2015

Genetic and Functional Analysis of Host Genes Involved in Pathogenic and Symbiotic Legume-Microbe Interactions

Fang Tang
University of Kentucky, fta223@uky.edu

Recommended Citation
https://uknowledge.uky.edu/pss_etds/60

This Doctoral Dissertation is brought to you for free and open access by the Plant and Soil Sciences at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Plant and Soil Sciences by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.
STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student’s advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student’s thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Fang Tang, Student

Dr. Hongyan Zhu, Major Professor

Dr. Arthur G. Hunt, Director of Graduate Studies
Genetic and functional analysis of host genes involved in pathogenic and symbiotic legume-microbe interactions

DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Agriculture, Food and Environment at the University of Kentucky

By
Fang Tang
Department of Plant and Soil Sciences
University of Kentucky

Director: Dr. Hongyan Zhu, Associate Professor of Plant and Soil Sciences

Lexington, Kentucky
2015

Copyright © Fang Tang 2015
ABSTRACT OF DISSERTATION

Genetic and functional analysis of host genes involved in pathogenic and symbiotic legume-microbe interactions

Legumes form symbiotic and pathogenic interactions with microbes. Understanding the molecular mechanisms underlying the legume-microbe interactions would help us to improve crop production in a sustainable manner. This thesis covers two independent research projects. The first project was to study the role of alternative splicing in RCT1-mediated disease resistance. RCT1 is a TIR-NBS-LRR-type plant resistance (R) gene in Medicago truncatula that confers broad-spectrum resistance to Colletotrichum trifolii, a fungal pathogen that causes anthracnose disease in Medicago. RCT1 undergoes alternative splicing at both coding and 3'-untranslated regions, thereby producing multiple transcript variants in its expression profile. Alternative splicing of RCT1 in the coding region results from the retention of intron 4. The transcript with retention of intron 4 is predicted to encode a truncated protein lacking the C-terminal domain of the full-length protein. We showed that the RCT1 function requires the combined presence of the regular and alternative transcripts. This study, in addition to the reports on the tobacco N and Arabidopsis RPS4 genes, adds another significant example showing the involvement of alternative splicing in R gene-mediated plant immunity. The second project was to study the symbiotic specificity in the soybean-rhizobial interaction. It is well known that legume plants can make their own nitrogen fertilizer by forming a root nodule symbiosis with nitrogen-fixing soil bacteria, called rhizobia. One remarkable property of this symbiosis is its high level of specificity, which occurs at both inter- and intra-species levels and takes place at multiple phases of the interaction, ranging from initial bacterial infection and nodulation to late nodule development associated with nitrogen fixation. In this study, we performed fine mapping of the Rj4 gene that controls nodulation specificity in soybean. The Rj4 allele prevents the host plant from nodulation with many strains of Bradyrhizobium elkanii, which are frequently present in soils of the southeastern USA. Since B. elkanii strains are poor symbiotic partners of soybean, cultivars containing an Rj4 allele are considered favorable. We have delimited the Rj4 locus within a 47-kb genomic region on soybean chromosome 1 and identified the candidate genes. We are in the process to validate the candidate genes.

Keywords: RCT1, alternative splicing, Medicago truncatula, Rj4, nodulation specificity, soybean
Fang Tang

Student’s Signature

April, 2015

Date
Genetic and functional analysis of host genes involved in pathogenic and symbiotic legume-microbe interactions

By
Fang Tang

Hongyan Zhu
Director of Dissertation

Arthur G Hunt
Director of Graduate Studies

2015
Acknowledgement

I would like to thank my advisor Dr. Hongyan Zhu for giving me the chance to join his group and providing me the opportunity to work on these interesting projects. Thanks for your encouragement and support in these years, and for your instructive suggestions and constructive criticisms during the writing of this thesis. I would like to thank Dr. Arthur Hunt, Dr. Sharyn Perry and Dr. Pradeep Kachroo for serving as member of my committee and being good teachers. Thanks for your valuable suggestions in improving my research work and comments in polishing my dissertation. I thank Dr. David F. Hildebrand for helping me to revise the thesis and sharing the knowledge beyond the research.

I would like to give many thanks to my friends and colleagues from the Zhu lab Dr. Shengming Yang, for his infinite patience, professional guidance and generous assistance in my research work and tremendous help in my thesis writing. Without your support and encouragement, I cannot complete the projects successfully. Dr. Jingge Liu and Dr. Qiaolin Zheng, for being good colleagues, constructive discussions and technical advices, and all the help in both work and everyday life. Dr. Muqiang Gao, for teaching me the techniques in plant transformation and providing me the soybean seeds.

Grateful acknowledgement gives to my parents, for their endless love and unconditional support over all these past years. Many many thanks to my husband Yuanheng Li. Your love and trust have been helping me to go through difficult times and giving me sufficient confidence to do anything. I would also like to give my special thanks to all my friends I met in Lexington, Ran An, Xile Li, Dandan Li, Qiulin Qin,
Qingming Gao, Pengpeng Lin, Yan Zhu, Xia Wu, Yuting Li, Jingchen Xu et al. Thanks for giving me generous support and help in life and all the encouragement during the past years. I am so lucky to have you as my family and friends and I enjoyed every moment that spent with you.

Thanks to all the people who helped me. Thanks to University of Kentucky, Department of Plant and Soil Sciences for providing the scholarship to support me and giving me an opportunity to become a good scientist.
# TABLE OF CONTENTS

Acknowledgement ........................................................................................................................................ iii

TABLE OF CONTENTS .................................................................................................................................... v

LIST OF TABLES ................................................................................................................................................. viii

LIST OF FIGURES ............................................................................................................................................... ix

Part 1 ................................................................................................................................................................. 1

Chapter 1 Alternative splicing in plant immunity ......................................................................................... 1

Introduction ......................................................................................................................................................... 1

Plant disease resistance ................................................................................................................................... 2

Alternative splicing of TIR-NBS-LRR R genes ................................................................................................. 7

Alternative splicing of CC-NBS-LRR R genes ................................................................................................. 11

Possible mechanisms of alternative splicing-mediated regulation of defense response ................................ 13

Regulation of alternative splicing of R genes ................................................................................................. 18

Summary ............................................................................................................................................................ 21

Chapter 2 Functional characterization of the role of alternative splicing in RCT1-mediated resistance in *Medicago truncatula* ............................................................................................................ 22

Introduction ......................................................................................................................................................... 22

Materials and methods ................................................................................................................................... 25

Plasmid construction ...................................................................................................................................... 25

Plant transformation ....................................................................................................................................... 30

Pathogen inoculation and phenotypic analysis ............................................................................................... 32

RNA isolation and analysis of gene expression by reverse-transcriptase (RT)-PCR ..................................... 33

Results ............................................................................................................................................................... 33

Alternative splicing of the *M. truncatula RCT1* gene .................................................................................... 33

The full-length or truncated open reading frame alone is not sufficient for resistance ................................ 35

Removal of introns 1-3 affects transgene expression but has no effect on alternative splicing of intron 4 ......................................................................................................................................................... 40

Overexpression of the intron 4-containing transgene is sufficient for resistance ..... 42
\textit{RCT1}-mediated resistance requires \textit{EDS1} ............................................. 42
Discussion....................................................................................................................... 45
Part 2................................................................................................................................ 48
Chapter 3 Legume-rhizobial symbiosis........................................................................ 48
Introduction ..................................................................................................................... 48
An overview of the legume-rhizobia symbiosis ............................................................. 49
Nodulation signaling in legumes ...................................................................................... 51
Host genes required for nitrogen fixation ...................................................................... 55
Host specificity in nodulation and nitrogen fixation ...................................................... 57
Summary........................................................................................................................... 67
Chapter 4 Map-based cloning of \textit{Rj4}, a gene controlling nodulation specificity in soybean .................................................................................................................. 69
Introduction ..................................................................................................................... 69
Materials and methods .................................................................................................. 74
Plant materials and nodulation assay .............................................................................. 74
DNA isolation, marker development, and PCR amplification ....................................... 74
Sequence analysis .......................................................................................................... 77
Association mapping ....................................................................................................... 77
Rapid amplification of cDNA ends (RACE) .................................................................. 79
Plasmid construction ....................................................................................................... 80
Soybean hairy-root transformation .................................................................................. 81
RNA isolation and analysis of gene expression by reverse-transcriptase (RT)-PCR .... 82
T3SS mutant inoculation ................................................................................................. 83
Results ............................................................................................................................. 83
Localization of the \textit{Rj4} locus ......................................................................................... 83
Fine mapping of the \textit{Rj4} locus ....................................................................................... 86
Identification of candidate genes for \textit{Rj4}...................................................................... 87
The candidate gene Glyma01g37060 is duplicated in the \textit{Rj4} genotypes .................. 90
The expression of Glyma01g37060 and Glyma01g37060-2 is root-specific and induced upon rhizobial inoculation ................................................................. 91
Complementation tests failed to validate the candidate genes ...................................... 93
Rj4 function is dependent on the bacterial type III secretion system (T3SS) ....... 95
Discussion........................................................................................................... 96
Appendix............................................................................................................. 101
References.......................................................................................................... 104
VITA.................................................................................................................... 133
LIST OF TABLES

Table 2.1 The time of anthracnose symptom onset in transgenic alfalfa plants ................ 39

Table 4.1 Molecular markers used for genetic mapping of the Rj4 locus. ....................... 76

Table 4.2 Soybean genotypes used for association mapping analysis ............................... 78

Table 4.3 Gene specific primers used in RACE for PCR amplification ......................... 79

Table 4.4 Predicted genes in the 47-kb genomic region of Williams 82 (rj4/rj4) ........... 86
LIST OF FIGURES

Figure 2.1 Schematic diagram of the RCT1 gene constructs used in this study. ............ 26
Figure 2.2 Diagrammatic representation of RCT1-Intronless construct.......................... 27
Figure 2.3 Diagrammatic representation of RCT1-Intron4 construct............................. 28
Figure 2.4 Diagrammatic representation of RCT1-Tr construct.................................... 29
Figure 2.5 Semi-quantitative reverse-transcriptase (RT)-PCR analysis of the RCT1 regular and alternative transcripts in the M. truncatula genotype Jemalong A17 within 24 h post inoculation (hpi). ................................................................. 34
Figure 2.6 Semi-quantitative reverse-transcriptase (RT)-PCR analysis of transgene expression in plants transformed with the RCT1-Intronless or RCT1-Tr constructs. ...... 37
Figure 2.7 Disease resistance assay for transgenic alfalfa plants expressing the native RCT1 gene and the intron-deleted constructs ................................................................. 38
Figure 2.8 Semi-quantitative RT-PCR analysis of transgene expression in plants transformed with the RCT1-Intron4 construct................................................................. 41
Figure 2.9 Silencing of EDS1 in RCT1-transformed alfalfa compromised disease resistance......................................................................................................................... 44
Figure 3.1 The symbiosis signaling pathway............................................................... 52
Figure 3.2 Host and bacterial signals that regulate host specificity in the legume-rhizobia symbiosis................................................................................................................ 59
Figure 3.3 The general structure of the Nod factors produced by rhizobia.................. 61
Figure 4.1 Nodulation phenotypes of Williams ($rj4/rj4$) and Hill ($Rj4/Rj4$) by $B. elkanii$ USDA61................................................................................................................................. 84

Figure 4.2 Genetic mapping of the $Rj4$ locus................................................................................................................................. 85

Figure 4.3 Alignment of amino acid sequences of Glyma01g37060 isoforms from Hill ($Rj4/Rj4$) and Williams ($rj4/rj4$). ................................................................................................................................. 88

Figure 4.4 Genotyping using the CAPS marker developed from Glyma01g37060 revealed that the sequence substitution was invariably associated with the nodulation phenotypes. ................................................................................................................................. 89

Figure 4.5 Alignment of amino acid sequences of Glyma01g37060 and Glyma01g37060-2 from Hill ($Rj4/Rj4$). ................................................................................................................................. 91

Figure 4.6 The expression of Glyma01g37060 and Glyma01g37060-2 is root-specific in Hill ($Rj4/Rj4$). ................................................................................................................................. 92

Figure 4.7 Semi-quantitative reverse-transcriptase (RT)-PCR analysis of the expression of (A) Glyma01g37060 and (B) Glyma01g37060-2 in Hill ($Rj4/Rj4$) and Williams ($rj4/rj4$). ................................................................................................................................. 92

Figure 4.8 Composite transgenic plants transformed with the Glyma01g37060 or Glyma01g37060-2 constructs possess both transgenic (blue) and wild-type (white) roots. ................................................................................................................................. 94

Figure 4.9 RT-PCR analysis of transgene expression in Williams roots transformed with the Glyma01g37060 or Glyma01g37060-2 constructs driven by the CaMV-35S promoter. ................................................................................................................................. 94
Part 1

Chapter 1 Alternative splicing in plant immunity

Introduction

Alternative splicing is a regulated process during gene expression that produces multiple distinct transcript isoforms from a single pre-mRNA (Nilsen and Graveley 2010). Genome-wide studies have shown that alternative splicing occurs frequently and widely in eukaryotes. In human, nearly 95% of multi-exon genes undergo alternative splicing (Pan et al., 2008; Wang et al., 2008). In Drosophila melanogaster, transcriptome analysis revealed that 7,473 genes contain at least one alternative splicing event, accounting for 60.7% of the total expressed multi-exon genes (Graveley et al., 2011). As an important mechanism to regulate gene expression, alternative splicing greatly increases transcriptome plasticity and proteome diversity in many eukaryotes (Brett et al., 2002; Kazan 2003; Pan et al., 2008; Ramani et al., 2011). In plants, analysis of Arabidopsis thaliana EST/cDNA libraries initially gave rise to estimates of alternative splicing rates as low as 1.2% (Zhu et al., 2003). Subsequently, improved EST coverage led to estimates of 11.6% (Iida et al., 2004), 21.8% (Wang and Brendel 2006), and 30% (Campbell et al., 2006). More recently, high-throughput sequencing has revealed that about 61% of intron-containing genes in Arabidopsis undergo alternative splicing (Marquez et al., 2012). Considering that these data were obtained from plants growing under normal conditions, the actual value for alternative splicing frequency is likely to be even higher. Environmental and biotic stresses can induce alternative splicing, and novel splicing sites have been identified in studies of alternative splicing under stress conditions (Ali and
Reddy 2008; Mastrangelo et al., 2012; Howard et al., 2013). A recent RNA-seq study of *Pseudomonas syringae*-infected *Arabidopsis* indicated that over 90% of the expressed genes were alternatively spliced (Howard et al., 2013). Moreover, expression of alternative transcript isoforms in tissue-specific and developmental stage-specific manner adds another layer of complexity to alternative splicing mechanisms and transcriptome annotation (Lopato et al., 1996; Lopato et al., 1999; Yoshimura et al., 2002; Loraine et al., 2013).

Proteins encoded by alternative splicing isoforms can have different activities, tissue distributions, or intracellular localizations (Lopato et al., 1996; de la Fuente van Bentem et al., 2003; Carvalho et al., 2012; Kriechbaumer et al., 2012; Remy et al., 2013). Although its biological function is not fully understood in plants, alternative splicing is involved in many physiological processes, including defense responses (Reddy 2007; Carvalho et al., 2013; Reddy et al., 2013; Staiger and Brown 2013). Plants have evolved sophisticated systems to detect pathogen attacks and trigger innate immunity. Recently, alternative splicing has been recognized as a crucial regulatory mechanism in plant defense against pathogen infections (Dinesh-Kumar and Baker 2000; Zhang and Gassmann 2003; Zhang and Gassmann 2007; Staiger et al., 2013).

**Plant disease resistance**

Two types of plant immunity operate to restrict pathogen colonization in the host. In the initial phase, a basal level of plant defense responses are activated by the microbe/pathogen-associated molecular patterns (MAMPs/ PAMPs), such as chitin, flagellin, and Elongation Factor-Tu (EF-Tu) (Nicaise et al., 2009). Perception of structurally conserved
PAMPs by host transmembrane pattern recognition receptors (PRRs) triggers PAMP-triggered immunity (PTI). However, pathogens can interfere with PTI and dampen basal resistance with secreted effector proteins. Accordingly, in the second layer of defense, the plant deploys resistance (R) proteins to perceive the presence or action of the corresponding effector proteins, known as Avirulence (Avr) proteins, leading to initiation of the stronger disease resistance, called effector-triggered immunity (ETI) (Jones and Dangl 2006). R proteins recognize Avr proteins either directly or indirectly. Direct R-Avr interaction is exemplified by the direct binding of the Linum usitatissimum (flax) L protein with its cognate effectors (Dodds et al., 2006). Indirect R-Avr interaction can be explained by the proposed “guard hypothesis” (Van Der Biezen and Jones 1998). In this model, R proteins detect pathogens through perception of the altered status of other host cellular proteins that are effector virulence targets, termed guardees.

The co-evolutionary “arms race” between host and pathogen has been extensively studied in the interaction between Arabidopsis and pathogenic P. syringae expressing EF-Tu. Direct binding of EF-Tu to its pattern receptor EFR induces phosphorylation on the tyrosine residues of EFR, and activates PTI (Boller and Felix 2009). However, the P. syringae-secreted effector HopA1 has phosphatase activity and reduces EFR phosphorylation, thus suppressing EF-Tu-triggered PTI (Macho et al., 2014). The Arabidopsis R protein RPS6 (Resistance to P. syringae 6) specifically recognizes HopA1 (Kim et al., 2009). The HopA1 target guarded by RPS6 is believed to be EDS1 (Enhanced disease susceptibility 1), a central regulator of basal resistance and of ETI mediated by R proteins (Bhattacharjee et al., 2011).
PTI cannot completely inhibit pathogen colonization, but can retard pathogen invasion (Glazebrook et al., 1997). In contrast, ETI can be considered as an amplified version of PTI and is often associated with a rapid, localized programmed cell death, known as the hypersensitive response (HR) (Cui et al., 2014). A chain of defense responses occur concomitant with the HR, including oxidative burst, accumulation of salicylic acid (SA), expression of pathogenesis-related (PR) genes, and defensin biosynthesis. PTI involves mitogen-activated protein kinase-signaling cascades and the accumulation of reactive oxygen species (Asai et al., 2002; Shen et al., 2007), and constitutive activation of PTI in the absence of pathogen results in deleterious effects on plant development. Since ETI induces long-lasting systemic immunity (systemic acquired resistance) (Cui et al., 2014), it must be fine-tuned to protect the plant from pathogen attack without excessive fitness costs.

1. Plant R genes

The majority of cloned R genes encode proteins containing a central nucleotide-binding site (NBS) and a C-terminal leucine-rich repeat (LRR) region. The NBS region normally is comprised of three subdomains, NBS, ARC1, and ARC2. The characteristic NBS subdomain includes a binding site for ATP or GTP and is active in initiation of signaling cascades leading to resistance responses (Traut 1994). The ARC subdomains (named for their presence in Apaf-1, R proteins, and CED-4) are highly conserved and essential for intramolecular interactions of R proteins (Rairdan and Moffett 2006). By contrast, the LRR motif confers recognition specificity to the plant defense response (Kobe and Deisenhofer 1994; Ellis et al., 1999; Jia et al., 2000; Leister and Katagiri 2000).
Based on their N-terminal structures, members of the NBS-LRR family of \( R \) genes can be further subdivided into two subfamilies. One subfamily comprises members with a domain homologous to the intracellular signaling domains of the \textit{Drosophila} Toll and mammalian Interleukin (IL)-1 receptor (TIR-NBS-LRR). TIR-NBS-LRR genes are exclusively present in dicot species. Members of this subfamily include tobacco \( N \), flax \( L6 \) and \( M \), \textit{Arabidopsis} \textit{RPP1}, \textit{RPP4} and \textit{RPS4}, and \textit{Medicago truncatula RCT1} (Whitham et al., 1994; Lawrence et al., 1995; Anderson et al., 1997; Botella et al., 1998; Gassmann et al., 1999; van der Biezen et al., 2002; Zhang and Gassmann 2003; Yang et al., 2008). Another subfamily is characterized by a putative coiled-coil domain in the N-terminal region (CC-NBS-LRR). CC-NBS-LRR genes are widely distributed in both dicots and monocots. Both the CC and TIR domains likely function in interaction with downstream factors in ETI signaling (Tao et al., 2000). Although most TIR- and CC-NBS-LRRs lack putative transmembrane domains or organelle-targeting signals and are predicted to be cytosolic, some show dynamic changes in subcellular localization (Boyes et al., 1998; Shen et al., 2007).

2. Signaling components in ETI

In addition to their structural differences, TIR-NBS-LRR and CC-NBS-LRR genes generally function through distinct signaling pathways, requiring either EDS1 or NDR1 (Non-race-specific disease resistance 1), respectively (Aarts et al., 1998). One exception is the \textit{Arabidopsis} \textit{HRT} gene that confers resistance to TCV (Turnip crinkle virus). \textit{HRT} is a CC-NBS-LRR gene but its signaling is dependent on EDS1 (Chandra-Shekara et al., 2004). Moreover, a few CC-NBS-LRR genes including \textit{RPP7}, \textit{RPP8}, and \textit{RPP13} can activate defense signaling independent of EDS1 and NDR1 (Aarts et al., 1998; McDowell
et al., 1998; Bittner-Eddy and Beynon 2001). Venugopal et al. (2009) proposed, however, that EDS1 and SA act redundantly to regulate ETI to viral, bacterial, and oomycete pathogens. As such, participation of EDS1 in signaling triggered by CC-NBS-LRR R proteins may be masked by SA, and vice versa. In such cases, the requirement for EDS1 would be observed only when disease resistance does not require SA accumulation. PAD4 (Phytoalexin deficient 4) and SAG101 are indispensable for EDS1-required signaling to restrict pathogen growth (Falk et al., 1999; Feys et al., 2005). EDS1, PAD4, SAG101 function independently, as well as in a ternary complex of SAG101-EDS1-PAD4, serving as signal transducers in HRT-mediated resistance to TCV (Zhu et al., 2011). However, the HR associated with TCV resistance conferred by HRT requires only EDS1, whereas the SA signaling induced by HRT requires only PAD4.

Genetic analysis of Arabidopsis mutants defective in systemic acquired resistance led to the isolation of NPR1 (Non-expressor of PR genes 1), which encodes a putative transcription factor regulating PR gene expression downstream of SA production (Cao et al., 1997). Further investigation of the regulator of NPR1 in Arabidopsis resulted in identification of the gain-of-function mutant snc1 (Suppressor of npr1-1, constitutive 1) (Li et al., 2001), which exhibits a dwarfed phenotype caused by constitutive activation of defense signaling in the absence of pathogen infection. Based on these mutants it can be concluded that wild type SNC1 suppresses NPR1 and to finely control autoimmune responses. Interestingly, SNC1 encodes a TIR-NBS-LRR R protein, and the snc1 mutant morphology is restored or suppressed to different extents in a series of mos (Modifier of snc1) mutants. Up to now, 13 MOS genes have been cloned, the gene products of which act in various cellular and molecular processes, including pre-mRNA splicing, nuclear
trafficking of serine-arginine rich (SR) proteins and protein modification, which is indicative of a highly complex network for regulation of R protein-mediated ETI (Palma et al., 2005; Zhang et al., 2005; Zhang and Li 2005; Goritschnig et al., 2007; Palma et al., 2007; Wiermer et al., 2007; Goritschnig et al., 2008; Cheng et al., 2009; Germain et al., 2010; Li et al., 2010; Xu et al., 2011; Xu et al., 2012).

**Alternative splicing of TIR-NBS-LRR R genes**

Most TIR-NBS-LRR genes have conserved gene structures in the coding region, which generally contains three or four introns. The first exon encodes the TIR domain, the second exon encodes the NBS domain, and the remaining exons encode the LRR region. Alternative splicing of TIR-NBS-LRR genes can result from intron retention, selection of alternative exons, or usage of alternative 5' or 3' splicing sites. Alternative isoforms have been reported for many TIR-NBS-LRR genes, such as tobacco N (Whitham et al., 1994), flax L, and M loci (Ayliffe et al., 1999), *Arabidopsis SNC1, RPS4, RPS6, RPP5*, and RAC1 (Parker et al., 1997; Gassmann et al., 1999; Borhan et al., 2004; Yi and Richards 2007; Kim et al., 2009), tomato *Bs4* (Schornack et al., 2004), potato *Y-J* (Vidal et al., 2002), and *M. truncatula RCT1* (Yang et al., 2008). The functional consequences of alternative splicing events have been characterized for only a few TIR-NBS-LRR R genes, including *Arabidopsis RPS4* and tobacco N.

The *Arabidopsis RPS4* gene confers resistance to *Pseudomonas syringae* pv. *tomato* strain DC3000 expressing *AvrRps4*. Alternative splicing produces six transcript isoforms of *RPS4* via retention of intron 2 and/or intron 3, and splicing of a cryptic intron in exon 3 (Zhang and Gassmann 2003). Due to premature stop codons introduced by frame shifts,
the alternatively spliced isoforms encode no or fewer LRR repeats. Experiments involving stable transformation of \textit{RPS4} genomic constructs lacking intron 2 and/or intron 3, under the control of the \textit{RPS4} promoter, showed that deletion of a single intron was sufficient to abolish \textit{RPS4} function, even though splicing of remaining intron was unaffected and the normally spliced transcript was also expressed (Zhang and Gassmann 2003). Therefore, resistance to DC3000 requires alternative splicing of \textit{RPS4}.

The biological function of these alternatively spliced isoforms as regulatory RNAs was unknown, but there is evidence that the truncated proteins which encoded by these alternative transcripts regulate the activity of full-length RPS4. An artificial combination of normal and alternatively spliced isoforms only partially restored \textit{RPS4}-mediated resistance (Zhang and Gassmann 2003). The molar ratio of \textit{RPS4} transcript isoforms in that experiment was altered compared to those naturally occurring, suggesting that the ratio is of functional importance. The abundance of the various alternative splicing isoforms of \textit{RPS4}, particularly the isoform retaining intron 3 (\textit{RPS4}_{AT4}), is under dynamic regulation in response to AvrRps4. Whereas the full-length transcript including all exons is the predominant splicing product in uninoculated leaves, pathogen inoculation induces a rapid, >100-fold increase of \textit{RPS4}_{AT4} (Zhang and Gassmann 2007). The truncated proteins encoded by \textit{RPS4} variants were detected in transient expression assays, confirming that the aberrant transcripts are functional.

Tobacco \textit{N} specifically recognizes a 50-kDa helicase protein (p50) of tobacco mosaic virus (TMV), and the \textit{N} gene is alternatively spliced (Whitham et al., 1994; Les Erickson et al., 1999). In addition to the major isoform (\textit{N}_{RT}), an alternative isoform (\textit{N}_{AT}) is generated via alternative splicing of a hidden exon containing a stop codon within intron
3, which yields a putative product lacking 13 of 14 LRR repeats. Similar to RPS4, a dynamic abundance ratio of \( N_{RT} \) to \( N_{AT} \) is also observed during TMV infection (Dinesh-Kumar and Baker 2000). Although \( N_{RT} \) is predominant before infection, \( N_{AT} \) is the more abundant isoform 6 h after TMV inoculation, and the original isoform ratio reappears 9 h after inoculation. Perturbing the ratio of \( N_{RT} \) to \( N_{AT} \) resulted in compromised TMV resistance. The boost in \( N_{AT} \) production may result from a signaling cascade induced by interaction between \( N_{RT} \) and p50. Because the accumulation of spliced variants occurs rapidly, the induced alternative splicing may regulate N function via feedback inhibition. Tobacco transformants expressing only \( N_{RT} \) displayed incomplete resistance manifested by delayed HR, suggesting \( N_{AT} \) is required for complete N-mediated resistance (van Rhijn et al., 2001). However, \( N_{AT} \) expressed alone was not sufficient for TMV-dependent HR, indicating both \( N_{RT} \) and \( N_{AT} \) are indispensable for full N-mediated resistance.

In contrast to RPS4 and N, alternatively spliced transcripts of flax L6 and tomato BS4 are not required for full resistance to the corresponding pathogens. For example, transgenic plants carrying an intronless L6 (\( L6_{RT} \)) exhibited complete rust resistance, similar to plants carrying the wild-type L6 (Ayliffe et al., 1999). L6 triggers flax rust resistance by direct interaction with its cognate effector \( AvrL567 \) (Dodds et al., 2006). The flax rust resistance gene \( M \), which is homologous to L6, is also alternatively spliced (Schmidt et al., 2007); therefore, it is possible that a truncated protein of M could functionally substitute for the truncated L6 protein (Gassmann 2008). However, transient co-expression of \( L6_{RT} \) and the cognate avirulence gene \( AvrL6 \) in tobacco gives rise to apparent HR, which argues against any interference by the \( M \) locus (Gassmann 2008). Likewise, transient expression of intronless Bs4 revealed that the normal Bs4 protein
alone could mediate AvrBs4 recognition, suggesting the alternative splicing of Bs4 is functionally dispensable (Schornack et al., 2004). Although such transient expression assays have served well for isolation of R genes (Bendahmane et al., 2000), whether this system can be reliably used to analyze functional roles for alternative splicing of R genes remains to be seen. It is possible that the observed HR could be due to partial resistance conferred by an endogenous full-length R protein, such as tobacco N. Recent analysis of truncated R genes containing TIR-NBS only (TN) in Arabidopsis showed that chlorosis was induced by transient overexpression of TN genes (Nandety et al., 2013). The alternative L6 and Bs4 isoforms were not tested in transient assays; therefore, these transient expression experiments may not fully reflect the physiological roles of alternative splicing in the process.

A stunted phenotype caused by constitutive defense responses was observed in transgenic tobacco carrying an L6 genomic construct, as well as in transgenic tobacco plants in which L6RT was under the control of the 35S promoter (Frost et al., 2004). This observation suggested that alternative splicing is irrelevant to L6-mediated resistance, with dwarfism serving as a reporter for activation of defense responses. However, the lack of tobacco transformants expressing L6RT from its native promoter precludes firm conclusions about this. Structural and functional analysis demonstrated that the TIR domain alone is necessary and sufficient for L6 immune signaling (Bernoux et al., 2011). More interestingly, with only one exception (L10-A), tobacco plants transformed with a genomic construct of L10 grew normally (Frost et al., 2004). Further analysis revealed that the stunted phenotype of L10-A is associated with the presence of an additional truncated L10 transcript resulting from an aberrant T-DNA integration (Frost et al., 2004).
This truncated transcript is predicted to encode a protein containing the TIR and 39 amino acids of the NBS domain of L10. These findings point to the possibility that the functional significance of alternative splicing in L6 has been undervalued.

**Alternative splicing of CC-NBS-LRR R genes**

Alternative splicing has been identified in many CC-NBS-LRR R genes, including LR10 and Sr35 in wheat (Sela et al., 2012; Saintenac et al., 2013), Mla in barley (Haltermann et al., 2001; Haltermann et al., 2003), Pi-ta and RGA5 in rice (Costanzo and Jia 2009; Cesari et al., 2013), and JA1tr in common bean (Ferrier-Cana et al., 2005), but the functional importance of this post-transcriptional modification for full disease resistance is largely unknown. Only the alternative transcripts of RGA5 have been functionally characterized in a robust system (Cesari et al., 2013).

Rice blast R protein RGA5 was found to cooperate with RGA4 in recognizing two sequence-unrelated effectors, Avr-pia and Avr1-CO39, through direct binding. Two transcript isoforms are generated by alternative splicing of the third of the three introns in the coding region of RGA5 (Okuyama et al., 2011). As in the case of M. truncatula RCT1, protein products of both the intronless, fully-spliced transcript (RGA5_{RT}) and the alternative splicing version (RGA5_{AT}) share the CC, NBS, and LRR domains, and differ only in the C-terminal region, which is related to the copper binding protein ATX1 (RATX1) (Cesari et al., 2013). Transformants carrying RGA5_{AT} are fully susceptible to Avr-pia- and Avr1-CO39-expressing Magnaporthe oryzae strains. Furthermore, in conjunction with RGA4, RGA5_{RT} is necessary and sufficient to confer dual recognition specificity (Cesari et al., 2013). Yeast two-hybrid assays demonstrated that Avr-pia and
Avr1-CO39 physically interact with the C-terminal RATX1 domain, which is present only in RGA5RT. The disruption of the RATX1 domain consequently renders RGA5AT inactive. These findings highlight the importance of the non-LRR regions near the C-termini of R proteins, indicating that they may deserve more attention when exploring the functions of R proteins in disease resistance.

Another rice blast resistance gene, Pi-ta, confers resistance to strains of M. oryzae containing cognate avirulence gene Avr-Pita. A total of 12 distinct transcript isoforms were identified as resulting from alternative splicing and are predicted to encode 11 proteins. Some of these transcripts are constitutively expressed while others show differential expression upon blast infection (Costanzo and Jia 2009). Their regulatory roles in disease resistance remain unknown.

The barley powdery mildew resistance genes Mla6 and Mla13 have very similar gene structures, including the conservation of two introns in the 5′-UTR and two introns in the coding region, as well as a large intron in the 3′-UTR. Notably, both genes exhibit alternative splicing of the 5′-UTR, which contains three upstream ORFs; Alternative splicing is also predicted to cause variation of one amino acid in the coding region of Mla13 (Halterman et al., 2003). The expression of Mla13 transcripts is induced upon pathogen penetration, and a dynamic change in the relative abundance of transcript isoforms has been observed. Inactivation of upstream ORF translation via mutagenesis suggests the upstream ORFs in the 5′-UTR downregulate Mla13 synthesis (Halterman and Wise 2006). Therefore, alternative splicing of upstream ORFs may finely tune Mla13 expression to achieve effective resistance while minimizing host cell damage. However,
it remains unknown whether full resistance mediated by *Mla13* or *Mla6* requires alternative splicing of the upstream ORFs.

**Possible mechanisms of alternative splicing-mediated regulation of defense response**

In the cases where alternative splicing is necessary for disease resistance, transgenic plants containing only the full-length transcript do not display auto-immunity or lesion mimic phenotypes induced by increased R protein activity, suggesting that alternative splicing is not likely to negatively regulate the *R* gene function. By contrast, the absence of alternative splicing impairs *R* gene-mediated resistance, indicating the positive roles for alternative splicing in defense responses. Therefore, R protein isoforms possibly function by suppressing the negative regulation of immunity activation, or by directly engaging in effector-trigged signaling, or by a combination of both.

1. Disruption of R protein auto-inhibition

Whether an R protein is active or inactive is determined by the binding of ATP or ADP to the NBS domain (Lukasik and Takken 2009). Since constitutive activation of R proteins leads to lethal effects on plant growth, negative regulation of R protein activity is essential (Bendahmane et al., 2002; Shirano et al., 2002; Zhang et al., 2003). Intramolecular interactions between R-protein domains may function as a regulatory switch, and several mechanistic models have been proposed to describe this R protein self-regulation, such as the “Jack-knife” model (Belkadir et al., 2004). These models are based largely on the *trans*-complementation of Rx CC-NBS and LRR domains (Moffett et al., 2002). From the crystal structures of the TIR and CC domains (Bernoux et al., 2011; Maekawa et al., 2011), Takken and Goverse (2012) proposed a model which is the
NBS domain interacts with the $N$-terminal half of the LRRs, maintaining the R protein as inactive in a closed conformation before pathogen invasion. An electrostatic interface that maintains the inactive conformation may be formed by interaction between the LRR and NBS domains. The C-terminal LRRs are exposed to serve as an antenna to detect charge changes induced by environmental perturbations. Since the TIR or CC domain can also interact with the NBS domain (Moffett et al., 2002), the R protein is stabilized in a compact structure in the absence of pathogens. Studies on intramolecular interactions of Rx have provided evidence that the NBS domain alone is not sufficient for stable binding, but instead requires the CC domain. Notably, the CC domain could also interact with the NBS domain, unless $N$-terminal LRRs were bound to the NBS domain (Moffett et al., 2002; Rairdan and Moffett 2006; Rairdan et al., 2008). As such, the interaction of LRRs and NBS domains seems to cause conformational changes in the latter that facilitate NBS binding with CC domain. It has also been demonstrated that the ARC1 subdomain is necessary for binding of the Rx $N$-terminal LRR domain, while the ARC2 subdomain is required to maintain an auto-inhibited state in the absence of elicitor, as well as for subsequent signaling (Rairdan and Moffett 2006). Mutation in LRRs or conserved ARC2 motifs of the NBS domain leads to the auto-activation of Rx and RPS5 (Moffett et al., 2002; Rairdan and Moffett 2006; Qi et al., 2012). The majority of truncated R protein variants generated by alternative splicing are presumably unstable, due to the lack of LRR domain, and thus it is speculated that the aberrant R protein isoforms induced by pathogen inoculation could form intermolecular interactions with their regular protein products. This would disrupt the closed conformation stabilized by intermolecular interactions and free active R proteins.
In addition to the auto-inhibition, R proteins are also subjected to negative regulation by *trans* factors (Belkhadir et al., 2004). RIN4, guarded by RPM1 and RPS2, is phosphorylated upon infection with *P. syringae* by AvrRpm1 and AvrB (Mackey et al., 2002; Mackey et al., 2003). The rin4 mutants cannot survive in the presence of wild-type RPM1 and RPS2, due to strong activation of defense responses independent of pathogen infection. However, the rin4 defective phenotype is suppressed in the triple mutant rin4 rps2 rpm1 (Belkhadir et al., 2004). It was deduced that interactions of RIN4 with RPM1 and RPS2 negatively regulate the activities of both of these R proteins. The down-regulation of R protein activity could also be achieved by limiting its accumulation to a steady level. SRFR1 (Suppressor of rps4-RLD1) interacts with SNC1 to negatively regulate production of several R proteins, such as RPS2, RPS4 and RPS6 (Li et al., 2010). Likewise, the F-box protein CPR1 (Constitutive expresser of PR genes 1) controls the stability of R proteins through SKP1-Cullin1-F-box (SCF)-mediated protein degradation (Cheng et al., 2011). Loss-of-function *cpr1* mutants displayed higher expression of SNC1 and RPS2, as well as autoimmunity responses. Excess R protein isoforms produced via alternative splicing upon effector recognition may compete with full-length R protein to interact with negative regulators and decrease the relative abundance of these suppressors, thereby releasing active R protein (Belkhadir et al., 2004). This assumption is in line with observations that the overexpression of some R genes, including *Rx*, *RPS2*, and *RPM1*, leads to constitutive activation of resistance signaling.

2. Function as signaling factors

Overexpression of the TIR or CC domain of some R proteins (e.g., RPS4, RPP1, MLA10, and L6) can induce HR in the absence of cognate effectors (Zhang et al., 2004; Swiderski
et al., 2009; Bernoux et al., 2011; Collier et al., 2011; Maekawa et al., 2011). In addition to the TIR-NBS-LRR-encoding R genes, plants also contain short pseudo-R gene homologs (TN and TX) (Meyers et al., 2002). TN proteins contain the TIR and NBS domains, but lack the LRR domain, while TX proteins have only the TIR domain followed by a small and variable C-terminal domain. *Arabidopsis* contains 21 TN and 30 TX genes (Meyers et al., 2003). Transient and stable overexpression of some TN and TX genes induced necrosis in tobacco leaves and reduced disease symptoms in *P. syringae*-infected *Arabidopsis* plants, respectively (Nandety et al., 2013). This suggests that the truncated R proteins resulting from alternative splicing may also confer disease resistance with or without recognition specificity.

The crystal structures of the TIR domain of L6 and CC domain of MLA10 indicated that two activated R proteins form a homodimer at the CC or TIR domain to constitute a minimal functional unit (Bernoux et al., 2011; Maekawa et al., 2011). In the presence of full-length R protein, the production of massive amounts of truncated proteins containing TIR or CC may serve as a rapid and energy-efficient mechanism to activate responses to pathogen infection. If so, the rapid increase of TIR or CC domain-dependent dimerization stimulated by alternative splicing of R genes might function to amplify the plant defense responses.

Protein function is associated with subcellular localization. It is possible that the alternative proteins generated by alternative splicing are localized to different compartments than the full-length R proteins, and numerous reports have demonstrated dynamic subcellular localization for R proteins such as RPS4 and N (Burch-Smith et al., 2007; Shen et al., 2007; Wirthmueller et al., 2007). Distinct signaling pathways can be
initiated by a single R protein in different subcellular localizations, and, thus, the coordinated trafficking of R proteins is required for the activation of full resistance (Heidrich et al., 2011). RPS4 is detected in both the endomembrane and nucleus in healthy and diseased leaves, with RPS4 accumulation in the nucleus appearing to be necessary for AvrRPS4-trigged immunity (Wirthmueller et al., 2007). AvrRPS4 also shows a nucleo-cytoplasmic distribution. Forcing AvrRPS4 to accumulate in cytoplasm through the C-terminal fusion of a nuclear export sequence led to moderate HR and partial suppression of bacterial growth. By contrast, sequestration of AvrRPS4 in the nucleus by fusion of nuclear localization sequence was sufficient for inhibition of bacterial growth, but cell death elicited by HR was abolished. HR signaling is therefore mediated by cytoplasmic RPS4-AvrRPS4 interaction, whereas the nuclear R-Avr interaction-induced resistance is not coupled to programmed cell death. This is in line with the findings that restriction of pathogen spread does not always correlate with HR (Bendahmane et al., 1999; Gassmann 2005; Coll et al., 2010). However, because the construct used for examination of RPS4 subcellular localization consisted of its genomic sequence with an upstream fusion of the reporter gene under the control of the 35S promoter, any differential targeting of full-length RPS4 compared to truncated variants could not be distinguished (Burch-Smith et al., 2007). It is likely that the truncated RPS4 proteins would accumulate in the endomembrane system, since their C-terminal lack a bipartite nuclear localization sequence, which is necessary for accumulation of full-length RPS4 in the nucleus. This could explain why only 6%–10% of RPS4 was observed in the nuclei. The distinct types of signaling triggered by nucleo-cytoplasmic distribution of R-
Avr interaction may be coordinated by alternative splicing and differential localization of the resultant protein isoforms.

**Regulation of alternative splicing of R genes**

Alternative splicing dramatically increases the diversity of the transcriptome, and alternative splicing of R genes plays crucial roles in regulating plant defense responses; therefore, the mechanisms that regulate alternative splicing must be finely tuned to control the levels of different alternative splicing transcripts. Removal of introns within pre-mRNA in eukaryotes is catalyzed by the spliceosome, a highly dynamic and complex macromolecule comprising five (U1, U2, U4, U5, and U6) small ribonucleoproteins (snRNPs) and numerous RNA binding proteins (RBPs), such as serine/arginine-rich (SR) proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs). The precise selection of intron/exons requires splicing factors to recognize four loosely conserved sequence features in pre-mRNA: (1) the 5' splicing site (SS) of GU paired with snRNP U1; (2) a branch point A for binding of splicing factor 1 at the 18 to 40 nucleotides upstream of the 3' SS; (3) the 3' SS of AG and (4) a poly-pyrimidine tract for recruitment of U2 auxiliary factor heterodimer (Reddy 2007; Chen and Manley 2009). It is noteworthy that a single intron may contain multiple sites for each of these four conserved sequence elements, adding more complexity in splicing site selection.

Differential selection of 5'- or 3'-SSs can be also affected by some short sequences of *cis*-elements in intronic and exonic region. According to the position and function, these *cis*-elements are grouped as exonic splicing enhancers, exonic splicing silencers, intronic splicing enhancers, and intronic splicing silencers. These splicing regulatory elements
bind to trans-acting splicing factors, such as SR proteins and hnRNPs, playing critical roles in both constitutive and alternative splicing through either inducing or suppressing selection of nearby 5’- or 3’-splicing sites (Day et al., 2012; Thomas et al., 2012). Interestingly, SR genes are also extensively alternatively spliced and alternative splicing of SR genes is affected by environmental stresses such as temperature, light and salt, which in turn induces splicing changes in the pre-mRNAs of other genes (Palusa et al., 2007).

As mentioned above, screening for suppressors of the gain-of-function mutation snc1 led to the identification of a set of MOS genes, some of which function in pre-mRNA processing. For example, Arabidopsis mutants carrying a loss-of-function mutation for MOS4, MOS12, or MOS14 show altered splicing patterns for SNC1 and RPS4, which indicate that those genes have regulatory roles in alternative splicing of R genes (Palma et al., 2007; Xu et al., 2011; Xu et al., 2012). MOS4, required for both ETI and PTI, is a nuclear localized CC homologous to human BCA2 (Breast cancer-amplified sequence 2). Together with the Myb-transcription factor CDC5L (Cell division cycle 5 like protein) and the WD-40 repeat PLRG1 (Pleiotropic regulator 1), BCA2 was isolated from humans as an important component of a multi-protein spliceosome complex that includes the E3 ubiquitin ligase Prp19 (Precursor RNA processing 19) (Ajuh et al., 2001; Zhou et al., 2002). Yeast two-hybrid and in planta assays confirmed that MOS4 interacted with the Arabidopsis homologs of CDC5L and PRLG1 (AtCDC5 and PRL1, respectively) to constitute a core structure for a spliceosome-associated complex termed the MOS4-associated complex (MAC) (Palma et al., 2007). MAC3A and MAC3B, two functionally redundant homologs of Prp19, contribute to proper splicing of SNC1, though their effects
on alternative splicing of \textit{RPS4} have not been investigated (Monaghan et al., 2009). Similarly, whether two other redundant homologs, MAC5A and MAC5B, function in \textit{R} gene alternative splicing has not been tested (Monaghan et al., 2010). However, given that its counterpart in human is RBM22, which interacts with U6 snRNP, it is possible that MAC5 participates in pre-mRNA splicing in plants.

\textit{MOS12} encodes an SR protein homologous to human cyclin L (Xu et al., 2012). Co-immunoprecipitation of MOS12 with MOS4 indicates that MOS12 is also associated with the MAC. The \textit{mos12} mutant displays compromised \textit{RPS4}-mediated resistance as well as an altered splicing pattern of \textit{RPS4}, leading to a different abundance ratio of \textit{RPS4} transcript isoforms. However, the splicing pattern of \textit{RPS6} is normal in the \textit{mos12} mutant, as is \textit{RPS6}-mediated resistance. This suggests that in addition to MAC, more spliceosomal complexes with distinct splicing specificities probably exist in plants.

Impaired \textit{SNC1}- and \textit{RPS4}-mediated PTI and ETI was also observed in the loss-of-function mutant of \textit{MOS14} (Xu et al., 2011). In addition to distorted splicing patterns, the \textit{mos14} mutants showed reduced expression of \textit{SNC1} and \textit{RPS4}. \textit{MOS14} encodes a nuclear protein homologous to transportin-SR, which functions in nuclear trafficking of the SR protein. MOS14 interacts with four different SR proteins through its C-terminus, while the N-terminus interacts with a GTP-binding protein AtRAN1 (Ras-related nuclear protein 1) which functions in many processes, including nuclear transport of proteins. The nuclear localization of these four proteins was disrupted in \textit{mos14} mutants, which consequently affects the splicing profiles for their targets. Defective splicing resulting from mis-localization of MOS14 cargos may cause the reduction in \textit{SNC1} and \textit{RPS4} expression (Xu et al., 2011).
Summary

The functional importance of alternative splicing in plant disease resistance has become increasingly clear. However, despite the substantial progress that has been made in the past decade, alternative splicing research in plant immunity is still in its infancy. The alternative splicing events characterized to date in CC-NBS-LRR genes appear not to be required for disease resistance. However, some truncated TIR-NBS-LRR proteins encoded by alternative transcripts are required for full R-gene mediated resistance. To further investigate the biological role of alternative splicing in plant disease resistance, we worked to characterize the role of alternative splicing in RCTI-mediated resistance in M. truncatula.
Chapter 2 Functional characterization of the role of alternative splicing in RCT1-mediated resistance in Medicago truncatula

Introduction

Alternative splicing generates multiple transcript variants from a single gene through selective use of different splice sites (Nilsen and Graveley 2010). It is a widespread mechanism to increase proteomic diversity and also contributes to regulation of gene expression in eukaryotic organisms (Brett et al., 2002; Kazan 2003; Nilsen and Graveley 2010; Syed et al., 2012). It was estimated that over 95% of multi-exonic genes in human are alternatively spliced (Pan et al., 2008). A recent analysis in Arabidopsis indicates that alternative splicing occurs in over 60% of the intron-containing genes (Marquez et al., 2012). It becomes increasingly evident that alternative splicing plays a key role in regulation of plant development and metabolism as well as in response to biotic and abiotic stresses (Syed et al., 2012; Staiger et al., 2013).

Alternative splicing appears to occur frequently in the TIR-NBS-LRR class of plant disease resistance (R) genes (Whitham et al., 1994; Lawrence et al., 1995; Gassmann et al., 1999; Dinesh-Kumar and Baker 2000; Jordan et al., 2002; Zhang and Gassmann 2003; Tan et al., 2007). The alternatively-spliced transcripts of R genes generally possess premature termination codons and thus encode putative truncated proteins lacking the LRR and/or C-terminal domains (Jordan et al., 2002). The best-studied examples are the tobacco N gene that confers resistance to tobacco mosaic virus (Whitham et al., 1994; Dinesh-Kumar and Baker 2000) and the Arabidopsis RPS4 gene that conditions resistance to Pseudomonas syringae strains expressing AvrRps4 (Zhang and Gassmann
2003; Zhang and Gassmann 2007). In the case of the tobacco N gene, the alternative transcript results from alternative splicing of a 70-bp exon within intron 3 (Whitham et al., 1994), while the alternative splicing of RPS4 involves the retention of intron 3 or introns 2 and 3 (Gassmann et al., 1999). In both of these examples, rapid and complete resistance requires the presence of both regular and alternative transcripts. Furthermore, the expression of alternative transcripts of N and RPS4 was up-regulated by pathogen infection, and such change in expression profiles appeared to be highly correlated with resistance responses (Dinesh-Kumar and Baker 2000; Zhang and Gassmann 2007). In line with these findings, loss-of-function mutations in splicing components that lead to altered alternative splicing pattern for RPS4 render plants susceptible to the pathogen (Xu et al., 2011; Xu et al., 2012). These studies indicated that truncated proteins or peptides encoded by alternative transcripts are important for triggering resistance responses, although their exact biochemical function remains to be determined. There are also cases, however, where the alternative transcript(s) appear not to be essential for full function of an R gene. For example, transgenic flax plants expressing an intronless L6 gene were completely resistant to the strain expressing the corresponding avirulence gene (Ayliffe et al., 1999).

RCT1 is a TIR-NBS-LRR gene in Medicago truncatula that confers resistance to multiple races of Colletotrichum trifolii, a hemi-biotrophic fungal pathogen that causes anthracnose disease in Medicago (Yang et al., 2008). RCT1, which shares a common gene structure with N and RPS4, undergoes alternative splicing at both coding and 3’-untranslated regions (UTR), thus generating multiple transcript variants in its expression profile. However, in contrast to N and RPS4, for which alternative splicing involves
intron 2 and/or intron 3, alternative splicing of *RCT1* in the coding region is caused by the retention of intron 4. Because the intron 4 lies downstream of the LRR-encoding exons and contains an in-frame stop codon, the alternative transcript is predicted to encode a truncated protein consisting of the entire portion of the TIR, NBS, and LRR domains but lacking the C-terminal domain of the full-length *RCT1* protein encoded by the regular transcript. Interestingly, the presence or absence of the alternative transcript appears to be strongly associated with resistance and susceptible phenotypes based on expression analysis of multiple, naturally occurring *RCT1* and *rct1* alleles. Moreover, the alternative splicing of *RCT1* orthologs appears to be conserved between *M. truncatula* and alfalfa (*M. sativa*). All these observations point to an important role of alternative splicing in *RCT1*-mediated disease resistance in Medicago.

In order to study the role of alternative splicing in *RCT1*-mediated resistance in *M. truncatula*, we developed several gene constructs that express the regular, the alternatively spliced, and a combination of both *RCT1* transcripts, respectively, and transferred these constructs into the susceptible alfalfa plants. The data showed that the *RCT1*-mediated disease resistance requires the combined presence of both regular and alternative transcripts. This study, in addition to those reports on the tobacco *N* and *Arabidopsis RPS4* genes (Dinesh-Kumar and Baker 2000; Zhang and Gassmann 2003; Zhang and Gassmann 2007), adds another significant example demonstrating the involvement of alternative splicing in *R* gene-mediated plant immunity.
Materials and methods

Plasmid construction

In a previous study, we created a genomic construct of \textit{RCT1} with its native promoter in pCAMBIA2300, named RCT1-G (Fig. 2.1A, Yang et al., 2008). This construct contained a 12.9 kb genomic fragment comprising of the ~5.0 kb \textit{RCT1} coding region plus ~3.6 kb upstream of the start codon and ~4.7 kb downstream of the stop codon. The transgene in alfalfa showed the same expression pattern as the endogenous gene in the resistance genotype Jemalong A17. In this study, we modified the RCT1-G plasmid to generate three intron-deleted constructs of \textit{RCT1} through sequence swaps (Fig. 2.1B-D). For this purpose, cDNA from Jemalong A17 was used as a template to amplify cDNAs of the regular (intron 4 spliced out) and alternative (intron 4 retained) transcripts, using the primer pair F1 and R1 designed from the 5’- and 3’-UTR, respectively (F1, forward, 5’-CCATAGATCTCTTCTTTTCTTTTCC-3’; R1, reverse, 5’-TGCCCTCAGCAATAAGAGCATT-3’). The shorter cDNA fragment, representing the regular, fully processed transcript (RT) without intron 4, was digested with \textit{MfeI} and \textit{Bbvcl} to obtain a 3,132-bp fragment that encompasses part of exon 1, exons 2, 3, and 4, and part of exon 5 of \textit{RCT1}. This fragment was then swapped to the RCT1-G construct digested with the same restriction enzymes (Fig. 2.2). The resulting gene construct, designated RCT1-Intronless (Fig. 2.1B), presumably generates only the regular transcript encoding the full length RCT1 protein.
Figure 2.1 Schematic diagram of the RCT1 gene constructs used in this study. Name of each gene construct is indicated on the right, with the transcript(s) it produces showing in the parentheses. Exons are drawn as boxes, with different styles showing the distinct protein domains or regions they code for. Retained introns are shown as black lines and spliced introns are shown as diagonal lines. Arrows indicate the positions of the primers used for expression analysis. Stars represent the premature termination codon identified in the alternative transcript.
Similarly, the longer cDNA fragment, representing the alternatively spliced transcript (AT) with retained 448-bp intron 4, was digested with MfeI to obtain a 2,529-bp fragment containing part of exon 1, exons 2, 3, and 4, and part of intron 4 of RCT1. This fragment was then ligated into the MfeI-digested RCT1-G construct (Fig. 2.3). We named this vector as RCT1-Intron4 (Fig. 2.1C). Because of the removal of introns 1-3 and retention of intron 4, RCT1-intron 4 is expected to encode both the regular and alternative transcripts, assuming that alternative splicing of intron 4 still occurs.

**Figure 2.2** Diagrammatic representation of RCT1-Intronless construct.
To generate a gene construct that only encodes the truncated protein, we modified the RCT1-Intron 4 construct by deleting part of the intron 4 (after the in-frame stop codon) and the entire exon 5, and named this construct as RCT1-Tr (Fig. 2.1D). To generate this construct, a 136bp fragment (ST) that contains introduced sequence with *MfeI* site to encode part of RCT1 truncated protein and 3’-UTR sequence was amplified from Jemalong A17 cDNA, using the primer pair Tr-F and R1 designed from the 3’-UTR (Tr-F, forward, 5’- CCAATTGTTACACTTTTTATATAGATATGG AAGGTGTG -3’; R1, reverse, 5’-TGCCTCAGCAATAAGAGCATT-3’). And then the 136bp PCR fragment was digested with *MfeI* and *Bbvcl* and ligated into the RCT1-Intron 4 construct digested
with the same restriction enzymes to obtain the RCT1-Tr construct (Fig. 2.4). The transcript generated by this construct is expected to encode a truncated (Tr) protein with 936 amino acids, identical to that predicted from the intron 4-containing alternative transcript. All these gene constructs are driven by the RCT1 native promoter because we maintained the entire 5’ genomic region upstream of the start codon and 3’ genomic region downstream of the stop codon.

**Figure 2.4** Diagrammatic representation of RCT1-Tr construct
We also created cDNA constructs of \textit{RCT1} driven by the CaMV 35S promoter in pCAMBIA2300. For this purpose, we amplified cDNAs corresponding to different types of transcripts using plasmid DNA of above mentioned constructs as templates. The primer pair used are F3 and R3 with introduced cloning sites of \textit{BamHI} and \textit{SacI} (F3: 5'-TCTCTCAAGCTTGGATCCCATAGATCTCTTTTTTCC-3'; R3: 5'-GCTCTAGAGGATCAATTCGTGCCTCAGCAATAAGAGCATT-3'). These fragments were digested with \textit{BamHI} and \textit{SacI} and ligated with the modified pCAMBIA2300 vector (with the added 35S promoter and transcription termination signals). Accordingly, we named the individual constructs as \textit{35S}:RCT1-Intronless, \textit{35S}:RCT1-Intron4, and \textit{35S}:RCT1-Tr, respectively.

For the generation of RNAi knockdown lines of alfalfa, a 470-bp inverted-repeat sequence from the third exon of \textit{Mt-EDS1} was cloned into the modified pHELLSGATE8 vector (Helliwell and Waterhouse 2003). We modified the pHELLSGATE8 by replacing the kanamycin-resistance gene (\textit{nptII}) with the hygromycin-resistance gene (\textit{hph}). The construct was introduced into \textit{Agrobacterium tumefaciens} strain GV3101 and transformed to a RCT1-G transformed alfalfa clone as described below.

**Plant transformation**

The susceptible alfalfa plants from the clone SY6 (Yang et al., 2008) were transformed with \textit{A. tumefaciens} GV3101 carrying various transgene constructs described above. Transformation of alfalfa followed the protocol developed by Samac and Austin-Phillips (Samac and Austin-Phillips 2006). Fresh alfalfa leaves were cut from Regen SY6 plants, surface sterilized with 70% ethanol for 10s and transferred to 20% bleach with 0.05%
Tween20 for 3 min, followed by three times rinse with sterile water. Leaf margins were removed and then the leaflets were cut into about 0.5 cm × 0.5 cm pieces and placed into SHO liquid medium (Schenk and Hildebrant salts, Schenk and Hildebrant vitamins, 30g/L sucrose, 0.5g/L MES, pH 5.7 with KOH). When sufficient leaf pieces had been collected, A. tumefaciens GV3101 cells carrying various transgene constructs from an overnight culture at 28°C in liquid YEP medium (10g/L protease peptone, 10g/L yeast extract, 5g/L NaCl, 50mg/L kanamycin) were added to SHO medium (1ml Agrobacterium/4 ml SHO). The cell density was adjusted to OD₆₀₀ ≈ 0.6. After 15min incubation with Agrobacterium cells, the leaf pieces were blotted briefly on sterile filter paper to remove excess liquid and then placed on B5h medium (3.1 g/L Gamborg's B5 salts, 1.0ml/L 1000x Gamborg's B5 vitamins, 0.5g/L KNO₃, 0.25g/L MgSO₄·7H₂O, 0.5g/L proline, 30g/L sucrose, pH 5.7 with KOH, 8g/L Phytagar) with 30ml/L stock aminos (26.6g/L L-glutamine, 3.32g/L serine, 0.016g/L adenine and 0.332g/L L-glutathione) and hormones (1mg/L 2,4-D and 0.1mg/L kinetin) for 3-5 days. After coculture period, the leaf pieces were rinsed three times, blotted gently on sterile filter paper and then transferred to B5hKTc selection medium (B5h medium with stock aminos and hormones plus 50mg/L kanamycin and 500mg/L ticarcillin). All the plates were incubated at 24°C, 16h photoperiod with light intensity of 60-80 μE/m2s and 8h dark period. After three weeks, the leaf pieces-derived calli were formed and moved to B5hOKTc regeneration medium (similar with B5hKTc but without hormones). After three to four weeks on B5hOKTc plates, embryos were separated and transferred singly to MMSKTc selection medium (4.3g/L Murashige and Skoog salts, 1ml/L 1000x Nitsch and Nitsch vitamin stock, 0.1g/L myo-inositol, 30g/L sucrose, pH 5.7 with KOH, 7.0g/L...
phytagar plus 500mg/L ticarcillin and 50mg/L kanamycin.). Over the next one to three weeks, embryos were formed a shoot and sometimes a root. Green plantlets were moved to MMSTc medium (similar with MMSKTc but lacking kanamycin) for further shoot and root development. Once a good root system was formed, the plants were removed from medium and placed into the soil for following inoculation experiments of *C. trifolii* race 4. All transformants were selected on 50mg/L kanamycin. At least 10 independent transformants were generated for each construct.

**Pathogen inoculation and phenotypic analysis**

*Colletotrichum trifolii* race 4 (isolate OH-WA-520) was used for inoculation as described by Yang et al. (2007). Mycelium was grown on ANM plates (20.0g/L Malt extract, 1.0g/L bactopeptone, 20.0g/L glucose, and 20.0g/L agar) in the dark at 23°C in petri dishes. Conidia were produced after one week at 23°C on Emerson’s YPSS medium (4.0g/L Yeast extract, 15.0g/L soluble starch, 1.0g/L K₂HPO₄, 0.5g/L MgSO₄·7H₂O, pH 6.8 with KOH, 20.0g/L Agar). Spores were collected and washed three times in sterile water with the final concentration being adjusted to 2×10⁶ spores/ml. Plants were inoculated by injection of spores into the stems of living plants using a latex free syringe with a thin needle (0.4 mm ×13 mm). At least two stems of each plant were inoculated. Inoculated plants were then transferred to a growth chamber programmed for a 16 h light, 23°C and 8 h dark, 20°C regime with >90 % humidity. Symptoms were recorded five days after inoculation. Plants were scored for development of a resistant (no symptom) or susceptible (stem collapse) response to *C. trifolii* 5–20 days post inoculation (dpi). For gene expression analysis of *RCT1*, spray inoculation was applied. Plants were sprayed
with *C. trifolii* race 4 spore suspension (2×10⁶/ml) and maintained in a growth chamber with >90% humidity. Leaves were collected at 0, 6, 12, and 24 h post inoculation.

**RNA isolation and analysis of gene expression by reverse-transcriptase (RT)-PCR**

Total RNA was isolated using the QIAGEN Plant RNAeasy miniprep Kit. Two micrograms of RNA was used to perform RT reactions using M-MLV reverse transcriptase (Invitrogen) in a 20-μl reaction mixture. Two microliters of the RT reaction was used as a template in a 20-μl PCR solution. The PCR primers were as follows: *MtActin*, F: 5’-GGAGAAGCTTGCATATGTTG-3’ and R: 5’-TTAGAAGCACTTCCTGTGGA-3’; *RCT1*, F1: 5’-CCATAGATCTCTTTTCTTTTCC-3’ and R1: 5’-TGCCTCAGCAATAAGAGCATT-3’, or F2: 5’ -CAAAAGCTTGTTGAGGACTG-3’ and R2: 5’-ATTTTCGACGACTGGTTTCATC-3’; and *EDS1*, F: 5’-AGCACGAATTTGTTGTTGGAGA-3’ and R: 5’-TTGGCAATATCAAGAGGCTCAA-3’.

**Results**

**Alternative splicing of the *M. truncatula RCT1* gene**

In a previous study, we reported that the *RCT1* gene undergoes alternative splicing at both coding and 3’-UTR regions, therefore producing multiple transcript variants in the *RCT1* expression profile (Yang et al., 2008). Alternative splicing in the coding region is caused by the retention of the 448-bp intron 4 (Fig. 2.1). Hereafter we refer to this alternative transcript as *RCT1^At* and the fully processed, regular transcript as *RCT1^Rt*. In contrast to *N* and *RPS4*, *RCT1^Rt* and *RCT1^At* transcripts were both abundantly expressed in the resistant genotypes, and the expression of *RCT1^At* was constitutive, not apparently up-regulated by pathogen infection within and beyond 24 h post inoculation (Fig. 2.5).
survey of multiple susceptible and resistant *M. truncatula* genotypes revealed that RCT1\textsuperscript{At} was expressed in all resistant genotypes, while most susceptible alleles show no or very weak expression of RCT1\textsuperscript{At} (Yang et al., 2008).

The RCT1\textsuperscript{Rt} transcript is predicted to encode a protein of 1098 aa, consisting of an N-terminal TIR domain, a centrally located NBS domain followed by seven degenerate LRRs, and a C-terminal domain that is highly conserved with other members of TIR-NBS-LRR genes in *M. truncatula* and closely related legumes. Retention of intron 4 in RCT1\textsuperscript{At} results in a shift in the open reading frame and a premature termination codon within intron 4. As such, RCT1\textsuperscript{At} is predicted to encode a truncated protein of 936 aa, of which the first 920 aa are identical to those of the full-length protein, encompassing the entire portion of TIR, NBS, and LRR domains but lacking the C-terminal domain.

**Figure 2.5** Semi-quantitative reverse-transcriptase (RT)-PCR analysis of the *RCT1* regular and alternative transcripts in the *M. truncatula* genotype Jemalong A17 within 24 h post inoculation (hpi). Primers used were F2 and R2 that span the intron 3 and intron 4. RCT1\textsuperscript{At} represents the alternative transcript that retained intron 4 (~2.0 kb); RCT1\textsuperscript{Rt} represents the regular transcript with intron 4 spliced out (~1.5 kb). The number in parentheses indicates the cycle number of the RT-PCR. The *M. truncatula* Actin gene was used as a control.
We also detected complex alternative splicing that occurs at the 3’-UTR region of \textit{RCT1}, represented by at least 3 transcript variants of 721, 734, and 801 bp beyond the stop codon (Yang et al., 2008). These transcript variants attribute to the alternative splicing of three additional introns. For instance, the 801-bp fragment results from retention of the 80-bp intron relative to the 721-bp fragment. In this study, we only focused on addressing the potential role of alternative splicing of the coding region in the \textit{RCT1}-mediated disease resistance.

\textbf{The full-length or truncated open reading frame alone is not sufficient for resistance}

To determine the role of individual \textit{RCT1} transcripts in conferring resistance to \textit{C. trifolii}, we created several \textit{RCT1} gene constructs in which all or part of the introns were removed (Fig. 2.1). These intron-deleted genes were derived from a genomic \textit{RCT1} construct (RCT1-G) driven by its native promoter (Fig. 2.1A, Yang et al. 2008). To generate the intronless gene construct, genomic sequences encompassing the entire coding region of \textit{RCT1} was swapped for cDNA sequences derived from the regular transcript (Fig. 2.1B). This construct, named RCT1-Intronless, presumably generates only RCT1\textsuperscript{Rt} that encodes the full-length RCT1 protein. Likewise, we developed a gene construct, called RCT1-Tr, by deletion of introns 1–3, part of intron 4, and entire exon 5 (Fig. 2.1D). The RCT1-Tr gene construct codes for an artificial transcript (RCT1\textsuperscript{Tr}) that translates to a truncated protein identical to that predicted from the intron 4-containing alternative transcript (RCT1\textsuperscript{At}). All these gene constructs were driven by the \textit{RCT1} native promoter since we retained the entire 5’ genomic region upstream of the start codon and 3’ genomic region downstream of the stop codon.
Since *M. truncatula* is recalcitrant to transformation and regeneration, we used alfalfa (*M. sativa*) as a study system (Yang et al., 2008). Because of the high degree of sequence identity and remarkably conserved genome structure and function between the two species, this strategy has been widely used to validate the function of cloned *M. truncatula* genes. We transformed the RCT1-Intronless and RCT1-Tr gene constructs to the susceptible alfalfa clone SY6, and at least 10 independent transformants from each construct were tested for anthracnose resistance from 5 to 20 days post inoculation. We also generated transgenic plants containing the empty vector and the RCT1-G construct for use as negative and positive controls, respectively. As expected, transgenic plants containing the RCT1-Intronless gene construct, in which introns 1–4 were all removed, generated only RCT1*R* predicted to encode full length RCT1 protein (Fig. 2.6A). Similarly, plants transformed by the RCT1-Tr gene construct produced the expected artificial transcript, RCT1*Tr*, predicted to code for the truncated protein (Fig. 2.6A).

To assay for resistance and susceptibility, we inoculated the plants by injection of the fungal spores into the stems of living plants (Yang et al., 2007). In the absence of resistance, inoculated stems of susceptible plants invariably form large lesions at the inoculation sites and collapse with severe anthracnose symptoms at 5 days post inoculation, while inoculated stems of resistant plants grow normally and are completely symptomless. As shown in Fig. 2.7, the control plants transformed with the RCT1-G construct were completely resistant to the *C. trifolii* race 4 (Fig. 2.7A), whereas plants transformed with the empty vector were fully susceptible (Fig. 2.7B). Strikingly, plants transformed with the RCT1-Intronless (Fig. 2.7C) or RCT1-Tr construct (Fig. 2.7E) were also susceptible.
Figure 2.6 Semi-quantitative reverse-transcriptase (RT)-PCR analysis of transgene expression in plants transformed with the RCT1-Intronless or RCT1-Tr constructs driven by A. the *RCT1* native promoter or B. the CaMV-35S promoter. Four independent transgenic lines from each construct were presented. The number in parentheses indicates the cycle number of the RT-PCR. The *M. truncatula Actin* gene was used as a control. S, susceptible; R, resistance.
Figure 2.7 Disease resistance assay for transgenic alfalfa plants expressing the native RCT1 gene and the intron-deleted constructs. The transgenes were driven by the 35S promoter if indicated; otherwise they were under the control of the RCT1 native promoter. Plants were inoculated with the C. trifolii race 4 as described in “Materials and methods”. Arrows indicate inoculated stems. S, susceptible; R, resistance.

However, the expression levels of the transgenes in the plants transformed with the RCT1-Intronless were overall low when compared with transgene expression levels of plants transformed by the RCT1-G construct (Fig. 2.6A), which was likely due to the removal of introns from the wild-type RCT1 gene (Le Hir et al., 2003; Moabbi et al., 2012). In order to test whether the observed phenotypic differences were the result of differential transgene expression levels, we developed gene constructs that express the same transcripts under the control of the CaMV 35S promoter. We did not observe any aberrant phenotypes in the 35S transgenic lines. As shown in Fig. 2.6B, the transgene expression in these plants was enhanced to a similar or even a higher level than plants...
transformed with RCT1-G. Despite the enhanced expression of the transgenes, all the transgenic plants were still susceptible to the pathogen (Fig. 2.7D, F). However, the symptom occurrence (branch collapsed) was delayed for the plants transformed with the 35S: RCT1-Intronless compared to those transformed with the empty vector, suggesting that the regular transcript could confer partial resistance to pathogen infection (Table 2.1). In contrast, we did not observe any partial resistance in 35S: RCT1-Tr transgenic plants, indicating that the truncated RCT1 protein alone was not functional.

Taken together, these data suggested that neither the regular nor alternative RCT1 transcript alone is sufficient to confer resistance against the pathogen, consistent with that reported for the tobacco N and Arabidopsis RPS4 genes (Dinesh-Kumar and Baker 2000; Zhang and Gassmann 2003).

Table 2.1  The time of anthracnose symptom onset in transgenic alfalfa plants. dpi, days post inoculation.

<table>
<thead>
<tr>
<th>Onset time(^a) (dpi)</th>
<th>(RCT1):RCT1-G Intronless</th>
<th>(RCT1):RCT1-Tr</th>
<th>(RCT1):RCT1-Intron4 (low expression level)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empty vector</td>
<td>5.00±0</td>
<td>7.42±3.62</td>
<td>5.04±0.20</td>
</tr>
<tr>
<td>Onset time(^a) (dpi)</td>
<td>None</td>
<td>5.39±0.71</td>
<td>5.04±0.19</td>
</tr>
</tbody>
</table>

\(^a\) The average time of symptom onset based on at least 10 inoculated transformants for each construct. (P<0.05)
Removal of introns 1-3 affects transgene expression but has no effect on alternative splicing of intron 4

To investigate the importance of intron 4 for RCT1 function, we generated a gene construct in which intron 4 was retained but introns 1, 2 and 3 were removed. We named this gene construct as RCT1-Intron4 (Fig. 2.1C). The RCT1-Intron4 construct was derived from RCT1-G by replacing the genomic sequences surrounding the coding region with cDNA sequences of the alternatively spliced transcript (RCT1Alt). RCT1-Intron4 was expected to encode either or both of the regular and alternative transcripts depending on the splicing pattern of the transgene.

We introduced the RCT1-Intron4 construct into the susceptible alfalfa clone SY6 to obtain stable transgenic plants. Expression analysis of the transgenic plants identified two transcripts, which, based on sequence analysis, were correspondent to the regular and alternatively-spliced transcripts encoded by the RCT1 genomic construct (Fig. 2.8A). This experiment indicated that the alternative splicing of intron 4 was independent of the presence or absence of other introns in the transgene.

However, the transgene was expressed at a low level relative to the transgene derived from the RCT1-G construct (Fig. 2.8A). In contrast to the plants transformed with the native RCT1 gene that were resistant to the pathogen, disease assays revealed that plants transformed with RCT1-Intron4 were susceptible (Fig. 2.7G). Since both RCT1-G and RCT1-Intron4 generated the same types of transcripts, we hypothesized that the susceptibility of the transgenic plants containing the RCT1-intron4 construct may be due to the low expression level of the transgene.
Figure 2.8 Semi-quantitative RT-PCR analysis of transgene expression in plants transformed with the RCT1-Intron4 construct driven by A. the RCT1 native promoter or B. the CaMV-35S promoter. Multiple independent transgenic lines from each construct were presented. The number in parentheses indicates the cycle number of the RT-PCR. The *M. truncatula* Actin gene was used as a control. S, susceptible; R, resistance. The faint bands between RCT1<sup>At</sup> and RCT1<sup>Rt</sup> transcripts are heteroduplex resulting from RT-PCR of alternatively spliced mRNAs of *RCT1* (Yang et al., 2008).
Overexpression of the intron 4-containing transgene is sufficient for resistance

To further clarify whether or not the combined presence of regular and alternative transcripts is sufficient for RCT1 function, we created a gene construct that expresses the RCT1-Intron4 gene under the control of the 35S promoter (35S: RCT1-Intron4). We introduced the construct into the susceptible alfalfa clone and obtained transgenic plants with various levels of transgene expression. Strikingly, such variation in expression levels was highly correspondent to resistance responses. Transgenic plants that showed a high level of transgene expression were resistant to the pathogen (Fig. 2.7H). In contrast, the plants with low-level expression of the transgene were susceptible (Fig. 2.8B), although symptom occurrence was postponed in these plants (Table 2.1). Interestingly, the 35S promoter appeared to greatly enhance the expression of the alternatively spliced transcript but not significantly for the regular transcript in the resistant transgenic plants. These data suggested that the expression level of the alternatively spliced transcript is pivotal for triggering defense response. This is in line with the observations for the N and RPS4 genes, for which enhancing the expression of alternative transcripts to certain threshold was essential for the resistance function (Dinesh-Kumar and Baker 2000; Zhang and Gassmann 2007).

RCT1-mediated resistance requires EDS1

Enhanced Disease Susceptibility 1 (EDS1) has been shown to be essential for basal resistance and for effector-triggered immunity mediated by TIR-NBS-LRR class of R genes, including N and RPS4 (Aarts et al., 1998; Liu et al., 2002; Peart et al., 2002; Zhang et al., 2004; Hu et al., 2005). In particular, the induction of RPS4 expression by
pathogen inoculation was dependent on *EDS1* (Zhang and Gassmann 2007). EDS1 can interact with pathogen effectors, and such interactions can trigger resistance signaling through disruption and then activation of a protein complex formed by EDS1, TIR-NBS-LRRs, and SRFR1, a negative regulator of R protein-mediated resistance (Bhattacharjee et al., 2011; Heidrich et al., 2011).

We identified the *M. truncatula EDS1* ortholog as MTR_3g079340, a single copy gene in the *M. truncatula* genome. To determine if *MtEDS1* is essential for *RCT1*-mediated resistance in Medicago, we generated an RNA interference (RNAi) construct consisting of a 470-bp inverted-repeat sequence from the third exon. We introduced the RNAi construct into a RCT1-G transformed alfalfa clone to obtain double-transgenic alfalfa plants. Eight independent transgenic RNAi lines were obtained, for which *EDS1* expression was dramatically down-regulated (Fig. 2.9A). Disease assay revealed that all these lines were susceptible to the *C. trifolii* race 4 (Fig. 2.9B). Thus, we conclude that *RCT1*-mediated resistance requires *EDS1*. 
Figure 2.9 Silencing of *EDS1* in *RCT1*-transformed alfalfa compromised disease resistance. **A.** Down-regulation of *EDS1* in the eight RNAi lines. The RNAi construct was transformed to the *RCT1*-transformed alfalfa clone to obtain double-transgenic plants. The number in parentheses indicates the cycle number of the RT-PCR. The *M. truncatula* Actin gene was used as a control. All eight lines were susceptible to the *C. trifolii* race 4. **B.** An example of *EDS1* silenced plants susceptible to the *C. trifolii* race 4. Arrows indicate inoculated stems. *S*, susceptible; *R*, resistance.
Discussion

We have shown in this study that the alternative splicing of RCT1 is essential for its resistance function. Neither the regular nor the alternative transcript alone is sufficient to confer resistance. The alternatively spliced transcript (with retained intron 4) was constitutively expressed, without apparent induction by pathogen infection. Removal of introns 1-3 had no effect on the alternative splicing of intron 4; however, the expression of the regular and alternative transcripts was low when driven by the native RCT1 promoter, and resulted in disease susceptibility. Nonetheless, over-expression of the intron 4-containing transgene under the control of the 35S promoter is sufficient for resistance. Therefore the combined presence of transcript variants at appropriate abundance is required for RCT1 function. Together with the previous reports on the tobacco N and Arabidopsis RPS4 (Dinesh-Kumar and Baker 2000; Zhang and Gassmann 2003; Zhang and Gassmann 2007), our data indicated that the alternative TIR-NBS-LRR gene transcripts that encode putative truncated proteins are of importance in R gene-mediated plant immunity.

The fact that the alternatively spliced transcript of RCT1 was constitutively expressed is in contrast to the expression of alternative transcripts of the tobacco N and Arabidopsis RPS4 genes (Dinesh-Kumar and Baker 2000; Zhang and Gassmann 2003; Zhang and Gassmann 2007). The alternatively spliced transcript of N greatly increases 6h after TMV infection, and decreases to the original state 9h after infection. An artificial combination of regular and alternative spliced isoforms at a 1:1 ratio in a single plant cannot restore complete N-mediated resistance to TMV (Dinesh-Kumar and Baker 2000), indicating the expression of both N messages at a certain ratio is crucial for conferring complete
resistance to pathogen. In the case of *RPS4*, the alternative splicing of *RPS4* undergoes dynamic changes during the resistance response. Zhang and Gassmann (2007) showed that the expression of *RPS4* transcript variants was induced by the presence of AvrRps4 and demonstrated that *RPS4*-mediated full resistance is dependent on the finely balanced or tightly regulated numbers of regular and alternatively spliced transcripts of *RPS4*. In contrast, *RCT1*<sup>RT</sup> and *RCT1*<sup>AT</sup> expressed at a consistent level during 24h after *C. trifolii* inoculation (Fig. 2.5). Together with the fact that the lower expression level of *RCT1* results in susceptible phenotype (Fig. 2.8), our data suggested that the abundance of the transcript variants is crucial for *RCT1*-mediated resistance.

Dinesh-Kumar and Baker (2000) observed partial resistance in transgenic plants carrying the full length of *N* cDNA only. In contrast to their observation, the transgenic plants expressing individual *RPS4* transcript variants, including the full-length RPS4 protein-encoding transcript, all failed to confer measurable pathogen resistance (Zhang and Gassmann 2003). Unlike *RPS4* but consistent with *N*, partial resistance was also observed in transgenic plants expressing the *RCT1*<sup>RT</sup> transcript only. However, plants transformed with the RCT1-Tr construct were fully susceptible.

Although not exclusively found in TIR-NBS-LRR genes, the functional importance of the alternatively spliced transcripts that encode putative truncated proteins has been mostly reported for this class of plant *R* genes. However, the functional mechanisms for these alternative transcripts remain unknown. Zhang and Gassmann (2007) proposed a model for the function of the truncated RPS4 protein. In this model, the sum of truncated RPS4 proteins is responsible for priming the *RPS4*-dependent resistance response, whereas the up-regulated *RPS4*<sup>TN4L</sup> containing the entire TIR and NBS domains and the first four
LRRs, specifically function in amplifying the plant defense response. After activation of
RPS4-mediated resistance, the unstable RPS4\textsuperscript{TN4L} leads to down-regulation of this
amplified resistance signal. In our study, we speculated that the truncated RCT1 protein
may function in amplifying the RCT1-mediated defense response as well. Transgenic
plants expressing the regular RCT1 transcript showed partial resistance, suggesting that
the RCT1-mediated defense response is activated in these transformants. However, the
defense signal is not strong enough to confer complete resistance to pathogen attack.
Therefore, the truncated RCT1 protein may function as a “helper” to amplify the defense
signal to activate full resistance in plants. This hypothesis is supported by the observation
that the transgenic plants expressing the truncated RCT1 protein-encoding transcript all
are fully susceptible to pathogen. To elucidate the biological function of the truncated
RCT1 protein in plant defense, future research should focus on its dynamic and
subcellular localization in plant cells, and interactions with Avr proteins and their host
targets. In addition, identify the cis-sequences and trans-acting factors that are required
for RCT1 alternative splicing could help us to understand the importance of alternative
splicing in plant immunity.
Part 2

Chapter 3 Legume-rhizobial symbiosis

Introduction

Nitrogen is the most limiting nutrient for plant growth. The biological nitrogen fixation provides a nitrogen supply route for both natural vegetation and crop plants (Vance 2002). Biological nitrogen fixation is a process in which the atmospheric dinitrogen (N\textsubscript{2}) is enzymatically reduced to ammonia, a biological form that can be directly consumed by the plant, with the participation of microorganisms under the action of nitrogenase. The largest contribution to biological nitrogen fixation is carried out by nitrogen-fixing soil bacteria, collectively called rhizobia, including a large group of α- and β-proteobacteria (Berrada and Fikri-Benbrahim 2014).

Legumes represent the second most important family of crop plants, accounting for ~27% of the world’s primary crop production (Graham and Vance 2003). A hallmark trait of legumes is their unique ability to establish a symbiotic relationship with rhizobia. The rhizobial mutualism promotes the formation of a new plant organ, the root nodule, within which the bacteria can convert the nitrogen gas into ammonia and provide the fixed nitrogen to the host. In return, the plant provides the bacteria with the carbohydrates and other nutrients (Lodwig and Poole 2003). On a world-wide basis, legumes can fix about 40 to 60 million metric tons of nitrogen worth an estimate of 7 to 10 billion dollars per year (Smil 1999). In comparison, the world’s nitrogen fertilizer usage in the form of ammonia and its compounds is ~100 million metric tons per year (FAO 2008). Even though the nitrogen fertilizers are effective, they are expensive and also pollute the
environment. Therefore, the legume-rhizobia symbiosis becomes more attractive to sustainable agriculture, because it not only reduces the need for exogenous nitrogen fertilizer but also provides an efficient way to improve crops yields as well as produce protein-rich foods. Understanding the nodulation and nitrogen fixation process represents a key objective for plant biologists, with significant implications for both agricultural and natural ecosystems.

An overview of the legume-rhizobia symbiosis

The development of nitrogen-fixing symbiosis in legumes is remarkable in both its complexity and its overriding importance in the biosphere nitrogen cycle. The symbiotic nitrogen fixation involves the conversion of atmospheric N\textsubscript{2} to NH\textsubscript{3}, a reaction that is catalyzed by the rhizobial enzyme nitrogenase (Postgate 1998). The conversion of N\textsubscript{2} to NH\textsubscript{3} requires large quantities of cellular energy derived from aerobic metabolism to generate ATP (Sprent and Raven 1985). Paradoxically, the nitrogenase is extremely sensitive to and irreversibly inactivated by free oxygen. This so-called “oxygen paradox” is solved by a specialized plant organ, the root nodule. The root nodule provides high flux of cellular energy in an environment where oxygen tension is controlled by the binding of free oxygen by leghemoglobin, a component that is synthesized by legume plants when the roots are colonized by rhizobia (Appleby 1984).

The legume-rhizobia interaction begins with a molecular dialogue between the symbiotic partners (Masson-Boivin et al., 2009). Starting in the rhizosphere, legume root exudates that belong to the flavonoids family are perceived by the bacterium, resulting in the activation of a suite of bacterial genes, termed the *nod* genes. The *nod* genes encode
enzymes that are required for the synthesis and secretion of a highly specific signal of bacterial origin, known as Nod factors or lipo-chitooligosaccharides (LCOs) (Lerouge et al., 1990). Perception of Nod factors by the plant induces a series of host responses, including ion fluxes, calcium spiking, root hair deformation and curling, entry of the bacteria into the root hairs, transcriptional reprogramming of the host symbiotic genes, and cortical cell divisions, that ultimately lead to the development of rhizobia-infected root nodules (Oldroyd and Downie 2008; Oldroyd et al., 2011).

The rhizobial infection and nodule development follow a well-defined morphological program (Oldroyd and Downie 2008). Rhizobia first attach the tip of root hairs, resulting in deformation and curling of the epidermal cells of root hair. Then the bacteria become entrapped in a pocket formed by the curl, within which a small rhizobial colony is formed. During the process, the cell wall of curled root hairs is locally degraded and the root hairs begin an inverse tip growth from the trap site, forming a long and narrow plant-derived tube-like structure, called infection thread (Oldroyd and Downie 2008). This infection thread continuously divides at the leading edge so that the bacteria can ‘travel’ in it. When the infection thread reaches the nodule primordium, the bacteria are released into the plant cytoplasm and enveloped by a plant-derived membrane via an endocytosis-like process, within which the bacteria enlarge and differentiate into nitrogen-fixing forms, called bacteroids (Oke and Long 1999). The bacteroids, the peribacteroid membranes, and the interface between them form an organelle-like structure, known as the symbiosome (Saalbach et al., 2002). Symbiosomes are the sites for nitrogen fixation and nutrient exchange between the two symbiotic partners. Key aspects of symbiotic metabolism include the supply of energy in the form of carbon from the plant to the
bacterium, and the return of reduced nitrogen in the form of ammonia from the bacterium to the plant.

Entry through root hair curling is not the only mechanism for rhizobial invasion. In many basal legume species, the rhizobium can enter the host plant through the breaks in the root epidermis or epidermal damage formed by the emergence of lateral roots, known as crack entry (Sprent and James 2007). In this situation, the intercellular or intracellular infection threads are originated from the infected cracks in the outer cortical cells and direct the bacteria to nodule primordia (Oldroyd and Downie 2008).

**Nodulation signaling in legumes**

Establishing a successful symbiosis relationship between the two symbiotic partners requires two tightly coordinated process: bacterial infection at the root epidermis and nodule development which initiates at the cortex and involves complex signal transduction at epidermal, cortical and pericycle cells (Oldroyd and Downie 2008). Genetic studies in the two model legumes *Medicago truncatula* and *Lotus japonicas* identified a number of host genes that are required for rhizobial infection and nodule development (Oldroyd et al., 2009; Madsen et al., 2010). Analysis of these genes has begun to reveal the nodulation signaling pathway that is universally conserved in legumes (Fig. 3.1). At the beginning, the Nod factor signal is perceived by the plant receptors located in the epidermis (Oldroyd et al., 2011). Two LysM (lysin motif) domain containing receptor-like kinases were identified as Nod factor receptors in several plants: LjNFR1 and LjNFR5 in *L. japonicas*, PsSYM37 and PsSYM10 in *P. sativum,*
Figure 3.1 The symbiosis signaling pathway. Epidermal cells are able to perceive Nod factors (NFs)/ Myc factors (MFs) through LysM domain containing receptor kinases that activate the common symbiosis signaling pathway in mycorrhizal and nodulation symbioses. The Nod factors-induced calcium spiking is required a suite of proteins, including leucine-rich repeat receptor kinase LjSYMRK/ MtDMI2, two nuclear-localized potassium channels LjCASTOR and LjPOLLUX/ MtDMI1, three nucleoporins LjNUP85, LjNUP133 and LjNENA. The Ca\(^{2+}\)/ calmodulin-dependent protein kinase LjCCaMK/ MtDMI3 and its interaction protein LjCYCLOPS/ MtIPD3 are presumably responsible for decoding and transmitting the calcium spiking signal and activate several nodulation-specific transcription factors, such as MtNSP1 and MtNSP2, leading to transcriptional reprogramming of the host symbiotic genes.
GmNFR1α/β and GmNFR5α/β in Glycine max and MtLYK3/ MtLYK4 and MtNFP in M. truncatula (Limpens et al., 2003; Madsen et al., 2003; Radutoiu et al., 2003; Arrighi et al., 2006; Smit et al., 2007; Haney et al., 2011; Lin et al., 2012; van Hameren et al., 2013). The LysM domains are responsible for perception and binding the rhizobial Nod factors, and the kinase domains are required for transmitting the receptor-activated nodulation signal. In L. japonicas, both LjNFR1 and LjNFR5 bind Nod factors directly at high-affinity binding sites and form a heterodimeric complex to transduce the signal into the cytosol (Madsen et al., 2011; Broghammer et al., 2012).

Downstream of the Nod factor perception are a set of proteins that play a dual role in mycorrhizal and nodulation symbioses, which define the so-called common symbiosis pathway (Parniske 2008). These common symbiosis proteins include the leucine-rich repeat receptor kinase LjSYMRK (symbiotic receptor kinase) / MtDMI2 (Endre et al., 2002; Stracke et al., 2002), two nuclear-localized potassium channels LjCASTOR and LjPOLLUX /MtDMI1(Ané et al., 2004; Imaizumi-Anraku et al., 2005), three nucleoporins LjNUP85, LjNUP133 and LjNENA (Kanamori et al., 2006; Saito et al., 2007; Groth et al., 2010), one nuclear-localized protein kinase LjCCaMK / MtDMI3 (Lévy et al., 2004; Mitra et al., 2004; Tirichine et al., 2006), and its facilitator protein LjCYCLOPS / MtIPD3 (Messinese et al., 2007; Yano et al., 2008). LjSYMRK/ MtDMI2 encodes an extracellular, leucine-rich repeat receptor kinase that belongs to a large LRR-RLK gene family in higher plants which are involved in a range of biological signaling, such as pathogen recognition, hormone perception, and plant development (Afzal et al., 2008). It was reported that MtDMI2 forms a high-molecular weight complex and localizes to plasma membrane-associated puncta and cytoplasmic vesicles in hairy roots.
(Riely et al., 2012). However, the molecular and biochemical function of DMI2 has not been fully understood yet. LjCASTOR and LjPOLLUX/MtDMI1 are two potassium channels located on nuclear membrane and presumably function as the targets of the secondary messenger derived from the Nod factor perception signal (Oldroyd and Downie 2008). Potassium changes in these channels may alter the membrane polarity, resulting in activation the voltage-gated calcium channels and leading to calcium spiking. Three nuclear pore proteins, LjNUP85, LjNUP133 and LjNENA may provide an entrance for the secondary messenger from cytoplasm to nucleoplasm. Alternatively, they may be required for targeting LjCASTOR and LjPOLLUX/ MtDMI1 in inner nuclear membrane (Oldroyd and Downie 2008).

LjCCaMK/ MtDMI3 and its interactor LjCYCLOPS/ MtIPD3 function downstream of the Nod/Myc-factor-induced calcium spiking and are presumably responsible for decoding and transmitting the calcium spiking signal. CCaMK protein contains three characteristic domains: one CaM (calmodulin) domain, three EF-hand domains and one kinase domain. Depending on Ca$^{2+}$ concentration, CCaMK can bind calcium directly through its EF-hand domains or indirectly in a complex with CaM to trigger the initiation of nodule development (Oldroyd and Downie 2008). Recently, Miller et al. (2013) proposed a model for CCaMK to decode calcium oscillations through using differential calcium binding affinities. At the basal levels of Ca$^{2+}$ concentration, Ca$^{2+}$ bind at least one EF-hand domains and induce phosphorylation in kinase domain with resultant inactive state of CCaMK. The elevated concentrations provoked by calcium spiking, triggers CaM binding to CCaMK which overrides the negative regulation caused by autophosphorylation and activates the protein (Miller et al., 2013). CYCLOPS contains a
coiled-coil domain and acts as the phosphorylation substrate of CCaMK to activate downstream transcriptional factors. It has been proposed that phosphorylated CYCLOPS directly binds to the \textit{LjNIN} promoter and activates \textit{LjNIN} expression (Suzaki and Kawaguchi 2014). In addition to \textit{LjNIN}, several other symbiosis-specific transcription factors, including two GRAS family transcriptional factors \textit{MtNSP1} and \textit{MtNSP2}, and one ethylene response factor \textit{MtERN1}, are also activated downstream of \textit{LjCCaMK/MtDMI3} and \textit{LjCYCLOPS/MtIPD3} and responsible for transcriptional reprogramming of the host symbiotic genes (Schauser et al., 1999; Kaló et al., 2005; Smit et al., 2005; Andriankaja et al., 2007; Marsh et al., 2007; Middleton et al., 2007).

Not surprisingly, phytohormones are also involved in nodulation signaling and play important role in nodule development, especially cytokinin and auxin (Traut 1994; Desbrosses and Stougaard 2011). It was reported that NIN induces cytokinin signaling through upregulating cytokinin receptor \textit{MtCRE1} expression in the cortex cells, leading to blocking auxin transport by suppression the expression of auxin efflux carriers. Subsequently, the cell divisions are initiated and ultimately lead to spontaneous nodulation (Oldroyd et al., 2011).

**Host genes required for nitrogen fixation**

In contrast to nodulation signaling, less is known about how plant genes facilitate the nitrogen fixation process after nodule inception. Several late nodulin genes, which are induced around the onset of nitrogen fixation, have been shown to be essential for nitrogen fixation. Examples include a leghemoglobin that plays an important role in maintaining a low oxygen concentration in the nodule-infected cells while also
facilitating O$_2$ supply to the bacteroids for their aerobic respiration (Ott et al., 2005); a sucrose synthase that breaks down photosynthate to provide carbon energy to nodules (Baier et al., 2007); and a phosphoenolpyruvate carboxylase that is involved in the carbon and nitrogen flux in nodules (Nomura et al., 2006).

Identification and characterization of Nod+Fix- plant mutants has led to the cloning of several host genes that are essential for nitrogen fixation. These genes include $SSTI$, $FENI$ and $IGNI$ in *L. japonicas* and $DNFI$ in *M. truncatula*. $SSTI$ encodes a symbiotic sulfate transporter that is expressed exclusively in rhizobial-infected nodules (Krusell et al., 2005). Since sulfur is a component of metal-sulfur clusters within the nitrogenase complex, $SSTI$ possibly functions to provide the high demand for sulfur by bacteroids. The $FENI$ gene encodes a homocitrate synthase (HCS) which is also specifically expressed in nodules (Hakoyama et al., 2009). Homocitrate is a known component of iron-molybdenum cofactor in the nitrogenase complex (Hoover et al., 1987), but most rhizobial species do not possess the HCS-encoding genes. Thus, the nodule-specific HCS encoded in the host genome could compensate for the lack of such genes in rhizobia. This finding also highlighted a complementary and indispensable partnership between legumes and rhizobia for symbiotic nitrogen fixation. $IGNI$ encodes for an ankyrin-repeat membrane protein (Kumagai et al., 2007). It was hypothesized that $IGNI$ may be required for preventing the host cells from inappropriately invoking defense responses against compatible microsymbionts, but the exact biochemical function of $IGNI$ remains to be elucidated.

The differentiation of rhizobia into bacteroids is essential for nitrogen fixation. In *M. truncatula*, nodule-specific cysteine-rich (NCR) peptides have been shown to be required
for the differentiation of symbiotic rhizobia into terminal bacteroids (Van de Velde et al., 2010). The NCR peptides are targeted to the bacteroids through a nodule-specific protein secretory pathway, which requires the \textit{DNF1} gene that encodes a component of the signal peptidase complex (Wang et al., 2010). Another symbiotic gene \textit{DNF2}, encoding a putative phosphatidylinositol phospholipase C-like protein, was also shown to be required for bacteroid differentiation and nitrogen fixation in \textit{M. truncatula} (Bourcy et al., 2013).

\textbf{Host specificity in nodulation and nitrogen fixation}

The legume-rhizobial symbiosis is highly specific, such that each rhizobial strain establishes an efficient symbiosis with only a limited set of host plants, and vice versa. Such specificity can occur both at the early stages of the interaction associated with bacterial infection and nodule development as well as at the late stages that are related to nitrogen fixation. Symbiotic specificity has been a subject of intensive studies for several decades. From a basic science perspective, specificity in this system is strikingly similar to host-pathogen interactions. Are specificity determinants shared between antagonistic and friendly interactions? If so, how are opposite outcomes achieved? If not, then what are the different features that a host would recognize to distinguish friend from foe? From an applied aspect, understanding the molecular mechanisms underlying symbiotic specificity can lead to improved crop yield through better practice, without the need for substantial input. It has been documented that domesticated crop species tend to have fewer compatible symbionts (higher specificity) than their wild counterparts (Mutch and Young 2004). Such a constraint can lead to decreased yield in soils where the favorable strains are absent. On the other hand, even though many legumes can nodulate with
indigenous soil bacteria, nitrogen fixation efficiency varies tremendously between different host-rhizobial combinations (Schumpp and Deakin 2010). Knowledge of genetic control of symbiosis specificity will improve our ability to predict and manipulate key genetic factors controlling the symbiotic interaction and allow researchers to develop new crop varieties or engineer novel rhizobial strains that are able to enhance the agronomic potential of the root nodule symbiosis.

1. Host specificity in nodulation

The nodulation specificity is regulated by a fine-tuned signal exchange between the two prospective symbiotic partners (Perret et al., 2000, Fig. 3.2). Several molecular players from both the host and the microsymbiont function as determinants of nodulation specificity in the legume-rhizobia symbiosis. At the beginning, flavonoids-NodD interaction is considered as an early checkpoint of the symbiosis. After that, Nod factor, a bacterial derived signal, is widely thought to play a key role in defining the host range. In addition, rhizobia also use surface polysaccharides or secreted proteins to modulate host range. Moreover, plant immune system also is involved in determining symbiosis specificity (Wang et al., 2012).

Flavonoids are host plant secreted phenolic molecules, which can passively diffuse across the bacterial membrane (Oldroyd et al., 2011). Perception of flavonoid signals by rhizobia leads to activation of bacterial NodD proteins, which are transcriptional regulators belonging to the LysR family (Sharon 1996). Activated NodDs then bind to conserved DNA motifs, known as $nod$ boxes, found in the promoter regions of nodulation ($nod$) genes, to initiate the transcription of $nod$ genes (Fisher and Long 1993). NodDs
Figure 3.2 Host and bacterial signals that regulate host specificity in the legume-rhizobia symbiosis. (1) At the first stage, the plant-derived flavonoid signals (such as the luteolin shown here, from *M. truncatula*) are perceived by free-living soil bacteria, activating the rhizobial NodD proteins. NodD proteins bind to the conserved nod-box promoter regions of bacterial nodulation genes to activate their expression. (2) The *nod* genes encode the enzymes required for the synthesis of Nod factors. Secreted Nod factors are recognized on the cell surface by plant transmembrane Nod factor receptors in a strain- and ecotype-specific manner. Modifications on the Nod factor such as the length and degree of saturation of the acyl group determine host specificity. Nod factor perception leads to growth changes in the root hair to trap a small number of bacteria, which would give rise to the entire population colonizing the resulting nodule. (3) Besides Nod factors, rhizobia also use surface polysaccharides (such as EPS from *S. meliloti*, depicted) to modulate host range. The plant receptor(s) are unknown, but may resemble animal receptors for surface polysaccharides from bacterial pathogens. (4) In certain rhizobial strains, NodD also activates the expression of *TtsI*, a gene encoding a transcriptional factor that binds to highly conserved promoter elements, called tts boxes, upstream of operons encoding the type III secretion machinery and effectors. Recognition of these effector proteins by plant *R* genes limits the host range in a genotype-specific manner (Wang et al., 2012).
from different rhizobial species respond to different sets of flavonoids. Similarly, different NodD homologues from the same strain may be activated by different groups of flavonoids (Broughton et al., 2000; Peck et al., 2006). Given that each host plant secretes different types of flavonoids, the ability to initiate NodD-dependent expression of nod genes in response to a specific spectrum of flavonoids defines an early checkpoint for the legume-rhizobial symbiosis. For example, point mutations in nodD from Rhizobium leguminosarum bv. trifolii extended the host range as a consequence of the induction of nod gene expression by flavonoid inducers, which are normally inactive (McIver et al., 1989). Another example is that transferring the nodD1 from the broad-host-range Rhizobium sp. strain NGR234 to narrow-host-range Rhizobium leguminosarum bv. trifolii strain ANU843 enabled the engineered rhizobia to nodulate the non-host plant of ANU843, Parasponia (Bender et al., 1988). The role of NodD in determining host specificity was also supported by expressing nodD genes from different species of rhizobia in a strain of S. meliloti lacking endogenous nodD activity. It was revealed that the initiation of nod gene expression in response to discrete sets of flavonoids is dependent on the source of NodD (Peck et al., 2006).

Expression of rhizobial nod genes results in synthesis and secretion of a highly specific bacterial derived signal, called Nod factors. Nod factors are diffusible signaling molecules that can activate diverse developmental process in the host plants. With the exception of certain photosynthetic rhizobia that are able to nodulate their legume hosts without producing Nod factors (Giraud et al., 2007), Nod factors are indispensable for nodulation in most legumes and function as a determinant of host specificity. Nod factors consist of a chitin-like N-acetyl-D-glucosamine oligosaccharide backbone with a fatty
acyl chain at the non-reducing end (Fig. 3.3). The length of the backbone, the size and saturation of the fatty acyl side chain, as well as additional modifications presented at the chitin backbone, such as sulfation and acetylation, vary among Nod factors from different rhizobia and this structural variation plays an important role in defining the host range. (Lerouge et al., 1990). The nodABC genes are required for the synthesis of the core structure of Nod factors. Mutations in these genes completely suppress the ability of Nod factors to nodulate their host (Roche et al., 1996). Additional genes are involved in determining the specific decorations on the Nod factor core, and alterations in these genes often change host specificity. For example, inactivation of the nodE gene in R. leguminosarum bv. trifolii changes the identity of the fatty acyl chain attached to the Nod factor (Spaink et al., 1991), and this change severely affects the rhizobial symbiosis with Trifolium species while enhancing symbiosis with Pisum sativum and Vicia sativa (Djordjevic et al., 1985; Spaink et al., 1989). In contrast, introducing the nodEFGHPQ gene cluster from S. meliloti into R. leguminosarum changed the modifications on Nod

![Fatty acyl](image)

**Figure 3.3** The general structure of the Nod factors produced by rhizobia (n=0-3). The substituents R1–R9 indicate various structural modifications on the chitin backbone in different strains of rhizobia (Spaink 2000; Janczarek et al., 2015).
factor into the *S. meliloti* type, resulting in the engineered rhizobium being able to nodulate the *S. meliloti* host *Medicago sativa* (Debellé et al., 1988; Faucher et al., 1989).

Corresponding to the Nod factor structure, Nod factor receptors are a host determinant of symbiosis specificity. This has been demonstrated by genetic and molecular analyses in pea (*Pisum sativum*), soybean (*Glycine max*) and *L. japonicas*. The first example is the cultivar-strain interaction reported in pea. In wild pea variety ‘Afghanistan’, a single recessive gene, *sym2* controls the restriction of nodulation by European *Rhizobium* strains. The *sym2* allele interacts with a specific rhizobial gene, *nodX*, present in all strains that are compatible with Afghanistan (Davis et al., 1988). The *nodX*-encoded protein acetylates a Nod factor specifically to induce nodulation (Firmin et al., 1993; Geurts et al., 1997). Thus, the *sym2* allele can only recognize the Nod factors with a *NodX*-dependent acetylation at their reducing end, but not non-acetylated ones. This ‘gene-for-gene’ interaction is in line with the finding that *SYM2* (likely allelic to *SYM37*) is located in an orthologous region of *Lj-NFR1* and *Mt-LYK3* (Limpens et al., 2003; Zhukov et al., 2008). Another example is the naturally occurring *rj1* gene in soybean, which restricts nodulation by most strains of *Bradyrhizobium*. However, some strains that are unable to nodulate the *Rj1/Rj1* genotypes have the ability to nodulate the *rj1/rj1* plants (Devine et al., 1980). Recently, it was confirmed that *Rj1* is a soybean orthologue of *Lj-NFR1* (Indrasumunar et al., 2011). The role of Nod factor receptors in defining host range was also evidenced by the observation that co-expression of the *Lj-NFR1* and *Lj-NFR5* in *M. truncatula* enables nodulation of the transformants by *Mesorhizobium loti*, normally infecting *L. japonicus* (Radutoiu et al., 2007). Furthermore, the specificity for different
rhizobial symbionts of two different Lotus species is determined by a single amino acid variation in the LysM domain of Lj-NFR5 (Radutoiu et al., 2007).

Besides Nod factors, rhizobia also use surface polysaccharides, including exopolysaccharides (EPS), lipopolysaccharides (LPS), capsular polysaccharides (KPS) and cyclic β-glucans, to modulate host range (Deakin and Broughton 2009). Surface polysaccharides likely play a role in the evasion or suppression of defense response in compatible hosts, a feature that is shared by pathogenic and symbiotic bacteria (D'Haeze and Holsters 2004). Depending on the specific system, defects in surface polysaccharides may cause failures of rhizobia-legume symbiosis at either early or late stages. In the narrow-host-range rhizobia strain S. meliloti, defects in EPS production tend to result in arrests of microcolony or infection thread formation (Finan et al., 1985; Leigh et al., 1985). Particularly, the severity of the defects is correlated with the degree of alteration in the EPS structure (Cheng and Walker 1998). Similarly, an arrested infection threads formation phenotype was observed in alfalfa upon inoculation with the R. meliloti mutants which were unable to produce or export cyclic β-glucans (Dylan et al., 1986). Surface polysaccharides are specificity determinants was also supported by the observation that the strain-ecotype specificity in S. meliloti-M. truncatula symbiosis was related to succinoglycan oligosaccharide structure (Simsek et al., 2007).

In addition to surface polysaccharides, the proteins secreted by rhizobial type III secretion systems (T3SSs) also play an important role in determining symbiosis specificity (Wang et al., 2012). T3SSs are complex apparatus derived from bacteria that help them to invade the host through injecting effector proteins directly into the host cells (Coburn et al., 2007). Similar to the synthesis of Nod factors, flavonoids and the
bacterial transcription activator NodD also trigger the synthesis of rhizobial T3SS and its secreted effectors (Wassem et al., 2008). In this case, NodD activates the expression of TtsI, a gene encoding a transcriptional regulator that binds to highly conserved promoter elements, called tts boxes, upstream of operons encoding the T3SS machinery and its secreted effectors (Wassem et al., 2008). In contrast to T3SSs of pathogenic bacteria that they are essential for causing disease in susceptible hosts and eliciting the hypersensitive response in resistance hosts (Büttner and He 2009), rhizobial T3SSs are dispensable for rhizobial infection and nodulation. Some rhizobial effectors are homologous to those secreted by pathogenic bacteria, suggesting that they share a similar strategy with pathogens to invade the hosts (Dai et al., 2008; Kambara et al., 2009). Thus, the type III effector proteins may also be perceived by plant R genes to trigger effector-triggered immunity, limiting the host range in a genotype-specific manner (Sadowsky et al., 1990; Deakin and Broughton 2009; Soto et al., 2009). In a compatible interaction, the effectors may promote rhizobial infection and nodulation due to absence of the corresponding R genes for recognition in host plants. However, when they are perceived by the host immune system, the effectors function negatively to suppress nodulation. This hypothesis is supported by our recent study of R gene-controlled host specificity in legume-rhizobia symbiosis (Yang et al., 2010). Two plant genes, Rj2 and Rfg1 that restrict nodulation with specific strains of Bradyrhizobium japonicum and Sinorhizobium fredii, respectively, were cloned from soybean. It was demonstrated that Rj2 and Rfg1 are allelic genes encoding a member of TIR-NBS-LRR class of plant resistance proteins. In the incompatible interactions controlled by the Rj2 or Rfg1 alleles, the rhizobial infection process was completely blocked, which was presumably caused by host defense
responses triggered by the recognition of yet unknown rhizobial effectors. This is in line with the observation that a T3SS mutant of *S. fredii* USDA257 can restore nodulation with the soybean genotypes carrying the *Rfg1* alleles. Our research is consistent with reports describing rhizobial T3SSs and its secreted effectors that play important role in modulating the host range, and suggests that establishment of a successful root nodule symbiosis requires the evasion of plant immune responses triggered by rhizobial effectors.

2. Host specificity in nitrogen fixation and natural variation in symbiotic efficiency

Symbiotic specificity also exhibits at the nitrogen fixation phase, which is independent of the early signal exchange leading to rhizobial infection and nodulation. For example, a rhizobial strain that is able to nodulate a host genotype may fail to fix nitrogen in the resulting nodules (Nod+Fix-); however, the same host genotype or rhizobial strain are capable of establishing a successful nitrogen-fixing symbiosis (Nod+Fix+) with alternative symbiotic partners (Tirichine et al., 2000; Simsek et al., 2007). Moreover, nitrogen fixation efficiency varies tremendously in root nodules derived from different host-rhizobial combinations, and there is no single plant genotype or bacterial strain that is always associated with the greatest nitrogen fixation efficiency (Snyman and Strijdom 1980; Rangin et al., 2008; Schumpp and Deakin 2010). The presence of natural variations in nitrogen fixation efficiency offers an opportunity to optimize the legume-rhizobial mutualism for better symbiosis. Theoretically, this can be achieved by means of a coordinated selection of the most effective plant-rhizobia combinations (Rengel 2002; Schumpp and Deakin 2010). A successful example is the Brazilian soybean crop, where a selection of native varieties as well as indigenous rhizobia that are both competitive and efficiently fix nitrogen has resulted in high-yield production without an external nitrogen
source (Alves et al., 2003). However, simultaneous selection of both plant and bacterial genotypes represents a big challenge, given that we lack an in-depth understanding of the plant and rhizobial genes that regulate host-specific nitrogen fixation.

Both bacterial and plant genes are involved in regulation of host-specific nitrogen fixation and nitrogen fixation efficiency. Despite a lack of knowledge about the gene networks that govern the host specificity in the nitrogen fixation process, several mechanisms could be envisioned that may be involved in the control of symbiosis specificity in nitrogen fixation. First, beyond initial molecular dialogue leading to bacterial infection and nodule formation, the symbiotic partners likely undergo an additional round of molecular communications within mature nodules. This later signal exchange could contribute positively or negatively (e.g. triggering host defense-like responses) to the differentiation and/or persistence of bacteroids and symbiosomes, and accordingly affect the nitrogen fixation efficiency. Second, there may exist host genetic factors that regulate host-specific expression of bacterial genes associated with nitrogen fixation. Additionally, since bacteroids and symbiosomes have complete metabolic dependence on their hosts, the status of metabolic changes (e.g., amino acid cycling) across the peribacteroid membranes may play a critical role in regulation of the efficiency of nitrogen fixation. Research on elucidating the complexity of this important, but currently overlooked aspect of the legume-rhizobial symbiosis will help us to develop novel strategies to enhance the agronomic potential of biological nitrogen fixation.
Summary

Multiple checkpoints are employed during the process of the legume-rhizobial symbiosis to define the host range. In addition to exported Nod factors, surface polysaccharides and secreted proteins derived from bacteria, the host receptors to these molecules are also of great importance in host-rhizobial interactions. Changes of the structures or sequences of these factors usually alter the host specificity. Recently, a large number of bacterial secreted proteins have been catalogued (Deakin and Broughton 2009). However, the host targets and receptors for the symbiotic bacterial effectors remain largely unknown. Therefore, identification host genes controlling symbiotic specificity will contribute substantially to our knowledge of the nature of the interaction between the two symbiotic partners.

Advances in understanding specificity in symbiosis will likely be facilitated by the shift from inter-species studies to intra-species studies. Ecotypes showing differential responses to the same collection of rhizobia can be crossed to characterize the host genes that contribute to ecotype specificity. Identification of these host determinants will provide candidate genes for inter-species specificity, which can be investigated much in the same way as structure-function studies in Nod factor receptors (Radutoiu et al., 2007). Studying symbiosis specificity within species morphs into the realm of natural variation. Such specificity has been well documented in soybean (Glycine max) (Devine T. E. and Kuykendall 1996). Several dominant genes, such as *Rj2*, *Rj4*, and *Rfg1*, have been identified to restrict nodulation with specific rhizobial strains through genetic analysis of these naturally occurring variations (Caldwell 1966; Vest and Caldwell 1972; Trese 1995). Recently, we cloned the *Rj2* and *Rfg1* genes that restrict nodulation with specific
strains of *Bradyrhizobium japonicum* and *Sinorhizobium fredii*, respectively and demonstrated that the *Rj2* and *Rfg1* are allelic genes encoding a member of the TIR-NBS-LRR class of plant resistance (R) proteins (Yang et al., 2010). This study reveals a common recognition mechanism underlying symbiotic and pathogenic host-bacteria interactions. However, it is unclear that if this R gene-controlled host specificity is common in the legume-rhizobial symbiosis. To address this question, we worked to clone and characterize the *Rj4* gene that restricts nodulation by specific rhizobial strains of *B. elkanii* in soybean.
Chapter 4 Map-based cloning of \(Rj4\), a gene controlling nodulation specificity in soybean

Introduction

Legumes are able to make their own nitrogen fertilizer by forming a symbiotic relationship with nitrogen-fixing soil bacteria, collectively called rhizobia. The symbiosis results in the formation of a specialized plant organ, called the root nodule, within which the rhizobia acquire carbohydrates from the host while providing the host with fixed nitrogen. This biological process represents an efficient and sustainable nitrogen-fixing system because it uses solar energy through plant photosynthesis. On a global scale, legume-rhizobial symbiosis can fix an amount of nitrogen nearly equivalent to that produced by the chemical fertilizer industry (Brockwell et al., 1995).

The legume-rhizobial symbiosis begins with a cross-kingdom molecular dialogue. Starting in the rhizosphere, flavonoids released by the legume root trigger the expression of a set of bacterial genes, termed the nod genes, which results in the synthesis and secretion of a highly specific signal of bacterial origin, known as Nod factors or lipo-chitooligosaccharides (Geurts and Bisseling 2002). Recognition of Nod factors by the plant in turn activates a suite of host responses that ultimately lead to the development of rhizobia infected root nodules (Oldroyd 2013). Recent studies in the two model legumes \(M. \) truncatula \((\text{Mt})\) and \(Lotus japonicas \((\text{Lj})\) identified a number of host genes that are required for rhizobial infection and nodule development (Oldroyd et al., 2009; Madsen et al., 2010). Analysis of these genes has begun to reveal the nodulation signaling pathway that is universal in legumes. The Nod factor signal is perceived by the receptor-like
kinases with LysM domains in the extracellular region (e.g., Lj-NFR1/Mt-LYK3/Gm-Rj1 and Lj-NFR5/Mt-NFP) (Limpens et al., 2003; Madsen et al., 2003; Radutoiu et al., 2003; Arrighi et al., 2006; Smit et al., 2007; Indrasumunar et al., 2011; Broghammer et al., 2012). Downstream of the Nod factor perception is a set of proteins that play a dual role in mycorrhizal and nodulation symbioses, which define the so-called common symbiosis pathway (Parniske 2008). These common symbiosis proteins include the leucine-rich repeat receptor kinase Lj-SYMRK/Mt-DMI2 (Endre et al., 2002; Stracke et al., 2002), the two nuclear-localized potassium channels Lj-CASTOR and Lj-POLLUX/Mt-DMI1 (Ané et al., 2004; Imaizumi-Anraku et al., 2005), the three nucleoporins Lj-NUP85, Lj-NUP133 and Lj-NENA (Kanamori et al., 2006; Saito et al., 2007; Groth et al., 2010), the Ca\(^{2+}\)/calmodulin-dependent protein kinase Lj-CCaMK/Mt-DMI3 (Lévy et al., 2004; Mitra et al., 2004; Tirichine et al., 2006), and the Lj-CCaMK/Mt-DMI3 interacting protein Lj-CYCLOPS/Mt-IPD3 (Messinese et al., 2007; Yano et al., 2008). Among these common symbiosis genes, CCaMK and CYCLOPS function downstream of the Nod factor-induced calcium spiking and are presumably responsible for decoding and transmitting the calcium spiking signal (Okazaki et al., 2009). Perception of the calcium spiking signal activates several nodulation-specific transcription factors, such as Mt-NSP1 and Mt-NSP2, leading to transcriptional reprogramming of the host symbiotic genes (Schauser et al., 1999; Kaló et al., 2005; Smit et al., 2005; Middleton et al., 2007).

One remarkable property of the legume-rhizobial symbiosis is its high level of specificity (Broughton et al. 2000; Perret et al. 2000; Wang et al. 2012). Such specificity occurs at both inter- and intra-species levels and takes place at multiple phases of the interaction, ranging from initial bacterial infection and nodulation (nodulation specificity) to late
nodule development associated with nitrogen fixation (nitrogen fixation specificity). Symbiotic specificity has been a subject of intensive studies for several decades. From an applied perspective, understanding the molecular mechanisms controlling symbiotic specificity would allow researchers to develop new crop varieties or engineer novel rhizobial strains that are able to enhance the agronomic potential of the root nodule symbiosis. It has been reported that modern crops tend to have fewer compatible symbiotic partners in the soil than their wild relatives (Mutch and Young 2004). In this case, broadening the host range can lead to increased yields in soils where the favorable strains are lacking. On the other hand, even though many legumes can nodulate with indigenous soil bacteria, nitrogen fixation efficiency varies tremendously between different host-rhizobial combinations (Schumpp and Deakin 2010); in this situation, developing genetic mechanisms for excluding nodulation with low-efficient bacterial strains is desirable.

Genetic control of symbiosis specificity is complex, involving complex signal communications between the two symbiotic partners (Wang et al., 2012). Despite recent advances in our understanding of the nodulation signaling pathway by studies in the model legumes (Oldroyd 2013), we know little about the genetic mechanisms that regulate symbiosis specificity because natural variation in these systems has not been thoroughly surveyed. In contrast, naturally occurring variation in strain-specific nodulation has been well-documented in soybean (Glycine max L.) (Devine T. E. and Kuykendall 1996). Several dominant genes, such as Rj2, Rj4, and Rfg1, have been identified that restrict nodulation with specific rhizobial strains (Caldwell 1966; Trese 1995; Vest and Caldwell 1972). Specificity in this system strikingly resembles gene-for-
gene resistance in plant-pathogen interactions (Sadowsky et al. 1990). We recently cloned the \textit{Rj2} and \textit{Rfg1} genes that restrict nodulation with specific strains of \textit{Bradyrhizobium japonicum} and \textit{Sinorhizobium fredii}, respectively (Yang et al. 2010). It turned out that \textit{Rj2} and \textit{Rfg1} are allelic genes encoding a member of the Toll-interleukin-1 receptor/nucleotide-binding site/leucine-rich repeat (TIR-NBS-LRR) class of plant resistance (R) proteins. This study reveals a common recognition mechanism underlying symbiotic and pathogenic host-bacteria interactions. Our finding is consistent with the discoveries that these bacterial strains possess a type III secretion system (T3SS) to deliver effectors into the host cells, which are presumably recognized by the host \textit{R} genes (Krishnan et al. 2003; Deakin and Broughton 2009; Tsukui et al. 2013).

In contrast to most \textit{R} genes, which function against specific pathogen isolates, a single \textit{Rj2} or \textit{Rfg1} allele could restrict nodulation by many distantly related rhizobial strains (Devine and Kuykendall 1996). A tantalizing question is why legumes evolved \textit{R} genes to prevent beneficial symbioses. One explanation is that some rhizobia deliver “virulence” effectors into the host cells to facilitate their infection; however, these effectors are homologous to those secreted by pathogenic bacteria, resulting in being recognized as pathogens (Yang et al., 2010). Alternatively, host plants may have evolved \textit{R} genes to selectively exclude certain rhizobial strains. In this latter case, the nodulation-restrictive \textit{R} genes could be used to prevent nodulation with those indigenous strains that are highly competitive but with very low nitrogen-fixing efficiency so that a host can selectively interact with rhizobial inoculants with high nitrogen-fixing efficiency (Devine and Kuykendall 1996; Yang et al. 2010).
It remains unclear if \( R \)-gene-controlled host specificity is common in the legume-rhizobial symbiosis. Addressing this question, we are working to clone additional dominant genes that restrict nodulation with specific rhizobial strains in soybean and \textit{Medicago truncatula}. In this study, we describe fine mapping of the soybean \textit{Rj4} locus and identification of the candidate genes. The \textit{Rj4} gene was identified more than 40 years ago (Vest and Caldwell 1972) and has been the subject of extensive study in the 1980s and 1990s (e.g., Devine and O’Neill 1986; Devine et al. 1990; Sadowsky and Cregan 1992). It restricts the host plant from nodulation with many strains of \textit{Bradyrhizobium japonicum} and \textit{Bradyrhizobium elkanii} (Sadowsky and Cregan 1992). \textit{B. elkanii} strains are frequently present in soils of the southeastern US and readily nodulate most soybean cultivars. However, \textit{B. elkanii} strains are considered as poor symbiotic partners of soybeans, and many of the strains also produce rhizobitoxine, a compound that induces chlorosis in the host plant (Devine and Kuykendall 1996). Thus, cultivars containing an \textit{Rj4} allele are favorable in soils where \textit{B. elkanii} strains are common. The \textit{Rj4} allele occurs with high frequency in \textit{Glycine soja}, the wild progenitor of soybean, and in soybean cultivars from Southeast Asia (>60%). In contrast, the \textit{Rj4} genotypes are less frequent in cultivars from North America (Devine and Breithaupt 1981). In this study, we delimited the \textit{Rj4} locus within a 47-kb region on soybean chromosome 1 and identified two candidate genes. The data reported here facilitates the development of genetic markers for marker-assisted selection in soybean.
Materials and methods

Plant materials and nodulation assay

An F2 mapping population was derived from the cross between the two soybean cultivars Hill (Rj4/Rj4) and Williams (rj4/rj4). Seedlings of parents and the segregating population were grown in sterile vermiculite in a growth chamber programmed for 16h light at 26°C and 8h dark at 23°C. Roots of 1-week-old seedlings were inoculated with B. elkanii USDA61, obtained from the National Rhizobium Germplasm Collection (US Department of Agriculture-Agriculture Research Service, Beltsville, MD, USA). The strain was first cultured on YEM agar plates (10g/L Mannitol, 0.4g/L Yeast extract, 0.2g/L MgSO₄·7H₂O, 0.1g/L NaCl, 0.5g/L K₂HPO₄, pH 6.9 with HCl, 10g/L agar) in the dark at 28°C for 5-7 days. And the bacterial paste was then collected from petri dish plates and diluted in sterile water. For nodulation assay, each seedling was flood-inoculated with 10ml of the bacterial suspension (optical density at 600 nm=0.1). Nodulation phenotypes were recorded 2-3 weeks after inoculation. The plants were scored as either nodulation or non-nodulation.

DNA isolation, marker development, and PCR amplification

Leaf DNA was extracted from 100mg fresh leaf tissue with 2×CTAB buffer (2% CTAB; 1.4M NaCl; 100mM pH 8.0 Tris-HCl; 20mM pH 8.0 EDTA) and purified with phenol/chloroform/isoamyl alcohol (25:24:1 mixture, pH 5.2±0.3) (Stewart C. Neal and Via 1993).

We first mapped SSR (simple sequence repeat) markers with known genetic position to localize the approximate position of Rj4. Additional markers were then developed based
on the genomic sequence of soybean genotype Williams 82 (rj4/rj4) surrounding the Rj4 locus. Markers were based on SNPs (single nucleotide polymorphisms) identified between the two parents. For this purpose, primers were designed for PCR amplification of genomic DNA from the two parents of the F2 mapping population, followed by sequencing the two PCR products to identify the sequence polymorphisms. Where possible, the SNPs markers were converted to CAPS (cleaved amplified polymorphic sequences) markers for genotyping, as described elsewhere in this thesis. Otherwise, they were genotyped by direct sequencing. The primers were designed using online software Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/primer3/). All markers described in this study are listed in Table 4.1.

For PCR amplification, a 10µl total volume PCR reaction contained 20ng DNA template, 1× PCR reaction buffer, 2.0 mM MgCl₂, 0.25 mM of dNTPs, 0.25 mM of each primer and 0.5 unit of DNA polymerase (New England Biolabs). PCR was performed with a 2 min initial denaturation step at 95°C, followed by 35 cycles of 94°C for 30 seconds, 55°C (adjusted with different primers) for 30 seconds, and 72°C for 1 min, and then followed by single final extension of 5 min at 72°C. For CAPS marker, 3µl PCR product was digested with corresponding restriction enzymes. Each digestion reaction contained 1 unit of the restriction enzyme, 1× compatible reaction buffer, and additional ddH₂O to the total volume of 10µl. The enzyme digestions were incubated in the PCR machine at the suitable temperature for at least 2 hours. Digestion products were applied on the agarose gel of appropriate percentage and recorded for the respective homozygous parental and heterozygous genotypes. Only nodulated plants (homozygous recessive alleles) were used for genetic mapping.
<table>
<thead>
<tr>
<th>Marker name</th>
<th>Marker type</th>
<th>Restriction enzyme</th>
<th>Hill restriction fragment pattern of CAPS</th>
<th>Williams restriction fragment pattern of CAPS</th>
<th>Forward primer</th>
<th>Reverse Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sat_036</td>
<td>SSR</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>GCGACTCCAAGTT</td>
<td>GCGGGAGTTAGAG</td>
</tr>
<tr>
<td>SNP&lt;sup&gt;49,280kb&lt;/sup&gt;</td>
<td>SNP</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>GCTAAACCGTT</td>
<td>GAAAGAACA</td>
</tr>
<tr>
<td>CAP&lt;sup&gt;49,320kb&lt;/sup&gt;</td>
<td>CAPS</td>
<td>BspHI</td>
<td>448</td>
<td>245+203</td>
<td>TCAATGCCTCAA</td>
<td>GCCATTAGTTAGGC</td>
</tr>
<tr>
<td>SNP&lt;sup&gt;49,370kb&lt;/sup&gt;</td>
<td>SNP</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>CCAAATAA</td>
<td>TCCGATGAAATCT</td>
</tr>
<tr>
<td>SNP&lt;sup&gt;49,399kb&lt;/sup&gt;</td>
<td>SNP</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>GCTAAACCGTT</td>
<td>GACAAC</td>
</tr>
<tr>
<td>SNP&lt;sup&gt;49,405kb&lt;/sup&gt;</td>
<td>SNP</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>TCAATGCCTCAA</td>
<td>ACCATTTCGACC</td>
</tr>
<tr>
<td>CAP&lt;sup&gt;49,415kb&lt;/sup&gt;</td>
<td>CAPS</td>
<td>HphI</td>
<td>429+185</td>
<td>429+76</td>
<td>ACAAAGCCGAGTT</td>
<td>ACAGTTA</td>
</tr>
<tr>
<td>SNP&lt;sup&gt;49,454kb&lt;/sup&gt;</td>
<td>SNP</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>ACAAAGCCGAGTT</td>
<td>TACCATCTACCCAT</td>
</tr>
<tr>
<td>CAP&lt;sup&gt;49,462kb&lt;/sup&gt;</td>
<td>CAPS</td>
<td>Hinfl</td>
<td>171+211</td>
<td>171+421</td>
<td>TGGCTTTTAAAT</td>
<td>TTTTGAG</td>
</tr>
<tr>
<td>SNP&lt;sup&gt;49,471kb&lt;/sup&gt;</td>
<td>SNP</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>GGGGAGGATAAAC</td>
<td>GACATTG</td>
</tr>
<tr>
<td>SNP&lt;sup&gt;49,500kb&lt;/sup&gt;</td>
<td>SNP</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>CATTTGGGTTCAT</td>
<td>CAGATAAAGCCAGA</td>
</tr>
<tr>
<td>CAP&lt;sup&gt;49,640kb&lt;/sup&gt;</td>
<td>CAPS</td>
<td>Hpy188I</td>
<td>269+108</td>
<td>377+144</td>
<td>ACCAACCGTGAAGA</td>
<td>CGAGAGGGAACT</td>
</tr>
<tr>
<td>CAP&lt;sup&gt;49,680kb&lt;/sup&gt;</td>
<td>CAPS</td>
<td>Apol</td>
<td>90+254</td>
<td>90+396</td>
<td>GAAAGGCAGAAGTT</td>
<td>CTTGAGAG</td>
</tr>
<tr>
<td>CAP&lt;sup&gt;50,190kb&lt;/sup&gt;</td>
<td>CAPS</td>
<td>BsmBI</td>
<td>172+465</td>
<td>637</td>
<td>TCAACATGGTGGC</td>
<td>GACGTC</td>
</tr>
<tr>
<td>CAP&lt;sup&gt;50,810kb&lt;/sup&gt;</td>
<td>CAPS</td>
<td>SspI</td>
<td>254+407</td>
<td>661</td>
<td>AACAAATCCTCG</td>
<td>TCTTGA</td>
</tr>
<tr>
<td>CAP&lt;sup&gt;50,940kb&lt;/sup&gt;</td>
<td>CAPS</td>
<td>TaqI</td>
<td>36+303</td>
<td>36+412</td>
<td>TTTTGCGCTGCCCA</td>
<td>TGTGGTAAG</td>
</tr>
<tr>
<td>Sat_414</td>
<td>SSR</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>GCGATTGCTTGTAG</td>
<td>GCGAAGGAAAGACT</td>
</tr>
</tbody>
</table>

<sup>* SNP and CAP markers were named based on their approximate chromosomal locations. SSR markers were based on [http://www.soybase.org](http://www.soybase.org)</sup>
Sequence analysis

Taking the advantage of the availability of the genomic sequences of the G. max Williams82 (http://www.soybase.org/) and the BAC-end sequences of G. soja PI468916, we identified two BAC clones, GSS_Ba124A02 and GSS_Ba201P23, containing the flanked region for Rj4 locus by BLAST analysis. Sequencing of these two BAC clones was carried out at the Advanced Genetic Technologies Center, College of Agriculture, University of Kentucky. The gene prediction was performed using online program FGENESH (Larkin et al., 2007). Functional domains were predicted using online tool Pfam 27.0 (http://pfam.xfam.org/) (Finn et al., 2014) with an initial E-value cutoff of 0.1. Sequence alignments were performed using ClustalX2 (Larkin et al., 2007).

Association mapping

We obtained 48 soybean genotypes, including 40 G. max genotypes and 8 G. soja genotypes, from the USDA Soybean Germplasm Collection. These genotypes represent a wide range of genetic diversity inferred from their geographic origin and were previously tested for the presence or absence of the Rj4 allele. The 48 soybean genotypes are listed in Table 4.2.

We phenotyped all these lines to confirm their nodulation phenotype by inoculation with B. elkanii USDA61. For genotyping, leaf DNA was extracted from the 48 soybean lines, and followed by PCR amplification. The primer pair is 5’- CATCGGTGGAAGTTTAGGAACAACAT-3’ (forward) and 5’- TAACAAAAGCACGGAGGGAAATGTTGC-3’ (reverse). 3µl PCR product was digested with PstI, followed by electrophoresis on 1.5 % agarose gel for genotyping analysis.
<table>
<thead>
<tr>
<th>Order</th>
<th>PI #</th>
<th>Species</th>
<th>Origin</th>
<th>Phenotype</th>
<th>Order</th>
<th>PI #</th>
<th>Species</th>
<th>Origin</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PI 548654</td>
<td><em>G. max</em></td>
<td>USA</td>
<td>Nod-</td>
<td>25</td>
<td>PI 157469</td>
<td><em>G. max</em></td>
<td>Japan</td>
<td>Nod+</td>
</tr>
<tr>
<td>2</td>
<td>PI 548631</td>
<td><em>G. max</em></td>
<td>USA</td>
<td>Nod+</td>
<td>26</td>
<td>PI 209340</td>
<td><em>G. max</em></td>
<td>Japan</td>
<td>Nod+</td>
</tr>
<tr>
<td>3</td>
<td>PI 69533</td>
<td><em>G. max</em></td>
<td>China</td>
<td>Nod-</td>
<td>27</td>
<td>PI 200522</td>
<td><em>G. max</em></td>
<td>Japan</td>
<td>Nod+</td>
</tr>
<tr>
<td>4</td>
<td>PI 70241</td>
<td><em>G. max</em></td>
<td>China</td>
<td>Nod-</td>
<td>28</td>
<td>PI 548666</td>
<td><em>G. max</em></td>
<td>USA</td>
<td>Nod+</td>
</tr>
<tr>
<td>5</td>
<td>PI 80461</td>
<td><em>G. max</em></td>
<td>Japan</td>
<td>Nod-</td>
<td>29</td>
<td>PI 548559</td>
<td><em>G. max</em></td>
<td>USA</td>
<td>Nod+</td>
</tr>
<tr>
<td>6</td>
<td>PI 82246</td>
<td><em>G. max</em></td>
<td>Korea</td>
<td>Nod-</td>
<td>30</td>
<td>PI 540556</td>
<td><em>G. max</em></td>
<td>USA</td>
<td>Nod+</td>
</tr>
<tr>
<td>7</td>
<td>PI 84668</td>
<td><em>G. max</em></td>
<td>Korea</td>
<td>Nod-</td>
<td>31</td>
<td>PI 553039</td>
<td><em>G. max</em></td>
<td>USA</td>
<td>Nod+</td>
</tr>
<tr>
<td>8</td>
<td>PI 85666</td>
<td><em>G. max</em></td>
<td>Japan</td>
<td>Nod-</td>
<td>32</td>
<td>PI 165673</td>
<td><em>G. max</em></td>
<td>China</td>
<td>Nod+</td>
</tr>
<tr>
<td>9</td>
<td>PI 86026</td>
<td><em>G. max</em></td>
<td>Japan</td>
<td>Nod-</td>
<td>33</td>
<td>PI 60272</td>
<td><em>G. max</em></td>
<td>China</td>
<td>Nod+</td>
</tr>
<tr>
<td>10</td>
<td>PI 86454</td>
<td><em>G. max</em></td>
<td>Japan</td>
<td>Nod-</td>
<td>34</td>
<td>PI 54818</td>
<td><em>G. max</em></td>
<td>China</td>
<td>Nod-</td>
</tr>
<tr>
<td>11</td>
<td>PI 91729</td>
<td><em>G. max</em></td>
<td>Korea</td>
<td>Nod-</td>
<td>35</td>
<td>PI 548489</td>
<td><em>G. max</em></td>
<td>China</td>
<td>Nod+</td>
</tr>
<tr>
<td>12</td>
<td>PI 227327</td>
<td><em>G. max</em></td>
<td>Japan</td>
<td>Nod-</td>
<td>36</td>
<td>PI 548480</td>
<td><em>G. max</em></td>
<td>China</td>
<td>Nod-</td>
</tr>
<tr>
<td>13</td>
<td>PI 291312</td>
<td><em>G. max</em></td>
<td>China</td>
<td>Nod-</td>
<td>37</td>
<td>PI 71570</td>
<td><em>G. max</em></td>
<td>China</td>
<td>Nod-</td>
</tr>
<tr>
<td>14</td>
<td>PI 71558</td>
<td><em>G. max</em></td>
<td>China</td>
<td>Nod-</td>
<td>38</td>
<td>PI 86078</td>
<td><em>G. max</em></td>
<td>Japan</td>
<td>Nod+</td>
</tr>
<tr>
<td>15</td>
<td>PI 192869</td>
<td><em>G. max</em></td>
<td>Indonesia</td>
<td>Nod-</td>
<td>39</td>
<td>PI 548472</td>
<td><em>G. max</em></td>
<td>China</td>
<td>Nod+</td>
</tr>
<tr>
<td>16</td>
<td>PI 331793</td>
<td><em>G. max</em></td>
<td>Vietnam</td>
<td>Nod-</td>
<td>40</td>
<td>PI 518671</td>
<td><em>G. max</em></td>
<td>USA</td>
<td>Nod+</td>
</tr>
<tr>
<td>17</td>
<td>PI 379623</td>
<td><em>G. max</em></td>
<td>Japan</td>
<td>Nod-</td>
<td>41</td>
<td>PI 479744</td>
<td><em>G. soja</em></td>
<td>China</td>
<td>Nod-</td>
</tr>
<tr>
<td>18</td>
<td>PI 181699</td>
<td><em>G. max</em></td>
<td>Suriname</td>
<td>Nod-</td>
<td>42</td>
<td>PI 468916</td>
<td><em>G. soja</em></td>
<td>China</td>
<td>Nod-</td>
</tr>
<tr>
<td>19</td>
<td>PI 205910</td>
<td><em>G. max</em></td>
<td>Thailand</td>
<td>Nod-</td>
<td>43</td>
<td>PI 407025</td>
<td><em>G. soja</em></td>
<td>Japan</td>
<td>Nod-</td>
</tr>
<tr>
<td>20</td>
<td>PI 281898</td>
<td><em>G. max</em></td>
<td>Malaysia</td>
<td>Nod-</td>
<td>44</td>
<td>PI 101404</td>
<td><em>G. soja</em></td>
<td>China</td>
<td>Nod-</td>
</tr>
<tr>
<td>21</td>
<td>PI 547895</td>
<td><em>G. max</em></td>
<td>USA</td>
<td>Nod-</td>
<td>45</td>
<td>PI 507785</td>
<td><em>G. soja</em></td>
<td>Russian</td>
<td>Nod+</td>
</tr>
<tr>
<td>22</td>
<td>PI 518668</td>
<td><em>G. max</em></td>
<td>USA</td>
<td>Nod-</td>
<td>46</td>
<td>PI 366121</td>
<td><em>G. soja</em></td>
<td>Japan</td>
<td>Nod+</td>
</tr>
<tr>
<td>23</td>
<td>PI 230974</td>
<td><em>G. max</em></td>
<td>Japan</td>
<td>Nod+</td>
<td>47</td>
<td>PI 245331</td>
<td><em>G. soja</em></td>
<td>Taiwan</td>
<td>Nod-</td>
</tr>
<tr>
<td>24</td>
<td>PI 548447</td>
<td><em>G. max</em></td>
<td>China</td>
<td>Nod-</td>
<td>48</td>
<td>PI 366122</td>
<td><em>G. soja</em></td>
<td>Japan</td>
<td>Nod-</td>
</tr>
</tbody>
</table>

Nod+ means the plant can nodulate with USDA61, Nod- indicates the plant cannot nodulate with the strain USDA61.
Rapid amplification of cDNA ends (RACE)

The full-length cDNA of candidate genes of *Rj4* was determined with rapid amplification of cDNA ends (RACE). One-week-old seedlings of Hill were inoculated with *B. elkanii* USDA61. Three days post inoculation, ~100mg root sample was collected for RNA extraction. RNA was isolated using the Plant RNAeasy miniprep Kit (QIAGEN). Two micrograms of RNA was used to perform reverse transcription reaction using Superscript™ II reverse transcriptase (Invitrogen) in a 20-μl reaction mixture. The 5’ and 3’ ends of the cDNAs were amplified with Smart Race cDNA Kit (Clontech) (Zhu et al., 2001). Final amplification product was obtained by performing two rounds of nested PCR followed by the primary PCR reaction. After the three rounds of PCR, the final PCR product was applied on the 1.5% agarose gel for analyze. After gel extraction with QIAquick Gel Extraction Kit (QIAGEN), the purified product was cloned into pGEM®-T easy vector (Promega) for sequencing. All the gene specific primers used in RACE were listed in Table 4.3.

**Table 4.3** Gene specific primers used in RACE for PCR amplification.

<table>
<thead>
<tr>
<th>Cycle</th>
<th>5’RACE</th>
<th>3’RACE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>ACGATAAGTTCAAGTCCATCAGCAACC</td>
<td>GCCCCACCCGGTTACTCTGGTGAGTT</td>
</tr>
<tr>
<td>2nd</td>
<td>AACTTCACCAGAGTAACCGGTGGG</td>
<td>TGCCTGATGGAACCTTATGGACGTTGTC</td>
</tr>
<tr>
<td>3rd</td>
<td>GGTGACACATGAGAAGTTGCTGGTA</td>
<td>TACTATCATTCCTTGACCACAC</td>
</tr>
</tbody>
</table>
**Plasmid construction**

We used both cDNA and genomic constructs for validation of candidate genes of *Rj4*. We created cDNA constructs of Glyma01g37060 and Glyma01g37060-2 driven by the CaMV 35S promoter in pCAMBIA1305.1-100 which was derived from pCAMBIA1305.1 by induction of a Gateway® cassette. For this purpose, we amplified cDNAs of the two candidate genes from Hill by RT-PCR. The primer pairs used for PCR amplification were attached with introduced Gateway® cloning sites (cDNA060-F: 5’-AAAAAAGCAGGCTTCATGGCATCCATCGTGACTAT-3’; cDNA060-R: 5’-AAGAAAGCTGGGTCTTTAAACAGTGACATCTTGGA-3’; cDNA060-2-F: 5’-AAAAAAGCAGGCTTCATGGCATCCATCGTGACTAT-3’; cDNA060-2-R: 5’-AAGAAAGCTGGGTCTTTAAACAGTGACATCTTGGA-3’). Purified PCR products were ligated with the pENTER vector by using Gateway BP clonase (Invitrogen) as an entry clone. The target genes then were transferred into the destination vector pCAMBIA1305.1-100 by using Gateway LR clonase (Invitrogen). The individual constructs were named as 35S: Glyma01g37060 and 35S: Glyma01g37060-2, respectively.

We also developed genomic constructs of Glyma01g37060 and Glyma01g37060-2 under the control of their native promoters in pCAMBIA1305.1. Due to the complexity of soybean genome and numerous small repeat sequences around Glyma01g37060 and Glyma01g37060-2, we cannot successfully amplify these two genes from Hill. Given that there is no allelic polymorphism of Glyma01g37060 between Hill and *G. soja* PI468916 (*Rj4/Rj4*), we amplified Glyma01g37060 from PI468916 using the DNA of BAC clone GSS_Ba201P23 as template. The PCR product contained the ~1.0 kb coding region plus
~3.1kb upstream of the start codon and ~1.0 kb downstream of the stop codon. Taking
the advantage of the highly homologous sequence between Hill and PI468916, overlap
PCR was performed to amplify Glyma01g37060-2 using Hill and GSS_Ba201P23 DNA
as template. The overlap PCR product contained the ~2.7 kb upstream of the start codon
derived from BAC clone GSS_Ba201P23 and ~1.0 kb coding region plus ~1.9 kb
downstream of the stop codon derived from Hill. The PCR primers we used were as
follows: 060-F: 5’- CCTGCAGGCATGCAAGCTGGGGCAGGCTGTGATGAGAACA
TACTT -3’; 060-R: 5’- GACGGCCAGTGCCAAGCTAATGGGAATGTGGGCCCTTTTG -
3’; 060-2-F1: 5’- ATTACGAATTTCAGCTCGTTGTACACATGGAGACACAAAAACGAG
CA -3’; 060-2-R1: 5’- GGAAAAGGAAGGACGAAAAACTTCCC -3’; 060-2-F2: 5’-
GGGAAGTTTTTCGTCCTTCCTTTCC -3’; 060-2-R2: 5’- ACTCTAGAGGATCCCG
GGTTGCATGAGAAATCGGGAGGAATATG -3’. The genomic fragments obtained
above were ligated into the HindIII or KpnI digested pCAMBIA1305.1 vector by using
the In-Fusion Advantage PCR Cloning Kits (Clontech). Accordingly, we named the
individual constructs as G: Glyma01g37060 and G: Glyma01g37060-2, respectively. In
case that restriction of USDA61 infection requires presence of both genes, we also
constructed a vector, G: Glyma01g37060+ Glyma01g37060-2 containing both candidate
genes, using the In-Fusion HD cloning kit (Clontech). All vectors were introduced into
the Agrobacterium rhizogenes strain K599 and transformed to Williams (rj4/rj4) by hairy
root transformation as described below.

**Soybean hairy-root transformation**

We performed Agrobacterium rhizogenes-mediated hairy root transformation based on
the protocol described by Kereszt et al. (2007). The K599 strain carrying various
transgene constructs was cultured at 28°C overnight in liquid LB medium (10g/L tryptone, 5g/L yeast extract, 10g/L NaCl, 50mg/L kanamycin). The culture was centrifuged and bacterial paste was collected and injected into the cotyledonary node of 5-days-old seedlings of Williams with a thin needle. The infected seedlings were maintained in a growth chamber with 90% humidity and watered with Hoagland solution as nutrient source. The newly developed hairy roots from the infection sites were covered with wet vermiculite when they were approximately 1-3 cm in length. The main roots were removed when the hairy roots were well developed and long enough to support growth of the plant (normally 3-5 days after covering with vermiculite). The composite plants were moved to the new pots with sterile vermiculite, followed by inoculation with USDA61 three days after transplanting. Thereafter, the plants were watered with nitrogen-free nutrient solution to maintain a nitrogen-limit condition. 2-3 weeks after inoculation, nodulation of the transgenic roots were examined. Transgenic roots were identified through GUS staining.

**RNA isolation and analysis of gene expression by reverse-transcriptase (RT)-PCR**

For gene expression analysis, tissues of roots, leaves and stems from two parental plants were collected at 0 and 3 dpi with *B. elkanii* USDA61. Total RNA was isolated by the QIAGEN Plant RNAeasy miniprep Kit. Two micrograms of RNA was used to perform RT reactions by using M-MLV reverse transcriptase (Invitrogen) in a 20-μl reaction mixture. Two microliters of the RT reaction were used as template in a 20-μl PCR solution. The PCR primers were as follows: Gm-Actin, F: 5’-GAGCTATGAATTGCCTGATGG-3’ and R: 5’-CGTTTCATGAATTCCAGTAGC-3’; Glyma01g37060 specific, F1: 5’- TAAAATGGCATCCATCGTGA-3’ and R1: 5’-ATAGGCCACAACACAAGCAA
GAGCA-3’; Glyma01g37060-2 specific, F2: 5’-TACTTCTCAACCCTCAGC-3’ and R2: 5’-CCGCTTCCCTGGATATTCTTGATA-3’.

**T3SS mutant inoculation**

Type III secretion system (T3SS) mutants of *B. elkanii* USDA61, *rhcC2* and *rhcJ*, were kindly provided by Dr. Shin Okazaki (Department of International Environmental and Agricultural Science, Tokyo University of Agriculture and Technology, Tokyo, Japan). The mutant strains were grown on YEM selection plates (YEM with 50mg/L kanamycin) in the dark at 28°C for 5-7 days. The bacterial paste was then collected from petri dish plates and diluted in sterile water. 1-week-old seedlings of 40 *G. max* genotypes listed in Table 4.2 were flood-inoculated with the bacterial suspension (OD$_{600}$≈0.1). Nodulation phenotypes were recorded 3 weeks post inoculation.

**Results**

*Localization of the Rj4 locus*

For genetic mapping of the *Rj4* locus, we used an F2 population derived from the cross between the two soybean cultivars Hill (*Rj4/Rj4*) and Williams (*rj4/rj4*). The plants were inoculated with the *B. elkanii* strain USDA61. USDA61 can nodulate Williams but not Hill (Fig. 4.1), and this specificity is controlled by the *Rj4* gene (Vest and Caldwell 1972). Out of a total of 4,765 inoculated F2 plants, 1,159 plants nodulated, which fits the 3:1 (non-nodulation : nodulation) ratio ($\chi^2 = 1.11$, $df = 1$, $P = 0.29$), consistent with the restriction of nodulation by USDA61 being controlled by a single dominant gene.
The *Rj4* locus was previously mapped to a linkage group containing numerous RAPD (rapid amplification of polymorphic DNA) and AFLP (amplified fragment length polymorphism) markers plus a RFLP marker, pBLT017 derived from a sequenced cDNA clone- AW160139 (Ude et al., 1999; Matthews et al., 2001). However, this linkage group was not assigned to a soybean chromosome. Blast analysis against the soybean genomic sequence database (Schmutz et al., 2010; [http://www.phytozome.net/soybean.php](http://www.phytozome.net/soybean.php)) using AW160139 as a query sequence revealed two nearly identical homologs on soybean chromosome 1 and chromosome 9, respectively, suggesting that the *Rj4* gene is most likely located on one of the two chromosomes. Therefore, we selected known SSR (simple sequence repeat) markers on the two soybean chromosomes to map the *Rj4* locus. Initial mapping of 48 nodulated F2 individuals confirmed that *Rj4* is located on chromosome 1, within a genomic region defined by the flanking markers Sat_036 and Sat_414 that span ~2.3 Mb (Gm01: 49158693…51449420; Fig. 4.2A).

![Nodulation phenotypes of Williams (rj4/rj4) and Hill (Rj4/Rj4) by *B. elkanii* USDA61](image)

**Figure 4.1** Nodulation phenotypes of Williams (*rj4/rj4*) and Hill (*Rj4/Rj4*) by *B. elkanii* USDA61
Figure 4.2 Genetic mapping of the \( Rj4 \) locus. A. Fine mapping of the \( Rj4 \) locus. The \( Rj4 \) locus was delimited to a 47-kb genomic region between markers SNP\(^{49,415kb}\) and SNP\(^{49,462kb}\). Numbers indicate the number of recombination breakpoints separating the marker from \( Rj4 \) based on genotyping 1,159 homozygous \( rj4/rj4 \) segregant from the F2 population. B. Annotation of the 47-kb genomic DNA of Williams 82 \((rj4/rj4)\) identifies four putative genes. This region is conserved with three additional homologous regions located on soybean chromosomes 2, 11, and 16, respectively. Homologs are drawn in the same colors and connected with lines.
Fine mapping of the *Rj4* locus

Taking advantage of the availability of the genome sequence of soybean cultivar Williams82 (Schmutz et al., 2010), we developed high-density SNP markers for fine mapping of the *Rj4* locus. All the SNPs markers were named according to their approximate chromosomal locations on the reference genome of Williams82. SNPs were genotyped either by converting to CAPS markers or by direct sequencing. Genotyping a total of 1,159 F2 nodulated plants using the SNP markers allowed us to delimit the *Rj4* locus within a 47-kb genomic region defined by SNP^{49.415kb} and SNP^{49.462kb} (Fig. 4.2A). The 47-kb genomic sequence of Williams82 (*rj4/rj4*) contains four predicted genes (Glyma01g37040, Glyma01g37051, Glyma01g37060, and Glyma01g37080), which are listed in Table 4.4. Surprisingly none of these genes are homologous to typical plant *R* genes. Glyma01g37040 and Glyma01g37060 encode a thaumatin-like protein (TLP); Glyma01g37080 codes for an armadillo repeat-containing protein; and Glyma01g37051 represents a truncated NADPH flavin oxidoreductase.

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Gene annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glyma01g37040</td>
<td>Thaumatin-like protein</td>
</tr>
<tr>
<td>Glyma01g37051</td>
<td>Truncated NADPH flavin oxidoreductase</td>
</tr>
<tr>
<td>Glyma01g37060</td>
<td>Thaumatin-like protein</td>
</tr>
<tr>
<td>Glyma01g37080</td>
<td>Armadillo repeat-containing protein</td>
</tr>
</tbody>
</table>
Identification of candidate genes for *Rj4*

In order to identify the candidate genes of *Rj4*, we designed primers to amplify the corresponding predicted genes from Hill (*Rj4/Rj4*) and Williams (*rj4/rj4*), followed by sequencing the PCR products. Sequence analysis did not identify any non-synonymous nucleotide substitutions between the parental alleles of Glyma01g37040 and Glyma01g37080. Glyma01g37051 is also a pseudogene in the *Rj4* genotype Hill. However, we did identify amino-acid sequence polymorphisms for Glyma01g37060.

BLAST analysis of Williams82 identified three additional homologous regions on chromosomes 2, 11, and 16, respectively. These duplicated regions share highly conserved gene content, gene order, and transcriptional orientation (Fig. 4.2B). However, the Glyma01g37060 homologs are not present in these three chromosomal regions. These observations suggest that Glyma01g37060 is possibly a candidate gene for *Rj4*.

As shown in Fig. 4.3, a total of six amino acid substitutions and two amino acid insertions/deletions were identified between the two parental protein isoforms of Glyma01g37060. We then developed two SNP markers based on the DNA sequence polymorphisms that cause the amino-acid substitutions at positions 154 and 185 (Fig. 4.3) to conduct association mapping experiment. Genotyping of the 48 soybean genotypes listed in Table 4.2, including 40 *G. max* and 8 *G. soja*, revealed that the sequence substitutions were invariably associated with the nodulation phenotypes. One of the examples was shown in Fig. 4.4. This association analysis indicated that the same locus controls nodulation specificity in both *G. max* and *G. soja*, and further supported that
Glyma01g37060 is a candidate gene of $Rj4$. The molecular markers developed here have broad applicability for marker-assisted selection of the $Rj4$ allele in soybean breeding.

Figure 4.3 Alignment of amino acid sequences of Glyma01g37060 isoforms from Hill ($Rj4/Rj4$) and Williams ($rj4/rj4$). Five amino acid substitutions and two amino acid insertions/deletions are highlighted. The non-synonymous nucleotide substitutions at positions 154 and 185 were used to develop CAPS markers for association mapping in 48 soybean genotypes listed in Table 4.2.
Figure 4.4 Genotyping using the CAPS marker developed from Glyma01g37060 (corresponding to position 154 in Fig. 4.3) revealed that the sequence substitution was invariably associated with the nodulation phenotypes. The primers amplify a 1,379-bp product from Hill and a 1,395-bp product from Williams. The PCR product from Hill can be digested by PstI to produce 666-bp and 713-bp fragments, which were difficult to separate in the agarose gel, while the product from Williams cannot be digested by this restriction enzyme. The numbers represent the soybean genotypes listed in Table 4.2. ‘‘+’’ = Nod+; ‘‘-’’ = Nod-.
The candidate gene Glyma01g37060 is duplicated in the Rj4 genotypes

Even though our data appear to support Glyma01g37060 as a candidate gene of Rj4, there is a caveat for this inference. Since the genotype of the reference genome (Williams 82) is rj4/rj4, it is possible that rj4 gene represents a null allele in the reference genome. To address this concern, we identified and sequenced two G. soja BAC clones, GSS_Ba124A02 and GSS_Ba201P23, derived from the Rj4 genotype PI468916 that contains the orthologous region of the 47-kb Williams82 genomic region. Intriguingly, sequence analysis identified a 10.5kb insertion between the genomic region of Glyma01g37060 and Glyma01g37051 in the Rj4 genotypes Hill and PI468916. Gene prediction revealed that this insertion contained only one gene which was highly homologous to Glyma01g37060. Hereafter, this gene was referred to Glyma01g37060-2. The two tandem copies of Glyma01g37060 are separated by ~7-kb repetitive sequences. The copy number and sequence variation around this locus was most likely due to unequal crossover events. A total of thirteen amino acid substitutions were identified between the two protein isoforms of Glyma01g37060 and Glyma01g37060-2 in Hill (Fig. 4.5), majority of which are located on the N-terminal region. Based on these data, we hypothesize that either or both of the two duplicated genes may be required for the Rj4 function.
Figure 4.5 Alignment of amino acid sequences of Glyma01g37060 and Glyma01g37060-2 from Hill (Rj4/Rj4). Thirteen amino acid substitutions are highlighted.

The expression of Glyma01g37060 and Glyma01g37060-2 is root-specific and induced upon rhizobial inoculation

We performed gene-specific semi-quantitative RT-PCR to examine the gene expression pattern of Glyma01g37060 and Glyma01g37060-2. Root tissue of Hill and Williams was collected before and after inoculation, and RNA was isolated to conduct RT-PCR experiments. It was revealed that the expression of these two genes was root-specific (Fig. 4.6) and induced upon inoculation with USDA61 (Fig. 4.7). Although the overall expression level is very low, both genes were apparently up-regulated by rhizobia infection in Hill. However, no detectable transcript of Glyma01g37060 was observed in Williams with 35 PCR cycles either before or after rhizobia inoculation. The inducible
and root-specific expression pattern further supported the hypothesis that Glyma01g37060 and Glyma01g37060-2 are candidate genes for Rj4.

**Figure 4.6** The expression of Glyma01g37060 and Glyma01g37060-2 is root-specific in Hill (Rj4/Rj4). Roots, leaves and stems RNA isolated from Hill (Rj4/Rj4) and Williams (rj4/rj4) were analyzed for Glyma01g37060 and Glyma01g37060-2 expression by RT-PCR within 3 days post inoculation. The number in parentheses indicates the cycle number of the RT-PCR. The *G.max Actin* gene was used as a control.

**Figure 4.7** Semi-quantitative reverse-transcriptase (RT)-PCR analysis of the expression of (A) Glyma01g37060 and (B) Glyma01g37060-2 in Hill (Rj4/Rj4) and Williams (rj4/rj4). The expression of both genes is induced upon *B. elkanii* USDA61 inoculation in Hill. The number in parentheses indicates the cycle number of the RT-PCR. The *G.max Actin* gene was used as a control. dpi, days post inoculation.
Complementation tests failed to validate the candidate genes

To validate the candidate genes, we developed cDNA constructs of Glyma01g37060 and Glyma01g37060-2 driven by the CaMV 35S promoter and transferred them into the Williams background by *A. tumefaciens*-mediated hairy root transformation (Kereszt et al., 2007). Since these experiments were conducted without antibiotic selection, the resulting hairy roots contained both transgenic and wild type, which can be readily distinguished by examining the expression of the GUSPlus gene in the binary vector pCAMBIA 1305.1-100. In contrast to what we expected, all the transgenic hairy roots formed nodules. Comparable number of nodules was formed on the transgenic and wild type roots (Fig. 4.8). Meanwhile, the transgene expression was examined in the transgenic roots by RT-PCR using gene-specific primers. As shown on Fig. 4.9, both transgenes were expressed.

It is possible that both of the two candidate genes are required for the *Rj4* function. To validate this hypothesis, we also developed three genomic constructs containing single candidate or a combination of both driven by their native promoters, as described in Material and Methods, and transferred them into the Williams background. Unfortunately, none of these genomic constructs successfully blocked nodulation by USDA61 on the transgenic hairy roots.
**Figure 4.8** Composite transgenic plants transformed with the Glyma01g37060 or Glyma01g37060-2 constructs possess both transgenic (blue) and wild-type (white) roots. All the composite transgenic plants were inoculated with *B. elkanii* USDA61.

**Figure 4.9** RT-PCR analysis of transgene expression in Williams roots transformed with the Glyma01g37060 or Glyma01g37060-2 constructs driven by the CaMV-35S promoter. The number in parentheses indicates the cycle number of the RT-PCR. The *G.max Actin* gene was used as a control. CK, non-transgenic roots as negative control.
Rj4 function is dependent on the bacterial type III secretion system (T3SS)

Type III secretion system (T3SS) is an essential apparatus in pathogenic bacteria that help them to invade the hosts through injecting effector proteins into the host cells (Büttner and He 2009). Many, but not all, rhizobial strains also possess a T3SS to deliver effectors, so-called nodulation out proteins (Nops), into the host cells (Deakin and Broughton 2009). In contrast to bacterial pathogens, a rhizobial T3SS and its secreted effectors are not required for rhizobial infection and nodulation. However, they are determinants of host specificity in legume-rhizobia symbiosis. Previous study has revealed that a T3SS mutant of USDA 257, called DH4, gained the ability to nodulate the soybean genotypes carrying an \( Rfg1 \) allele and maintained the ability to nodulate the soybean genotypes carrying an \( rfg1 \) allele (Yang et al., 2010). The mutant strain can increase, decrease, or have no effect on nodule numbers in comparison with the wild-type strain depending on the genetic background. Similar results were also reported for other rhizobial strains such as \( Rhizobium \) sp. NGR234 (Krishnan et al., 2003). All these studies suggested that rhizobial T3SS possibly function as a facilitator superimposed on the Nod-factor signaling pathway. In the absence of recognition by the host \( R \) genes, the T3SS effectors may play a positive role in facilitating rhizobial infection, but function negatively if perceived by the host \( R \) genes.

To determine if T3SS is involved in host specificity controlled by \( Rj4 \), we examined the effects of two T3SS mutants of \( B. \) elkanii USDA61, \( rhcC2 \) and \( rhcJ \), on symbiotic properties of 40 \( G. \) max genotypes listed in Table 4.2. The \( rhcC2 \) and \( rhcJ \) mutants failed to secret effector proteins due to the disruption of transcriptional activator TtsI (Okazaki et al., 2009). Similar to the DH4 mutant of the \( S. \) fredii strain USDA257, the \( rhcC2 \) and
rhcJ mutants are able to nodulate the soybean genotypes that carry an Rj4 allele, suggesting that the function of Rj4 is also dependent on the bacterial T3SS and Rj4 is involved in a gene-to-gene interaction between host and its microsymbiont.

Discussion

In the present study, we finely mapped the Rj4 gene that controls nodulation specificity to B. elkanii in soybean. We delimited the Rj4 locus to a 47-kb genomic region flanked by the molecular markers SNP at 49.415kb and SNP at 49.462kb on soybean chromosome 1. A total of four predicted genes were found in the rj4 genomic region of Williams 82 (rj4/rj4). Sequence analysis only identified non-synonymous nucleotide substitutions between the two parental alleles of Glyma01g37060, which encodes a thaumatin-like protein (TLP). In addition, association mapping of 48 soybean genotypes revealed that the sequence substitutions are invariably associated with nodulation phenotypes, suggesting that Glyma01g37060 is most likely the Rj4 gene. However, sequencing two BAC clones derived from G. soja PI468916 (Rj4/Rj4) identified a duplicated gene of Glyma01g37060, named Glyma01g37060-2, in the Rj4 genotypes. The induced and root-specific expression pattern suggested that Glyma01g37060 and Glyma01g37060-2 are candidate genes for Rj4. Using four different mapping populations, Hayashi et al. (2014) also conducted map-based cloning of Rj4 and delimited this locus to a 53-kb genomic region that overlaps with the flanking region we identified. However, they only considered Glyma01g37060 as the candidate gene of Rj4. In addition, based on the real-time PCR results, they concluded that Glyma01g37060 is constitutively transcribed in roots and expressed in stems and leaves at low levels. The discrepancy may be caused by different protocols. The quantitative RT-PCR may have the sensitivity to detect the gene
expression at lower levels, but the contradictive patterns of gene expression derived from these two studies cannot be reconciled.

We developed several cDNA and genomic DNA constructs of Glyma01g37060 and Glyma01g37060-2 from Rj4/Rj4 genotype for complementation tests. Although transgenes expressed in transgenic hairy roots, they failed to block nodulation by USDA61. In contrast to our results, Hayashi et al. (2014) concluded that Glyma01g37060 was indeed Rj4 gene. The Rj4/Rj4 genotypes have been reported to exclude nodulation by B. elkanii USDA61 and B. japonicum Is-34 (Vest and Caldwell, 1972, Ishizuka et al. 1991). Hayashi et al. used Is-34-compatible Japanese cultivar Enrei (rj4/rj4) for validation of the Rj4 candidate. In their study, hairy roots of Enrei were transformed with Glyma01g37060 cDNA from Hill under the control of the ubiquitin promoter, followed by inoculation with Is-34. Because no reliable reporter gene was used as an indicator of transformation, they selected transformants by PCR with the Rj4 primers for the DNA samples prepared form the individual hairy roots. Based on the observation that fewer nodules were formed on transgenic roots, Hayashi et al. concluded that Glyma01g37060 functions as Rj4 to inhibit nodulation by Is-34.

We failed to complement the Rj4 phenotype with the candidate genes, while Hayashi et al. successfully validated Rj4 with Glyma01g37060. However, there were some pitfalls in their study. First, difference in nodule numbers between transgenic and wild type hairy roots during complementation usually is not a strong evidence for validation of a dominant gene. I participated in cloning and validation of Rj2 and Rfg1 in our lab using the same strategy. In that experiment, no or few nodules were produced in Rj2/Rfg1-trangenic roots, while the non-transgenic roots produced hundreds of nodules due to a
lack of auto-regulation of nodule numbers. Furthermore hairy roots vary dramatically in their size; larger roots normally form more nodules than the smaller ones. The nodule number per wild-type hairy root reported by Hayashi et al. ranged 10-15, which is questionable.

Given that our complementation tests were unsuccessful, we are using reverse genetics tools to test if the candidate genes are responsible for \textit{Rj4} function. As an evolutionary technique, CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9), with increased genetic editing efficiency, offers a fast and easy means to generate desirable gene-knockout mutants (Pennisi 2013; Segal 2013). This technology has been widely applied in various organisms, including bacteria, yeast, plants, animals and human cell lines (Xing et al., 2014). We have developed several CRISPR/Cas9 constructs containing one or two gRNAs to knock-out either single or both of the two candidate genes. All the vectors are being transferred to Hill with \textit{A. tumefaciens}-mediated hairy root transformation. Based on the phenotype of resulted mutants, we will determine if Glyma01g37060 and/or Glyma01g37060-2 are responsible for the \textit{Rj4}-mediated nodulation restriction.

Similar to the \textit{Rj2} and \textit{Rfg1} genes that encode plant R proteins, the function of \textit{Rj4} is dependent on the bacterial T3SS. Mutation of the T3SS of USDA61 enables it to nodulate soybean genotypes carrying an \textit{Rj4} allele. Another intriguing fact is that USDA61 can nodulate soybean plants defective in Nod-factor perception, and this nodulation ability relies on the T3SS of USDA61 (Okazaki et al., 2009; Okazaki et al., 2013). These data suggested that certain bacterial effectors secreted by T3SS of USDA61 can activate the nodulation signaling pathway without perception of Nod factors, while
others may trigger defense responses upon recognition by plant R genes resulting in blocking nodulation (Okazaki et al. 2013). It led to our expectation that Rj4 is most likely another plant R gene. Surprisingly, this appears to be unlikely the case. If we validate the candidate gene Glyma01g37060 and/or its duplicated copy as Rj4, it is going to be a big challenge for us to explain how a thaumatin-like protein is involved in perception of type III effectors and triggers “gene-for-gene” resistance against a rhizobial strain. Plant TLPs are classified as the pathogenesis related (PR) protein family 5 (PR-5), due to their inducible expression by pathogen attack (Petre et al., 2011). Overexpression of PR-5 genes has been shown to be able to enhance disease resistance against various pathogens. Several members of the plant TLP family have been reported as food allergens from fruit and pollen allergens from conifers. Some TLPs also hydrolyze beta-1,3-glucans of the type commonly found in fungal cell walls. However, Glyma01g37060 is a very unique gene: 1) it lacks a putative immunoglobulin E (IgE)-binding epitope (~30 amino acids long) that is universally present in all other TLPs, and we were unable to detect any TLP genes with similar deletion in other plants; 2) it does not present in other homeologous regions in the soybean genome; and 3) the orthologs of this gene appear to be present only in the syntenic regions of the sequenced tropical legumes pigeon-pea and common bean (data not shown). Thus, it is likely that this gene has been specifically evolved in the lineage of tropical legumes (Phaseoleae). Consistent with this prediction is that B. elkanii USDA61 is able to nodulate other tropical legumes and there exist genotypes that restrict nodulation with this strain (Okazaki et al. 2009). This raises the question whether the same gene restricts nodulation with this strain in different legume species. To answer this question, we plan to map the locus that restricts nodulation with USDA61 in Vigna.
*radiata* (mung bean) using an F2 population derived from the cultivars CN36 (Nod+) and KPS1 (Nod-) using the orthologous gene markers.
## Appendix

### Table A1 Abbreviations and acronyms

<table>
<thead>
<tr>
<th>Abbreviation/Acronym</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFLP</td>
<td>amplified fragment length polymorphism</td>
</tr>
<tr>
<td>Avr</td>
<td>Avirulence</td>
</tr>
<tr>
<td>CaM</td>
<td>calmodulin</td>
</tr>
<tr>
<td>CAPS</td>
<td>cleaved amplified polymorphic sequences</td>
</tr>
<tr>
<td>Cas9</td>
<td>CRISPR-associated protein 9</td>
</tr>
<tr>
<td>CC</td>
<td>coiled-coil</td>
</tr>
<tr>
<td>CDC5L</td>
<td>Cell division cycle 5 like protein</td>
</tr>
<tr>
<td>CPR1</td>
<td>Constitutive expresser of PR genes 1</td>
</tr>
<tr>
<td>CRISPR</td>
<td>clustered regularly interspaced short palindromic repeats</td>
</tr>
<tr>
<td>dpi</td>
<td>days post inoculation</td>
</tr>
<tr>
<td>EDS1</td>
<td>Enhanced disease susceptibility 1</td>
</tr>
<tr>
<td>EF-Tu</td>
<td>Elongation factor-Tu</td>
</tr>
<tr>
<td>EPS</td>
<td>exopolysaccharides</td>
</tr>
<tr>
<td>ETI</td>
<td>Effector-triggered immunity</td>
</tr>
<tr>
<td>HCS</td>
<td>homocitrate synthase</td>
</tr>
<tr>
<td>hnRNPs</td>
<td>heterogeneous nuclear ribonucleoproteins</td>
</tr>
<tr>
<td>HR</td>
<td>hypersensitive response</td>
</tr>
<tr>
<td>IgE</td>
<td>immunoglobulin E</td>
</tr>
<tr>
<td>KPS</td>
<td>capsular polysaccharides</td>
</tr>
<tr>
<td>LCOs</td>
<td>lipo-chitooligosaccharides</td>
</tr>
<tr>
<td>Lj</td>
<td><em>Lotus japonicas</em></td>
</tr>
<tr>
<td>LPS</td>
<td>lipopolysaccharides</td>
</tr>
<tr>
<td>LRR</td>
<td>leucine-rich repeat</td>
</tr>
<tr>
<td>LysM</td>
<td>lysin motif</td>
</tr>
<tr>
<td>MAC</td>
<td>MOS4-associated complex</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>MAMPs/ PAMPs</td>
<td>Microbe/pathogen-associated molecular patterns</td>
</tr>
<tr>
<td>MFs</td>
<td>Myc factors</td>
</tr>
<tr>
<td>MOS</td>
<td>Modifier of <em>snc1</em></td>
</tr>
<tr>
<td>Mt</td>
<td><em>M. truncatula</em></td>
</tr>
<tr>
<td>NBS</td>
<td>nucleotide-binding site</td>
</tr>
<tr>
<td>NCR</td>
<td>nodule-specific cysteine-rich</td>
</tr>
<tr>
<td>NDR1</td>
<td>Non-race-specific disease resistance 1</td>
</tr>
<tr>
<td>NFs</td>
<td>Nod factors</td>
</tr>
<tr>
<td>nod</td>
<td>nodulation</td>
</tr>
<tr>
<td>Nops</td>
<td>nodule-specific cysteine-rich protein</td>
</tr>
<tr>
<td>NPR1</td>
<td>Non-expressor of <em>PR</em> genes 1</td>
</tr>
<tr>
<td>PAD4</td>
<td>Phytoalexin deficient 4</td>
</tr>
<tr>
<td>PLRG1</td>
<td>Pleiotropic regulator 1</td>
</tr>
<tr>
<td>PR</td>
<td>Pathogenesis-related</td>
</tr>
<tr>
<td>PR-5</td>
<td>pathogenesis related protein family 5</td>
</tr>
<tr>
<td>Prp19</td>
<td>Precursor RNA processing 19</td>
</tr>
<tr>
<td>PRRs</td>
<td>Pattern recognition receptors</td>
</tr>
<tr>
<td>PTI</td>
<td>PAMP-triggered immunity</td>
</tr>
<tr>
<td>R</td>
<td>resistance</td>
</tr>
<tr>
<td>RACE</td>
<td>rapid amplification of cDNA ends</td>
</tr>
<tr>
<td>RAN1</td>
<td>Ras-related nuclear protein 1</td>
</tr>
<tr>
<td>RAPD</td>
<td>rapid amplification of polymorphic DNA</td>
</tr>
<tr>
<td>RBPs</td>
<td>RNA binding proteins</td>
</tr>
<tr>
<td>RNAi</td>
<td>RNA interference</td>
</tr>
<tr>
<td>RPS6</td>
<td>Resistance to <em>P. syringae</em> 6</td>
</tr>
<tr>
<td>RT</td>
<td>reverse-transcriptase</td>
</tr>
<tr>
<td>SA</td>
<td>salicylic acid</td>
</tr>
<tr>
<td>SCF</td>
<td>SKP1-Cullin1-F-box</td>
</tr>
<tr>
<td>SNC1</td>
<td>Suppressor of <em>npr1-1</em>, constitutive 1</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>SNPs</td>
<td>single nucleotide polymorphisms</td>
</tr>
<tr>
<td>snRNPs</td>
<td>small ribonucleoproteins</td>
</tr>
<tr>
<td>SR</td>
<td>serine/arginine-rich</td>
</tr>
<tr>
<td>SRFR1</td>
<td>Suppressor of rps4-RLD1</td>
</tr>
<tr>
<td>SS</td>
<td>splicing site</td>
</tr>
<tr>
<td>SSR</td>
<td>simple sequence repeat</td>
</tr>
<tr>
<td>SYMRK</td>
<td>symbiotic receptor kinase</td>
</tr>
<tr>
<td>T3SSs</td>
<td>type III secretion systems</td>
</tr>
<tr>
<td>TCV</td>
<td>Turnip crinkle virus</td>
</tr>
<tr>
<td>TIR</td>
<td>Toll-interleukin-1 receptor</td>
</tr>
<tr>
<td>TLP</td>
<td>thaumatin-like protein</td>
</tr>
<tr>
<td>TMV</td>
<td>tobacco mosaic virus</td>
</tr>
<tr>
<td>TN</td>
<td>TIR-NBS</td>
</tr>
<tr>
<td>Tr</td>
<td>truncated</td>
</tr>
<tr>
<td>UTR</td>
<td>untranslated regions</td>
</tr>
</tbody>
</table>
References


domain reveals interfaces for self-association, signaling, and autoregulation." Cell Host & Microbe 9(3): 200-211.


VITA
FANG TANG

Place of birth
Bohu County, Xinjiang Province, China

Education
Graduate student, Sep, 2005- July, 2008
Bio-engineering College,
Chongqing University, China

B.S., Sep, 2001- July, 2005
Bio-engineering College,
Chongqing University, China

Publications


