Expression, Regulation and Evolutionary patterns of a large family of Defensin- Like genes found in the nodules of *Medicago truncatula*

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Sumitha Nallu

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Kathryn A. VandenBosch (Advisor)

Kevin A.T. Silverstein (Co-Advisor)

May, 2011

ACKNOWLEDGEMENTS

I'm grateful to my advisor Kate VandenBosch for her patience and guidance. Our conversations throughout these years helped me foster my scientific thinking and writing. My co-advisor Kevin Silverstein has been a great mentor with a friendly attitude. I have gained a lot from his in-depth knowledge in various subjects. My advisors allowed a great deal of freedom to follow my own path, and always made time to lend an ear or advise me when I needed help.

My committee members, Jane Glazebrook, Debby Samac and Jim Bradeen are the most understanding group of advisors. They offered timely advice and encouraged me when the road got rocky.

I'm thankful to the past and present members of the VandenBosch lab. Past post-docs, Michelle Graham and Dashrath Lohar helped me during my initial years as a graduate student with advice on experimental work. Addition of Mesfin Tesfaye to our "DEFL group" helped our group and specifically me to accelerate my research work, his various expertise were handy in the lab. Mike Greene, our undergraduate with his enthusiastic zeal for science provided hands-on support especially for my work in the last chapter.

I benefitted from the various collaborations our lab had with other research groups. Our neighbors, the Gantt group, which we fondly called the "VandenBosch-Gantt family", made the seventh floor truly a "seventh heaven" for work. All the combined lab meetings and fun endeavors made the research work more enjoyable. My special thanks to Colby Starker, the post-doc of Gantt lab for all those words of motivation, scientific conversations and in particular for

patiently providing answers to all my naive questions. I would also like to thank the post-docs and graduate students from our other collaborators, Bruna Buccaleri (Carrol Vance lab), Tim Paape, Peng Zhou from Nevin Young's group and Ryoko Oono (Ford Denison lab).

My family has always been a constant source of love and support. All those long distance telephonic conversations with my parents on the other side of the globe reminded me of the "childhood dreams" and helped me not to get distracted even during the toughest times of this process. My brother, who has been my pillar of support, provided the encouragement that kept me going. This thesis would have never come through if it was not for the patience and support of my spouse. He deserves a special doctorate for being the most tolerant partner.

ABSTRACT

A very large number of Defensin-like (*DEFL*) genes were predicted in analyses of the genome sequences from the model plants *Medicago truncatula* and *Arabidopsis thaliana*. *DEFL*s share common characteristics with antimicrobial defensins, including cysteine-rich sequence motifs, gene structure and genome organization. However, because many of the *DEFL*s were only recently discovered, comparatively little information is known about their expression, regulation and evolutionary patterns. I explored the expression patterns of *DEFL*s among various tissues, symbiotic and pathogen treatments in legume *M. truncatula* and found that the majority of *DEFL*s are highly expressed in nitrogen-fixing root nodules. A further indepth investigation with various nodule developmental stages and rhizobial mutants indicated that the expression of nodule *DEFL*s is dependent on the number and morphology of rhizobia in the nodule. I found conserved motifs in the upstream regions that occur uniquely in nodule *DEFL*s. Promoter deletion assays demonstrated that these *cis* motifs are required for nodule *DEFL* expression. Few small regulatory elements known to be involved in the spatial and temporal regulation of various genes were contained within the unique conserved nodule *DEFL* motifs. Reverse genetic approaches suggested redundancy of function in this large family of genes. Lastly, I examined the evolutionary patterns of nodule *DEFL*s with four ecotypes of *M. truncatula*. The presence of expression, sequence variation and signatures of diversifying selection in nodule *DEFL*s within the Medicago species indicates

rapid and recent evolution and suggests that this family of genes is constantly evolving to adapt to different environments and acquiring new functions.

Supplementary Information:

Supplementary File 1: Supplementary tables from Chapter 2.

Supplementary File 2: Supplementary figures from Chapter 2.

Supplementary File 3: Supplementary tables from Chapter 3.

Supplementary File 4: Supplementary figures from Chapter 3.

Supplementary File 5: Supplementary tables from Chapter 4.

Supplementary File 6: Data analysis of transgenic plant assays at 0 dpi.

Supplementary File 7: Data analysis of transgenic plant assays at 14 dpi.

Supplementary File 8: Data analysis of transgenic plant assays at 40 dpi.

Supplementary File 9: Data analysis of pathogen assays on 14 dpi nodulated transgenic plants at 7 dpi and 12 dpi.

Supplementary File 10: Data analysis of pathogen assays on non-nodulated transgenic plants at 7 dpi and 12 dpi.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: Introduction

One of the biggest paradoxes in nature is the limited availability of nitrogen in soil, despite the fact that nitrogen is the most abundant element in earth's atmosphere. Only certain prokaryotic organisms are capable of reducing nitrogen into a form that can be biologically assimilated. Legumes, belonging to the family Fabaceae, form a mutualistic partnership with nitrogen-fixing bacteria. These plants provide carbon, an energy source to the nitrogen-fixing bacteria and harbor them in specialized root organs called nodules, in return for the nitrogen fixed by the bacteria. This association is governed by a complex interplay of molecular signals between the host and the microbe.

For the past half-century, researchers have made significant strides in understanding the genetic and physiological changes that occur in both the plant host and the symbiotic bacteria during the interaction (Dixon and Kahn 2004; Jones et al., 2007; Oldroyd and Downie 2008). This dissertation focuses on the expression, regulation and evolutionary patterns of one of the largest plant gene families encoding cysteine-rich proteins, the nodule Defensin-Like genes (nodule *DEFL*s), which are expressed in nodules during the symbiosis of *Medicago truncatula – Sinorhizobium meliloti*. Here, I present the background on *DEFL*s and discuss the specific questions and hypotheses that I formulated to guide the execution of my dissertation research.

Discovery of *DEFL*s

Expression of genes encoding cysteine-rich proteins in nodules was first reported in *Pisum sativum* (Scheres et al., 1990), followed by *Vicia faba* (Frühling et al., 2000), *M. truncatula* (Györgyey et al., 2000) and *Galega orientalis* (Kaijalainen et al., 2002). Three independent research groups then identified a large gene family (>300 members) encoding nodule-specific cysteine cluster proteins (*CCP*s), also known as nodule-specific cysteine-rich genes (*NCR*s), from the expressed sequence tags (ESTs) of the model legume *M. truncatula* (Fedorova et al., 2002; Mergaert et al., 2003; Graham et al., 2004). Expression of the nodule *CCP* family was found to be restricted to developing and mature nodules, except for two members that were also discovered to be expressed during mycorrhizal symbiosis (Mergaert et al., 2003). The expression patterns of the *CCP*s suggested at the time that they might function primarily in nodule and mycorrhizal symbioses, however, their exact roles were unknown. While conventional sequence homology searches could identify related sequences from nodules of other legumes in the Inverted Repeat Loss Clade (IRLC), the functions of these sequences were also unknown (Scheres et al., 1990, Kardailsky et al., 1993).

While the first *CCP*s were identified from nodule transcripts, Graham et al. (2004) also identified similar sequences from seed ESTs of *Glycine max* and *M. truncatula*. Although BLAST was used to find similar genes, this sequence comparison tool was insufficient to find similarity to genes of known function. Therefore, Graham et al. (2004) identified conserved motifs among the translated

sequences that were similar to sequence motifs in plant defensins. These motifs included conserved Cys residues that form dicysteine bridges important for protein folding in defensins (Thomma et al., 2002).

Defensins are a component of an ancient immune response and have been identified from vertebrates, invertebrates, plants and fungi (Boman, 1995; Mygind et al., 2005). They are small, cysteine-rich, cationic proteins, some of which have been found to be active against bacterial, fungal, or viral pathogens and/or insects (Boman, 1995; Stotz et al., 2009) and some have recently been found to play a role in regulating plant growth and development (Stotz et al., 2009).

Spurred by the discovery of such a large family of *CCP*s in *M. truncatula*, Silverstein et al. (2005) used the motif models developed in *M. truncatula* to iteratively search the near-complete *Arabidopsis thaliana* genome for similar Defensin-like (*DEFL*) sequences. Previously, 15 *DEFL* genes had been identified in the *A. thaliana* genome (Thomma et al., 2002). However, the motif searches used by Silverstein et al. (2005) identified 317 *DEFL* sequences, including all previously identified defensins. They characterized the expression patterns for a small subset of *DEFL*s via RT-PCR and observed specificity in nutrient-rich and reproductive tissues, in agreement with limited high-throughput expression results that were available for this gene family at the time. When the motif searches were expanded to include the EST collections from 25 plants, over 1,100 *DEFL*s were identified (Silverstein et al., 2005). More recently, 2,300

*DEFL*s were counted among scans of ESTs from 33 plant species plus the Arabidopsis and rice genomes (Silverstein et al., 2007).

The *DEFL*s share many characteristics with plant defensins (Graham et al., 2004, Silverstein et al., 2005). These characteristics include: [1] a signal sequence that is relatively well conserved, [2] a mature peptide sequence that is divergent but has conserved cysteine residues, [3] a gene structure that often has two exons with a conserved intron position and size, and [4] a genome organization that features clusters of genes likely derived from gene duplication. Because of their small size, sequence diversity, and the short sequence-length cutoffs imposed by existing annotation pipelines, the full extent of the defensin family had previously gone unrecognized (Graham et al., 2004).

I. Global expression patterns of Defensin-Like genes: How do *DEFL*s differ in their expression patterns between the model dicot *A. thaliana* and model legume *M. truncatula*?

Over the last decade, there have been significant efforts to create resources for gene expression atlases for model species, including *A. thaliana* and *M. truncatula*. In the model dicot *A. thaliana*, through an international effort AtGenExpress has been created using the ATH1 Affymetrix array, it contains expression data from 79 different samples across different tissues, developmental stages, and responses to biotic or abiotic stresses, hormones and light conditions (Schmid et al., 2005). More recently, At-TAX has been created using tiling arrays to accommodate the expression data for all those genes

missing from the ATH1 array (Laubinger et al., 2008). Massively Parallel Signature Sequencing (MPSS) is a technological advancement that has the expression profiles of small sized genes from several transcripts (Meyers et al., 2004). In spite the availability of these resources, at the time of this thesis research work more than 50% of *DEFL*s lacked expression data in *A. thaliana*.

In the model legume *M. truncatula*, the *Medicago truncatula* Gene Index (MtGI) was launched as an integrative database from international Expression Sequence Tag (EST) sequencing projects (http://compbio.dfci.harvard.edu/cgi-bin/tgi/gimain.pl?gudb=medicago), this database has substantial expression data sampled across different parts, biotic and abiotic treatments of *M. truncatula*. In 2008, *Medicago truncatula* Gene Expression Atlas was released. Using the Affymetrix Medicago gene chip, Benedito et al., (2008) collected transcript information from major organ systems and developmental series covering 64 different conditions. Recently, using custom microarrays Maunoury et al., (2010) studied the detailed transcriptomics of genes expressed in nodules of *M. truncatula*. Similar to *A. thaliana*, despite the availability of such vast information less than 50% of Medicago *DEFL*s are represented on those microarrays, and so the expression data is lacking for the rest of the *DEFL*s in *M. truncatula*.

A comprehensive study that includes the majority of *DEFL*s from a wide range of morphological parts and conditions for both species was required. This study would not only provide expression evidence for many of the recently identified *DEFL*s, but also would enable us to draw comparisons in *DEFL* transcriptomics between these two very different and important model species.

In Chapter 2, a collaborative effort explores the expression patterns of *DEFL*s in many different plant parts of *A. thaliana* and *M. truncatula*. Plants inoculated with different pathogens and plant mutants are also used for induction of *DEFL*s.

II. Expression, regulatory mechanisms and role of nodule *DEFL*s: What factors and elements regulate the expression patterns of nodule *DEFL*s and what is their functional role in *M. truncatula*?

Nodule *DEFL*s are the largest subfamily among *DEFL*s in *M. truncatula* (Chapter 2). Prior to this thesis research it was known from the nodule Expressed Sequenced tags (ESTs) that they are expressed in early and late nodules. In 2003, Mergaert et al. discovered more than 300 members of this family (they called them NCRs-nodule specific cysteine rich peptides). They profiled the expression of 14 *NCR* genes across different nodule developmental stages and in response to rhizobial mutants, and concluded that all the *NCR* genes were involved in nodulation, and probably acted at different nodule stages and in different tissues and cell types. In 2008, Benedito et al. generated a gene expression atlas that provided a global view of the expression patterns of genes in all organs of *M. truncatula*. This atlas described the nodule specific expression of 322 *NCR*s. However, a detailed and comprehensive investigation on the expression patterns of nodule *DEFL*s was yet to be reported.

Expression patterns in eukaryotes are governed by regulatory elements present in the non-coding regions of genes. A few nodule-specific regulatory

elements have been reported from various legumes. The promoters from the *G. max N23* and leghemoglobin *lbc3* genes (Stougaard et al., 1990), along with the *Sesbania rostrata Srglb3* (Szczyglowski et al., 1994), have been most extensively analyzed to date. In *Pisum sativum*, a short 139 bp promoter region was found to be sufficient to drive nodule expression by the early nodulin promoter from *Enod12B* (Hansen et al., 1999). A regulatory element called the NF box was identified in the promoter of Mt *ENOD11* (Andriankaja et al., 2007). Exploration for regulatory elements in large gene families like nodule *DEFLs* not only helps in understanding their expression patterns, but also highlights any possible alliance with other nodulins for recruitment of these elements during evolution.

Because of their similarity to the defensins, DEFLs have been hypothesized to have antimicrobial functions (Graham et al., 2004, Silverstein et al., 2005). But there are recent reports of some plant defensins and DEFLs functions as signal molecules in plant development (Allen et al., 2008), regulation of reproduction (Stotz et al., 2009), pollen tube guidance (Okuda et al., 2009) and pollen self-incompatibility (Schopfer et al., 1999, Takayama et al., 2001). Thus, one can hypothesize that the potential role of nodule DEFLs in *M. truncatula* is probably not restricted to defense but also in nodule development and physiology. Recently, it has been reported for a few NCRs (nodule DEFLs) that the nodule DEFL proteins have a lethal effect on free-living rhizobia in *in vitro* assays (Van de Velde, et al., 2010). They observed that like some defensins the NCRs (nodule DEFLs) triggered membrane modifications and inhibited bacterial cytokinesis. *Lotus japonicus* is a non-IRLC legume, the nodule *DEFL*s are absent

in its genome (Alunni et al., 2007) and its nodules are filled with rhizobia that can reproduce like the free living forms. When one of the *NCR* (nodule *DEFL*) was expressed in *L. japonicus,* it resulted in terminal bacteroid differentiation. In addition, similar effects were found when a few NCRs were ectopically expressed in cultured rhizobia (Van de Velde, et al., 2010). Reverse genetic assays like RNAi (knock down) and over expression can further determine the function or suggest functional redundancy of this large family of genes.

In Chapter 3, the expression patterns of nodule *DEFL*s are investigated in different nodule developmental stages and in nodules inoculated with various rhizobial mutants. Presence of putative regulatory elements is explored in different regions of the gene, their significance in governing the expression of nodule *DEFL*s is evaluated, and finally the functional role of nodule *DEFL*s is assayed by knocking them down and over expressing them in the nodules of *M. truncatula.*

III. Evolutionary patterns of nodule *DEFL*s in *M. truncatula*: Are nodule *DEFL*s a recent innovation in the legume lineage?

In legume-rhizobia symbiosis, there are two different rhizobial forms and two different nodule structures found depending on the species of the legume host. In the nodules of hosts like *M. truncatula*, *P. sativum*, *Vicia faba*, rhizobia differentiate into bacteroids that start fixing nitrogen and lose their reproductive ability. These bacteroids undergo major tranformations, they are swollen and branched with their genome amplified (Mergaert et al., 2006). In contrast in some

hosts like *G. max*, *L. japonicus*, *Vigna unguiculata*, the rhizobia are transformed into nitrogen fixing bacteroids but have no swelling or genome amplification and continue to reproduce. In the case of nodule dimorphism, they are classified broadly into indeterminate or determinate and further subtypes (Sprent et al., 2001, 2008). Indeterminate nodules have a persistent meristem and are elongated while the determinate nodules have a transient meristem and are spherical. There are examples of legumes with correlated bacteroid and nodule types, like in *M. truncatula*, *P. sativum*, *V. faba* (closely related species of IRLC) have indeterminate nodules with swollen bacteroids, whereas *G. max*, *Phaseolus vulgaris*, *L. japonicus* have determinate nodules and non-swollen bacteroids. However, there are also reports that conflict the idea of correlation (Chandler et al., 1982, Sprent and Thomas 1984).

BLAST searches in National Center for Biotechnology Information (NCBI) database indicated the presence of nodule *DEFL*s to be restricted to legumes belonging to IRLC clade like *M. truncatula*, *P. sativum*, *V. faba*, *Trifolium repens* and *Galega orientalis* which have indeterminate nodules with swollen bacteroids (Mergaert et al., 2003). A complete genome search of non-IRLC legumes with determinate and non-swollen bacteroids like *L. japonicus* (Alunni et al., 2007) and *G. max* (Michelle Graham, unpublished) could not detect the orthologs of nodule *DEFL*s. In contrast, seed *DEFL*s (i.e., *DEFL*s that specifically expressed in seeds) were found in *G. max* (Graham et al., 2004), suggesting the clade specificity of nodule *DEFL*s. These reports led to the hypothesis of correlation of

the occurrence of nodule *DEFL*s with nodules that are of indeterminate type and contain swollen bacteroids (Alunni et al., 2007).

Recently, a study of phylogeny of 40 legume species of Papilionoideae revealed that there is no consistent relationship between nodule type and rhizobial form, and that association with non-swollen bacteroids is an ancestral trait of legumes (Oono et al., 2010). This study suggested that the swollen bacteroid is a derived trait with five independent origins in Papilionoideae. Because of the correlation of nodule *DEFL*s with the swollen bacteroid trait (Van de Velde, et al., 2010), it can be hypothesized that nodule *DEFL*s are also a derived trait and had five independent origins. Availability of sequence information for rest of the clades will help in testing this hypothesis.

To understand the evolutionary patterns of nodule *DEFL*s, it is important to determine their expression and sequence polymorphisms between the closely related species and within species. Another tool for studying molecular evolution is through the analysis of rates of non-synonymous and synonymous substitutions in the coding region of the genes (Romero-Herrera et al., 1973). Among defensins, a closely related gene family to *DEFL*s (Graham et al., 2004), the evolutionary patterns within the paralogous subgroups of β-defensins in mammals indicate the evidence of purifying or diversifying selection, depending on the subgroup (Maxwell et al., 2003, Semple et al., 2003, Semple et al., 2006). Another functionally related NBS/LRR family of R genes also show similar trends among subgroups in evolutionary pressures within the paralogous subgroups (Baumgarten et al., 2003, Meyers et al., 2003) as well as orthologous groups in

A. thaliana (Bakker et al., 2006). Studies of the *DEFL* family paralogs in *A. thaliana* (Silverstein et al., 2005) and a limited number of nodule *DEFL* paralogs in *M. truncatula* (Alunni et al., 2007) indicate that the signal peptide is predominantly under purifying selection, whereas the mature polypeptide has signatures of both purifying and diversifying depending on the cluster. By estimating and comparing the selection pressures in orthologous groups of nodule *DEFL*s to closely related gene families, one can draw conclusions on their evolutionary rates.

In Chapter 4, the expression and sequence polymorphism of nodule *DEFL*s in ecotypes of *M. truncatula* are examined and their evolutionary rates are estimated in comparison to seed *DEFL*s.

CHAPTER 2: Defensin-Like genes in *Arabidopsis thaliana* and *Medicago truncatula* differ markedly in their spatio-temporal expression patterns[1]

M Tesfaye, KAT Silverstein, S Nallu, L Wang, CJ Botanga, SK Gomez, M Harrison, D Samac, J Glazebrook, F Katagiri, JF Gutierrez-Marcos and KA VandenBosch.[2]

Summary

Plant genomes typically contain up to several hundred defensin-like (*DEFL*) genes that encode short proteins resembling defensins, which are antimicrobial polypeptides. Little is known about the expression patterns or functions of many

---

[1] This chapter will be submitted for publication to the journal *Plant Physiology*, and is written in the style of the journal.

[2] The authors M Tesfaye, KAT Silverstein, S Nallu and KA VandenBosch designed the experiments and carried out the majority of research and analysis. As part of the collaborative research, the other authors provided resources required for a part of the study and contributed conceptually to aspects of its design. Specifically, S Nallu was involved in the design and execution of assays involving the inflorescences in *A. thaliana* and flower buds, flowers, seeds, seedlings, roots, nodules, mycorrhizal roots and pathogen infected roots and leaves in *M. truncatula*.

*DEFL*s, since most were recently discovered and many are not well represented on standard microarrays. For expression analysis in two model plant species, we designed a custom Affymetrix chip for transcript profiling of nearly all *DEFL*s from *Arabidopsis thaliana* and *Medicago truncatula*, as well as additional genes for data normalization. Among three normalization algorithms evaluated, the Stable-Based Quantile (SBQ) approach gave the best performance for the custom chip. The DEFL chip analysis provided evidence for the transcription of many of the recently identified *DEFL* genes that previously lacked expression information. Most *DEFL*s are expressed in highly condition-specific patterns, and these differ notably between Arabidopsis and Medicago. *DEFL* expression in Arabidopsis was most pronounced in inflorescences, a pattern not found in Medicago, where *DEFL*s were most prominently expressed in nitrogen-fixing nodules. Both species contain a subset of *DEFL*s specifically expressed in seeds or fruits. A few *DEFL*s, including some defensins, were significantly up-regulated in pathogen-infected tissues of Arabidopsis and Medicago. Using plant and bacterial mutants with defects in defense signaling and co-expression analysis with three marker genes for early defense signaling and JA-mediated host response, we identified *DEFL*s that appear to be associated with putative defense responses.

Introduction

Defensins are endogenous antimicrobial polypeptides present in most eukaryotic life forms (Boman, 1995; Mygind et al., 2005). These small, charged,

cysteine-rich polypeptides are secreted or sequestered within compartments of the endomembrane system, and have broad-spectrum anti-fungal, anti-bacterial, anti-viral, and/or insecticidal activity via a variety of proposed molecular mechanisms (Boman, 1995; Terras et al, 1995; Thomma et al., 1998; 2002, Stotz et al., 2009a).

Plants have a large repertoire of genes encoding defensin-like (DEFL) polypeptides that have a conserved pattern of cysteine residues, but that are otherwise highly variable in the mature protein (Graham et al., 2004; Silverstein et al., 2005; Silverstein et al., 2007). The model legume *Medicago truncatula* (hereafter called Medicago) has at least 800 *DEFL* genes, with more likely to be discovered once its genome is fully sequenced (Silverstein et al., 2007). Fedorova et al., 2002; Maergaert et al., 2003; Graham et al., 2004). Arabidopsis has 317 *DEFL* genes (Silverstein et al., 2005), while rice has 93 (Silverstein et al., 2007).

DEFL gene families appear to have expanded and diversified in plant genomes via tandem and ectopic duplication, followed by successive rounds of diversifying selection (Graham et al., 2004; Silverstein et al. 2005; Alunni et al., 2007). The expansion and diversification of the *DEFL* family appear to have led to divergent subgroups of *DEFL*s, including some characteristic of different plant lineages, while other subgroups, such as the defensins themselves, have remained broadly conserved among lineages (Silverstein et al., 2007).

The phylogenetic diversity among *DEFL*s suggests that different gene family members may play distinct roles in divergent taxa, but knowledge of the

functions of this large gene family is incomplete. Some defensins are constitutively expressed or developmentally regulated, where they act as antimicrobial agents that protect nutrient-rich and vulnerable tissues, such as in seed development and seedling growth (Hanks et al., 2005; Terras et al., 1995). Other defensins are deployed specifically during defense responses, and are induced downstream of jasmonate or other defense signaling molecules (Manners et al., 1998; Fujita et al., 2006). Roles for DEFLs in plant/microbe interactions are likely to be more varied. For example, nodule cysteine-rich proteins belonging to the DEFL family were recently demonstrated to effect the terminal differentiation of nitrogen-fixing bacteroids in certain legume nodules (Van de Velde, et al., 2010).

By contrast, some members of the highly diverse DEFL group have also been shown to function in development. For example the S-locus cysteine-rich proteins (SCRs), which make up a subgroup of DEFLs, are expressed in pollen tubes and act as signal molecules in self-incompatibility in some dicots (Schopfer et al., 1999, Takayama et al., 2001). Recently some DEFLs, known as LUREs, that are expressed in the female gametophyte of *Torenia* have been found to function in pollen tube guidance (Okuda et al., 2009). Other DEFLs act in a variety of other plant development processes (Allen et al., 2008; Stotz et al., 2009a). However, the characterization of DEFLs as functioning either in microbial interactions or in developmental regulation may be a misleading dichotomy. A tomato defensin has been found to show evidence of bifunctionality, acting both to regulate development and to confer disease resistance (Stotz et al., 2009b).

Despite great progress toward creating gene expression atlases for Arabidopsis (Schmid et al., 2005) and Medicago (Benedito et al., 2008), expression data for *DEFL* genes are far from comprehensive in either species. This is because standard Affymetrix arrays, utilized to construct gene atlas projects, contain only 12% (in the case of the Arabidopsis ATH1 array) and 50% (in the case of the Medicago genome array) of known *DEFL* genes (Silverstein, unpublished). Until whole genome tiling arrays or *de novo* next-generation transcript sequencing data are more widely available and cost-effective, it may be difficult to ascertain comprehensive expression patterns of *DEFL* genes as a family, and to determine whether DEFL functions are conserved or divergent among angiosperms.

To address this gap, we have designed a custom Affymetrix microarray (here after called the AtMtDEFL array) with known and predicted Arabidopsis and Medicago *DEFL* genes, and the closely related maternally expressed gene family (*MEG*s; Gutierrez-Marcos et al., 2004; Silverstein et al., 2007). The custom array was utilized for hybridizations of mRNAs from a wide range of conditions from both species, enabling a comparative analysis in the two model plant species. The custom chip analysis of *DEFL* transcript accumulation provided the first expression evidence for many of the recently identified *DEFL*s, and indicated both common and unique patterns of expression between the two plant species.

Results

Development of the custom AtMt*DEFL* array and robust normalization methods*.*

The AtMt*DEFL* array was commissioned from Affymetrix in March 2007 and includes one probe set each for all 317 Arabidopsis *DEFL*s, and 684 previously identified Medicago *DEFL*s (Silverstein et al., 2007), plus additional marker genes. The array also contains 401 probe sets (229 and 172 from Arabidopsis and Medicago, respectively) with invariant levels of expression (hereafter called invariant genes), to aid normalization of data. All probe sets were represented by 11 Perfect Match (PM) and 11 Mis Match (MM) probes, and 82% of Arabidopsis and 52% of Medicago *DEFL* probe sets do not cross-hybridize to other genes. Probe sets were interspersed on the custom array, although chip descriptions and library files (cdfs) were created separately for the two species by Affymetrix, with the intent that chip hybridizations and microarray data analyses would be performed for only one plant species at a time.

Although Robust Multi-array Average (RMA; Irizarry et al., 2003) has been frequently applied to genome-wide Affymetrix arrays, the primary assumption that the distribution of probes responding in each quantile level of expression is similar across experiments has not been met by the expression data for the *DEFL* family of genes on our custom array (Supplemental Figure 2-S1). Distribution plots of the raw signal intensity data were clearly skewed to much higher signal intensity levels in Arabidopsis siliques and especially inflorescences than they are in leaves and roots. If we attempted to apply RMA globally across At*DEFL* custom arrays (the subset of the AtMtDEFL array made up of probe sets

corresponding to Arabidopsis genes), these highly variable distribution plots would be coerced into an average distribution plot, which would significantly skew absolute expression values, and simultaneously normalize out real biological differential expression between arrays. These caveats in the distribution of expression values often don't exist for genome-wide Affymetrix arrays such as the ATH1, because the sheer number and random nature of probe sets on these arrays tend not to show marked tissue- and condition-specific effects. Hence, the assumptions underlying RMA are not violated for such larger arrays as ATH1.

The Affymetrix ATH1 and the custom At*DEFL* arrays share probe sets for 299 genes (171 invariants, 37 *DEFL*s, and 91 markers and other genes). As a test for comparing microarray data normalization approaches described in the methods, we downloaded publicly available microarray data and carried out RMA normalization on 36 ATH1 datasets (Supplemental Figure 2-S2). The public data sets were chosen to include biological samples in which various common *DEFL*s were known to be differentially expressed (e.g., flowers, siliques, roots, leaves, seeds, stamens, stems, cotyledons, and seedlings). Restricting our view to the subset of the 299 genes on the ATH1 array that were shared by our custom At*DEFL* array, we sought a normalization method that, when applied only to these probe sets, achieved high correlation with the expression values obtained via RMA performed on the entire ATH1 array.

Figure 2-1 shows that the Stable-Based Quantile (SBQ, Sato et al., 2007) method (see Methods and Materials for details) has an outstanding

correlation (Pearson R > 0.998) with RMA, even though it only includes expression data from the restricted set of *DEFL*s, invariants, and other genes common to the At*DEFL* custom array. Two other microarray data normalization methods we proposed, RMA with Invariant Median Scaling (RIMS) and RMA with Median Absent Probe set Scaling (RMAPS) (see Materials and Methods for details), also showed high correlation values of 0.992 and 0.998, respectively. Heatmaps of At*DEFL*s, invariants, and additional genes produced by all four normalization methods examined are presented in Supplemental Figures 2-S3 to 2-S5.

Similarly, Mt*DEFL* custom array (the Medicago subset of the AtMt*DEFL* array) raw probe set intensity distributions in nodules are highly skewed to higher intensities relative to uninoculated roots and other tissues and conditions examined (Supplemental Figure 2-S6). We concluded that the assumption underlying quantile normalization across *DEFL* genes using the custom array does not apply. The Affymetrix Medicago genome array and the custom Mt*DEFL* array share probe sets for 565 genes (172 invariants, 370 *DEFLs*, and 23 additional genes). If we apply the SBQ, RMAPS and RIMS normalizations to the standard Affymetrix Medicago array, restricting the analysis to the 565 genes that are shared with the custom Mt*DEFL* array, we again achieve outstanding correlation with the expression values obtained from RMA applied over all 61,000 probe sets (Supplemental Figure 2-S7). Again, SBQ shows the best concordance among the three normalization methods examined. But all three methods have high correlation values (all with R>0.99).

We performed SBQ, RIMS, and RMAPS normalization on our custom At*DEFL* chip for untreated roots, leaves, and inflorescences, and computed the correlation with data deposited at NCBI from similar Arabidopsis samples across the subset of shared genes on the two platforms. We found that the At*DEFL* custom arrays perform similarly to the Affymetrix ATH1 array***.*** All raw public data from NCBI was re-normalized using RMA prior to comparison. SBQ correlations were outstanding (Figure 2-2), ranging from a low of R=0.93 for inflorescences to R=0.95 for roots. RIMS and RMAPS At*DEFL* correlations with the ATH1 RMA datasets were nearly as good, each ranging from 0.93-0.94 and across these morphological structures. Consequently, we used SBQ in the subsequent study.

Many computationally predicted Arabidopsis *DEFL* family genes show constitutive and condition-specific expression patterns.

Based on MAS5 detection call criteria ('presence' vs 'absence'), a total of 206 *DEFL*s (about 65%) could be regarded as expressed in one or more Arabidopsis organs examined (Figure 2-3A). The number of expressed *DEFL*s ranged from 41 in roots to 170 in inflorescences. We also observed the expression of 76 *DEFL*s in pathogen-infected leaves. Approximately 36 *DEFL*s were expressed in all infected and healthy tissues and organs examined (Figure 2-3A). There were many *DEFL*s (approximately 36% of all *DEFL*s) with unique gene expression patterns, ranging from one uniquely expressed *DEFL* in seedlings (7- and 14-day-old) to 59 uniquely expressed *DEFL*s in inflorescences.

To further improve our information base on the number of Arabidopsis *DEFL*s for which we have gene expression evidence, we cross-referenced the list of our expressed *DEFL*s in the present study to that of two other sequencing based transcript profiling approaches: MPSS (Massively-Parallel Signature Sequences) database, and the Arabidopsis Gene Index (AGI) at the Dana Farber Cancer Institute (http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=arab). The Massively-Parallel Signature Sequences (MPSS) database (http://mpss.udel.edu/at/) is based on sequencing technology that captures very small sequence signature tags (17 bp and 20 bp) for expressed transcripts. The AGI is a culmination of sequencing efforts involving a collection of several thousand expressed sequence tags. From our comparison, the total number of *DEFL*s and Maternally Expressed Genes (*MEG*s) with evidence of expression in one or more organs and conditions has been increased to 255 genes (79% of the 322 *DEFL*s and *MEG*s on the chip) (Figure 2-3B). Of this total, the custom chip analysis provides new insights into gene expression profiles of 89 recently identified *DEFL*s previously lacking any evidence of expression.

The SBQ-normalized signal intensity values of *DEFL*s collected from a variety of Arabidopsis organs, including healthy and pathogen-infected material, presented in Supplemental Table 2-1A, are schematically displayed as a heat map (Supplemental Figure 2-8). From this figure, it is immediately apparent that a great majority of expressed *DEFL*s show very strong transcript abundance in inflorescences. Interestingly, approximately 14 *DEFL*s, including a plant defensin (AT1G55010) and an S-locus cysteine-rich protein gene (*SCR*) (AT1G60989)

were expressed largely in seedlings and/or roots, either at much higher amplitude than in other Arabidopsis organs or in a seedling- and root-specific manner (Figure 2-4).

Expression atlas of *SCR*, *MEG* and *PDF* gene families.

Gene expression patterns are known only for a small number of *SCR*s and *MEGs*, mainly because a great majority of the *SCR*s and *MEGs* were not represented on the Affymetrix ATH1 array. Those *SCR*s and *MEG*s studied were shown to be expressed specifically in inflorescences and siliques, where they are expected to play reproductive regulatory roles (Vanoosthuyse et al., 2001; Gutierrez-Marcos et al., 2004). Moreover, three of the classical *PDF*s (*PDF1.2c*, *PDF1.3* and *PDF1.5*) were not present in the ATH1 array. The custom array contains probe sets representing 28 *SCR*s, 15 *MEG*s and all 15 classical *PDF*s. We evaluated the tissue-specific and developmental regulated expression patterns of these *DEFL* gene families using the microarray data collected from several plant tissues and organs during normal physiological conditions and pathogen infected tissues. It was found that only 20 *SCR*s were expressed in the organs and experimental conditions tested. Except for AT1G60989, which showed enhanced expression in 21 days-old roots, the rest of the *SCR*s were expressed primarily in inflorescence tissues (Figure 2-5A). Of the 15 *MEG*s on the custom chip, 10 were expressed in at least one of the tissues tested, although a great majority showed enhanced expression patterns primarily in inflorescence tissues (Figure 2-5B).

Cluster analysis revealed three main expression groups of co-regulated *PDF*s with distinct organ- and/or condition-specific expression patterns (Figure 2-5C). The first PDF expression cluster consisted of *PDF1.5*, *PDF2.1*, *PDF2.2*, *PDF2.3* and *PDF2.5* that showed up-regulated expression primarily in Arabidopsis seedlings and root tissues, but down regulated expression in pathogen infected tissues. The second *PDF* expression cluster comprised of *PDF2.4*, *PDF2.6*, *PDF3.1* and *PDF3.2* that showed upregulated expression in inflorescence tissues. The third *PDF* expression cluster included those genes that showed significantly up-regulated expression patterns in pathogen infected leaves. These group of classical defensins included *PDF1.1*, *PDF1.2a-c*, *PDF1.3* and *PDF1.4*.

Identification of fungal pathogen-responsive Arabidopsis *DEFL*s.

In an attempt to identify pathogen-responsive *DEFL*s in Arabidopsis, wild-type Col-0 plants and a *dde2-2* mutant in the Col-0 background were challenged with the necrotrophic fungal pathogen *Alternaria brassicicola*, the causal organism for early blight, and patterns of gene expression were studied in leaf samples collected 24 hpi. Leaves from mock-inoculated plants were included for comparison. *DDE2-2* encodes the allene oxide synthase, which is essential for jasmonate biosynthesis (von Malek et al., 2002). In agreement with previous findings (van Wees et al., 2003; AbuQamar et al., 2006), *dde2-2* plants were found to be more susceptible to *A. brassicicola* than wild-type Col-0 plants (Supplemental Figure 2-S9). The microarray data showed that a total of 21

genes including eight *DEFL*s and five *PDF*s were significantly (P<0.05) differential expressed (Table 2-1). Of these, two *DEFL*s (AT3G05730 and AT2G43550) were significantly repressed 24 hpi with *A. brassicicola* inoculation, regardless of the plant genotype. Transcript abundance of one *DEFL* (AT2G43530) was also significantly reduced by *A. brassicicola* inoculation in the *dde2-2* mutant, but not in the wildtype Col-0 (Table 2-1). Differentially regulated genes include many known JA-dependent defense response genes such as *PDF1.1*/*PDF1.2*, *PDF1.2a-c*, and *PDF1.3* (AT1G75830, AT5G44420, AT2G26020, AT5G44430, AT2G26010),a hevein-like protein precursor (*HEL1*) (AT3G04720), and a gene encoding basic chitinase (AT3G12500) whose transcript abundance in the *dde2-2* mutant was not increased to the level of transcript accumulation in wildtype Col-0 plants (Table 2-1).

***DEFL* expression patterns in Arabidopsis coincide with bacterial pathogen recognition.**

Two modes of pathogen recognition mechanisms are known in plants: recognition of molecular patterns (pathogen-associated molecular patterns, PAMPs, or microbe-associated molecular patterns, MAMPs) common in some plant pathogens, and recognition of certain effectors delivered inside the plant cell by the cognate resistance proteins. The effector proteins (type III effectors) are said to interfere with host immune signaling. Recognition by PAMPs and type III effectors by host plants leads to induction of PAMP-triggered immunity (PTI) and effector-triggered immunity (ETI), respectively. Gram-negative bacterial

pathogens, such as the hemibiotroph *Pseudomonas syringae* pv. *tomato* DC3000 (Pto DC3000), employ the type III secretion system to translocate effector proteins into the host cytoplasm. The bacterial *hrcC* gene is an essential component of the type III secretion system, so inoculant bacteria with the *hrcC* mutation cannot deliver type III effectors into the host cytoplasm. Therefore, the strain *Pto* DC3000 *hrcC-* can induce PAMP-triggered immunity (PTI) but cannot interfere with PTI, whereas the wild-type *Pto* DC3000 induces and interferes with PTI. Consequently, *Pto* DC3000 reproduces well in the wild-type Col-0, but *Pto* DC3000 *hrcC-* cannot. On the other hand, the *Pto* DC3000 AvrRpt2 bacterial strain induces effector-triggered immunity (ETI) as the wild-type Col-0 plants expresses the resistance protein RPS2, which recognizes the type III effector *AvrRpt2* gene. In this study, leaves collected at 3 and 9 hpi with bacterial strains *Pto* DC3000, *Pto* DC3000 AvrRpt2, and *Pto* DC3000 *hrcC-* were used to study gene expression responses in wild-type Col-0. At 3 hpi, six *PDF*s and 12 *DEFL*s were significantly ($P<0.05$) differentially regulated among the three Pseudomonas bacterial strains used, although the number of significantly ($P<0.05$) differentially regulated genes was increased to 8 *PDFs* and 29 *DEFLs* at 9 hpi (Table 2-2). All but two genes that were differentially regulated at 3 hpi were also differentially regulated at 9 hpi. *Pto* DC3000-induced genes at 9 hpi included six *PDF*s and 13 *DEFL*s, and *Pto* DC3000 repressed genes included two *PDF*s and 16 *DEFL*s (Table 2-2).

We expected that transcript accumulation of PTI- and ETI-responsive genes in plants inoculated with *Pto* DC3000 *hrcC-* and *Pto* DC3000 AvrRpt2

strains would not be expressed to a similar degree as in plants inoculated with wildtype *Pto* DC3000. By this criterion, a novel *DEFL* (AT5G43285), five classical defensins *PDF1.2a-c*, *PDF1.3* and *PDF1.4* (AT5G44420, AT2G26020, AT5G44430, AT2G26010 & AT1G19610), and several other genes including a hevein-like protein (*HEL/PR-4*)(AT3G04720), and two high-affinity nitrate transporters (AT3g45060, AT1G08090) could be considered as MAMP-responsive genes. Genes that showed evidence of ETI-response expression patterns (gene-for-gene interactions) included two *DEFLs* (AT3G61185, AT5G4687) , a single *PDF* (AT1G75830), as well as genes encoding *PR-3* (AT3G12500), *PR-1* (AT2G14610), 4CL (AT3G21230) and a MYB108 transcription factor (AT3G06490) (Table 2-2).

Co-expression analysis with known markers for defense signaling pathways verified expression patterns of pathogen-responsive *DEFL*s:

We extended our analysis of the Arabidopsis custom microarray data for co-expression patterns using three marker genes for host defense response. Genes used for the co-expression analysis include *PDF1.2a* (AT5G44420), which is a marker for host defense via the JA-mediated pathway (Glazebrook, 2005; Penninckx et al., 1998), *coronatine-insensitive 1* (*COI1*, AT2G39940), also known to be induced by methyl jasmonate (Lorenzo and Solano, 2005), as well as the MAMPs-inducible gene, FLG22-inducible receptor-like kinase 1(*FRK1*; AT2G19190), which is implicated in early defense signaling associated with the perception of flagellin (Asai et al., 2002). Our analysis identified 16 genes

including four *DEFL*s that showed co-regulated (R2 ≥0.7) expression patterns with the *FRK1* gene (Figure 2-6A). Other co-regulated genes with *FRK1* included *PR-1* (AT2G14610/AT1G64280), *PR-5* (AT1G75040), *APX* (AT1G07890), β-glucanase (AT3G57260), and glycerol kinase (AT1G80460).

We also identified two *DEFL*s (AT2G16367/AT5G18407), and four *PDF*s (AT1G19610, AT2G26010, AT2G26020, AT5G44430) that showed co-regulated (r2 ≥0.7) expression patterns with *PDF1.2a* and/or *COI1* (Figure 2-6B). Except for AT2G16367, transcript abundance of the remaining co-regulated *DEFLs* and other genes was already shown to be significantly regulated by *Pseudomonas* and/or *A. brassiciocola* inoculations in our study (Tables 2-1 and 2-2).

RT-PCR validation of microarray data.

We employed quantitative reverse-transcription polymerase chain reaction (qRT-PCR) to verify the *DEFL* microarray data trends that were collected from several Arabidopsis tissues and pathogen challenged leaves. First we needed to identify suitable reference genes that show constant expression levels for qRT-PCR analysis of gene expression data. We initially identified four candidates as potential reference genes as described in Materials and Methods. We then validated the gene expression stability of these selected genes by qRT-PCR assay. Gene expression stability analysis of these four genes using the average threshold cycle (Ct) from the qRT-PCR assays showed that any one of these four is equally good (Supplemental Figure 2-S10). Although all four Arabidopsis genes can be equally used as internal control

genes by RT-PCR analysis of gene expression patterns, we chose to use two genes, Metallothionein 2B (AT5G02380) and Yellow-leaf-specific gene 8 (AT5G08290) as reference genes for follow up with qPCR validation of our microarray data.

In an effort to validate our microarray data, we chose ten Arabidopsis *DEFL*s as reference points for validation of organ- and/or condition-specific gene expression patterns. As shown in Supplemental Table 2-S2, fold- change values among selected morphological structures and conditions obtained via RT-qPCR were most quantitatively consistent with tissue-specific and/or condition-enhanced expression patterns obtained with the SBQ normalized custom array data. In most cases however, fold changes obtained by custom array data were much smaller in amplitude than those obtained via qRT-PCR, as commonly reported in the literature.

Mt*DEFL*s are categorized based on their tissue- and condition-specific expression patterns

In order to survey the expression profiles of 684 *DEFL*s in *M. truncatula* genotype A17, we screened different parts of the plant and various treatments with pathogens and its symbiotic partners using the custom AtMt*DEFL* chip. The SBQ-normalized signal intensity values for the various Medicago developmental and microbe challenged tissues are schematically displayed as a heatmap (Supplemental Figure 2-S11). It is immediately apparent that a great majority of *DEFL*s show very strong gene expression in root nodules and seeds.

Approximately 79% of the total Mt*DEFL*s on the custom chip were expressed in microbe-associated tissues. Based on MAS5 gene expression detection call criteria, a total of 176 Mt*DEFL*s (26% of the Mt*DEFL*s on the custom chip) were expressed in one or more Medicago organs and tissues that comprised of germinating seeds, leaves, stems, flower buds, opened flowers, immature seeds and whole roots of uninfected and/or uninoculated (supplied with artificial fertilizers as needed) from plants grown under normal conditions (Figure 2-7). The number of expressed Medicago *DEFL*s (Mt*DEFL*s) ranged from 19 in stems to 103 in seeds. Approximately 40 Mt*DEFL*s were expressed in all vegetative, reproductive and microbe-infected tissues (Figure 2-7). Interestingly, 52 Mt*DEFL*s were uniquely expressed in Medicago seeds, where as only three Mt*DEFL*s (MtgAC142396_38300,MtgCU302334_101329, MtgAC149130_78022) were specifically expressed in flower buds and flowers.

Expression of *DEFL*s in symbiotic tissues of Medicago

To study the Mt*DEFL* gene expression during symbiotic interactions of *M. truncatula* A17 with arbuscular mycorrhizal (AM) fungus and root nodule forming soil bacteria, we selected the vastly studied species *Glomus intraradices* and *Sinorhizobium meliloti* Sm1021 strain respectively in Medicago-symbiont interactions. For AM symbiosis, *Glomus intraradices* colonized roots (Myc+) were collected at 28 dpi which is a substantially infected stage, a total of 44 *DEFL*s were found to be expressed in Myc+ roots. For statistical analysis of gene expression during AM symbiosis, signal intensity values in Myc+ colonized roots

were compared with signal intensity values with non-mycorrhizal control roots (Myc−). Twelve Mt*DEFL*s were significantly (P<0.05) differentially regulated between Myc+ and Myc- control roots, and all but one of the differentially regulated Mt*DEFL*s was significantly upregulated in Myc+ roots compared to Myc- roots (Supplemental Table 2-S3). With regard to unique expression patterns, there were 11 Mt*DEFL*s that were expressed uniquely in Myc+ roots.

Our preliminary analyses of data from Bendito 2008 et al., showed that majority of *DEFL*s in nodules are expressed at 14 dpi. As expected, a great majority (77%) of the *DEFL*s on the custom chip were found to be expressed in root nodules formed by *Sinorhizobium meliloti* Sm1021 strain at 14 dpi. Statistical analysis of SBQ normalized data revealed that 517 Mt*DEFL*s were significantly (P<0.05) differentially regulated between nitrogen-fixing root nodules and uninoculated root controls of similar age (Supplemental Table 2-S4).

Mt*DEFL* gene expression during pathogenic microbial interactions

Colletotrichum trifolii, a foliar pathogen and a root pathogen, *Phytophthora medicaginis,* are the common pathogens associated with Medicago pathogenesis. We investigated *DEFL* expression patterns during their interactions with Medicago. The fungal pathogen *C. trifolii* is a causal organism for athracnose in many plant species, including alfalfa (Barnes et al., 1969). However, *M. truncatula* genotype A17 displays a rapid hypersensitive response (HR) at 12-24 h after inoculation with *C. trifolii* and is thus highly resistant (Torregrosa et al., 2004). In an effort to identify Mt*DEFL*s whose expression is

coincident with the HR response, the custom AtMt*DEFL* array was used to evaluate gene expression patterns at 24 hpi with *C. trifolii* race 1 of *M. truncatula* A17. Approximately 36 Mt*DEFL*s were detected in infected leaves of Medicago in response to *C. trifolii* inoculation, although only four *DEFL*s were found to be significantly (P<0.05) differentially regulated between *C. trifolii* inoculated and mock-inoculated leaf samples (Supplemental Table 2-S5).

M. truncatula A17 is moderately susceptible to *P. medicaginis.* Lesions are typically small, occurring about 5 dpi, and don't kill the plant but will stunt it (Samac, unpublished). Our preliminary timeline quantitative PCR results indicated the highest significant differential expression of two antimicrobial peptide genes (TLP and PR1) at 4 dpi during *M. truncatula* and *P. medicaginis* interaction. Further investigations of Mt*DEFL* expression using AtMt*DEFL* array revealed approximately 41 Mt*DEFL*s to be expressed in infected root segments of Medicago in response to *Phytophthora* inoculation. Eight *DEFL*s were (P<0.05) significantly differentially regulated between *Phytophthora* inoculated and mock inoculated root tissues at 4 dpi.

Discussion

The *DEFL* array's design is robust and accurate for expression profiling of *DEFL*s in Arabidopsis and Medicago

We designed an Affymetrix custom chip for *DEFL* gene transcript analysis in two model plant species, *A. thaliana* and *M. truncatula*, in a variety of organs and developmental stages, and during symbiotic and pathogenic associations.

This custom array study provides expression evidence for 64% and 90% of the *DEFLs* identified in Arabidopsis and Medicago, respectively. Various normalization methods were examined for analysis of the data, among all the methods explored, the SBQ method was found to be superior.

Microarray data normalization algorithms such as Affymetrix's MAS 5.0 (Liu et al., 2002) or loess (Smyth and Speed, 2003) normalization require that the majority of genes show unchanged pattern of gene expression among the conditions under consideration, or at least that an equivalent proportion of genes are up- and down expressed. Alternatively, Robust Multichip Average (RMA, Irizarry et al., 2003) and other quantile normalization based schemes (Bolstad et al., 2003) require that the density distribution of intensities is at least qualitatively similar across samples. These essential probe intensity distribution assumptions for microarray data are clearly violated when boutique arrays such as the custom AtMt*DEFL* chip used in this study. For example, we have observed more than five hundred Medicago *DEFL*s to be expressed in nitrogen-fixing nodules, yet only a handful were detectable in control uninoculated roots. These extreme expression patterns pose a challenge to typical microarray data normalization algorithms used routinely (Wilson et al., 2003; Sato et al., 2007). Hence we have included more than 170 invariant genes for each plant species to aid in the microarray normalization steps, plus numerous condition-specific marker genes to serve as positive controls. Wilson et al. (2003) proposed that such boutique arrays should include a set of invariant housekeeping genes for accurate data normalization purposes. Ideally, invariant genes should be chosen such that the

full range of observed expression levels for the genes of interest is uniformly covered. Wilson et al. (2003) removed intensity-dependent nonlinear trends in the data by applying loess normalization to the invariant genes, and then transforming the variable genes of interest by interpolating from the flanking invariant genes. Genes whose expression fell outside the invariant genes range could be normalized via linear extrapolations off the lower and upper edges in the transformation curves. Sato et al. (2007) also used a similar approach for their mini-array, the Stable-gene's Based Quantile (SBQ) normalization uses a quantile normalization scheme in place of the loess normalization used by Wilson et al. (2003). In this study, we have shown that the SBQ method performs superbly, closely mirroring the expression values that RMA produces (when the latter is used with sufficient additional genes to avoid violation of its central assumption of similar intensity distribution across conditions).

Two additional normalizations were examined: RIMS and RMAPS. Both of these methods perform standard RMA separately within biological replicates, and then scale the values obtained between these independently normalized sets differently. RIMS scales each biological replicate set using the median expression value of the set of ~170 invariant genes. RMAPS uses the median expression value of the set of genes that are “Absent” (MAS 5.0, Liu et el., 2002) across all conditions in the study. SBQ proved to be superior to both these methods, but not by a large margin. Also, both RIMS and RMAPS have the significant drawback that they artificially reduce the within biological-group variance, and correspondingly accentuate the between-group variance. Hence if

accurate differential expression between groups is sought, use of these methods should be discouraged.

A small set of *DEFL*s are pathogen-responsive in both the model systems

We earlier reported that *DEFL*s generally share common features in their gene structure and genomic organization to defensins, which make up one sub-family within DEFLS (Graham et al., 2004; Silverstein et al., 2005). If some additional sub-families of DEFLs have antimicrobial activity as do the defensins, we would expect DEFLs would be up-regulated in response to pathogens. To examine this hypothesis, we profiled the expression of *DEFL*s across various mutants of pathogens and plants across different time points in Arabidopsis and a few selected pathogen treatments in Medicago. In Medicago, very few DEFLs were expressed in roots and leaves when they were infected with Phytophthora and Colletotrichum pathogens respectively.

In Arabidopsis, a few novel *DEFL*s were found to be expressed in pathogen-infected tissues. Through co-expression analysis with known marker genes for early defense signalling and JA-mediated host response, we also identified additional pathogen-responsive Arabidopsis *DEFL*s. Plants activate several mechanisms of defense responses upon recognition of invading pathogens. Several classical defensins with demonstrated protease inhibitory or microbicidal activity have been shown to be induced by pathogens via jasmonate signalling pathway (Penninckx et al., 1998; Thomma and Broekaer, 1998; Glazebrook, 2005). Of the 15 plant defensins previously described, *PDF1.1*,

PDF1.2a-c, *PDF1.3* and *PDF1.4* are induced by pathogens and showed co-regulated expression patterns across a variety of morphological structures and pathogen infected leaves. Transcript abundance of *PDF2.1*, *PDF2.2*, *PDF2.3* & *PDF2.5* were down-regulated by pathogen infections, but showed higher transcript abundance in seedlings, roots and inflorescence tissues. These expression trends of classical *DEFL*s were consistent with previous reports that showed the upregulated expression of *PDF1.2* by infection with *A. brassicicola*, while *PDF2.2* was down-regulated by the same pathogen (Thomma and Broekaer, 1998).

Plant resistance against necrotrophic pathogens such as *Botrytis cinerea* and *A. brassicicola* appears to be related to the JA-mediated signalling, leading to the activation of pathogenesis-related (*PR*) genes including classical plant defensins (van Wees et al., 2003; AbuQamar et al., 2006). The use of the *dde2-2* mutant with impaired JA synthesis and increased susceptibility to the nectrophic fungus in our analysis is expected to provide baseline data concerning the transcriptional regulation of *DEFL*s in defense responses and allow us to identify those *DEFL*s that may show co-expressed patterns of gene expression with known JA marker genes. Our microarray results, based on the *dde2-2* mutant challenged with *A. brassicicola* infection that invokes the JA-mediated defense response, as well as co-expression analysis with known marker genes for the JA-mediated pathway are highly suggestive that a few novel and classical *DEFL*s are induced via JA-mediated signalling. In fact *PDF1.2a-c* (At5g44420/ At5g44430/At2g26020), are commonly used marker genes for jasmonate

signalling mediated host defense. Interestingly, our list of interesting genes included *HEL1* (AT3G04720) and *COR1* (AT1G19670) that are commonly used marker gene transcripts for jasmonate signalling mediated host defense. The fact that these *DEFL*s and *PDF*s were found to be coregulated with two of the known JA-signalling marker genes (*PDF1.2a* & *COR1*) used in our analysis suggested that a few of the novel *DEFL*s are putatively involved in plant defense and their expression is related to genes associated with JA-signalling defense pathway. Nevertheless, future experiments using several Arabidopsis mutants including the JA signalling mutant *coi1* and ET signalling mutant *ein2* should help clarify the role of *DEFL*s in JA mediated host defense.

Complex transcriptional reprogramming involving a large number of genes has been shown to be involved during PTI and ETI. The *Pto* DC3000 used in this study causes necrotic lesions that are often surrounded by chlorotic halos in susceptible tomato and Arabidopsis plants (Katagiri et al., 2002). We also used two Pto DC3000 virulence mutants to characterize *DEFL* transcriptional responses as proxy to PTI and ETI. In total, our global At*DEFL* gene expression analysis identified 37 Arabidopsis *DEFL*s that are differentially regulated in response to bacterial pathogen inoculations. We also found that considerably more *DEFL*s were responsive to *Pseudomonas* treatments at 9 hpi compared to either 3 hpi with *Pseudomonas* treatments or 24 hpi with *Alternaria* inoculation. Our microarray data suggested that several novel and classical *DEFL*s may be important in PTI and/or ETI. Furthermore, co-expression analysis using the MAMPs induced marker gene, *FLK1*, suggested that the expression of up to nine

*DEFL*s may be related to early defense signaling to a hemibiotrophic pathogen. Identification and characterization of pathogen-responsive *DEFL*s provide a molecular signature for Arabidopsis basal defense to plant pathogenic bacteria.

The other putative roles of *DEFL*s in reproductive regulation and symbiotic microbial associations

Since only small subsets of DEFLs are expressed in pathogen treatments, we hypothesize that DEFLs have additional/ alternative functions in both Arabidopsis and Medicago. Among our various treatments, it is seen that the Arabidopsis *DEFL*s are primarily expressed in inflorescensces which is consistent with earlier observations of previously known *DEFL*s (Meyers et al., 2004; Silverstein et al. 2007; Jones-Rhoades et al., 2007). More than 100 *DEFL*s are specifically expressed in flower buds and/or siliques. Some of the *DEFL*s were shown to be expressed solely in siliques (Meyers et al., 2004) or ovules (Jones-Rhodes et al., 2007) detected via MPSS or whole-genome tiling arrays. The recent findings in *Torenia fournieri* (Okuda et al., 2009), and the specific expression of Arabidopsis *DEFL*s in synergid cells that were found to be the origin of pollen tube attractants in that species, suggest that some of the Arabidopisis DEFLs are likely to be pollen-tube attractants (Dresselhaus and Marton, 2009).

Highly polymorphic, rapidly evolving endogenous peptides such as DEFLs serve as ideal molecular signals, especially in the area of reproductive regulation. DEFLs have been co-opted for the purpose of preventing self-pollination in

Brassicacae (Shiba et al., 2002), in which the defensin-like SP11 (a.k.a. SCR) was shown to bind to the S-locus receptor kinase to initiate a cascade of events that leads to self-pollen rejection. DEFLs have also been co-opted to erect species barriers. In *Torenia*, a pair of DEFLs has been shown to serve as the elusive molecular signal exuded from the female gametophyte that attracts pollen tubes to the embryo sac (Okuda et al., 2009). These peptide chemoattractants have no effect on pollen tubes from other species. Interestingly, in mammalian systems, DEFLs have been shown to erect species barriers as well (Zhou et al., 2004). We are currently carrying out GeneChip hybridization experiments involving wild type and mutant Arabidopsis ovules, along with various stages of embryos, on the At*DEFL* custom array to further characterize and document the diversity of reproductive-specific *DEFL* expression.

From our custom chip analysis and previous in silico analysis of *Medicago* ESTs have shown that large numbers of Medicago *DEFL*s are expressed almost exclusively in nitrogen-fixing nodules (Mergaert et al., 2003; Graham et al., 2004).Recently, it has been reported for a few DEFLs found in nodules (called nodule-specific cysteine-rich (NCR) peptides in that study), that some nodule DEFL peptides similar to some defensins, triggered membrane modifications and inhibited cytokinesis in bacteria (Van de Velde et al., 2010). In addition, ectopic expression of these peptides in legume species lacking these peptides and cultured rhizobia triggered terminal differentiation. Based on these findings and additional reports of fungal hyphae branching on application of defensin peptides (Spelbrink et al., 2004), it may be speculated that the DEFLs might induce similar

changes in mycorrhizal hyphae growth.To further characterize the nodule DEFLs, we are conducting in-depth examinations of the expression patterns of nodule *DEFL*s using different symbiotic bacterial mutants and developmentally arrested nodules.

The custom AtMt*DEFL* array reveals comprehensive and specific spatio-temporal patterns of *DEFL* gene expression that differ markedly between the two model plant species

Our study revealed that *DEFL* expression in Arabidopsis and Medicago plants is highly condition specific, with subgroups of *DEFL*s expressed in different morphological structures or conditions. The patterns of *DEFL*s expression revealed notable differences in *DEFL* expression specificity in the two model plant species. Arabidopsis and Medicago are the sister clades within the dicot subclass Rosidae. Medicago belongs to the Fabaceae or legume species family and lacks an extensive genomic macrosynteny with Arabidopsis, especially the gene families involved in symbiotic nitrogen fixation lack homologs in Arabidopsis (Zhu et al., 2003). Our earlier in silico genomic analysis of the two species revealed that the majority of *DEFL*s are present in the reproductive parts of Arabidopsis and additionally in nodules in Medicago (Graham et al., 2004; Silverstein et al., 2005). As per our predictions, the microarray results reveal that most Arabidopsis *DEFL*s were expressed in inflorescences before anthesis, whereas in Medicago, majority is expressed in the nitrogen fixing nodules. It is surprising to see that in Medicago only a single *DEFL* was expressed in flower

buds or open flowers. A modest-sized subset (14 *DEFL*s) of At*DEFL*s were specifically expressed or exhibit enhanced expression in roots, yet similar root-specific or root-enhanced *DEFL*s were absent from Medicago. Another subset of *DEFL*s was specifically elicited in roots that have been colonized by mycorrhizal fungi, yet this type of symbiotic association is non-existent in Arabidopsis. The one commonality across both species was the large number of *DEFL*s expressed in seeds of Arabidopsis and Medicago. These results raise an intriguing question about the evolution of *DEFL*s in dicots and their diverse functions. As discussed above, one of the functions of Arabidopsis DEFLs in inflorescences might be to prevent self pollination. It can be hypothesized that in Medicago, a self pollinating plant, the DEFLs were lost from the flowers and a novel set of DEFLs appeared in the nodules.

In conclusion, our custom chip provides a rich source of information on the expression patterns of the large family of DEFLs in the two model systems. Further studies on the sub families found in Arabidopsis and Medicago using reverse genetic tools will be useful in understanding the diverse functional roles of DEFLs. Additionally, comparative genomic studies will add a new perspective on the evolution of this interesting gene family among dicots.

Materials and Methods

AtMt*DEFL* array design features and description

The AtMt*DEFL* array includes one probe set each for all 317 Arabidopsis *DEFL*s and 15 *DEFL*-related Maternally Expressed Gene (*MEG*) family members

(Gutierrez-Marcos et al., 2004; Silverstein et al., 2007). It also includes 684 probe sets (some probe sets exactly matched an entire subfamily of related *DEFL*s) representing all 756 Medicago *DEFL* genes that are not pseudogenes previously identified (Silverstein et al., 2007). The custom chip also includes probe sets of 31 Arabidopsis genes to serve as markers for expression in particular morphological structures and/or biological processes (e.g., pathogenesis-related proteins *PR-1* through *PR-5*; *ABI1*; *EIN2*; seed- and floral-specific genes). There were also 23 Medicago probe sets that were expected to be specific to seeds, other morphological structures, or to pathogen induction, based on the MtDB2 query interface (Lamblin et al., 2003). In addition, the custom chip included 171 probe sets directly from the Arabidopsis ATH1 array, which we have designated as 'invariant genes'. These probe sets were chosen because they had a coefficient of variance (cv) of less than 10% across a collection of more than 5,000 AtGeneExpress array hybridizations (http://arabidopsis.org/info/expression/ATGenExpress.jsp).

Similarly, 172 invariant probe sets were taken directly from the Affymetrix Medicago Genome Array, because they showed less than 10% cv across more than 100 GenChip hybridizations (He et al., 2009). Beyond these core invariants, 60 Arabidopsis probe sets were also included that had less than 3% cv across select tissue subsets. However, 20 Medicago invariant probe sets, chosen based on the sparse set of experimental conditions available at the time of custom chip design (March, 2007), failed the 10% cv threshold when their expression profiles were re-analyzed against the 100 Medicago hybridizations

available after the custom chip was produced. Probe sets for marker genes and invariant genes were designed exactly the same way as they appeared on the standard Affymetrix ATH1 and Medicago Genome arrays.

Probe sets from the two species are interspersed on the array although separate chip description files (cdfs) were created at Affymetrix with the intent that hybridization and analysis of the array would be performed using only one species at a time. All probe sets were represented by 11 perfect match (PM) and 11 single mismatch (MM) probes, as was typical of Affymetrix array designs. In general, cross-hybridizing probe sets are designated with '_x_at' suffixes at Affymetrix, and one should exercise caution with the expression values of these probe sets.

Arabidopsis plant material and growth conditions

Arabidopsis thaliana (L.) Columbia (Col-0) and *dde2-2* line (AT5G42650) in the Col-0 background (von Malek et al. 2002) were used in this study. Seeds were imbibed and stored at 4°C for 3 days before planting in an appropriate growth medium. Arabidopsis plants used for harvesting inflorescences (-2,-1 and 0 dp anthesis), siliques, and pathogen-treated leaves were grown in SUNGRO sunshine potting mix (SUNGRO Horticulture, Pine Bluff, AZ). For harvesting root tissues, Arabidopsis plants were grown in vermiculite (SUNGRO Horticulture). Seeds for seedling tissue harvest were sterilized by treating with 10% (v/v) bleach for 20 min, rinsed in sterile water several times and kept at 4°C for 3 days before planting on rectangular plastic plates containing 1x Murashige and Skoog salt mixture supplemented with 1% sucrose in 1% agar. Plates were kept vertical

in a growth chamber and seedlings were collected 7 and 14 days after sowing. All Arabidopsis experiments were conducted in growth chambers set at light intensity of approximately 170 µmol m^{-2} s^{-1}, temperature of 22° C, with a 12-h photoperiod. Plants were watered with tap water as needed. Three biological replicates were produced for each treatment.

Pathogen infection of Arabidopsis plants

Alternaria brassicicola strain ATCC96866 was cultured on 0.5X potato dextrose agar (PDA) at 22 °C. Spores were harvested from 10-day-old PDA plates by washing with water and filtering through four layers of cheesecloth. Spore concentrations were determined using a hemocytometer. For pathogen challenge, leaves of 21-day-old Arabidopsis plants were inoculated with *A. brassicicola* spores according to Nafisi et al. (2007). Briefly, the third, fourth, and fifth true leaves of each plant were inoculated for monitoring disease symptoms as well as for microarray analysis of DEFL gene expression. The three inoculated leaves from each plant were collected and bulked for RNA isolation. The mock-control plants were treated in a similar manner with sterile water containing 0.02% Tween-20 used as inoculum. Leaf samples were collected 24 h after inoculation and were immediately frozen in liquid nitrogen for storage at -80 ° C until needed for total RNA extraction.

For experiments with bacterial pathogens, virulent *Pseudomonas syringae* pv. tomato DC3000 (*Pto* DC3000), a derivative of *Pto* DC3000 lacking *hrcC* (*Pto* DC3000 *hrcC*-, an avirulent strain), and *Pto* DC3000 carrying *AvrRpt2* (*Pto* DC300 *AvrRpt2*) were cultured as described previously (Glazebrook and

Ausubel, 1994; Wang et al., 2008; 2009). Four- week- old Arabidopsis plants were inoculated with bacterial strains essentially as described before (Wang et al., 2008). Control plants were mock-inoculated with 5 mM $MgSO_4$. Leaves were collected 3 and 9 h after bacterial inoculations, and were immediately frozen in liquid nitrogen for storage at -80℃ until needed fo r total RNA extraction.

Medicago plant culture and tissue harvest

Seeds of *Medicago truncatula* genotype A17 were sterilized and germinated as described previously (Lohar et al. 2006). Unless indicated otherwise, Medicago plants were grown from 24 h germinated seedlings transplanted in 6 inch pots containing an equal mixture of Turface (Profile Products LLC, Buffalo Grove, IL) and metro mix (SUNGRO - 200 series) and fertilized with Osmocot PLUS (15-9-12). All Medicago experiments were conducted in controlled growth chambers set at 16 h light / 8 h day, 25-21°C, 200-300 µmol m^{-2} s^{-1} and 50% relative humidity. All Medicago tissues and organs were collected from three biological replicates, with samples collected and pooled from multiple plants in each replicate. Plant samples were immediately frozen in liquid nitrogen during harvest and were stored at -80° C for subsequent RNA extraction.

Medicago flower buds that were 6-8 mm in size (measured from the base of the bud excluding the petiole to the tip of the sepals), tightly closed, and whose petals had not emerged beyond the sepals were collected from 80-day-old plants. For flower samples, closed flowers (with petals bigger than the sepals) and open flowers were harvested from 80-day-old plants and pooled. For seed

samples, flowers were tagged on the date of opening as described by Wang and Gussak (2005), and a pool of seeds were collected 10-21 days after pollination (dap) from 98 day old plants. Seedlings, with radicles 8 – 12 mm long, were collected at 24 h post-germination. Stem segments, consisting of the 5th and 6th internodes from the shoot apex were collected, from 8-week- old Medicago plants as described (Tesfaye et al. 2009).

To obtain nodule tissues, plants were grown on sterile plates (Corning 245 mm x 245 mm dishes) containing buffered nodulation medium (BNM; Ehrhardt et al., 1992), pH 6.5, 1.2% agar (Plant Tissue culture grade, Sigma). The radicles of sterile germinated seeds were placed on a moist, sterile germinating paper and wrapped with a sterile black cotton cloth (Cotton Club Black, #074300603820, Wal Mart). Plates were maintained vertically in a growth chamber as described above. At 5 d after planting, plants were inoculated with 100 µL/root of a washed suspension of *Sinorhizobium meliloti* Sm1021 (Meade et al., 1982) [OD600] = 0.05 in sterile water. Control plants were mock-inoculated with 100 µL/root of sterile water. At 14 dpi, approximately 5 cm-long, nodule-bearing root segments were harvested from inoculated plants. At the same time, control roots consisting of approximately 5-cm root segments corresponding to the regions harvested for 14 dpi nodules were collected from mock-inoculated plants. Root tips were removed from all plants at the time of harvest to reduce transcript contribution from the RNA-rich, root meristematic cells.

For mycorrhizal roots, germinated seedlings were transplanted into Turface (Profile Products LLC), and were grown in growth chambers. *Glomus intraradices* spores were prepared from plates as previously described (St-Arnaud et al., 1996). Twenty-one-day-old plants were inoculated with 1,000 *G. intraradices* spores per plant as described previously (Liu et al., 2003). Control plants were mock-inoculated with distilled water. At 28 dpi, mycorrhizal roots without the root tip were harvested and immediately frozen in liquid nitrogen. Roots of mock-inoculated plants were similarly harvested. Throughout the experiment, Medicago plants were fertilized with 100 mL/pot of ½ strength Hoagland solutions with 1x nitrogen, as needed.

For *Phytophthora*-infected roots, germinated seeds were placed in CYG seed germination pouches, (Mega International, West St. Paul, MN), with six seedlings/pouch, and then placed in a growth chamber set at conditions as described above. After 5 days, each growth pouch was inoculated with 2 ml of inoculum consisting of comminuted mycelium from a 7-day-old culture of *P. medicaginis* M2019 grown on V8 agar (1 g culture/5 ml water) at room temperature. For inoculum preparation, the agar with the mycellium was cut, weighed and homogenized in 0.5X weight distilled water using a blender. Control plants were mock-inoculated with 2 ml sterile water. All plants were fertilized every 2 days with 2 ml per growth pouch of 5x Peter's Lite (10:10:10). At 4 dpi, infected hypocotyl and root segments were harvested from both the infected and mock inoculated plants.

For *Colletotrichum*-infected Medicago leaves, pre-germinated seeds were transplanted in pots containing an equal mixture of Turface (MVP) and metro mix (SUN GRO- 200 series), kept in growth chambers described above. *C.trifolii* was cultivated on potato dextrose agar (Difco) plates to collect a spore suspension for inoculation. The *Colletotrichum* spore suspension @10^6 /ml with a drop of Tween 20 was sprayed onto the leaves of 19 day old Medicago plants until run off. Control plants were mock-inoculated by spraying leaves with sterile water containing a drop of Tween 20. After 24 hours, leaves from pathogen-inoculated or mock-inoculated plants were collected for RNA extraction.

RNA extraction and GeneChip hybridization

Approximately 150 mg of Arabidopsis or Medicago tissue ground in liquid nitrogen was used for total RNA extraction. Total RNA from Medicago seeds and germinating seedlings was extracted using the Spectrum Plant Total RNA Kit (Sigma) following the manufacturer's instructions. Total RNA from Arabidopsis siliques was extracted using the hot phenol protocol. Total RNA from all other Arabidopsis and Medicago tissues was extracted using TRIZOL reagent (Invitrogen, Carlsbad, CA) following the manufacturer's instructions. During the RNA extraction, contaminating genomic DNA was removed by incubating samples with DNase TURBO™ DNA-free Kit following standard procedures suggested by the supplier (Applied Biosystems). The integrity and quality of total RNA was verified using the Agilent 2100 Bioanalyzer RNA 6000 Nano LabChip (Agilent Technologies, Santa Clara, CA). Ten micrograms of total RNA was used to produce biotin-labeled cRNA using Affymetrix kits following the manufacturer's

suggested procedures for 1-cycle eukaryotic reactions, except for siliques (Affymetrix, Santa Clara, CA). For Arabidopsis siliques, 100 ng of total RNA was labeled using the 2-cycle labeling procedure (Affymetrix). Ten micrograms of biotin-labeled cRNA, fragmented as suggested by Affymetrix, was hybridized to the AtMtDEFL custom Array. The integrity and quality of labeled and fragmented biotin-labeled cRNA was verified using the Agilent 2100 Bioanalyzer RNA 6000 Nano LabChip (Agilent Technologies). GeneChips were hybridized, washed, stained and scanned as previously described (Tesfaye et al. 2009).

Microarray data normalization and analysis

Microarray normalizations and initial analyses were performed in R using custom scripts that made use of Bioconductor routines (Gentleman et al., 2004). All scripts are available upon request. In addition to RMA (Irizarry et al., 2003) from the Bioconductor package, three microarray data normalization methods were compared: (1) To implement the Stable-genes Based Quantile (SBQ) normalization (Sato et al., 2007), we first combined expression data from all 11 probes within a probe set into a single expression measure using the Bioconductor expresso command, using RMA-style background correction and median polish probe summarization. These summarized values were output from R and used as input to the SBQ perl script (Sato et al., 2007); (2) For RMA with Median Absent Probeset Scaling (RMAPS), we performed RMA separately on each group of biological replicates (within which the density distribution assumptions hold) using the RMA routine within the Bioconductor package. We then rescaled the $\log_2$ expression levels for each biological replicate group such

that the median values among the set of probe sets that had an absent call (MAS5 calls routine in the affy package) across all sampled conditions were equal from one replicate group to the next (3) For RMA with Invariant Median Scaling (RIMS), we performed RMA separately on each group of biological replicates as above. Then we rescaled the expression levels for each replicate group according to the median value of the set of invariant genes.

We also calculated a detection call ('present' vs 'absent') for each probe set using default MAS5 parameters, except changing α_1= 0.05 as suggested for 11 probe pairs per probe set by Liu et al. (2002). For statistical analysis, SBQ normalized signal intensity values were Log_2 transformed and statistical analysis was based on a t-test or ANOVA ($P<0.05$) with Benjamini and Hochberg false discovery rate multiple testing correction (Benjamini and Hochberg, 1995), and the corrected p values were designated as q values. To identify *DEFL*s and marker genes that showed co-expression patterns, Pearson correlation coefficients were calculated for each probe set to identify co-regulated genes. Co-expression analysis and Pearson correlation coefficients were calculated using the GeneSpring Expression analysis platform version 7.3 (Agilent Technologies). For gene expression data clustering analysis, the average linkage clustering algorithm and the Pearson correlation similarity measure were applied as provided in GeneSpring Expression analysis software version 7.3 (Agilent Technologies). Heatmaps were also generated with the heatmap2 Bioconductor package.

Real-time RT-PCR assays

We employed quantitative reverse-transcription polymerase chain reaction (qRT-PCR) to validate the *DEFL* microarray data trends that were collected from several Arabidopsis tissues and pathogen challenged leaves. Initially, we evaluated the expression patterns and detection calls of candidate invariant genes included in the custom chip. Based on the microarray data collected using the custom AtMt*DEFL* chip, we chose ten of the invariant genes for further evaluation of gene expression data stability using geNorm software (Vandesompele et al. 2002). Genes were selected to cover a range of signal intensity values. The raw signal intensity values were imported into geNorm software for gene expression stability analysis. Based on their stability score, we then selected four reference genes (metallothionein 2B (AT5G02380), yellow-leaf-specific gene 8 (AT5G08290), sumo conjugating enzyme 1 (AT3G57870), and polyubiquitin (UBQ11)) for validation of their gene expression stability using RT-PCR assays. Gene-specific primers were designed using the PrimerExpress software (Applied Biosystems, Foster City, CA.). The list of genes and primer sequences used for the qRT-PCR analysis and microarray data validation are shown (Supplemental Table 2-S7). Gene expression stability analysis was performed using the average threshold cycle (Ct) of the four candidate reference genes and the geNorm software (Vandesompele et al., 2002). Total RNA extraction procedures were as described above and first-strand cDNA was prepared from 2 µg of total RNA with the Superscript RT II kit (Invitrogen) and oligo dT primers (Sigma-Aldrich) at 200 ng/reaction, according to the manufacturer's instructions. Real-time PCR conditions were as described

previously (Tesfaye et al., 2006), with the following modifications: A PCR master mix using SYBR Green PCR and RT-PCR reagents (Applied Biosystems) was mixed with 1 µl of first-strand cDNA as template. Each assay also included a no-template control.

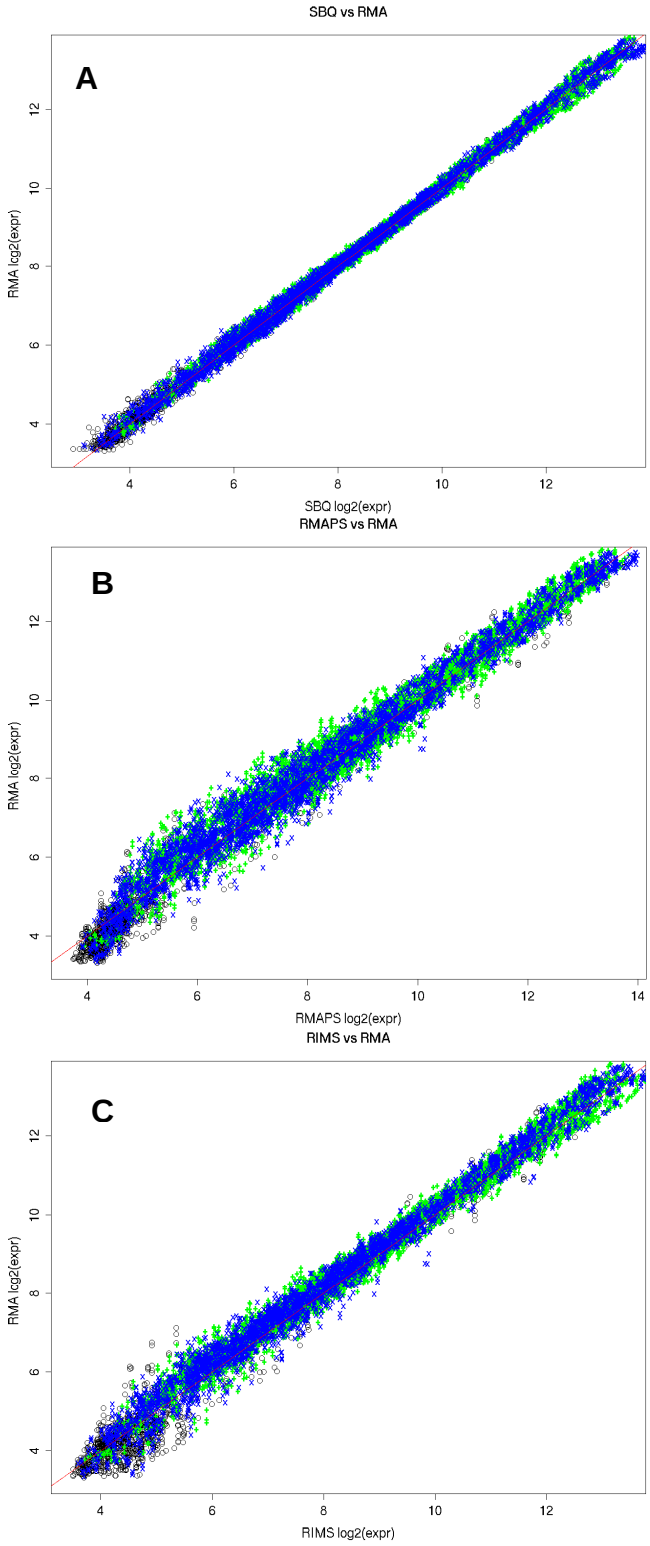


Figure 2-1 Scatter plots showing correlation of normalized expression values obtained from 3 boutique array methods as compared to whole-array RMA across 36 ATH1 arrays. (A) SBQ vs RMA; (B) RMAPS vs RMA; (C) RIMS vs RMA. The red line indicates the y=x axis. All three boutique array normalizations used only a subset of 299 probe sets that correspond to genes represented on the AtDEFL array (37 DEFLs - black circles, 171 invariants - green plus signs, 91 marker genes - blue crosses). Each was compared to the reference RMA normalization, which included all 22,810 probe sets on the Affymetrix Arabidopsis ATH1 array in the normalization process. Expression values have been log transformed (base 2). All (37) probe sets on the ATH1 array that matched an At DEFL with at least 6 of 11 probe sets were included in the analysis. The 36 arrays included 3 biological replicates for a wide variety of morphological structures and conditions obtained from GEO: GSE1491 (seedlings), GSE5630 (cotyledons, leaves, senescent leaves), GSE5631 (roots), GSE5632 (carpels, stage 9 flowers, stamens), GSE5633 (stems), GSE5634 (old and young siliques) and GSE7227 (seeds).

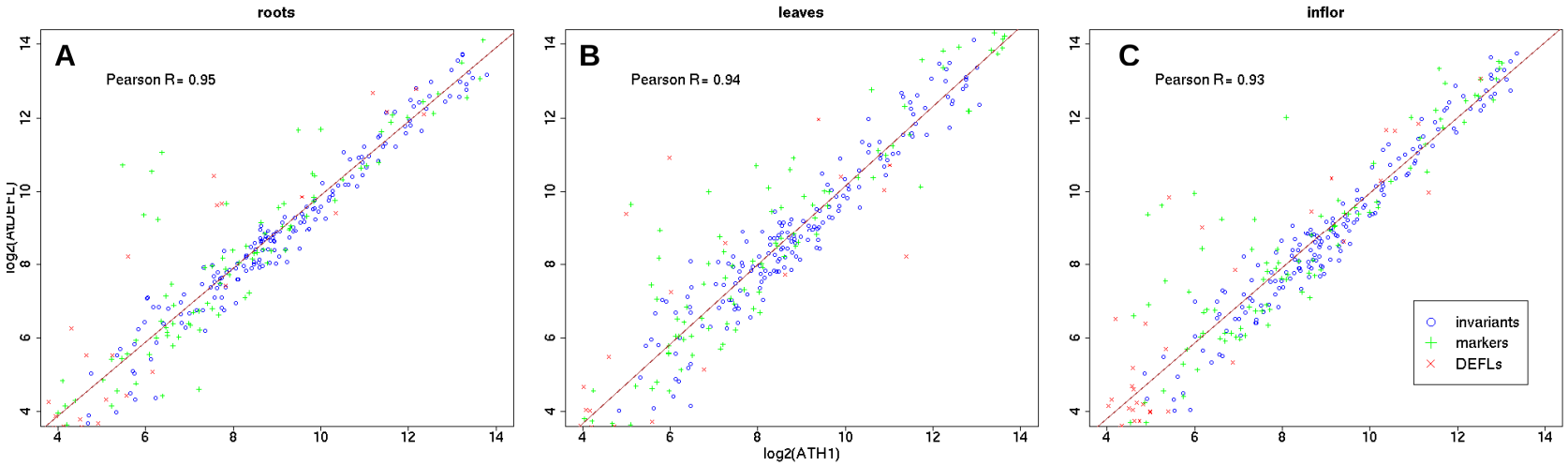


Figure 2-2 Scatter plots showing correlation of SBQ normalized AtDEFL array (y-axis) expression values obtained for 3 morphological structures as compared to GEO ATH1 arrays (x-axis) hybridized to similar biological specimens. (A) roots; (B) leaves; (C) inflorescences. The red line indicates the y=x axis. Probe sets for invariants (blue circles) and markers (green plus signs) are identical across the two platforms (though distributed in different physical locations on the respective arrays). The 37 DEFL probe sets (red crosses) were represented with different probes on the two arrays, with differences in some cases in the underlying gene model. Log2 signals for biological replicates for a platform were averaged prior to cross-platform comparison. GEO accessions for ATH1 arrays included: GSM131558, GSM131559, GSM131560 (roots); GSM131498, GSM131499, GSM131500 (leaves); GSM62694, GSM227612 (inflorescences).

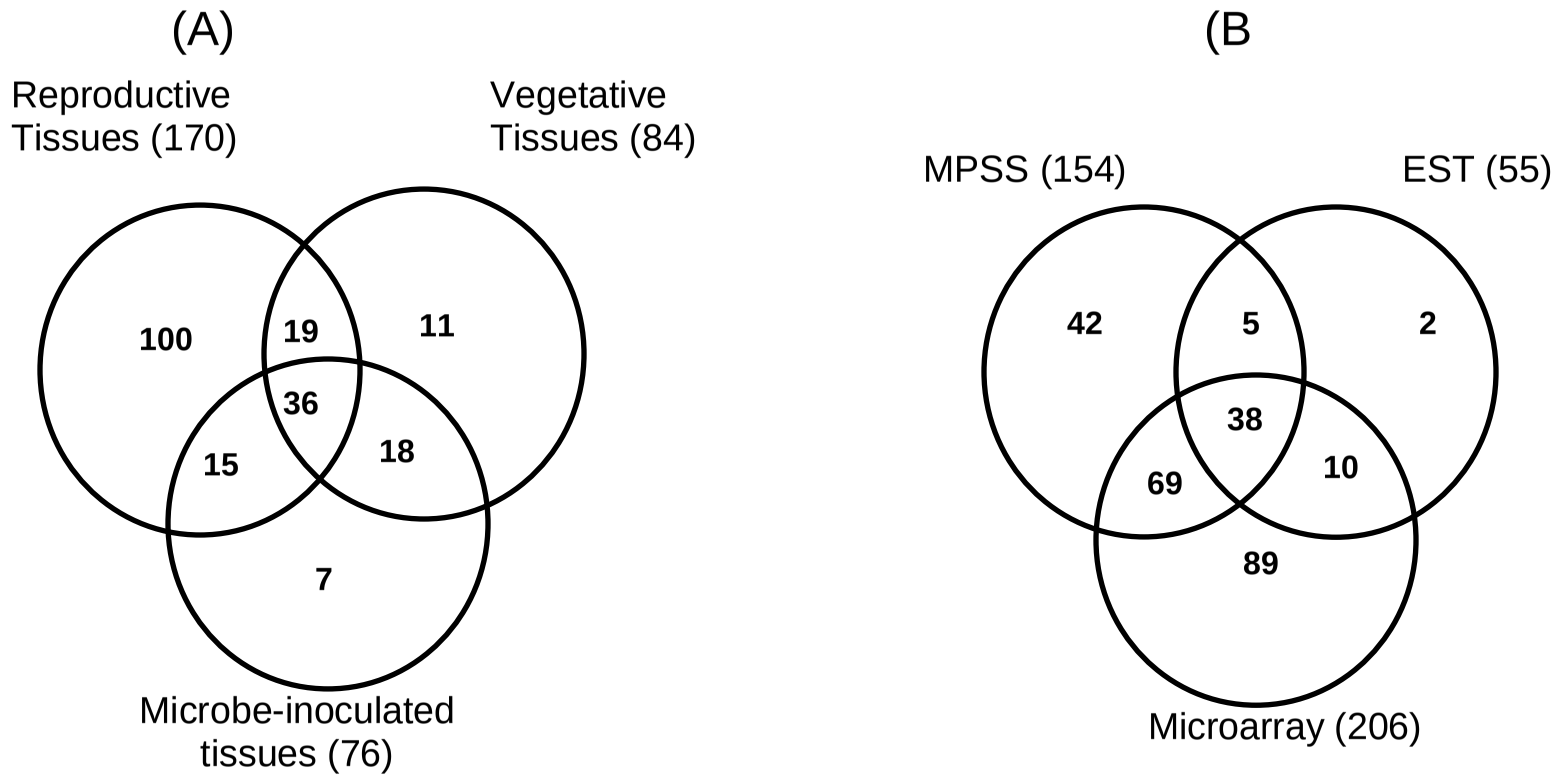


Figure 2-3 Venn diagrams of expressed DEFLs in different treatments. (A) Number of Differentially expressed *DEFL*s and *MEG*s genes in various tissues in Arabidopsis (B) Comparison of different techniques for uniquely identified expressed *DEFL*s in each technique.

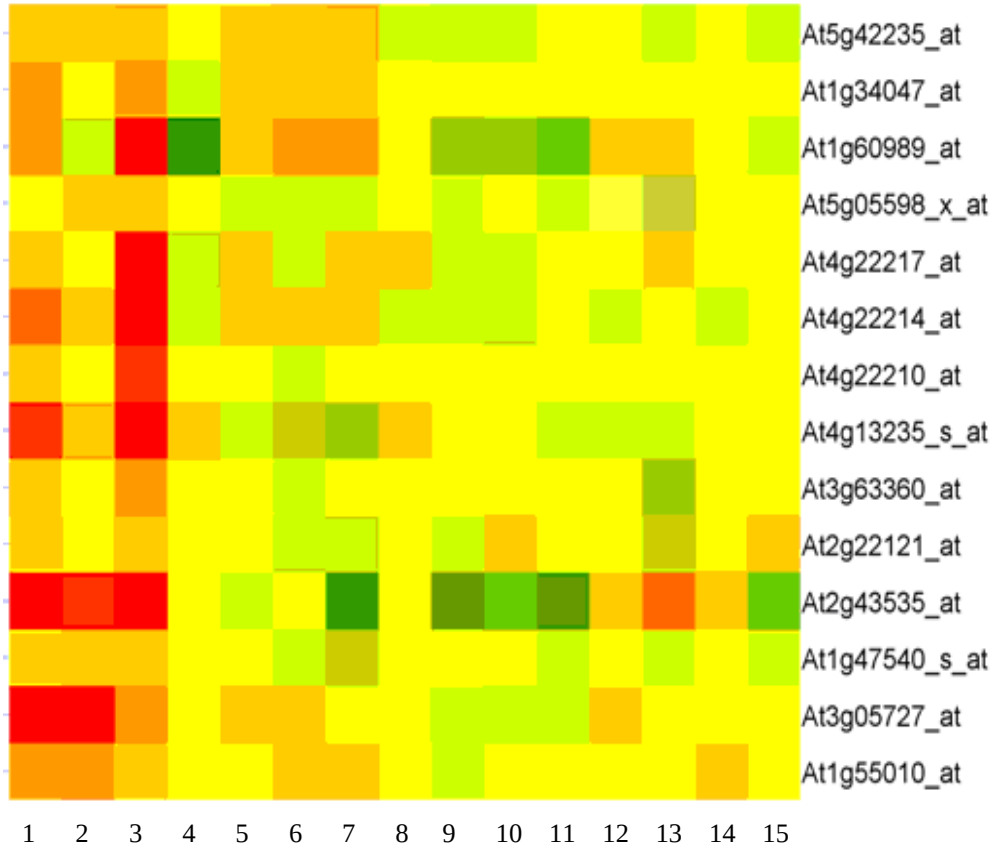
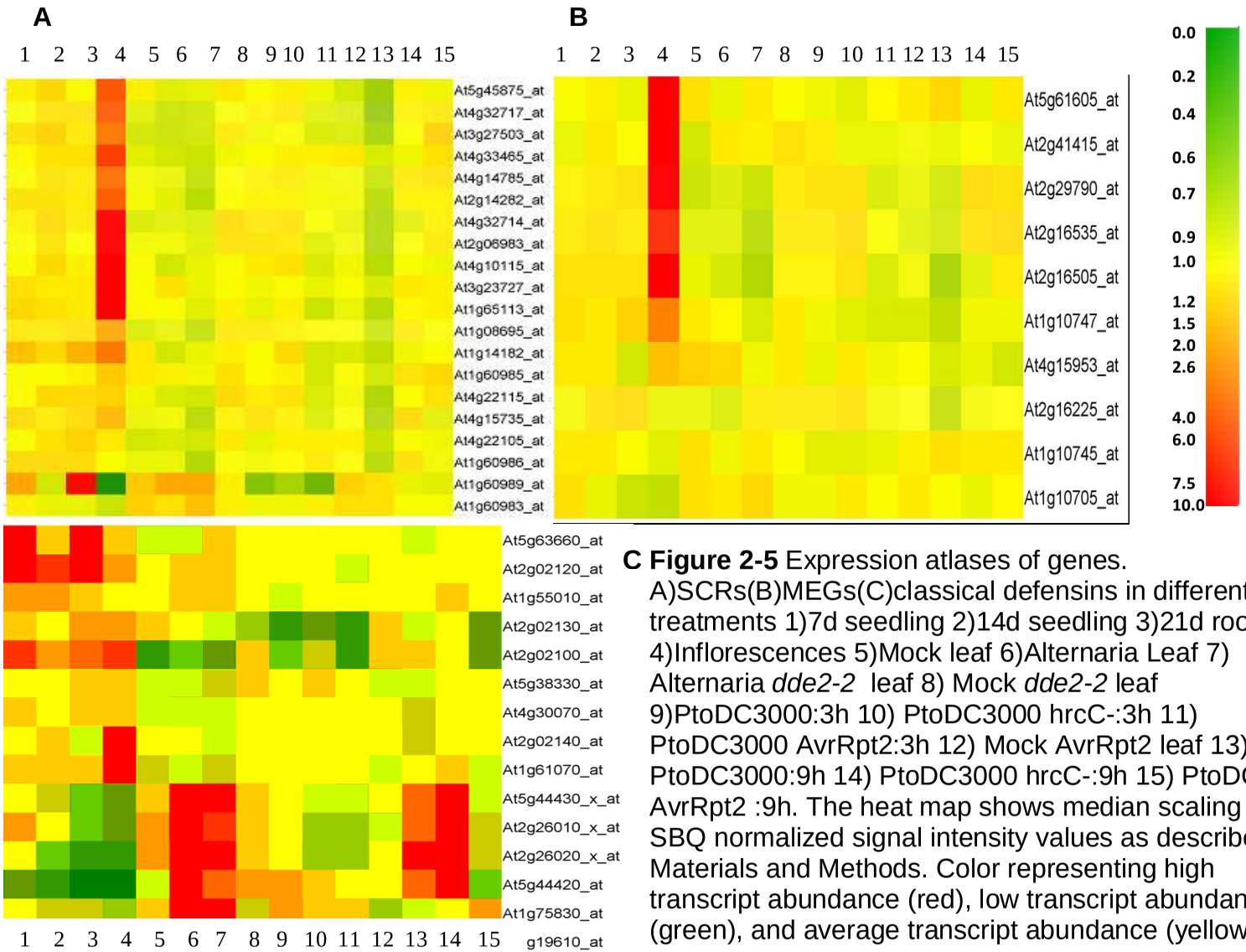


Figure 2-4 Transcript profiles of seeding and root specific/enhanced *DEFL*s in different treatments. 1)7d seedling 2)14d seedling 3)21d root 4)Inflorescences 5)Mock leaf 6)Alternaria Leaf 7) Alternaria *dde2-2* leaf 8) Mock *dde2-2* leaf 9)PtoDC3000:3h 10) PtoDC3000 hrcC-:3h 11) PtoDC3000 AvrRpt2:3h 12) Mock AvrRpt2 leaf 13) PtoDC3000:9h 14) PtoDC3000 hrcC-:9h 15) PtoDC3000 AvrRpt2 :9h. The heat map shows median scaling of SBQ normalized signal intensity values as described in Materials and Methods. Color representing high transcript abundance (red), low transcript abundance (green), and average transcript abundance (yellow).



C Figure 2-5 Expression atlases of genes. A)SCRs(B)MEGs(C)classical defensins in different treatments 1)7d seedling 2)14d seedling 3)21d root 4)Inflorescences 5)Mock leaf 6)Alternaria Leaf 7) Alternaria *dde2-2* leaf 8) Mock *dde2-2* leaf 9)PtoDC3000:3h 10) PtoDC3000 hrcC-:3h 11) PtoDC3000 AvrRpt2:3h 12) Mock AvrRpt2 leaf 13) PtoDC3000:9h 14) PtoDC3000 hrcC-:9h 15) PtoDC3000 AvrRpt2 :9h. The heat map shows median scaling of SBQ normalized signal intensity values as described in Materials and Methods. Color representing high transcript abundance (red), low transcript abundance (green), and average transcript abundance (yellow).

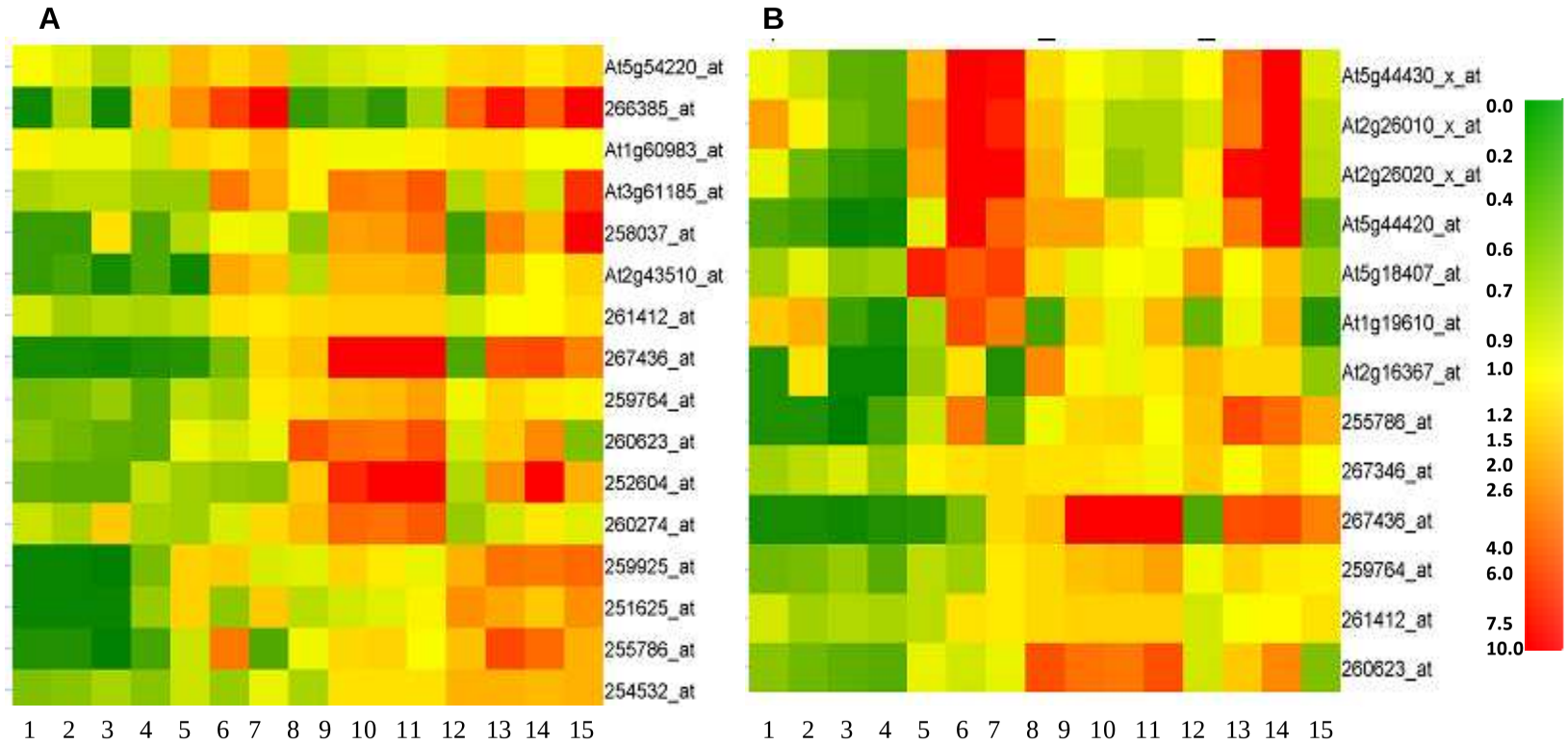


Figure 2-6 Expression patterns of Arabidopsis co-regulated *DEFL*s (A)receptor-like protein kinase (FRK1/AT2G19190) involved in early defense signaling (B) JA marker genes PDF1.2a (AT5G44420) and/or coronatine-insensitive 1 (COI1/AT2G39940) indifferent treatments1)7d seedling 2)14d seedling 3)21d root 4)Inflorescences 5)Mock leaf 6)Alternaria Leaf 7) Alternaria *dde2-2* leaf 8) Mock *dde2-2* leaf 9)PtoDC3000:3h 10) PtoDC3000 hrcC-:3h 11) PtoDC3000 AvrRpt2:3h 12) Mock AvrRpt2 leaf 13) PtoDC3000:9h 14) PtoDC3000 hrcC-:9h 15) PtoDC3000 AvrRpt2 :9h. The heat map shows median scaling of SBQ normalized signal intensity values as described in Materials and Methods. Color representing high transcript abundance (red), low transcript abundance (green), and average transcript abundance (yellow).

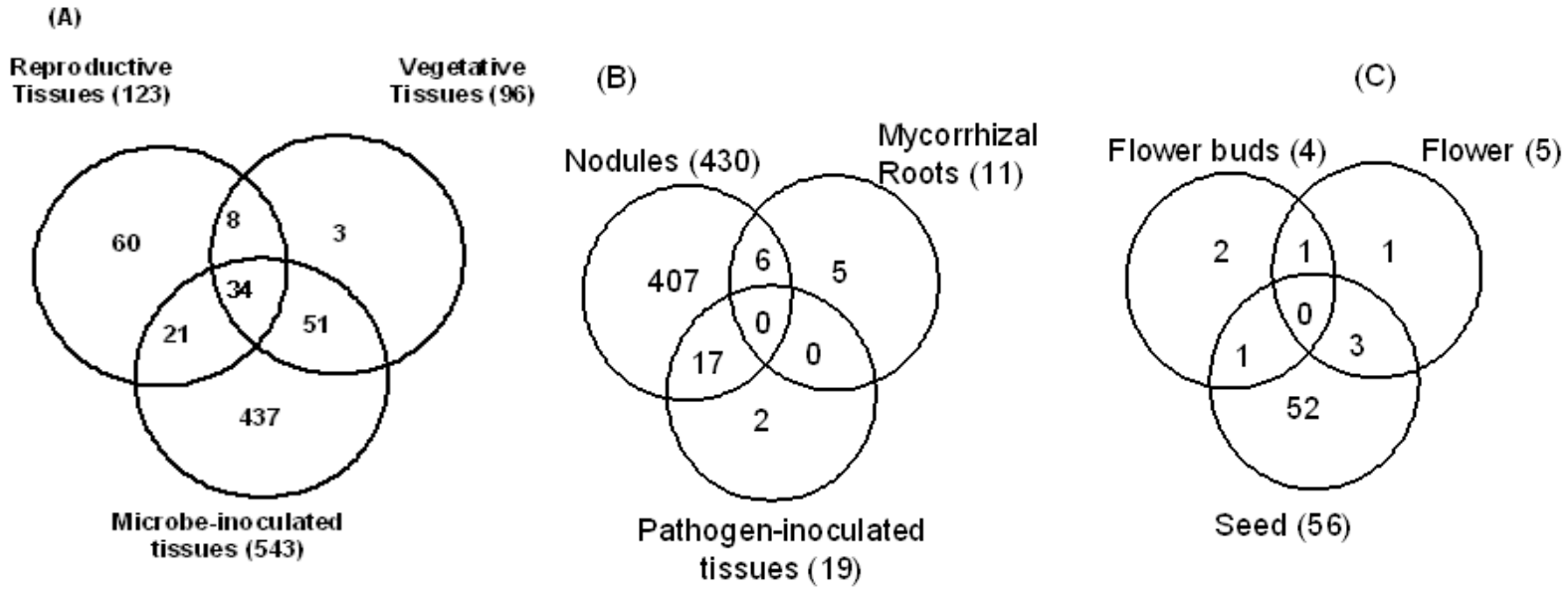


Figure 2-7 Venn diagrams of Medicago *DEFL*s expressed in different treatments. Comparison among (A) all tissue types (B) Tissues inoculated with symbionts and pathogens (C) all reproductive tissues.

Table 2-1 Differentially regulated Arabidopsis *DEFL*s after 24 hours post inoculation with *Alternaria brassicicola.*

AGI	probe set identifier	Annotation	COL_24/Mock	dde2-2_24/Mock
AT5G19890	246149_at	Peroxidase	9.77	0.95
AT5G39520	249454_at	Expressed protein	13.39	44.51
AT4G26080	253994_at	Abscisic acid (ABA) INSENSITIVE 1	1.60	2.73
AT1G19670	255786_at	Coronatine-induced protein 1 (CLH1, CORI1)	4.39	0.40
AT3G12500	256243_at	Basic chitinase (PR-3)	38.63	6.01
AT3G04720	258791_at	Hevein-like (HEL) (PR-4)	13.11	4.58
AT1G07890	261412_at	Ascorbate peroxidase 1 (APX 1)	1.51	1.48
AT1G75750	262947_at	GA-responsive GAST1 protein homolog (GAST1)	0.58	1.08
AT1G68735	At1g68735_at	DEFL	1.11	2.03
AT1G75830	At1g75830_at	PDF1.1, PDF1.2, LCR67	132.53	16.27
AT2G26010	At2g26010_x_at	PDF1.3	52.74	2.51
AT2G26020	At2g26020_x_at	PDF1.2b	66.50	4.42
AT2G43510	At2g43510_at	DEFL	20.49	16.37
AT2G43530	At2g43530_at	DEFL	1.14	0.30
AT2G43550	At2g43550_at	DEFL	0.87	0.21
AT3G05730	At3g05730_at	DEFL	0.53	0.55
AT3G59930	At3g59930_at	DEFL	43.43	3.46
AT5G33355	At5g33355_s_at	DEFL	70.67	2.30
AT5G44420	At5g44420_at	PDF1.2a, LCR77	36.78	4.59
AT5G44430	At5g44430_x_at	PDF1.2c	77.50	5.01
AT5G46871	At5g46871_at	DEFL	2.84	4.27

Table 2-2. Pathogen responsive Arabidopsis *DEFL*s

			Expression ratios in leaves inoculated with P. Syringae strains[a]					
AGI	**Probe set Id**	**Annotation**	**3-DC3000**	**3-hrcC-**	**3-AvrRpt2**	**9-DC3000**	**9-hrcC-**	**9-AvrRpt2**
Differentially regulated only at 3 hpi								
At1g11055	At1g11055_at	DEFL	1.2	1.2	2.1	0.89	1.28	1.33
At2g16505	At2g16505_at	MEG	1	1.1	0.8	0.71	0.96	1.14
Differentially regulated at 3 & 9 hpi								
At4g26080	253994_at	Abscisic acid-insensitive 1 (ABI1)	0.9	0.9	1.2	2.65	0.85	2.95
At1g19670	255786_at	Coronatine-induced protein 1 (CORI1)	1.2	1.3	1.1	3.85	2.86	1.26
At3g12500	256243_at	Basic chitinase (PR-3)	1.6	1.4	1.8	2.06	1.79	7.82
At3g04720	258791_at	hevein-like protein (HEL) (PR-4)	2.5	1.1	2.1	2.16	5.55	2.26
At1g75750	262947_at	GA-responsive GAST1 protein homolog	0.5	0.5	0.4	0.37	0.64	0.25
At1g75830	At1g75830_at	PDF1.1, PDF1.2, LCR67, LCR67	3.2	2.5	2.4	1.46	1.57	3.5
At2g26010	At2g26010_x_at	PDF1.3	0.6	0.4	0.4	4.25	20.65	0.93
At2g26020	At2g26020_x_at	PDF1.2b	0.5	0.3	0.4	9.1	31.66	0.73
At2g43510	At2g43510_at	DEFL	2.3	2.2	2.6	4.19	3.19	4.03
At2g43530	At2g43530_at	DEFL	0.6	0.9	0.6	1.23	1.03	0.47
At2g43550	At2g43550_at	DEFL	0.3	0.6	0.3	1.38	0.78	0.15
At3g05730	At3g05730_at	DEFL	0.2	0.4	0.3	0.62	0.62	0.44
At2g33070	245161_at	Nitrile-specifier protein	1.5	1.3	1.6	1.72	1.09	1.53

		NSP2						
At3g45060	252604_at	High-affinity nitrate transporter 2.6	5.2	9.2	17.9	3.72	18.51	2.62
At4g21960	254386_at	Peroxidase 42 (PER42) (P42) (PRXR1)	0.7	0.7	0.6	0.7	0.96	0.24
At4g19660	254532_at	NPR1-like protein 4 (NPR4) (NPR1)	1.7	1.6	1.7	0.99	0.89	0.99
At3g26520	257313_at	Tonoplast intrinsic protein 2	0.6	0.6	0.5	0.49	0.61	0.15
At3g21230	258037_at	4-coumarate::CoA ligase (4CL)	4	4.1	6.6	12.77	7.26	49.23
At3g06490	258516_at	MYB108 transcription factor	1.7	2	2.1	3.77	1.51	20.23
At1g64280	259764_at	Nonexpresser of PR genes 1 (NPR1) (SA insensitive 1)	1.3	1.4	1.9	1.32	1.12	1.13
At1g75040	259925_at	pathogenesis-related protein 5 (PR-5)	1.4	1.2	1	2.18	1.95	2.44
At1g80460	260274_at	glycerol kinase (GLR1).	2.5	2.3	3	1.36	1.73	1.53
At2g19190	267436_at	FLG22-induced Receptor-like kinase 1	10.1	9.3	13.4	16.85	18.16	11.06
At1g60989	At1g60989_at	DEFL	0.5	0.6	0.4	0.9	0.75	0.74
At2g16367	At2g16367_at	DEFL	0.3	0.3	0.4	0.71	0.72	0.36
At2g43520	At2g43520_at	DEFL	2.1	2.1	2.3	6.21	4.86	1.57
At2g43535	At2g43535_at	DEFL	0.3	0.4	0.3	2.7	0.83	0.32
At3g61185	At3g61185_at	DEFL	3.4	3.2	4.7	2.11	1.16	10.09
At5g18407	At5g18407_at	DEFL	0.7	0.8	0.8	0.43	0.69	0.29
At1g19610	At1g19610_at	pdf1.4,LCR78	4.6	3.3	6.2	2.19	4.14	0.39

At2g02100	At2g02100_at	pdf2.2,LCR69	0.3	0.5	0.2	0.66	0.6	0.2
At2g02130	At2g02130_at	PDF2.3,LCR68	0.4	0.5	0.3	0.65	0.81	0.27
Differentially regulated at 9 hpi								
At5g03280	250928_at	Ethylene-insensitive 2 (EIN2)	0.7	0.9	0.7	1	0.97	0.73
At3g57260	251625_at	beta-1,2-glucanase (PR-2)	1	1.1	1.3	0.77	0.56	1.1
At1g08090	260623_at	high-affinity nitrate transporter 2.1 (ACH1)	0.7	0.6	0.9	1.74	3.66	0.67
At2g14610	266385_at	pathogenesis-related protein 1 (PR-1)	1.2	0.6	2.3	2.21	1.11	6.66
At2g39940	267346_at	coronatine-insensitive 1 (COI1)	1	0.9	0.9	0.72	0.92	0.75
At5g39520	249454_at	Expressed protein	1	1	1.1	3.47	0.85	22.28
At1g07890	261412_at	Ascorbate peroxidase 1 (APX1)	1.1	1	1.1	1.19	1.24	1.43
At3g59930	At3g59930_at	DEFL	1.1	1.1	1.4	3.17	1.32	1.13
At5g33355	At5g33355_s_at	DEFL	1	1.1	0.9	2.15	1.4	0.98
At5g44420	At5g44420_at	PDF1.2a, LCR77	0.8	0.5	0.4	3.76	11.53	0.48
At5g44430	At5g44430_x_at	PDF1.2c	0.8	0.7	0.7	3.77	19.95	0.89
At5g46871	At5g46871_at	DEFL	1	0.9	1	1.5	1.59	6.02
At1g13609	At1g13609_at	DEFL	0.9	0.9	0.8	0.56	0.82	0.71
At1g49435	At1g49435_at	DEFL	1.1	1.1	1	0.76	1.19	1.18
At1g64195	At1g64195_at	DEFL	1	0.9	1	1.51	1.26	1.54
At1g69828	At1g69828_at	DEFL	0.9	0.9	1.1	1.19	1.21	1.59
At2g02147	At2g02147_at	DEFL	0.9	1.1	1.1	0.81	0.93	1.39
At2g04425	At2g04425_at	DEFL	1.1	1	1	0.7	1.05	1.14

At2g17723	At2g17723_at	DEFL	0.9	1	1	0.85	1.1	8.51
At3g13403	At3g13403_at	DEFL	0.8	0.8	0.9	0.38	0.59	0.27
At3g27503	At3g27503_at	DEFL	0.9	0.9	0.8	0.78	1.16	1.49
At4g10115	At4g10115_at	DEFL	1.1	1.1	0.8	0.78	1.09	1.06
At4g13955	At4g13955_at	DEFL	1	1	0.9	0.8	1.08	1.05
At4g22230	At4g22230_s_at	DEFL	1.2	1.5	1.3	1.49	1.41	3.72
At4g29280	At4g29280_at	DEFL	1	1.1	1	0.81	1.12	1.14
At5g42223	At5g42223_at	DEFL	1	1	0.9	0.89	1.18	1.53
At5g43285	At5g43285_x_at	DEFL	1.1	1	1.1	1.33	2.01	1.17
At5g55132	At5g55132_at	DEFL	1	1	0.8	0.63	0.98	1.17

[a] Expression Ratios are the relative values of *A. thaliana* Col leaves inoculated with the indicated *P. Syringae* strains compared to mock-inoculated leaves, based on means from three biological replicates. Fold change values greater or equal to 1 represent genes that were up-regulated in response to inoculation, while fold change value <1 represent down-regulation.

CHAPTER 3: A Large Family of Defensin-Like genes found in Nodules of *Medicago truncatula* has Unique Expression and Regulatory Patterns.[3]

Sumitha Nallu[4], Kevin A. Silverstein, Bruna Bucciarelli, Carroll P. Vance, Kathryn A. VandenBosch

Summary

Root nodules are the symbiotic organ of legumes that house nitrogen-fixing bacteria. Multiple genes are induced in the nodules during the interactions between the host and its symbiotic partner. Information regarding the regulation of expression for most of these genes is lacking. One of the largest gene families expressed in the nodules of model legume *Medicago truncatula* are the nodule Defensin-Like (*DEFL*s) genes. We used a custom Affymetrix oligonucleotide microarray to examine the expression changes of nodule *DEFL*s in different stages of nodule development. Additionally, various rhizobial mutants were used to understand the importance of the rhizobial components in induction of nodule *DEFL*s. The early nodule *DEFL*s were detected during the extensive infection of rhizobia in nodules and continue to be expressed into the late stages of nodule development. The late nodule *DEFL*s were induced when there is differentiation

---

[3] This chapter will be submitted for publication to the journal *Plant Physiology*, and is written in the style of the journal.

[4] The first author, Sumitha Nallu, has executed all of the experimental work and analysis reported in this chapter, and has prepared this draft for publication.

of the rhizobia in the nodules. The induction of these groups of genes was dependent on the number and morphology of rhizobia in the nodule. Conserved motifs found in the upstream 1000 bp regions of nodule *DEFL*s were required to drive the expression. These *cis*-element motifs were found to be unique to the nodule *DEFL* family among all the annotated genes in *M. truncatula* genome. Few small regulatory elements known to be involved in the spatial and temporal regulation of various genes were contained within the unique conserved nodule *DEFL* motifs.

Introduction

In legumes, biological nitrogen fixation results from the mutualistic interaction of roots with rhizobia, soil bacteria that are the most efficient nitrogen fixers, producing specialized organs called nodules (Mylona et al., 1995). This interaction leads to a cascade of modified gene expression in both the host and the invading microbe (Jones et al., 2007). Various tools, including mutant screening, reverse genetics, suppressive subtractive hybridization, large scale EST sequencing, macroarray and microarray gene expression analysis (Ané et al., 2008; Maunoury et al., 2010) have been used to identify plant genes governing nodule development and function in the model legume *Medicago truncatula.*

In some legumes, a strikingly large number of genes encoding small, cysteine-rich proteins show conspicuous expression in nodules. Expression of members of this family in nodules was first reported in *Pisum sativum* (Scheres

et al., 1990) followed by *Vicia faba* (Frühling et al., 2000), *M. truncatula* (Györgyey et al., 2000) and *Galega orientalis* (Kaijalainen et al., 2002). Three independent studies then found that these cysteine-rich proteins appeared to be legume-specific and were part of a large (>300) gene family (Fedorova et al., 2002; Mergaert et al., 2003; Graham et al., 2004). Defensins are a highly variable gene family found in vertebrates, invertebrates, plants and fungi involved in defense mechansims (Boman, 1995; Mygind et al., 2005). Graham et al., (2004) found these groups of cysteine-rich peptides to be similar to defensins in their gene structure and genome organization and hence called them Defensin-Like genes (*DEFL*s). Interestingly, a search for the nodule-specific *DEFL*s among ESTs of *Glycine max* and *Lotus japonicus* failed to identify their presence, which led to the hypothesis that nodule-specific *DEFL*s are specific to the Inverted-Repeat Loss Clade (IRLC) of legumes (Mergaert et al., 2003; Graham et al., 2004), although *DEFL*s in seeds appear to be more widely conserved (Graham et al., 2004).

Previously, it was known from the nodule expressed sequenced tags (ESTs) that some nodule *DEFL*s are expressed as early as four days post-inoculation with rhizobia. Mergaert et al. (2003) discovered more than 300 members of this family (they were called NCRs-nodule specific cysteine-rich peptides). They profiled the expression of 14 *NCR* genes across different nodule developmental stages and rhizobial mutants and concluded that all the *NCR* genes were involved in nodulation and probably act at different nodule stages, in different tissues and cell types. The Affymetrix *M. truncatula* Genome Array was

used to create an atlas of gene expression in more than 30 different conditions (Benedito et al., 2008). This array contains a subset of the 322 nodule *DEFL*s (called *NCR*s in their study) that exhibited a nodule-specific expression pattern (Benedito et al., 2008). Using a more comprehensive *DEFL* array, we observed a similar pattern of expression for 571 nodule *DEFL*s (Chapter 2).

Since >80% of *M. truncatula DEFL*s are expressed in nodules at 14 days post-inoculation (dpi) (Chapter 2), we have undertaken a detailed study to identify mechanisms regulating expression of nodule *DEFL*s. We used the custom designed Affymetrix chip to explore the expression patterns of nodule *DEFL*s in nodules inoculated with *Sinorhizobium meliloti* 1021(Sm1021) at marked developmental stages and nodules inoculated with various Sm1021 mutants. The mutants are helpful in dissecting the expression patterns observed within the nodules at different time points. They also provide information on the role of various components of rhizobia in induction of nodule *DEFL*s. An analysis of the expression patterns of nodule *DEFL*s should be carried out in tandem with a careful examination of the upstream promoter sequences that regulate those patterns. Since nodule *DEFL*s are a large family of genes with different expression patterns, contrasting DNA motifs present upstream from some members but not others may provide insights on the factors that regulate their expression.

Here, we describe the different expression patterns of the nodule *DEFL*s. They are broadly grouped into early *DEFL*s and late *DEFL*s based on their expression patterns in nodules at various developmental stages and nodules

inoculated with Sm1021 mutants. Their expression levels are dependent on the volume of rhizobia present in the nodule. The upstream 1000 bp promoter regions of the nodule *DEFL*s have conserved DNA motifs which correlate with a few known *cis*-regulatory elements. Using promoter deletion assays we report that the upstream 1000 bp region is required for the expression of nodule *DEFL*s in the nodules.

Results

Nodule *DEFL* expression patterns are dependent on rhizobial development and nodule maturation

To characterize differential expression of *DEFL*s in nodules, *DEFL* mRNA accumulation was analyzed in nodules inoculated with Sm1021 (WT) at 3, 4, 7, 14 and 40 dpi and in mock-inoculated roots at 0, 4, 7, 14 and 40 dpi. A list of the nodule *DEFL*s that are differentially expressed both in comparison with mock roots 0 dpi and their respective mock roots is presented (Supplemental Table 3-S1). We used 0 dpi mock-inoculated roots as a common control because very few *M. truncatula DEFL*s were differentially expressed in roots over the time course covered by the study (Supplemental Table 3-S2). In roots inoculated with WT under our experimental conditions, infection threads have penetrated the root cortex by 3 dpi and have proliferated within the nodule primordia by 4 dpi. Acetylene reduction assays indicated that the onset of nitrogen fixation occurs at 7 dpi (Supplemental Fig.3-S1). Nodules are fully mature at 14 dpi and by 40 dpi

a senescence zone appears. During nodule development, the percent of nodule *DEFL*s that were expressed increased gradually (Table 3-I, Fig.3-1A).

To understand the role of the symbiotic partner in inducing nodule *DEFL*s, we used *S. meliloti* mutants, derived from the WT strain Sm1021, that affect nodule development at different stages. The *nodC* mutant cannot induce nodules due to the absence of Nod factors (Long et al., 1989b), the *exoY* mutant induces uninfected nodules (Cheng et al., 1998), and the *bacA* mutant induces nodules in which rhizobia senesce before differentiating into bacteroids (Glazebrook et al., 1993). Nodules formed when inoculated with the *nifH* mutant lack nitrogen fixation (Hirsch et al., 1983). In all the treatments listed above, the nodules were harvested at 14 dpi for expression studies. Among the mutants, the *nifH* mutant induced the highest level of *DEFL* expression in terms of both numbers of genes induced and intensity. As seen in the time series, the expression of nodule *DEFL*s gradually increased with rhizobial development and nodule maturation (Supplemental Table 3-S3, Fig.3-1A).

The nodule *DEFL*s were further divided into early and late groups based on their presence and absence calls of expression in the nodules formed by *bacA* mutant. There were a total of 346 early nodule *DEFL*s observed, and all the genes were expressed at lower levels in nodules formed by *bacA* at 14 dpi in comparison to the WT nodules at 14 dpi with the exception of one gene which exhibited a slight elevated expression level in the nodules of *bacA* than in WT nodules. The late nodule *DEFL* group was comprised of a set of 79 genes and was expressed in WT nodules but not in nodules of *bacA* (Supplemental Table 3-

S4). Most of the early nodule *DEFL* expression was first detected in 4 dpi nodules and the intensity levels gradually increased with the age of the nodule (Fig.3-1B). Whereas the late nodule *DEFL* expression was first detected in the nodules after 14 dpi and like the early nodule *DEFL*s, the intensity levels increased in the older nodules (40 dpi) (Fig.3-1C).

In comparing nodule *DEFL* expression in nodules induced by different rhizobial mutants with the various time points following inoculation with WT *S. meliloti*, nodules blocked at different stages of development resembled corresponding time points in the development of WT nodules. The nodule *DEFL* expression in nodules induced by the rhizobial mutants *exoY, bacA* and *nifH* at 14 dpi most closely resembled the expression in WT nodules at 3, 7 and 14 dpi respectively (Fig.3-2 A, B and C). A significant correlation coeffiecient (R^2=0.93) in nodule *DEFL* expression between the nodules formed by *nifH* and WT at 14 dpi suggested a very subtle difference in the expression patterns in these two types of nodules. It has been previously reported that the nodules formed by the *nifH* mutant closely resemble the WT in structure and contain differentiated bacteroids (Hirsch et al., 1983). Through flow cytometry assays we found that the total number of rhizobia in both types of nodules at 14 dpi is not significantly different (Fig.3-3).

The *dnf1* (defective in nitrogen fixation) mutant is a plant mutant, where bacteria enter the nodule via infection thread but do not differentiate into bacteroids (Wang et al., 2010) this is similar to the phenotype observed in the nodules formed by *bacA* rhizobial mutant (Glazebrook et al., 1993). We analyzed

for differentially expressed nodule *DEFL*s in the *dnf1* mutant in comparison to the WT from the previously published data (Starker et al., 2006). None of the 113 nodule *DEFL*s found on their chip was upregulated in the nodules of *dnf1* at 7 dpi, and additionally 89 of the 113 nodule *DEFL*s were down regulated in comparison to the WT nodules at 7 dpi. The 113 nodule *DEFL*s exhibited a similar trend in their expression patterns in the nodules of *bacA* mutant at 14 dpi when compared to the 14 dpi WT nodules in our study (Supplemental Table 3-S5).

Conserved motifs occur uniquely in the upstream regions of the nodule *DEFL*s

To search for common *cis*-elements among nodule *DEFL*s, we mapped the position of 209 nodule *DEFL*s onto the sequenced BACs (MT2.0 version) and used the region from the start site to 2000 bp upstream for motif discovery using the Multiple Em for Motif Elicitation algorithm (MEME, Bailey and Elkan, 1994). This approach identified five conserved motifs that each occurred in more than half the input sequences. These five motifs were selected (Supplemental Fig.3-S2) and searched against different databases using Motif Alignment and Search Tool (MAST) (Bailey and Gribskov, 1998). The conserved motifs occurred in the first upstream 1000 bp region and were especially densely clustered approximately in the upstream 400 bp region relative to the putative translational start site. Previously, Silverstein et al. (2007) grouped *M. truncatula DEFLs* into subgroups based on sequence similarity. Here, supplemental Table 3-S6 lists

the nodule *DEFL* genes, the subgroups to which they belong, combined E-values of all the motifs and the upstream motif patterns found in each gene. Nodule *DEFL*s that belong to nodule-specific subgroups (Silverstein et al., 2007) have more significant E-values when compared to nodule *DEFL*s subgroups that are also expressed in other parts of the plant, suggesting that some of the motifs regulate the nodule specificity of the gene. As described earlier, the nodule *DEFL*s were divided into early and late groups based on their expression in the nodules of *bacA* mutant. MEME was used to identify conserved motifs for each group, and no additional motifs were observed between early and late nodule *DEFL*s, suggesting that both the groups have the same motif pattern.

To search for additional conserved motifs outside of the 5' upstream region, the introns and the putative 3' UTR regions (1000 bp downstream region from the translational stop site) were used to generate motifs using MEME. The software did not reveal any significant new motifs. In addition, the five conserved upstream motifs were also found to be absent from these regions. The conserved nodule *DEFL* motifs were also absent from the upstream 2000 bp, introns and 3' UTR regions of the 88 non-nodule *DEFL*s (*DEFL*s not expressed in the nodules). Furthermore, this motif pattern was absent among the 1000 bp upstream regions of the 33,131 annotated genes in the *M. truncatula* genome sequence (IMGAG 2.0) excluding the nodule *DEFL*s. Thus, this unique combination of motifs is confined to the upstream 1000 bp region, and clustered in the 400 bp upstream regions of nodule *DEFL*s in *M. truncatula*.

Known plant regulatory elements resemble components of the conserved motifs found in nodule *DEFL*s

We used Clover (Cis-eLement OVERrepresentation, Frith et al., 2004), an algorithm that detects both under and over-represented DNA motifs using statistical models, to identify the occurrence of known elements in the upstream regions of the nodule *DEFL*s. The upstream 1000 bp regions from (1) the 209 nodule *DEFL*s, (2) the 88 non-nodule *DEFL*s and (3) 3000 non *DEFL*s from IMGAG 2.0 were searched for all 104 plant regulatory elements from TRANSFAC database (release 12.1). The latter two groups were used as two separate background models for comparison to the nodule *DEFL*s. Six known elements were found to be overrepresented in the 1000 bp upstream region of the nodule *DEFL*s (Table II). No statistically significant underrepresented nodule *DEFL* promoter motifs were identified.

In order to compare the nodule motifs generated by MEME with the overrepresented known elements from TRANSFAC, we used STAMP, a web tool that explores similarities between DNA motifs (Mahony and Benos 2007, http://www.benoslab.pitt.edu/stamp). MEME motifs 1 and 2, which are 41 bp long, exhibited a high correlation with the six small known motifs (6-12 bp in length) that mapped to sites within the longer MEME motifs. MEME motif 3 exhibited a strong correlation with a known element, ELEMENT1GMLBC3 (S000319) that has been found in the promoter region of leghemoglobin in *G. max* (Jensen et al., 1988). This element is from the plant *cis*-acting regulatory DNA elements (PLACE 30.0) database (Higo et al., 1999). A further search for

nodule-specific *cis*-elements revealed that MEME motifs 1 and 4 have the CTCTTT or the NICE element motif 2 (Sandal et al., 1987; Szczyglowski et al., 1994) that have been shown to confer nodule specificity to *Srglb3,* leghemoglobin in *Sesbania rostrata*. We did not find any significant correlation between MEME motif 5 and the known elements evaluated in this study. Table 3-III lists the E- values for the correlations between the MEME nodule *DEFL* motifs and known *cis*-elements.

Upstream 1000 bp region is required to drive nodule *DEFL* expression in nodules

To determine the promoter region required to drive the expression of nodule *DEFL*s, we generated promoter::GUS fusion constructs using the upstream 1000 bp region of the gene corresponding to TC100321 (MtTC100321), a nodule *DEFL* with a very high expression value and three tandem duplicated gene copies. GUS staining revealed that this promoter is highly active in the interzone region (zone II-III, Vasse et al., 1990), where the cell differentiation of both the host and rhizobia occurs and the nitrogen fixing zone (zone III) (Fig.3-4A). To validate the localization pattern, we hybridized the antisense mRNA probe of MtTC100321 to nodule sections and found that the *in situ* hybridization pattern was similar to the pattern observed using the promoter::GUS fusion (Fig.3-4B). Earlier it was reported that early nodule *DEFL* (*NCR084*) transcripts mainly accumulate in the interzone II-III and the late nodule *DEFL* (*NCR001*) are in zone III (Mergaert et al., 2003). MtTC100321 is an early

nodule *DEFL* with very high expression values compared to *NCR084* (MtTC94567) in both early and late stages of nodule development, MtTC100321 might have a different or extended functional role compared to *NCR084* and hence a different spatial localization pattern. In recent reports the late nodule DEFL peptides were observed only in the infected cells (Van de Velde et al., 2010). We also found that in the nitrogen fixing (III) zone, transcripts of MtTC100321 accumulated only in the infected cells (Fig.3-4C).

Because the approximate upstream 400 bp region of the nodule *DEFL*s typically contains the 5 conserved nodule *DEFL* motifs, we performed promoter deletion assays to test whether this region was sufficient to drive the nodule *DEFL* expression. Three nodule *DEFL*s were chosen for this assay. The genes corresponding to TC103606 (MtTC103606) and TC95126 (MtTC95126) were selected as examples of the most highly conserved motif pattern and significant E-values and a medium level of expression. MtTC100321 was selected because its 400 bp region had a less typical pattern of motifs. We used three overlapping segments of the promoter region upstream of the translational start site to drive the GUS construct: segment 1, approximately 0 to 400 bp upstream, segment 2, 0 to 1000 bp upstream, and segment 3, 400-2000 bp upstream (Fig.3-5A). Out of all the constructs, only the upstream 1000 bp regions of the three putative promoters of nodule *DEFL*s drove the expression of the GUS reporter gene (Fig. 3-5B, 3-5C, 3-5D), indicating that the 400 bp region upstream of the translation start site is necessary, but not sufficient, to drive gene expression in nodules.

Nodule *DEFL*s are redundant in function

Five different nodule *DEFL*s were selected based on their expression profiles and sequences for functional analysis. MtTC100321 is a nodule *DEFL* with the highest level of expression at 14 dpi, differential expression following infection with the root rot pathogen *Phythophthora medicaginis* (Chapter 2) and belongs to the group of early nodule *DEFL*s. Genes corresponding to TC100264 (MtTC100264), TC94214 (MtTC94214), TC108430 (MtTC108430) and AW775198 (MtAW775198) were also selected for functional assays. MtTC100264 and MtTC94214 are ubiquitously expressed in vegetative tissues, including nodules, (Chapter 2), MtTC100264 belongs to the subgroup of *M. truncatula* specific classical defensins and MtTC94214 is a defensin and a member of the subgroup that has homologs in various plant species (Silverstein et al., 2007). MtTC108430 and MtAW775198 are late nodule *DEFL*s, and are only known to be expressed in nodules. Each gene was individually knocked down using RNAi technology and over-expressed using CsVMV promoter. Plants were assayed for various phenotypes with/without rhizobial inoculation as described in the materials and methods. Silencing or over-expression of single genes did not result in an observable phenotype, under our conditions (Supplemental Fig.3-S3). Using quantitative RNA methods the gene expression levels were assayed in the transgenic plants, the RT-PCR results confirmed the efficiency of our knock down and over-expression techniques (Supplemental Fig.3-S4). Similarly, transgenic plants were evaluated for their disease severity

when interacting with *P. medicaginis* as described in the materials and methods, and no significant difference was observed in comparision to the control.

Discussion

Expression of nodule *DEFL*s is proportional to the volume occupied by rhizobia in the nodules

Based on expression patterns, nodule *DEFL*s can be divided into two major groups: early and late genes. The early group of nodule *DEFL*s was induced after the invasion of bacteria into nodules. This is interpreted by comparing the expression of nodule *DEFL*s at 4 dpi with those of 3 dpi and Sm1021 mutants *exoY* and *nodC*. In our experimental conditions we observed that at 4 dpi, there is an extensive proliferation of infection threads compared to 3 dpi. In the *nodC* mutant, because of the absence of both root hair growth deformation and cortical cell division (Long et al., 1989), the plant is Nod$^-$.In *exoY* mutants there is no infection thread development due to the defect in succinoglycan production (Cheng et al., 1998; Jones et al., 2007). We have observed minimal nodule *DEFL* expression in the nodules of the WT at 3 dpi, as well as in the nodules formed by the Sm1021 mutants *nodC* and *exoY* relative to 4 dpi nodules. This suggests that the early nodule *DEFL*s are induced when there is an extensive infection of rhizobia in the nodules. These early nodule *DEFL*s are also expressed in the subsequent nodule development stages of our

study with expression levels increasing as the nodule develops and more rhizobia enter and fill in the different nodule zones.

The transcripts of the late nodule *DEFL* group are detected following bacteroid formation. This is inferred from the nodule *DEFL*s that are not expressed in *dnf1* and *bacA* mutant nodules but expressed at low levels in WT nodules at 7 dpi and medium to high levels at 14 dpi. When rhizobia are released from the infection threads into the nodules, they undergo differentiation in five different steps and can be located in different zones of nodules (Vasse et al., 1990).The *dnf1* mutant, like the *bacA* mutant forms Fix$^-$ nodules that have bacterial differentiation arrested in early stages of their release from the infection threads (Wang et al., 2010) and thus both the mutants form nodules that are impaired in development and have rhizobia scantly distributed in the nodule.

In *M. truncatula*, which has indeterminate nodules, there is a persistent meristem that remains constant in size as the nodule matures. As new meristematic cells are produced there are an equal number of post-mitotic cells that start to differentiate. These differentiated layers, which harbor the bacteroids, increase in size during indeterminate growth. Throughout the nodule development the rhizobia are released into the submeristematic cells, undergo differentiation and fill the growing nitrogen fixation zone (Mergaert et al., 2006). We observed that the nodule *DEFL*s are induced when there is an extensive release of rhizobia from infection threads into the WT nodules at 4 dpi, and the number as well as the level of expression continues to grow as rhizobia differentiate and occupy more volume of the nodule. In the sparsely populated

nodules formed from inoculation with Sm1021 mutants like *bacA* and the plant mutant *dnf1*, there seems to be a sudden reduction in the number and level of expression of nodule *DEFL*s. This sudden shift may occur when the plant senses the disruption of the nodule development and shuts down the transcription of the machinery required at later stages including other late nodulins like *MtLB1* and *MtCAM1,* which are not expressed in either *bacA* or *dnf1* nodules (Mitra and Long, 2004). In addition, the nodule *DEFL* expression patterns in nodules with *nifH* mutant rhizobia were similar to the WT Fix^+ nodules. This implies that the expression of nodule *DEFLs* is not dependent on nitrogen fixation as the nodules produced by *nifH* mutant cannot fix nitrogen but are similar in size to the WT nodules. They have well developed nodule zones with bacteroids filling up the nitrogen fixation zone (Hirsch et al., 1983). Our flow cytometry assays also indicate that there is no significant difference between the numbers of rhizobia present in the nodules formed by *nifH* mutant at 14 dpi when compared to a WT nodule at 14 dpi. This further supports our hypothesis that nodule *DEFL* expression is dependent on the number of bacteria present and the volume occupied by them in the nodule.

Regulation of nodule *DEFL* expression in nodules

It was earlier reported that there are large repeat regions in between the nodule *DEFL*s present on the same BAC and right in front of the predicted translational start site there are conserved mini repeats which are similar to each other in between the repeats (Graham et al., 2004). The clustered conserved

motifs found in the upstream 400 bp of the nodule *DEFL*s overlapped with those minirepeats. These five conserved motifs are found to contain previously known *cis* regulatory elements, the known elements could be divided into two groups, the first group includes gene expression regulators ID-1 binding site, ARF binding site, Dof protein binding site and MADS box genes binding site. ID1, a zinc finger transcription factor binds to an 11 bp DNA domain present in the promoter region of genes and regulates their expression (Kozaki et al., 2004). An ID-1 like transcription factor has been shown to be up regulated in *M. truncatula/S. meliloti* interaction (Godiard et al., 2007), which might be involved in regulation of nodule *DEFL*s. Auxin is known to play a major role in plant development, in *M. truncatula/S. meliloti* symbiosis it is involved in the initiation of nodule primordia and regulation of nodule number (Mathesius, 2008). It is also implicated to be involved in regulation of gene transcription through the binding of Auxin Response Factors to the *cis*-element (TGTCTC) present in the promoters of auxin response genes (Hagen and Guilfoyle, 2002). Dof proteins like Dof2 and PBF are transcriptional factors unique to plants. They are known to bind to diverse plant promoters and have been speculated to participate in the regulation of expression of genes involved in photosynthesis, defense mechanisms, seed specific genes and an oncogene (Yanagisawa et al., 1999). MADS box genes are transcription factors found in both plants and animals. In plants they have been shown to regulate flower development (Huang et al., 1996), in *M. truncatula* they are nodule specific and localized to the infected cells and are probably involved in regulation of nodule specific genes (Heard and Dunn, 1995).

Additionally, based on the data reported in the *Medicago truncatula* Gene Expression Atlas (http://mtgea.noble.org/v2/) all the above four transcription factors are expressed in nodules of *M. truncatula* and ID1 in particular exhibits a highly correlated expression profile to nodule *DEFL*s.

The second group of the conserved motifs includes the elements involved in nodule specificity like the ELEMENT1GMBLC3, CTCTTT and NICE2 which drive the expression of nodule genes into nodules (Jensen et al., 1988; Sandal et al., 1987; Szczyglowski et al., 1994). It is not surprising that large families of genes like the nodule *DEFL*s which have different expression patterns and expression levels are regulated by various transcription factors during the different stages of nodule development. The presence of common motifs between nodule *DEFL*s and leghemoglobin suggests that these regulatory motifs might have been recruited into nodule *DEFL*s from the ancient nodule specific leghemoglobins during the evolution of nodule *DEFL*s which probably was recent. Further studies on evolutionary mechanisms of nodule *DEFL*s are required to test this hypothesis.

In our Promoter::GUS deletions assays it was observed that the 400 bp upstream segment did not show any GUS expression while the upstream 1000 bp segment could drive the nodule *DEFL* expression into the nodules. The 1600 bp upstream segment without the clustered motif region did not show any GUS expression which suggests that the upstream region with clustered motifs (approximately 400 bp) was not sufficient but may be required to drive the nodule DEFL expression into the nodules. Further detailed analysis of the upstream

1000 bp region using serial deletions or site directed mutagenesis would reveal the significance of each motif in the regulation of nodule *DEFL*s.

Possible biological roles of nodule *DEFL*s

It has been recently reported that a few NCRs (nodule DEFLs) have a lethal effect on free-living rhizobia in *in vitro* assays. Like some defensins, they triggered membrane modifications and inhibited bacterial cytokinesis. When one *NCR* (nodule *DEFL*) was expressed in *L. japonicus,* it resulted in terminal bacteroid differentiation (Van de Velde et al., 2010). The legumes outside the IRLC, like *L. japonicus* and *G. max,* have nodules with rhizobia that do not undergo any changes in their cell and genome size and can reproduce within the nodule. These plants do not have nodule *DEFL*s (*NCR*s). In contrast, the IRLC legumes like *M. truncatula, P. sativa, V. faba* have elongated terminally differentiated bacteria with amplified genomes and have nodule *DEFLs* (Mergaert et al., 2006). Host sanctions have been reported in *G. max* to prevail over the cheating rhizobia that take up the carbon resources without fixing nitrogen. Such host sanctions are yet to be reported in the IRLC legumes (Oono et al., 2009). It has been speculated that the nodule *DEFL* (*NCR*) family was recruited by the IRLC legumes to overcome the cheating mechanisms of rhizoba (Van de Velde et al., 2010). It is observed from this study that the nodule *DEFL* expression levels correlate with the number of rhizobia present in the nodule. Very little is known about perception of rhizobium signal molecules after initial perception of Nod factor. However, our results suggest that perception of rhizobial surface

components or other signal molecules by the plant may trigger nodule *DEFL* expression.

Nodule *DEFL*s are classified into 35 subgroups based on sequence similarity (Silverstein et al., 2007). Due to their large numbers and sequence diversity, it is possible that they are involved in multiple functions in nodules. Recent reports demonstrate that some plant defensins and DEFLs can function as signal molecules in plant development (Allen et al., 2008), regulation of reproduction (Stotz et al., 2009), pollen tube development and guidance (Okuda et al., 2009; Amien et al., 2010) and pollen self-incompatibility (Schopfer et al., 1999; Takayama et al., 2001). Similarly, it can be hypothesized that the nodule *DEFL*s could act as signals for the different nodule development stages. Additionally, because of their sequence similarity to defensins, they could be acting as anti-microbial peptides acting against the plethora of soil pathogens. Alternatively, they may be functional against rhizobium by triggering senescence or containing the rhizobia to these indeterminate nodules, and preventing a systemic infection.

We speculate that nodule DEFLs may have multiple roles to play in this complex network of communiqué between the microbe and the plant. DEFLs have been reported to play dual roles in defense and developmental signaling of plants (Dresselhaus and Marton, 2009), and so there is a possibility for a nodule *DEFL* to have more than one function. In whatever role(s) the nodule DEFLs are involved, our reverse genetic assays suggest functional redundancy among the many nodule DEFLs. A detailed study of molecular interactions between the

host-microbe components, their regulatory factors and selection pressures is required to understand this large, fast-evolving redundant family of genes.

Materials and Methods

Plant material and growth conditions

Seeds of *Medicago truncatula* accession A17 were sterilized and germinated as described previously (Lohar et al., 2006). Germinated seedlings were grown on sterile plates (Corning 245mm*245mm dishes) containing buffered nodulation medium (BNM, Ehrhardt et al., 1992), pH 6.5, 1.2% agar (Plant Tissue culture grade, Sigma). The radicals were placed on a moist sterile germination paper and wrapped with a sterile black cotton cloth (Cotton Club Black, #074300603820, Wal Mart). Plates were maintained vertically in a growth chamber set at 16 h light / 8 h day, 25 to 21 ^{o}C, 200 to 300 µmol m^{-2} s^{-1} and 50% humidity. At 5 d after planting, plants for the nodulation studies were inoculated with 100µL/root of a washed suspension of *Sinorhizobium meliloti*, Sm1021(Meade et al., 1982) culture [OD600] = 0.05 in sterile water. For the entire nodule developmental stages Sm1021 was used for inoculation and for the mutant series *nodC*, *exoY*, *bacA* and *nifH* Sm1021 mutant strains were used for inoculation. Mock-inoculated plants were inoculated with 100µL/root of sterile water. For 40 dpi nodules and mock-inoculated roots, the germinated seedlings were transplanted in pots with Turface (MVP) and inoculated with Sm1021 culture100µL/root as described above. The plants were fertilized once a week

with aeroponic nutrient medium (LIPM formula, Lullien et al., 1987) without the nitrogen source (for mock inoculated plants the nutrient medium included the nitrogen source). Approximately 5-cm nodule-bearing root segments were harvested from inoculated plants and mock inoculated plants at 3, 4, 7, 14 dpi and 40 dpi, root tips were removed from all plants at the time of harvest to reduce any transcript contribution from the meristematic cells. The root segments from all treatments were collected from three biological replicates with multiple plants in each replicate, were frozen immediately in liquid nitrogen during harvest and stored at -80°C for subsequent RNA extraction. Total RNA extraction, probe preparation and array hybridization was performed as described in Chapter 2.

Microarray data analysis

Data were normalized across the different nodule treatments using SBQ normalization (Sato et al., 2007; Chapter 2). Differentially expressed genes were identified using the Empirical Bayes method within the LIMMA (Smyth, GK 2005) package distributed with R/Bioconductor (Gentleman, RC et al., 2004). Both a fold change cut off (2.0) and a Benjamini-Hochberg FDR correction ($P < 0.05$) were applied.

Characterization of nodule phenotypes

For measuring the nitrogenase enzyme activity, acetylene reduction assays (Vance et al., 1979) were performed on 6, 7 and 8 dpi plants grown on sterile BNM plates. Briefly, the plants were wiped dry and placed in 250 ml

sealed jars. Twenty five ml of air was withdrawn and replaced with an equal amount of acetylene. After 1 h of incubation, 1 ml of sample was injected into a Photovac 10S Plus Gas C chromatograph. The nitrogenase activity was expressed as nmoles of ethylene generated per h/plant.

To quantify the number of rhizobia, nodules from Sm1021 and *nifH* plants grown on sterile BNM plates were harvested at 14 dpi. The bacterial suspension was prepared as previously described (Oono et al., 2010) and quantified on a Becton Dickson FACScalibur. A Student's t-test was used to determine the p-value between the two different treatments.

Identification and analysis of conserved motifs

Nodule *DEFL*s were mapped onto MT 2.0 using PASA (Haas et al., 2003). The upstream 1000 bp to start site, 2000 bp to 1000 bp, intron regions and stop site to downstream 1000 bp were extracted using custom Perl scripts. For motif discovery, several runs with different parameters of locally installed MEME algorithm (Bailey and Elkan, 1994) were executed to find the best possible motifs, the motifs generated by using zero or one occurrence per site module were chosen. MAST (Bailey and Gribskov, 1998) was used to scan for the presence of motif models generated by MEME in different regions of nodule *DEFL*s, non-nodule *DEFL*s and 33,131 non *DEFL*s in IMGAG 2.0. The sequences with E-values $<10^{-3}$ were considered significant.

To scan for known elements in the 1000 bp upstream regions of the nodule *DEFL*s, 104 plant motif matrices were extracted out of the TRANSFAC

12.1 database and using the locally installed Clover algorithm (Frith et al., 2004) overrepresented motifs in nodule *DEFL*s were identified. The motifs were called overrepresented if they had scores greater than 1 and p-value less than 1, the p-value was generated against the background sequence sets. The upstream 1000 bp regions of 88 non-nodule *DEFL*s and 3000 non *DEFL* (IMGAG 2.0) genes were used as background sets.

Correlation among the motif matrices from MEME versus the matrices of overrepresented elements from TRANSFAC was statistically calculated using STAMP (Mahony and Benos 2007, http://www.benoslab.pitt.edu/stamp). For elements from PLACE 30.0 database (Higo et al., 1999) and the nodule specific elements, their consensus sequences were statistically evaluated against the matrices of MEME motifs.

Generation and evaluation of promoter deletion constructs.

The segment 1 (400 bp), segment 2 (1000 bp) of MtTC103606, MtTC95126 and MtTC100321 as well as the segment 3 (1600 bp) of MtTC103606 and MtTC95126 were PCR amplified from their respective BAC DNA (primer and template details in Supplemental Table 3-S7) and cloned into pENTR-D/TOPO (Invitrogen). LR recombination was performed with the Gateway compatible vector pKGW-R:EGFP-GUS which is a modified pKGW-R (Smit et al., 2005) that includes EGFP-GUS fusion gene. All the plasmids were transformed into *A. rhizogenes* Arqua1 (Quandt et al., 1993). *M. truncatula* A17 seedlings were transformed via hairy root transformation (Boisson-Dernier et al.,

2001) with 20 µg/ml of kanamycin selection. After generation of hairy roots (~2.5 cm), all the roots except one transgenic root were removed and the plants were transferred to 0.5X Gamborg's B5 Basal Salt medium (Sigma) with 1% plant agar (Sigma) to recover from antibiotic selection. After a week the plants were transferred to 2.25 square pot (Landmark Plastics) filled with Turface (MVP) and inoculated 100µL/root of a washed suspension of rhizobial culture [OD600] = 0.05 in sterile water.

After 14 dpi the plants were screened for the transgenic *DsRED1* marker expressed in the pKGW-R:EGFP-GUS vector using a Nikon SMZ 1500 microscope with DsRed filter. The nodules were harvested from the screened transgenic plants and assayed for GUS activity, the plants were not screened for GFP because of the interference of auto florescence. The nodules were infiltrated with 2 mM X-Gluc, Triton X-100, 50 mM NaPO4, pH 7.2, 2 mM potassium ferrocyanide, 2 mM potassium ferricyanide under vaccum for 30 min and incubated at 37ºC for overnight.

For confirmation of the expression pattern of promoter::GUS fusions, *in situ* hybridization of MtTC100321 was performed. The coding region of MtTC100321 was PCR amplified (primer and template details in Supplemental Table 3-S7) and cloned into pGEMT easy vector (Promega, USA), subcloned into pBlueScript KS+ vector (Stratagene Inc.) for digoxigenin (DIG) labeling. The *in vitro* transcription using DIG-11-UTP (Roche Applied Science) was carried out on linearized plasmid using T7 and T3 polymerases. The 14 dpi nodules were harvested from plants grown on sterile BNM plates as described above. The

tissue fixation, sectioning, hybridization and signal detection were carried out as mentioned in Sbadou et al. (2010).

RNAi and over-expression construct design and screen.

For RNAi constructs, 150-200 bp from the coding regions of 5 nodule *DEFL*s were amplified from their corresponding EST libraries along with human myosin gene, which was used as a control (primer and template details in Supplemental Table 3-S7). The PCR products were cloned into pENTR-D/TOPO (Invitrogen). LR recombination was performed with the binary vector, modified pHellsgate8 (Pumplin et al., 2010).

For over-expression, the entire coding regions of five nodule *DEFL*s were amplified from their cDNAs with *Xba*I and *Bam*HI adapter sites at 5' and 3' ends respectively (primer and template details in Supplemental Table 3-S7). The PCR products were cloned into pENTR-D/TOPO (Invitrogen). The cDNA was excised using *Xba*I and *Bam*HI endonucleases and ligated into the binary vector pILTAB381 (Samac et al., 2004) that was digested with *Xba*I and *Bam*HI endonucleases which removed the GUS gene. The result was a construct with the entire coding region of the nodule *DEFL* driven by CsVMV promoter. The pILTAB381 vector with intact GUS gene was used as a control.

All the constructs were transformed into *A. rhizogenes* Arqua1 followed by A17 hairy root transformation, inoculation with SM1021, and growth as described above. After 14 dpi the plants were screened for the transgenic *DsRED1* marker expressed in the modified pHellsgate8 vector under Nikon SMZ 1500 microscope

with DsRed filter. Plants to be assayed at 40 dpi were transferred to 2.25 Square pot (Landmark Plastics) filled with Turface (MVP) and fertilized with 0.5X BNM once per week. The height of the plant, color of leaves, root and root hair length, architecture and nodule number, size, distribution and color were measured at 0 dpi, 14 dpi and 40 dpi. Significant ($P<0.05$) difference between the transgenic plant and control was determined by the Mann-Whitney U test. For Quantitative RT-PCR, primers were designed for TC100321 gene region and used to quantify RNA levels in both the RNAi and OE transgenic lines (primer and template details in Supplemental Table 3-S7). Quantitative RT-PCR, data quantification, and analysis were performed as described in Chapter 2.

For evaluation of susceptibility to *Phytophthora medicaginis*, the non-nodulated and nodulated 14 dpi RNAi and over-expression transgenic lines along with their controls were transferred to 2.25 Square pot (Landmark Plastics) filled with Turface (MVP) and inoculated with 1 ml *P. medicaginis* M2019 inoculum prepared as described in Chapter 2. The flats were flooded with sterile water and covered for 2 days. After 2 days the excess water and the covers were removed and fertilized with Peters Professional 10:10:10 fertilizer (5X, Scotts) every 3 days, and monitored for symptoms of infection (Moussart et al., 2006).

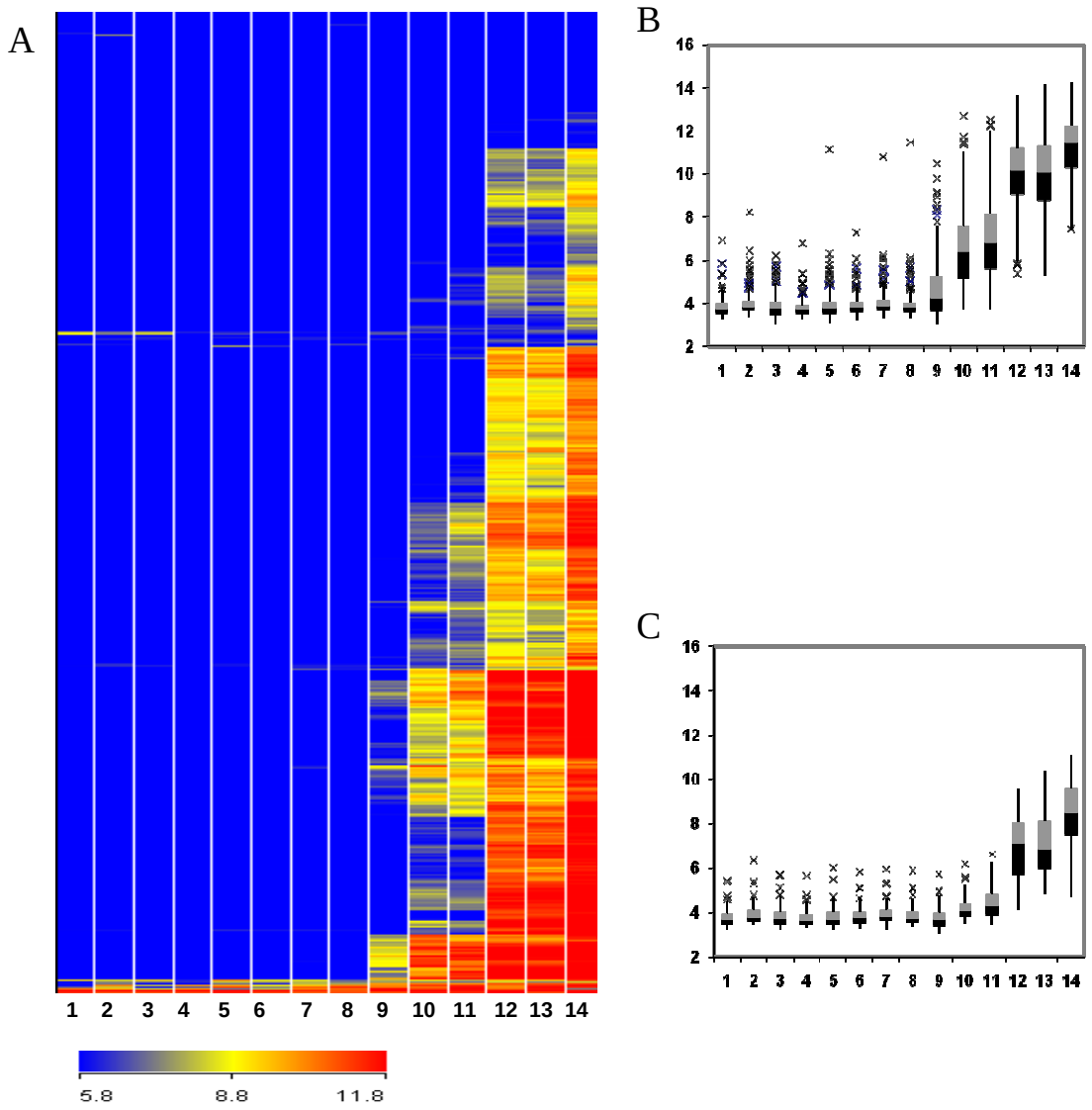


Figure 3-1. Expression profiles of nodule *DEFLs*. Expression values are $\log_2$ transformed. A) Hierarchical clustering (Euclidian-average) of 571 nodule *DEFLs* in 1,2,3,4 and 5 are mock inoculated roots at 0, 4,7,14 and 40 dpi, 6,7,10 and 12 are roots inoculated with rhizobial mutants *nodC*, *exoY*, *bacA* and *nifH* at 14dpi, 8,9,11,13 and 14 are roots inoculated with WT Sm1021 at 3,4,7,14 and 40 dpi respectively.Color scales representing signal intensities are shown at the bottom. B and C are the expression profiles of early and late nodule *DEFL*s respectively.

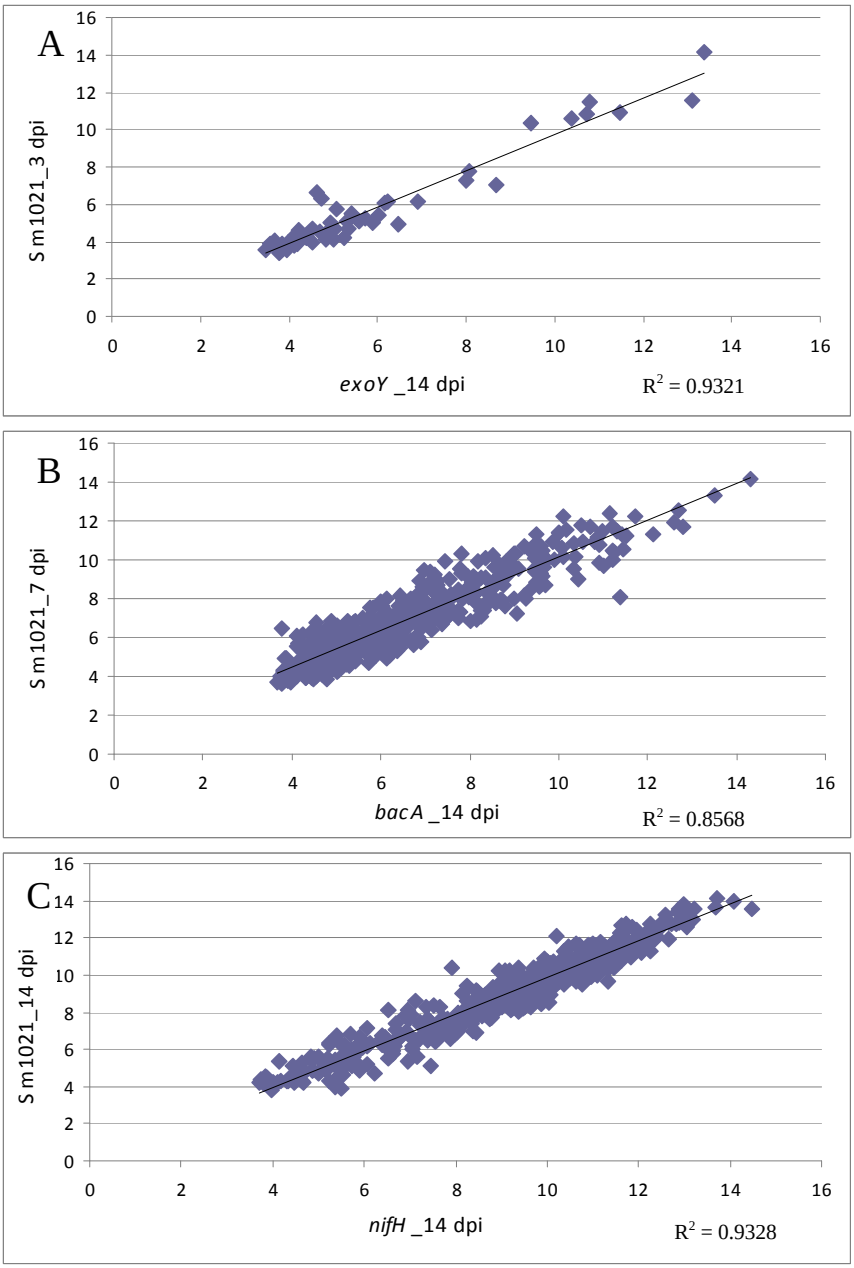
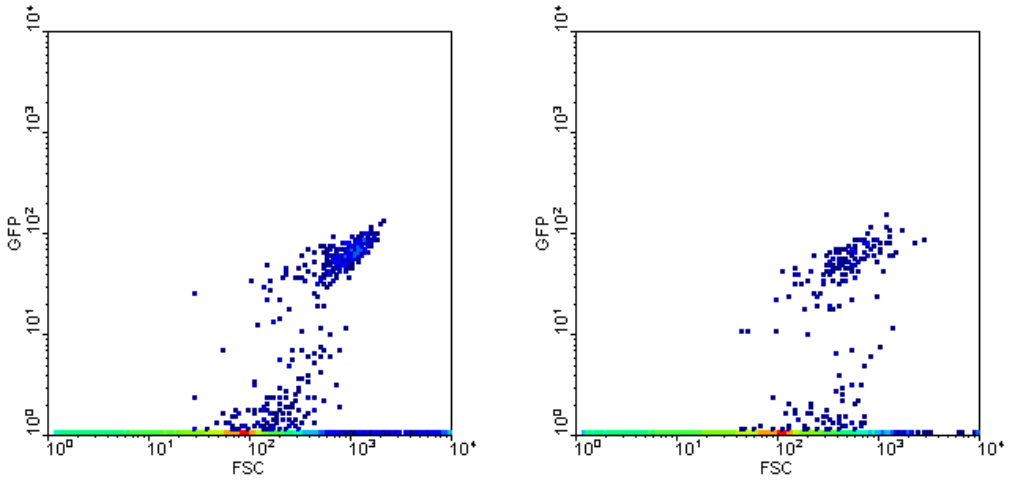


Figure 3-2. Correlations among the nodules induced by mutants and WT nodules at various developmental time points. All nodule *DEFLs* with "present" calls among the two treatments plotted were used to generate the scatter plots. Linear regression plots with significant correlation (R^2) are shown. A, is the comparision between nodules induced by *exoY* mutant at 14 dpi VS WT nodules at 3 dpi. B, nodules induced by *bacA* mutant at 14 dpi VS WT nodules at 7 dpi and C, nodules induced by *nifH* mutant at 14 dpi VS WT nodules at 14 dpi.



nifH nodules_14dpi

Sm1021 nodules_14dpi

Figure 3-3. Quantification of rhizobia in flow cytometry assays. Density plots showing no significant difference in the total number of rhizobia in *nifH* and Sm1021 (WT) nodules at 14 dpi. Twenty biological replicates (nodules from 20 different plants) were used for each treatment. The t-test p-value was >0.95.

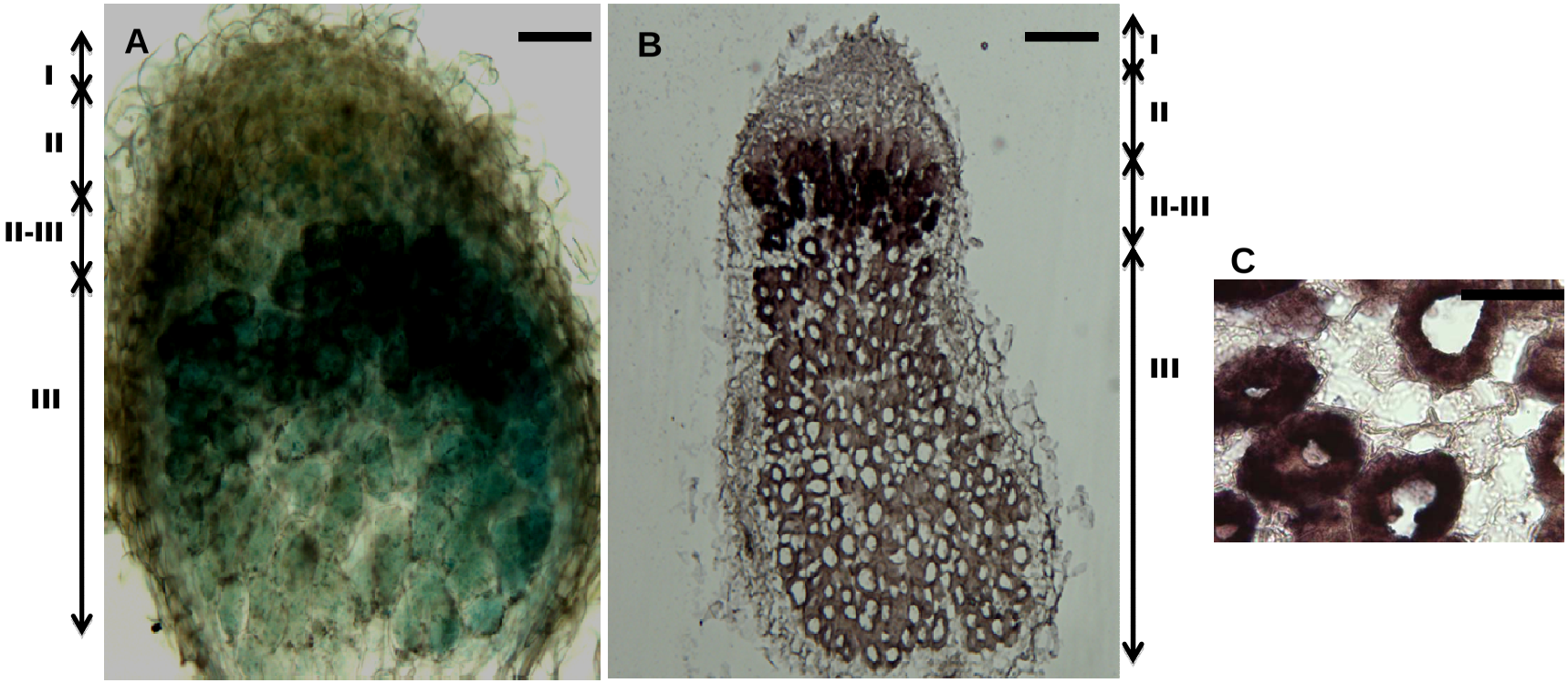


Figure 3-4. Localization of a nodule *DEFL* in *M.truncatula* nodules. A, is the GUS staining of 50 micron section of a promoter (TC100321):GUS transgenic nodule. B, is the *in situ* hybridization of the antisense probe of TC100321 on a 10 micron nodule section. Bars equal 200 µm. C, is magnified inset of nodule from *in situ* hybridization. Bar equals 100 µm. All nodules =14 dpi.

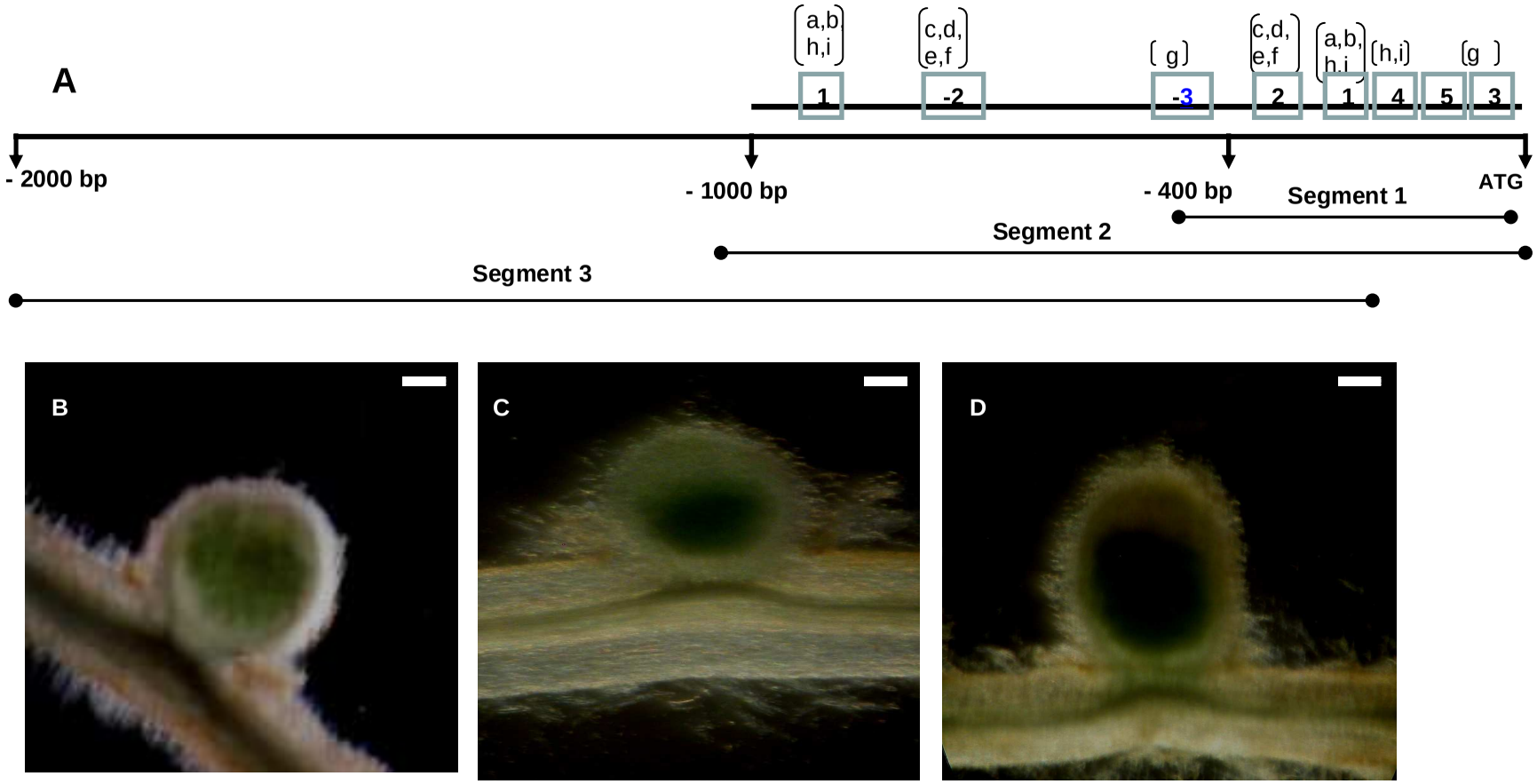


Figure 3-5. Promoter deletion assays. A, illustrates the different segments used for promoter deletion assays for the gene corresponding to TC103606.1,2,3,4 and 5 are the conserved MEME nodule motifs and a to i are the known elements - P$ID1_01, P$ARF_Q2, PPBF_Q2, PPBF_01, P$DOF2_01, $AGL1_01, ELEMENT1GMBLC3, TTGTCTCTT and CTCTTT, repectively, mapped on the MEME motifs. B, C, D are transgenic nodules of genes corresponding to TC103606, TC95126 and TC100321 respectively with 1000 bp (Region 2) promoter:GUS fusions. 14 dpi nodules were stained for GUS expression. Bars equal 200 µm.

Nodule stage[a]	Inoculum[b]	DE genes[c]
3 dpi	Sm1021	15
4 dpi	Sm1021	166
7 dpi	Sm1021	391
14 dpi	Sm1021	511
14 dpi	*nodC*	9
14 dpi	*exoY*	15
14 dpi	*bacA*	374
14 dpi	*nifH*	510
40 dpi	Sm1021	527

Table 3-I. Differentially expressed nodule *DEFL*s in various nodule treatments. [a]Time point of nodule harvest. [b]Rhizobial culture used for inoculation. [c]Number of Differentially expressed nodule *DEFL*s in different nodule treatments in comparison to mock inoculated roots at 0 dpi, both +/- induced genes are included.

Identifier[a]	Accession[b]	p-value[c]	Reference
P$ID1_01	M01021	0	Kozaki et al., 2004
P$ARF_Q2	M00438	0	TRANSFAC Reports 2:0002 (2001)
P$PBF_Q2	M01130	0	TRANSFAC Reports 103:0001 (2006)
P$PBF_01	M00355	0	Yanagisawa et al., 1999
P$DOF2_01	M00353	0.08	Yanagisawa et al., 1999
P$AGL1_01	M01059	0.005	Huang et al., 1996

Table 3-II. Overrepresented TRANSFAC elements in nodule *DEFL* promoters. [a]The known element identifier. [b]Corresponding TRANSFAC Accession number of the known element. [c] p-values obtained from randomizing the elements using the non *DEFL* IMGAG genes as background sequence sets. All p-values were 0.0 when using non-nodule *DEFL*s as a background set.

Known element[a]	MEME Motif[b]	Orientation[c]	E- value[d]
P$ID1_01	Motif 1	forward	3.17E-09
P$ARF_Q2	Motif 1	reverse complement	6.42E-05
P$PBF_Q2	Motif 2	forward	3.99E-05
P$PBF_01	Motif 2	forward	7.16E-04
P$DOF2_01	Motif 2	reverse complement	3.13E-02
P$AGL1_01	Motif 2	reverse complement	4.05E-04
ELEMENT1GMBLC3	Motif 3	reverse complement	0.00E+00
TTGTCTCTT	Motif 1	reverse complement	5.09E-08
TTGTCTCTT	Motif 4	reverse complement	8.09E-08
CTCTTT	Motif 1	forward	1.11E-03
CTCTTT	Motif 4	reverse complement	1.94E-03

Table 3-III. Mapping of known elements onto the nodule *DEFL* MEME motifs. [a]Known elements were mapped onto the 5 nodule *DEFL*s.[b]Motifs generated by MEME. [c]Orientation of alignment with the MEME motifs and the most significant. [d]E-values of correlation.

CHAPTER 4: Evolutionary patterns of nodule Defensin-Like genes in ecotypes of *Medicago truncatula*.[5]

Sumitha Nallu[6], Peng Zhou, Timothy Paape, Kevin A.T. Silverstein, Nevin D. Young, Kathryn A. VandenBosch

Summary

- Defensin-Like genes (*DEFL*s) in nodules have been reported only in the Inverted Repeat Loss Clade (IRLC) of legumes, including *Medicago truncatula*. Nodule *DEFL*s have a conserved, unique pattern of regulatory elements in upstream regions of genes. To understand their evolutionary patterns, we analyzed their expression and sequence polymorphisms, and estimated selection pressures shaping evolution of this unique class of genes among ecotypes of *Medicago truncatula*.
- Using custom *DEFL* arrays, expression and single feature polymorphisms (SFPs) were assessed for four ecotypes. Single nucleotide polymorphic (SNP) calls were used to estimate the variation in non-coding regions of

---

[5] This chapter will be submitted for publication to the journal *New Phytologist*, and is written in the style of the journal.

[6] The first author, Sumitha Nallu, has executed all of the experimental work and analysis reported in this chapter, and has prepared this draft for publication. And as a part of collaborative research, Peng Zhou and Nevin D. Young provided the SNP information for *DEFL*s used in this research.

genes. The nucleotide diversity at non-synonymous and synonymous sites in coding region was estimated using Nei and Gojobori method.

- Significant expression and nucleotide variation was observed among nodule *DEFL*s in the ecotypes of *M. truncatula.* The mature peptides of orthologous nodule *DEFL*s had signatures of both purifying and diversifying selection pressures, unlike the seed *DEFL*s, which predominantly exhibited purifying selection.
- The expression, sequence variation and the presence of diversifying selection in nodule *DEFL*s within the Medicago species indicates rapid and recent evolution and suggests that this family of genes is constantly evolving to adapt to different environments and acquiring new functions.

Introduction

Recent advances in genomic analysis of model plants and availability of their sequences have facilitated new prospects for understanding plant function and evolution. *Medicago truncatula* was chosen as a model species for study of legume symbiotic nitrogen fixation because of its relatively small, diploid genome, rapid generation time and self-pollination (Cook 1999). Through an international collaborative effort, multiple tools and resources are currently available for molecular, genetic and evolutionary studies in *M. truncatula* (Young and Udvardi 2009).

Availability of sequence information facilitates the study of large and diverse gene families. A large diverse family of cysteine-rich proteins was discovered in the nodules of *M. truncatula* (Fedorova et al., 2002; Mergaert et al., 2003; Graham et al., 2004), and subsequently found to be expressed in other developmental stages, including seeds, of legumes and other dicots (Silverstein et al., 2005; Silverstein et al., 2007). Because of the similarity of the gene sequence and structure of these cysteine-rich proteins to defensins, this gene family has been called Defensin-Like genes (*DEFL*s) (Graham et al., 2004; Silverstein et al., 2005; Silverstein et al., 2007). Using the Affymetrix *M. truncatula* Genome Array and later a more comprehensive *DEFL* array, the majority of the *DEFL*s were found to be expressed in the nodules of *M. truncatula* (Benedito et al., 2008; Chapter 2). Nodule *DEFL*s have unique expression and regulatory patterns (Chapter 3). Using a custom *DEFL* array for expression profiling across various nodule developmental stages and in response to different rhizobial mutants, we observed the nodule *DEFL* expression levels to be dependent on the volume of rhizobia present in the nodule. We also reported the presence of unique conserved regulatory elements in the 1 Kb upstream region of nodule *DEFL*s.

Nodule *DEFL*s have a distinctive phylogenetic lineage among legumes. To date, they have been found only in legumes belonging to the Inverted Repeat Loss clade (IRLC) within the subfamily Papilionoideae, and appear to be absent from *Lotus japonicus* and *Glycine max,* which occur in other clades within the Papilionoideae (Graham et al., 2004; Alunni et al., 2007), and which are

estimated to have diverged from the IRLC approximately 39 million years ago (Cannon et al., 2009). When rhizobia are released into the symbiosome compartment in determinate nodules of *L. japonicus* and *G. max,* the rhizobia retain the cell and genome size of free-living rhizobia, whereas in *M. truncatula* and related indeterminate nodule-forming legumes, the rhizobia terminally differentiate into enlarged bacteroids with an amplified genome (Mergaert et al., 2006). The correlation between *DEFL* expression in nodules and bacteroid differentiation appears to be causally related. Recently it was demonstrated that when one of the nodule *DEFL* (called nodulin cysteine rich peptide (*NCR*) in their study) was expressed in *L. japonicus,* terminal bacteroid differentiation resulted (Van de Velde, et al., 2010).

Since known nodule *DEFL*s are restricted to a single clade within legumes and paralogs show evidence of diversification (Graham et al., 2004; Alunni et al., 2007), we initiated a study to examine the expression and genomic polymorphism of nodule *DEFL*s within the ecotypes of *M. truncatula.* In the last decade, studies have addressed DNA polymorphisms within species using DNA markers like micro satellites, single nucleotide polymorphisms (SNPs) and single feature polymorphisms (SFP) in *Arabidopsis thaliana* (Zhu et al., 2003; Schmid et al., 2005; Borevitz et al., 2007) and *M. truncatula* (Ronfort et al., 2006). However, identifying sequence variation is only the first step in studying the evolutionary patterns of gene families. Availability of technology in recent years has prompted researchers to correlate the DNA variation with variation in the transcriptome and metabolome.

Gene expression variation among ecotypes can be caused by host-environment effects, sequence polymorphism in coding region, *cis* and *trans* regulatory elements and transcription factor binding sites. There are various reports correlating *cis* and *trans* acting promoter polymorphisms with expression differences (Cowles et al., 2002; Rockman et al., 2002; Caicedo et al., 2004). Nucleotide variation in the binding sites of mi- and siRNAs or in the 3'-UTR of the gene can also affect the gene expression levels (Mignone et al., 2002; Wilkie et al., 2003). Nucleotide polymorphism in coding regions leads to synonymous and non-synonymous substitutions in amino acids. There have been various studies on the intraspecific variation in the loci belonging to the NBS-LRR gene family in *A. thaliana* (Tian et al., 2002; Mauricio et al., 2003; Rose et al., 2004). Nucleotide diversity measurements in different regions of the NBS-LRR loci have been helpful to understand the selection pressures and evolutionary mechanisms of this large family of defense genes.

In this study, we evaluated expression and sequence polymorphisms of nodule *DEFL*s among four different ecotypes of *M. truncatula*. We also used two different rhizobial strains to assess the nodule *DEFL* expression differences caused by host-symbiont specificity. The nodule *DEFL*s displayed significant expression differences among the ecotypes but had subtle differences in gene expression because of strain difference. The sequence variation was estimated in terms of SFPs and SNPs. The pattern of DNA variation among the ecotypes correlated with the pattern observed in expression polymorphism. In addition to polymorphisms in the coding region that may lead to expression differences there

was variation observed in the upstream 1 kb putative promoter, downstream 1 kb putative 3'UTR and in a few introns. When examined closely some of the upstream SNPs mapped to the earlier reported *cis*-acting elements in upstream regions of nodule *DEFL*s. By estimating the nucleotide diversity of nodule *DEFL*s among the four ecotypes and comparing them to the seed *DEFL*s, we observed that the mature peptide of nodule *DEFL*s has signatures of both diversifying and purifying selection in *M. truncatula*.

Materials and Methods

Plant material and growth conditions

Seeds of *Medicago truncatula* ecotypes Jemalong A17 (hereafter called A17), DZA315-16, F83005-5 and R108-C3 were sterilized and germinated as described previously for A17 (Lohar et al., 2006). One day old germinated seedlings were grown on sterile plates (Corning 245mm * 245mm dishes) containing buffered nodulation medium (BNM, Ehrhardt et al., 1992), at pH 6.5, and 1.2% agar (plant tissue culture grade, Sigma). The radicals were placed on a moist, sterile germination paper on the surface of the medium and covered with a sterile black cotton cloth (Cotton Club Black, #074300603820, Wal Mart). Plates were maintained vertically in a growth chamber set at 16 h light / 8 h day, 25 to 21 ^{o}C, 200 to 300 µmol m^{-2} s^{-1} and 50% relative humidity. At 5 d after planting, plants were inoculated with 100 µL/root of a washed suspension of *Sinorhizobium meliloti*, Sm1021 (Meade et al., 1982) or *Sinorhizobium medicae* A321 (Rome et al., 1996) culture (OD600 = 0.05) in sterile water. Approximately

5 cm-long nodule-bearing root segments were harvested from inoculated plants at 14 days post-inoculation (dpi), root tips were removed from all plants at the time of harvest to reduce transcript contribution from the root tip. The root segments from all treatments were collected from three biological replicates, with a pool of multiple plants in each replicate, were frozen immediately in liquid nitrogen and stored at -80°C for subsequent RNA extraction. Total RNA extraction, probe preparation and array hybridization was performed as described in Chapter 2.

Microarray data analysis

Data were normalized across the different nodule treatments using SBQ normalization (Sato et al., 2007, Chapter 2). For all further analysis, only probes with at least one present call among the treatments were considered as nodule *DEFL*s. The ecotypes, ecotype-strain interaction and strain effects on nodule *DEFL* expression were calculated using 2-way ANOVA (P <= 0.01) using GeneSpring GX11 software from Agilent.

Genes that were differentially expressed among the four *M. truncatula* ecotypes and amongst or within a ecotype between inoculation treatments were identified using the Emperical Bayes for variance shrinkage (Cui et al., 2005) within the LIMMA (Smith, 2005) package distributed with R/Bioconductor (Gentleman, RC et al., 2004). Both a 2 fold-change cut off and Storey's False Discovery Rate (FDR) correction (P < = 0.01) were applied.

Single feature polymorphism (SFP) detection

The Probe Affinity Shape Power (PASP) method (Xu et al., 2009, https://dbw10.msi.umn.edu:8443/sfp) was used to detect SFPs from the probes of nodule *DEFL*s. Briefly, the normalized signal intensities were used to calculate the Perfect Match (PM) affinities for each probe and the expression index of the probe set was based on the affinity model (Li and Wong, 2001; Hubbell et al., 2002; Irizarry et al., 2003; Cui et al., 2005). SFP weight scores was calculated as described in Xu et al. (2009). A more stringent cut off was applied by filtering off all the SFPs with weight scores lower than 2.5 and finally only the common SFPs calls from the two strains between the pairs of ecotypes was reported.

Single Nucleotide Polymorphism (SNP) detection

The coding sequences of nodule *DEFL*s were mapped onto MT3.0 using BLAT (Kent 2002) with a 90% identity threshold. Using the mapped coordinates, SNPs were extracted in the coding region, 1 Kb upstream region, 1 Kb downstream region and introns of *DEFL* genes for ecotypes DZA315-16, F83005-5 and R108-C3 against the reference genome A17 Jemalong (www.medicagohapmap.org) using Perl scripts.

Perl scripts were also used for validation of a few random SFPs against SNPs, to map SFPs and SNPs to coding regions of differentially expressed nodule *DEFL*s and to map the SNPs to *cis*-regulatory elements in the differentially expressed nodule *DEFL*s.

Estimation of nodule *DEFL*s nucleotide diversity

All of the mapped *DEFL*s (as explained in SNP detection) were categorized as nodule and seed *DEFL*s if they had a present call in at least one treatment across the different nodule treatments (Chapter 3) and seed treatments (Chapter 2). The complete coding sequences of the nodule and seed *DEFLs* from the four ecotypes were recovered from the A17 reference sequence (Mt3.0) using SNP and Indel data (www.medicagohapmap.org). The sequences were aligned using Perl scripts. A total of 157 nodule *DEFL*s and 33 seed *DEFL*s with at least 75% sequence information were used to estimate the total (θ) and average (π) nucleotide diversity. The coding sequences of nodule and seed *DEFL*s were divided into signal peptide and mature peptide regions by SignalP 3.0 (Bendtsen et al., 2004). The (θ) and (π) values were estimated using the Nei-Gojobori method (Nei and Gojobori, 1986) from DnaSP v5.10 (Librado and Rozas 2009). To test the statistical significance of the differences in θ values among different regions in nodule and seed *DEFL*s, all the genes with $\theta_{non}/\theta_{syn} > 1$ were assigned as 1, $\theta_{non} = \theta_{syn}$ as 0 and $\theta_{non}/\theta_{syn} < 1$ as -1 and examined with Mann-Whitney U-test. All the steps were repeated for testing the statistical significance of the π values.

Results

Nodule *DEFL* expression is polymorphic among different ecotypes of *M. truncatula*.

To characterize the difference in expression levels of nodule *DEFL*s, we selected four different ecotypes of *M. truncatula*: the reference genome Jemalong A17, plus DZA315-16, F83005-5 and R108-C3. They were inoculated with strains of two rhizobial species: *S. meliloti* Sm1021 (Meade et al., 1982) and *S. medicae* A321 (Rome et al., 1996) to evaluate the expression differences caused due to host-symbiont interaction. The other strain, *S. medicae* was also chosen on the basis of an earlier report that *S. meliloti* 1021 is a poor nitrogen fixing partner for A17(Terpolilli et al., 2008). A two factorial analysis of variance (2-way ANOVA) indicated expression differences in nodule *DEFL*s as an effect of ecotype, ecotype x strain interaction and a subtle effect of strain alone (Fig. 4-1). Using the variance shrinkage method (Cui et al., 2005) we further assessed differentially expressed genes in the ecotype x strain effect, and found that the differentially expressed genes had subtle fold changes (all very close to a fold change of 2), supporting a conclusion that a difference in strains used for inoculation contributes minimally to expression differences among nodule *DEFL*s.

To assess the number of nodule *DEFL*s that were differentially expressed among *M. truncatula* under each of the inoculation regimes, we performed the variance shrinkage analysis on different pairs of ecotypes. R108-C3 showed the highest number of differentially expressed genes in comparison to A17 under both inoculation regimes followed by F83005-5 and the least differentially expressed genes are seen in DZA315-16 (Fig. 4-2).

Coding sequence polymorphisms among orthologous nodule *DEFL*s.

We used the 25-bp oligonucleotide features of the custom *DEFL* chip to estimate sequence polymorphisms in nodule *DEFL* coding sequences in pair-wise comparisons among the four ecotypes. A common set of genes with SFPs was extracted from the analysis of the two inoculation regimes for each ecotype pair. As seen in the case of expression polymorphism patterns, R108-C3 exhibited the highest sequence polymorphism followed by F83005-5 and the least sequence polymorphism seen in DZA315-16 in comparison to A17 (Table 4-1). From the current *M. truncatula* hapmap project, we were able to extract information on coding region SNPs for 242 nodule *DEFL*s for DZA315-16, F83005-5 or R108-C3 in reference to the A17 genome. The pattern of sequence polymorphism seen with SNP analysis was consistent with the pattern observed by SFP detection and expression polymorphism. R108-C3 again exhibited the highest occurrence of polymorphisms in comparison to A17 (Fig. 4-3). A random set of 20 nodule *DEFL*s with SFPs from A17 vs. DZA315-16 ecotype comparison were selected to confirm the SFP method of detection. We were able to confirm 68% of SFP calls by comparing them to the available SNP data (Table 4-S1). However, the 32% of SFPs which could be not validated do not represent the true false discovery rate (FDR). The Medicago hapmap project is in progress and there are a few missing sequence gaps especially in highly duplicated genes like nodule *DEFL*s. Hence, with the availability of updated sequence information there is a possibility of validating a higher percentage of SFPs. In addition, even

with the above mentioned limitations, our FDR is lower than the previously reported 40% FDR (Rostoks et al., 2005; Xu et al., 2009).

Sequence polymorphisms in *cis*-regulatory regions of nodule *DEFL* genes.

The expression polymorphisms observed among the pairwise comparisons of the ecotypes could be true expression level polymorphisms (ELPs) (Doerge 2002) or could be due to sequence polymorphisms that cause a probe mismatch between the ecotypes used and the probe sequence on the *DEFL* array, which results in a signal difference that inaccurately predicts expression differences. We categorized the nodule *DEFL*s into 4 groups: a) differentially expressed genes with SFPs b) differentially expressed genes with no SFPs c) genes with SFPs but not differentially expressed and d) genes with no SFPs and no differential expression for pairwise comparisons of ecotypes to A17, when inoculated with two different strains (Table 4-2). We further filtered the group of differentially expressed genes with no SFPs using the SNP data. The results from the *DEFL* array should be the most accurate representation of gene expression for the group of *DEFL*s that lack evidence of sequence polymorphisms in their coding regions. For those nodule *DEFL*s that are differentially expressed but have no SFPs or SNPs in their coding regions, we mapped the occurrence of SNPs onto their 1 kb upstream, downstream regions from the start and stop translational sites of the nodule *DEFL*s respectively and also included the intronic regions if present in the genes. This set of differentially expressed genes exhibited SNPs and Indels in other regions of the gene. We

had earlier reported the presence of conserved putative *cis*-acting regulatory elements in the upstream region of nodule *DEFL*s that are required for nodule-specific expression (Chapter 3). Some of the SNPs in the upstream regions of the nodule *DEFL*s were mapped onto the *cis*-acting regulatory elements suggesting that these nucleotide changes in the regulatory motifs might be one of the reasons for the observed differential expression of nodule *DEFL*s (Fig. 4-4, Table 4-S2).

Nucleotide diversity of nodule *DEFL*s in *M. truncatula*

The *DEFL*s were classified as nodule and seed *DEFL*s sets based on expression patterns, following the criteria described in materials and methods. To determine the evolutionary pressures acting on nodule *DEFL*s, the total nucleotide diversity (θ) and the average nucleotide diversity (π) were estimated at synonymous and nonsynonymous sites for the four ecotypes of *M. truncatula.* Since the Medicago hapmap project is in progress, the SNP data especially for highly duplicated genes like *DEFL*s is incomplete. We obtained sequence information for a total of 157 nodule *DEFL*s and 33 seed *DEFL*s, a significant representative set of genes for each group. The θ_{non}, π_{non} and θ_{syn}, π_{syn} were calculated for the entire coding region, and separately for regions that code for the signal and mature peptides of the nodule and seed *DEFL*s. From the nonsyn/syn ratios observed in different regions of nodule and seed *DEFL*s, the signal peptides of both the nodule and seed DEFLs were conserved with more than half displaying no substitutions and the majority of the other half exhibiting

purifying selection (Fig. 4-5a). Mature polypeptides of nodule DEFLs exhibit evidence of both purifying and diversifying selection whereas the mature polypeptides of seed DEFLs are mostly under purifying selection (Fig. 4-5b). To test the statistical significance of the differences in rates of substitutions among different regions in nodule and seed *DEFL*s, we designated all the genes with nonsyn/syn >1 as 1, nonsyn=syn as 0 and nonsyn/syn < 1 as -1 and employed the Mann-Whitney U-test. We could not use the raw nonsyn/syn values for analysis because many of the genes had no synonymous substitutions and in such cases nonsyn/syn would give an undefined number. The Mann-Whitney U tests indicated a significant difference within the full coding regions and the mature polypeptides of nodule vs. seed DEFLs. There was no significant difference in substitution rates observed between the signal peptides of nodule vs. seed DEFLs, or between the signal peptides vs. mature polypeptides of the seed DEFLs. Among nodule DEFLs, the mature polypeptides were significantly different in their selection pressures from the signal peptides (Table 4-3).

Discussion

Nodule *DEFL*s have similar trends in expression and sequence polymorphism in *M. truncatula*

The four ecotypes used in this study were selected because they were from different ecological zones (Ronfort et al., 2006). A17 is the reference genome for *M. truncaula* sequencing project (http://www.medicago.org/genome/),

R108 has been the common choice for generating Medicago insertion mutants and both DZA315-16, F83005-5 along with A17 have been used as parents for Recombinant Inbred Lines (RILs). Additionally all the four ecotypes were resequenced under the Medicago Hapmap Project (http://www.medicagohapmap.org). We found a significant number of differentially expressed nodule *DEFL*s among the four ecotypes of *M. truncatula.*

A difference in nodule *DEFL* gene expression patterns in *M. truncatula* ecotypes could possibly result from differences in host/endophyte specificity, resulting in global differences in nodule effectiveness and nodulin expression, rather than specific differences in *DEFL* regulation. To address this possibility, we inoculated each ecotype with two different strains of rhizobia to evaluate the difference in expression patterns of nodule *DEFL*s in response to different symbiotic partners. *S. meliloti* 1021 has been a common strain of study for symbiosis in *M. truncatula,* due to extensive analysis of symbiotic genes in this genetic background, and with an available genome sequence (Galibert et al., 2001). *S. meliloti* 1021 strain was recently reported to be a poor partner compared to two other *Sinorhizobium* strains for nitrogen fixation in *M. truncatula* (Terpolilli et al., 2008). We therefore chose *S. medicae* A321 as a second strain for inoculation of the four ecotypes. Inoculation with these two strains resulted in minimal differences in the expression patterns of nodule *DEFL*s, suggesting either that the two strains are equally effective in nodulation of *M. truncatula* under our experimental conditions, or that strains differing in nitrogen fixing abilities are not a significant factor in nodule *DEFL* induction. In support of the

second possibility, we have previously seen in Chapter 3 that the nodule *DEFL* expression levels differ little between nodules infected by wild type Sm1021 and nodules infected with a *nifH* mutant that is deficient in fixing nitrogen (Hirsch et al., 1983).

In addition to expression polymorphisms among the ecotypes, we hypothesized that the *DEFL*s would also demonstrate sequence polymorphism among the ecotypes. We used two approaches to evaluate *DEFL* sequence polymorphism. Single-Feature Polymorphism (SFP) is a rapid and cost-effective technique for scanning genomic polymorphism, especially for the species whose genomic sequence is unknown or under construction (Hazen and Kay et al., 2003). First, we used the hybridization results from our analysis of *DEFL* expression using the custom *DEFL* array to scan for SFPs among the four ecotypes. Secondly, we used available genome sequence from the Medicago HapMap project for the four ecotypes. Among the four ecotypes, R108-C3 had the highest sequence variation in comparison to A17 under both inoculation regimes followed by F83005-5 and the least sequence variation was seen with DZA315-16. This trend was similar to that observed in expression polymorphism. As of date, there are no reports available on the evolutionary distances among the four ecotypes. By examining the trends in expression and sequence variation of nodule *DEFL*s, it can be speculated that the reference genome A17 is closest to DZA315-16 followed by F83005-5 and R108-C3. R108-C3 probably has the highest evolutionary divergence from the other three ecotypes while, DZA315-16 and F83005-5 are the closest in evolutionary distance.

Borevitz et al. (2007) examined the genomic DNA diversity in 23 wild strains (accessions) of *A. thaliana* and reported that the frequency of sequence variation (SFPs) is higher in gene families like NBS-LRR (R genes), receptor-like proteins (RLP) and S-locus protein kinase genes (SRK) compared to basic Helix-Loop-Helix (bHLH) transcription factor gene family, RING-type E3 Ubiquitin ligase proteins and F-box-containing subunits of the SCF proteasome. Members of R and RLP gene families are involved in disease resistance with a pathogen-recognition domain that is a well documented target of diversifying selection and rapid evolution (Mondragon-Palomino and Gaut, 2005)**.** The SRK genes are receptors for S-locus cysteine rich (SCR) genes that are a sub family of *DEFL* genes expressed in inflorescences and siliques in *A. thaliana* and are involved in reproductive regulatory roles (Nasrallah et al., 2002; Chapter 2). Since these families of genes are closely related to nodule *DEFL*s in gene structure, genome organization and potential functional roles, it can be suggested that nodule *DEFL*s probably are rapidly evolving in *M. truncatula*. Comparison of nucleotide diversity to other gene families in *M. truncatula* would help in further testing of the hypothesis.

Polymorphism is observed in regulatory elements of nodule *DEFL*s

Expression variation within a species may correlate with nucleotide heterozygosity (Ronald et al., 2005, Kliebenstein et al., 2006). However, in addition to expression variation due to true expression level polymorphisms (ELPs) (Doerge 2002), sequence differences between the target cDNA from

various ecotypes and the probe set on the array can result in a probe mismatch. In the latter case, array hybridization may not be an accurate report of expression differences among ecotypes. We defined a subset of differentially expressed nodule *DEFL*s that had no evidence of sequence polymorphism (no SFPs and SNPs) in their coding regions, these set of nodule *DEFL*s were likely to be true ELPs and were used to identify putative regulatory differences. Our report of the presence of SNPs in the *cis*-acting regulatory elements in the 1 Kb upstream region of nodule *DEFL*s (Chapter 3) is in agreement with the studies showing the importance of *cis* elements in regulation of expression of genes (Cowles et al., 2002, Rockman et al., 2002) though additional research is required to establish the effect of SNPs in these conserved regulatory elements on the expression patterns of nodule *DEFL*s.

We also observed substantial levels of nucleotide polymorphism in the 1 kb downstream region which is a putative 3'UTR of nodule *DEFL*s. The 3'UTR contains binding sites for miRNAs and AU-rich elements which affect the mRNA stability and thus influence the gene expression (Wilkie et al., 2003). A recent report identified 100 novel miRNA in roots and nodules of *M. truncatula* and two of the nodule *DEFL*s (called *NCR* in the study) were found to be putative targets for a few miRNA's (Lelandais-Briere et al., 2009). We predict that in addition to the upstream regulatory elements found in nodule *DEFL*s there are additional binding sites found in the 3'UTR regions of nodule *DEFL*s, which could not be predicted by the methods used in our earlier report on regulation of nodule *DEFL*s (Chapter 3).

Nodule *DEFL*s differ from seed *DEFL*s in evolutionary rates

*DEFL*s are closely related to the diverse defensin family of genes (Graham et al., 2004). The evolutionary patterns within the paralogous subgroups of β-defensins in mammals indicate patterns of purifying or diversifying selection (Maxwell et al., 2003, Semple et al., 2003, Semple et al., 2006). Another functionally related NBS/LRR family of R genes also show similar trends in selective pressures within the paralogous subgroups (Baumgarten et al., 2003, Meyers et al., 2003) as well as orthologous groups in *A. thaliana* (Bakker et al., 2006). We observed that equal proportion of orthologous groups of nodule *DEFL*s exhibit patterns of diversifying or purifying selection in their coding regions, which is in contrast to nearly 73% of seed *DEFL*s exhibiting purifying selection. Further analysis of the mature peptide regions of nodule and seed DEFLs had similar results as observed with the entire coding region. However, the signal peptides of both the nodule and seed DEFLs were under purifying selection. Previously, it has been reported that paralogous clusters of DEFLs in *A. thaliana* (Silverstein et al., 2005) and *M. truncatula* (Alunni et al., 2007) show patterns of diversifying or purifying selection in the mature peptide region of the gene whereas the signal peptide is predominantly under purifying selection. Nodule DEFLs (called NCRs in their study) are targeted into the secretory pathway by their signal peptides (Mergaert et al., 2003), and were later found to be targeted specifically to the bacterial membrane and cytosol (Van de Velde et al., 2010). Absence of DNF1-1, a signal peptidase expressed in *M. truncatula*

nodules (Wang et al., 2010) in the *dnf1-1* mutant disrupted the localization of the nodule DEFLs (NCRs) indicating the importance of signal peptide in proper post translational processing and subcellular localization of the nodule DEFL peptides. The presence of purifying selection in signal peptides among orthologous nodule and seed DEFLs indicates that the signal peptide is the most conserved part of DEFLs irrespective of the species.

The differences in the selection pressures exhibited by the mature peptides of nodule and seed DEFLs may reflect their different times of origin and diverse path of gene duplications. Seed *DEFL*s are seen in different clades of legumes and non-legumes (Graham et al., 2004, Silverstein et al., 2007), whereas nodules *DEFL*s are reported only in a narrow clade within the legume family (Mergaert et al., 2003, Graham et al., 2004). Because the seed *DEFL*s have a more ancient lineage and more conserved sequences, their functions may have become fixed before the diversification of the legumes. Nodule *DEFL*s have been reported only in the legumes with swollen bacteroids, a trait where the rhizobia are differentiated into swollen and branched nitrogen fixing forms. They are found to be absent in legumes with non swollen bacteroids. It has been recently reported that the swollen bacteroid is a derived trait and non swollen bacteroid is an ancestral trait (Oono et al., 2010). This suggests that nodule *DEFL*s, though extremely numerous in the *M. truncatula* genome, are a recently derived innovation in legumes. Our analysis of orthologous nodule *DEFL*s in four ecotypes of *M. truncatula* indicates that many loci are under purifying selection, while a large proportion of others exhibit patterns of diversifying selection, which

may result in new functions by introducing non synonymous substitutions into the coding region for the mature polypeptide. There are different ways a gene pair may evolve following duplication, such as through partitioning the original function between the two paralogs (subfunctionalization) or by acquisition of a new function by one copy while the other retains the original function (neofunctionalization). It has been reported that subfunctionalization and neofunctionalization can coexist, a condition referred to as subneofunctionalization (He and Zhang 2005, Rastogi and Liberles 2005). We predict from our observations of purifying and diversifying selection signatures in the mature peptide regions of orthologous nodule DEFLs that these genes are one of the good examples for subneofunctionalization. The set of orthologous nodule DEFLs under purifying selection probably share the same function (subfunctionalization) and the other set under diversifying selection probably acquired new functions (neofunctionalization) though additional functional studies on nodule DEFLs are necessary to establish it.

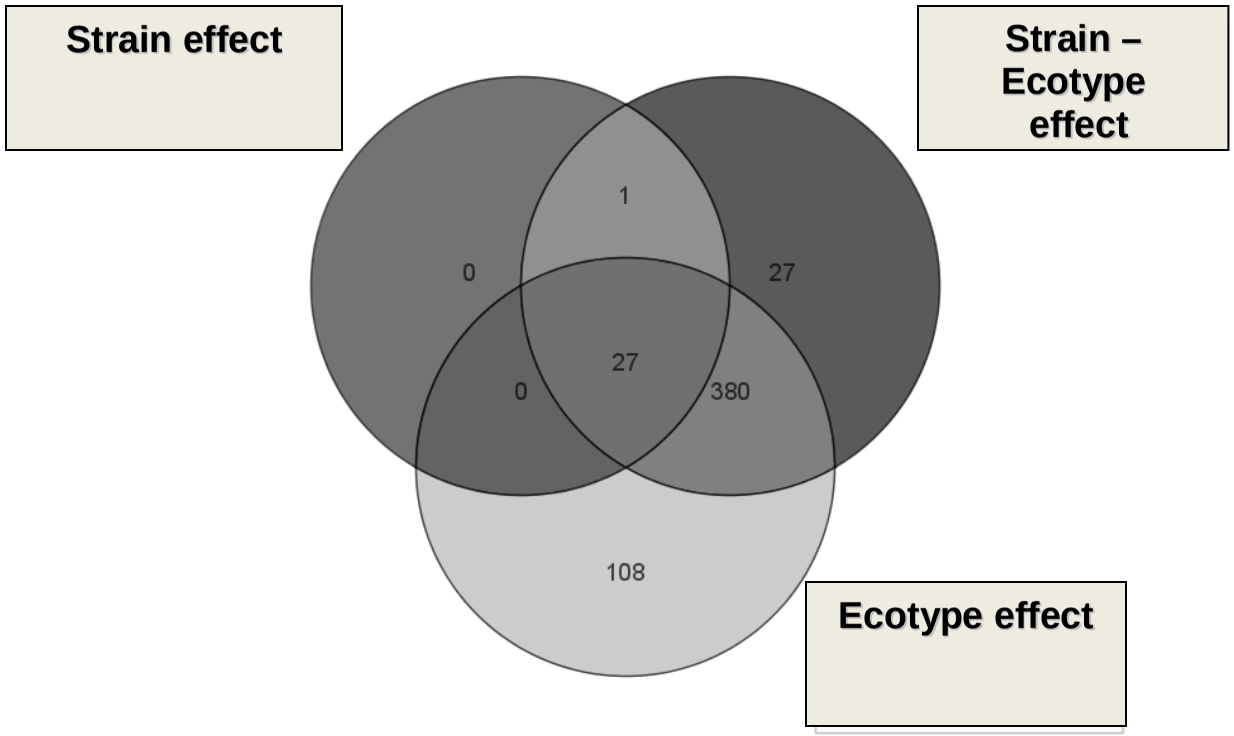


Figure 4-1 2-way ANOVA for nodule *DEFL* expression differences. The numbers of differentially expressed genes, at p <= 0.01, under different effects are represented as a Venn diagram.

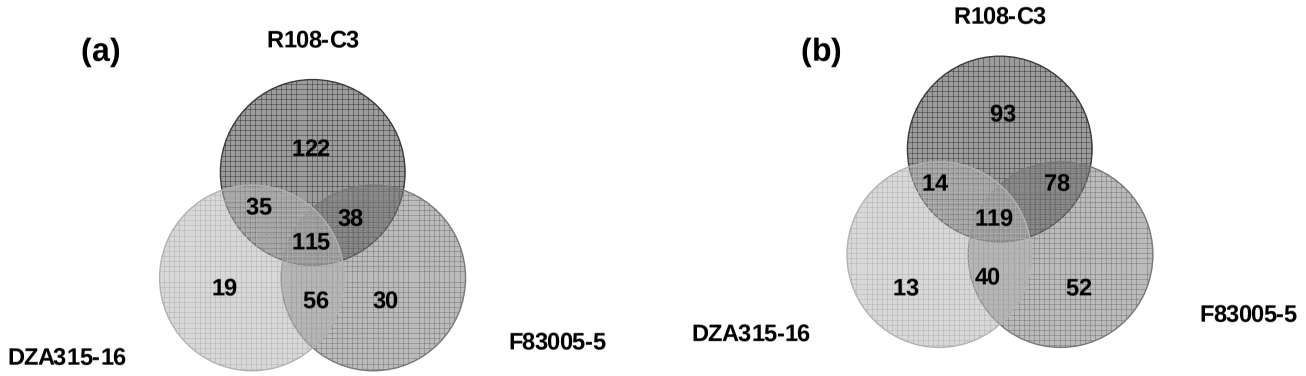


Figure 4-2 Differential expression of nodule *DEFL*s among *M. truncatula* ecotypes. Numbers indicate genes that are differentially expressed in reference to A17 when inoculated with (a) *S. medicae* or (b) *S. meliloti*. The differentially expressed genes were estimated using variance shrinkage method with a p-value <= 0.01 and a fold change cut-off of 2.0.

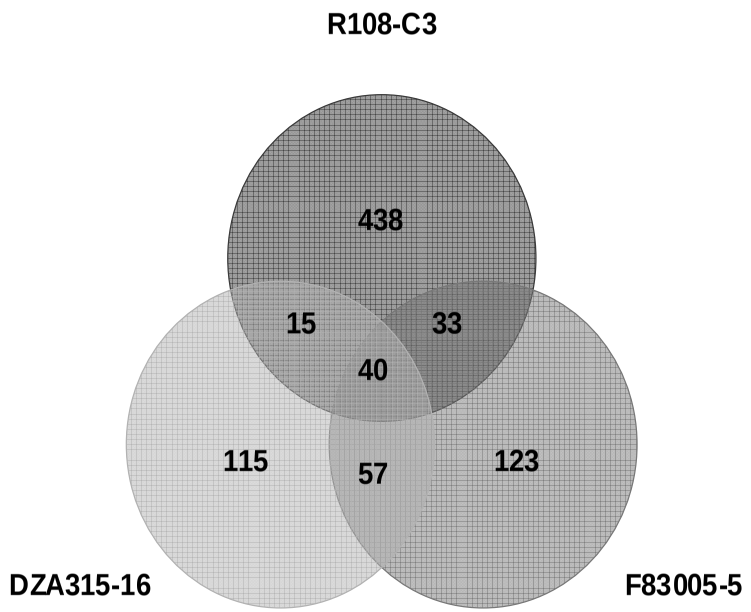


Figure 4-3 Number of Single Nucleotide Polymorphism (SNPs) observed in different ecotypes against the reference genome A17 in the coding region of nodule *DEFL*s.

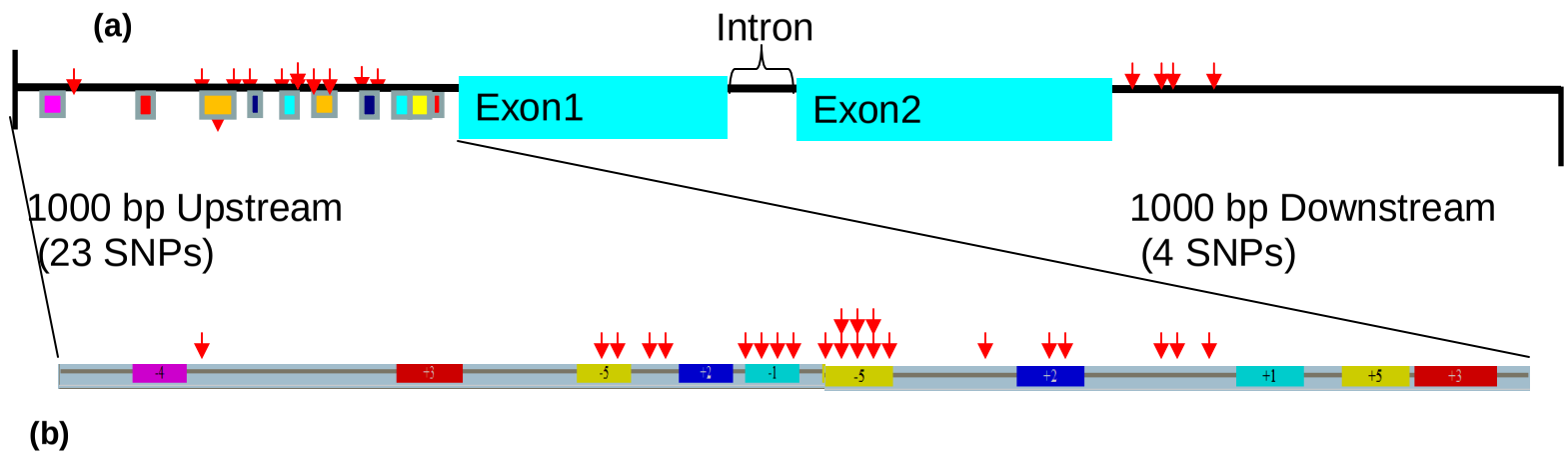


Figure 4-4 Mapping of SNPs to non-coding regions of a nodule *DEFL* that is differentially expressed among ecotypes. a) MtgAC171166_66001 is absent from R108 but expressed at moderate levels in A17 with no SNPs and SFPs in coding region but SNPs (red arrows) in upstream and downstream regions. b) 1 kb region upstream region of MtgAC171166_66001 with SNPs mapped onto the *cis*-regulatory elements.

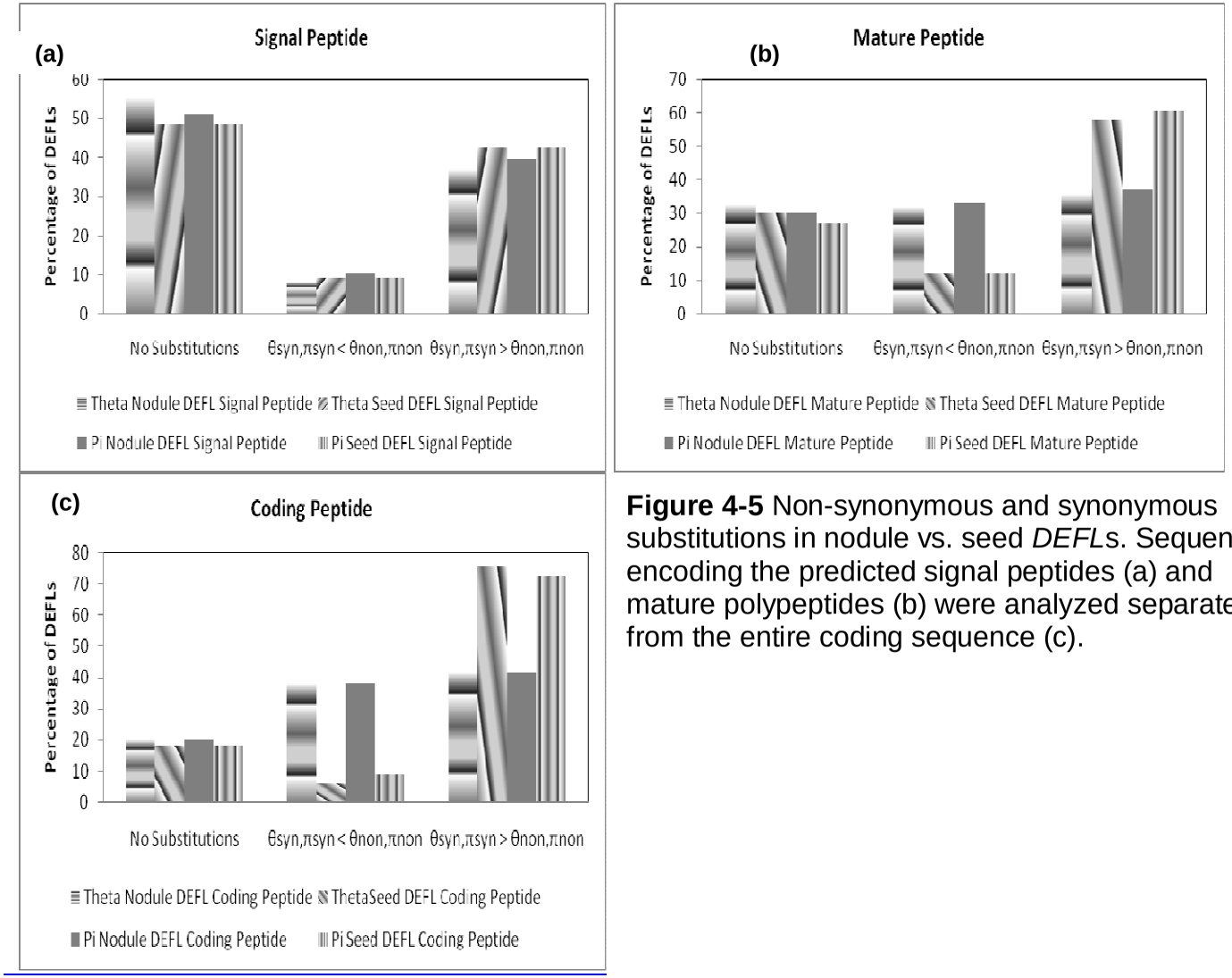


Figure 4-5 Non-synonymous and synonymous substitutions in nodule vs. seed *DEFL*s. Sequences encoding the predicted signal peptides (a) and mature polypeptides (b) were analyzed separately from the entire coding sequence (c).

Table 4-1 Nodule *DEFL*s with Single Feature Polymorphisms observed between pairs of ecotypes

	A17	DZA315-16	F83005-5
DZA315-16	202		
F83005-5	220	156	
R108-C3	388	416	411

The data indicate the set of nodule *DEFL*s containing SFPs that were common between the two inoculation regimes for each of the pair-wise comparison of *M. truncatula* ecotypes.

Table 4-2 Nodule *DEFL*s categorized into different groups based on expression and sequence polymorphisms.

	Differentially expressed	**SFP**	***S. meliloti***[a]	***S. medicae***[b]
A17 VS DZA315-16	**+**	**+**	138	119
	+	**-**	72	76
	-	**+**	165	164
	-	**-**	197	207
A17 VS F83005-5	**+**	**+**	156	150
	+	**-**	80	86
	-	**+**	164	186
	-	**-**	123	154
A17 VS R108-C3	**+**	**+**	263	254
	+	**-**	41	57
	-	**+**	204	215
	-	**-**	68	51

+/- indicates presence/absence in the differentially expressed genes group or Single feature polymorphism (SFP) group. Numbers of nodule *DEFL*s observed in each group when inoculated with [a]*S. meliloti* and [b]*S. medicae* are represented in two different columns.

Table 4-3 Mann-Whitney p-values of nodule and seed *DEFL*s

	nodule *DEFL*s - seed *DEFL*s	nodule *DEFL*s - nodule *DEFL*s	seed *DEFL*s - seed *DEFL*s
Coding region	0.00034368 ***	——	——
Signal Peptide	0.722444	——	——
Mature Peptide	0.0035406 ***	——	——
Signal-Mature peptide	——	0.000768054 ***	0.400906

*** denotes significance at p<= 0.05, two-tailed test. It is observed that the total coding and mature peptide of nodule *DEFL*s is significantly different from seed *DEFL*s in their θ and π values.

BIBLIOGRAPHY

AbuQamar S, Chen X, Dhawan R, Bluhm B, Salmeron J, Lam S, Dietrich RA, Mengiste T (2006) Expression profiling and mutant analysis reveals complex regulatory networks involved in Arabidopsis response to Botrytis infection. Plant J **48:** 28-44

Allen A, Snyder AK, Preuss M, Nielsen EE, Shah DM, Smith TJ (2008) Plant defensins and virally encoded fungal toxin KP4 inhibit plant root growth. Planta **227:** 331-339

Alunni B, Kevei Z, Redondo-Nieto M, Kondorosi A, Mergaert P, Kondorosi E (2007) Genomic organization and evolutionary insights on GRP and NCR genes, two large nodule-specific gene families in Medicago truncatula. Mol Plant Microbe Interact **20:** 1138-1148

Ameline-Torregrosa C, Cazaux M, Danesh D, Chardon F, Cannon SB, Esquerre-Tugaye MT, Dumas B, Young ND, Samac DA, Huguet T, Jacquet C (2008) Genetic dissection of resistance to anthracnose and powdery mildew in Medicago truncatula. Mol Plant Microbe Interact **21:** 61-69

Andriankaja A, Boisson-Dernier A, Frances L, Sauviac L, Jauneau A, Barker DG, de Carvalho-Niebel F (2007) AP2-ERF transcription factors mediate Nod factor dependent Mt ENOD11 activation in root hairs via a novel cis-regulatory motif. Plant Cell **19:** 2866-2885

Ane JM, Zhu H, Frugoli J (2008) Recent Advances in Medicago truncatula Genomics. Int J Plant Genomics **2008:** 256597

Asai T, Tena G, Plotnikova J, Willmann MR, Chiu WL, Gomez-Gomez L, Boller T, Ausubel FM, Sheen J (2002) MAP kinase signalling cascade in Arabidopsis innate immunity. Nature **415:** 977-983

Bailey PC, Martin C, Toledo-Ortiz G, Quail PH, Huq E, Heim MA, Jakoby M, Werber M, Weisshaar B (2003) Update on the basic helix-loop-helix transcription factor gene family in Arabidopsis thaliana. Plant Cell **15:** 2497-2502

Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol **2:** 28-36

Bailey TL, Gribskov M (1998) Combining evidence using p-values: application to sequence homology searches. Bioinformatics **14:** 48-54

Bakker EG, Stahl EA, Toomajian C, Nordborg M, Kreitman M, Bergelson J (2006) Distribution of genetic variation within and among local populations of Arabidopsis thaliana over its species range. Mol Ecol **15:** 1405-1418

Barnes DK, Ostazeski SA, Schillinger JA, Hanson CH (1969) Effect of antrhracnose (*Colletotrichum trifolii*) infection on yield, stand and vigor of alfalfa. Crop Sci **9:** 344-346

Baumgarten A, Cannon S, Spangler R, May G (2003) Genome-level evolution of resistance genes in Arabidopsis thaliana. Genetics **165:** 309-319

Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol **340:** 783-795

Benedito VA, Torres-Jerez I, Murray JD, Andriankaja A, Allen S, Kakar K, Wandrey M, Verdier J, Zuber H, Ott T, Moreau S, Niebel A, Frickey T, Weiller G, He J, Dai X, Zhao PX, Tang Y, Udvardi MK (2008) A gene expression atlas of the model legume Medicago truncatula. Plant J **55:** 504-513

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol **57**: 289–300

Boisson-Dernier A, Chabaud M, Garcia F, Bécard G, Rosenberg C, Barker GD (2001)
Agrobacterium rhizogenes-Transformed Roots of *Medicago truncatula* for the Study of Nitrogen-Fixing and Endomycorrhizal Symbiotic Associations. Mol Plant Microbe Interact **14:** 695-700

Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics **19:** 185-193

Boman HG (1995) Peptide antibiotics and their role in innate immunity. Annu Rev Immunol **13:** 61-92

Borevitz JO, Hazen SP, Michael TP, Morris GP, Baxter IR, Hu TT, Chen H, Werner JD, Nordborg M, Salt DE, Kay SA, Chory J, Weigel D, Jones JD, Ecker JR (2007) Genome-wide patterns of single-feature polymorphism in Arabidopsis thaliana. Proc Natl Acad Sci U S A **104:** 12057-12062

Busch W, Lohmann JU (2007) Profiling a plant: expression analysis in Arabidopsis. Curr Opin Plant Biol **10:** 136-141

Caicedo AL, Stinchcombe JR, Olsen KM, Schmitt J, Purugganan MD (2004) Epistatic interaction between Arabidopsis FRI and FLC flowering time genes generates a latitudinal cline in a life history trait. Proc Natl Acad Sci U S A **101:** 15670-15675

Cannon SB, May GD, Jackson SA (2009) Three sequenced legume genomes and many crop species: rich opportunities for translational genomics. Plant Physiol **151:** 970-977

Chandler MR, Date RA, Roughley RJ. (1982) Infection and root-nodule development in *Stylosanthes* species by *Rhizobium*. J Exp Bot **33**: 47-57

Cheng HP, Walker GC (1998) Succinoglycan is required for initiation and elongation of infection threads during nodulation of alfalfa by Rhizobium meliloti. J Bacteriol **180:** 5183-5191

Cook DR (1999) Medicago truncatula--a model in the making! Curr Opin Plant Biol **2:** 301-304

Cowles CR, Hirschhorn JN, Altshuler D, Lander ES (2002) Detection of regulatory variation in mouse genes. Nat Genet **32:** 432-437

Cui X, Hwang JT, Qiu J, Blades NJ, Churchill GA (2005) Improved statistical tests for differential gene expression by shrinking variance components estimates. Biostatistics **6:** 59-75

Cui X, Xu J, Asghar R, Condamine P, Svensson JT, Wanamaker S, Stein N, Roose M, Close TJ (2005) Detecting single-feature polymorphisms using oligonucleotide arrays and robustified projection pursuit. Bioinformatics **21:** 3852-3858

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

Doerge RW (2002) Mapping and analysis of quantitative trait loci in experimental populations. Nat Rev Genet **3:** 43-52

Dresselhaus T, Marton ML (2009) Micropylar pollen tube guidance and burst: adapted from defense mechanisms? Curr Opin Plant Biol **12:** 773-780

Ehrhardt DW, Atkinson EM, Long SR (1992) Depolarization of alfalfa root hair membrane potential by Rhizobium meliloti Nod factors. Science **256:** 998-1000

Fedorova M, van de Mortel J, Matsumoto PA, Cho J, Town CD, VandenBosch KA, Gantt JS, Vance CP (2002) Genome-wide identification of

nodule-specific transcripts in the model legume Medicago truncatula. Plant Physiol **130:** 519-537

Frith MC, Fu Y, Yu L, Chen JF, Hansen U, Weng Z (2004) Detection of functional DNA motifs via statistical over-representation. Nucleic Acids Res **32:** 1372-1381

Frühling M, Albus U, Hohnjec N, Geise G, Pühler A, Perlick MA (2000) A small gene family of broad bean codes for late nodulins containing conserved cysteine clusters. Plant Science **152:** 67-77

Fujita M, Fujita Y, Noutoshi Y, Takahashi F, Narusaka Y, Yamaguchi-Shinozaki K, Shinozaki K (2006) Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. Curr Opin Plant Biol **9:** 436-442

Gagne JM, Downes BP, Shiu SH, Durski AM, Vierstra RD (2002) The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in Arabidopsis. Proc Natl Acad Sci U S A **99:** 11519-11524

Galibert F, Finan TM, Long SR, Puhler A, Abola P, Ampe F, Barloy-Hubler F, Barnett MJ, Becker A, Boistard P, Bothe G, Boutry M, Bowser L, Buhrmester J, Cadieu E, Capela D, Chain P, Cowie A, Davis RW, Dreano S, Federspiel NA, Fisher RF, Gloux S, Godrie T, Goffeau A, Golding B, Gouzy J, Gurjal M, Hernandez-Lucas I, Hong A, Huizar L, Hyman RW, Jones T, Kahn D, Kahn ML, Kalman S, Keating DH, Kiss E, Komp C, Lelaure V,

Masuy D, Palm C, Peck MC, Pohl TM, Portetelle D, Purnelle B, Ramsperger U, Surzycki R, Thebault P, Vandenbol M, Vorholter FJ, Weidner S, Wells DH, Wong K, Yeh KC, Batut J (2001) The composite genome of the legume symbiont Sinorhizobium meliloti. Science **293:** 668-672

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol **5:** R80

Glazebrook J (2005) Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. Annu Rev Phytopathol **43:** 205-227

Glazebrook J, Ausubel FM (1994) Isolation of phytoalexin-deficient mutants of Arabidopsis thaliana and characterization of their interactions with bacterial pathogens. Proc Natl Acad Sci U S A **91:** 8955-8959

Glazebrook J, Ichige A, Walker GC (1993) A Rhizobium meliloti homolog of the Escherichia coli peptide-antibiotic transport protein SbmA is essential for bacteroid development. Genes Dev **7:** 1485-1497

Godiard L, Niebel A, Micheli F, Gouzy J, Ott T, Gamas P (2007) Identification of New Potential Regulators of the *Medicago truncatula*–*Sinorhizobium meliloti* Symbiosis Using a Large-Scale Suppression Subtractive Hybridization Approach. Mol Plant Microbe Interact **20:** 321-332

Graham MA, Silverstein KA, Cannon SB, VandenBosch KA (2004) Computational identification and characterization of novel genes from legumes. Plant Physiol **135:** 1179-1197

Gutierrez-Marcos JF, Costa LM, Biderre-Petit C, Khbaya B, O'Sullivan DM, Wormald M, Perez P, Dickinson HG (2004) maternally expressed gene1 Is a novel maize endosperm transfer cell-specific gene with a maternal parent-of-origin pattern of expression. Plant Cell **16:** 1288-1301

Gyorgyey J, Vaubert D, Jimenez-Zurdo JI, Charon C, Troussard L, Kondorosi A, Kondorosi E (2000) Analysis of Medicago truncatula nodule expressed sequence tags. Mol Plant Microbe Interact **13:** 62-71

Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK,Jr, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, Salzberg SL, White O (2003) Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res **31:** 5654-5666

Hagen G, Guilfoyle T (2002) Auxin-responsive gene expression: genes, promoters and regulatory factors. Plant Mol Biol **49:** 373-385

Hanks JN, Snyder AK, Graham MA, Shah RK, Blaylock LA, Harrison MJ, Shah DM (2005) Defensin gene family in Medicago truncatula: structure, expression and induction by signal molecules. Plant Mol Biol **58:** 385-399

Hansen AC, Busk H, Marcker A, Marcker KA, Jensen EO (1999) VsENBP1 regulates the expression of the early nodulin PsENOD12B. Plant Mol Biol **40:** 495-506

Hazen SP, Kay SA (2003) Gene arrays are not just for measuring gene expression. Trends Plant Sci **8:** 413-416

He J, Benedito VA, Wang M, Murray JD, Zhao PX, Tang Y, Udvardi MK (2009) The Medicago truncatula gene expression atlas web server. BMC Bioinformatics **10:** 441

He X, Zhang J (2005) Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics **169:** 1157-1164

Heard J, Dunn K (1995)
Symbiotic induction of a MADS-box gene during development of alfalfa root nodules. Proc Natl Acad Sci USA **92:** 5273-5277

Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Res **27:** 297-300

Hirsch AM, Bang M, Ausubel FM (1983) Ultrastructural analysis of ineffective alfalfa nodules formed by nif::Tn5 mutants of Rhizobium meliloti. J Bacteriol **155:** 367-380

Huang H, Tudor M, Su T, Zhang Y, Hu Y, Ma H (1996) DNA binding properties of two Arabidopsis MADS domain proteins: binding consensus and dimer formation. Plant Cell **8:** 81-94

Hubbell E, Liu WM, Mei R (2002) Robust estimators for expression analysis. Bioinformatics **18:** 1585-1592

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature **335:** 167-170

Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP (2003) Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res **31:** e15

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics **4:** 249-264

Jensen EO, Marcker KA, Schell J, Bruijn FJ (1988) Interaction of a nodule specific, trans-acting factor with distinct DNA elements in the soybean leghaemoglobin Ibc(3) 5' upstream region. EMBO J **7:** 1265-1271

Jones KM, Kobayashi H, Davies BW, Taga ME, Walker GC (2007) How rhizobial symbionts invade plants: the Sinorhizobium-Medicago model. Nat Rev Microbiol **5:** 619-633

Jones-Rhoades MW, Borevitz JO, Preuss D (2007) Genome-wide expression profiling of the Arabidopsis female gametophyte identifies families of small, secreted proteins. PLoS Genet **3:** 1848-1861

Kafatos FC, Efstratiadis A, Forget BG, Weissman SM (1977) Molecular evolution of human and rabbit beta-globin mRNAs. Proc Natl Acad Sci USA **74:** 5618-5622

Katagiri F, Thilmony R, He SY (2002) The *Arabidopsis thaliana-Pseudomonas syringae* interaction. The Arabidopsis book. American Society of Plant Biologists. doi:10.1199/tab.0039

Kaijalainen S, Schroda M, Lindstrom K (2002) Cloning of nodule-specific cDNAs of Galega orientalis. Physiol Plant **114:** 588-593

Kardailsky I, Yang WC, Zalensky A, van Kammen A, Bisseling T (1993) The pea late nodulin gene PsNOD6 is homologous to the early nodulin genes PsENOD3/14 and is expressed after the leghaemoglobin genes. Plant Mol Biol **23:** 1029-1037

Kent WJ (2002) BLAT--the BLAST-like alignment tool. Genome Res **12:** 656-664

Kimura M (1983) *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge, U.k.

Kliebenstein DJ, West MA, van Leeuwen H, Loudet O, Doerge RW, St Clair DA (2006) Identification of QTLs controlling gene expression networks defined a priori. BMC Bioinformatics **7:** 308

Kling J (2005) The search for a sequencing thoroughbred. Nat Biotechnol **23:** 1333-1335

Kozaki A, Hake S, Colasanti J (2004) The maize ID1 flowering time regulator is a zinc finger protein with novel DNA binding properties. Nucleic Acids Res **32:** 1710-1720

Lamblin AF, Crow JA, Johnson JE, Silverstein KA, Kunau TM, Kilian A, Benz D, Stromvik M, Endre G, VandenBosch KA, Cook DR, Young ND, Retzel EF (2003) MtDB: a database for personalized data mining of the model legume Medicago truncatula transcriptome. Nucleic Acids Res **31:** 196-201

Laubinger S, Zeller G, Henz SR, Sachsenberg T, Widmer CK, Naouar N, Vuylsteke M, Scholkopf B, Ratsch G, Weigel D (2008) At-TAX: a whole genome tiling array resource for developmental expression analysis and transcript identification in Arabidopsis thaliana. Genome Biol **9:** R112

Lee YH, Ota T, Vacquier VD (1995) Positive selection is a general phenomenon in the evolution of abalone sperm lysin. Mol Biol Evol **12:** 231-238

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

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

Li C, Wong WH (2001) Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad Sci U S A **98:** 31-36

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25:** 1451-1452

Liu J, Blaylock LA, Endre G, Cho J, Town CD, VandenBosch KA, Harrison MJ (2003) Transcript profiling coupled with spatial expression analyses reveals genes involved in distinct developmental stages of an arbuscular mycorrhizal symbiosis. Plant Cell **15:** 2106-2123

Liu WM, Mei R, Di X, Ryder TB, Hubbell E, Dee S, Webster TA, Harrington CA, Ho MH, Baid J, Smeekens SP (2002) Analysis of high density expression microarrays with signed-rank call algorithms. Bioinformatics **18:** 1593-1599

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

Long SR (1989) Rhizobium genetics. Annu Rev Genet **23:** 483-506

Lorenzo O, Solano R (2005) Molecular players regulating the jasmonate signalling network. Curr Opin Plant Biol **8:** 532-540

Lullien V, Barker DG, de Lajudie P, Huguet T (1987)
Plant gene expression in effective and ineffective root nodules of alfalfa (*Medicago sativa*). Plant Mol Biol **9:** 469-478

Mahony S, Benos PV (2007) STAMP: a web tool for exploring DNA-binding motif similarities. Nucleic Acids Res **35:** W253-8

Manners JM, Penninckx IA, Vermaere K, Kazan K, Brown RL, Morgan A, Maclean DJ, Curtis MD, Cammue BP, Broekaert WF (1998) The promoter of the plant defensin gene PDF1.2 from Arabidopsis is systemically activated by fungal pathogens and responds to methyl jasmonate but not to salicylic acid. Plant Mol Biol **38:** 1071-1080

Mathesius U (2008)
Auxin: at the root of nodule development? Functional Plant Biology **35:** 651-668

Maunoury N, Redondo-Nieto M, Bourcy M, Van de Velde W, Alunni B, Laporte P, Durand P, Agier N, Marisa L, Vaubert D, Delacroix H, Duc G, Ratet P, Aggerbeck L, Kondorosi E, Mergaert P (2010) Differentiation of symbiotic cells and endosymbionts in Medicago truncatula nodulation are coupled to two transcriptome-switches. PLoS One **5:** e9519

Maxwell AI, Morrison GM, Dorin JR (2003) Rapid sequence divergence in mammalian beta-defensins by adaptive evolution. Mol Immunol **40:** 413-421

Meade HM, Long SR, Ruvkun GB, Brown SE, Ausubel FM (1982) Physical and genetic characterization of symbiotic and auxotrophic mutants of Rhizobium meliloti induced by transposon Tn5 mutagenesis. J Bacteriol **149:** 114-122

Mergaert P, Nikovics K, Kelemen Z, Maunoury N, Vaubert D, Kondorosi A, Kondorosi E (2003) A novel family in Medicago truncatula consisting of more than 300 nodule-specific genes coding for small, secreted polypeptides with conserved cysteine motifs. Plant Physiol **132:** 161-173

Mergaert P, Uchiumi T, Alunni B, Evanno G, Cheron A, Catrice O, Mausset AE, Barloy-Hubler F, Galibert F, Kondorosi A, Kondorosi E (2006) Eukaryotic control on bacterial cell cycle and differentiation in the Rhizobium-legume symbiosis. Proc Natl Acad Sci U S A **103:** 5230-5235

Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW (2003) Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. Plant Cell **15:** 809-834

Meyers BC, Lee DK, Vu TH, Tej SS, Edberg SB, Matvienko M, Tindell LD (2004) Arabidopsis MPSS. An online resource for quantitative expression analysis. Plant Physiol **135:** 801-813

Mignone F, Gissi C, Liuni S, Pesole G (2002) Untranslated regions of mRNAs. Genome Biol **3:** REVIEWS0004

Mitra RM, Long SR (2004) Plant and bacterial symbiotic mutants define three transcriptionally distinct stages in the development of the Medicago truncatula/Sinorhizobium meliloti symbiosis. Plant Physiol **134:** 595-604

Miyata T, Yasunaga T (1980) Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. J Mol Evol **16:** 23-36

Mondragon-Palomino M, Gaut BS (2005) Gene conversion and the evolution of three leucine-rich repeat gene families in Arabidopsis thaliana. Mol Biol Evol **22:** 2444-2456

Mygind PH, Fischer RL, Schnorr KM, Hansen MT, Sonksen CP, Ludvigsen S, Raventos D, Buskov S, Christensen B, De Maria L, Taboureau O, Yaver D, Elvig-Jorgensen SG, Sorensen MV, Christensen BE, Kjaerulff S, Frimodt-Moller N, Lehrer RI, Zasloff M, Kristensen HH (2005) Plectasin is a peptide antibiotic with therapeutic potential from a saprophytic fungus. Nature **437:** 975-980

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

Moussart A, Tivoli B, Samac D, D'Souza N (2006) *Medicago truncatula* resistance to Oomycetes. In: The *Medicago truncatula* handbook. **http://www.noble.org/MedicagoHandbook/**

Nafisi M, Goregaoker S, Botanga CJ, Glawischnig E, Olsen CE, Halkier BA, Glazebrook J (2007) Arabidopsis cytochrome P450 monooxygenase 71A13 catalyzes the conversion of indole-3-acetaldoxime in camalexin synthesis. Plant Cell **19:** 2039-2052

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol **3:** 418-426

Okuda S, Tsutsui H, Shiina K, Sprunck S, Takeuchi H, Yui R, Kasahara RD, Hamamura Y, Mizukami A, Susaki D, Kawano N, Sakakibara T, Namiki S, Itoh K, Otsuka K, Matsuzaki M, Nozaki H, Kuroiwa T, Nakano A, Kanaoka MM, Dresselhaus T, Sasaki N, Higashiyama T (2009) Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells. Nature **458:** 357-361

Oldroyd GE, Downie JA (2008) Coordinating nodule morphogenesis with rhizobial infection in legumes. Annu Rev Plant Biol **59:** 519-546

Oono R, Denison RF, Kiers ET (2009) Controlling the reproductive fate of rhizobia: how universal are legume sanctions? New Phytol **183:** 967-979

Oono R, Schmitt I, Sprent JI, Denison RF (2010) Multiple evolutionary origins of legume traits leading to extreme rhizobial differentiation. New Phytol

Penninckx IA, Thomma BP, Buchala A, Metraux JP, Broekaert WF (1998) Concomitant activation of jasmonate and ethylene response pathways is required for induction of a plant defensin gene in Arabidopsis. Plant Cell **10:** 2103-2113

Pumplin N, Mondo SJ, Topp S, Starker CG, Gantt JS, Harrison MJ (2010) Medicago truncatula Vapyrin is a novel protein required for arbuscular mycorrhizal symbiosis. Plant J **61:** 482-494

Quandt HJ, Pühler A, Broer I (1993) Transgenic root nodules of *Vicia hirsuta*: a fast and efficient system for the study of gene expression in indeterminate-type nodules. Mol Plant Microbe Interact **6:** 699-706

Rastogi S, Liberles DA (2005) Subfunctionalization of duplicated genes as a transition state to neofunctionalization. BMC Evol Biol **5:** 28

Rockman MV, Wray GA (2002) Abundant raw material for cis-regulatory evolution in humans. Mol Biol Evol **19:** 1991-2004

Rome S, Fernandez MP, Brunel B, Normand P, Cleyet-Marel JC (1996) Sinorhizobium medicae sp. nov., isolated from annual Medicago spp. Int J Syst Bacteriol **46:** 972-980

Romero-Herrera AE, Lehmann H, Joysey KA, Friday AE (1973) Molecular evolution of myoglobin and the fossil record: a phylogenetic synthesis. Nature **246**:389–395

Ronald J, Akey JM, Whittle J, Smith EN, Yvert G, Kruglyak L (2005) Simultaneous genotyping, gene-expression measurement, and detection of allele-specific expression with oligonucleotide arrays. Genome Res **15:** 284-291

Ronfort J, Bataillon T, Santoni S, Delalande M, David JL, Prosperi JM (2006) Microsatellite diversity and broad scale geographic structure in a model legume: building a set of nested core collection for studying naturally occurring variation in Medicago truncatula. BMC Plant Biol **6:** 28

Rostoks N, Borevitz JO, Hedley PE, Russell J, Mudie S, Morris J, Cardle L, Marshall DF, Waugh R (2005) Single-feature polymorphism discovery in the barley transcriptome. Genome Biol **6:** R54

Samac DA, Tesfaye M, Dornbusch M, Saruul P, Temple SJ (2004) A comparison of constitutive promoters for expression of transgenes in alfalfa (Medicago sativa). Transgenic Res **13:** 349-361

Sandal NN, Bojsen K, Marcker KA (1987) A small family of nodule specific genes from soybean. Nucleic Acids Res **15:** 1507-1519

Sato M, Mitra RM, Coller J, Wang D, Spivey NW, Dewdney J, Denoux C, Glazebrook J, Katagiri F (2007) A high-performance, small-scale microarray for expression profiling of many samples in Arabidopsis-pathogen studies. Plant J **49:** 565-577

Sbabou L, Bucciarelli B, Miller S, Liu J, Berhada F, Filali-Maltouf A, Allan D, Vance C (2010) Molecular analysis of SCARECROW genes expressed in white lupin cluster roots. J Exp Bot **61:** 1351-1363

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

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU (2005) A gene expression map of Arabidopsis thaliana development. Nat Genet **37:** 501-506

Schopfer CR, Nasrallah ME, Nasrallah JB (1999) The male determinant of self-incompatibility in Brassica. Science **286:** 1697-1700

Semple CA, Rolfe M, Dorin JR (2003) Duplication and selection in the evolution of primate beta-defensin genes. Genome Biol **4:** R31

Semple CA, Taylor K, Eastwood H, Barran PE, Dorin JR (2006) Beta-defensin evolution: selection complexity and clues for residues of functional importance. Biochem Soc Trans **34:** 257-262

Shiba H, Iwano M, Entani T, Ishimoto K, Shimosato H, Che FS, Satta Y, Ito A, Takada Y, Watanabe M, Isogai A, Takayama S (2002) The dominance of alleles controlling self-incompatibility in Brassica pollen is regulated at the RNA level. Plant Cell **14:** 491-504

Silverstein KA, Graham MA, Paape TD, VandenBosch KA (2005) Genome organization of more than 300 defensin-like genes in Arabidopsis. Plant Physiol **138:** 600-610

Silverstein KA, Moskal WA,Jr, Wu HC, Underwood BA, Graham MA, Town CD, VandenBosch KA (2007) Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. Plant J **51:** 262-280

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

Smyth GK, Speed T (2003) Normalization of cDNA microarray data. Methods **31:** 265-273

Spelbrink RG, Dilmac N, Allen A, Smith TJ, Shah DM, Hockerman GH (2004) Differential antifungal and calcium channel-blocking activity among structurally related plant defensins. Plant Physiol **135:** 2055-2067

Sprent JI. (2008) Evolution and diversity of legume symbiosis. Pages 1-21 In: Dilworth MJ, James E, Sprent J, and Newton WE, eds. *Nitrogen-fixing leguminous symbioses.* New York:Springer-Verlag, 1-21

Sprent JI. (2001) *Nodulation in legumes.* Royal Botanic Gardens, Kew: The Cromwell Press Ltd.

Sprent JI, Thomas RJ. (1984) Nitrogen nutrition of seedling grain legumes: some taxonomic, morphological and physiological constraints. Plant, Cell and Environment **7**: 637-645

Starker CG, Parra-Colmenares AL, Smith L, Mitra RM, Long SR (2006) Nitrogen fixation mutants of Medicago truncatula fail to support plant and bacterial symbiotic gene expression. Plant Physiol **140:** 671-680

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

Stotz HU, Spence B, Wang Y (2009) A defensin from tomato with dual function in defense and development. Plant Mol Biol **71:** 131-143

Stotz HU, Thomson JG, Wang Y (2009) Plant defensins: defense, development and application. Plant Signal Behav **4:** 1010-1012

Stougaard J, Jorgensen JE, Christensen T, Kuhle A, Marcker KA (1990) Interdependence and nodule specificity of cis-acting regulatory elements in the soybean leghemoglobin lbc3 and N23 gene promoters. Mol Gen Genet **220:** 353-360

Szczyglowski K, Szabados L, Fujimoto SY, Silver D, de Bruijn FJ (1994) Site-specific mutagenesis of the nodule-infected cell expression (NICE) element

and the AT-rich element ATRE-BS2* of the Sesbania rostrata leghemoglobin glb3 promoter. Plant Cell **6:** 317-332

Takayama S, Shimosato H, Shiba H, Funato M, Che FS, Watanabe M, Iwano M, Isogai A (2001) Direct ligand-receptor complex interaction controls Brassica self-incompatibility. Nature **413:** 534-538

Terpolilli JJ, O'Hara GW, Tiwari RP, Dilworth MJ, Howieson JG (2008) The model legume Medicago truncatula A17 is poorly matched for N2 fixation with the sequenced microsymbiont Sinorhizobium meliloti 1021. New Phytol **179:** 62-66

Terras FR, Eggermont K, Kovaleva V, Raikhel NV, Osborn RW, Kester A, Rees SB, Torrekens S, Van Leuven F, Vanderleyden J (1995) Small cysteine-rich antifungal proteins from radish: their role in host defense. Plant Cell **7:** 573-588

Tesfaye M, Samac DA, Vance CP (2006) Insights into symbiotic nitrogen fixation in Medicago truncatula. Mol Plant Microbe Interact **19:** 330-341

Tesfaye M, Yang SS, Lamb JF, Jung HJ, Samac DA, Gronwald J, Vance CP, VandenBosch KA (2009) *Medicago truncatula* as a model for dicot cell wall development. BioEnergy Research **2**: 59-76

Thomma BPHJ, Broekaer WF (1998) Tissue-specific expression of plant defensin genes *PDF2.1* and *PDF2.2* in *Arabidopsis thaliana*. Plant Physiol Biochem **36**: 533-537

Thomma BP, Cammue BP, Thevissen K (2002) Plant defensins. Planta **216:** 193-202

Torregros CA, Cazaux M, Danesh D, Chardon F, Cannon SB, Esquerre-Tugaye M-T, Dumas B, Young ND, Samac DA, Huguet T, Jacquet C (2008) Genetic dissection of resistance to Antrhacnose and powdery mildew in *Medicago truncatula.* Mol Plant-Microbe Interact **21**: 61-69

Torregrosa C, Cluzet S, Fournier J, Huguet T, Gamas P, Prosperi JM, Esquerre-Tugaye MT, Dumas B, Jacquet C (2004) Cytological, genetic, and molecular analysis to characterize compatible and incompatible interactions between Medicago truncatula and Colletotrichum trifolii. Mol Plant Microbe Interact **17:** 909-920

Van de Velde W, Zehirov G, Szatmari A, Debreczeny M, Ishihara H, Kevei Z, Farkas A, Mikulass K, Nagy A, Tiricz H, Satiat-Jeunemaitre B, Alunni B, Bourge M, Kucho K, Abe M, Kereszt A, Maroti G, Uchiumi T, Kondorosi E, Mergaert P (2010) Plant peptides govern terminal differentiation of bacteria in symbiosis. Science **327:** 1122-1126

van Wees SC, Chang HS, Zhu T, Glazebrook J (2003) Characterization of the early response of Arabidopsis to Alternaria brassicicola infection using expression profiling. Plant Physiol **132:** 606-617

Vance CP, Heichel GH, Barnes DK, Bryan JW, Johnson LE (1979) Nitrogen Fixation, Nodule Development, and Vegetative Regrowth of Alfalfa (Medicago sativa L.) following Harvest. Plant Physiol **64:** 1-8

Vandemark JG, Barker MB (2003)
Quantifying Phytophthora medicaginis in Susceptible and Resistant Alfalfa with a Real-Time Fluorescent PCR Assay. J Phytopathology **151:** 577-583

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol **3:** RESEARCH0034

Vanoosthuyse V, Miege C, Dumas C, Cock JM (2001) Two large Arabidopsis thaliana gene families are homologous to the Brassica gene superfamily that encodes pollen coat proteins and the male component of the self-incompatibility response. Plant Mol Biol **46:** 17-34

Vasse J, de Billy F, Camut S, Truchet G (1990)
Correlation between Ultrastructural Differentiation of Bacteroids and Nitrogen Fixation in Alfalfa Nodules. J Bacteriol **172:** 4295-4306

von Malek B, van der Graaff E, Schneitz K, Keller B (2002) The Arabidopsis male-sterile mutant dde2-2 is defective in the ALLENE OXIDE SYNTHASE gene encoding one of the key enzymes of the jasmonic acid biosynthesis pathway. Planta **216:** 187-192

Wang D, Griffitts J, Starker C, Fedorova E, Limpens E, Ivanov S, Bisseling T, Long S (2010) A nodule-specific protein secretory pathway required for nitrogen-fixing symbiosis. Science **327:** 1126-1129

Wang HL, Grusak MA (2005) Structure and development of Medicago truncatula pod wall and seed coat. Ann Bot **95:** 737-747

Wang L, Mitra RM, Hasselmann KD, Sato M, Lenarz-Wyatt L, Cohen JD, Katagiri F, Glazebrook J (2008) The genetic network controlling the Arabidopsis transcriptional response to Pseudomonas syringae pv. maculicola: roles of major regulators and the phytotoxin coronatine. Mol Plant Microbe Interact **21:** 1408-1420

Wang L, Tsuda K, Sato M, Cohen JD, Katagiri F, Glazebrook J (2009) Arabidopsis CaM binding protein CBP60g contributes to MAMP-induced SA accumulation and is involved in disease resistance against Pseudomonas syringae. PLoS Pathog **5:** e1000301

Wilkie GS, Dickson KS, Gray NK (2003) Regulation of mRNA translation by 5'- and 3'-UTR-binding factors. Trends Biochem Sci **28:** 182-188

Wilson DL, Buckley MJ, Helliwell CA, Wilson IW (2003) New normalization methods for cDNA microarray data. Bioinformatics **19:** 1325-1332

Xu WW, Cho S, Yang SS, Bolon YT, Bilgic H, Jia H, Xiong Y, Muehlbauer GJ (2009) Single-feature polymorphism discovery by computing probe affinity shape powers. BMC Genet **10:** 48

Yamaguchi-Shinozaki K, Shinozaki K (2005) Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. Trends Plant Sci **10:** 88-94

Yanagisawa S, Schmidt RJ (1999) Diversity and similarity among recognition sequences of Dof transcription factors. Plant J **17:** 209-214

Young ND, Udvardi M (2009) Translating Medicago truncatula genomics to crop legumes. Curr Opin Plant Biol **12:** 193-201

Zhou CX, Zhang YL, Xiao L, Zheng M, Leung KM, Chan MY, Lo PS, Tsang LL, Wong HY, Ho LS, Chung YW, Chan HC (2004) An epididymis-specific beta-defensin is important for the initiation of sperm maturation. Nat Cell Biol **6:** 458-464

Zhu YL, Song QJ, Hyten DL, Van Tassell CP, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB (2003) Single-nucleotide polymorphisms in soybean. Genetics **163:** 1123-1134

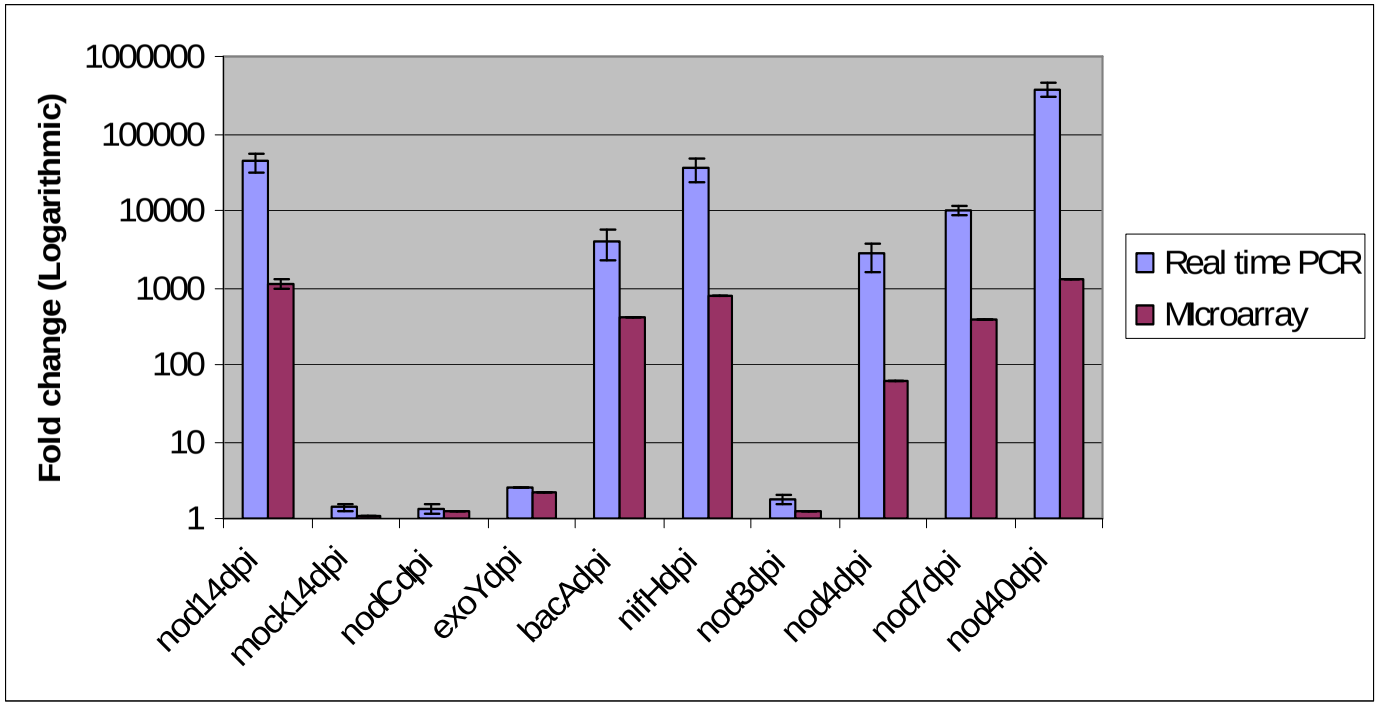


Figure A-1. Real time PCR verification of microarray data. The blue bars represent the fold change values (logarithmic scale) of MtTC100321 probe in different treatments obtained from Real time PCR and the maroon bars represent corresponding fold change values of the same probe from microarray analysis. All the fold changes were calculated as treatment vs. mock 14 dpi roots except for mock 14 dpi roots, where the fold change was against mock 0 dpi roots. Error bars indicate SE of the three biological replicates.

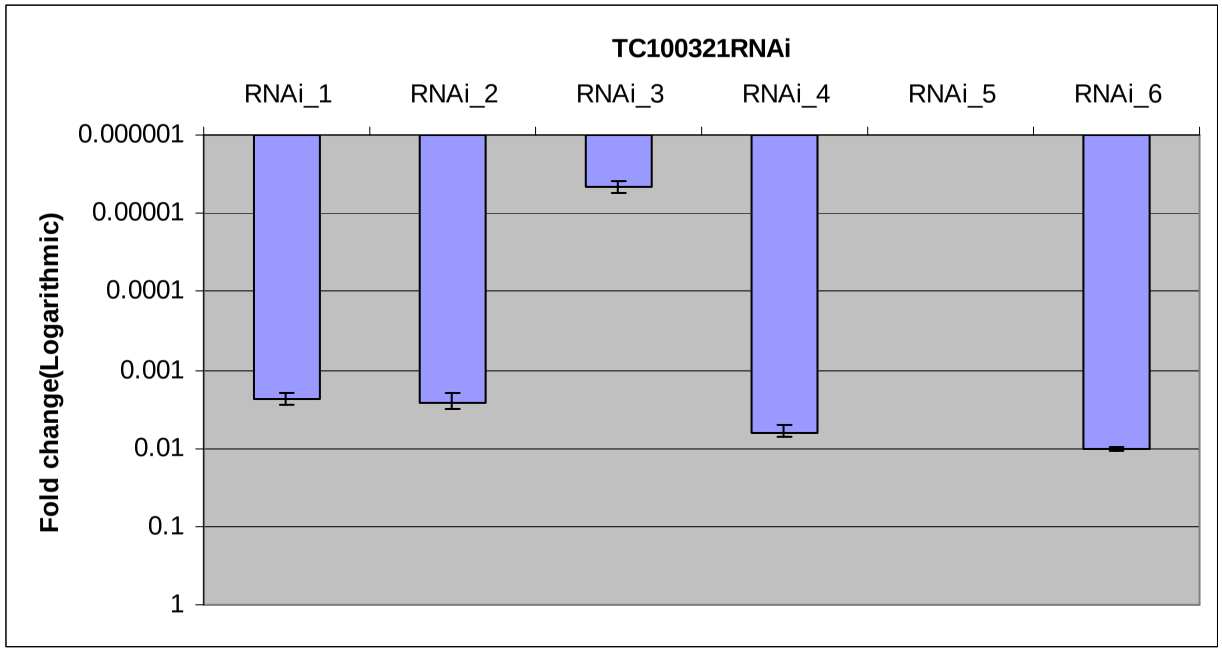


Figure A-2. Real time PCR verification of expression levels in transgenic RNAi plants. Six transgenic RNAi plants of gene corresponding to TC100321 (MtTC100321) were assayed for expression levels using real time PCR. All the fold changes (logarithmic scale) were calculated as transgenic MtTC100321 RNAi roots vs. transgenic Myosin RNAi roots (human myosin gene was used as a control) at 14 dpi. Error bars indicate SE of the three technical replicates.

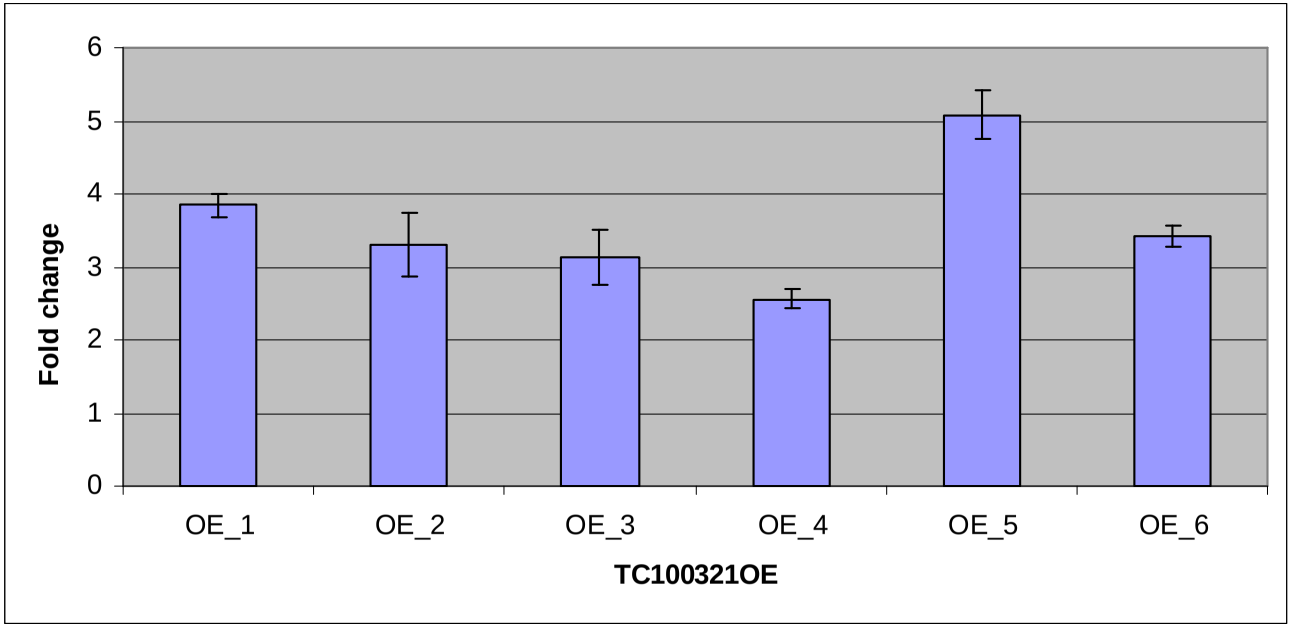


Figure A-3. Real time PCR verification of expression levels in transgenic over expressed (OE) plants. Six transgenic plants where the gene corresponding to TC100321 (MtTC100321) was OE were assayed for expression levels using real time PCR. All the fold changes (logarithmic scale) were calculated as transgenic MtTC100321 OE roots vs. transgenic empty vector (pILTAB381 vector with CsVMV promoter driving the GUS gene) roots at 14 dpi. Error bars indicate SE of the three technical replicates.

Table A-1. Quantification of *Phytophthora medicaginis* in transgenic over expressed (OE) plants at 7dpi.

	***P. medicaginis* (gm)**[a]	**SE** [b]
MtTC100321 OE	2.580555992	0.0885785
	1.737008185	0.1575284
	7.448798858	0.2830938
	5.42211356	0.2830938
	1.105628356	0.0507662
Control (pILTAB381)	0.493159687	0.0746919
	7.30037196	0.4191367
	4.48344944	0.4883892
	4.980363851	0.3836892
	1.929526113	0.1454227

[a]*P. medicaginis* (gm) is the amount of pathogen calculated in five transgenic nodulated plants where the gene corresponding to TC100321 (MtTC100321) was OE and five transgenic nodulated control (pILTAB381 vector with CsVMV promoter driving the GUS gene) plants. All the transgenic plants were inoculated with *P. medicaginis* and the amount of the pathogen was quantified after 7 dpi using TaqMan QPCR. No significant difference between the amount of pathogen in MtTC100321 OE plants and the controls was observed using Student's t-test ($p <= 0.01$).
[b]SE is the Standard Error from the three technical replicates.

Table A-2. Quantification of *Phytophthora medicaginis* in transgenic over expressed (OE) plants at 12dpi.

	***P. medicaginis* (gm)**[a]	**SE** [b]
MtTC100321 OE	4.027098293	0.3117292
	3.545119872	0.3224981
	4.636401538	0.3167573
	4.316194531	0.275071
	6.557298315	0.6230559
Control (pILTAB381)	6.949882577	0.1750552
	4.192516181	0.5730981
	2.615416098	0.6054775
	5.956088606	0.3455074
	4.48344944	0.2257595

[a]*P. medicaginis* (gm) is the amount of pathogen calculated in five transgenic nodulated plants where the gene corresponding to TC100321 (MtTC100321) was OE and five transgenic nodulated control (pILTAB381 vector with CsVMV promoter driving the GUS gene) plants. All the transgenic plants were inoculated with *P. medicaginis* and the amount of the pathogen was quantified after 12 dpi using TaqMan QPCR. No significant difference between the amount of pathogen in MtTC100321 OE plants and the controls was observed using Student's t-test ($p <= 0.01$). [b]SE is the Standard Error from the three technical replicates.

Table A-3. Quantification of *Phytophthora medicaginis* in non-nodulated transgenic over expressed (OE) plants at 7dpi.

	***P. medicaginis* (gm)**[a]	**SE** [b]
	2.049625127	0.2755686
	4.913982058	0.2017476
MtTC100321 *OE*	3.024628957	0.4394432
	2.167483055	0.2576642
	3.3448635	0.2086313
Control (pILTAB381)	2.512223507	0.4299142

[a]*P. medicaginis* (gm) is the amount of pathogen calculated in three transgenic nodulated plants where the gene corresponding to TC100321 (MtTC100321) was OE and three transgenic non- nodulated control (pILTAB381 vector with CsVMV promoter driving the GUS gene) plants. All the transgenic plants were inoculated with *P. medicaginis* and the amount of the pathogen was quantified after 7 dpi using TaqMan QPCR. No significant difference between the amount of pathogen in MtTC100321 OE plants and the controls was observed using Student's t-test ($p <= 0.01$). [b]SE is the Standard Error from the three technical replicates.