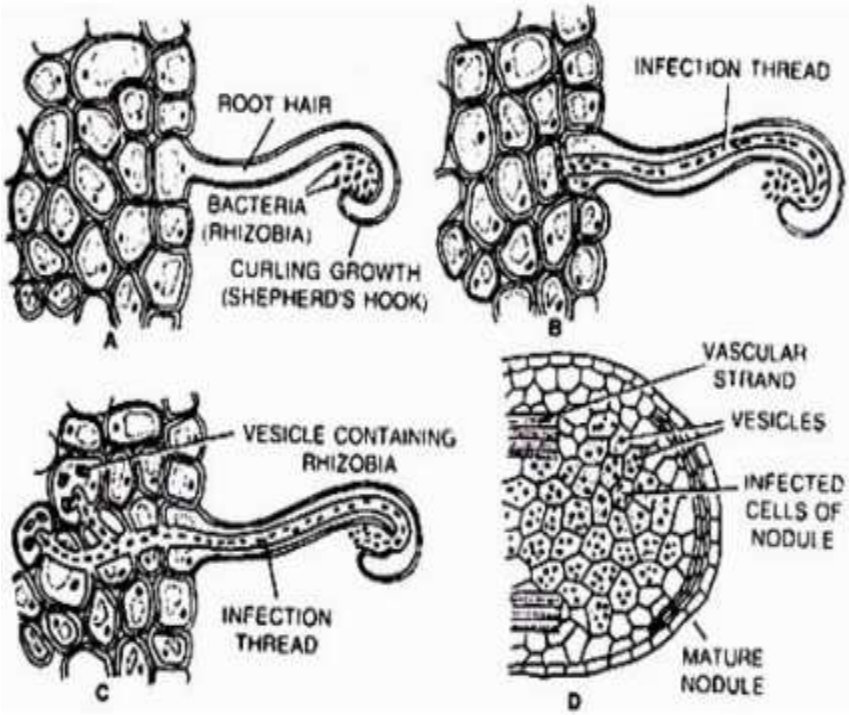


Figure 1.1 Stages of infection during nodule development. Panel A illustrates the attachment of rhizobia in root hair tip resulting in curling of root hair and a shepherd's hook-like structure. Panel B illustrates the formation of infection thread and passage of rhizobia through infection thread. Panel C illustrates the release of rhizobia into plant cortex cells. Panel D illustrates a mature nodule with infected cells at the center (Biology Discussion, 2019).

Based on the site of primordium initiation and their meristem types, nodules are classified into two main categories.

1.4.1 Indeterminate nodules

During initiation of indeterminate nodules, cell division occurs anticlinally in the inner cortex followed by periclinal divisions in the endodermis and pericycle, resulting in nodule primordium formation. Legumes such as *Pisum* (pea), *Medicago* (alfalfa), *Trifolium* (clover), and *Vicia* (vetch) form indeterminate nodules (Bond 1948; Newcomb 1976). As they have persistent meristem, indeterminate nodules have elongated oblong structure with multiple functional zones. The nodule meristem in Zone I consists of continuously dividing cells; Zone II is known as the infection zone; Zone III is the nitrogen fixation zone; Zone IV is the senescence zone; and Zone V is the saprophytic zone (Gage 2004). All these zones in the central tissue are covered at the periphery by the nodule parenchyma tissue, with vascular bundles traversing it (Bond 1948; Newcomb 1976). Due to the continuous cell division, matured nodules contain a heterogeneous population of nitrogen-fixing bacteria. This results in a gradient of developmental stages, giving them an elongated structure. Indeterminate nodules have a less branched vasculature system (Ferguson et al., 2010) (Figure 1.2).

1.4.2 Determinate nodules

Determinate nodules are spherical in shape and lack a persistent meristem and an obvious developmental gradient (Turgeon and Bauer 1982). Cell-division generally occurs sub-epidermally in the outer cortex. However, in *L. japonicus*, there is no sub-epidermal cell division. The cells dividing in the root outer cortex

differentiate into the central tissue whereas dividing pericycle and inner cortex give rise to the parenchyma tissue that surrounds the central zone (Hirsch 1992). The nodule vascular tissues traverse the parenchyma tissue in the periphery of the mature nodule. Mature determinate nodules contain a homogenous population of nitrogen fixing bacteroids because infected cells are differentiated synchronously and senescence. Such nodules only last for a few weeks. When old nodules senesce, new nodules are formed on newly developed roots (Rolfe et al. 1988). In determinate nodules, lenticels are also present which helps in gas exchange. Such nodules are mainly found in tropical and sub-tropical species like soybean (*Glycine max*), bean (*Phaseolus vulgaris*), while some are also found in temperate species (*L. japonicus*)

Even though there is a difference in nodule morphology between determinate and indeterminate nodules, bacterial infection follows a similar molecular mechanism (Ferguson et al., 2010, Newcomb et al. 1979).

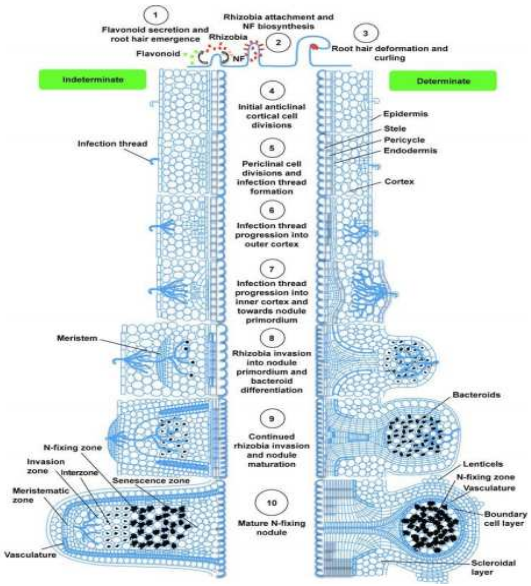


Figure 1.2 Two major types of root nodules. The different stages of development for two major types of nodule in legumes, indeterminate (left) and determinate (right) are illustrated (Ferguson et al., 2010).

1.5 Nutrient exchange between host plant and nitrogen fixing bacteroids

Bacteroids convert atmospheric nitrogen into ammonia. It is predicted that a large concentration gradient of ammonia facilitates the nitrogen efflux from the symbiosome space to the plant cytosol via diffusion. (Udvardi and Day, 1990). However, it is also known that acidification of the symbiosome space generates membrane potential across the symbiosome membrane (Tyerman et al., 1995; Mouritzen and Rosendahl, 1997). The membrane potential activates the membrane and allows the movement of NH_4^+ into the plant cytosol through voltage-gated non-selective cation channel. The ammonium released by bacteroids is converted to glutamine by glutamine synthase, and glutamine is further converted to glutamate by glutamate synthase (Patriarca et al., 2002; Barsch et al., 2006a). Indeterminate nodules, such as pea, clover, and alfalfa, mainly export Asparagine (Asn), while determinate nodules, such as soybean and Phaseolus bean, export ureides (Temple et al., 1998) out of root to shoot. In determinate nodules, infected cells are in contact with at least one uninfected cell (Selker and Newcomb, 1985) resulting in symplastic movement of solutes from infected cells to uninfected cells and hence from uninfected cells to the vascular bundles. There is very little symplastic exchange between infected cells without transit through uninfected cells, suggesting that uninfected cells primarily control the transport among tissues within the nodules. However, in indeterminate nodules such as in *V. faba nodules*, as most of the infected cells are not in contact with uninfected cells, infected cells release amino acids into the apoplast which is accumulated by uninfected cells (Peiter et al., 2004).

Now, uninfected cells can transfer the aminoacids symplastically to the vascular system (Abd-Alla et al., 2000; Peiter et al., 2004).

The carbon required to process the nitrogenase activity in the bacteroid is derived from plant photosynthate which is transported to the nodules via the phloem as Suc (Gordon et al., 1999). Even though several sugar transporters have been known to express at the symbiosome membrane of *Lotus japonicus* and *Medicago truncatula* (Colebatch et al., 2002; Kouchi et al., 2004), transport of sugars has only been demonstrated across the symbiosome membrane of *Phaseolus* beans (*Phaseolus vulgaris*) (Day et al., 2001). In other plants nodules, transport occurs via diffusion. The rate up the uptake of sucrose via diffusion is not sufficient to support nitrogen fixation (Day et al., 2001). This supports the fact that dicarboxylic acids and not sugars are supplied to bacteroids. Nitrogen-fixing bacteroids are present in the center of the nodule. But the phloem is located within the nodules vascular network system in the inner cortex. The phloem is enclosed within an endodermis which acts as an apoplastic barrier (Hartmann et al. 2002). Uninfected cells as in *Vicia faba* have the ability to actively take up sucrose from apoplast. However, symplastic transport could be necessary to transport carbon to infected cells from uninfected cells (Peiter and Schubert 2003). Similarly, in matured *L. japonicus* nodules, the expression of Suc/H^+ cotransporter (LjSUT4) was restricted mainly to the vascular bundles and nodule parenchymatous cells, but not in the central region (Colebatch et al. 2004). This suggested that uninfected cells accumulate sugars and convert into organic acids. These organic acids are then released to the apoplast (Kavroulakis et al. 2000) and might be transferred symplastically to infected cells. Organic acids are

then supplied to bacteroids as the source of carbon required for metabolism and nitrogen fixation.

Plant and bacteria are mutually benefited as bacteroids use the photosynthates (dicarboxylic acids) assimilated into the nodule and fixed nitrogen (glutamine) is exported into the root.

1.6 Molecular mechanism of bacterial infection and nodule organogenesis

1.6.1 Rhizobial infection and nod factor (NF) perception

Symbiosis is initiated when rhizobia recognize and then respond to the presence of host plant roots. Host plant roots secrete the phenolic flavonoid compounds which are perceived by the bacteria and get attracted to the root (Cooper 2004; Redmond et al. 1986). As the bacterial cell perceives the appropriate flavonoid, it activates nodulation genes resulting in secretion of lipochitooligosaccharide NFs (Mergaert, Van Montagu, and Holsters 1997; Spaink 2000). Two receptors kinases (RLK) such as LjNFr1 and LjNFR5 in *Lotus japonicus*, PsSYM2A and PsSYM10 in *Pisum sativum*, and GmNFR1 α/β and GmNFR5 α/β in *Glycine max* are involved in NF binding (Madsen et al. 2003; Limpens et al. 2003; Indrasumunar et al. 2009). These receptors are in the epidermis and contain three important domains: intracellular kinase, transmembrane domain, and an extracellular portion having LysM domains (Steen et al. 2003). LjNFR1, PsSYM2A, and GmNFR1 α/β have serine/threonine kinase domain while others do not have the activation loop suggesting that these two receptors may bind as a heterodimeric receptor and activate kinase functions in downstream signal transduction (Limpens et al. 2003; Radutoiu et al. 2003). Some RLK have leucine-

rich receptors (LRR) and serine/threonine kinase domains (Indrasumunar et al. 2009; Stracke et al. 2002; Capoen et al. 2005). They are found in the plasma membrane and on the IT membrane. Activation of the LysM RLKs could be necessary for the activation of this LRR RLK. LRR RLK is necessary for initial root hair response and predicted to have an important function in NF perception and initial bacterial infection events. Whereas, the LysM RLKs is predicted to have a major role in the NF signaling cascade (Ferguson et al. 2010). Mere treatment with NFs is enough to initiate the sequence of morphological events associated with nodule development.

1.6.2 NF signaling cascade

When plant roots perceive the NF, they initiate the downstream signaling cascade. Potassium ion-channel proteins encoded by MtDMI1, LjCASTOR, and LjPOLLUX (Ané, Kiss et al. 2004, Imaizumi-Anraku, Takeda et al. 2005), two nucleoporins encoded by LjNup133 and LjNup85 (Kanamori et al. 2006) and CALCIUM CALMODULIN KINASE (CCaMK) encoded by MtDMI3/PsSYM9 (Lévy et al. 2004, Mitra et al. 2004) are the major actors involved in signaling cascade. They activate calcium oscillation in the nuclear region. Immediately after the NF application, there occurs the rapid influx of Ca^{2+} ions into the root hair cells and is followed by the efflux of Cl^- and K^+, inducing the calcium spiking. The ion-channel proteins and the nucleoporins are important for these Ca^{2+} spiking events whereas CCaMK is predicted to perceive the Ca^{2+} spiking signals (Oldroyd and Long 2003). Mutation in genes encoding the NF LRR RLK, the putative ion channels or the nucleoporins although blocked Ca^{2+} spiking and nodule development events, the Ca^{2+} fluxes and root hair deformation events were maintained. However,

mutations in CCaMK didn't affect Ca^{2+} fluxes and spiking events but nodule development was blocked (Lévy et al. 2004; Miwa et al. 2006). This experiment suggests that the ion channels and the nucleoporins act downstream of NF perception and upstream of Ca spiking in root hairs while CCaMK acts downstream of Ca spiking.

NODULE SIGNALING PATHWAY 1 and NODULE SIGNALING PATHWAY 2, Ets2 REPRESSOR FACTOR (ERF) required for nodulation (ERN), and NODULE INCEPTION (NIN) are the transcription factors that are activated downstream of CCaMK (Cerri et al. 2012). Although nsp1 and nsp2 mutants show normal Ca^{2+} responses on NFs treatment, expression of the early nodulation (ENOD) genes in the epidermis is not observed. NSP1 and NSP2 are co-localized with CCaMK in the nucleus (Smit et al. 2005; Oldroyd and Long 2003). This suggests that NSP1 and NSP2 are likely activated after Ca^{2+} spiking and could act directly downstream of CCaMK. NSP1, NSP2, ERN1, and NIN are believed to work together and regulate the expression of ENODs in the epidermis. Various protein components, such as DMI3 (MtIPD3) and LjCYCLOPS interact with CCaMK and are important for NF signaling and nodule development (Yao et al. 2005; Messinese et al. 2007). These proteins interact through a C-terminal and possibly regulate NSP1 expression (Smit et al. 2005; Zhu et al. 2008). LjSIP1 is a transcription factor that binds to the promoter of NIN and regulates bacterial infection events (Zhu et al. 2008). Similarly, MtRPG is localized in the nucleus and is also necessary for bacterial infection. LjCERBERUS and ethylene response factor 1 (LjERF1) also

play an important role in bacterial infection event (Yano et al. 2009; Diédhiou et al. 2014).

1.6.3 The early stages of the nodulation

Root cortex consists of cytokinin receptors which play an important role during the cell division process. Loss of function mutants of the cytokinin receptor do not form nodule primordia (Murray et al. 2007). Although bacterial infections occur, infection threads are not capable of growing towards cortex but spread laterally instead, (Murray et al. 2007; Gage and Reviews 2004) suggesting that initial bacterial infection may not require a cytokinin receptor but is necessary to direct the infection threads. However, in recent paper published by (Miri et al. 2019), they showed that expression of epidermal LOTUS HISTIDINE KINASE1 (LHK1) cytokinin receptor is dispensable for the establishment of the *L. japonicus–M. loti* symbiosis, challenging the earlier model which stated that that the root epidermis-localized LHK1is necessary to regulate *M. loti* infection (Held et al. 2014). As the infection threads penetrate the root cells, the root cortical cells divide and form the nodule primordium. Both the epidermal and the cortical cell events require CCaMK. However, CCaMk follows different signaling pathways. NIN, NSP1, and NSP2 are activated downstream of CCaMK and the cytokinin receptor (Tirichine et al. 2007; Madsen et al. 2010). NIN has a positive role during cortical cell division as cytokinin/NF application induces the expression of NIN. As the cytokinin receptor is a significant actor during nodule formation, it explains the role of cytokinin as well. Cytokinin is a plant hormone, predicted to act as a mobile signal to relay the NF perception from the epidermis to the cortex. Similarly, abscisic acid (ABA) is also a

plant hormone and could be that mobile signal (Ding et al. 2009) (Figure 1.4). However, the direct role and functions of these signals during nodule development are yet to be fully understood.

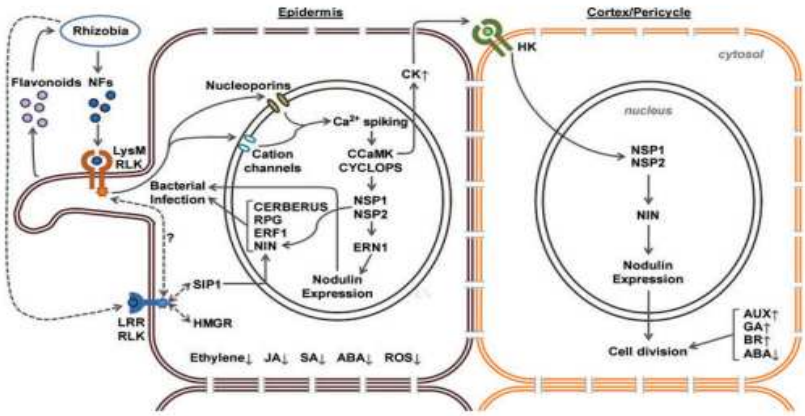


Figure 1.3 Early molecular signals in root nodule development. Legume roots secrete flavonoids into the rhizosphere which are perceived by the compatible rhizobia. This perception of flavonoids stimulates the production nod factors (NF) in rhizobia. Leucine-rich repeat receptor-like kinase (LRR RLK) perceived the NF at the epidermis together with LysM RLKs and activate downstream signaling events which result in cell division in cortex and pericycle. Cortical cell division results nodule primordia formation. Cytokinin is believed to act as a mobile signal communicating from the epidermis to the cortex. Apart from cytokinin, many other hormones play a vital role in regulating nodule initiation and development and their levels are also precisely regulated throughout nodule organogenesis. (Ferguson, B.J., et al,2010).

1.7 Determinate nodule- zones differentiation

Fifty percent of the global area is cultivated by soybean, contributing to 8% of the total global legume production (Vance 2001; Wagner 2011) . Moreover, up to 337 kg nitrogen per hectare can be supplied to soybean through atmospheric nitrogen which is 98% of the total nitrogen uptake in soybean (Salvagiotti et al. 2008). This shows the importance of biological nitrogen fixation in soybean production. Therefore, we are trying to gain more insight into soybean root nodule which is a determinate nodule.

Nodule can be broadly subdivided into central and peripheral tissues. Central tissue is composed of cytoplasmically-rich cells. The peripheral tissues consist of nodule parenchyma with vascular bundles embedded in it, and nodule cortex. Nodule cortex is separated from parenchyma by the nodule endodermis. The early nodulin gene GmENOD40 is expressed in the central tissues, especially in uninfected cells. In fact, GmENOD40 has a complex expression pattern during soybean nodule development. One day after inoculation, the first cell divisions are induced in the sub-epidermal cell layer. At this stage of development, GmENOD40 is induced in dividing root cortical cells. Three days after inoculation, a small nodule primordium is formed and at this stage, GmENOD40 mRNA is expressed in all cells of the primordium (Yang et al. 1993). Seven days after inoculation, nodule primordium is differentiated into central region consisting of infected cells. Whereas, dividing inner cortical cells are differentiated into a vascular bundle that connects central tissue and root stele. At this stage of development, GmENOD40 is expressed in central tissues (especially in uninfected cells) and in cell layer connecting vascular bundles. However, at this stage, GmENOD40 is no longer expressed in root pericycle. In matured nodules (14 days onwards), GmENOD40 is expressed specifically in uninfected cells and in pericycle of nodule vascular bundles (Newcomb et al. 1979).

GmENOD2 and GmENOD13 are expressed in nodule parenchyma cells. GmENOD13 is 50% homologous with GmENOD2. Both ENOD2 and GmENOD13 are predicted to be cell wall components (Franssen et al. 1992). In soybean nodules, parenchyma cells are known to block the oxygen into the central part of the nodule.

The oxygen level dropped abruptly across the nodule parenchyma zone and into the central zone that consisted of several layers of cells separated from each other by a very small intercellular space (Tjepkema and Yocum 1974). Therefore, early nodulins such as GmENOD2 and GmENOD13 may be involve in limiting oxygen diffusion into the central tissue. Additionally, ENOD2 is a cell wall-based hydroxyproline-rich glycoprotein (HRGP), thus may contribute to the formation of an oxygen barrier layer (Hirsch 1992). Expression of the ENOD2 gene is spatiotemporally regulated. The earliest stage ENOD2 gene can be detected is 6 days after inoculation (van de Wiel et al. 1990). At 6 days after inoculation, cell divisions are induced in the inner cell layers of the root cortex and the central part of these dividing cells is developing into vascular tissue that connects the root nodule with the central cylinder of the root. At this stage of development, GmENOD2 is expressed in the newly formed tissue surrounding the procambial strand between the primordium and the root central cylinder and in inner cortical cells at the proximal and lateral sides of the nodule primordium. In a 10 days old nodule, globular meristem is developed into a central and a cortical tissue, and at this stage GmENOD2 gene is expressed in the nodule inner cortex as well as in the tissue surrounding the vascular strand that connects the nodule with the central cylinder. In matured nodules from 21 days after inoculation, ENOD2 transcript are present in all parts of the nodule inner cortex (van de Wiel et al. 1990). The innermost cells of the nodule parenchyma may also be involved in ureide production (Newcomb et al. 1989). Immuno-localization study in these cells reveals the presence of nodule-specific uricase. This suggests that nodule parenchyma in the determinate nodules

are involved in different functions and require more experiment to fully understand them.

Although we know some nodule parenchyma and central tissue specific genes and their expression patterns during soybean nodule development, we lack the information on signaling mechanism that directs the differentiation of nodule parenchyma and central tissues. Nodule parenchyma and central tissue specific gene expression study will allow us to understand the mechanism that directs the nodule zone differentiation.

1.8 Cell type-specific gene expression

1.8.1 Regulation of gene expression

Living organisms regulate the expression of certain genes to increases their versatility and adaptability. To understand the complexity and development of a certain organ, it is important to understand how cells obtain specific properties. Eukaryotic gene expression is regulated by multiple mechanisms. Cells use a different mechanism to alter the expression of a particular gene, a process known as gene regulation (McAdams and Arkin 1997). A single gene can be regulated in different ways. Changing the number of RNA copies that are transcribed is one of the ways to regulate gene expression. Similarly, gene expression is also regulated through temporal control when the gene is transcribed (Cvekl et al. 2007). Plant organs also gain their specific identity from the respective transcriptome (Libault et al. 2010). Various experiments have been conducted on transcriptome profiling during nodulation in various leguminous plants. During nodulation, a major change in the expression of the host legume genes is observed and this is due to bacterial

infection and nodule organogenesis. Especially, nodulin genes are highly expressed during nodulation process and have been isolated and studied from many legume species.

Similarly, transcriptomic analysis has been conducted and successfully identified differentially expressed genes during soybean nodulation (Brechemacher et al. 2008; Libault et al. 2010; Damodaran et al. 2017; Adhikari 2016). However, a comprehensive study of cell-type specific gene expression profile during the soybean nodule development is necessary to have better understanding about the mechanisms involved in nodule formation and nitrogen fixation.

1.8.2 Cell type-specific gene expression

Several methods have been developed to study transcriptome profiling at the cellular level such as FACS (isolation of cells of interest after tissue dissociation), through manual sorting of fluorescent cells (Sugino et al. 2006; Hempel et al. 2006), and laser-capture microdissection (Luo et al. 1999; Yao et al. 2005).

In an experiment conducted by (Limpens et al. 2013), they performed the cell-type specific study at different stages of nodule development in *Medicago truncatula* using laser-capture microdissection and successfully identified genes enriched in different cells or tissues. The nodule-specific signal peptidase subunit MtDNF1/DAS12, the putative metallo-peptidase MtMMPL1 and the AP2/ERF transcription factor MtEFD were enriched in the infection zone of the nodule, suggesting their role in controlling infection and symbiosome development. Nodule specific cysteine-rich peptides (NCRs) were enriched in infected cells of the infection zone. Genes such as ENOD11, ERN1, ERN2 were highly expressed in the

distal infection zone (where symbiosome are formed and divided) compared to the proximal infection zone (where symbiosomes are differentiated). Auxin signaling related genes such as AUX/IAA's, ARF's, PIN auxin efflux carrier genes, Auxin responsive genes such as GH3-like show a meristem specific expression in the nodule. This paper demonstrated that laser-capture microdissection was successfully used to isolate the specific cells/tissues at different stages of symbiosome formation from nodules of the model legume *Medicago truncatula*. Although these methods such as FACS and laser-capture microdissection are widely used, they require tissue fixation and tissue dissociation and have relatively low throughput (Heiman et al. 2014). INTACT (Isolation of Nuclei Tagged in specific Cell-types) and TRAP (Translating Ribosomes Affinity Purification) overcome such limitations.

In INTACT and TRAP methods, fixation or dissociation of tissue is not necessary to capture cell type-specific mRNA (Heiman et al. 2014). In these methods, the cell type of interest is labeled with EGFP, thus, facilitating visualization in living cells. However, they also possess certain limitations. For these methods, it is necessary to generate transgenic lines for each cell type of interest. Also, an appropriate promoter is necessary to drive the INTACT/TRAP transgene in the cell type of interest. If the problem occurs, the other methods such as laser capture microdissection are preferable. However, for cell type-specific study in soybean, we have the well-adapted technique to create transgenic lines, and genetic material specific to our cell type of interest is also well characterized. Therefore, INTACT (Deal and Henikoff 2011) and TRAP (Heiman et al. 2014) are

the best alternatives to study the cell type-specific gene expression in soybean root nodules.

1.8.3 INTACT

INTACT gene cassette consists of a binary vector containing a biotin ligase cassette and a nuclear targeting fusion protein (NTF). In biotin ligase cassette, a constitutive promoter drives the expression of a biotin ligase gene (BirA). The NTF protein consists of WPP domain of AtRanGAP1 (nuclear envelope tagging sequence); the Green fluorescent protein (GFP) is used to visualize the target cells; and the Biotin ligase recognition peptide is the substrate for *E. coli* biotin ligase encoded by BirA gene. Cell-type specific promoter drives the expression of this construct in target cells whose nuclear envelope are tagged with biotin. The biotin-tagged nuclei are isolated by the affinity purification using streptavidin magnetic beads (Deal and Henikoff 2011). Subsequently, the isolated nuclei are used to determine the gene expression profiles in the cell-type of interest.

1.8.4 TRAP

Cell type-dependent patterns of transcriptional activity determine the maintenance of cellular identities and this is complemented by post-transcriptional control. For example, miRNA regulates mRNA stability and translation (Lee et al. 2006). miRNAs are 21 to 22 nucleotides long sequences that regulate the transcripts either by cleaving mRNA or through the inhibition of mRNA translation (Axtell 2013). For instance, during soybean nodule development, miR160 promotes auxin activity by suppressing the levels of the ARF10/16/17 family of repressor ARF transcription factors. Similarly, miR169 is expressed in the infection zone and

confines the nodule-meristem specific expression of MtHAP2-1 (Combier et al. 2006). Differences in mRNA and protein levels in plant cells are due to the differential mRNA translation and protein degradation. Actively translated mRNAs are associated with multiple ribosomes in large polyribosome complexes, while non-translated mRNAs are found in association with only one or a few ribosomes or also could remain as ribonucleoprotein (RNP) complexes (Preiss et al. 2003). Therefore, the association of mRNA with polyribosomes gives the idea about the translational status of an mRNA. For this, we use TRAP, an approach that combines cell type-specific transgene expression with affinity purification of ribosomes. Since this approach provides the profile of translated mRNA content of a cell, a more precise estimate of the protein content could be ascertained.

In TRAP, mRNAs are indirectly tagged with epitopes such as FLAG/HIS and cell type-specific genetic element is used for transgene expression. The use of epitope aids in rapid immunopurification of ribosomal complexes. If ribosomes are maintained on mRNA, purification of the cell type-specific tagged ribosomes gives cell type-specific translated mRNAs (Serafini and Ngai 2000; Zanetti et al. 2005) which can subsequently be analyzed by qPCR and sequencing.

Several experiments have been conducted which show the successful application of TRAP approach to study the gene expression profiles. The experiment done by Zanetti et. al. 2005 showed the successful application of TRAP method for global analysis of gene expression in Arabidopsis. In this experiment, the Cauliflower Mosaic Virus 35S promoter was used to drive the expression of TRAP gene cassette. Immunopurification was performed using anti-FLAG agarose bead,

and 60-S ribosomal subunits, intact 80-S monosomes and polysome were obtained. They also carried out the sucrose density gradient fractionation and found similar distribution patterns for crude cell extracts and the purified complexes. When immunopurified and total cellular RNA samples were compared, it was found that mRNAs of many genes were found in association with the epitope-tagged polysomal complexes. These results suggested that TRAP approach is a useful tool to study actively translating mRNAs in plant cells and can also be used to study the tissue-specific mRNA populations using tissue-specific promoters. Similarly, to understand the translational regulation of mRNAs during root nodule formation, TRAP approach was applied in *Medicago truncatula* during the symbiotic interaction with *Sinorhizobium meliloti* (Reynoso et al. 2013). Using the TRAP approach, they obtained 400–500 ng of RNA from 1ml of pulverized tissue, which was comparable with the yield reported for immunopurification of polysomal RNA from Arabidopsis root tissue (Mustroph et al. 2009). To confirm the presence of mRNAs in the immunopurified sample, semiquantitative RT-PCR analyses of ACTIN11 were conducted. Total RNA samples from empty vector and 35S: FLAG-RPL18 roots, and immunopurified RNA sample from the 35S: FLAG-RPL18 roots showed the single band of the expected size. Whereas, there was no amplification in the immunopurified samples from empty vector roots. These results demonstrated the efficiency and purity of the TRAP method, further confirming that TRAP method which was initially developed in Arabidopsis can be applied in other plants as well.

As the TRAP method is well adapted and overcome the limitations in existing methods, we used the TRAP method to study the gene expression profile of nodule parenchyma and central tissue (also known as infection zone) in soybean. Parenchyma cells are known to block the oxygen into the central part of the nodule, and holds the vascular bundles required for nutrient exchange (Tjepkema and Yocum 1974). Similarly, central tissue contains the nitrogen-fixing bacteria (Yang et al. 1993). Understanding the mechanism with which these two nodule zones are formed in soybean will allow us to optimize the nodule number and increase the nitrogen fixation efficiency. As discussed in the section **1.7**, GmENOD2 is known to be expressed in nodule parenchyma (van de Wiel et al. 1990) and GmENOD40 is expressed in infection zone (Yang et al. 1993). Therefore, to study the gene expression profiles of these two nodule zones, we used ENOD2 and ENOD40 promoter to drive the TRAP gene cassettes in nodule parenchyma and infection respectively. Moreover, transcriptomic data from parenchyma and infection zones using INTACT approach had been already generated in our lab. Thus, combining results from INTACT and TRAP will give a better insight into the complexity of the regulatory pathway of two nodule zones.

2 MATERIAL AND METHODS

2.1 Gene cassettes preparation

2.1.1 Verification of destination vector

A binary vector containing a Gateway attR1-attR2 destination sites to clone a promoter of choice in front of the TRAP gene cassette (pK7WG-TRAP) was obtained from Dr. Shioban Brady (University of California, Davis, CA) (Ron et al. 2014). The TRAP cassette consists of a fusion protein with a 6xHis/Flag (HF) epitope followed by Green Fluorescent protein (GFP) and Ribosomal protein L18 (RPL18). RPL18 encodes a protein L18 which is incorporated into the large subunit (60s) of ribosomes. The HF epitope on RPL18 allows for rapid affinity- or immuno-precipitation of ribosomes, and this epitope can be detected by using anti-flag-antibodies. Green Fluorescent protein (GFP) can be used to visualize the expression of the fusion protein. The vector was received as bacterial stab in *E. coli* ccdB Survival 2T1 host cells. *E. coli* cells containing the vectors were streaked in LB + agar plate with spectinomycin (100 µg/mL) and incubated at 37 ^{0}C for 16 hours. A single colony was picked and cultured in 3 mL of LB media with spectinomycin (100 µg/mL) at 37 ^{0}C for 16 hours shaking (220 rpm). A glycerol stock was prepared by mixing 1 mL of cultured cells and 1 mL of 50% glycerol and stored at – 80 ^{0}C. The plasmid of pK7WG-TRAP vector was isolated from the rest of the culture (2 mL) using PureYield plasmid miniprep kit (Catalogue no: A1222). The diagnostic restriction digest was then carried out to verify the plasmid using three different restriction enzymes (NotI - Catalogue no: R0189S, EcoRI - Catalogue no: R3101S, and NdeI - Catalogue no: R0111S) which cleaved the plasmid at specific

sites. Restriction digestions were performed as described in product's manual. The resulting fragments were analyzed by gel electrophoresis and expected restriction digest patterns were verified (Appendix A).

2.1.2 Verification of expression vector for Enod2 and Enod40

TRAP constructs containing GmENOD2 (pK7WG-ENOD2p-TRAP) and GmENOD40 promoters (pK7WG-ENOD40p-TRAP) were generated by Gateway LR clonase-mediated recombination reactions of the destination vector (pK7WG-TRAP) with pMH40-GmENOD2p and pMH40-GmENOD40p entry vectors respectively. The entry vectors were already available in the lab and were maintained in *E. coli* as described by Pathak 2016. *E. coli* cells containing pK7WG-Enod2-TRAP and pK7WG-Enod40-TRAP vectors were separately cultured in 3 mL of LB media with spectinomycin (100 µg/mL) at 37 ^{0}C for 16 hours with shaking at 220 rpm. The plasmid was isolated from the culture using PureYield plasmid miniprep kit (Catalogue no: A1222). Isolated plasmids from two different constructs [(pK7WG-ENOD40p-TRAP) and (pK7WG-ENOD2p -TRAP)] were verified through restriction digestion using three different restriction enzymes (NotI - Catalogue no: R0189S, EcoRV - Catalogue no: R3195S and NdeI - Catalogue no: R0111S for pK7WG-ENOD2p-TRAP, and EcoRV - Catalogue no: R3195S, EcoRI - Catalogue no: R3101S and NdeI - Catalogue no: R0111S for pK7WG-ENOD40p - TRAP) which cleaved the plasmid at specific sites. Restriction digestions were performed as described in product's manual. The resulting fragments were analyzed by gel electrophoresis and expected restriction digest patterns were verified (Appendix B and C).

2.1.3 Construction of PMH40-CsVMV entry vector

To amplify CsVMV CVP2 promoter, 5 µL of 5x Q5 reaction buffer, 5 µL of 10 mm dNTP, 1.25 µL of 10 µm forward (containing NcoI recognition site) and reverse primer (containing EcoRV recognition site) (Appendix M), 20 ng of DNA of pCAMGFP-GWOX vector (Fisher et al. 2018) and 0.25 µL of Q5 polymerase were mixed in reaction. RNase free water was added to make the final reaction volume to 25 µL. The reaction mixture was incubated in a thermocycler at 98 ^{0}C for 1 minute; 15 cycles at 98 ^{0}C for 15 seconds, 58 ^{0}C for 15 seconds, and 72 ^{0}C for 45 seconds; and 72 ^{0}C for 3 minutes. The PCR product of 500bp size was obtained and was confirmed using gel electrophoresis. PCR cleanup was done by using Wizard gel and PCR clean-up system (Catalogue no: A9282). To prepare for cloning, the clean PCR product and pMH40 entry vector were digested using NCoI and EcoRV enzymes. The digestion reaction of pMH40 entry vector was analyzed by gel electrophoresis. Two distinct bands were formed having 3.8 kb and 515 bp size. Gel band of 3.8 kb was cut and eluted using gel and PCR clean-up system (Catalogue no: A9282). Ligation was done using 75 ng of the insert (digested CsVMV CVP2 PCR product) and 185 ng of vector (digested pMH40 entry vector), 2 µL of 10x buffer and 1 µL of T4 ligase enzyme (Catalogue no: M0202T). RNase free water was added to make the final reaction volume 20 µL. The mixture was incubated at room temperature for 2 hours. Bacterial transformation was carried using 5 µL of the ligation reaction product into 50 µL of Top10 competent cells. The mixture was incubated for 30 minutes on ice followed by heat shock at 42 ^{0}C for 30 seconds. Two hundred and fifty µL of S.O.C medium was added to the reaction and incubated at 37 ^{0}C for 1

hour in a shaker. One hundred µL of the reaction was plated in LB + agar plate with ampicillin (100 µg/mL) and incubated at 37 ^{0}C for 14 hours. A single colony was picked and cultured in 3 mL of LB media with ampicillin (100 µg/mL) at 37 ^{0}C for 14 hours shaking (220 rpm). A glycerol stock was prepared by mixing 1 mL of cultured cells and 1 mL of 50% glycerol and stored at -80 ^{0}C. The plasmid of pMH40-CsVMV entry vector was isolated from the rest of the culture (2 mL) using PureYield plasmid miniprep kit (Catalogue no: A1222). The isolated plasmid was verified through restriction digestion using three different restriction enzymes (NotI-Catalogue no: R0189S, BSRGI - Catalogue no: R0575S and PvuI - Catalogue no: R0150S) which cleaved the plasmid at specific sites. Restriction digestions were performed as described in product's manual. The resulting fragments were analyzed by gel electrophoresis and expected restriction digest patterns were verified (Appendix D). The isolated plasmid was also verified by sequencing using T7 primer (Appendix M).

2.1.4 Construction of expression vector for CsVMV

A Gateway LR clonase reaction was performed between the entry vector carrying CsVMV CVP2 promoter (pMH40-ENTR-CsVMV) and destination clone (PK7WG-TRAP). The quantity of entry vector used was 190 ng (1 µL) and that of the destination vector was 380 ng (3 µL). One µL of LR clonase II enzyme (catalogue no: 11791020) was added, and TE buffer was added to make the final reaction volume 6 µL and the reaction was incubated for 1 hour at room temperature. One µL of proteinase K solution was added to terminate the LR clonase reaction and the reaction was incubated at 37 ^{0}C for 10 minutes. For bacterial transformation, 4

µL of LR reaction product was transferred to 50 µL of Top10 competent cells. Bacterial transformation was done by heat shock at 42 ^{0}C for 30 seconds. Two hundred and fifty µL of S.O.C medium was added to the reaction and incubated for 37 ^{0}C for 1 hour in a shaker. One hundred and fifty µL of the reaction was plated in LB + agar plate with spectinomycin (100 µg/mL) and incubated at 37 ^{0}C for 16 hours. A single colony was picked and cultured in 3 mL LB + Spec media in the tube for 16 hours. Glycerol stock was prepared by mixing 1 mL of cultured cells and 1 mL of 50% glycerol and stored at -80 ^{0}C. The plasmid of pK7WG-CsVMV-TRAP vector was isolated from the rest of the culture (2 mL) using PureYield plasmid miniprep kit (Catalogue no: A1222). The isolated plasmid was verified through restriction digestion using three different restriction enzymes (EcoRV - Catalogue no: R3195S, SphI- Catalogue no: R3182 and PvuI - Catalogue no: R0150S) which cleaved the plasmid at specific sites. Restriction digestions were performed as described in product's manual. The resulting fragments were analyzed by gel electrophoresis and expected restriction digest patterns were verified (Appendix E).

2.2 Growth and maintenance of soybean plants

Soybean plants (*Glycine max* cvWilliam82) were used for the experiment. The seeds were sterilized by washing with 10% Clorox for two minutes followed by 70% ethanol for two minutes. Sterilized seeds were washed with distilled water to remove the residues of Clorox and ethanol. Seeds were sown in 4" plastic pot (Catalogue no: 14335600) filled with the autoclaved potting mixture (vermiculite and perlite in the ratio of 1:3). Seeds were regularly watered with Hoagland solution (Appendix F). The seedlings were grown in a growth chamber using the following

growth conditions 16 hours of daylight and 8 hours of night, 50% humidity with 25^0C during the daytime and 20^0C during the nighttime.

2.3 Hairy Root Transformation

2.3.1 Preparation of competent cells of *Agrobacterium rhizogenes* K599 strain

Agrobacterium rhizogenes K599 strain were already available in the lab and were stored in glycerol stock as described by Pathak 2016. The *Agrobacterium rhizogenes* K599 strain was streaked in a LB plate and incubated for at 30 ^{0}C for 36 hours. A single colony was picked and grown in 5 mL LB at 30 ^{0}C on a shaker for 16 hours. Two hundred mL of LB was inoculated with 2 mL of above-grown culture and kept at 30 ^{0}C on a shaker until the optical density (O.D) was 0.5. The cells were split into 4 conical tubes (50 mL each) and centrifuged at 5000 rpm for 10 minutes at 4 ^{0}C. The cells produced a pink pellet. The pellet in each tube was resuspended in 20 mL ice-cold 10% glycerol. The cells were then again centrifuged, and the pellets were resuspended in 10 mL ice-cold 10% glycerol. The cells were then centrifuged again, and pellets were resuspended with 2ml of ice-cold 10% glycerol. Fifty µL of the cells were aliquoted into cold 1.75 mL Eppendorf tube and stored at -80 ^{0}C.

2.3.2 Electroporation-mediated transfer of construct in *Agrobacterium rhizogenes* K599 strain

Competent cells were thawed on ice for 10 minutes. One µL of plasmid DNA was added to 50 µL of competent cells and mixed well. The mixture was kept on ice for 30 minutes. Electroporation cuvettes were also kept on ice. Twenty-five µL of the competent cell-DNA mixture was transferred into electroporation cuvette. The mixture was placed between electrodes and was electroporated at 25 µF capacitance,

400 ohms resistance, voltage of 1.8 KV in a Biorad gene pulser Xcell electroporation system. One mL of LB was immediately added to the cuvette after the electroporation and gently mixed. The cells were then transferred to a 2 mL eppendorf tube and incubated at 30 ^{0}C for 2 hours at shaking. Cells were centrifuged at 8000 xg for 30 seconds and the pellet was resuspended in 200 µL of LB. The cells were then plated on LB + spectinomycin (100 µg/mL) and incubated at 30 ^{0}C for 36 - 48 hours. Four different individual colonies were picked and grown in LB+ spectinomycin (100 µg/ml) medium at 30 ^{0}C for 16 hours in a shaker with 200 rpm. Glycerol stocks were maintained for each culture and stored at - 80 ^{0}C.

2.3.3 *Agrobacterium rhizogenes* mediated hairy root plant transformation

Agrobacterium rhizogenes with the plasmid of interest was grown in LB + spectinomycin (100 µg/mL) media at 30 ^{0}C for 16 hours in a shaker with 200 rpm. The culture was centrifuged at 3300 xg for 9 min at 4 ^{0}C. The pink color pellet was resuspended in nitrogen free plant nutrient solution (N^{-} PNS) (Appendix G) to a final concentration of O.D600 0.3. Rockwool plug (Hummert International, MO) was cut into ~1”x1”x1” cubes and autoclaved before use. Prepared rockwool plug was kept in Petri dishes held on a sterile tray. A small hole was made in the center of rockwool plug using a micropipette tip. The prepared Agrobacterium culture was poured on the hole of rockwool plug using a serological pipette (VWR, catalog no:89130-900) until it was completely soaked. Soybean plants that were grown on growth chamber for 2 weeks and have fully opened first trifoliate leave were used for plant transformation. A slanted cut was made around 2 cm below the trifoliate leaf and the shoot segment was inserted into the hole of rockwool plug (Figure 2.1).

This ensured exposure of plant at the culture (Collier et al. 2005). The tray was then covered with a sterile transparent lid and grown in 16-hour day and 8-hour night light condition until root emerged from the plant (around 21 days). These plants formed both adventitious roots and transgenic roots.

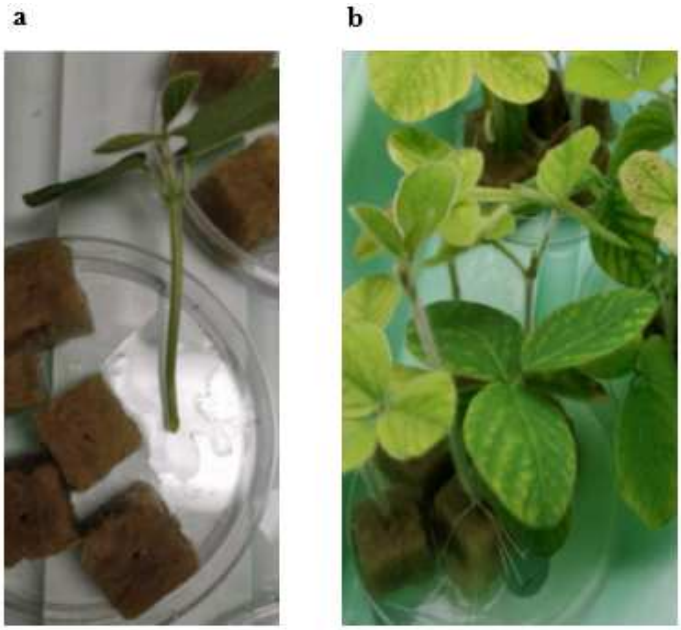


Figure 2.1 Hairy rot plant transformation in soybean. Figure shows a) Rockwool plug with a hole in the center and soaked with Agrobacterium culture, and a 2 weeks old soybean plant with slanted cut at the bottom b) 3 weeks old transformed soybean plants.

2.4 Plant screening and transplanting

Plants that were transformed with pK7WG-ENOD2-TRAP and pK7WG-ENOD40-TRAP constructs were directly transplanted whereas plants that were transformed with pK7WG-CsVMV-TRAP were first screened using fluorescence stereomicroscope (Leica, Model no: MSV269), and only the GFP positive plants were transplanted. Non-GFP plants were discarded. Transplanted plants were grown in 4" pot (Catalogue no: 14335600) containing autoclaved potting mixture that was prepared by mixing vermiculite and perlite in the ratio of 1:3 and was kept in a growth chamber with 16-hour day and 8-hour night light condition for 5 to 7 days. Plants were regularly watered with nitrogen free plant nutrient solution.

2.5 Inoculation of rhizobia

Bradyrhizobium japonicum (USDA 110) was grown in Vincent's rich medium (Appendix H) with antibiotic chloramphenicol (20 µg/mL) at 30^0C on the shaker at 200 rpm for 4-5 days (O. D 600 < 0.5). The culture was then centrifuged at 3300 xg for 9 min at 4 ^{0}C and resuspended in nitrogen free plant nutrient solution (N^- PNS) to a final concentration of O. D600 0.08. Each plant was inoculated with 25 mL of this culture. Plants were harvested after 7- and 10-days post inoculation (dpi) with *Bradyrhizobium japonicum* for microscopy and ribosome isolation.

2.6 Microscopy

2.6.1 Fluorescence microscopy

GFP positive nodules were identified using fluorescence stereomicroscope (Leica, Model no: MSV269) at 2 different time points of nodule development, 7 and 10 dpi. After screening, only GFP positive nodules were harvested. Nodules were

collected in labeled 50 mL falcon tubes kept in liquid nitrogen, and later stored at -80 ^{0}C for affinity purification of ribosomal complexes.

2.6.2 Vibratome sectioning

Fresh GFP positive nodules at two different time points (7 and 10 dpi) were harvested for microscopy. Samples were embedded in the 4% agarose and sectioned using Vibrating-blade microtome (Catalogue No: VT1000S). Sections of 50 µm thickness were obtained.

2.6.3 Confocal microscopy

Sectioned nodules were imaged using a laser scanning confocal microscope (Olympus fluoview 1200) using FITC filter (495 nm laser excitation wavelength, 530 nm emission wavelength and 15% transmissivity). Images were taken using a 10X magnification objective lens. Merged images of white light and GFP fluorescence were obtained using the Fluoview software.

2.7 Translating ribosomes affinity purification (TRAP)

Polysomes were purified from the transgenic nodules as described previously for Arabidopsis root (Zanetti et al. 2005) and Cold Spring Harbor laboratory protocol for plant cytoplasmic ribosomes and polysomes isolation (Rivera, Maguire, and Lake 2015) with some modifications. In this method, the tissue stored in -80 ^{0}C was ground in liquid nitrogen (about 2 g) using a mortar and pestle. Ten mL of freshly prepared polysome extraction buffer (PEB: 50 mM Tris pH. 9.0, 30 mM MgCl2. 400 mM KCl, 17% (w/w) sucrose, 1mM PMSF, 50 µg/mL cycloheximide, 50 µg/mL chloramphenicol) was added and the mixture was thawed on ice. Tissue was transferred to a glass homogenizer. The mixture was homogenized with five

strokes of pressing the plunger down and centrifuged at 3000 xg for 7 minutes at 4 ^{0}C. The supernatant was passed through sterile and cold miracloth (Catalogue no: 2668144). Twenty% triton X-100 (0.1 parts) was added to the supernatant and centrifuged at 12,000 xg for 30 minutes. The supernatant ("clarified extract") was collected and applied to a Ni sepharose slurry as follows. One mL of nickel sepharose (Ni sepharose high-performance resin, Catalogue no: 17-5268-010) was transferred to a 50 mL tube and washed with 5 mL of 1x PBS (Catalogue No:1157C274) by incubating for 1 minute at 4 ^{0}C with gentle back-and-forth shaking on a nutator, then centrifuged at 3000 xg for 3 minutes at 4 ^{0}C. The supernatant was discarded. Eight mL of clarified extract was added to the washed Ni sepharose slurry and incubated for 2 hours at 4 ^{0}C with gentle back-and-forth shaking on a nutator. The sample was centrifuged at 3000 xg for 4 min at 4 ^{0}C. The supernatant (unbound fraction) was collected. The pellet containing the protein of interest was resuspended in 1 mL 1x PBS. The polysome suspension was transferred into a cellulose acetate filter column (Catalogue no: 60702) and centrifuged at 800 rpm for 1 minute. The washing steps were repeated three times and all washed fractions were collected. The column was transferred to a new 2 mL collection tube. Hundred µL of elution buffer (500 mM imidazole, 500 mM NaCl, 20 mM sodium phosphate buffer) was added to the column and centrifuged at 800 rpm for 1 minute. Three successive elutions were performed and collected in different collection tubes. RNase inhibitor (Catalogue no: M0253S) was added to the elutions at 0.5µL/ 100 µL and stored at -80 ^{0}C for further use.

2.8 Western blotting

Sample was denaturized by boiling each cell lysate in 2x Laemmli sample buffer (Catalogue no: 1610737) at 100 °C for 2 minutes. Equal amount of protein (30 µg) for each sample was loaded into the wells of the mini-protean TGX stain-free precast gels (Catalogue No: 4568024) along with molecular weight marker. Electrophoresis was performed in 1x SDS-PAGE running buffer (25 mM Tris,192 mM glycine and 0.1% SDS) for 1 hour at 100 V. Polyvinylidene difluoride (PVDF) membrane (Catalogue no: IPFL00010) was activated with methanol for 1 min and rinsed with transfer buffer (25mM Tris,192 mM glycine, 0.1% SDS and 20% methanol) before preparing the membrane gel stack. The membrane was then blocked with 1:1 Odyssey blocking buffer (Catalogue no: 927-40000) and PBS (150 mM NaCl and 50 mM phosphate buffer pH 7.4) for 2 hours at room temperature. The blot was briefly rinsed with wash buffer (150 mM NaCl and 50mM phosphate buffer pH 7.4, Tween-20 0.1%). GFP antibody (Catalogue no: SC-9996) diluted in the wash buffer (1:500) was added to the blot, followed by incubation overnight at 4 °C. The blot was extensively washed in wash buffer (6 x 5 minutes) with gentle agitation. Secondary antibody, IRDye 680LT goat anti-mouse IgG (Catalogue no: P/N 925-68020), diluted in wash buffer (1:5000) was added and incubated for 1 hour at room temperature with gentle agitation. The blot was extensively washed in wash buffer (6 x 5 minutes) with gentle agitation. The image was acquired using Licor Odyssey FC with 700 nm and 800 nm channels.

2.9 RNA isolation

RNA was isolated from TRAP elution using Quick-RNA microprep kit (Catalogue no. R1050) as follows. Four volumes of RNA lysis buffer were added to each volume of sample (4:1). The mixture was vortexed for 1 min, and equal volume of ethanol (95 - 100%) was added to the sample in RNA lysis buffer (1:1). The mixture was transferred to a zymo-spin column (provided in the kit) in a collection tube and centrifuged for 30 seconds. Flow-through was discarded. Four hundred of RNA prep buffer was added to the column and centrifuged for 30 seconds and flow through was discarded. Seven hundred µL RNA A wash buffer was added to the column and centrifuged for 30 seconds. Flow-through was discarded. Second washing was done by adding 400 µL RNA wash buffer B and centrifuging the column for 2 minutes to ensure the complete removal of the wash buffer. The column was transferred into an RNase-free tube. RNA was eluted using 10 µL of RNase free water and 0.2 µL RNase inhibitor (Catalogue no: M0253S) was added to prevent RNA from degradation. RNA was immediately stored at -80 °C until further use.

2.10 DNase treatment

DNase treatment of the total RNA was done using Turbo DNA-free kit (Catalogue No: AM1907). One µL turbo DNase and 0.1 volume 10x turbo DNase buffer was added to the RNA and mixed gently. The mixture was incubated at 37 °C for 30 min. DNase inactivation reagent (typically 0.1 volume) was added and further incubated for 5 min at room temperature. The tube was flicked 2–3 times during the incubation period to redisperse the DNase inactivation reagent. The mixture was

centrifuged at 10,000 ×g for 1.5 min and the RNA was transferred to a fresh tube. The DNase treated RNA was stored at -80 °C until further use.

2.11 Library preparation

ScriptSeq™ v2 RNA-seq Library preparation kit (Catalogue no: SSV21106) was used to prepare the library for sequencing as follows.

2.11.1 mRNA enrichment

2.11.1.1 mRNA isolation

Dynabeads mRNA direct micro kit (Catalog No. 61021) was used for mRNA isolation from Total RNA. Twenty-two µL of dynabeads oligo dT per sample was pipetted into a new 1.5 mL tube and placed the tube in a magnetic stand. The clear supernatant was discarded without disturbing the beads. The tube was removed from the magnetic stand and an equivalent volume of binding buffer was added to the beads and mixed thoroughly. Total RNA was heated at 70 °C for 2 minutes and 50 µL of binding buffer was added to each 50 µL of the prepared total RNA sample. Twenty µL of washed dynabeads oligo dT was pipetted into a new 1.5 mL tube. Hundred µL of the heat-denatured RNA mixture was added to the tube containing washed beads. The mixture was pipetted up and down 10 times, then incubated at room temperature for 5 minutes. The tube was placed on the magnetic stand. After the clear solution is obtained, the supernatant was discarded without disturbing the pellet. The tube was removed from the magnetic stand. Hundred µL washing buffer was added to each tube and mixed properly. The tube was again placed on the magnetic stand. After the clear solution is obtained, the supernatant was discarded without disturbing the pellet. The tube was removed from the magnetic stand.

Twenty-five µL of the pre-heated (80 °C) nuclease-free water was added to each well, mixed thoroughly and incubated for 30 seconds at room temperature. Re-binding of mRNA to the beads was done by adding 25 µL of binding buffer to each tube. The mixture was incubated at room temperature for 5 minutes. The tube was then placed on the magnetic stand. The supernatant was discarded without disturbing the pellet. Washing was done as described earlier. Ten µL of the warmed (80 °C) nuclease-free water was added to each tube. The tube was placed on the magnetic stand. The supernatant containing the mRNA was transferred to a new tube without disturbing the pellet. The isolated mRNA was stored at –80 °C.

2.11.1.2 rRNA Removal

For rRNA removal, Ribo-Zero® rRNA removal kit (MRZSR116) was used. For each reaction, 225 µL magnetic beads were added to a 1.5 mL microcentrifuge tube. The tube was placed on a magnetic stand, with the cap open, and waited until the liquid is clear (~1 minute). The supernatant was discarded, and the tube was removed from the magnetic stand. The beads were washed by adding 225 µL RNase-free water. The tube was placed on a magnetic stand and supernatant was discarded. Magnetic bead resuspension solution (65 µL) was added to the tube and vortexed to resuspend. Riboguard RNase inhibitor (1 µL) was added to the tube and then pipetted to mix. For each reaction, 10 µL of RNA sample, 4 µL of ribo-zero reaction buffer and 8 µL ribo-zero removal were combined in a tube. RNase-free water was added to make the final volume 40 µL. The tube was placed on the preheated heat block incubated for 10 minutes. The tube was removed from heat incubate at room temperature for 5 minutes. For each reaction, 40 µL RNA sample was added to a 1.5

mL tube containing 65 µL washed magnetic beads. Vortexed for 10 seconds, and then incubated at room temperature for 5 minutes. The tube was placed on the preheated heat block and incubated for 5 minutes. The tube was then immediately placed on a magnetic stand and transferred 85 – 90 µL supernatant containing depleted RNA to a fresh 1.5 mL tube. For each tube of the depleted sample, RNase-free water to bring the volume to 180 µL, 18 µL of 3 M sodium acetate and 22 µL of glycogen (10 mg/mL) were mixed thoroughly. Six hundred µL of 100% ethanol was added and incubated at -20 °C for at least 2 hours. The sample was centrifuged at 10,000 ×g for 30 minutes at 4 °C. All supernatant was discarded. The sample was washed with 200 µL freshly prepared 70% ethanol by centrifuging at 10,000 × g for 5 minutes at 4 °C. The supernatant was discarded, and the tube was dried at room temperature for 5 minutes. The pellet was dissolved in the 10 µL of RNase-free water and stored at –80°C.

2.11.2 Anneal the cDNA synthesis primer

Nine µL rRNA-depleted RNA was combined with 2 µL cDNA primer making the total volume of 11 µL. The reaction was incubated at 65 °C for 5 minutes in a thermocycler. The reaction was stopped by placing the tube on ice.

2.11.3 Synthesize cDNA

One µL fragmentation Solution, 3.0 µL cDNA synthesis premix, 0.5 µL 100 mM DTT and 0.5 µL starscript AMV reverse transcriptase were added to each reaction. The mixture was thoroughly mixed by pipetting 10 times and incubated at 25 °C for 5 minutes followed by 42 °C for 20 minutes. The reaction was cooled to 37 °C and paused the thermocycler. The tube was removed from the thermocycler,

and 1.0 µL of finishing solution was added. The mixture was mixed gently by pipetting. The tube was placed back in the thermocycler and incubate at 37 °C for 10 minutes followed by 95 °C for 3 minutes. The reaction was cooled to 25 °C and thermocycler was paused before proceeding to the next step.

2.11.4 Synthesize 3′-tagged DNA

Tube was removed from the thermocycler and 7.5 µL terminal tagging premix and 0.5 µL DNA polymerase was added. The mixture was thoroughly mixed by pipetting. After mixing, the tube was returned to the thermocycler and incubated at 25 °C for 15 minutes followed by 95 °C for 3 minutes. Then, the reaction was cooled to 4 °C.

2.11.5 Purify the cDNA

The cDNA was purified using the Minelute kit (catalogue No: 28004). Five volumes of buffer PB was added to 1 volume of the PCR reaction and mixed. Minelute column was placed in a provided 2 ml collection tube. The sample was transferred to the Minelute column and centrifuged for 1 min at 17,900 x g. Flow-through was discarded. To wash, 750 µL buffer PE was added to the Minelute column and centrifuged for 1 minute at 17,900 x g. Flow-through was discarded. The column was centrifuged for an additional 1 minute at 17,900xg. Minelute column was placed in a clean 1.5 ml microcentrifuge tube. To elute DNA, 25 µL Buffer EB (10 mM Tris·Cl, pH 8.5) was added to the center of the membrane, let the column stand for 1 minute, and then centrifuged for 1 minute at 17,900 xg.

2.11.6 PCR amplify the library and add an index (barcode)

One µL forward PCR primer, 1 µL scriptseq index PCR primer, 25 µL failsafe PCR premix E and 0.5 µL failsafe PCR enzyme was added to the 22.5 µL of di-tagged cDNA making the total volume 50 per reaction. The mixture was incubated at 95 °C for 1 minute (to denature DNA), followed by 15 cycles of 95 °C for 30 seconds, 55 °C for 30 seconds and 68 °C for 3 minutes. After the 15 PCR cycles, it was incubated at 68 °C for 7 minutes.

2.11.7 Purify the RNA-Seq library

Excess PCR primers were removed by adding 1 µL of exonuclease I to each reaction and incubated the reactions at 37 °C for 15 minutes. The library was purified using the Minelute kit (Catalogue No: 28004) as described earlier in section **10.5**.

2.11.8 Assess library quantity and quality

A qualitative check of the prepared library was performed using bioanalyzer.

2.12 cDNA synthesis

DNase treated RNA was reverse-transcribed using M-MuLV reverse transcriptase (Catalogue no: M0253S) to synthesize first-strand complementary DNA. DNase treated RNA (14.5 µl), 1 µL of 10Mm dNTP mix and 10 µM oligodT were mixed in a PCR tube. The mixture was incubated at 75 ^{0}C for 5 minutes in a thermocycler and immediately placed on ice for 5 minutes. Two µL of 10x M-MuLV RT buffer, 0.5 µL of M-MuLV reverse transcriptase and 1 µL of M-MuLV RNase inhibitor were added to each reaction and incubated at 42 ^{0}C for 1 hour. The reaction was inactivated by heating the mixture to 90 ^{0}C for 5 minutes.

2.13 DNA contamination test

DNA contamination test was done to check the purity of RNA before proceeding with quantitative RT-PCR (qRT-PCR) for gene expression analysis. Equivalent amount of total RNA from three samples (ENOD2p-, ENOD40p- and CsVMVp-TRAP derived samples) at two time points (7 and 10 dpi) was used as template and qRT-PCR was performed to check the expression all genes used in this experiment. Reactions conditions used for qRT-PCR are explained below in section **2.14.** Absence of amplification indicated the absence of any contaminating DNA in the total RNA.

2.14 Reverse Transcription – quantitative Polymerase Chain Reaction (RT-qPCR)

Eighteen genes that showed the most differential expression in INACT samples were selected (Appendix N). Primers were designed using primique and IDT Oligoanalyzer tool, and the parameter for primer design are listed in (Appendix I). A standard curve was generated based on a serial dilution from 1:3 to 1:51 at 3-fold intervals and efficiency was calculated for all primers (Appendix J). SYBR advantage qPCR premix (catalogue No: 639676) was used, and PCR was performed on Quantstudio Q6 qPCR system using SYBR Green detection chemistry. The reaction conditions were: 95 °C for 2 minutes, 40 cycles of 95 °C for 10 seconds and 62°C for 20 seconds. The dissociation curve was determined using the thermal cycle at 55°C for 30sec followed by heating at 0.1°C/sec to 95°C for 1min. Three biological replications were included, and each replicate was assayed in triplicate. Each replicate sample was harvested from approx. 30-50 independent transgenic

roots. Data were normalized to cons7, and dCt value was obtained for each gene. Statistical significance of differences in gene expression was evaluated using the Duncan test, $P<0.05$ using R version 3.3.0. The output of Duncan test was verified using the Tukey test, $P<0.05$ using R version 3.3.0 and similar results were obtained from both tests. Hence, all the statistical analysis was performed using the Duncan test, $P<0.05$ using R version 3.3.0. Log_2 fold change (ENOD40p-TRAP/ENOD2p-TRAP) was calculated to find the relative enrichment of each gene in two different tissue at 7 dpi (Appendix K) and 10 dpi (Appendix L).

2.15 Gene Annotation

The peptide sequences of genes were obtained from Soybean Knowledge Base (http://soykb.org/search/gene.php). These peptide sequences were used as a query in a TBLASTN search against the soybean genome in LegumeIP. Transcription factors and a list of genes involved in biosynthesis and signaling of plant hormones such as auxin were obtained from the lab.

3 RESULTS

3.1 Localization of the tissue-specific promoter activity

ENOD2 and ENOD40 promoter driven gene cassettes were transformed into *Agrobacterium rhigozenes* (K599 strain), and subsequently into soybean hairy root composite plants to tag ribosomes in nodule parenchyma and infection zone respectively. A TRAP gene cassette driven by the constitutive promoter, CsVMV was used as comparison control as well as to optimize the TRAP method in soybean root tissues (Figure 3.1).

Figure 3.1 TRAP constructs used in the study. This illustration shows the arrangement of TRAP gene cassettes that encode a fusion protein consisting of Flag/His epitope (for affinity purification), followed by GFP (for visualization), and RPL18 (Large Subunit Ribosomal protein for tagging ribosomes). The cassettes were driven by ENOD2, ENOD40, or CsVMV promoter in independent vector constructs.

Soybean plants subject to *Agrobacterium rhigozenes* mediated hairy root plant transformation consist of adventitious (non-transgenic roots), and transgenic roots with and without the binary vector cassette. The roots were screened using a fluorescence microscope for GFP fluorescence to identify transgenic roots with the binary vector cassette. Green fluorescence was specific to nodules in roots transformed with the ENOD2 and ENOD40 promoter driven TRAP constructs (Figure 3.2) and could only be observed in rhizobium-inoculated (and nodulated) roots. Roots transformed with the CsVMV promoter driven TRAP construct showed

green fluorescence in the entire root system of uninoculated and inoculated plants (Figure 3.3).

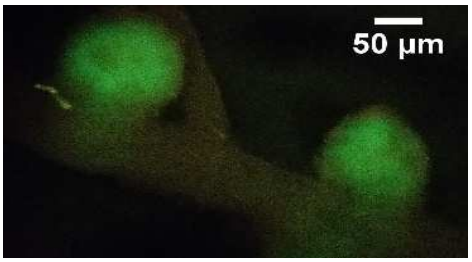


Figure 3.3. A ENOD40p-TRAP transgenic root containing nodules. The figure shows GFP fluorescence in nodules on a root transformed with the ENOD40p-TRAP cassette.

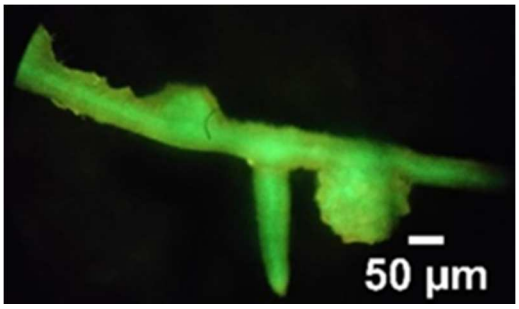


Figure 3.2 A CsVMVp-TRAP transgenic root containing nodules. The figure shows GFP fluorescence in a root transformed with the CsVMVp-TRAP cassette.

Transverse sections of several transgenic nodules showing GFP expression were imaged under a laser scanning confocal microscope to evaluate the sites of expression of TRAP cassettes. At 7 dpi, inner cortical cells have divided to form a clearly visible nodule primordium. At this stage, nodules appeared as small bulges on the roots. At 10 dpi, infection zone and parenchyma zone are clearly differentiated: infection zone in the center surrounded by parenchyma region that holds the vascular bundles. Infection zone contains infected and uninfected cells. Transgenic nodule from a root transformed with ENOD2p-TRAP cassette showed GFP expression in the parenchyma zone (Figure 3.5) and transgenic nodule from a root transformed with ENOD40p-TRAP cassette showed the GFP expression in the infection zone at 7 and 10 dpi (Figure 3.4). Tissue/Nodule zone-specific GFP epifluorescence suggested that the cassettes indeed drove expression of the TRAP cassette in the expected tissue types.

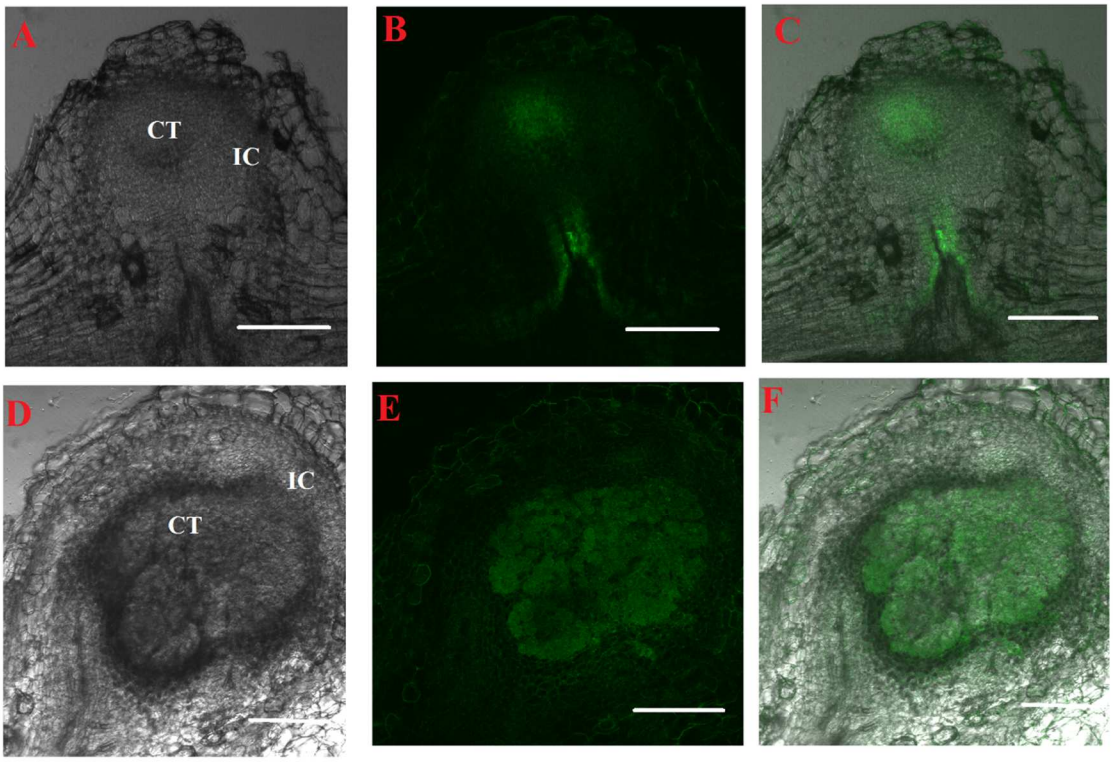


Figure 3.4 Spatio-temporal localization of GFP expression in transverse sections of nodules from a soybean root transformed with ENOD40p-TRAP cassette. The figure shows GFP fluorescence in cross sections of transgenic nodules from a soybean root transformed with ENOD40p-TRAP cassette at 2 different time points, 7 dpi and 10 dpi. Panels A and D shows bright field image; B and E shows GFP image; C and F shows white and GFP merged image.; IC: Inner cortex; CT: Central tissue. Panels A to C represent the nodule at 7 dpi and panels D to F represent nodules at 10 dpi. The bar represents a 200µm scale.

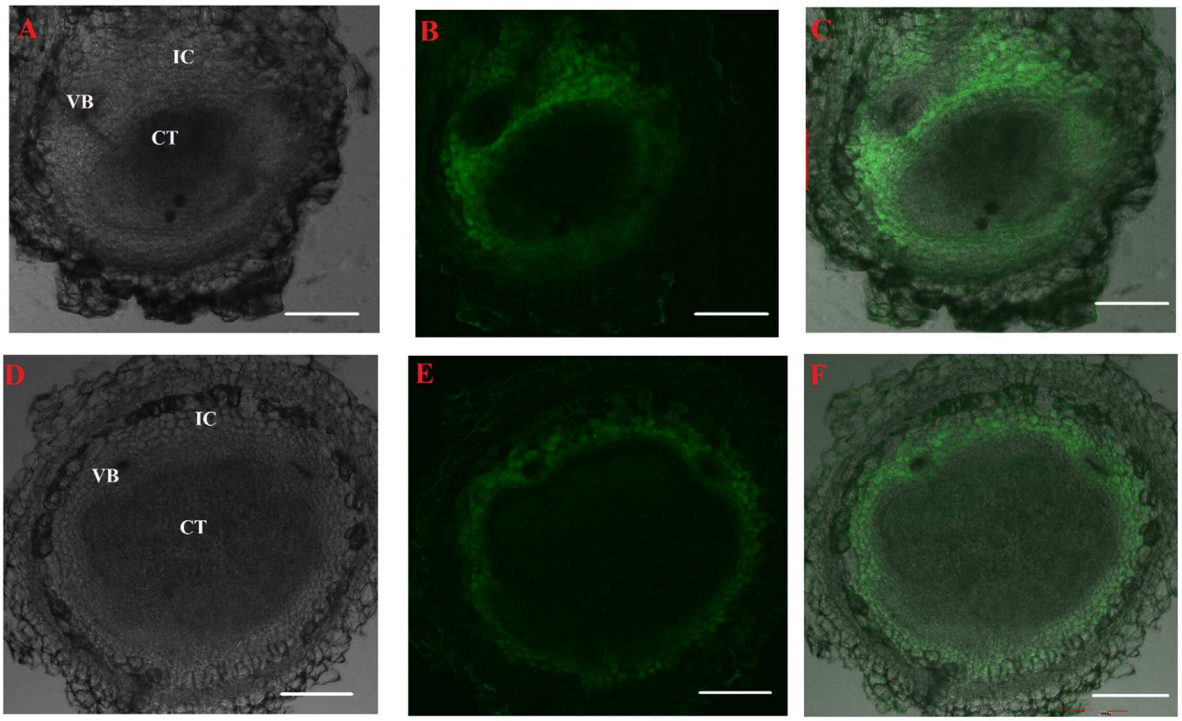


Figure 3.5 Spatio-temporal localization of GFP expression in transverse sections of nodules from a soybean root transformed with ENOD2p-TRAP cassette. The figure shows GFP fluorescence in cross sections of transgenic nodules from a soybean root transformed with ENOD2p-TRAP cassette at 2 different time points, 7 dpi and 10 dpi. Panels A and D shows bright field image; B and E shows GFP image; C and F shows white and GFP merged image.; IC: Inner cortex; CT: Central tissue; VB: nodule vascular bundle. Panels A to C represent the nodule at 7 dpi and panels D to F represent nodules at 10 dpi. The bar represents a 200µm scale.

Tagged ribosomes from roots transformed with the CsVMVp-TRAP cassette were purified using immobilized metal affinity chromatography (IMAC) as described in sub-section **2.7**. The crude extract and different fractions were separated on a SDS-PAGE gel and transferred onto a PVDF membrane via Western blotting. The membrane was probed with GFP antibody to monitor the presence of tagged RPL18 in the crude protein (clarified extract), column flow through, and final elution. A 50 kDa band (expected size of RPL18) was detected in crude extract and eluted fraction. No band was observed on the column flow through (Figure 3.7). This demonstrated the successful isolation of tagged ribosomes from roots transformed

with the CsVMVp-TRAP cassette. To evaluate if intact ribosomes were purified, the preparations were subject to agarose gel electrophoresis. Clarified extract, column flow through, washed fraction and three subsequent elutions were loaded in an agarose gel to check the quality of RNA (Figure 3.6). The agarose gel images confirmed that TRAP methods yielded intact ribosomes. These observations suggested that TRAP is an efficient method to isolate intact ribosomal complexes from transgenic soybean hairy roots.

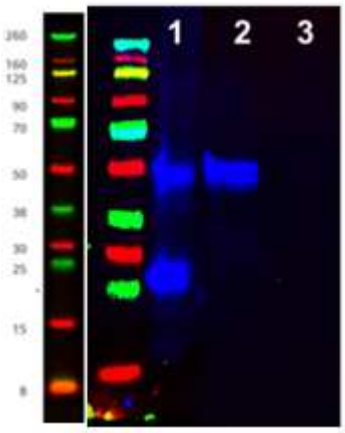


Figure 3.7 Detection of RPL18:GFP fusion protein. The figure shows a Western blot probed with anti-GFP (to detect RPL18:GFP) in (1) crude extract, (2) Column flow through, and (3) an eluted fraction (2) obtained from roots transformed with the CsVMVp-TRAP cassette 10 dpi with rhizobium. The image was acquired using Licor Odyssey FC with 700 nm and 800 nm channels.

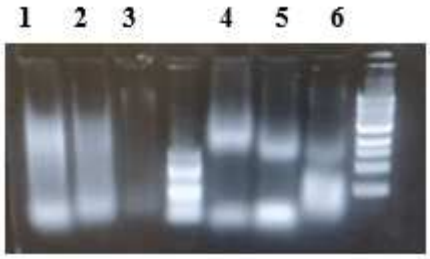


Figure 3.6 Agarose gel image of different fractions obtained from affinity purification. Agarose gel electrophoresis of 1) Clarified extract 2) column flow through 3) washed fraction and (4-6) the three subsequent eluted fractions from the affinity purification of roots transformed with the CsVMVp-TRAP cassette 10 dpi with rhizobium. The gel was stained with Ethidium bromide and imaged on a UV-transilluminator.

3.3 Library Preparation

RNA was isolated from the purified ribosome preparations using Quick-RNA microprep kit as described in **2.9**, and mRNA was isolated from the purified RNA using Dynabeads® mRNA direct micro kit as described in **2.11.1.1**. mRNA was also

isolated from total RNA fractions of nodules from non-transgenic roots as positive control samples for mRNA isolation and library construction. RNA-Seq libraries were prepared using the Scriptseq v2 library preparation kit as described in 2.11. Quality of the prepared libraries was evaluated using a Bioanalyzer in SDSU Genome Sequencing Core Facility. The library prepared from the positive control samples had fragments sizes of 200-400 as expected. However, the library from CsVMVp-TRAP derived RNA sample had much shorter fragments suggesting that library synthesis was ineffective (Figure 3.8). As the quality (260/280 and 260/230 ratios) and quantity of RNA preparations (absorbance at 260nm) were checked and confirmed by nanodrop reading, it was likely that the mRNA isolation kit used for library preparation might not be effective for TRAP derived RNA samples.

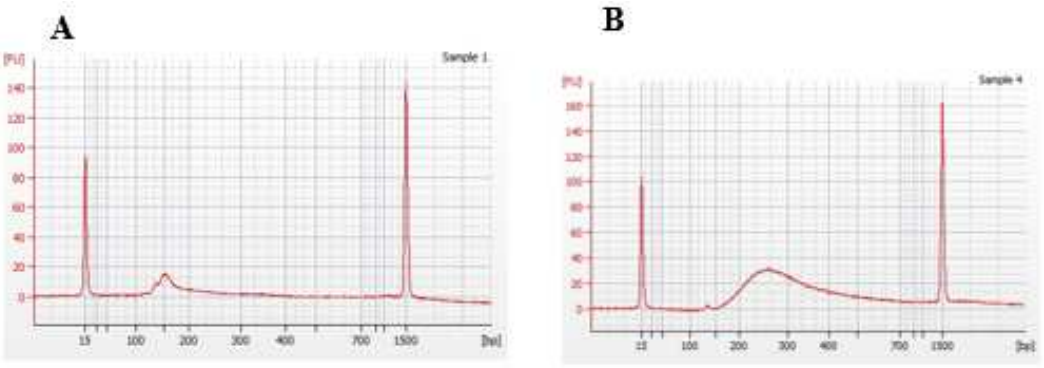


Figure 3.8 Evaluation of RNA-Seq library quality using Bioanalyzer. Figure shows the Bioanalyzer profiles of RNA-Seq libraries prepared from (A) CsVMVp-TRAP derived RNA at 10 dpi, and (B) Non-transgenic mature nodules at 21 dpi.

As an alternative to mRNA isolation, r-RNA depletion was used to enrich mRNA in the RNA preparations. RNA isolated from the roots transformed with CsVMVp-TRAP cassette (at 10 dpi stage) were processed using the Ribo-Zero rRNA removal kit as described in 2.11.1.2. RT-qPCR was conducted using equal amounts of input RNA to compare the efficiency of each approach by evaluating the

expression of Actin and Cons 7 genes. For both genes tested, C_t values for rRNA depleted samples were significantly lower than that for total ribosomal RNA and mRNA preparations (Figure 3.9). The results indicated that rRNA depletion was more efficient in enriching mRNA preparations for cDNA synthesis (as equal amounts of input RNA were used for both methods). RNA isolated using mRNA isolation kit showed significantly higher C_t values for Actin and Cons7 than the total ribosomal RNA preparations suggesting that either there was lower abundance of mRNA or that these preparations are poorly suited for cDNA synthesis. These results suggested that rRNA depletion was more efficient than mRNA isolation to prepare cDNA from TRAP-derived RNA preparations.

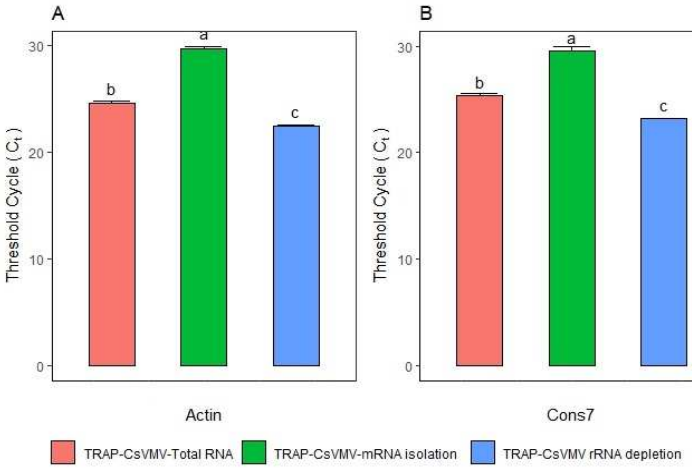


Figure 3.9 Evaluation of rRNA depletion and mRNA isolation approach. The figure shows the levels of (A) Actin, and (B) Cons7 in total RNA, mRNA and rRNA depleted RNA from the CsVMVp-TRAP derived RNA at 10 dpi assayed by RT-qPCR. C_t values are plotted on the y-axis, and genes methods the x-axis. Data shown are the average of 3 technical replicates and error bars indicate SD. Samples marked with different letters are significantly different from each other based on Duncan test. ($P<0.05$).

3.5 Library preparation with rRNA-depleted RNA

mRNA enrichment was done by depleting the rRNA from total ribosomal RNA preparations as described in **2.11.1.2**, followed by library preparation using the

Scriptseq v2 library preparation kit as described in **2.11**. Total RNA fractions of nodules from non-transgenic roots as positive control samples were also used for mRNA enrichment by depleting the rRNA and for library construction. Quality of the prepared libraries was evaluated using a Bioanalyzer in SDSU Genome Sequencing Core Facility. The library prepared from the positive control samples had fragments sizes of 200-400 as expected. However, the library from CsVMVp-TRAP derived rRNA depleted RNA had much shorter fragments suggesting that library synthesis was ineffective (Figure 3.10).

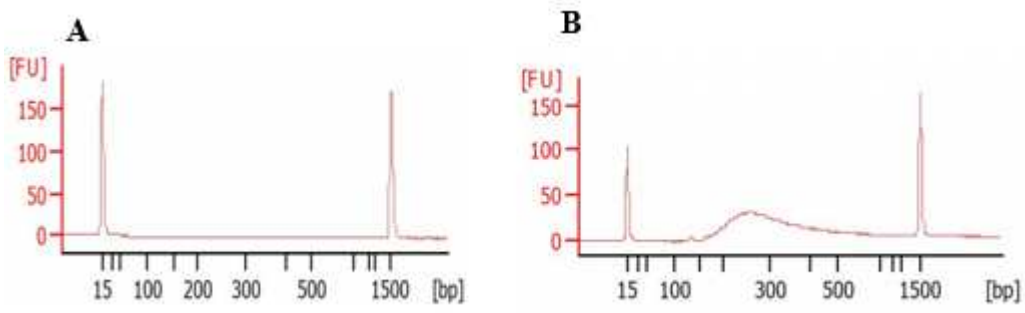


Figure 3.10 Evaluation of RNA-Seq library quality using Bioanalyzer. Figure shows the Bioanalyzer profiles of RNA-Seq libraries prepared from (A) CsVMVp-TRAP derived RNA after 10 dpi, and (B) Non-transgenic matured nodules at 21 dpi after rRNA depletion.

3.6 Analysis of marker gene expression patterns at 7 and 10 dpi time points

Since we were able to successfully evaluate gene expression using RT-qPCR, we evaluated the TRAP RNA preparations for expression patterns of known nodule zone-specific marker genes using this approach. ENOD2 is known to be expressed in the nodule parenchyma region (van de Wiel et al. 1990) and ENOD40 gene is known to be expressed in developing nodule primordium and in the uninfected cells of the infection zone in a matured nodule (Yang et al. 1993). CYP83B1 is known to be

expressed in the nodule parenchyma, primarily in the inner cortex (Damodaran et al. 2018).

3.6.1 Analysis of marker gene expression patterns at 7 dpi time point

At 7 dpi, average Ct values for Cons 7 in Enod2p-, Enod40p- and CsVMVp-TRAP derived RNA were 25.9 ± 0.42, 24.3 ± 0.44 and 25.9 ± 0.45 in respectively (based on three biological replicates). Similarly, for Actin, average Ct values were 24.9± 0.5, 25.4± 0.5 and 25.8± 0.6 in Enod2p-, Enod40p- and CsVMVp- TRAP derived RNA respectively (based on three biological replicates). As the expression of Cons 7 was more consistent than the expression of Actin at both 7 and 10 dpi (see below) time points among different biological replicates, Cons 7 was used for normalization.

At 7 dpi, the abundance of ENOD2 and CYP83B1 mRNAs was significantly higher in ENOD2p-TRAP derived RNA (nodule parenchyma) compared to ENOD40p-TRAP derived RNA (nodule infection zone). Abundance of ENOD2 and CYP83B1 mRNAs in nodule parenchyma was around 1000 and 32 times more than in infection zone respectively. Abundance of ENOD40 mRNA was significantly higher (~32-fold) in nodule infection zone compared to nodule parenchyma. Expected patterns of marker gene (ENOD2, ENOD40, and CYP83B1) enrichment suggested that the samples were enriched in RNA from the targeted nodule zones. In addition, there was a significantly reduced abundance of ENOD2 and CYP83B1 mRNAs in nodule infection zone compared to entire nodule (CsVMVp-TRAP derived RNA). Similarly, there was a significantly reduced abundance of ENOD40 mRNA in nodule parenchyma compared to entire nodule (Figure 3.11). These

observations suggested that using the ENOD2p- and ENOD40p- TRAP constructs, we were able to obtain tissue/nodule zone-enriched ribosomal preparations at 7dpi.

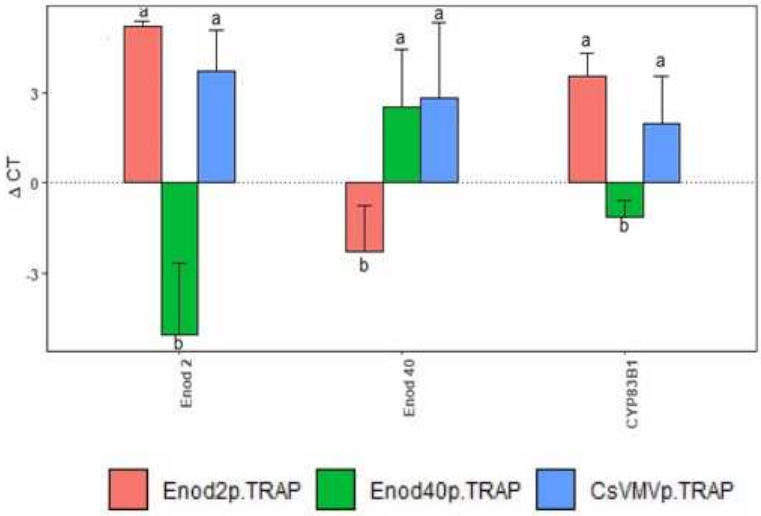


Figure 3.11 Marker genes expression pattern in ENOD2p-, ENOD40p- and CsVMVp-TRAP derived RNA at 7 dpi. The figure showed the difference in the expression pattern of marker genes in ENOD2p-, ENOD40p-, and CsVMVp- TRAP derived RNA assayed by RT-qPCR. Delta Ct values (normalized to Cons7) are plotted on the y-axis and marker genes at x-axis. Data shown are the average of 3 biological replicates and error bars indicate SD. Samples marked with different letters are significantly different from each other based on Duncan test. ($P<0.05$).

3.6.2 Analysis of marker gene expression patterns at 10 dpi time point

At 10 dpi, the average Ct values for Cons 7 were 23.6±0.43, 22.3±0.38 and 23.8±0.6 in Enod2p-, Enod40p- and CsVMVp- TRAP derived RNA respectively (based on 3 biological replicates). The average Ct values for Actin were 24.5±0.6, 22.8± 0.9 and 23.3± 1.0 in Enod2p-, Enod40p- and CsVMVp- TRAP derived RNA respectively.

At 10 dpi, the abundance of ENOD2 and CYP83B1 mRNAs was significantly higher in ENOD2p-TRAP derived RNA (nodule parenchyma) compared to ENOD40p-TRAP derived RNA (nodule infection zone). Abundance of ENOD2 and

CYP83B1 mRNAs in nodule parenchyma was around 8000 and 64 times more than in infection zone respectively. Similarly, abundance of ENOD40 mRNA was significantly higher (~500-fold) in nodule infection zone compared to nodule parenchyma. Expected patterns of marker gene (ENOD2, ENOD40, and CYP83B1) enrichment suggested that the samples were enriched in RNA from the targeted nodule zones. In addition, there was a significantly reduced abundance of ENOD2 and CYP83B1 mRNAs in the infection zone (ENOD40p-TRAP derived RNA) compared to the entire nodule. Similarly, there was a significantly reduced abundance of ENOD40 mRNA in nodule parenchyma compared to entire nodule (Figure 3.12). These observations suggested that using the ENOD2p- and ENOD40p- TRAP constructs, we were able to obtain tissue/nodule zone-enriched ribosomal preparations at both 7 and 10 dpi.

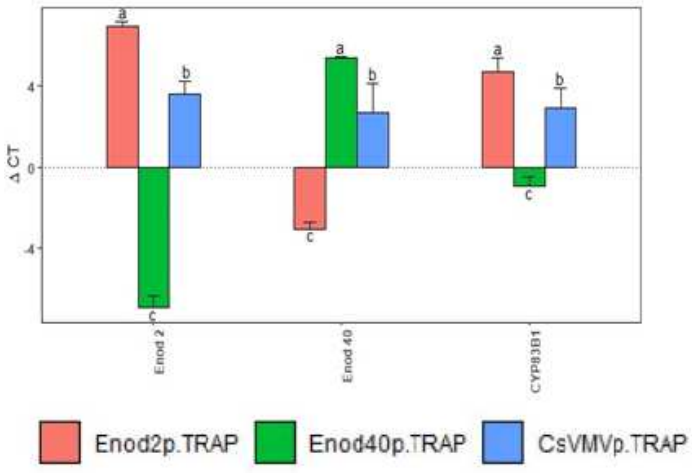


Figure 3.12 Marker genes expression pattern in ENOD2p-, ENOD40p- and CsVMVp-TRAP derived samples at 10 dpi. The figure showed the difference in the expression pattern of marker genes in ENOD2p-, ENOD40p- and CsVMVp-TRAP derived RNA assayed by RT-qPCR. Delta Ct values (normalized to Cons7) are plotted on the y-axis and marker genes at x-axis. Data shown are the average of 3 biological replicates and error bars indicate SD. Samples marked with different letters are significantly different from each other based on the Duncan test. ($P<0.05$).

3.7 Evaluation of precursor microRNA and non-coding RNA

To evaluate the purity of the TRAP RNA preparations and the efficiency of TRAP method, we evaluated the TRAP RNA preparations for the abundance of precursors of miRNA 160 and miRNA 166. Since microRNAs precursors are non-coding RNAs that are processed in the nucleus and not loaded onto the ribosomes, their absence or reduced abundance would indicate high quality TRAP RNA preparations. On the other hand, total RNA comprised of both nuclear and cytoplasmic RNAs is expected to contain non-coding RNAs. We used total RNA from non-transgenic nodules as positive control for miRNA precursor qPCR. Ct values of 25.6 and 26.8 were observed respectively for miRNA160 and miRNA166 suggesting that these precursors were successfully assayed by qPCR. We did not detect precursor miRNA 160 and 166 in ENOD2p-, Enod40p-, or CsVMVp-TRAP derived RNA at both 7 and 10 dpi (Figure 3.13) suggesting that they were absent or reduced below detectable levels in TRAP-derived RNA.

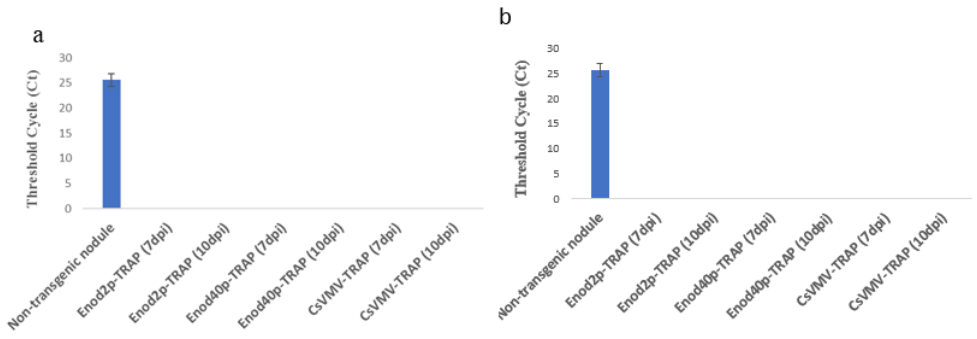


Figure 3.13 Abundance of miRNA 160 and miRNA 166 in TRAP samples. Threshold cycle (Ct) values of (a) miRNA 160 and (b) in non- transgenic mature nodule and TRAP RNA samples (ENOD2p-, ENOD40p-, CsVMVp-TRAP derived RNA) at 7 and 10 dpi. Data shown are the average of 3 biological replicates and error bars indicate SD.

We also evaluated the expression of a putative non-coding RNA (GLYMA19G06330) in ENOD2p-, ENOD40p- and CsV MVp-TRAP derived RNA at both time points. This non-coding RNA was detectable in all TRAP samples, but at very low levels. In TRAP derived RNAs, the abundance of this non-coding gene mRNA was around 200 times less compared to that of Cons 7 (Figure 3.14 and 3.15). However, the expression of this non-coding gene in RNA from non-transgenic matured nodules (as positive RNA) was only 12 times less compared to housekeeping gene, Cons7. Although the gene is characterized as non-coding, it may be possible that some part of the gene is still loaded in the ribosomes for translation. However, the complete absence of precursor miRNA 160 and miRNA 166 in TRAP derived RNAs confirmed that TRAP RNA preparations were free of total or nuclear RNA contamination.

3.8 Differential gene expression between the two nodule zones

To identify genes enriched in nodule parenchyma and infection zones, we selected a set of genes based on functional significance (auxin biology-related genes, transcription factors, and transporters) and nodule zone-enrichment determined by INTACT. Selected genes, their functional characterization and their expression pattern in INTACT samples are presented in Appendix N.

3.8.1 Differential gene expression between two nodule zones at 7 dpi

At 7 dpi, the abundance of ABC transporter G family member (GLYMA19g35270) mRNA was significantly higher in ribosomes of nodule parenchyma (ENOD2p-TRAP derived RNA) compared to infection zone (ENOD40p-TRAP derived RNA) with $\log_2$ fold change value 8.9. Similarly, the

abundance of *CUP-SHAPED COTYLEDON3* (GLYMA19G34881) mRNA was significantly higher in nodule parenchyma compared to infection zone with log_2 fold change value 4.3. The abundance of auxin efflux carrier (GLYMA01G36190.1) mRNA, auxin- responsive GH3 family protein (GLYMA05G21680) mRNA, and auxin response factor, ARF5 (GLYMA17G37580.1) mRNA were significantly higher in nodule parenchyma compared to infection zone with log_2 fold change values of 9.2, 2.5 and 4.3 respectively. However, the abundance of NAC domain (GLYMA12G09670) mRNA was significantly higher in infection zone compared to nodule parenchyma with log_2 fold change value of 5.9. mRNAs of nitrate transporter (GLYMA11G04500.2) and the *PHOSPHATE 2* gene (GLYMA13G24810.1; encoding ubiquitin-conjugating enzyme 24) were significantly higher in infection zone compared to nodule parenchyma with log_2 fold change values of 2.7 and 3.5 respectively. There was no significant difference between nodule parenchyma and infection zone in the abundance of the mRNAs of the transcription factors MyB-related transcription factor (GLYMA03G42260.1), bZIP transcription factor (GLYMA19G43420.1), and bHLH transcription factor (GLYMA08G04661.1). Similarly, there was no significant difference in the abundance of auxin- induced (GLYMA19G30640) mRNA and *ETHYLENE INSENSITIVE 3* family (GLYMA05G31410) mRNA between nodule parenchyma and nitrogen fixation zone. The abundances of *LONGIFOLIA* protein (GLYMA17G37580) mRNA and disease resistance protein of TIR-NBS-LRR class (GLYMA16G23800.2) mRNA were not significantly different between nodule parenchyma and infection zone. There was no significant difference in the abundance of the phosphate transporter1

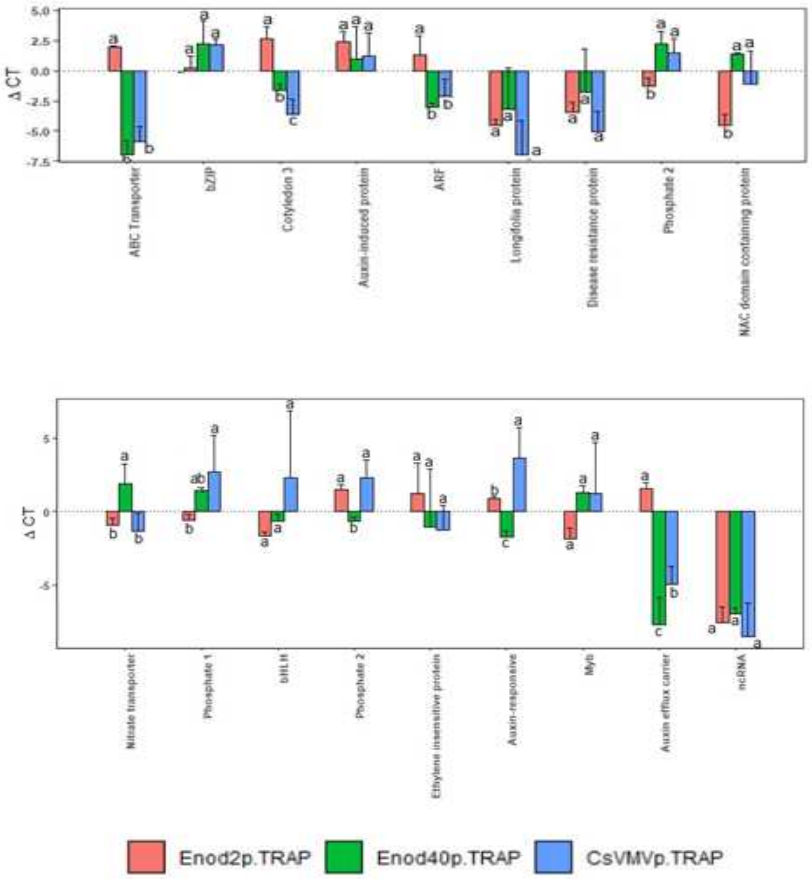


Figure 3.14 Gene expression pattern in ENOD2p-, ENOD40p- and CsVMVp-TRAP derived samples at 7 dpi. The figure showed the differences in the expression pattern of 18 genes in ENOD2p-, ENOD40p- and CsVMVp-TRAP derived RANA assayed by RT-qPCR. Delta Ct values (normalized to Cons7) are plotted on the y-axis and genes at x-axis. Data shown are the average of 3 biological replicates and error bars indicate SD. Samples marked with different letters are significantly different from each other based on Duncan test. ($P<0.05$).

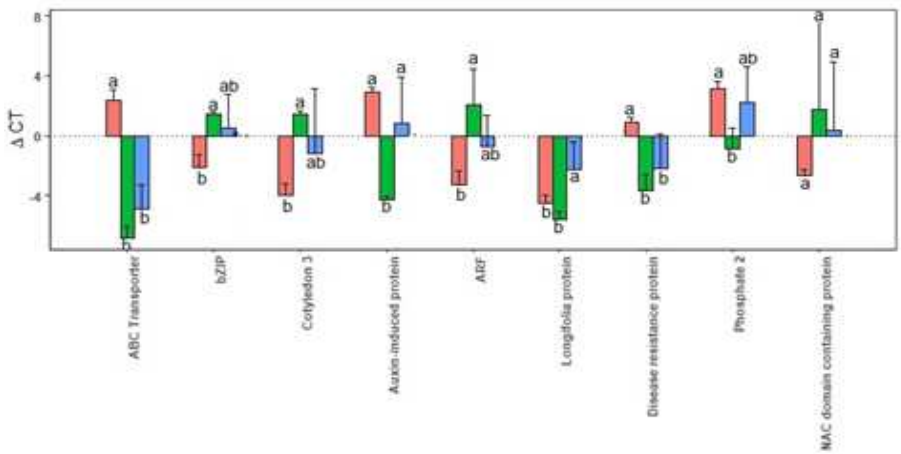
(GLYMA10G00720.1) mRNA between nodule parenchyma and infection zone.

3.8.2 Differential gene expression between two nodule zones at 10 dpi

At 10 dpi, the abundance of ABC transporter G family member (GLYMA19g35270) mRNA was significantly higher in nodule parenchyma (ENOD2p-TRAP derived RNA) compared to infection zone (ENOD40p-TRAP

derived RNA) with log_2 fold change value of 9.1. The abundance of *ETHYLENE INSENSITIVE - 3* family (GLYMA05G31410) mRNA, auxin efflux carrier (GLYMA01G36190.1) mRNA, auxin-induced (GLYMA19G30640) mRNA and auxin-responsive GH3 family protein (GLYMA05G21680) mRNA were significantly higher in nodule parenchyma compared to infection zone with log_2 fold change values of 7.3, 3.5, 7.1 and 4.2 respectively. Similarly, the abundance of TIR-NBS-LRR class (GLYMA16G23800.2) mRNA was significantly higher in nodule parenchyma compared to infection zone with log_2 fold change value of 4.5. The abundance of phosphate transporter 1 (GLYMA10G00720.1) mRNA was significantly higher in nodule parenchyma compared to infection zone with log_2 fold change value of 6.5. However, the abundance of *CUP-SHAPED COTYLEDON3* (GLYMA19G34881) mRNA was significantly higher in infection zone compared to nodule parenchyma with log_2 fold change value of 5.4. The abundance of the mRNAs of transcription factors such as MyB- related transcription factor (GLYMA03G42260.1), bZIP transcription factor (GLYMA19G43420.1), and bHLH transcription factor(GLYMA08G04661.1) were significantly higher in infection zone compared to nodule parenchyma (Figure 3.8) with log_2 fold change values of 6.3, 3.5 and 4.4 respectively. Similarly, the abundance of mRNA of ARF5 (GLYMA17G37580.1) was significantly higher in infection zone compared to nodule parenchyma with log_2 fold change value of 5.3. The abundance of NAC domain (GLYMA12G09670) mRNA and nitrate transporter (GLYMA11G04500.2) mRNA was significantly higher in infection zone compared to nodule parenchyma with log_2 fold change values of 4.3 and 5.3 respectively. There was no significant

difference in the abundance of mRNA of *LONGIFOLIA* gene (GLYMA17G37580) between nodule parenchyma and infection zone. Similarly, there was no significant difference in the abundance of the *PHOSPHATE 2* gene (GLYMA07G31630.1, encoding ubiquitin-conjugating enzyme 24) mRNA, between nodule parenchyma and infection zone.



3.9 Comparative enrichment analysis between nodule parenchyma and infection zone

mRNAs of ABC transporter G family member 34-related, Auxin efflux carrier family protein, Auxin-responsive GH3 family protein, Auxin-induced protein 5NG4-like, *PHOSPHATE 2*, *ETHYLENE INSENSITIVE* 3 family protein were enriched in nodule parenchyma region (ENOD2p-TRAP derived RNA) at both 7 and 10 dpi. Similarly, nitrate transporter, bZIP, MYB homeodomain-like superfamily

protein, NAC domain mRNAs were enriched in infection zone (ENOD40 p-TRAP derived RNA) at both 7 and 10 dpi. mRNAs of bHLH transcription factor family, ARF5 and *PROTEIN CUP-SHAPED COTYLEDON* 3 genes were enriched in nodule parenchyma at 7 dpi. However, they were enriched in infection zone at 10 dpi. Similarly, *LONGIFOLIA PROTEIN*, phosphate transporter 1 and disease resistance protein (TIR-NBS-LRR class) mRNAs were enriched in infection zone at 7 dpi and were enriched in nodule parenchyma at 10 dpi.

These observations suggested that using the ENOD2p- and ENOD40p- TRAP constructs, we were able to identify spatio and/or temporal specific expression of the selected genes in nodule parenchyma and infection zone at 7 and 10 dpi.

4 DISCUSSION

4.1 Nodule zone-specific promoters

An experiment carried by Yang et al. 1993 showed that ENOD40 is expressed specifically in developing nodule primordium, and in the uninfected cells of the infection zone in a mature soybean nodule. Similarly, ENOD2 is specifically expressed in soybean nodule parenchyma as demonstrated by (van de Wiel et al. 1990). Therefore GmENOD40 and GmENOD2 were used as promoters to drive TRAP constructs respectively in parenchyma and infection zone in soybean. As expected, ENOD40p-TRAP expression was localized to the infection zone and that of ENOD2p-TRAP to the nodule parenchyma zone based on GFP fluourescence. The native mRNAs of ENOD40 and ENOD2 were enriched respectively in ENOD40p-TRAP and ENOD2p-TRAP derived RNA preparations suggesting that the promoters were suitable to obtain translating ribosomes from the target tissues. In addition, independent confirmation was provided by the enrichment of GmCYP83B1 in nodule parenchyma.

At 7 dpi, transgenic nodule from a root transformed with the ENOD40p-TRAP cassette showed the expression of GFP in central tissue and in the vascular bundle that connects central tissue and root stele. The expression level of GFP was similar in both tissues, suggesting that using ENOD40 promoter, at 7dpi, will yield translating ribosomes not only from the nodule central tissue but also from connecting vascular bundles. This is a limitation and needs to be overcome by the use of a more suitable promoter. At 10 dpi, in transgenic nodules from roots transformed with the ENOD40p-TRAP gene cassette, GFP was localized to central

tissue (infection zone). However, it is also known that GmENOD40 mRNA is expressed specifically in the uninfected cells (Yang et al.1993). Even though we used ENOD40 promoter to study the gene expression profiles of infection zone, gene expression profiles from infected cells are missing in our study. It is known that Nodulin-93 transcripts appeared only in the infected cells of mature nodules (Kouchi et al. 1993). Similarly, Nodulin-35 is found only in the specialized uninfected cells which process fixed nitrogen into ureides for transport to the rest of the plant (Bergmann et al. 1983). The use of these cell type-specific promoters is likely to give us detailed information about genes specifically involved in nitrogen fixation and transport in the infection zone. Similarly, ENOD2 promoter-driven TRAP gene cassette showed the GFP expression in inner cortex surrounding central tissue, and in the tissue surrounding the connecting vascular bundle. To study the gene expression profiles of pericycle cells in vascular bundles, Nodulin-36 can be used as a promoter as it was known to be expressed specifically in the pericycle in vascular bundles (Kouchi et al.1993). Although these genes are tissue-specific, it is important to consider their expression levels in order to efficiently label the cell-type specific ribosomes to study the gene expression profiles.

4.2 Tagged ribosomes affinity purification

Initially anti-FLAG agarose beads were used to pull down the FLAG-tagged ribosomal complexes (Zanetti et al. 2005). In 2016, the protocol was updated with α-FLAG M2 coupled Protein G Dynabeads as the use of Protein G Dynabeads was more efficient to pull down the FLAG-tagged ribosomal complexes (Reynoso et al .2016). Although the use of anti-FLAG beads to pull down FLAG-tagged ribosomal

complexes were well adapted in Arabidopsis (Mustroph et al. 2009), Medicago (Reynoso, 2015) and Tomato (Ron, 2014) plants, it did not yield tagged ribosomes from soybean roots. The concentration of FLAG3 peptide used for eluting the FLAG-tagged ribosomes (@ 200 ng/mL) as described in the protocol may not be enough to elute the FLAG-tagged ribosomes from soybean roots. Affinity purification using different concentrations of FLAG3 peptide can be done to find out the optimum concentration of FLAG3 peptide required to isolate the FLAG-tagged ribosomal complexes from soybean roots. However, as the TRAP construct has both 6xHis and FLAG epitope, we successfully purified tagged ribosomes using affinity purification of His-tagged fusion proteins.

Actively translating mRNAs from these nodule zones were isolated using the TRAP method. Ribosomal protein L18 (RPL18) was detected in clarified extract and in the final elution. No band was detected on flow through suggesting that TRAP approach was an efficient way to isolate ribosomal complexes. Different patterns of RNA migration of TRAP elutions were observed on agarose gel. Similar RNA migration pattern was observed when RNA isolated from non-transgenic matured nodules was diluted in the same elution buffer (Appendix O), suggesting that the presence of NaCl in elution buffer and their differential recovery at subsequent elutions might have affected the migration pattern of RNA. Western blotting and agarose gel image confirmed that the TRAP method was effective for isolation of tagged ribosomes, while qPCR assays of marker genes confirmed that we were able to isolate nodule zone specific ribosomal complexes from transgenic soybean roots.

A non- coding RNA (GLYMA19G06330) was detected in all ENOD2p-, ENOD40p-, and CsVMVp-TRAP derived RNA samples, but at very low level. Although the gene is characterized as non-coding, it may be possible that some part of the gene is still loaded in the ribosomes for translation. However, the complete absence of precursor miRNA 160 and miRNA 166 in TRAP samples confirmed that TRAP samples are free of contamination from total or nuclear RNA.

4.3 mRNA enrichment

Two different approaches, mRNA isolation and rRNA depletion were used for mRNA enrichment from ribosomal preparations. qPCR assays for Actin and Cons7 house-keeping genes suggested that rRNA depletion was found to be more efficient than mRNA isolation for cDNA synthesis. In rRNA depletion approach, the capture of the rRNA does not affect the mRNA sequences in the original RNA preparation. This could result in the more quantitative recovery of mRNA from rRNA depletion than from mRNA isolation approach, thereby providing a higher amount of starting mRNA for cDNA synthesis for qPCR. The presence of a higher amount of rRNA in total RNA could have interfered during cDNA synthesis. rRNA removal step might have also removed the cDNA synthesis inhibitors present in total RNA. This observation suggested that rRNA removed RNA samples are more suited for cDNA synthesis for qPCR and potentially library construction using a similar approach.

4.4 Library Synthesis

The library was prepared using the method described earlier and qualitative check of the prepared library was performed using bioanalyzer. Library having fragments size of 200-300 bp was obtained in the positive sample. However,

synthesizing the library from TRAP samples was not possible. This might be because of the low quality and quantity of RNA. Because of the higher input requirement of Bioanalyzer, the quality of the purified RNA was not assessed using this method. Although we confirmed the quality and quantity of total RNA through nanodrop, it may not be reliable because of its lack of accuracy with low amounts of sample. In addition, Nanodrop measurement does not take into account the changes in RNA integrity e.g. if RNA samples are degraded, because single nucleotides also will contribute to absorbance at 260nm. Many other methods are available for quantification and analysis of RNA samples, such as Fluorescent Dye-Based Quantification. Dye-based methods such as the QuantiFluor™ RNA System requires less template RNA than other systems. Although this method is more sensitive than absorbance methods for low-concentration samples, this also fails to provide the purity and integrity information (Wieczorek, 2012). Therefore, checking the quality of total RNA using Bioanalyzer would give us more confidence for the downstream experiments. Additionally, we were able to perform the qPCR with the RNA. For qPCR, cDNA was prepared via reverse transcription using oligo (dT) primers. So, replacing the random priming with oligo-dT priming during library preparation could possibly provide a good library. For, oligo-dT based library construction, poly-adenylated RNA can be reverse transcribed with an anchored oligo-dT primer carrying a universal primer sequence at its 5′ end. Then poly-nucleotide tailing can be used to add a poly (A) tail to the 3′ end of the cDNA. This cDNA is amplified with universal PCR primers containing an oligo-dT sequence at the 3′ end.

Amplified cDNA can now be used in a standard DNA library construction protocol (Head et al. 2014).

4.5 Expression of auxin-related genes

Abundance of auxin- responsive GH3 family protein (GLYMA05G21680) mRNA was significantly higher in nodule parenchyma compared to infection zone. This expression pattern was similar to that of the soybean GH3 promoter in Lotus nodules (Takanashi et al. 2011), where they showed that the gene was expressed at the nodule cortex in a young nodule, and at nodule vascular bundles in a matured nodule, suggesting distinct auxin involvement in the determinate nodule development. In an experiment conducted by (Damodaran et al. 2017), it was observed that suppression of GH3 protein activity led to alterations in nodule number and nodule size in soybean. These observations suggested that GH3 play important role in soybean nodule development likely via their effect on auxin homeostasis. mRNA of auxin efflux carrier family protein (GLYMA01G36190) was significantly enriched in nodule parenchyma compared to infection zone at both time points. PILS7 gene (At5g65980), the orthologue of GLYMA01G36190, belongs to the putative auxin transport facilitator family, called PIN-LIKES (PILS) and was transcriptionally upregulated by auxin application in wild-type seedlings (Barbez, 2012). Similarly, the abundance of GmARF5 (GLYMA17G37580) mRNA was significantly enriched in nodule parenchyma compared to infection zone at 7dpi. Arabidopsis ARF5, the ortholog of GmARF5 (GLYMA17G37580) was found to act both cell-autonomously and noncell-autonomously to control embryonic vascular tissue formation and root initiation (Möller, 2017). Therefore, enrichment of this

gene in parenchyma may be associated with nodule vascular development. Indeed, this is consistent with relatively lower auxin activity in nodule central tissues vs. the nodule vasculature (Turner et al. 2013). However, at 10 dpi, the gene was significantly enriched in infection zone vs. parenchyma. This is an unexpected expression pattern and needs to be studied further to have a clear understanding of its role during soybean nodule development.

4.6 Expression of transcription factor families

Abundance of Myb-related transcription factor (GLYMA03G42260.1) mRNA was significantly higher in infection zone compared to nodule parenchyma. This result was consistent with the research data demonstrated by (Duangkhet et al. 2016), where LjMYBr (the orthologue of GLYMA03G42260.1) and ENOD40 promoter-driven GUS showed the similar expression pattern: Ljmybr- or Enod40-promoter induced GUS expression in central tissues of emerging nodules of *Lotus japonicus*. The research data also emphasized the fact that the higher expression levels of ENOD40 genes were observed in MYB overexpressing nodules and lower in MYB RNAi-treated nodules. A BLAST search showed that LjMYBR is highly conserved to At5g56840 in Arabidopsis. These genes are members of the MYB-related CCA1 group and are involved in regulation of circadian rhythm and flower development in Arabidopsis, maize and soybean (Fujiwara et al. 2008; Schaffer et al. 1998). Similarly, the other transcription factor, bHLH (GLYMA08G04661.1) was highly enriched in infection zone at 10 dpi. AT1G73830, the ortholog of GLYMA08G04661.1 in Arabidopsis, encodes the brassinosteroid signaling component BEE3 (BR-ENHANCED EXPRESSION 3) and is known to positively

modulates the shade avoidance syndrome in Arabidopsis seedlings. Further study on the function of these transcription factors and their potential targets would help to understand the mechanism behind their tissue-specific expression in soybean nodule.

4.7 Nitrate and phosphate related genes

The abundance of nitrate transporter 1 (GLYMA11G04500.2) mRNA was significantly higher in infection zone at 7 dpi and 10 dpi. This is consistent with the research done by (Criscuolo et al. 2012), where it was demonstrated that, in *L. japonicus,* CM0826.370, a member of the nitrate transporter (NRT1) family, the ortholog of GLYMA11G04500.2, was localized in nodule primordia and in infection zone where ENOD40 promoter was expressed, and was not detected in the inner cortex and vascular bundle zones (Takanashi et al., 2012). These observations suggested that nitrate transporter 1 (GLYMA11G04500.2) might be specifically associated with nitrogen fixation in soybean. The abundance of phosphate transporter 1 (GLYMA10G00720) was significantly higher in nodule parenchyma compared to infection zone at 10dpi. PHO1 gene in Arabidopsis (At3g23430), the orthologue of soybean phosphate transporter 1 (GLYMA10G00720) was found to be highly expressed in the vascular cylinder of roots and was involved in Pi loading to the xylem. (Stefanovic, 2007). This suggested that the soybean phosphate transporter 1 (GLYMA10G00720) which was enriched in nodule parenchyma might be involved in transporting Pi from roots to nodules. The abundance of *PHOSPHATE 2* gene (GLYMA13G24810.1) encoding ubiquitin-conjugating enzyme 24, was significantly enriched in nodule parenchyma at 7 dpi, while it was significantly enriched in infection zone at 10dpi. In an experiment conducted by (Liu, 2012), it

was found that ubiquitin conjugase activity of AtPHO2, the orthologue of soybean phosphate 2 gene (GLYMA13G24810.) is required for PHO1 (the orthologue of soybean phosphate transporter 1) degradation to maintain Pi homeostasis in plants. At 10 dpi, phosphate transporter 1 gene was expressed at low level in infection zone whereas phosphate 2 was highly expressed in infection zone. While this is consistent with negative regulation of PHO1 by PHO2, no evidence exists for regulation of PHO1 at the transcript level by PHO2.

5 CONCLUSION

TRAP method was used to evaluate a selected set of genes for enrichment in specific nodule zones: nodule parenchyma and infection zone at two different time points during soybean nodule development: 7 and 10 dpi. The expected expression pattern of tissue-specific marker genes at 7 and 10 dpi was obtained validating the suitability of our system and methods to evaluate nodule zone-specific gene expression profiles. Quality controls tests demonstrated efficient purification of tagged ribosomes, absence of or minimal contamination by total or nuclear RNAs, and expected enrichment of nodule zone-specific marker genes suggesting that our adapted method was optimized for use in soybean composite transgenic plant system.

Three transcription factors, a MyB-related transcription factor (GLYMA03G42260.1), a bZIP transcription factor (GLYMA19G43420.1), and a bHLH transcription factor (GLYMA08G04661.1) were significantly enriched in infection zone at 10 dpi, suggesting that these genes might be involved in nitrogen fixation process. Similarly, nitrate transporter (GLYMA11G04500.2) was significantly enriched in infection zone at 7 dpi and 10 dpi suggesting that the gene might be involved in transporting nitrogen from nodule to shoot. A phosphate transporter 1 (GLYMA10G00720) which was enriched in nodule parenchyma might be involved in transporting Pi from roots to nodules. GmARF5 (GLYMA17G37580) was significantly enriched in nodule parenchyma at 7 dpi and was significantly enriched in infection zone at 10 dpi. Change in expression of this gene from the nodule parenchyma to infection zone suggest that the gene might be involved in

tissue differentiation and patterning during nodule development and highlights the importance of plant hormones in nodule tissue specification.

The gene expression analysis helped in the identification of genes that were differentially expressed in nodule parenchyma and infection zone. Differential expression of transcription factors suggested that different signaling components might be involved in providing the distinct identities to nodule parenchyma and infection zone. Detail study on transcription factors and their targets is necessary to have a clear idea about their regulatory mechanism during nodule development and maturation. Similarly, the differential expression of auxin highlights the potential importance of this hormones in nodule tissue specification.

Although we determine the spatiotemporal expression of many genes, global transcriptomic analysis and evaluation of gene function in nodule zone differentiation and/or function is necessary to gain crucial mechanistic insights. The knowledge can be used to alter nodule numbers and/or maturity to enhance nitrogen fixation in soybean and other legume nodules.

REFERENCES

Abdel-Lateif, Khalid, Bogusz, Didier, Hocher, Valérie. The role of flavonoids in the establishment of plant roots endosymbioses with arbuscular mycorrhiza fungi, rhizobia and *Frankia* bacteria. *J Plant signaling, & behavior*. (2012), 7(6), 636-641.

Adhikari, Sajag. Transcriptional and post-transcriptional regulation of nodule-specific gene expression in soybean. *Unpublished doctoral dissertation.* (2016). South Dakota State University, Brookings, SD

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

Beringer, JE, Brewin, NJ, Johnston, AWB, Schulman, HM, & Hopwood, DA. The Rhizobium-legume symbiosis. *J Proceedings of the Royal Society of London. Series B. Biological Sciences*. (1979), 204(1155), 219-233.

Bond, Lora. Responses of pea roots to application of certain growth-regulating substances. *J Botanical Gazette.* (1948),109(4), 435-447.

Canfield, Donald E, Glazer, Alexander N, & Falkowski, Paul G. The evolution and future of Earth's nitrogen cycle. *J science.* (2010), 330(6001), 192-196.

Capoen, Ward, Goormachtig, Sofie, De Rycke, Riet, Schroeyers, Katrien, & Holsters, Marcelle. SrSymRK, a plant receptor essential for symbiosome formation. *J Proceedings of the National Academy of Sciences*. (2005), 102(29), 10369-10374.

Cerri, Marion R, Frances, Lisa, Laloum, Tom, Auriac, Marie-Christine, Niebel, Andreas, Oldroyd, Giles ED,de Carvalho-Niebel, Fernanda. *Medicago truncatula* ERN transcription factors: regulatory interplay with NSP1/NSP2 GRAS factors and expression dynamics throughout rhizobial infection. *J Plant physiology.* (2012), 160(4), 2155-2172.

Colebatch, Gillian, Desbrosses, Guilhem, Ott, Thomas, Krusell, Lene, Montanari, Ombretta, Kloska, Sebastian, Udvardi, Michael K. Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus*. *J The Plant Journal.* (2004), 39(4), 487-512.

Collier, Ray, Fuchs, Beth, Walter, Nathalie, Kevin Lutke, William, & Taylor, Christopher G. Ex vitro composite plants: an inexpensive, rapid method for root biology. *J The Plant Journal*. (2005), 43(3), 449-457.

Combier, Jean-Philippe, Frugier, Florian, De Billy, Françoise, Boualem, Adnane, El-Yahyaoui, Fikri, Moreau, Sandra. MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*. *Development.* (2006), 20(22), 3084-3088.

Cooper, James E. Multiple responses of rhizobia to flavonoids during legume root infection. *In Advances in Botanical Research.* (2004), (Vol. 41, pp. 1-62).

Criscuolo, Giuseppina, Valkov, Vladimir Totev, Parlati, Aurora, Alves, Ludovico Martin, Chiurazzi, Maurizio. Molecular characterization of the *Lotus japonicus* NRT1 (PTR) and NRT2 families. *J Plant, cell, & environment.* (2012), 35(9), 1567-1581.

Damodaran, Suresh, Westfall, Corey, Kisely, Brian, Jez, Joseph, & Subramanian, Senthil. Nodule-enriched GRETCHEN HAGEN 3 enzymes have distinct substrate specificities and are important for proper soybean nodule development. *J International journal of molecular sciences.* (2017), 18(12), 2547.

Damodaran S., Adhikari S., Lin J., Pathak S., Ge X., Blakeslee L., Subramanian S. Comparative transcriptomics reveals a role for GmCYP83B1 in regulating auxin homeostasis during soybean nodule development. *J Plant Physiol.* (2017), (submitted).

de Bruijn, Frans J. "Biological Nitrogen Fixation*". J Advances in Microbiology.* (2016), Book Summary, 6(06), 407.

Deal, Roger B, & Henikoff, Steven. The INTACT method for cell type–specific gene expression and chromatin profiling in *Arabidopsis thaliana*. *J Nature protocols.* (2011), 6(1), 56.

Diédhiou, Issa, Tromas, Alexandre, Cissoko, Maïmouna, Gray, Krystelle, Parizot, Boris, Crabos, Amandine Svistoonoff, Sergio. Identification of potential transcriptional regulators of actinorhizal symbioses in *Casuarina glauca* and *Alnus glutinosa*. *J BMC plant biology*. (2014), 14(1), 342.

Duangkhet, Mallika, Thepsukhon, Apiraya, Widyastuti, Rahayu, Santosa, Dwi Andreas, Tajima, Shigeyuki, & Nomura, Mika. A MYB-related transcription factor affects nodule formation in *Lotus japonicus*. *J Plant Biotechnology*. (2016), 16.0905 a.

Ferguson, Brett J, Indrasumunar, Arief, Hayashi, Satomi, Lin, Meng-Han, Lin, Yu-Hsiang, Reid, Dugald E, & Gresshoff, Peter M. Molecular analysis of legume nodule development and autoregulation. *J Journal of integrative plant biology*. (2010), 52(1), 61-76.

Fisher, Jon, Gaillard, Paul, Fellbaum, Carl R, Subramanian, Senthil, Smith, SteveQuantitative 3D imaging of cell level auxin and cytokinin response ratios in soybean roots and nodules. *J Plant, cell, & environment.* (2018), 41(9), 2080-2092.

Follett, Ronald F, & Hatfield, Jerry L. Nitrogen in the environment: sources, problems, and management. *J The Scientific World Journal*. (2001), 1, 920-926.

Franssen, Henk J, Vijn, Irma, Yang, Wei Cai, & Bisseling, Ton. Developmental aspects of the Rhizobium-legume symbiosis. *J Plant molecular biology*. (1992). 19(1), 89-107.

Gage, Daniel J. Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *J Microbiol. Mol. Biol. Rev.* (2004), *68(2), 280-300*.

Gage, Daniel J. Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *J Microbiology, & Reviews, Molecular Biology*. (2004), 68(2), 280-300.

Glick, Bernard R. The enhancement of plant growth by free-living bacteria. *J Canadian journal of microbiology*. (1995), 41(2), 109-117.

Hartmann, Klaus, Peiter, Edgar, Koch, Kerstin, Schubert, Sven, & Schreiber, Lukas. Chemical composition and ultrastructure of broad bean (Vicia faba L.) nodule endodermis in comparison to the root endodermis. *J Planta.* (2002), 215(1), 14-25.

Head, Steven R, Komori, H Kiyomi, LaMere, Sarah A, Whisenant, Thomas, Van Nieuwerburgh, Filip, Salomon, Daniel R, & Ordoukhanian, Phillip. Library construction for next-generation sequencing: overviews and challenges. *J Biotechniques.* (2014), 56(2), 61-77.

Heiman, Myriam, Kulicke, Ruth, Fenster, Robert J, Greengard, Paul, & Heintz, Nathaniel Cell type–specific mRNA purification by translating ribosome affinity purification (TRAP). *J Nature protocols.* (2014), *9(6), 1282*.

Held, Mark, Hou, Hongwei, Miri, Mandana, Huynh, Christian, Ross, Loretta, Hossain, Md Shakhawat, Wang, Trevor L. *Lotus japonicus* cytokinin receptors work partially redundantly to mediate nodule formation. *J The Plant Cell.* (2014), 26(2), 678-694.

Hirsch, Ann M. Developmental biology of legume nodulation. *J New Phytologist.* (1992), 122(2), 211-237.

Huntley, ME, Mosier, AR, Smith, KA, Winiwarter, W. N2O release from agro-biofuel production negates global warming reduction by replacing fossil fuel. *J Atmosphere Chemist, & Physic.* (2007), 12, 573-608.

Indrasumunar, Arief, Kereszt, Attila, Searle, Iain, Miyagi, Mikiko, Li, Dongxue, Nguyen, Cuc DT. Inactivation of duplicated nod factor receptor 5 (NFR5) genes in recessive loss-of-function non-nodulation mutants of allotetraploid soybean (*Glycine max* L. Merr.). *J Cell.* (2009), 51(2), 201-214.

James, EK. Nitrogen fixation in endophytic and associative symbiosis. *J Field crops research.* (2000), 65(2-3), 197-209.

Kavroulakis, Nektarios, Flemetakis, Emanouil, Aivalakis, Georgios, & Katinakis, Panagiotis. Carbon metabolism in developing soybean root nodules: the role of carbonic anhydrase. *J Molecular plant-microbe interactions*. (2000), 13(1), 14-22.

Kumar S Nitrogen Fixation Types: Physical and Biological Nitrogen Fixation (With Diagram) http://www.biologydiscussion.com/nitrogen-fixation/types-nitrogen-fixation/nitrogenfixation-types-physical-and-biological-nitrogen-fixation-with-diagram/14969.

Lee, Ji-Young, Colinas, Juliette, Wang, Jean Y, Mace, Daniel, Ohler, Uwe, & Benfey, Philip N. Transcriptional and posttranscriptional regulation of transcription factor expression in Arabidopsis roots. *J Proceedings of the National Academy of Sciences.* (2006), 103(15), 6055-6060.

Lerouge, Patrice, Roche, Philippe, Faucher, Catherine, Maillet, Fabienne, Truchet, Georges, Promé, Jean Claude, & Dénarié, Jean. Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated and acylated glucosamine oligosaccharide signal. *J Nature*. (1990), 344(6268), 781.

Libault, Marc, Farmer, Andrew, Joshi, Trupti, Takahashi, Kaori, Langley, Raymond J, Franklin, Levi D,Stacey, Gary. An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. *J The Plant Journal.* (2010), 63(1), 86-99.

Limpens, Erik, Franken, Carolien, Smit, Patrick, Willemse, Joost, Bisseling, Ton, & Geurts, René. LysM domain receptor kinases regulating rhizobial Nod factor-induced infection. *J Science.* (2003), 302(5645), 630-633.

Limpens, Erik, Moling, Sjef, Hooiveld, Guido, Pereira, Patrícia A, Bisseling, Ton, Becker, Jörg D, & Küster, Helge. Cell-and tissue-specific transcriptome analyses of *Medicago truncatula* root nodules. *J PloS one*. (2013), 8(5), e64377.

Luo, Lin, Salunga, Ranelle C, Guo, Hongqing, Bittner, Anton, Joy, KC, Galindo, Jose E, Jackson, Michael R. Gene expression profiles of laser-captured adjacent neuronal subtypes. *J Nature medicine.* (1999), 5(1), 117.

Madsen, Esben Bjørn, Madsen, Lene Heegaard, Radutoiu, Simona, Olbryt, Magdalena, Rakwalska, Magdalena, Szczyglowski, Krzysztof,Sandal, Niels. A receptor kinase gene of the LysM type is involved in legumeperception of rhizobial signals. *J Nature.* (2003), 425(6958), 637.

Madsen, Lene H, Tirichine, Leïla, Jurkiewicz, Anna, Sullivan, John T, Heckmann, Anne B, Bek, Anita S, Stougaard, Jens. The molecular network governing nodule organogenesis and infection in the model legume *Lotus japonicus*. *J Nature communications.* (2010), 1, 10.

McAdams, Harley H, & Arkin, Adam. Stochastic mechanisms in gene expression. J *Proceedings of the National Academy of Sciences.* (1997), 94(3), 814-819.

McCauley, Ann, Jones, Clain, & Jacobsen, Jeff. Soil pH and organic matter. *J Nutrient management module.* (2009), 8, 1-12.

Mergaert, Peter, Van Montagu, Marc, & Holsters, Marcelle. Molecular mechanisms of Nod factor diversity. *J Molecular microbiology.* (1997), 25(5), 811-817.

Messinese, Elsa, Mun, Jeong-Hwan, Yeun, Li Huey, Jayaraman, Dhileepkumar, Rougé, Pierre, Barre, Annick,Cook, Douglas R. A novel nuclear protein interacts with the symbiotic DMI3 calcium-and calmodulin-dependent protein kinase of *Medicago truncatula*. *J Molecular plant-microbe interactions.* (2007), 20(8), 912-921.

Murray, Jeremy D, Karas, Bogumil J, Sato, Shusei, Tabata, Satoshi, Amyot, Lisa, & Szczyglowski, Krzysztof . A cytokinin perception mutant colonized by Rhizobium in the absence of nodule organogenesis. *J Science.* (2007)*,* 315(5808), 101-104.

Mustroph, Angelika, Zanetti, M Eugenia, Jang, Charles JH, Holtan, Hans E, Repetti, Peter P, Galbraith, David W, Bailey-Serres, Julia. Profiling translatomes of discrete cell populations resolves altered cellular priorities during hypoxia in Arabidopsis. *J Proceedings of the National Academy of Sciences.* (2009), 106(44), 18843-18848.

Newcomb, William. A correlated light and electron microscopic study of symbiotic growth and differentiation in *Pisum sativum* root nodules. *J Canadian Journal of Botany.* (1976), 54(18), 2163-2186.

Oldroyd, Giles ED, & Downie, J Allan. Coordinating nodule morphogenesis with rhizobial infection in legumes. *J Annu. Rev. Plant Biol.* (2008), 59, 519-546.

Oldroyd, Giles ED, & Long, Sharon R. Identification and characterization of nodulation-signaling pathway 2, a gene of *Medicago truncatula* involved in Nod factor signaling. *J Plant Physiology.* (2003), 131(3), 1027-1032.

Oldroyd, Giles ED, Murray, Jeremy D, Poole, Philip S, & Downie, J Allan. The rules of engagement in the legume-rhizobial symbiosis*. J Annual review of genetics.* (2011), 45, 119-144.

Patriquin, DG, Döbereiner, J, & Jain, DK. Sites and processes of association between diazotrophs and grasses. *J Canadian Journal of Microbiology.* (1983), 29(8), 900-915.

Pathak, S. Differential Gene Expression in Two Nodule Zones of Soybean (*Unpublished master's thesis*). (2017). South Dakota State University, Brookings, SD.

Peiter, Edgar, & Schubert, Sven. Sugar uptake and proton release by protoplasts from the infected zone of *Vicia faba* L. nodules: evidence against apoplastic sugar supply of infected cells. *J Journal of Experimental Botany.* (2003), 54(388), 1691-1700.

Peoples, MB, Brockwell, J, Herridge, DF, Rochester, IJ, Alves, BJR, Urquiaga, S, Maskey, SL. The contributions of nitrogen-fixing crop legumes to the productivity of agricultural systems. *J Symbiosis. (2009),* 48(1-3), 1-17.

Preiss, Thomas, Baron-Benhamou, Julie, Ansorge, Wilhelm, Hentze, Matthias W. Homodirectional changes in transcriptome composition and mRNA translation induced by rapamycin and heat shock. *J Nature Structural, & Biology, Molecular.* (2003), 10(12), 1039.

Radutoiu, Simona, Madsen, Lene Heegaard, Madsen, Esben Bjørn, Felle, Hubert H, Umehara, Yosuke, Grønlund, Mette, Sandal, Niels. Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *J Nature.* (2003), 425(6958), 585.

Redmond, John W, Batley, Michael, Djordjevic, Michael A, Innes, Roger W, Kuempel, Peter L, & Rolfe, Barry G. Flavones induce expression of nodulation genes in Rhizobium. *J Nature.* (1986), 323(6089), 632.

Reynoso, Mauricio Alberto, Blanco, Flavio Antonio, Bailey-Serres, Julia, Crespi, Martín, & Zanetti, María Eugenia. Selective recruitment of m RNAs and miRNA s to polyribosomes in response to rhizobia infection in *Medicago truncatula*. *J The plant journal. (2013),* 73(2), 289-301.

Rivera, Maria C, Maguire, Bruce, & Lake, James A. Isolation of ribosomes and polysomes. *J Cold Spring Harbor Protocols.* (2015), (3), pdb. prot081331.

Ron, Mily, Kajala, Kaisa, Pauluzzi, Germain, Wang, Dongxue, Reynoso, Mauricio A, Zumstein, Kristina,Inagaki, Soichi. Hairy root transformation using *Agrobacterium rhizogenes* as a tool for exploring cell type-specific gene expression and function using tomato as a model. *J Plant physiology.* (2014), 166(2), 455-469.

Roth, LE, & Stacey, G. Bacterium release into host cells of nitrogen-fixing soybean nodules: the symbiosome membrane comes from three sources. *J European journal of cell biology.* (1989), 49(1), 13-23.

Saikia, SP, & Jain, Vanita Biological nitrogen fixation with non-legumes: An achievable target or a dogma? *J Current science.* (2007), 317-322.

Salvagiotti, Fernando, Cassman, Kenneth G, Specht, James E, Walters, Daniel T, Weiss, Albert, & Dobermann, A. Nitrogen uptake, fixation and response to fertilizer N in soybeans: A review. *J Field Crops Research.* (2008), 108(1), 1-13.

Savci, Serpil. An agricultural pollutant: chemical fertilizer. *J International Journal of Environmental Science, & Development.* (2012), 3(1), 73.

Seniczak, Stanisław, Dabrowski, Janusz, Klimek, Andrzej, & Kaczmarek, Sławomir. Effects of air pollution produced by a nitrogen fertilizer factory on the mites (Acari) associated with young Scots pine forests in Poland. *J Applied Soil Ecology.* (1998), 9(1-3), 453-458.

Serafini, Tito, & Ngai, John. Method of defining cell types by probing comprehensive expression libraries with amplified RNA. *In: Google Patent.* (2000).

Shridhar, Bagali Shrimant. Nitrogen fixing microorganisms. *J Int J Microbiol Res.* (2012), 3(1), 46-52.

Smit, Patrick, Raedts, John, Portyanko, Vladimir, Debellé, Frédéric, Gough, Clare, Bisseling, Ton, & Geurts, René. NSP1 of the GRAS protein family is essential for rhizobial Nod factor-induced transcription. *J Science.* (2005), 308(5729), 1789-1791.

Spaink, Herman P. Root nodulation and infection factors produced by rhizobial bacteria. *J Annual Reviews in Microbiology.* (2000), 54(1), 257-288.

Steen, Anton, Buist, Girbe, Leenhouts, Kees J, El Khattabi, Mohamed, Grijpstra, Froukje, Zomer, Aldert L, Kok, Jan. Cell wall attachment of a widely distributed peptidoglycan binding domain is hindered by cell wall constituents. *J Journal of Biological Chemistry*. (2003), 278(26), 23874-23881.

Stracke, Silke, Kistner, Catherine, Yoshida, Satoko, Mulder, Lonneke, Sato, Shusei, Kaneko, Takakazu, Szczyglowski, Krzysztof. A plant receptor-like kinase required for both bacterial and fungal symbiosis. *J Nature.* (2002), 417(6892), 959.

Sutton, Mark A, Oenema, Oene, Erisman, Jan Willem, Leip, Adrian, van Grinsven, Hans, & Winiwarter, Wilfried. Too much of a good thing. 472(7342), 159. *J Nature.* (2011).

Tirichine, Leïla, Sandal, Niels, Madsen, Lene H, Radutoiu, Simona, Albrektsen, Anita S, Sato, Shusei, Stougaard, Jens. A gain-of-function mutation in a cytokinin receptor triggers spontaneous root nodule organogenesis. *J Science.* (2007), 315(5808), 104-107.

Tjepkema, JD, Schwintzer, CR, & Benson, DR. Physiology of actinorhizal nodules. *J Annual Review of Plant Physiology.* (1986), 37(1), 209-232.

Tjepkema, John D, & Yocum, CS %J Planta. (1974). Measurement of oxygen partial pressure within soybean nodules by oxygen microelectrodes. 119(4), 351-360.

Turgeon, B Gillian, & Bauer, Wolfgang DEarly events in the infection of soybean by *Rhizobium japonicum*. Time course and cytology of the initial infection process. . J *Canadian Journal of Botany.* (1982), 60(2), 152-161.

Turner, Marie, Nizampatnam, Narasimha Rao, Baron, Mathieu, Coppin, Stéphanie, Damodaran, Suresh, Adhikari, Sajag, Subramanian, Senthil. Ectopic expression of miR160 results in auxin hypersensitivity, cytokinin hyposensitivity, and inhibition of symbiotic nodule development in soybean. *J Plant physiology.* (2013), 162(4), 2042-2055.

van de Wiel, Clemens, Scheres, Ben, Franssen, Henk, van Lierop, Marie-José, Van Lammeren, A, Van Kammen, A, & Bisseling, T. The early nodulin transcript

ENOD2 is located in the nodule parenchyma (inner cortex) of pea and soybean root nodules. *J The EMBO Journal.* (1990), 9(1), 1-7.

Vanbleu, E, & Vanderleyden, Jozef.Molecular genetics of rhizosphere and plant-root colonization. In Associative and endophytic nitrogen-fixing bacteria and cyanobacterial associations. (2007), (pp. 85-112): Springer.

Vance, Carroll P. Symbiotic Nitrogen Fixation and Phosphorus Acquisition. Plant Nutrition in a World of Declining Renewable Resources. *J Plant Physiology (2011)*, 127(2), 390-397. doi:10.1104/pp.010331

Wagner, Stephen C. Biological Nitrogen Fixation. the nature EDUCATION Knowledge Project. (2011).

Wang, Dong, Yang, Shengming, Tang, Fang, & Zhu, Hongyan. Symbiosis specificity in the legume–rhizobial mutualism. *J Cellular microbiology.* (2012), 14(3), 334-342.

Westhoff, Patrick. The economics of biological nitrogen fixation in the global economy. *J Nitrogen fixation in crop production. Agronomy Monograph.* (2009), 52, 309-328.

Yang, Wei-Cai, Katinakis, Panagiotis, Hendriks, Peter, Smolders, Arie, de Vries, Floris, Spee, Johan, Franssen, Henk. Characterization of GmENOD40, a gene showing novel patterns of cell-specific expression during soybean nodule development. *J The Plant Journal.* (1993), 3(4), 573-585.

Yano, Koji, Shibata, Satoshi, Chen, Wen-Li, Sato, Shusei, Kaneko, Takakazu, Jurkiewicz, Anna, Kojima, Tomoko. 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. *J The Plant Journal.* (2009), 60(1), 168-180.

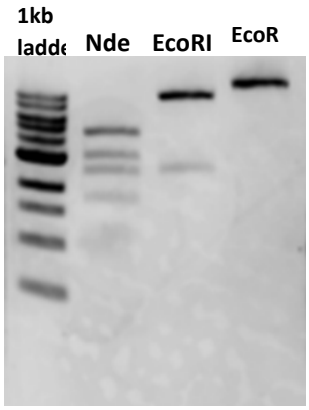
Yao, Fayi, Yu, Fei, Gong, Lijie, Taube, David, Rao, Donald D, & MacKenzie, Robert G. Microarray analysis of fluoro-gold labeled rat dopamine neurons harvested by laser capture microdissection. *J Journal of neuroscience methods.* (2005), 143(2), 95-106.

Zanetti, María Eugenia, Chang, Feng, Gong, Fangcheng, Galbraith, David W, & Bailey-Serres, Julia. Immunopurification of polyribosomal complexes of Arabidopsis for global analysis of gene expression. *J Plant physiology.* (2005), 138(2), 624-635.

Zhu, Hui, Chen, Tao, Zhu, Maosheng, Fang, Qing, Kang, Heng, Hong, Zonglie, & Zhang, Zhongming. A novel ARID DNA-binding protein interacts with SymRK and is expressed during early nodule development in *Lotus japonicus*. *J Plant Physiology.* (2008), 148(1), 337-347.

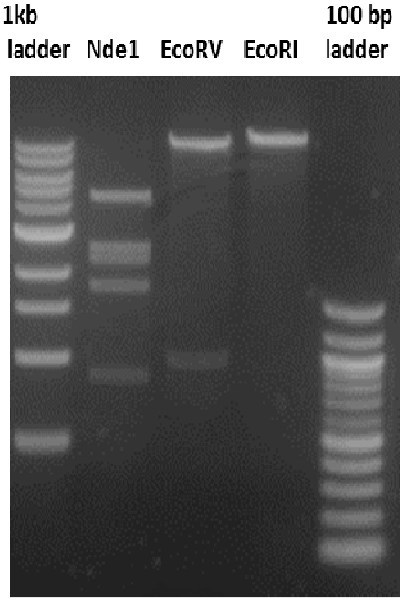
APPENDICES

Appendix A: Verification of destination vector



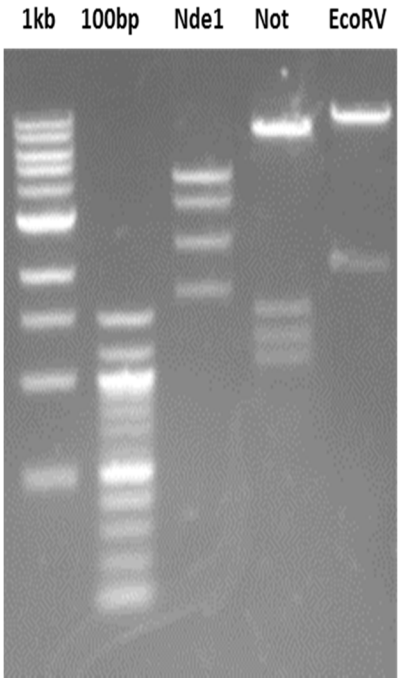
Appendix A: Verification of destination vector. The gel shows banding pattern of destination vector ***(PK7WG-TRAP)*** *after restriction digestion using three enzymes (lanes labeled as 'Nde1', 'ECOR1' and 'EcoRV'). The left most lane is 1kb ladder. Plasmid replicate=1.*

Appendix B: Verification of expression vector



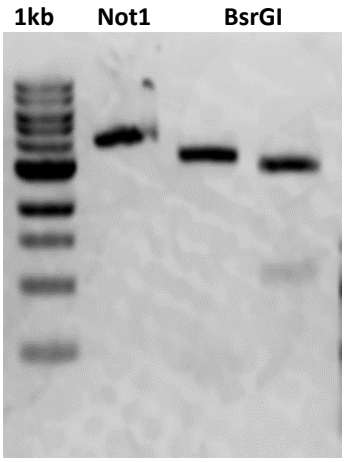
Appendix B: Verification of expression vector. The gel shows banding pattern of expression vector ***(PK7WG-GmENOD40-TRAP)*** *after the restriction digestion using three enzymes (labeled as 'Nde1', 'ECORV' and 'EcoR1'). The left most lane is 1kb and the right most lane is100bp ladder. Plasmid replicate=1.*

Appendix C: Verification of expression vector



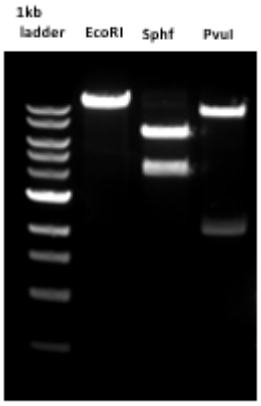
Appendix C: Verification of expression vector. The gel shows banding pattern of expression vector ***(PK7WG-GmENOD2-TRAP)*** *after the restriction digestion using three enzymes (labeled as 'Nde1', 'Not' and 'EcoRV'). The two most left lanes are 1kband 100bp ladder. Plasmid replicate=1.*

Appendix D: Verification of entry vector



Appendix D: Verification of entry vector. Gel image showing banding pattern of entry clone ***(PMH40-CsvMV)*** *using different restriction digestion enzymes (labeled as 'NotI', 'BSRGI' and 'PvuI). The left most lane is 1kb ladder. Plasmid replicate=1*

Appendix E: Verification of expression vector



Appendix E: Verification of expression vector. Gel image showing banding pattern of expression vector ***(PK7WG-CsVMV-TRAP)*** *using different restriction digestion enzymes (labeled as 'EcoRI', 'SphI' and 'PvuI). The left most lane is 1kb ladder. Plasmid replicate=1*

Appendix F: Composition of Hoagland solution

Table 1 Composition of Hoagland solution

Concentration for stock solution (1L)

	Component	**Mol.wt**	**Molarity (mM)**	**Amount (gm)**
Solution -I	$Ca(NO_3)_2.4H_2O$	236.15	892.7614	210.8253
Solution -II	$MgSO_4.7H_2O$	246.5	500	123.25
Solution-III	KNO_3	101.1032	1250	126.37
	KH_2PO_4	174.2	200	34.85
Soultion-IV	$Na_2FeEDTA$	372.24	11.5	4.28
Soultion-V	$MnCl_2$	125.84	3.6	0.453
	$ZnSO_4$	161.47	0.34	0.054
	H_3BO_3	61.83	11.5	0.711
	$CuSO_4$	159.6	0.125	0.0195
	H_2MoO_4	85%		0.085

Concentration for final solution (1L)

	Volume used (ml)	Final molarity(mM)
Solution -I	5.6	5
Solution -II	4	2
Solution-III	4	5
Soultion-IV	8	0.092
Soultion-V	4	1x

Appendix G: Composition of nitrogen free plant nutrient solution

Table 2 Composition of nitrogen free plant nutrient solution.

A. Macronutrient stocks:

Stock	**Stock vol**	**Amount(gm)**	**ml Stock/liter PNS**
$MgSO_4.7H_2O$	200ml	(12.3g)	2
$CaCl_2.2H_2O$	400ml	(29.4g)	4
$K_2HPO_4.3H_2O$	100ml	(3.4g)	1
K_2SO_4	400ml	(22.0g)	4
$FeCl_3.6H_2O$	250ml	(0.62g)	2.5

B. Micronutrients (10000x)

Stock	gm per 1 liter
H_3BO_3	1.42
$MnSO_4.\ H_2O$	0.77
$ZnSO_4.7H_2O$	1.73
$CuSO_4.5H_2O$	0.37
$NaMoO_4.2H_2O$	0.24
$CoCl_2.6H_2O$	0.025
$NiSO_4$	0.01

Appendix H: Composition of Vincent - rich media

Table 3 Composition of vincent - rich media

Chemical	Amount/Liter
K_2HPO_4	0.5 g
NaCl	0.1 g
$MgSO_4.7H_2O$	0.2 g
Yeast Extract	0.4 g
Mannitol	10.0 g, PH =6.8

Appendix I: qPCR primer design parameters

Table 4 qPCR primer design parameters

Parameters	**Criteria**
Primer length (nucleotides):	18-24
Product length (nucleotides):	50-150
Primer melting temperature (Celcius):	59-64
Max primer melting temperature difference:	3
GC content (%):	40-60
GC content in 3' tail (%):	40-60
Max base repeat:	3
Check for primer self-hybridization:	Yes
Check for primer cross hybridization:	Yes

Appendix J: List of qPCR primers used in this study

Table 5 qPCR primers used in this study

Primer Name	Primer Sequence (5' – 3')	Efficiency (%)
qP_Glyma19g35270.1 Fw	CTGTTGGCTCTGGCTGGAAGAC	98
qP_Glyma19g35270.1 RV	CTGACATAAGCAGCAGTTCGTTGG	
qP_Glyma19g43420.1 Fw	AGTCAGCCAGACGCTCAAGAAGG	98
qP_Glyma19g43420.1 Rv	CCGTGAATCGCTTTAACAACGAAG	
qP_Glyma19g34881.1 FW	CAGAGGTGGCTAAGCTGAATGCG	97
qP_Glyma19g34881.1 RV	TGCGGTCCCGGAAACTGAAGAAG	
qP_Glyma19g30640.1 Fw	AGGTTTCCCTCAACCGTGGGATG	102
qP_Glyma19g30640.1 RV	CTTGGGCCTCACTTTCCTCTCTAG	
qP_Glyma17g37580.1 Fw	CTGGAGGAGTTGGAGGAGGAATGG	98.5
qP_Glyma17g37580.1 Rv	CAAGAGGGCCTGCACAAGCATG	
qP_Glyma17g08770.1 Fw	CAATGCAGGTGATGCCTCTACCTC	103.7
qP_Glyma17g08770.1 Rv	TGGGATGCTGACGGTTGTTGTCG	
qP_Glyma16g23800.2 Fw	GAGAAGCATTGGCTAAGCATGAGG	101.3
qP_Glyma16g23800.2 RV	CGCAACAGGCAAAGGAGCATG	
qP_Glyma13g24810.1 Fw	TCCGTACTATCCAGGGCAGAGAG	93
qP_Glyma13g24810.1 Rv	ACCAAGCCTGCTTCCACAGCAC	
qP_Glyma12g09670.1 Fw	AGCCATCCATGTCCCTTCCCAATC	85
qP_Glyma12g09670.1 Rv	CGCGACTCTGCTCCTTCTTTCTC	
qP_Glyma11g04500.2 Fw	GGATCGGCGTGAACCTAGTGTTG	89
qP_Glyma11g04500.2 Rv	TGCTCACATTGTTGGCTGCATCAG	
qP_Glyma10g00720.1 Fw	AGCTGGATGGAGAGCTTAACAAGG	104.9
qP_Glyma10g00720.1 Rv	TTGCGTCGGTCACTAAGAATTTGC	
qP_Glyma08g04661.1 Fw	GATGCCATTCTCCTGTGAAAGC	99.79
qP_Glyma08g04661.1 Rv	CCTGGAAATTCAGCTTCTGGAG	
qP_Glyma07g31630.1 Fw	CTGTGGAAGCAGGCTTGGTGTATG	100
qP_Glyma07g31630.1 Rv	TTGCCAGCATTTGGGAGCACTC	
qP_Glyma05g31410.1 Fw	TAGGCGGAAGAAGATGTCAAGAGC	97
qP_Glyma05g31410.1 Rv	CGCAAGCTGTCAGAAGAACCAGTC	
qP_Glyma05g21680.1 FW	TCCATCGCATTGCTAATGGTGACC	101.4
qP_Glyma05g21680.1 RV	CGACGGTCCATCTCTTGACGAATG	
qP_GLYMA03G42260.1 Fw	TTGCCTGTCCTCCTGCATCAGC	102.3
qP_GLYMA03G42260.1 Rv	TTGAGCTTCCGAGTGTTCTGGATC	
qP_Glyma19g06330.1 Fw	TCAGATCTGCCCTCTGTCCT	104.7
qP_Glyma19g06330.1 RV	TCAATACCAGCTTTTCCCTATGTTG	
qP_Glyma01g36190.1 Fw	AGGAGACAATGCTCCACTGC	93
qP_Glyma01g36190.1 RV	ACCACCAAGCAAAAGGGTGA	
qP_miRNA 160 Fw	ATGCTTGGCTCCTCATACGC	93.9
qP_miRNA 160 Rv	TATGTGCCTGGCTCCCTGTA	
qP_miRNA 166 Fw	GGGAATGAAGCCTGGTCCG	91.8
qP_miRNA 166 Rv	GAGGGGAATGTTGTCTGGCT	

Appendix K: Log_2 fold change (ENOD40p-/ENOD2p-TRAp derived samples) at 7dpi

Table 6 Log_2 fold change (ENOD40p-/ENOD2p-TRAp derived samples) value at 7 dpi

Gene ID	Gene Name / Annotation	Log2 fold change value	p-value
GLYMA08G14023	ENOD2	-10.25	0.00046
GLYMA02G04180	ENOD40	4.80	0.0373
GLYMA01G17330	Cyp83B1	-4.68	0.00455
GLYMA01G36190.1	Auxin efflux carrier family protein	-9.28	0.000344
GLYMA19G35270	ABC transporter G family member 34-related	-8.90	37e-05
GLYMA17G37580.1	ARF5	-4.36	0.0117
GLYMA19G34881	PROTEIN CUP-SHAPED COTYLEDON 3	-4.29	0.000517
GLYMA05G21680	Auxin-responsive GH3 family protein	-2.59	0.00465
GLYMA05G31410	ETHYLENE INSENSITIVE 3 FAMILY PROTEIN	-2.30	0.512
GLYMA07G31630.1	PHOSPHATE 2	-2.14	0.00807
GLYMA19G30640	Auxin-induced protein 5NG4-like	-1.41	0.755
GLYMA08G04661.1	bHLH	0.99	0.258
GLYMA17G08770	LONGIFOLIA PROTEIN	1.27	0.285
GLYMA16G23800.2	Disease resistance protein (TIR-NBS-LRR Class), putative	1.71	0.302
GLYMA10G00720.1	Phosphate transporter 1	1.97	0.0867
GLYMA19G43420.1	BZIP	2.05	0.172
GLYMA11G04500.2	Nitrate transporter	2.78	0.0219
GLYMA03G42260.4	MYB homeodomain-like superfamily protein	3.14	0.179
GLYMA13G24810.1	PHOSPHATE 2	3.56	0.011
GLYMA12G09670	NAC DOMAIN CONTAINING PROTEIN 90	5.93	0.0138

Appendix L: Log_2 fold change (ENOD40p-/ENOD2p-TRAP derived samples) at 10dpi

Table 7 Log_2 fold change (ENOD40p-/ENOD2p-TRAp derived samples) value at 10 dpi

Gene ID	Gene Name / Annotation	Log2 fold change value	p-value
GLYMA08G14023	ENOD2	-13.83	3e-07
GLYMA02G04180	ENOD40	8.41	4.85e-05
GLYMA01G17330	CYP	-5.63	0.000262
GLYMA19G35270	ABC transporter G family member 34-related	-9.19	0.000145
GLYMA05G31410	ETHYLENE INSENSITIVE 3 FAMILY PROTEIN	-7.38	0.0322
GLYMA19G30640	Auxin-induced protein 5NG4-like	-7.19	0.00706
GLYMA10G00720.1	Phosphate transporter 1	-6.56	0.00539
GLYMA16G23800.2	Disease resistance protein (TIR-NBS-LRR class), putative	-4.55	0.0282
GLYMA05G21680	Auxin-responsive GH3 family protein	-4.20	0.0311
GLYMA13G24810.1	PHOSPHATE 2	-4.07	0.0555
GLYMA01G36190.1	Auxin efflux carrier family protein	-3.59	0.000316
GLYMA17G08770	LONGIFOLIA protein	-1.058	0.493
GLYMA19G43420.1	bZIP	3.57	0.0395
GLYMA12G09670	NAC domain containing protein 90	4.36	0.00283
GLYMA08G04661.1	bHLH	4.40	0.00263
GLYMA07G31630.1	PHOSPHATE 2	4.44	0.224
GLYMA17G37580.1	ARF5	5.34	0.00555
GLYMA11G04500.2	Nitrate transporter	5.38	0.00454
GLYMA19G34881	CUP-SHAPED COTYLEDON 3	5.41	0.00287
GLYMA03G42260.4	MYB homeodomain-like superfamily protein	6.33	0.00231

Appendix M: List of cloning primers used in this study

Table 4 List of cloning primers used in this study

Primer Name	**Primer Sequence (5' – 3')**
pCsVMV-Fw (NcoI)	GCGCCATGGCCAGAAGGTAATTATCCAAGATG
pCsVMV-Rv (EcoRV)	GCGGATATCCAAACTTACAAATTTCTCTGAAG
T_7 Fw	TAATACGACTCACTACTATAGGG

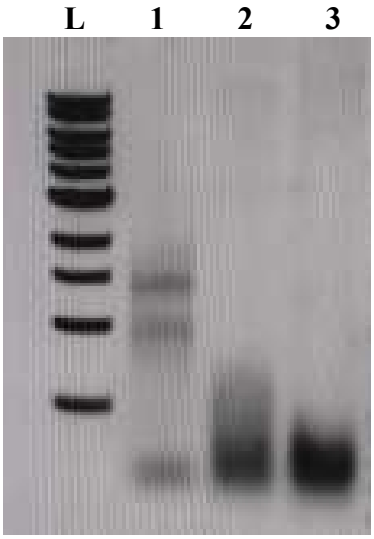
Appendix N: List of candidate genes for gene expression analysis

Table 5 List of candidate genes and their expression patterns in INTACT samples

Gene ID	Functional Annotation	Expression in INTACT samples	Reference
GLYMA19G35270	ABC TRANSPORTER G FAMILY MEMBER 34-RELATED	Enriched in nodule parenchyma	Phytozome v12.1
GLYMA19G43420.1	bZIP, Transcription factor family	Enriched in nodule parenchyma at 7 dpi and in infection zone at 10 dpi	Plant Transcription Factor Database, v5.0
GLYMA19G34881	CUP-SHAPED COTYLEDON 3	Enriched in infection zone	Soybean Knowledge Base (SoyKB), BLAST
GLYMA19G30640	Auxin-induced protein 5NG4-like	Enriched in infection zone	Peptide sequences of genes obtained from The Arabidopsis Information Resource (TAIR) and TBLASTN search against the soybean genome in LegumeIP
GLYMA17G37580.1	ARF5	Enriched in parenchyma at 7 dpi and in infection zone at 10 dpi	Plant Transcription Factor Database, v5.0
GLYMA17G08770	LONGIFOLIA protein	Enriched in nodule parenchyma	The Arabidopsis Information Resource (TAIR) and LegumeIP
GLYMA13G24810.1	PHOSPHATE 2, encodes UBIQUITIN-CONJUGATING ENZYME 24	Enriched in infection zone	The Arabidopsis Information Resource (TAIR), and Gene ontology and GO Annotations
GLYMA12G09670	NAC domain containing protein 90	Enriched in nodule parenchyma	The Arabidopsis Information Resource (TAIR),
GLYMA11G04500.2	Nitrate transporter	Enriched in infection zone	Soybean Knowledge Base (SoyKB) and BLAST
GLYMA10G00720.1	Phosphate transporter 1	Enriched in infection zone	Soybean Knowledge Base (SoyKB)
GLYMA08G04661.1	bHLH, Transcription factor family	Enriched in infection zone	Plant Transcription Factor Database, v5.0
GLYMA07G31630.1	PHOSPHATE 2	Enriched in infection zone	BLAST
GLYMA6G23800.2	Disease resistance protein (TIR-NBS-LRR class), putative	Enriched in infection zone	Soybean Knowledge Base (SoyKB) and BLAST
GLYMA05G31410	ETHYLENE INSENSITIVE 3 family protein	Enriched in infection zone	The Arabidopsis Information Resource (TAIR) and LegumeIP
GLYMA05G21680	Auxin-responsive GH3 family protein	Enriched in infection zone	The Arabidopsis Information Resource (TAIR) and LegumeIP

GLYMA03G42260.4	Myb Homeodomain-like superfamily protein	Enriched in nodule parenchyma	Plant Transcription Factor Database, v5.0
GLYMA01G36190.1	Auxin efflux carrier family protein	Enriched in nodule parenchyma	The Arabidopsis Inforation Resource (TAIR) and Legume IP
GLYMA19G06330	Uncharacterized LOC102668038 (LOC102668038), ncRNA	Enriched in infection zone	Soybean Knowledge Base (SoyKB) and BLAST

Appendix O: Agarose gel image of RNA



Appendix O: Agarose gel image of RNA

Figure shows migration pattern of 1) RNA (500ng) from non-transgenic nodule diluted in elution buffer @ 25ng/ul 2) RNA (200ng) from non-transgenic nodule diluted in elution buffer @ 10ng/ul (elution buffer was loaded in affinity column along with 500ul of Sepharose slurry. Three subsequent elutions were performed and RNA was diluted in the third elution) and 3) RNA (200ng) from third elution obtained from the affinity purification of roots transformed with the CsVMVp-TRAP cassette. The left most lane is 1kb ladder.