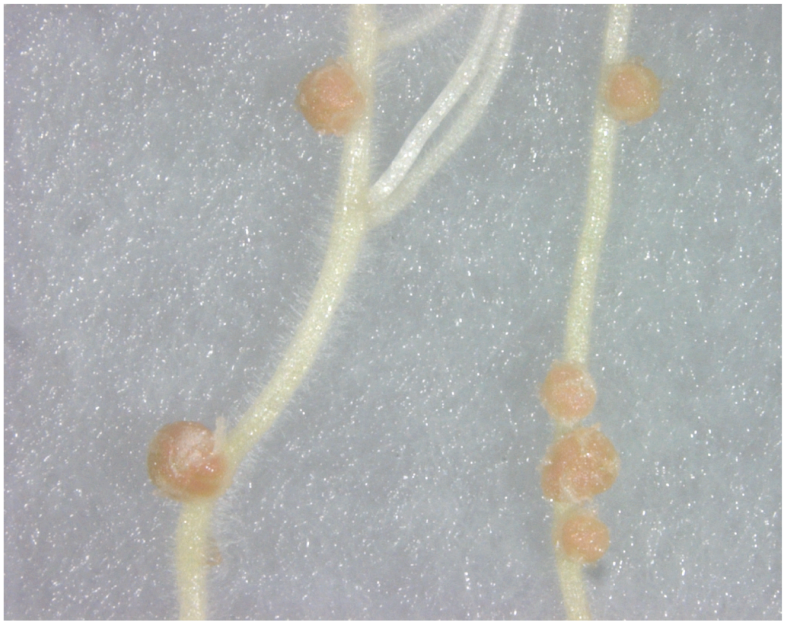
Novel and symbiosis-related microRNAs in *Lotus japonicus*



Ana de Luis Margarit

PhD dissertation

Université de Strasbourg
January 2010

UNIVERSITÉ DE STRASBOURG

Novel and symbiosis-related microRNAs in *Lotus japonicus*

Thèse présentée pour obtenir le grade de docteur de l'Université de Strasbourg

Discipline: Sciences du Vivant, Aspects Moléculaires de la Biologie

par

Ana de Luis Margarit

Soutenue publiquement le **29 janvier 2010** devant la Comission d'Examen:

Pr. Martin Crespi	**RAPPORTEUR EXTERNE**
Pr. Javier Paz-Ares	**RAPPORTEUR EXTERNE**
Dr. Thomas Ott	**EXAMINATEUR INVITÉ**
Pr. Mario Kéller	**RAPPORTEUR INTERNE**
Dr. Olivier Voinnet	**DIRECTEUR DE THÈSE**

Institut de Biologie Moléculaire des Plantes
UPR-CNRS 2357

When you set out on your journey to Ithaca,
then pray that the road is long,
full of adventure, full of knowledge.
K. Kaváfis

Dedicated to Guille

RÉSUMÉ DE THÈSE EN FRANÇAIS

Les légumineuses représentent un tiers de la production agricole mondiale et constituent les bases de l'agriculture durable, en raison de leur capacité d'assimilation de N_2 par l'établissement d'une symbiose avec les bactéries du genre *Rhizobium*. Cette interaction connue comme Fixation Symbiotique d'Azote (FSA) est un processus biologique complexe dans lequel la plante reconnaît spécifiquement les bactéries *Rhizobium*. Celles-ci pénètrent les poils racinaires, envahissent le cortex et induisent la formation de nouveaux organes dédiés à la Fixation Symbiotique d'Azote, les nodules. Ces structures nouvelles accueillent des formes différentiées des bactéries et un échange intense des nutriments commence entre les deux symbiontes.

La recherche génétique dans les légumineuses a beaucoup avancé ces dernières années grâce aux initiatives de séquençage de deux espèces modèles, *Medicago truncatula* et *Lotus japonicus*. L'étude parallèle de deux systèmes a permis des découvertes importantes pour la compréhension de la Fixation Symbiotique d'Azote et ses variantes, comme la nodulation déterminée et indéterminée. De plus, le séquençage génomique a rendu possible des analyses d'expression génique à grande échelle et, plus récemment, des analyses de petits ARNs.

Les petits ARNs (sRNAs) sont des molécules de 18-25nt qui régulent l'expression génique par similarité de séquence avec ses molécules d'ARN ou ADN cible, à travers des mécanismes connus sous le terme général de *RNA silencing*. Il y a différents types de petits ARNs décrits chez les eucaryotes, qui remplissent des taches régulatrices et activité biochimique différentes. Néanmoins, il existe un mécanisme général conservé pour la production et l'action de la plupart des petits ARNs dans la cellule. Les sRNAs sont produits à partir d'ARN double brin par des enzymes de type DICER, ayant une activité RNAse-III. Ces sRNAs sont ensuite chargés dans des complexes protéiques contenant une protéine de la famille ARGONAUTE. Ces deux facteurs clés du RNA silencing sont extrêmement conservés aussi bien dans le règne animal et végétal. Cependant une grande diversité dans le nombre de membres par famille est observée, ce qui reflète la complexification des voies du RNA silencing. Parmi les différents types de sRNAs, les microRNAs (miRNAs ou miRs) sont des régulateurs post-transcriptionnels spécifiquement dérivés des précurseurs d'ARN en tige-boucle, sous forme de duplex miR/miR*. Chez les plantes, les microRNAs jouent un rôle clé à chaque étape du développement, dans la régulation des métabolites et les interactions plante-microorganisme, y compris la symbiose.

Dans ce contexte scientifique, nous avons choisi la plante modèle *Lotus japonicus* pour étudier la diversité et la contribution des voies du RNA silencing dans les différentes facettes de la Fixation Symbiotique d'Azote. Premièrement, nous avons cherché à identifier les gènes principaux du RNA silencing chez *Lotus japonicus*, et principalement ceux impliqués dans la voie des microRNAs. Une analyse bioinformatique approfondi a révélé que le génome de *L. japonicus* contient 5 gènes *DICER* et 9 gènes *ARGONAUTE*. En même temps, nous avons aussi identifiés les gènes impliqués dans la voie canonique des miRNA chez *Arabidopsis*, tels que *LjDDL, LjHYL1, LjSE* ou *LjDSN*. L'analyse a été faite à partir des données génomiques disponibles chez *Lotus japonicus*, qui couvrent jusqu'à présent le 91,3% de l'espace génique. La complétion du séquençage pourrait permettre de découvrir de nouveaux membres de ces familles géniques. Enfin, l'analyse transcriptomique à grande-échelle sur chip Affymetrix a révélé une expression espécifique aux nodules des gènes *DCL2*, *AGO2* et *AGO5*, ce qui suggère que, en plus des miRNAs, d'autres classes de petits RNAs jouent un rôle dans ces organs symbiotiques.

Nous avons ensuite initié une stratégie basée sur la méthode TILLING pour obtenir le knock-down du gène producteur des miRNAs, DCL1. Cette approche n’ayant pas permis l’identification de bonnes lignées knock-down, nous

avons décidé de créer des lignées transgéniques de *L. japonicus* exprimant P19, un suppreseur virale du RNA silencing, pour interférer spécifiquement avec l'activité des miRNAs. Ces lignées ont montré une forte expression de P19, une accumulation plus elevé des brins miR* et la surexpression de plusieurs cibles de miRNA. Ces observations valident ces lignés P19 comme un nouvel outil pour l'étude du rôle des miRNAs chez les légumineuses.

Ensuite, nous avons entrepris une stratégie de clonage et séquençage de sRNAs, avec l'objectif d'identifier des miRNA impliqués dans la symbiose. Nous avons généré des librairies de sRNAs à partir des racines inoculés par *Rhizobia* et des nodules développés, qui ensuite ont été séquencées par la méthode haut-débit du 454. Une série d'environ 250.000 séquences obtenue, a été employée pour l'analyse de gènes de miRNAs avec l'algorithme miRCat. Nous avons identifié 27 familles de miRNAs conservés, incluses 3 familles de miRNA spécifiques aux légumineuses, et prédit 28 nouveaux miRNA putatifs. La plupart de ces miRNAs potentiels n'a pas été trouvé dans d'autres espèces de légumineuses et ont été considérés comme étant spécifiques à *L. japonicus*. Au moins 3 de ces 27 miRNAs putatifs ont été détectés par l'analyse en *Northern* blot et leur expression a été caractérisée dans différents organes. Suite au clonage des miRNA* de 2 de ceux miRNA potentielles, ils ont été annotés

comme véritables miRNAs: Lja-10 et Lja-12. Il est intéressant de noter que les ciblés prédites pour ces miRNAs putatifs incluent un grand nombre des gènes de résistance, comme précédemment décrit pour d'autres miRNAs identifiés chez des légumineuses et d'autres espèces. L'analyse différentielle de sRNAs clonés à partir des échantillons de racine ou nodule a révélé que l'expression d'un groupe de miRNA conservés est induite dans les nodules développés. Cette accumulation spécifique a été confirmée expérimentalement par *Northern blot*. C'est ainsi que nous avons déterminé que l'expression de *MIR167*, *MIR172, MIR390* et *MIR397* est induite dans les nodules. En plus, des analyses d'expression à grande échelle (Affymetrix) on permis d'identifier un membre de la famille de miRNAs conservés *MIR171*, Lja-miR171c, qui est conservé dans les légumineuses et non pas dans *Arabidopsis*. Ce gène cible potentiellement un des gènes clés dans la signallisation initiale de la symbioses, *NSP2*, et donc est défini comme un vrai candidat comme riborégulateur de la symbiose.

Ces observations nous ont amenés à nous demander si l'expression spécifique de ces gènes était liée au développement du nodule comme organe et ceci indépendamment ou non de l'infection par *Rhizobium*. Pour répondre à cette question, nous avons utilisé des mutants qui produisent des nodules spontanément, sans inoculation préalable, *snf (spontaneous*

nodulation mutant). Étonnamment, miR167 et miR172, qui ont été clonés en haute fréquence à partir des nodules d'autres légumineuses, sont exprimés au même niveau dans les nodules spontanés, et donc sont indépendants de la présence de *Rhizobium*. Chez les légumineuses, ainsi que dans d'autres espèces végétales, miR167 cible des facteurs de réponse à l'auxine (ARFs), alors que miR172 régule l'expression de facteurs de transcription de type APETALA-2 (AP2). La forte accumulation de miR167 dans les nodules non infectés indique que l'homéostasie de l'hormone auxine pourrait être spécifiquement importante pour l'organogénèse ou la maintenance du nodule, avec indépendance de la présence des bactéries. Concernant miR172, qui est exprimé dans la zone d'infection des nodules indéterminés de *M. truncatula*, nos observations suggèrent soit un rôle différent dans la nodulation déterminée, soit une fonction générale dans les événements cellulaires préalables à l'infection du nodule.

Contrairement à miR167 et miR172, l'expression de miR397, très abondante dans les nodules, était sous les limites de détection dans les nodules spontanés. MiR397 est le premier miRNA spécifique de l'infection qui a été caractérisé dans tout type des nodules étudiés jusqu'à présent. L'analyse de tissu systémique a révélé que l'interaction avec *Rhizobia* entraîne une augmentation de l'expression du miR397 simultanément dans

les nodules, racines et feuilles de la plante. Dans le but de rechercher si cette induction de miR397 est dépendante de la présence de *Rhizobium* ou de l'activation du métabolisme de la fixation d'azote, nous avons analysé des intéractions symbiotiques incomplètes de *L. japonicus*, notamment les mutants *sen1 (stationary endosymbiont nodule)*, *sst1* (*symbiotic sulfate transporter*) et l'infection avec l'espèce incompatible *Bradyrhizobium* sp. (*Lotus*) strain NZP2309. Nous avons pu observer que, surtout au niveau systémique, l'induction du miR397 est réduite dans ces mutants. Ce miRNA a été récemment impliqué dans la régulation de la redistribution de Cu^{2+} chez les plantes en cas des carences. En parallèle, les nodules matures deviennent une source de demande de Cu^{2+} très important, en raison de l'utilisation d'une chaine de transport électronique alternative basé sur des cytochromes spécifiques adaptés aux conditions anaérobiques de la fixation d'azote. Ces résultats suggèrent l'implication de miR397 dans la redistribution de nutriment entre les deux symbiontes qui se produit lors de l'établissement de la Fixation Symbiotique d'Azote, ceux-ci sont en accord avec nos études en cours sur les cibles potentielles de miR397.

Nos observations confirment que les miRNAs sont des régulateurs importants à différentes étapes de la fixation symbiotique de l'azote. Des analyses plus approfondies

permettront d'obtenir des détails additionnels sur le rôle de ces gènes dans la régulation de ce processus biologique aux implications économiques et environnementales majeures.

i. Table of Contents

ii. Abbreviations

AGO	ARGONAUTE-LIKE
APS	ammonium persulfate
ARF	auxin response factor
At	*Arabidopsis thaliana*
B&D	Broughton and Dilworth (medium)
bp	basepairs
CaMV	cauliflower mosaic virus
ca-siRNA	cis-acting siRNA
cDNA	complementary DNA
CDS	coding sequence
DCL	DICER-LIKE
dpi	days post-inoculation
dpg	days post-germination
dsRNA	double-stranded RNA
DTT	dithiothreitol
EDTA	ethylene diamine sodium tetraacetate
EMS	ethyl metane-sulfonate
EtBr	ethydium bromide
FP	Fahraeus (medium)
g	gram
GFP	Green Fluorescent Protein
GUS	β-glucuronidase
h	hour
HEPES	N-2 hydroxyethylpiperazine acid N'-2 ethane sulfonate
HMW	High Molecular Weight
hpi	hours post-inoculation
kb	kilobase
L	litre
LB	Luria Bertoni
Lj	*Lotus japonicus*
LMW	Low Molecular Weight
LNA	Locked Nucleic Acid
m	mock
M	molar
mA	miliamperes
MG20	*Lotus japonicus* accession miyakojima MG-20
min	minutes
miRNA (miR)	microRNA
MIR	microRNA gene
ml	milliliters
mm	nucleotide mismatch
Mt	*Medicago truncatula*
mya	millions years ago

μCi	microCurie
natsiRNA	natural-antisense transcript siRNA
nt	nucleotide
OD_{600}	optical density at 600 nanometres
o/n	overnight
ORF	open reading frame
PAGE	poly acrylamide gel electrophoresis
PVP	polyvinylpyrrolidone
qRT-PCR	quantitative RT-PCR
rasiRNA	repeat-associated siRNA
RISC	RNA-Induced Silencing Complex
RNAi	RNA interference
rpm	rotation per minute
RNA	Ribonucleic Acid
RNS	Root Nodule Symbiosis
rRNA	ribosomal RNA
RT	reverse transcriptase
RT	room temperature
SDS	sodium dodecyl sulfate
snRNA	small nuclear RNA
snoRNA	small nucleolar RNA
siRNA	small interfering RNA
SNF	symbiotic nitrogen fixation
sRNA	small RNA
SSC	salt sodium sitrate
tasiRNA	trans-acting siRNA
TBE	tris borate EDTA
TBSV	Tomato Bushy Stunt Virus
T-DNA	Transfer DNA
TEMED	NNNN' tetramethyl ethylene diamine
TF	Transcription Factor
TGS	Transcriptional Gene Silencing
TILLING	Targeting Induced Local Lesions in Genomes
u	enzyme units
UV	ultraviolet
V	volts
VSR	Viral Silencing Suppressor
wpg	weeks post-germination
wpi	weeks post-inoculation
wt	wild-type
3hm	3-hour mock-inoculated
3dm	3-day mock-inoculate

iii. Summary

Legumes are able to establish a mutualistic association with bacteria from the genus *Rhizobia*, known as Root Nodule Symbiosis (RNS). This is a complex biological process whereby, upon plant recognition of a compatible symbiont, bacteria penetrate the roots and colonize newly formed organs, called nodules, were N_2 fixation takes place. Genetic research in legumes has been fastened in the recent years by the genome sequencing initiatives from two legume model species, *Medicago truncatula* and *Lotus japonicus*. The parallel study of two model systems has yielded important discoveries for the understanding of the complex RNS pathway, and genome sequencing now opens the door for large-scale studies as small RNA profiling. Small RNAs (sRNAs) are 18-25nt molecules that regulate gene expression through sequence-specific base-pairing to target RNA or DNA molecules *via* a mechanism known as 'RNA silencing'. sRNAs are generated from double-stranded RNA by the action of RNAse-III proteins, called Dicers, and subsequently loaded into effector protein complexes containing a member of the Argonaute family of enzymes. Among the different types of smRNAs, microRNAs (miRNAs or miRs) are post-transcriptional gene regulators that are specifically diced as miR/miR* duplexes from stem-loop RNA precursors. In plants, miRNAs play key regulatory roles in development, in metabolite homeostasis and in the integration of biotic as well as abiotic stresses including symbiotic responses.

In this scientific context, we chose the determinate legume model *Lotus japonicus* to assess the diversity and potential contribution of RNA silencing to RNS. First, we sought to identify the main RNA silencing machinery components in *Lotus japonicus* and exhaustive bioinformatic analysis revealed that *L. japonicus* contains at least 5 Dicer-like and 9 confirmed Argonaute-like genes. Additionally, canonical orthologs of the genes involved in the *Arabidopsis* miRNA pathway were also indentified. Analysis of large-scale transcriptome data showed that a set of RNA Silencing-related genes are upregulated in nodules, including *LjDCL2*, *LjAGO2* and *LjAGO5*, suggesting that a wide range of small RNA types are involved in RNS. Additionally, in an

attempt to modulate the *L. japonicus* miRNA pathway, we initiated a TILLING-based strategy to knockdown the miRNA-processor gene *LjDCL1*. This attempt being unsuccessful, we thus decided to generate *L. japonicus* transgenic lines expressing the tombusviral silencing suppressor P19, as a means of interfering with miRNA activity. These transgenic lines showed a high expression of P19, as well as impaired general growth, higher accumulation of miRNA* passenger strands and increased expression of several miRNA targets. These results establish P19 transgenic plants as a new tool for novel miRNA characterization in legumes.

We next generated and sequenced sRNA libraries corresponding to *L. japonicus Rhizobia*-inoculated roots and mature nodules. The pooled sequencing data consisting of around 250.000 reads was used for the identification of miRNA genes in *Lotus japonicus*. 27 known miRNA families, including 3 legume-specific, were identified and 28 potential novel miRNAs were predicted. Molecular analysis confirmed 2 *Lotus*-specific bona-fide miRNAs. Interestingly, predicted targets for many of the potentially novel miRNAs included disease resistance or disease-related genes. On the other hand, differential analysis of cloned smRNAs from root and nodule samples revealed that a discrete number of conserved miRNAs are highly expressed in mature nodules, as opposed to roots, which was further confirmed by molecular assays and large-scale transcriptome analysis of pri-miRNAs. Hence, *MIR167, MIR171c, MIR172, MIR390, MIR397* and *MIR408* genes are differentially expressed in determinate *L. japonicus* mature nodules. Moreover, miR171c was pinpointed as divergent miR171 isoform being conserved in legumes and predicted to target the important nodulation signaling pathway regulator gene *NSP2*.

Next, through the analysis of spontaneous nodulation mutants (*snf*), we demonstrated that miR167 and miR172 accumulation in nodules is independent, not only of the nodule type, but also of bacterial infection, placing these miRNAs at the nodule developmental stages. On the other hand, we showed that, upon the development of a successful symbiosis, the high miR397 accumulation observed in nodules is extended to leaf tissue, and this local and systemic induction are simultaneously inhibited in spontaneous *Rhizobia*-free

nodules. Further analysis of a series of ineffective symbiotic interactions indicated that the systemic accumulation of miR397 is mostly dependant on the onset of a successful N_2 fixation rather than on bacterial colonization. Considering the reported involvement of miR397 in Cu^{+2} homeostasis regulation, these results suggest a role of miRNAs in the overall plant-bacterial nutrient redistribution that takes place in RNS, and further experiments following these lines of evidence are suggested.

Our findings confirm that miRNA are important gene regulators at different stages of the establishment of Symbiotic Nitrogen Fixation. We are hopeful that our future work will shed more light on the specific roles of these miRNAs in the dynamics and overall complexity of the genetic regulations involved in this economically and biologically important process.

I. INTRODUCTION

1. *Lotus japonicus*: a model legume

Legumes (*Fabaceae; Papilionoideae*) represent a third of the world's crop production and an important protein source for humans and livestock. Species from this family constitute the basis for a sustainable agriculture, since they are able to assimilate N_2 by the establishment of symbiosis with bacteriae from the genus *Rhizobia*, through a process called Symbiotic Nitrogen Fixation (SNF) or Root Nodule Symbiosis (RNS) (Udvardi et al., 2005; Markmann and Parniske, 2009). Research in legumes has been recently fastened by the parallel adoption in the early 90's of two complementary model species: *Lotus japonicus* and *Medicago truncatula*. This was due to their small diploid genomes, self-fertility, short generation times and ability to be genetically transformed (Barket et al., 1990; Handberg and Stougaard, 1992; Cannon et al, 2009). Important genetic tools have since been developed, including mutant collections (Perry et al., 2003), genetic linkage maps (Sandal et al., 2006) and large-scale gene expression analysis studies as Affymetrix® GeneChips (Mitra et al., 2004; Benedito et al., 2008; Guether et al., 2009; Hogslund et al., 2009). These have led to significant scientific breakthroughs in the recent years, especially regarding the understanding of the molecular basis of Root Nodule Symbiosis. Furthermore, genome sequencing initiatives from these two model species plus soybean (*Glycine max*) are attaining completion, which will indeed accelerate the rate of discoveries and their application to crop legumes (Cannon et al., 2009). *Lotus japonicus* genome sequencing project, which has followed both clone-to-clone and shotgun sequencing approaches, already covers 67% of the 472Mb genome, which is estimated to include 91,3% of the gene space (Sato et al., 2008).

2. The Root Nodule Symbiosis (RNS)

The Root Nodule Symbiosis (RNS) is a complex endosymbiotic process whereby, upon plant recognition of *Rhizobia*-secreted lipochito-oligosaccharide molecules (Nod factors, NF), bacteriae are allowed to penetrate the roots and colonize newly formed organs called nodules. These novel structures host differentiated nitrogen-fixing forms of bacteria, the bacteroids, and intense nutrient exchange is initiated between the two symbionts (Szczyglowski et al., 1998; White et al., 2007; Crespi and Frugier, 2008; Oldroyd and Downie, 2008). One of the distinctive features of 'legume-*Rhizobia*' symbiosis is its specificity, in that each legume species associates with a particular *Rhizobia* species. Thus, *L. japonicus*, *M. truncatula* and soybean respectively associate with *Mesorhizobium loti*, *Sinorhizobium meliloti* and *Bradyrhizobium japonicum*, in some cases in a strain-specific manner (Bek et al., 2009). However, the Root Nodule Symbiosis is not restricted to the association between the *Fabaceae* and *Rhizobia*, as certain non-legume species can associate with bacteriae from the genus *Frankia* (Markmann and Parniske, 2009). For convenience, the terms 'Root Nodule Symbiosis', 'Symbiotic Nitrogen Fixation' and 'legume-*Rhizobia* symbiosis' will be used equivalently in this text.

2.1. The RNS at the microscopic level

Throughout the establishment of a successful legume-*Rhizobia* symbiosis, a series of infection and nodule organogenesis steps take place at the tissue level. Here we will focus on the interactions between *L. japonicus* and *Mesorhizobium loti*. The perception of a compatible *M. loti* strain, previously attached to the root system, is followed by the deformation and curling of plant root hairs, entrapping bacterial colonies. These infection foci result in the development of the so-called 'infection threads', which are plant-derived tubular invaginations that allow the bacterial invasion of the root tissue through the epidermis and later on through the cortex (Figure I-1; Oldroyd and Downie, 2008). In parallel to these primary infection steps, the outer cortical cells

immediately below the bacterial entry point initiate anticlinal cell divisions, forming a nodule primordia that will further develop into a defined meristem (Figure I-2D). The *L. japonicus* nodule meristem ceases its activity early in nodule development so that nodule growth is the result of cell expansion instead of cell division, leading to round-shaped 'determinate' nodules (Hirsch, 1992). At 5 to 7 days post inoculation (dpi), the young nodule emerges, resulting in the splitting of the epidermis (Figure I-2E). Simultaneously, infection threads progress towards the nodule meristem. Bacteriae are finally released from infection threads into young nodules by migrating to the so-called 'infection droplets', which are in turn encapsulated by a plant-derived 'peribacteroid membrane' and released into the cytosol. Membrane-bound endosymbiotic bacteria, or 'symbiosomes', eventually undergoing cell divisions, finally differentiate into nitrogen-fixing bacteroids (Szczyglowski et al., 1998; Oldroyd and Downie, 2008). At 10 dpi, colonized mature nodules start N_2 fixation (Figure I-2F) and display several clear layers: a central infected zone, a vascular bundle-containing nodule parenchyma, a nodule endodermis and a peripheral nodule cortex (Figure I-3; Hirsch, 1992; Puppo et al., 2005). Approx. 6-7 weeks post inoculation (wpi), a senescence process is started at the centre of the tissue and progressively extended to the rest of the organ (Puppo et al., 2005).

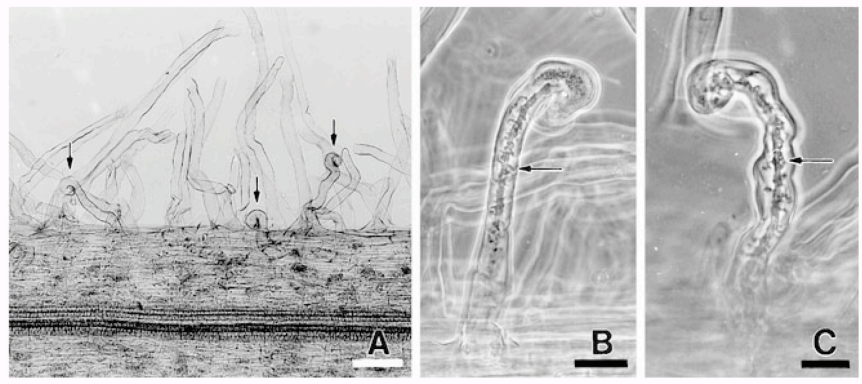


Figure I-1. Early infection of *Lotus japonicus* roots by *Mesorhizobium loti*. A. Initial root hair curling following NOD factor perception. **B-C**. Intracellular infection threads within curled root hairs (From Szczyglowski et al., 1998; Bar, 70 µm (A), 25 µm (B) and 20 µm (C)).

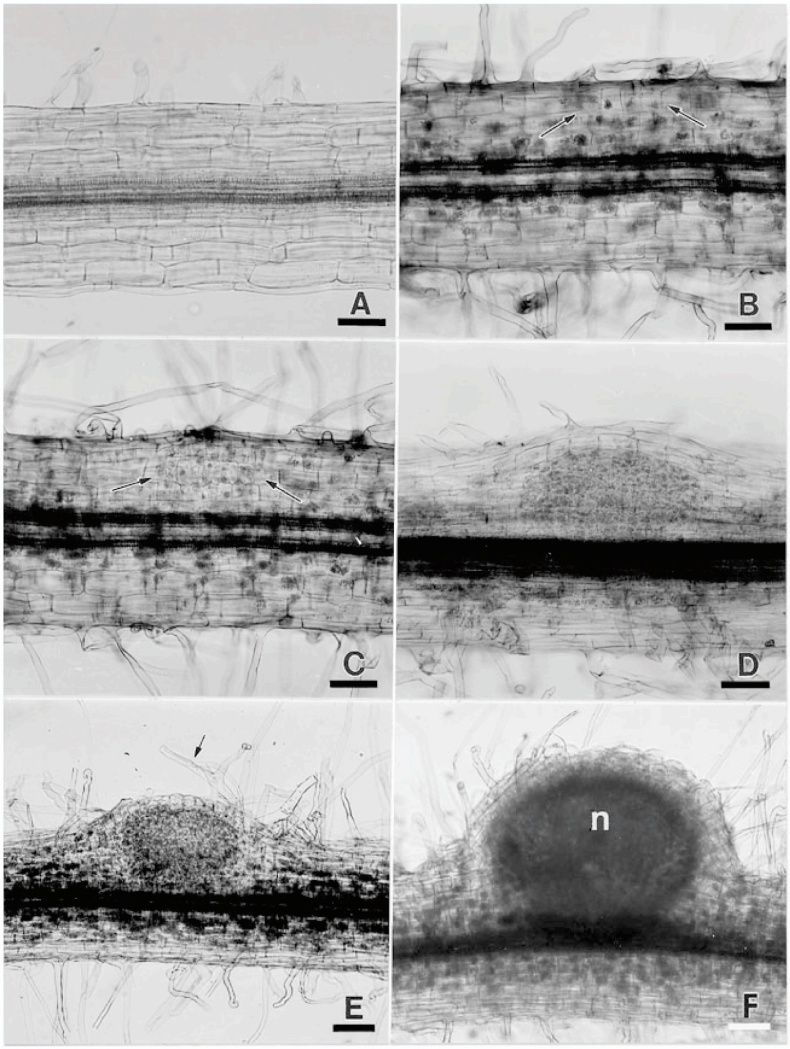


Figure I-2. Nodule developmental stages in *Lotus japonicus*. A. Normal histology of the root cortex containing three to four layers of cylindrical cells. **B.** Early developmental stage in which anticlinal cell divisions are restricted to a localized area of outer cortex (arrows). **C.** More anticlinal cortical cell divisions in a localized area containing several layers of the cortex (arrows). **D.** Clearly defined meristem consisting of the nodule primordium that caused the overlying root epidermis to bulge locally. **E.** Splitting of the overlying epidermis (arrow) as the young nodule emerged. **F.** A fully emerged nodule (n) (Brightfield micrographs from Szczyglowski et al., 1998; Bar, 70 μm (A–D) and 100 μm (E and F)).

Tropical legumes as *L. japonicus*, soybean and beans develop globular-like determinate nodules. Conversely, temperate legumes such as *Medicago* or pea, form nodules containing a permanent meristem leading to indeterminate growth. The resulting long-shaped 'indeterminate nodules' display several clear zones: apical meristematic (zone I), cell differentiation and *Rhizobia* infection (zone II), nitrogen-fixing (zone III) and senescence (zone IV) regions (Vasse et al., 1990). Further differential features are seen in indeterminate nodules: (i) the nodule meristem derives from the inner cortex cells instead of outer cortex, (ii) inside the symbiosomes, bacteroids undergo genomic endoreduplication, but not cell division, and (iii) the senescent area is increased as nodule develops, towards a whole senescent nodule (Hirsch, 1992; Puppo et al., 2005; den Herder and Parniske, 2009)(Figure I-3). The parallel study of two model systems as *Lotus japonicus* and *Medicago truncatula*, representative of the two major nodule types, has yielded important discoveries for the understanding of the complex RNS pathway.

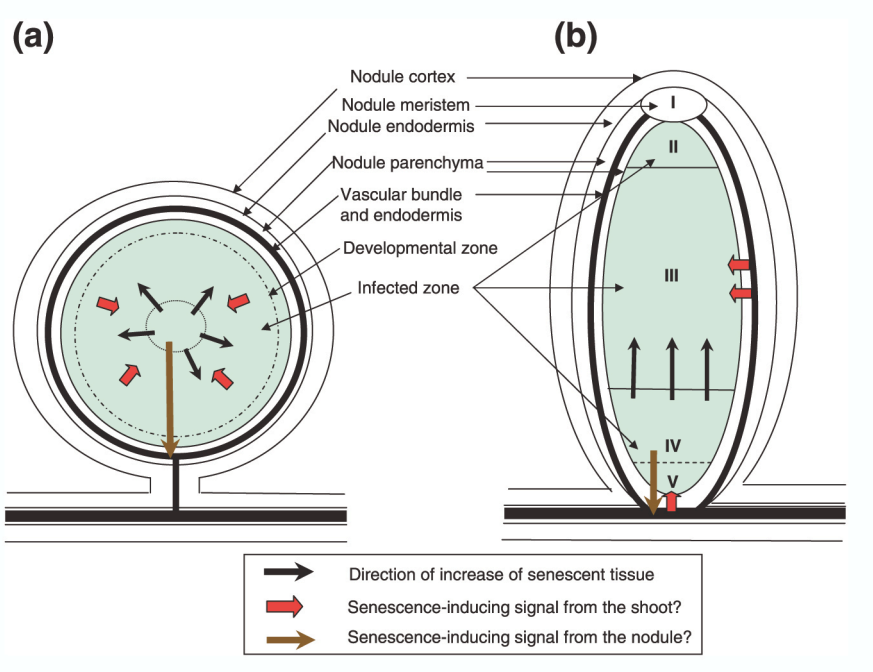


Figure I-3. Scheme of organ structure and senescence in determinate and indeterminate nodules. (a) Determinate globular-like nodules with a radial progress of senescence (b) Indeterminate nodules, including meristematic (I), differentiation and infection (II) , nitrogen-fixing (III) and senescent (IV) zones. The senescence region increases as nodule develops. (From Puppo et al., 2005)

2.2. The RNS at the molecular level

The formation of successful N_2-fixing nodules requires the coordination of the physically separated early bacterial infection and nodule meristem organogenesis (Oldroyd and Downie, 2008). Major advances have been made in the last decade towards the understanding of these molecular pathways, yet many aspects remain elusive.

2.2.1. Epidermal responses and early infection: Nod factor signalling

M. loti-synthesized Nod factors (NF) are perceived by *L. japonicus* through two LysM domain receptor-like kinases (RLK): Nod-factor receptor 1 (NFR1) and 5 (NFR5) (Madsen et al., 2003; Radutoiu et al., 2003), which defines the endosymbiont specificity (Radutoiu et al., 2007). The downstream signal transduction pathway involves the so-called 'common symbiosis signalling pathway' or 'Sym' genes. Indeed, in parallel to Root Nodule Symbiosis, legumes undergo a distinct type of endosymbiosis, Arbuscular Mycorrhiza (AM), by which they associate with fungus from the taxonomic group of Glomeromycota for an increased uptake phosphate and water (Parniske, 2008). Mutant studies have revealed that RNS and AM are genetically linked by a set of at least eight genes that constitute the common 'Sym' pathway (den Herder and Parniske, 2009): the leucine-rich receptor-like kinase SYMRK (Stracke et al., 2002), the potassium channels CASTOR and POLLUX (Charpentier et al., 2008), the nucleoporins NUP85 (Saito et al., 2007), NUP133 (Kanamori et al., 2003) and NENA (den Herder and Parniske, 2009), the calcium- and calmodulin-dependent protein kinase CCaMK (Gleason et al., 2006; Tirichine et al., 2006a) and finally the coiled-coil domain protein CYCLOPS (Yano et al., 2008). The initial recognition of Nod factors by NFR genes leads to the activation of this common Sym pathway, which in turn activates Ca^{2+} spiking. These Ca^{2+} signals are transduced by CCaMK through the activation of two GRAS domain transcription regulators, nodulation signalling pathway 1 and 2 (NSP1, NSP2) and the ERF transcription factors ERN1, 2 and 3 (Middleton et al., 2007), which regulate gene expression at the nucleus. In *M. truncatula*, NSP1 and NSP2 form a complex leading to the

activation of early nodulin 11 (ENOD11) expression (Hirsch et al., 2009) (Figure I-4). Genes involved in actin rearrangement, *LjPir1* and *LjNap1*, are also important for infection thread formation and progress (Yokota et al., 2009).

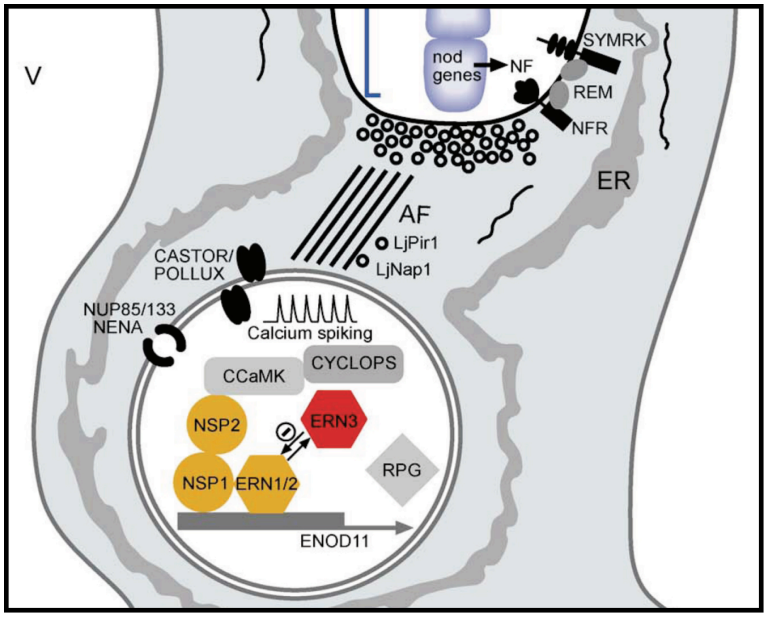


Figure I-4. Signal transduction from Nod factor (NF) perception to gene expression regulation at the nucleus. See text for further details. AF, actin filaments; ER, endoplasmic reticulum; REM, remorins; V, vacuole (Adapted from den Herder and Parniske, 2009).

2.2.2. Nodule organogenesis at the root cortex

Concomitantly to early bacterial infection, the root cortex initiates the formation of nodule primordiae. In opposition to epidermal responses, the nodule organogenesis program is controlled entirely by the plant, as reflected by the existence of mutants developing genuine nodules in the absence of *Rhizobia*: the spontaneous nodulation mutants (*snf*). *L. japonicus snf1* mutants harbor a gain-of-function mutation in the *CCaMK* gene, leading to an autoactivation of the Nod factor recognition signaling cascade (Tirichine et al.,

2006). Similarly, a dominant mutation in the cytokinin receptor *Lotus Histidine Kinase 1 (LHK1)* gene, present in the *snf2* mutant, triggers nodule formation regardless of bacterial presence (Tirichine et al., 2007). In addition, several transcription factors are required for nodule organogenesis: the DNA binding and membrane spanning protein NODULE INCEPTION (NIN), presumably a transcriptional regulator (Schauser et al., 1999), the GRAS domain proteins NSP1 and NSP2, ERN1, 2 and 3, and finally the microRNA-regulated HAP2.1 transcription factor (see below) (Combier et al., 2006; Crespi and Frugier, 2008).

Combining previous experimental evidences with gain-of-function mutants, it is clear that cytokinins are central players in coordinating epidermal and cortical responses. A linear pathway from Nod factor recognition to nodule development through CCaMK and LHK1 has been suggested (Frugier et al., 2008). However, it has also been proposed that the signaling pathways involving CCaMK and LHK1 are separated, as the combination of both dominant mutations results in enhanced nodule number (Oldroyd and Downie, 2008). Future studies will shed light on this poorly understood but relevant step in RNS.

2.2.3. A hormonal control of nodulation

Beyond the central role of cytokinins in nodule organogenesis, the entire collection of plant hormones are involved in fine-tuning the complex regulatory pathways of RNS. Auxin, cytokinins, giberellins (GA) and brassinosteroids (BR) are positive regulators of nodule formation. On the other hand, abscisic acid (ABA), jasmonic acid (JA), ethylene and salicylic acid (SA) negatively repress nodulation, presumably as a response to biotic or abiotic stresses (Oldroyd and Downie, 2008; Ding and Oldroyd, 2009). In *L. japonicus*, GA also suppresses nodulation, and GA treatment phenocopies *nsp2* mutations (Maekawa et al., 2008). Despite the fact that the specific molecular pathways underlying this complex regulation are poorly understood, it is clear that hormones are key players in the control of nodule development.

2.3. Solving the *Lotus* nodule puzzle through mutant analysis

Forward genetics has been decisive to understand the molecular basis of nodulation. *L. japonicus* mutant collections have been classified into distinct categories: *nod*$^-$, absence of nodulation; *hist*$^-$, affected in infection thread formation; *fix*$^-$, nodule-like structures are formed but show inefficient N_2 fixation; *snf*, spontaneous nodules (Szczyglowski et al., 1998; Kawaguchi et al., 2002; Tirichine et al., 2006). The understanding of the Nod factor signal transduction pathway was greatly benefited from the positional cloning of SYMRK (Stracke et al., 2002), NFR1 and NFR5 (Madsen et al., 2003; Radutoiu et al., 2003), among others. Similarly, the *snf* mutants uncoupled the endogenous nodule organogenesis pathways from the presence of *Rhizobia* (Tirichine et al., 2006 and 2007).

Genetic studies are also potent tools for the study of mature nodule biology. In this sense, *L. japonicus* determinate nodules become an excellent model system, as the different stages, including nodule development, infection or senescence, are spatio-temporally separated. Different *fix*$^-$ mutants have been identified in this species such as *sen1* (Suganuma et al., 2003), *sst1* (Krussel et al., 2005), *ign1* (Hossain et al., 2006) and *sym105* (Kumagai et al, 2007). These mutants develop *Rhizobia*-colonized nodules were N_2 fixation is at least partially impaired, when not abolished, and show premature senescence. In addition, some rhizobial strains as *Bradyrhizobium loti* strain NZP2309 establish a partial symbiosis with *L. japonicus* where N_2 fixation is strongly reduced (Scott et al., 1985; Bek et al., 2010). This set of ineffective symbioses represent unique tools for the understanding of the partial contribution of each symbiont to nodule biology. Indeed, the *Lotus fen1* mutant was recently mapped, revealing that *FEN1* encodes homocitrate, a crucial component of the iron-molybdenum cofactor needed for the bacterial nitrogenase complex activity (Hakoyama et al., 2009). This study, hence, emphasized the idea that plant roots are not only a physical platform hosting bacteriae, which would perform nitrogen fixation in exchange for other nutrients. Rather, the plant directly contributes and makes possible those metabolic processes required for nitrogen fixation.

3. Small RNAs in plants.

Small RNAs (sRNAs) are 18-30 nt molecules that regulate gene expression through sequence-specific base-pairing to target RNA or DNA molecules *via* a pan-eukaryotic mechanism known as 'RNA silencing'. First identified in plants as a mechanism for the defense against exogenous RNA derived from virus or transgenes, it was soon discovered that endogenous small RNA pathways represent one of the main ways for gene expression modulation in plants, and play crucial roles in virtually all important processes ranging from development to response to stress (Voinnet, 2001; Ruiz-Ferrer and Voinnet, 2009; Chen, 2009).

3.1. The core RNA Silencing mechanism

Different types of small RNAs have been described in plants according to their regulatory roles and/or biochemical activities, though a core biochemical mechanism accounts for their processing and action in the cell. Small RNAs are excised, or 'diced', from double-stranded RNA (dsRNA) by the action of RNAse-III proteins, called Dicers, with RNA helicase, RNAse-III, PAZ and double-stranded RNA-binding (dsRBD) domains. The resulting 18-25 nt small RNAs are subsequently loaded into effector protein complexes termed 'RNA Silencing-Induced Complex' (RISCs), which consistently include a member of the Argonaute enzymes, containing both an sRNA-binding PAZ and a PIWI domains, the latter with potential slicer activity (Tolia and Joshua-Tor, 2007)(Figure I-5). This ultimately leads to a sequence-specific modulation of gene expression at the transcriptional, mRNA or translational level. These key processor (Dicer) and effector (Argonaute) proteins, respectively associate with distinct partners in the cell to form the complex set of RNA Silencing pathways (Brodersen and Voinnet, 2006; Ruiz-Ferrer and Voinnet, 2009).

While both Dicer and Argonaute genes are greatly conserved across plants and animals, their diversity in terms of gene numbers and variants reflects the high complexity of RNA silencing pathways across and within kingdoms (Figure I-5; Vazquez, 2006). In plants, the model species *Arabidopsis thaliana* contains 4 Dicer-like (*DCL*) and 10 Argonaute-like

(*AGO*) paralogs, with either unique or redundant functions (Ghildiyal and Zamore, 2009). Nonetheless, studies in other plant species have revealed a divergent number of *DCL* genes across organisms, with a stronger diversification of plant *DCL2* and *DCL3*-like (Margis et al., 2006). On the other hand, the more abundant *AGO* genes in *Arabidopsis* are phylogenetically grouped into 3 clades (Vazquez, 2006; Vaucheret, 2008). These effector genes are even more diversified in rice, containing 19 paralogs and including a new phylogenetic clade, *MEL1*, whose representative member is specifically expressed in germ cells (Nonomura et al., 2007; Kapoor et al., 2008).

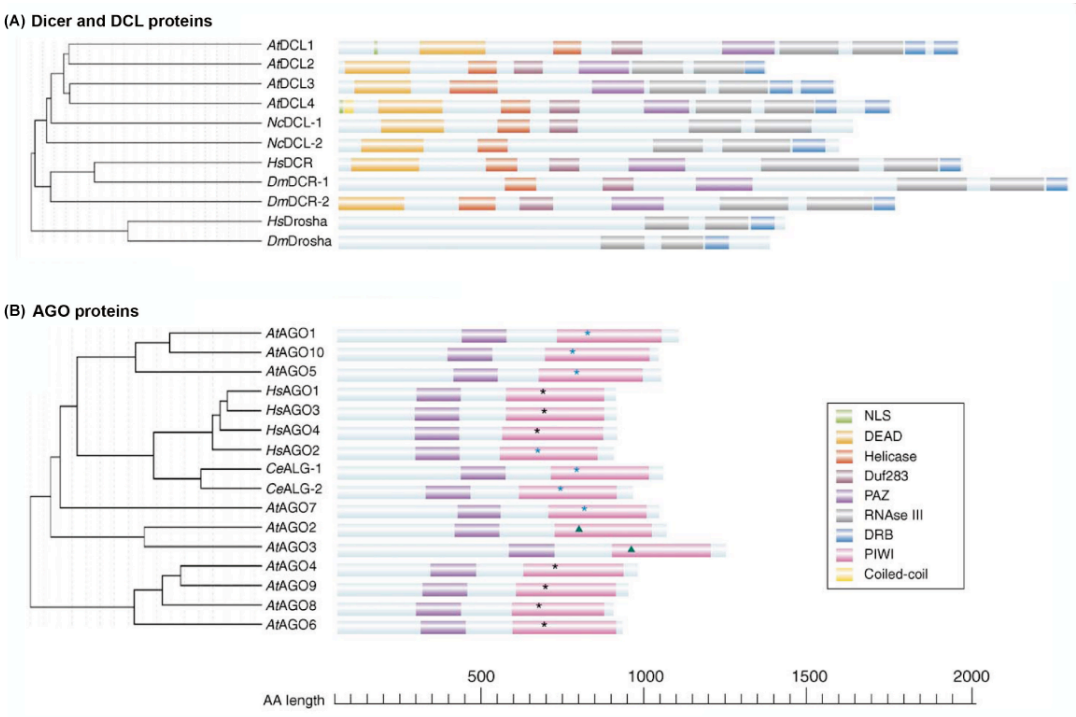


Figure I-5. Phylogenetic and functional domain conservation of *Arabidopsis* Dicer (*DCL*) and Argonaute (*AGO*) genes. Protein domains are represented on the right. **A.** Dicer and *DCL* proteins. **B.** *AGO* proteins. At, *Arabidopsis thaliana*; Ce, *Caenorhabditis elegans;* Dm, *Drosophila melanogaster;* Hs, *Homo sapiens;* Nc, *Neurospora crassa*. Slicer activity in the PIWI domain is represented: blue asterisks indicate slicer catalytic sites (DDH triad); green triangles indicate hypothetical slicer-potent catalytic sites (a DDD triad instead of the DDH triad); black asterisks indicate slicer-impotent catalytic sites in-vitro (Adapted from Vazquez, 2006).

3.2. The diversity of plant small RNAs

Plant small RNAs can be divided into two major subclasses: microRNA (miRNA) and short-interferring RNAs (siRNAs), though the differences among them are increasingly disappearing (Voinnet, 2009). Roughly, microRNAs (miRNAs; miRs) are excised as discrete miRNA/miRNA* duplexes from single-stranded stem-loop RNA precursors, while siRNAs originate from exogenous or endogenous long double-stranded RNA molecules perfectly or near-perfectly paired, generating a population of small RNAs from both precursor strands (Chen, 2009). Insight into these two plant small RNA classes and their biogenesis pathways in the model plant *A. thaliana* is given below.

3.2.1. MicroRNAs and their regulatory roles

3.2.1.1. The microRNA pathway in *Arabidopsis*

Plant microRNA genes, predominantly intronic or intergenic, are transcribed by RNA polymerase II, producing a capped and polyadenylated transcript termed 'pri-miRNA'. The latter is subsequently processed by DICER-LIKE 1 (DCL1) into a stem-loop structure, the 'pre-miRNA', which is in turn processed into the 19-22 nt miRNA/miRNA* duplex also by DCL1 (Jones-Rhoades et al., 2006). Yet, several partner proteins assist this main processor enzyme (Ruiz-Ferrer and Voinnet, 2009; Figure I-6). The forkhead-associated RNA-binding protein DAWDLE (DDL) interacts directly with DCL1 and contributes to pri-miRNA accumulation, presumably by its stabilization (Yu et al., 2008). Both pri- and pre-miRNA processing also require the double-stranded RNA-binding domain HYL1, and the zinc-finger protein SERRATE (SE) (Kurihara et al., 2006; Yang et al., 2006). In addition, the nuclear heterodimeric cap-binding complex (CBP) promotes the accumulation of certain miRNAs (Gregory et al., 2008; Laubinger et al., 2008).

Diced miRNA/miRNA* duplexes are subsequently 2'O-methylated by the methyltransferase HUA ENHANCER 1 (HEN1) (Yu et al., 2005; Huang et al., 2009). This feature prevents further degradation and distinguishes plant microRNAs from their animal counterparts, where only certain types of small

RNAs, but not microRNAs, undergo methylation (Li et al., 2005; Voinnet, 2009). Mature duplexes are exported from the nucleus to the cytoplasm, in certain cases with the help of factors as the exportin HASTY (Park et al., 2005). Next, one strand of the miRNA/miRNA* duplex is specifically loaded onto the AGO-containing RISCs, leading to sequence-dependent regulation of gene expression by mRNA endonucleolytic cleavage and/or translational inhibition (Brodersen et al., 2008). In parallel, miRNA steady-state levels are regulated by the SMALL RNA DEGRADING NUCLEASES (SDNs) exonucleases (Ramachandran and Chen, 2008)(Figure I-6).

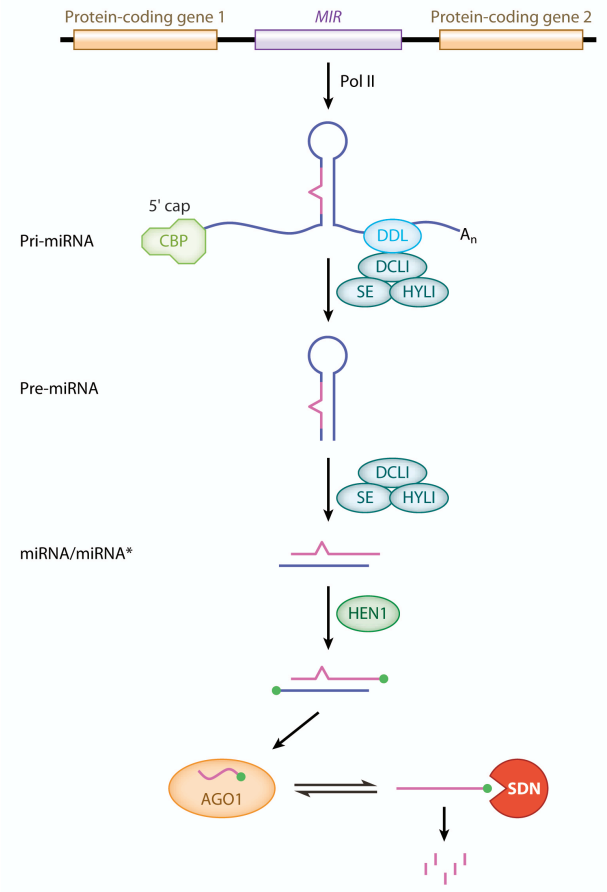


Figure I-6. The microRNA pathway in Arabidopsis. Biogenesis and stability-related steps are shown. Mature miRNAs loaded onto AGO1-containing RISC complex direct sequence-specific mRNA cleavage and/or translational inhibition of target genes. See text for further details. DDL, DAWDLE; DCL1, DICER-LIKE1; SE, SERRATE; HYL1, HYPONASTIC LEAVES1; AGO, ARGONAUTE1; SDN, SMALL RNA DEGRADING NUCLEASE (From Chen, 2009).

3.2.1.2. Regulatory roles of microRNAs

For a long time it was though that the main regulatory mechanism for miRNA-loaded RISC was the cleavage of target genes, coined 'slicing'. In fact, despite widespread miRNA-mediated translational inhibition of target genes in animals, only two examples had been described in plants (Aukerman and Sakai, 2003; Chen, 2004; Gandikota et al., 2007). Yet, a recent work proved that target translational inhibition is widely extended in plants and it is now established that miRNA action is generally a combination of these two regulatory mechanisms. The use of one or the other could depend on, for example, the AGO-interactor proteins present in specific cells (Brodersen et al., 2008; Voinnet, 2009).

At the cellular level, miRNA and target gene expression is not always mutually exclusive, as initial studies suggested. Indeed, the study of various specific miRNA/target pairs has revealed at least three distinct modes of action (Figure I-7) (Garcia, 2008; Chen, 2009; Voinnet, 2009). First, miRNAs can spatially restrict the expression of their target genes by total clearance of their corresponding mRNA, as in miR166 regulation of *HD-ZIP III* genes in *Arabidopsis* and maize (Kidner and Martienssen, 2004; Juarez et al., 2004). Secondly, miRNAs can be expressed simultaneously to their target genes, in a way that they contribute to modulate, or buffer, their expression. Lastly, miRNAs can follow a temporal gradient that controls the levels of their target gene through time. This mode of regulation is observed when a threshold of a target gene induces developmental transitions, as miR172 regulation of *APETALA-2*, controlling flowering (Aukerman and Sakai, 2003). Indeed, the study of the specific regulatory mode of each miRNA/target couple is important to understand the biological relevance of this regulation (Garcia, 2008).

Considered initially as the linchpin of plant development, miRNAs also play key roles in hormone signalling, metabolite homeostasis and the integration of biotic as well as abiotic stresses (Jones-Rhoades et al., 2006; Voinnet, 2009). Particularly, different miRNAs involved in plant-microbe interactions have been recently discovered, such as *Pseudomonas syringae*-induced miR393, repressing auxin signalling (Navarro et al., 2006) or

symbiotic-related miRNAs (see below). Indeed, in a similar way as transcriptional regulation, miRNA-mediated gene regulation cannot be restricted to any particular process (Kidner and Martienssen, 2005).

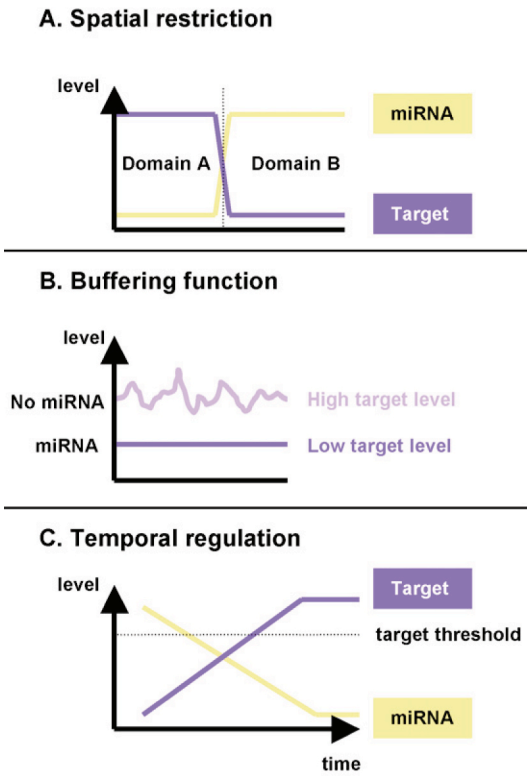


Figure I-7. Modes of action of microRNAs.
See text for further details (From Garcia, 2008).

3.2.1.3. Evolutionary aspects on microRNAs

It is not surprising that the first microRNA study in *Arabidopsis*, involving a total of approx. 200 sequences, identified at a shot 13 *MIRNA (MIR)* gene families that were soon found in other distant plant species (Reinhardt et al., 2002; Jones-Rhoades and Bartel, 2004). In fact, a set of highly expressed *MIR* genes is conserved among all plant species studied and shape the 'conserved' or 'ancient' *MIR* genes (Axtell and Bowman, 2008). Those whose sequences are identical or near identical constitute a *MIR* family (Meyers et al., 2008). While the identification of conserved miRNAs is

reaching saturation, recent deep sequencing assays are revealing an increasing number of nonconserved, also termed 'novel' or 'young' *MIR* genes (Rajagopalan et al., 2006; Fahlgren et al., 2007). One of the hypotheses for *MIR* origin argues that these genes derive from inverted duplication of transcribed genes, resulting in a perfectly matched hairpin, or 'proto-*MIR*', that generates a population of siRNAs. Sequence mutations would introduce mismatches that would allow the processing by DCL1 into uniform 'young' miRNA species. Further drifts would lead to a stem-loop where only the miRNA is similar to the original duplicated gene, as seen in 'ancient' *MIR* genes. Precursor stem-loop duplications would ultimately lead to different members of the same *MIR* family (Allen et al., 2004; Rajagopalan et al., 2006; Fahlgren et al, 2007). Accordingly, 'young' miRNAs are generally unigenic and considerably less expressed than their conserved equivalents, presumably as a mechanism to avoid off-targeting effects from residual siRNAs (Voinnet, 2009). However, miRNAs can also originate from random stem-loops on the genome or repeat-transposable elements that would, by chance, encounter compatible and biologically advantageous target candidates (de Felippes et al., 2008; Piriyapongsa and Jordan, 2008). The rising number of small RNA sequencing studies has led to a homogenization of *MIR* gene annotation in plants, in order to avoid misannotation of siRNAs as true miRNAs. Currently, the necessary and sufficient criteria for the annotation as a genuine *MIR* is miRNA* cloning or confirmation of precise excision from the stem-loop precursor (Meyers et al, 2008).

3.2.2. The short-interferring (siRNA) pathways

Short-interferring RNAs (siRNAs) can either be originated by exogenous or endogenous long RNA molecules. In both cases, the dsRNA precursor molecule is either present in the cell, i.e. viral origin, or synthesized by cellular RNA-dependent RNA-polymerases (RdRPs). Exogenous siRNAs are processed from viral or transgene-derived long RNAs, through the so-called 'antiviral' or 'RNAi' pathway, and they hold the title of being the first small RNAs identified in plants (Hamilton and Baulcombe, 1999). In turn, plant

endogenous small RNAs, coined endo-siRNAs (Ghildiyal and Zamore, 2009), are more diverse in their biosynthetic and regulatory pathways and will be reviewed below.

3.2.2.1. Heterochromatic siRNAs

SiRNAs derived from heterochromatic regions are also referred to as 'rasiRNAs' (repeat-associated siRNAs) because they are mostly derived from repetitive and transposable elements, but not exclusively, as they can also be found in intergenic regions. They are also called 'casiRNAs' (*cis*-acting siRNAs), as they modulate the expression of the genomic loci that originate them (Chen, 2009; Ghildiyal and Zamore, 2009). These 24-nt heterochromatic siRNAs are the most abundant small RNA class in plants, a feature that is easily observed in high-throughput sequencing assays (Lu et al., 2005). With the initial requirement of the polymerase Pol-IV, a chromatin remodelling protein, CLASSY1, and the RNA polymerase RDR2, heterochromatic siRNAs are diced preferentially by DCL3. Upon 2'-O-methylation by HEN1, they are incorporated into AGO4- or AGO6-containing complexes to guide cytosine methylation and histone modification at the originating loci, thus promoting heterochromatin formation (Chen, 2009). Thus, this process is also called RNA-directed DNA methylation (RdDM) (Figure I-8). These ca-siRNAs not only contribute to genome integrity, but also regulate the expression of genes such as *FLOWERING LOCUS C* (*FLC*) (Liu et al., 2004).

3.2.2.2. ta-siRNAs

The trans-acting siRNAs (ta-siRNAs) pathway derives from the cellular convergence of miRNA and siRNA pathways in plants (Ghildiyal and Zamore, 2009). A specific miRNA guides the cleavage of each *TAS* non-coding transcript and, upon presumed stabilization by the SGS3 protein, the RNA-dependent RNA-polymerase RDR6 is recruited and synthesizes a dsRNA molecule. This dsRNA *TAS* precursor is in turn processed at precise 21-nt phases by DCL4. Ta-siRNAs are then methylated and incorporated into AGO-containing RISCs, which ultimately mediates target mRNA cleavage (Allen et al., 2005; Yoshikawa et al., 2005) (Figure I-8). When compared to other small RNAs, ta-siRNAs are relatively rare in *Arabidopsis* genome, where only six

TAS loci have so far been identified: *TAS*1a, 1b, 1c, 2, 3 and 4 (Rajagopalan et al., 2006). Certain differences can be found between these loci, for instance at the miRNA-directed cleavage, as *TAS1* and *TAS3* require AGO1 and AGO7, respectively (Montgomery et al., 2008a and 2008b). *TAS3*-derived ta-siRNAs targets Auxin Response Factors (ARFs) 3 and 4 and, thus, play important developmental roles (Allen et al., 2005). These ta-siRNAs are conserved in *Arabidopsis*, grasses and legumes, and are critical for patterning in leaves and lateral root development in different organisms, including *L. japonicus* (Garcia, 2008; Shen et al., 2009; Yan et al., 2009; Yoon et al., 2009). Current methods for ta-siRNA discovery involve clustering algorithms that, based on sequenced small RNA reads, reveal 21-nt phased patterns (Rajagopalan et al., 2006; Moxon et al., 2008).

3.2.2.3. nat-siRNAs

Natural antisense siRNAs (nat-siRNAs) derive from overlapping sense and antisense transcripts, 'natural antisense' (*NAT*), that coexist in biotic or abiotically stressed cells. In general, one of the transcripts is constitutively expressed, while the other is induced upon stress perception. The region of overlap becomes the dsRNA precursor from which 21-24 nt nat-siRNAs are produced (Borsani et al., 2005; Katiyar-Agarwal et al., 2006). Additional 30-40 nt long siRNAs (lsiRNAs) are induced by bacterial challenge-initiated *NAT* expression (Katiyar-Agarwal et al., 2007). Some divergent RNA Silencing components are required for the biogenesis of each of the only three nat-siRNAs characterized so far, but all involve RDR6 and Pol IV activity, plus at least one of the *DCL* paralogs (DCL1, DCL2 or DCL4) (Borsani et al., 2005; Katiyar-Agarwal et al., 2006; Katiyar-Agarwal et al., 2007). Though only several examples have been characterized, the *Arabidopsis* genome contains more than a thousand *NAT* pairs (Wang et al., 2005).

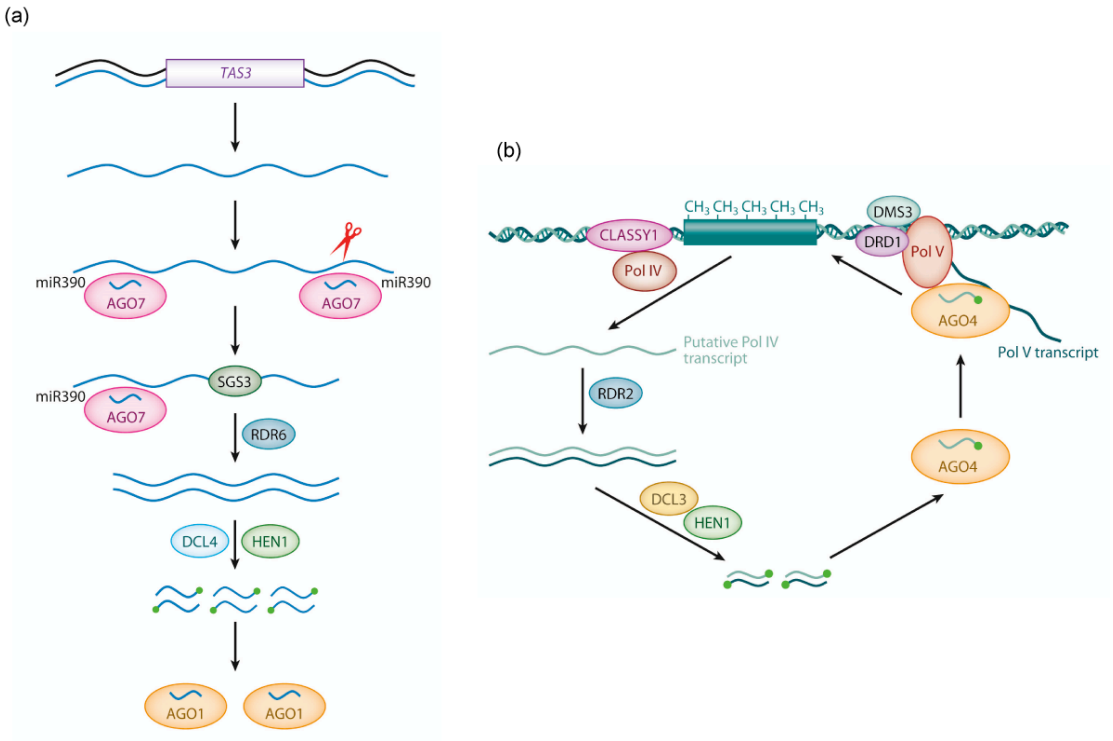


Figure I-8. Endogenous siRNA pathways: ta-siRNAs and heterochromatic siRNAs. (a) Biogenesis of ta-siRNAs. The *TAS3* precursor is recognized by miRNA-guided AGO7 complex at two sites and cleaved at the 3' site. RDR6 generates a double-stranded precursor that is cleaved in 21-nt phases by DCL4. The *TAS3* siRNAs are represented here, whereas in *TAS1*, miRNA-guided cleavage is mediated by AGO1 and recognized at a single site. **(b)** Biogenesis of heterochromatic siRNAs. With the coordination of CLASSY1, Pol IV and RDR2, DCL3 dices HEN1-methylated 24-nt ca-siRNAs, which are in turn incorporated into AGO4-complexes to guide methylation and hystone modifications. Pol V, the chromatin remodelator DRD1 and the DMS3 gene, generate non-coding transcripts also required for DNA methylation. Nat-siRNAs, not depicted here, represent the third plant endogenous siRNA pathway. Pol IV, DNA-dependent RNA polymerase IV; Pol V, DNA-dependent RNA polymerase V; CLASSY, a protein similar to chromatin remodeling proteins; AGO4, ARGONAUTE4; RDR2, RNA-DEPENDENT RNA POLYMERASE2; DCL3, DICER-LIKE3; HEN1, HUA ENHANCER1; dsRNAs, double-stranded RNAs; siRNAs, small interfering RNAs, RISCs, RNA-induced silencing complexes; DRD1, DEFECTIVE IN RNA-DIRECTED DNA METHYLATION1; DMS3, DEFECTIVE IN MERISTEM SILENCING3 (Adapted from Chen, 2009).

3.3. Interference of small RNA functions by Viral Silencing Suppressors

As a counterdefensive response to the antiviral role of RNA Silencing, plant viruses synthesize multiple proteins that block small RNA action at distinct steps (Ruiz-Ferrer at al., 2009). Viral Suppressors of RNA-silencing (VSRs) were first identified in the late 1990s through studies with potyviral HcPro (Anandalakshmi et al., 1998; Brigneti et al., 1998; Kasschau et al., 1998). Shortly after, the tomato bushy stunt virus (TBSV) P19 was also identified as a VSR by Voinnet and co-workers (Voinnet et al 1999). Crystal structures revealed that P19 homodimers act as molecular calipers that seize and measure especifically 21-nt siRNA duplexes (Vargason et al., 2003; Ye et al., 2003) (Figure I-9). Further studies showed that P19 is also able to sequester microRNA duplexes: in P19-expressing transgenic *Arabidopsis* plants, miRNA* accumulation is enhanced, miRNA-mediated target mRNA cleavage is reduced and general plant development is affected (Dunoyer et al. 2004; Chapman et al., 2004). These molecular hallmarks define P19 expression as a unique tool for the characterization of novel miRNAs, as specific dicing as miRNA/miRNA* is the main criteria for *bona-fide* annotation (Meyers et al., 2008). Nonetheless, other silencing suppressors have evolved to interfere the small RNA pathway at different steps, for instance the AGO1-interacting 2b (Cucumber Mosaic Virus) and P0 (polerovirus) (Zhang et al., 2006; Bortolamiol et al., 2007).

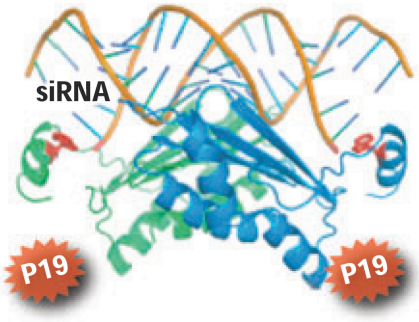


Figure I-9. P19 homodimers sequester small RNA duplexes in vivo. 21-nt siRNA and miRNA duplexes are specifically measured and bound (From Ruiz-Ferrer and Voinnet, 2009).

4. Identification of microRNAs and their targets in legume species.

MiRNAs being involved in nearly all steps of plant development and responses to stress, studies on the identification of small RNAs in species other than *Arabidopsis*, including legumes, have proliferated in the recent years. MicroRNAs can be identified through computational or experimental strategies (Simon et al., 2009). Pure computational strategies have been widely used, where conserved mature miRNA sequences are used to scan sequenced plant genomes or ESTs for miRNA precursor-like hairpin structures (Jones-Rhoades and Bartel., 2004; Dezulian et al., 2006). Following this strategy, 19 conserved families were identified in *L. japonicus*, *M. truncatula* and soybean (Sunkar and Jagadeeswaran, 2008). However, the study of non-conserved or novel miRNAs using this approach is greatly limited, as non-miRNA hairpin structures can indeed be found in plant genomes (Voinnet, 2009). Experimental methods as small RNA cloning and sequencing have revealed a great number of novel miRNAs and siRNAs in *Arabidopsis* and other plant species, with the help of the fast development of large-scale sequencing methods such as 454 pyrosequencing (Roche) or Solexa technology (Illumina) (Rajagopalan et al., 2006; Fahlgren et al., 2007). In fact, the combination of small RNA cloning, high-throughput sequencing and powerful computational tools is the current successful strategy for plant novel small RNA identification. In a recent study, $4x10^6$ reads from *M. truncatula* leaves yielded 25 conserved miRNA families and 8 genuine novel *MIR* loci using the potent University of East Anglia (UEA) plant sRNA bioinformatic toolkit (Simon et al., 2008; Szittya et al., 2008). Parallel studies have identified further conserved and candidate novel *MIR* families in different legume species (Table I-1). However miRNA* cloning or evidence for precise excision from the stem-loop precursor, necessary to discriminate between *bona-fide* miRNAs and siRNAs (Meyers et al, 2008), has not been confirmed for all these novel *MIR* families. Despite the identification of a set of legume-specific microRNAs, it is surprising that the degree of overlap of cloned novel miRNAs remains low, indicating that novel miRNA sequencing in legumes is still far from saturation (Subramanian et al., 2008;

Arenas-Huertero et al., 2009; Jagadeeswaran et al., 2009; Lelandais-Brière et al., 2009).

Species	Reads obtained[(1)]	Candidate novel miRNA	Cloned miRNA*[(2)]	Conserved miRNA families	References
G. max	350.000	35	3/35	20	Subramanian et al., 2008
G. max	375	3	-	4	Wang et al., 2008
M. truncatula	840.000	106	34/106	27	Lelandais-Brière et al., 2009
M. truncatula	22.000	8	4/8	20	Jagadeeswaran et al., 2009
M. truncatula	$4x10^6$	26	8/26	25	Szittya et al., 2008
P. vulgaris	2.000	6	-	18	Arenas-Huertero et al., 2009; Simon et al., 2009

TABLE I-1. Conserved and novel miRNA families cloned so far from legume species.
(1) Approximate number
(2) Cloned miRNA* from each candidate novel miRNA family

So far the main mode of confirmation of miRNA-mediated regulation has been the detection of sliced target mRNAs by rapid amplification of 5' complementary DNA ends assay (5' RACE). This is not surprising when considering that, until recently, slicing was thought to be the main plant miRNA target regulatory mechanism. However, even plant miRNA targets that are regulated by translational inhibition accumulate to a certain extent miRNA-derived cleavage products (Aukerman and Sakai, 2003; Chen, 2004; Brodersen at el., 2008), indicating that this is a qualitative method for target identification, regardless of the predominant regulatory mode of action of the miRNA (Voinnet, 2009). Using this technique, a set of miRNA targets has been identified in legumes (Table I-2). It is not surprising that most of these miRNA:target pairs are conserved in *Arabidopsis*, as it is more than often observed that conserved miRNAs regulate homologous targets at identical target sites (Axtell and Bowman, 2008). However, divergent targets for miR168 and miR396 have been identified in soybean, as it is the case in rice (Table I-2; Subramanian et al., 2008; Jones-Rhoades et al., 2006). Indeed, both conserved and young *MIR* genes are subjected to evolution, and point mutations which can lead to the acquisition of new target genes (Voinnet, 2009).

In most cases, miRNA target confirmation in legumes, as in other species, has been preceded by computational prediction using algorithms such as Target Finder (Allen et al., 2005; Falhgren et al., 2007), miRU (Zhang, 2005) or miRANDA (John et al., 2004). These scripts scan genomic or cDNA sequences based on a set of base-pairing rules derived from the observation of previously known miRNA:target pairs. Despite the fact that these prediction criteria have been crucial for the identification of important miRNA targets, they have recently been revised in the light of experimental evidences, extending the scope of miRNA regulation (Brodersen and Voinnet, 2009). Additionally, it has been shown that miRNA:target regulation does not depend exclusively on nucleotide pairing, but on other factors such as target site accesibility (Kertesz et al., 2007). Therefore, sequence homology-based computational methods are useful though not sufficient to identify all targets. Recently, large-scale miRNA target identification strategies have been developed, namely the Parallel Analysis of RNA ends (PARE), based on direct sequencing of 5'RACE products derived from miRNA-directed cleavage (Addo-Quaye et al., 2008; German et al., 2008). Together with a dramatic reduction of false positives, as compared to computational prediction methods, this technique has helped to discover unexpected miRNA:target pairs in *Arabidopsis* and rice. If applied to legume species, the PARE method will indeed help to extend the scope of potential miRNA targets, beyond those pinpointed by bioinformatic algorithms. In any case, it should be taken into account that later target confirmation should include both the assessment of protein and mRNA levels (Voinnet, 2009).

Species	miRNA	Confirmed targets	Annotation	*At* target family[2]	RNS[4]	References
G. max	**miR166**	TC221756	HD-ZIPIII	HD-ZIPIII	pot	Subramanian et al., 2008
	miR168	BG882680	Protein kinase[1]	AGO1	pot	Subramanian et al., 2008
	miR393	TC255843	TIR1-like	F-box, bZIP	pot	Subramanian et al., 2008
	miR396	TC206710	Cysteine protease[1]	GRF	pot	Subramanian et al., 2008
M. truncatula	**miR160**	ES612384, BQ148941	ARF-like	ARF (10,16,17)	pot	Jagadeeswaran et al., 2009
	miR162	AC150443_32.2	DCL1	Dicer	-	Jagadeeswaran et al., 2009
	miR164	AC203553_1.1	NAC-like	NAC	-	Jagadeeswaran et al., 2009
	miR166	CNA1,CNA2, HB8	HD-ZIPIII	HD-ZIPIII	yes	Boualem et al., 2008
	miR167	AC144478_44.4, CU326393_14.1	ARF-like	ARF (6, 8)	pot	Jagadeeswaran et al., 2009
	miR168	AW773594	AGO1	Argonaute		Jagadeeswaran et al., 2009
	miR169	HAP2-1	CCAAT-bind TF	HAP2	yes	Combier et al., 2006
	miR170	AC121238_43.2	SCL-like	Scarecrow-like	-	Jagadeeswaran et al., 2009
	miR172	AL383429	AP2-like	Apetala-2	pot	Jagadeeswaran et al., 2009
	miR390	-	TAS3 precursor	ta-siRNA precursor	-	Jagadeeswaran et al., 2009
	miR393	AC133780_22.2	TIR1-like	F-box, bZIP	-	Jagadeeswaran et al., 2009
	miR395	AC146721_16.4	Sulfate transporter	Sulfate transporter, ATP-sulfurylase	-	Jagadeeswaran et al., 2009
	miR397	AC203224_21.2, AC135467_30.2	Laccase-like	Laccase	-	Jagadeeswaran et al., 2009
	miR408	BG583436, AC161863_13.2	Cu-containing	Laccase, plantacyanin[3]	-	Jagadeeswaran et al., 2009
	miR399	AC144658	UBC E2-like	UBC E2-like, P transporter	pot	Jagadeeswaran et al., 2009
	miR2118	AC202360_18.1, AC143338_38.2, AC203224_17.1	TIR-NBS-LRR	-	-	Jagadeeswaran et al., 2009
P. vulgaris	**miR156**	17153936	SPL-related	SBP	-	Arenas-Huertero et al., 2009
	miR164	171589244	NAC1-related	NAC	-	Arenas-Huertero et al., 2009
	miR171	62704692	SCL6-related	Scarecrow-like	-	Arenas-Huertero et al., 2009
	miR172	171595931	AP2-related	Apetala-2	-	Arenas-Huertero et al., 2009
	miR319	6271620	TCP-related	TCP	-	Arenas-Huertero et al., 2009
	miR393	171655172	TIR1-related	F-box, bZIP	-	Arenas-Huertero et al., 2009
	miR396	17162341	GRL2-related	GRF	-	Arenas-Huertero et al., 2009

TABLE I-2. Validation of miRNA targets by 5'RACE in legume species
[1] Confirmed targets not conserved in other plant species
[2] Reviewed in Jones-Rhoades et al. (2006)
[3] Abdel-Ghany et al. (2008)
[4] Role in legume-*Rhizobia* Root Nodule Symbiosis. miR398 is also a potential symbiotic miRNA, though target confirmation is still lacking (Lelandais-Brière et al., 2009).
ARF, Auxin Response Factor; *At*, *A. thaliana*; bZIP, basic Leucine-Zipper; GRF, Growth Regulator Factor; pot., potentially; SBP, SQUAMOSA promoter binding protein; SCL, Scarecrow-like; TF, Transcription Factor.

5. MicroRNAs and Symbiotic Nitrogen Fixation

5.1. *M. truncatula* miR166 and miR169 are involved in nodule development

Preceding the increasing wave of small RNA cloning in diverse plant species, two microRNAs have recently found as involved in root nodule symbiosis. The first miRNA implicated in *Rhizobia* symbiosis was *M. truncatula* miR169. Combier and co-workers identified a CCAAT-binding transcription factor, MtHAP2-1, and its regulatory miRNA, miR169, as crucial for nodule development. Target RNAi was phenocopied by miRNA overexpression, leading to arrest of both nodule growth and bacterial release from infection threads. Besides, MtHAP2-1 and miR169 showed non-overlapping expression patterns in meristematic and infection zones, respectively, suggesting that this miRNA is a main player in the spatial restriction of MtHAP2-1 expression and cell differentiation within the nodule (Combier et al., 2009). On a different miRNA study, it was confirmed once again that lateral root architecture and nodule development share some common features (Hirsch, 1992). Indeed, miR166 was found to target several *HD-ZIP III* transcription factors in *M. truncatula*. Accumulation of this miRNA was restricted to vascular bundles and apical regions from both roots and nodules, whereas miRNA overexpression led to reduced lateral rooting and nodule number. Additionally, expression domains of miR166 and the *HD-ZIP III* genes were overlapping, indicating that miR166 contributes to the fine-tuning of the levels of these transcription factors (Boualem et al., 2008). Thus, two conserved miRNAs were found as involved in nodule development and/or nodule cell differentiation.

5.2. Candidate miRNAs on the pipeline

Deep sequencing of small RNAs from the past two years have also contributed to delineate a set of miRNAs as candidates for their involvement in symbiosis, though their biological relevance still remains elusive. Early symbiotic studies were undertaken in soybean, where *Bradyrhizobium japonicum*-inoculated roots at 3hpi were subjected to small RNA cloning

(Subramanian et al., 2008). Several miRNAs showed slight changes in cloning frequencies when comparing inoculated versus mock-inoculated small RNA libraries, and these differences could be reproduced at some extent in northern blot assays. According to the authors, distinct patterns of expression were observed, including transient up-regulation at 1-3hpi (miR168, miR172), induction and further maintenance of high accumulation (miR159, miR393) and down-regulation (miR160, miR169, miR396). Further studies will help to validate these patterns and investigate the role of these miRNAs in early signal transduction in soybean symbiosis. Still in soybean, a small-scale sequencing initiative from 28dpi determinate nodules showed that the conserved miR167, miR172, miR398 and miR399, plus several legume-specific miRNAs, are accumulated in this tissue. Among them, miR172 is profiled as the highest *MIR* expressed in nodules, as compared to other tissue types analyzed (Wang et al., 2009).

As part of a comprehensive and bioinformatically solid small RNA cloning study, Lelandais-Brière and colleagues have recently uncovered the landscape of mature nodule-specific miRNA genes and family isoforms in *M. truncatula.* Cloning data was supported with molecular detection and *in-situ* hybridization assays in a way that consistent candidates have been defined. A set of miRNAs is expressed in the meristematic region of mature nodules: miR160, miR167, miR172, miR398 and the two novel miRNAs mtr-2586 and mtr-107. In addition, Auxin Response Factor-controlling miR167 is also expressed in the nodule vascular bundles, whereas miR172 and miR398 accumulate at the differentiation/infection zones. A sole miRNA, miR399, is found at the nitrogen-fixing zone (Lelandais-Brière et al., 2009).

Hence, small RNA deep sequencing coupled with comparative molecular studies has emerged as a powerful approach to pinpoint symbiosis-related miRNAs. Interesting questions arise now, as to whether these miRNAs are general regulators of gene expression in nodules or specific for a certain nodule type. Besides, are these miRNAs exclusively related to organogenesis or are they involved to any extent in the intracellular accomodation of the endosymbiont or N_2 fixation itself? Indeed, functional analysis will ultimately

reveal the real biological contribution of these miRNAs to any of the steps of legume-*Rhizobia* symbiosis.

II. AIM OF THE STUDY:

This work was divided into three major parts. The main objectives were (i) the elucidation of the small 'RNAome' of the roots and nodules of the determinate model legume *L. japonicus*, and (ii) the assessment of its potential contribution to the various facets of Root Nodule Symbiosis (RNS), focusing our study on microRNAs and related small RNAs. As a third line of work, efforts were deployed to establish a genetic framework for the investigation of small RNAs in legumes by assessing the diversity of RNA silencing pathways in *L. japonicus* and by generating genetic tools for molecular studies.

III. RESULTS & DISCUSSION

CHAPTER 1. A framework for the study of microRNAs and other small RNAs in legumes

1.1. RESULTS

1.1.1. The RNA silencing machinery in *L. japonicus*

At the start of this project, the genome sequence of *Lotus japonicus* was not available to the public. Through authorized access to the partially sequenced and annotated genome data at Kazusa Institute (www.kazusa.or.jp/lotus/), we initiated an exhaustive bio-informatic survey of the main RNA silencing processor and effector genes in *Lotus japonicus*, with a particular emphasis on the microRNA pathway. *Lotus japonicus* genomic data, obtained both by clone-to-clone and shotgun sequencing, was published in 2008. It covers 67% of the 472Mb genome and 91,3% of the gene space (Sato et al., 2008). Using the corresponding available gene predictions and previously in-house generated gene models (see Materials & Methods) we carried out a characterization of RNA silencing-related genes in *Lotus*.

1.1.1.1. Identification of DICER-LIKE genes

We first concentrated our efforts on the analysis of the DICER-LIKE (DCL) genes in *Lotus japonicus*. Using the 4 members of this RNAse-III protein family found in *Arabidopsis thaliana*, we identified 5 *DCL* paralogs in the *Lotus japonicus* genome (Table R1). Predicted genes sequences were translated and analyzed phylogenetically using the maximum likelihood-based algorithm, PhyML 3.0 (Guindon et al., 2003) with a Dayhoff matrix model for aminoacid substitution and bootstrap with 100 replications. Phylogenetic trees showed that the five identified Lotus DCL genes strongly align to the different DCL paralog groups from rice, poplar and *Arabidopsis* (Figure R1-A), showing that *Lotus japonicus* genome harbors single *DCL1*, *DCL3* and *DCL4* genes, plus a double copy of the *DCL2* gene. As can be seen from the figure, *LjDCL2a*

and *LjDCL2b* are very closely related. The predicted proteins show 85,3% aminoacid similarity from their predicted sequences and their nucleotide sequences are organized in direct tandem, only 3,3 Kb apart on chromosome 6 (data not shown). These are indications of a recent gene duplication. All five Lotus *DCLs* are effectively transcribed, as demonstrated by 100% matching to Expressed-Sequence Tags (ESTs) clones (Table R1).

1.1.1.2. ARGONAUTE-LIKE genes in *L. japonicus*

The *Arabidopsis* genome contains 10 Argonaute-like (AGO) genes that are phylogenetically grouped into 3 clades (Vazquez, 2006; Vaucheret, 2008). We searched the *Lotus japonicus* genome and found 9 members of this PAZ/PIWI domain-containing protein family. Phylogenetic analysis (Figure R1-B) showed that Lotus encodes clear orthologs of the canonical *AtAGO1*, *AtAGO5* and *AtAGO10*, which constitute the first clade. In contrast, Lotus members of the AGO4/6/8/9 clade include 3 closely related proteins whose identity cannot be assigned easily by sequence homology to their Arabidopsis counterparts. Consequently, these were temptatively named *LjAGO4*-like 1 to 3. In *Arabidopsis* there is a third Argonaute clade formed by *AGO2, 3* and *7*. We could find a clear *LjAGO7* gene candidate together with 2 predicted genes close to *AtAGO2*, hereby called *LjAGO2*-like genes *1* and *2*. The former, similarly to its *Arabidopsis AGO2* and *AGO3* counterparts, does not contain the DDH motif in the PIWI domain (data not shown), a domain associated with the slicer activity (Vazquez, 2006; Tolia and Joshua-Tor, 2007). The latter was found in an unassembled genomic contig where the PIWI domain was deleted and, hence, could not be analyzed. Expressed Sequence Tags were found for all these AGO-like genes except for *LjAGO7* and *LjAGO2-like2* (Table R1). However, a recent forward genetics study identified the functional *LjAGO7* gene (renamed *LjREL3*) and confirmed our bioinformatic predictions on this gene (Yan et al., 2009). Besides, we could identify 3 PIWI domain-containing shotgun clones (Table R1). No further genomic data from these sequences was retrieved; therefore further sequencing of these clones (in progress) or RACE analysis will indicate whether these are actually part of additional Lotus *AGO-like* genes. ESTs could be found for at least 2 of these PIWI proteins, suggesting their genuine expression (Table R1).

1.1.1.3. Other silencing effectors in *L. japonicus*

RNA-dependent RNA-polymerases (RDRs) form also an important protein family for RNA silencing pathways in plants and some metazoans. Of the 6 *RDR* genes identified in *Arabidopsis*, only *RDR2* and *RDR6* functions have been characterized in depth (Chapman and Carrington, 2007), yet there are now clear evidences for an antiviral role for *RDR1* (Qi et al, 2009). Using the Arabidopsis protein sequences and the previously identified *Medicago truncatula RDR* genes (Zong et al., 2009), we could identify *LjRDR2* and *LjRDR6* genes. ESTs could be found for the former, but not *LjRDR6*, maybe because it was found on a partial shotgun sequence (LjSGA_072772.1; Table R1). Due to the short length of the other *RDR-like* clones found, a systematic classification of all Lotus *RDR* genes could not be carried out.

As our main interest was focused on microRNAs, we proceeded to identify the suite of enzymes known to be involved in the canonical *Arabidopsis* miRNA pathway. Table R1 provides a view of Lotus-encoded genes corresponding to the forkhead-associated domain protein DAWDLE (LjDDL), the zinc finger-C2H2 protein SERRATE (LjSE), the methyltransferase LjHEN1 and the exportin-5 homolog HASTY, all found to be involved in the miRNA pathway in plants (Voinnet, 2009). In parallel, a shotgun clone sequence was found to correspond to the 5'-end sequence of *LjHYL1*. The recently identified exonuclease SMALL RNA DEGRADING NUCLEASE (SDN)(Ramachandran and Chen, 2008) was also identified in Lotus: a single gene located on chromosome 1, named *LjSDN*, is highly similar to all 3 Arabidopsis *SDN*s (Table R1). All the above full or partial predicted protein sequences contained the catalytic domains specific to their cognate enzymatic role, as revealed by analysis using the InterPro databases.

Table R1. RNA silencing machinery genes identified in *Lotus japonicus*. *Arabidopsis thaliana* genes were blasted against updated genomic data available at Kazusa Institute (www.kazusa.or.jp/lotus/blast.html). Predicted cds were translated and scanned against the InterPro database (www.ebi.ac.uk/InterProScan). Corresponding 100% matching *L. japonicus* ESTs were obtained at the Gene Index Project interface (compbio.dfci.harvard.edu), using last version from May 25, 2009. *Lj*, *Lotus japonicus*; *At*, *Arabidopsis thaliana*; *DCL, DICER-LIKE; AGO, ARGONAUTE-LIKE; RDR, RNA-dependent RNA-polymerase; HEN1, HUA ENHANCER 1; HYL1, HYPONASTIC LEAVES 1; SE, SERRATE; DDL, DAWDLE; SDN, SMALL RNA DEGRADING NUCLEASE; dsRBD, double-stranded RNA-binding*.

Lj genes	Close At genes	Lj predicted genes at Kazusa Institute (2009)	InterPro predicted protein domains	Lj Gene Index 5.0 (2009)	Lj Affymetrix Chip IDs
LjDCL1	*AtDCL1*	chr1.CM0105.760.nc, chr1.CM0105.750.nc	DEAD/DEAH, Helicase-C, Dicer, PAZ, RNAseIII (2), dsRBD (2)	BP042378	chr1.CM0105.70(ID#1), chr1.CM0105.68(ID#2)
LjDCL2a	*AtDCL2*	LjT40C18.tp90, chr6.CM0437.20.n, chr6.CM0437.180.nd	DEAD/DEAH, Helicase-C, Dicer, PAZ, RNAseIII (2)	TC20681	chr6.CM0437.1.2
LjDCL2b	*AtDCL2*	LjT40C18.tp100	DEAD/DEAH, Helicase-C, Dicer, PAZ, RNAseIII (2)	BP053643	chr6.CM0437.3(ID#1), chr6.CM0437.7.1(ID#2)
LjDCL3	*AtDCL3*	chr1.CM0133.910.nd, chr1.CM0133.900.nd, chr1.CM0133.920.nd	DEAD/DEAH, Helicase-C, Dicer, PAZ, RNAseIII (2), dsRBD (1)	BP042852, TC29631	chr1.CM0133.54.1(ID#1), chr1.CM0133.54.2(ID#2)
LjDCL4	*AtDCL4*	LjSGA_047097.1	DEAD/DEAH, Helicase-C, Dicer, PAZ, RNAseIII (2), dsRBD (2)	BP062906, TC21684, AV426701	Ljwgs_047097.1
LjAGO1	*AtAGO1*	LjSGA_035516.1, LjSGA_014186.1	PAZ, PIWI	TC33857, TC20928, AU088885, TC30303, AV410769	Ljwgs_014186.1(ID#1), Ljwgs_035516.1(ID#2)
LjAGO2-like1	*AtAGO2/3*	chr4.CM0229.30.nc	PAZ, PIWI	BW598563, TC26781	chr4.CM0229.1(ID#1), chr4.CM0229.3(ID#2)
LjAGO2-like2	*AtAGO2/3*	chr6.CM0066.110.nd	PAZ, PIWI(*)	-	chr6.CM0066.14
LjAGO4-like1	*AtAGO4/6/8/9*	chr2.CM0031.300.nc	PAZ, PIWI	TC31337, TC28097, TC21711	chr2.CM0031.33
LjAGO4 -like2	*AtAGO4/6/8/9*	chr6.LjT28D11.70.nc	PAZ, PIWI	TC22027, AV407626, TC21452, AV421492	chr6.TM1650.9.2
LjAGO4-like3	*AtAGO4/6/8/9*	CM1092.250.nc	PAZ, PIWI	TC25339, BP052529	Ljwgs_030852.1
LjAGO5	*AtAGO5*	chr3.CM0396.280.nd	PAZ, PIWI	TC20447, BW595647, BW598349, CB828396, TC20727	chr3.CM0396.7
LjAGO7	*AtAGO7*	chr2.CM0641.180.nc	PAZ, PIWI	-	Ljwgs_040155.1(ID#1), Ljwgs_023703.1(ID#2)
LjAGO10	*AtAGO10*	chr5.CM0357.210.nc	PAZ, PIWI	TC22726	Ljwgs_019180.1
LjPIWI1	*AtAGO1/5/10*	LjSGA_026639.1	PIWI	TC28627	Ljwgs_026639.1
LjPIWI2	*AtAGO4/6/8/9*	LjSGA_005241.2	PIWI	TC29214	Ljwgs_005241.2
LjPIWI3	*AtAGO4/6/8/9*	LjSGA_017462.1	PIWI	-	-
LjRDR2	*AtRDR2*	LjT28B05.70.nc	RNA recognition motif (RNP-1), RNA-dependent RNA polymerase	TC27701, TC22266	TM1762.14
LjRDR6	*AtRDR6*	LjSGA_072772.1	RNA-dependent RNA polymerase	-	Ljwgs_072772.1
LjHASTY	*AtHASTY*	chr1.CM0544.60.nc	Exportin-1/Importin-beta-like	TC34976	chr1.CM0544.9
LjHEN1	*AtHEN1*	chr4.CM0007.1100.nc, chr4.CM0007.1070.nc	dsRBD, Methyltransferase type 11	AV409723, TC30178	chr4.CM0007.31
LjHYL1	*AtHYL1(DRB1)*	LjSGA_028943.1	Double-stranded RNA binding	CB826912	Ljwgs_028943.1
LjSE	*AtSE*	LjT02O17.70.nc	Zinc finger C2H2 type	BP051467, CN825698, AV773404	TM1725.11
LjDDL	*AtDDL*	LjT16B12.70.nd	Forkhead-associated (FHA)	-	-
LjSDN	*AtSDN1/2/3*	chr1.CM0378.180.nd	Exonuclease	-	chr1.CM0378.23

Table R1. RNA silencing machinery genes identified in *Lotus japonicu*

1.1.1.4. Expression changes in RNA silencing factors in symbiotic conditions

The availability of sequences from the main RNA silencing effectors of Lotus allowed us to analyze their expression, and possible contribution, along the process of RNS. To that aim, we took advantage of an available study that employed the *L. japonicus* Affymetrix GeneChip®. This study describes the transcriptome data obtained in 38 different inoculation conditions and Lotus mutants. It also includes a highly stringent statistical analysis from three spatially and temporally separated biological replicates (Hogslund et al., 2009). Transcriptome changes in the RNA silencing machinery were not significant when roots from early times post-inoculation or from Lotus early RNS signal transduction mutants were compared to non-inoculated roots (statistical significance criterion of FDR corrected p-value ≤ 0,05). By contrast, we found a consistent up-regulation of Probe Sets (≥ 2-fold) corresponding to a set of RNA silencing genes in both 14dpi and 21dpi nodules when compared to mock-inoculated root tissue (Figure R1-C). This group of genes was formed by *LjDCL2b*, *LjAGO2-like1* and, surprisingly, the Lotus ortholog of the uncharacterized *AtAGO5* gene. Conversely, *LjAGO7* was consistently downregulated, suggesting a root-specific, as opposed to nodule-specific role for this gene. We also analyzed expression patterns of the nodulation mutants *sen1* (Suganuma et al., 2003) and *sst1* (Krussel et al., 2005), whose nodules are colonized by *Rhizobia* but where nitrogen fixation is abolished or reduced at 90%, respectively. *Sst1* expression profiles reproduced our previous observations, whereas the induction of *LjAGO5* expression was reproduced in *sen1* nodules, when compared to *sen1* uninoculated roots (Figure R1-C). This set of data highlights the diversity of RNA silencing genes that might be involved in RNS, particularly in nodule development or bacterial infection, and indicates that gene regulation in root and nodule biology might utilize differential small RNA pathways.

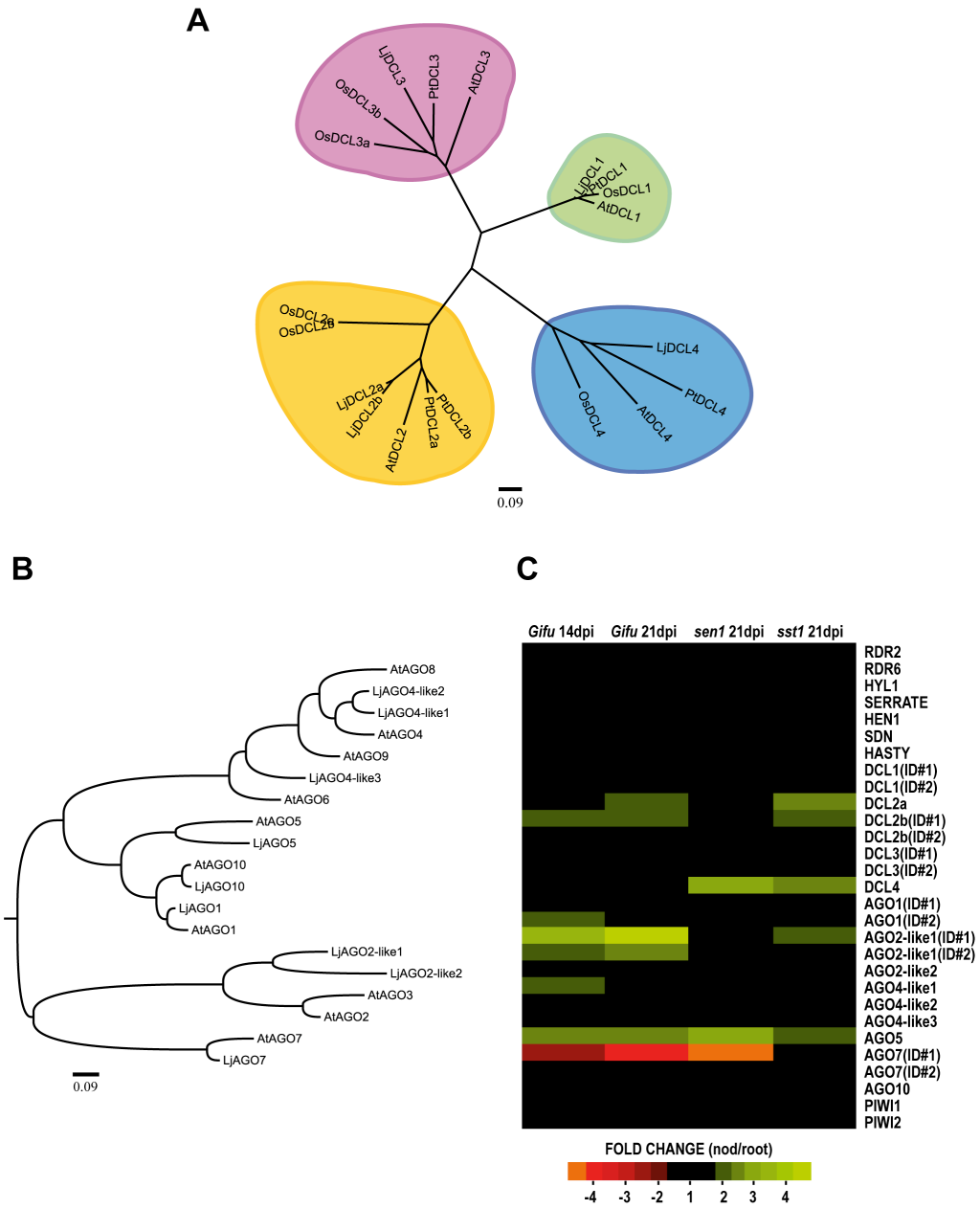


Figure R1. *L. japonicus* DICER and ARGONAUTE-LIKE. A. Phylogenetic analysis of DICER-LIKE proteins from *L. japonicus (Lj), poplar (Pt), rice (Os) and A. thaliana (At)*. The *Pt* and *Os DCL* genes used here were identified by Margis et al., (2006). Two additional and low expressed OsDCL1 paralogs recently identified are not included here (Kapoor et al., 2008). **B.** Phylogenetic tree from argonaute-like proteins in *L. japonicus and A. thaliana.* Amino-acid sequences were aligned using MUSCLE (Edgar et al., 2004) and analyzed using PhyML 3.0 (Phylogeny.fr; bootstrap with 100 replications). **C.** Fold changes in expression of *Lotus* RNA silencing genes in nodules as compared to uninoculated roots of the same genotype. For each respective Affymetrix ID (Table R1), fold change values were calculated as the ratio of normalised signal intensity values from nodules vs. uninoculated roots (False Discovery Rate (FDR) corrected p-value ≤ 0,05) (Hogslund et al., 2009). Only ≥ 2-fold changes are represented in the figure.

1.1.1.5. TILLING: an attempt to knock-down DICER-LIKE genes in *Lotus japonicus*

Since *Lotus japonicus* was established as a model legume (Handberg and Stougaard, 1992), diverse tools for forward and reverse genetics have been generated. Accordingly, the Sainsbury Laboratory generated a large population of M2 plants subjected to ethyl metane-sulfonate (EMS)-mutagenized TILLING (Targeted Induced Local Lesions in Genomes; Perry et al., 2003). This population of mutants was used to screen for lesions affecting the above-mentioned *LjDCL1*, *LjDCL2s* and *LjDCL4* genes. The fact that the sequenced Lotus genome contains mostly gene-rich regions (Sato et al, 2008) precluded us from searching for *LjDCL3* mutants or analyzing heterochromatic siRNAs. *DICER-LIKE* genes are long, multi-domain genes whose coding sequences, at least in Lotus, range from 4 to 5kB. To increase the chances of obtaining knockdown lines, we screened the gene regions corresponding to the nucleic acid-binding fold PAZ domains (Lingel et al., 2003; Yan et al., 2003).

The various mutant lines obtained are presented in Table R2 below. The nucleotide point mutations were analyzed using actual CDS nucleotide stretches obtained by sequencing cDNA clones (www.legumebase.agr.miyazaki-u.ac.jp). Screening of the PAZ domain yielded two independent missense alleles in which conserved aminoacids had been substituted in *LjDCL1* (Figure R2-A). Though these lines did not show any obvious developmental phenotype, accumulation of diverse miRNAs was studied (Figure R2-B). Unfortunately, no significant differences were found in miRNA levels when comparing the homozygous mutants to heterozygous or wild-type individuals. Given this result, *Ljdcl1-1* and *Ljdcl1-2* mutations do not seem to interfere at least with DCL1 processing activity and, thus, further backcrosses were not carried out.

Regarding *LjDCL2*, three missense mutants and one splicing mutant were found, all corresponding to the *LjDCL2a* paralog. Both homozygous and heterozygous *Ljdcl4* mutants were simultaneously found. Genotyping and molecular characterization of these lines is currently ongoing. As a parallel reverse genetics approach, we also generated RNAi constructs to knock-down

LjDCL1 and *LjDCL2* genes, though their *in vivo* root expression did not give successful results.

Mutant name	Mutant line	Mutation type	Copies
Ljdcl1-1	SL693-1	missense	heterozygous
Ljdcl1-2	SL2749-1	missense	heterozygous
Ljdcl2-1	SL595-1	missense	heterozygous
Ljdcl2-2	SL1758-1	splice site	heterozygous
Ljdcl2-3	SL1702-1	missense	heterozygous
Ljdcl2-4	SL1818-1	missense	heterozygous
Ljdcl4-1	SL2015-1	missense	homozygous
Ljdcl4-2	SL1130-1	missense	heterozygous

Table R2. DICER-LIKE TILLING mutants found in *L. japonicus.*

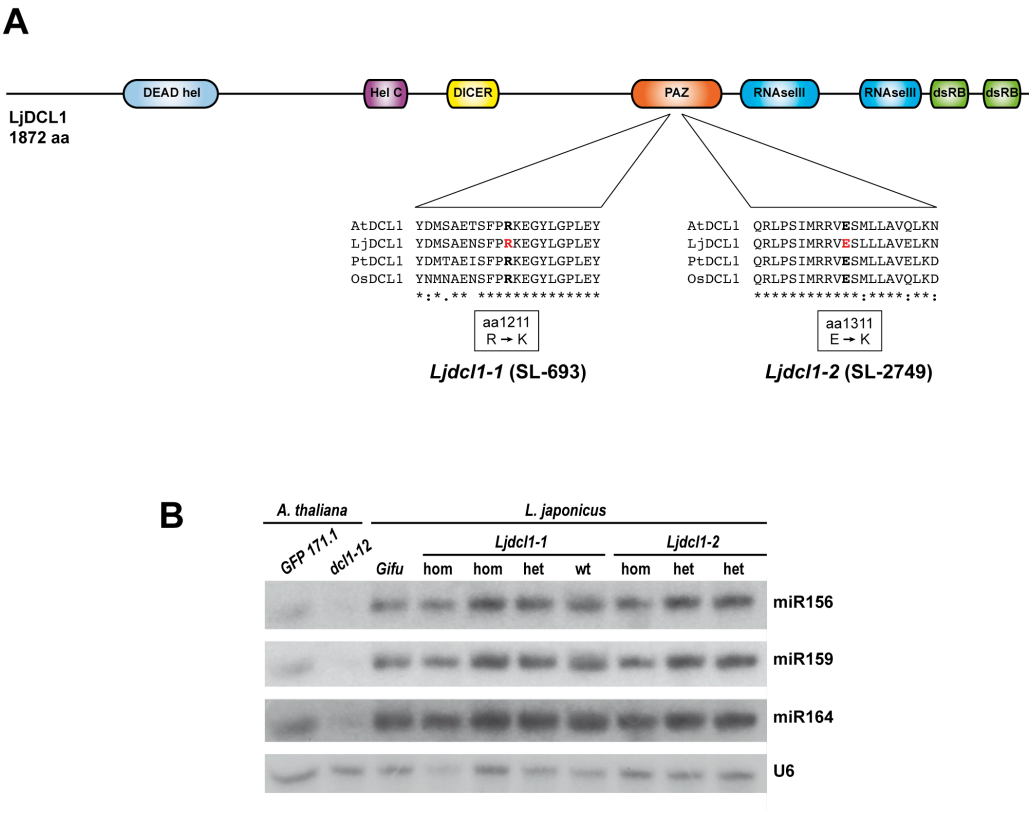


Figure R2. Analysis of TILLING Ljdcl1 mutants. A. *L. japonicus* TILLING *dcl1* mutants in the PAZ domain. Protein domains depicted as predicted by Pfam (pfam.sanger.ac.uk). **B.** miRNA accumulation in *Ljdcl1-1* and *Ljdcl1-2* mutant lines. Homozygous (*hom*), heterozygous (*het*) and wild-type (*wt*) individuals from each of the mutant lines were analyzed. *A. thaliana dcl1-12* EMS mutant and its parental line GFP 171.1 (Brodersen et al., 2008) were used as controls. Hybridization with U6 was used as a loading control. Quantification and normalization of Northern blot signals (ImageJ64 version 10.2) showed no significant differences of mutants versus wild-type or heterozygous individuals. *DEAD hel*, DEAD/DEAH box helicase; *Hel C*, helicase conserved C-terminal domain; *DICER*, dicer double-stranded RNA-binding domain; *RNAseIII*, ribonuclease III family domain; *dsRB*, double-stranded RNA-binding domain; *SL*, Sainsbury Laboratory mutant lines (Perry et al., 2003).

1.1.2. Expression of the P19 viral silencing suppressor in *L. japonicus*

As the strategies undertaken to knockdown *LjDCL1* were unsuccessful, we decided to express the *Tomato bushy stunt virus* (TBSV) silencing suppressor P19 (Voinnet et al., 1999) in *L. japonicus,* as a means of interfering with miRNA activity. Indeed, this tombusviral protein sequesters *in vivo* 21bp small RNA duplexes, including miRNA/miRNA* pairs and, thus, interferes with their activity (Silhavy et al., 2002; Vargason et al., 2003; Ye et al., 2003; Dunoyer et al., 2004).

1.1.2.1. Generation of P19 transgenic lines

TBSV P19 was cloned under the CaMV 35S promoter into the Gateway® vector pK7WG2D (Figure R3-A), which includes GFP, a useful visual marker for plant transformation. Though the 35S promoter has been widely used in *L. japonicus* molecular symbiotic studies (Handberg and Stougaard, 1992; Charpentier et al., 2008; Yano et al, 2008) it has been shown to be weak in roots and nodules, when compared to other promoters (Maekawa et al., 2008). Therefore, to confirm and enable monitoring 35S-driven expression in our system, the β-glucuronidase (GUS) was used for generating control plant lines (Figure R3-A). Root explants from the Miyakojima-20 (MG20) accession were transformed with these constructs. As previously described (Handberg and Stougaard, 1992), along the transformation process we could observe that a percentage of untransformed plants were able to grow on kanamycin medium, which made the selection of *bona-fide* transgenic lines more difficult. Nevertheless, we took advantage of GFP-based visual selection and, thus, several independent lines could be obtained (Figure R3-B). T1 primary transformants P19 #8 and #13 and control GUS #1 and #2 were selected for their higher transgene expression. Next, T2 lines derived from these primary transformants were selected: GUS#1.6, GUS#2.2, P19#8.2 and P19#13.4. Transgene inheritance in these T2 lines followed mendelian 3:1 segregation (χ^2<3,84) except for P19#13.4, homozygous for the transgene insertion locus. For the following phenotypical and molecular analyses, pooled GFP-positive progeny from these selected T2 lines was analyzed. In parallel, X-Gluc staining performed on the GUS lines (Figure R3-C) confirmed that the 35S promoter is expressed in leaf, root and nodule tissues, thus enabling the study of P19 action in transgenic plants.

1.1.2.2. Characterization of P19 transgenic lines

Lotus P19-expressing lines did not show any leaf morphology phenotype, compared to the characteristic leaf serration found in *Arabidopsis* P19 lines (Dunoyer et al., 2004). Nevertheless, the shoot size of *Lotus* P19 plants was considerably smaller than the GUS control lines (Figure R3-D and E): statistical analysis revealed highly significant differences between the GUS

control and P19 experimental groups (Unpaired two-tailed *t*-test; P-value<0.0001). By contrast, root length measurements showed no statistical differences (Unpaired two-tailed *t*-test; P-value=0.1989). Thus growth reduction was restricted to the aerial tissues. Nodulation tests were also performed and both P19 lines were able to form functional pink nodules (Figure R3-D).

To assay the ability of P19 to interfere with miRNA function *in vivo*, different tissues form the P19-expressing plants were subjected to small RNA analysis (Figure R4-A). We first analyzed miR160 because its ccumulation is high leaf, root and nodule tissues. In all samples tested, the P19-expressing lines showed a strong stabilization of the miR160* passenger strand, which normally undergoes rapid degradation upon its separation from the miR/miR* duplex. This observation is in accordance with the results of previous studies in Arabidopsis (Dunoyer et al., 2004) and strongly suggests that P19 is able to sequester miRNA/miRNA* duplexes *in vivo* in our transgenic system. To test if miRNA sequestration by P19 had an effect on Lotus miRNA activity, we decided to analyze miRNA target gene levels by qRT-PCR. As a control, we confirmed that several conserved miRNA target genes are induced in *Arabidopsis* P19 plants as compared to wild-type, including the highly conserved targets of miR160, miR164 and miR167 (F. Jay, unpublished data). To enable our analysis in Lotus, we carried out target prediction from the Lotus orthologs of the above-mentioned miRNAs (Target Finder; Allen et al., 2005; Falhgren et al., 2007). We selected those predicted targets showing the highest sequence homology to conserved targets found in *Arabidopsis* and *Medicago*. *Arabidopsis* miR160 regulates Auxin Response Factors (ARFs) *AtARF16* (At4g30080) and *AtARF17* (At1g77850), among others (Jones-Rhoades et al, 2006), and similar *ARF*-related ESTs (ES612384; BQ148941) have been confirmed as miR160 targets in *M. truncatula* (Jagadeeswaran et al, 2009). Through sequence analysis we isolated a *LjARF16-like* gene (chr3.LjT10E18.60.nc) harboring the highly conserved miRNA target sequence for miR160. miR164 targets NAC domain containing genes, including *NAC1* (At1g56010) and *ANAC08* (At5g07680) in Arabidopsis (Jones-Rhoades et al, 2006) as well as AC203553_1 in Medicago (Jagadeeswaran et al, 2009).

Corresponding Lotus NAC-domain genes were found in accessions chr3.CM0279.430.n (*LjNAC-like1*) and chr4.CM0087.800.n *(LjNAC-like2*). Lastly, we retrieved a *LjARF6-like* gene (chr4.LjT16L01.06.nd) which is highly similar to miR167 targets *AtARF6* (At1g30330), *AtARF8* (At5g37020) and *Medicago ARF-like* ESTs (AC14478_44; CU326393_14) (Jones-Rhoades et al, 2006; Jagadeeswaran et al., 2009). In all three cases, the 20-22t miRNA target sites were highly conserved (Figure R4-B). Thus, primers were designed either upstream or at both sides of the predicted conserved cleavage site and data were normalized to the three housekeeping genes encoding the GPI-anchored protein (LjGPI), ubiquitin-conjugating enzyme 10 (LjUBC10) and polyubiquitin (LjUBIQ) (Sanchez et al., 2008). qRT-PCR data revealed that both P19 lines tested consistently showed a 2-fold increase in the miR160 and miR164 predicted target transcript levels, whereas *LjARF6-like* mRNA levels remained unchanged (Figure R4-C). These expression differences, though modest, are in accordance with the 2-fold increase in miRNA target expression found in *Arabidopsis* P19, HcPro and P15 transgenic lines (F. Jay, unpublished) and in Arabidopsis *dcl1-7* mutants (Xie et a., 2005). The above results indicate that miRNA activity is at least partially impaired in Lotus P19 lines, which therefore, constitutes a potentiall tool for miRNA validation in legumes.

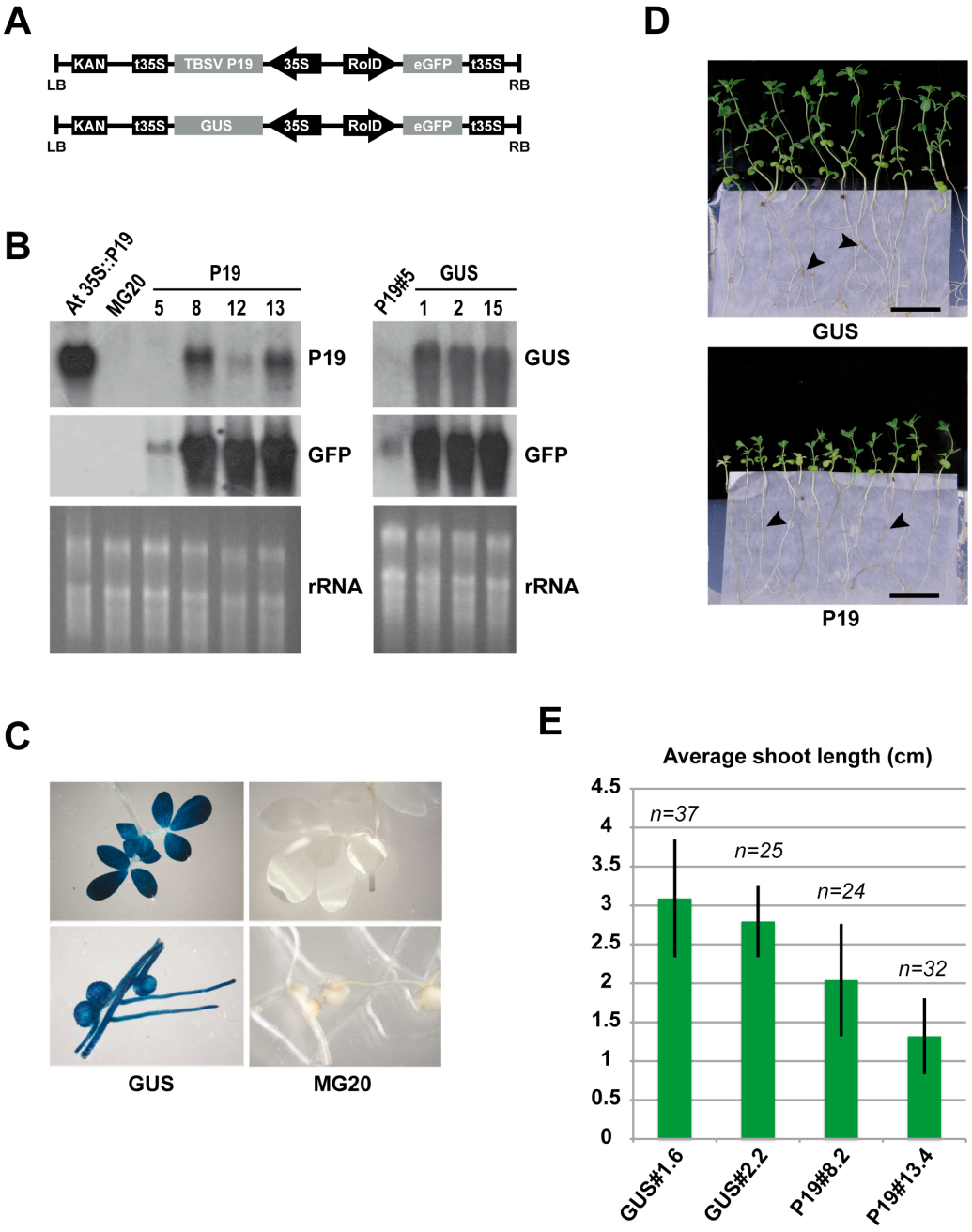


Figure R3. Characterization of Lotus japonicus P19 transgenic lines. A. Schematic of pK7WG2D vector expression cassette used for plant transformation. Arrows represent promoter regions. **B.** HMW Northern blot analysis of transgene expression. GFP-positive progeny seedlings from indicated independent lines were analyzed. **C.** X-Gluc histochemical staining of leaf and nodulated roots from 35S::GUS and wild-type MG20 plants. **D.** Phenotype of 5 week-old P19 and GUS plants. Plants were grown on N_2-media where nodules were the only N_2 source. Black arrows show mature pink nodules. Scale bars represent 2 cm. **E.** Measure of plant shoot length. GFP-positive offspring from indicated parental T2 lines was analyzed. Distance from cotyledons to last leaf node was measured. Error bars show standard deviation from the population measured. **F.** Nodule number per plant. (n). KAN, kanamycin; GUS, beta-glucuronidase; t, terminator; LB, left border; RB, right border; *At; A. thaliana; Lj; L. japonicus*; rRNA, ethidium bromide staining of ribosomal RNAs.

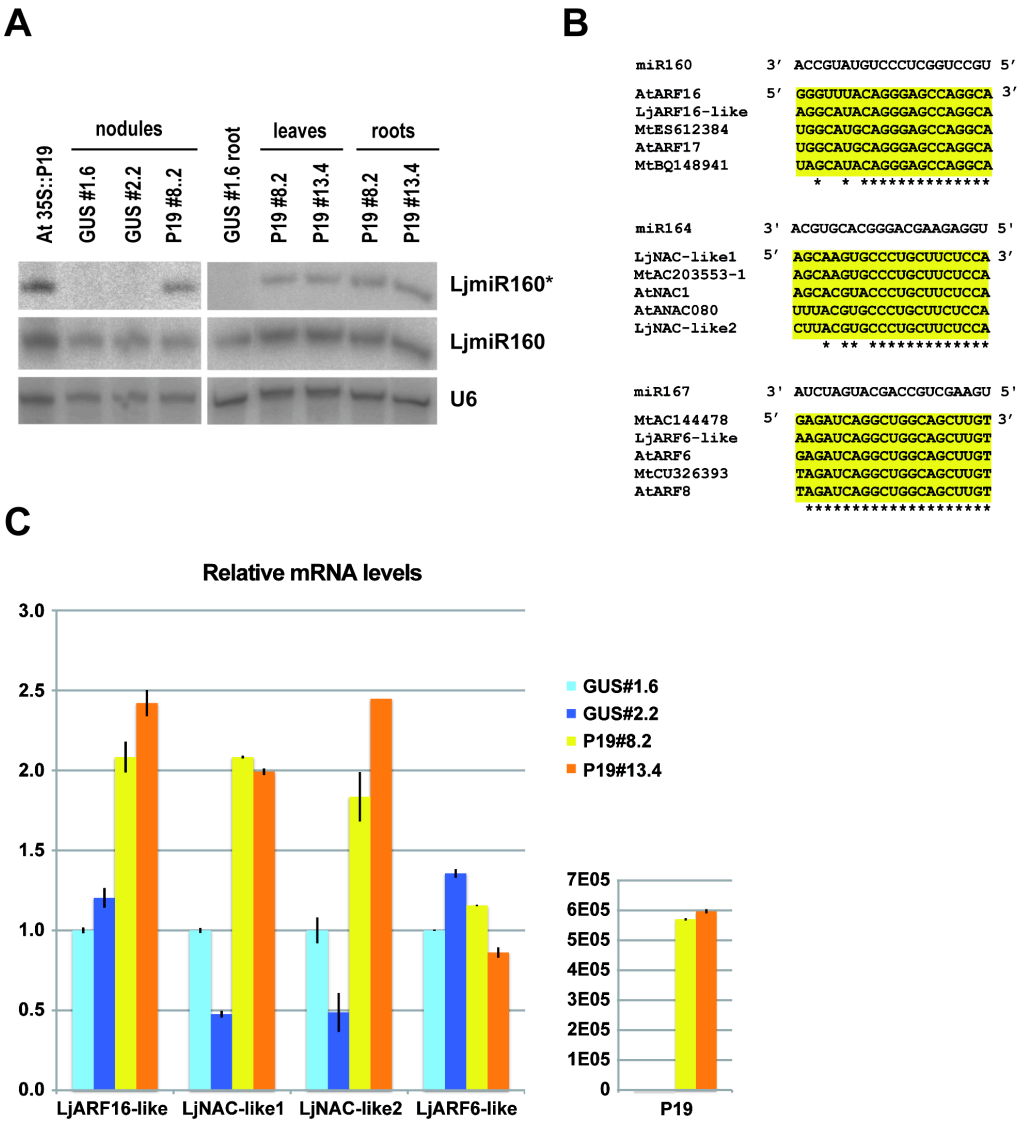


Figure R4. P19 is active in Lotus japonicus transgenic lines. A. Accumulation of passenger miRNA* strand. Different tissues were obtained from pooled 7 week-old progeny. 5wpi nodules were harvested. U6 provides a loading control. **B.** Sequence comparison of several conserved miRNA and miRNA target genes. Target mRNA fragments are shown on 5'-3' orientation and gene accession numbers are given on the text. **C.** Quantitative RT-PCR analysis of several conserved miRNA target genes. Leaf tissue from pooled 7 week-old plants was analyzed. Primers were designed either 5' upstream (*LjNAC-like2, LjARF6-like*) or at both sides of conserved cleavage site (*LjARF16-like, LjNAC-like1*). Transcript levels were normalized to that of 3 housekeeping genes *LjUBC10, LjGPI* and *LjUBIQ* and then to the value of GUS#1.6 line, arbitrarily set to 1. Error bars represent the standard deviations from three PCR results, and similar results were obtained in two independent experiments. *Lj, Lotus japonicu; At, Arabidopsis thaliana; Mt, Medicago truncatula.*

1.2. DISCUSSION

1.2.1. Diversity of the RNA silencing machinery in *L. japonicus*

Through thorough bioinformatic data mining, we have uncovered the set of RNA silencing effector genes encoded by the available *L. japonicus* genome, which covers more than 90% of the gene space (Sato et al., 2008). *DCL1* and *DCL4* remain unigenic in Arabidopsis, poplar and Lotus, while *DCL2* and *DCL3* orthologs have undergone a higher degree of diversification (Margis et al., 2006). A recent study has identified two additional copies of *DCL1* in rice, though they are low expressed and specific to panicles and seeds, as compared to ubiquitously and highly expressed *OsDCL1a* (Kapoor et al., 2008). These observations can be rationalized by the following ideas. New mechanisms for miRNA feedback regulation are currently being discovered. These include miRNA- and/or siRNA-mediated modulation of *DCL1* and *AGO1* expression in Arabidopsis (Rajagopalan et al., 2006; Mallory and Vaucheret, 2009) or the so-called Target Mimicry (Franco-Zorrilla et al, 2007; Voinnet, 2009), reflecting the importance of a tight control of the miRNA gene regulatory pathway. Therefore, one could speculate that a higher degree of sequence diversity or increase in paralog numbers in *DCL1-LIKE* genes would not favor the efficiency of feedback control mechanisms, at least at the processing step. On the other hand, Lotus, poplar and rice encode 2 copies of *DCL2*. Gene duplications at the origin of these paralogs are estimated to have ocurred 1 million years ago (mya) and 8-13 mya, respectively (Margis et al, 2006), that is, independently after the separation of the eudicots and monocots groups (Soltis et al, 1999). The established antiviral role of DCL2 (Deleris et al, 2006) could explain this observation, because diversification of the *DCL2* gene family would confer species-specific defensive advantage against rapidly evolving proteins encoded by viruses to counteract RNA silencing (Chapman and Carrington, 2007; Ding and Voinnet 2007). Rapid evolution and diversification of *DCL* genes as a result of defensive strategies, as opposed to *DCL* gene maintenance, a consequence of cellular regulatory purposes, has been best documented in Drosophila (Obbard et al., 2006).

Arabidopsis Argonautes are the most characterized plant *AGO* genes. Studies have revealed a role for *AGO1* and *AGO10* in post-transcriptional gene silencing (PTGS) and miRNA activity (Baumberger and Baulcombe, 2005; Brodersen et al., 2008; Mallory et al., 2009) as well as a role for both *AGO4* and *AGO6* in RNA-directed DNA methylation (RdDM) that may cause transcriptional gene silencing (TGS)(Zilberman et al., 2003; Zheng et al, 2007). However, the biological functions of the other Argonautes remain unknown. In fact, *Arabidopsis ago2, ago3*, *ago5*, *ago8* and *ago9* single mutants do show neither strong developmental phenotypes nor a reduction in miRNA and siRNA levels (Katiyar-Agarwal et al, 2007; Takeda et al., 2008). Identification of Lotus orthologs of these genes therefore provides a useful tool for comparative functional studies. In addition, the phylogenetic analysis of the Lotus Argonaute family confirms the hypothesis that *AtAGO8* and *9* have arisen from a very recent gene duplication in *Arabidopsis* (Vaucheret, 2008), which would have taken place after the separation of the Brassicales and Fabales orders (Soltis et al, 1999). Moreover, Lotus orthologs of *AtAGO4*, *8* and *9* are indistinguishable due to their high aminoacid sequence similarity (Figure R1-B), a feature that could be used to infer a role for *AtAGO8* and *9* in TGS. This high diversification of Argonaute family members, while not exclusive to plants, is thought to be at the origin of new small RNA pathways (Chapman and Carrington, 2007).

Using sequence data from the Lotus RNA silencing machinery and results from a recently published transcriptome analysis (Hogslund et al., 2009) we observed that a set of identified RNA silencing-related genes is specifically upregulated in nodules as compared to root tissues, and reproduced in *sst1* mutant, where N_2 fixation is impaired at 90% (Krusell et al., 2005). This observation has three implications. First, the complex RNA silencing machinery is likely to be an important contributor to gene regulation events taking place during RNS. Indeed, a set of RNA silencing genes was also significantly upregulated in 8 and 16dpi *Bradyrhizobium japonicum*-infected soybean roots (Brechenmacher et al., 2008). Secondly, RNA silencing contribution to RNS does not seem to be restricted to miRNA action, as suggested by recent work (Combier et al., 2006; Boualem et al., 2009), but

might also include additional siRNA pathways in *L. japonicus* involving DCL2, AGO2, AGO4 and AGO5 orthologs through as yet uncharacterized pathways. Thirdy, nodule-specific up-regulation of the poorly characterized *DCL2* and *AGO5* genes suggests their *bona fide* contribution to gene regulation processes, beyond putative surrogate functions that these genes might exert for other well-characterized *DICER* or *ARGONAUTE-like* genes. It is noteworthy that ESTs corresponding to *LjAGO5-like* were also found upregulated at 4dpi in nodule primordia tissues in a previous transcriptome analysis in Lotus (Kouchi et al., 2004). Though the cellular functions of *AtAGO5* remain unknown (Vaucheret, 2008), biochemical assays have revealed that AtAGO5 has a preference for small RNAs with a 5'-terminal cytosine and also binds preferentially to small RNAs derived from intergenic regions (Mi et al., 2008) suggesting a possible role in RNA-dependant DNA Methylation (RdDM). Similarly, it was shown that AtAGO2 binds nearly all types of cellular small RNAs, ranging from heterochromatic siRNAs to miRNAs, provided that they start with a 5' adenosine, yet the biological relevance of this protein remains elusive (Mi et al., 2008; Vaucheret, 2008). Lastly, DCL2 is involved in endogenous natural antisense siRNAs (nat-siRNA) processing (Borsani et al., 2005), while acts as a surrogate of DCL4 in siRNA production from endogenous, transgene-derived and viral siRNAs (Deleris et al., 2006; Dunoyer et al., 2007). Assessing the role of these genes in nodule biology might thus shed light on their overall contribution to RNA silencing.

Shotgun genome sequencing has fastened the availability of the Lotus *japonicus* genome (Sato et al., 2008), though, in our case, the short sequences generated by this technique have become a limitation for the characterization of some protein family members in the gene silencing pathways. While an overview of RNA silencing effectors in Lotus is presented here, complete genome sequencing and annotation could allow the discovery of new members of these gene families in the future. Distinct and common features of Lotus-encoded and previously known genes can become a useful tool for protein functional analysis. Moreover, the availability of the spectrum of RNA silencing gene sequences in legumes now provides a basis for the generation of mutants potentially useful for the functional characterization of symbiosis-

related small RNAs. Nonetheless, our *LjDCL1* mutant study indicates that EMS mutations are probably not the most straightforward method to obtain *DICER-LIKE* knockdown mutants, due to their long and multi-domain protein nature. Similar results were obtained upon TILLING of *M. truncatula DCL* genes (M. Crespi, personal communication). Indeed, *Arabidopsis dcl1* mutants found in the literature are mostly T-DNA or X-ray deletion mutants, while EMS alleles give milder phenotypes and lower reductions in miRNA accumulation (Schauer et al., 2002; Brodersen et al., 2008; Tagami et al., 2009). In this context, the availability of a comprehensive T-DNA or transposon insertion collection would both improve and fasten genetic studies in *Lotus japonicus*.

1.2.2. *L. japonicus* P19 plants: a genetic tool for legume miRNA studies

Although we initially expressed P19 ectopically in roots through a *hairy root* transient assay, our aim was to establish stable transgenic tools for small RNA pathway dissection in legumes, and therefore we embarked on a transgenic strategy. The stable lines obtained where characterized both phenotypically and molecularly. Compound leaves from *L. japonicus* P19 plants display a normal appearance, whereas analog *A. thaliana* leaves show a characteristic serration (Dunoyer et al., 2004). Besides, *N. benthamiana* transgenic lines expressing a related tombusviral P19 protein show curled leaf margins (Silhavy et al., 2002). These dissimilar leaf phenotypes suggest a diversification of the role of miRNA-mediated regulation along the development of each leaf morphology type. Indeed, miRNAs are important leaf development regulators: miR319 is required for composite leaf development in tomato (Ori et al., 2007), while miR164 is involved in leaf serration in *Arabidopsis* (Nikovics et al., 2006). Alternatively, different levels of P19 expression in the transgenic lines from these species could entail the distinct phenotypes observed. Indeed, we noticed that P19 expression in our *L. japonicus* lines could not attain those found in *Arabidopsis* lines, maybe due to a lower tolerance to P19 expression at early developmental stages, where miRNA are likelyto play important roles. Nonetheles, Lotus P19 plants display

an important decrease in stature, especially line #13.4, as compared to the control GUS lines. This could be expalined by a lower nutrient assimilation or translocation caused by miRNA sequestration, as miRNAs are important metabolite regulators (Sunkar et al., 2007; Ding and Zhu, 2009). A slight decrease in plant size can also be found in *Arabidopsis* P19 transgenic lines (P. Dunoyer, unpublished), though separate shoot and root data are not available in this case. At the molecular level, we observed that miRNA/miRNA* sequestration takes place in various P19 tissues, and this likely leads to an interference with miRNA target regulation. Overall, these results confirm that Lotus P19 plants are functional and provide new means for in-depth miRNA study in legumes. In a near future, these lines will be used for the validation of novel miRNAs and their targets, as well as in transcriptome analyses.

CHAPTER 2. Cloning, sequencing and annotation of *L. japonicus* miRNAs

One of the main objectives of this thesis work was to uncover the small RNAome of *L. japonicus* and study its role in the determinate legume-*Rhizobia* symbiosis. We thus embarked on a small RNA cloning strategy covering three major stages in the establishment of symbiosis, including early recognition events (3hpi), bacterial infection and nodule meristem initiation (3dpi), and maintenance of mature nodules (3wpi). Therefore, several small RNA libraries were constructed and sequenced. The resulting data were pooled and analyzed bioinformatically for the identification and study of *L. japonicus* small RNA populations, with a focus on potentially novel microRNAs.

2.1. RESULTS

2.1.1. Cloning and sequencing of small RNAs from *L. japonicus* roots and nodules

3hpi and 3dpi *Rhizobia*-inoculated *L. japonicus* whole roots, together with the corresponding mock-inoculated controls (3h-mock and 3d-mock) were harvested and their 19-24nt small RNA fraction isolated and cloned. A fifth small RNA library was constructed from mature nodules (3wpi) inoculated with *Mesorhizobium loti*. A high-yield small RNA extraction protocol suitable for cloning was adapted to *L. japonicus* tissues using a hot borate buffer extraction procedure (Wan and Wilkins, 1994). The resulting five small RNA libraries were then subjected to medium-scale sequencing using the 454 technology. Following the sequence quality filtering, a total number of approx. 250.000 reads were obtained from all combined libraries. Next, we proceeded to the annotation of the population of cloned small RNAs from each library, based on sequence homology to public databases (Figure R5-A and B). Whole-root libraries showed a 60% representation of rRNA and tRNA, hereby called "miscRNA" for "miscellaneous RNA" (Figure R5-B, Table R3). Though striking at first, this high proportion of miscRNAs is in agreement with previous root small RNA studies in Arabidopsis and legumes (Hewezi et al.,

2008; Subramanian et al., 2008), indicating that this feature might be intrinsic to whole root tissues. On average, 1% of all cloned small RNAs matched plant conserved miRNAs (miRBase v13; ≤ 2 mm), in accordance with soybean whole-root studies (Subramanian et al., 2008), and 4% matched plant conserved repetitive DNA sequences (Repbase), including COPIA-like retrotransposons. The remaining reads, representing approximately 20% of the cloned sequences, were derived from the Lotus genome (≤ 1 mm). Lastly, an average of 12% reads had an unknown origin, most probably corresponding to unsequenced stretches of the Lotus genome. When compared to whole-root libraries, *Rhizobia*-containing mature nodules showed a similar pattern except that nearly 50% of the reads were annotated as deriving from the *M. loti* bacterial genome (≤ 1 mm) (Figure 5-B). Further analysis of these sequences revealed that 4,85% corresponded to *M. loti* rRNA or tRNA (*M. loti* genome, GenBank).

	3h-mock	3hpi	3d-mock	3dpi	Average roots	3wpi nodules
Total reads obtained	**11932**	**18413**	**66945**	**83054**	-	**71802**
% conserved miRNA[(1)]	0,81	0,97	1,24	1,14	1,04	0,51
% miscRNA[(2)]	55,53	56,65	64,25	66,63	60,77	20,63
% *M. loti* genome[(4)]	0,34	0,47	0,39	0,62	0,46	44,86
% DNA repeats[(3)]	4,83	4,57	3,58	3,35	4,08	1,80
% *Lotus* genome[(5)]	24,11	24,67	17,83	16,53	20,78	13,27
% Unknown[(6)]	14,37	12,68	12,72	11,72	12,87	18,92

Table R3. Composition of small RNA libraries from *L. japonicus roots* and nodules. small RNA libraries were sequenced by 454 technology and annotated using in-house small RNA bioinformatic platform (bioimage.u-strasbg.fr/bioinfo/), allowing 1 mismatch to *L. japonicus* and *M. loti* databases and 2 mismatches to the rest.
(1) miRBase v13, plant database.
(2) miscRNA includes rRNA, tRNA, snRNA, snoRNA, scRNA (GeneBank v169).
(3) Reads matching eukaryotic repeat database (Repbase v14.02).
(4) Reads matching *M. loti* genome (Rhizobase database). small RNAs matching to highly conserved sequences between prokaryote and eukaryote are annotated as *M. loti*-derived.
(5) Comprises reads not matching to the previous annotation subclasses but still present in *L. japonicus* genome (Kazusa, January 2008).
(6) Reads from unknown origin, probably unsequenced *L. japonicus* genome regions.

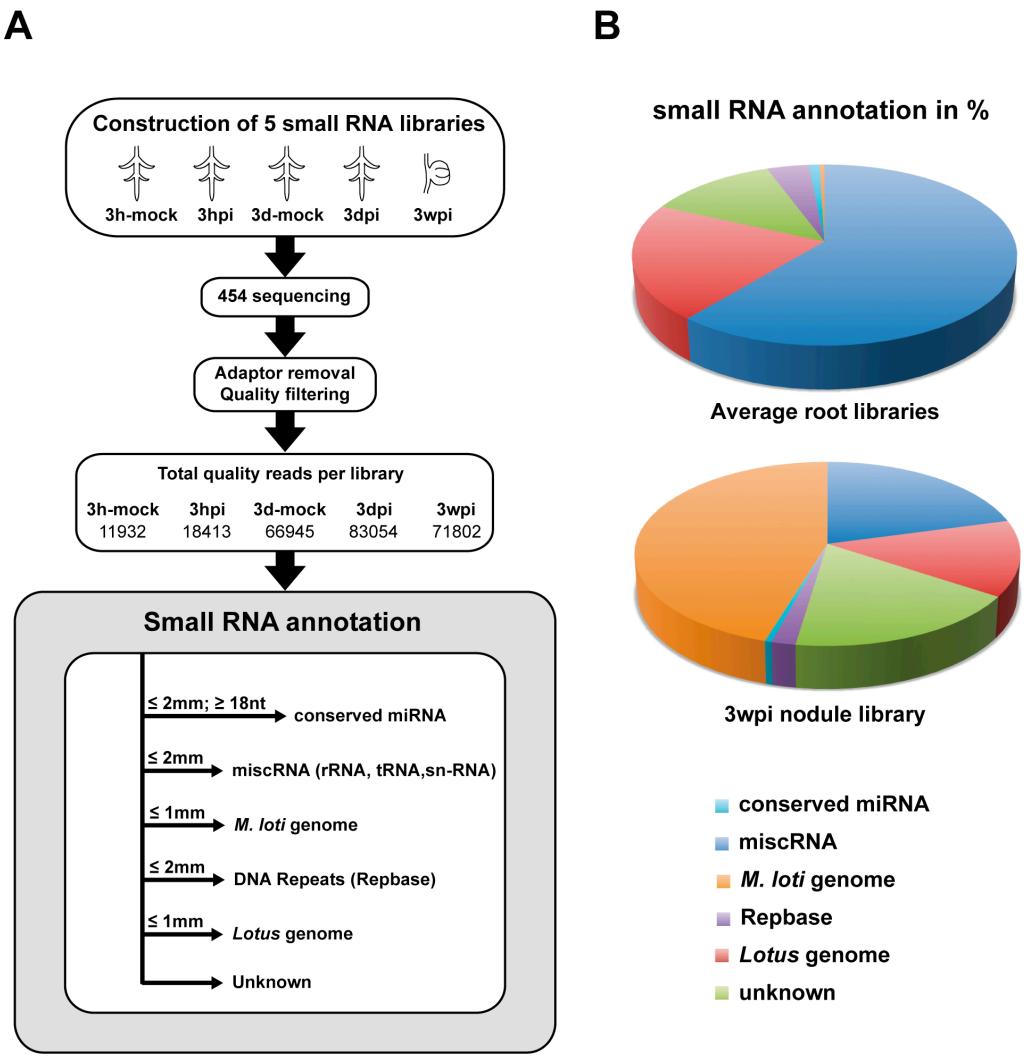


Figure R5. Composition of *L. japonicus* roots and 3wpi nodules small RNA libraries. A. Flowchart for bioinformatic small RNA library annotation. In-house sRNA annotation platform was used (bioimage.u-strasbg.fr/bioinfo). See Materials and Methods for database references. **B.** Graphic representation of small RNA library annotation. Sequences from the four root-derived libraries were averaged for simplicity (see Table R2). Non-unique (redundant) reads were analyzed. miscRNA, miscellaneous RNA (rRNA, tRNA, snoRNA, snRNA).

Next, we analyzed the size distribution of cloned small RNAs, an indicator of the degree of degradation undergone by the RNA samples. *L. japonicus* genome-matching reads, including conserved miRNAs and repeat-associated sRNAs, displayed two marked peaks at 21nt and 24nt, the latter being the most pronounced (Figure R6A-left panel). This pattern is observed in most plant species except conifers, which are depleted in 24-nt species (Dogolsheina et al., 2008). When analyzed separately, sRNAs matching to conserved miRNAs showed two main peaks at 20nt and 21nt, in agreement

with the average miRNA nucleotide length found in all plants inspected so far, including legumes (Figure R6A-right panel) (Subramanian et al., 2008). Considering that half of the small RNA populations cloned from 3wpi nodules mapped to *M. loti* genome (≤1 mm), we analyzed their nucleotide size. Surprisingly, the vast majority of these RNA reads were in a discrete 18-19nt size range (Figure R6-B). This is in contrast with r/tRNA-derived reads, which showed and heterogenous distribution in the 18-26nt range (data not shown) expected from random degradation products. Besides, sRNAs annotated as "unknown" formed a peak at 24-nt (data not shown), indicating that they are likely to be heterochromatic siRNAs whose genomic loci of origin remain unsequenced. These preliminary analyses thus indicated that our small RNA libraries presented good quality parameters and were thus suitable for further studies.

2.1.2. Identification of *L. japonicus* microRNAs.

2.1.2.1. Prediction of small RNA loci in *L. japonicus* using the sequenced sRNA data set.

Following the validation of small RNA libraries, we proceeded to the study of *L. japonicus* small RNA loci, focusing our efforts first on microRNAs and tasi-RNAs. To that aim, data from the 5 small RNA libraries were combined, generating a set of approximately 250.000 reads. Prediction of *L. japonicus* ta-siRNA loci was carried out with the ta-siRNA prediction tool (University of East Anglia, Norwic, UK) using default parameters (21nt sRNAs; ≥ 2 reads per sRNA) (Moxon et al., 2008). However, no tasi-RNA loci were identified in our data set, most probably due to technical limitations such as the fragmented genome sequences available (average contig length of 3Kb) or an insufficient sequencing depth. Indeed, *TAS3* ta-siRNAs have been recently experimentally identified in *L. japonicus* (Yan et al., 2009). Next, we focused on the analysis of microRNAs. We selected the miRNA prediction algorithm miRCat to analyze our set of 250.000 reads (Figure R7-A; Moxon et al., 2008). Through a series of highly stringent filtering steps that evaluate each sequencing read and its potential precursor hairpin, we identified 28 potential

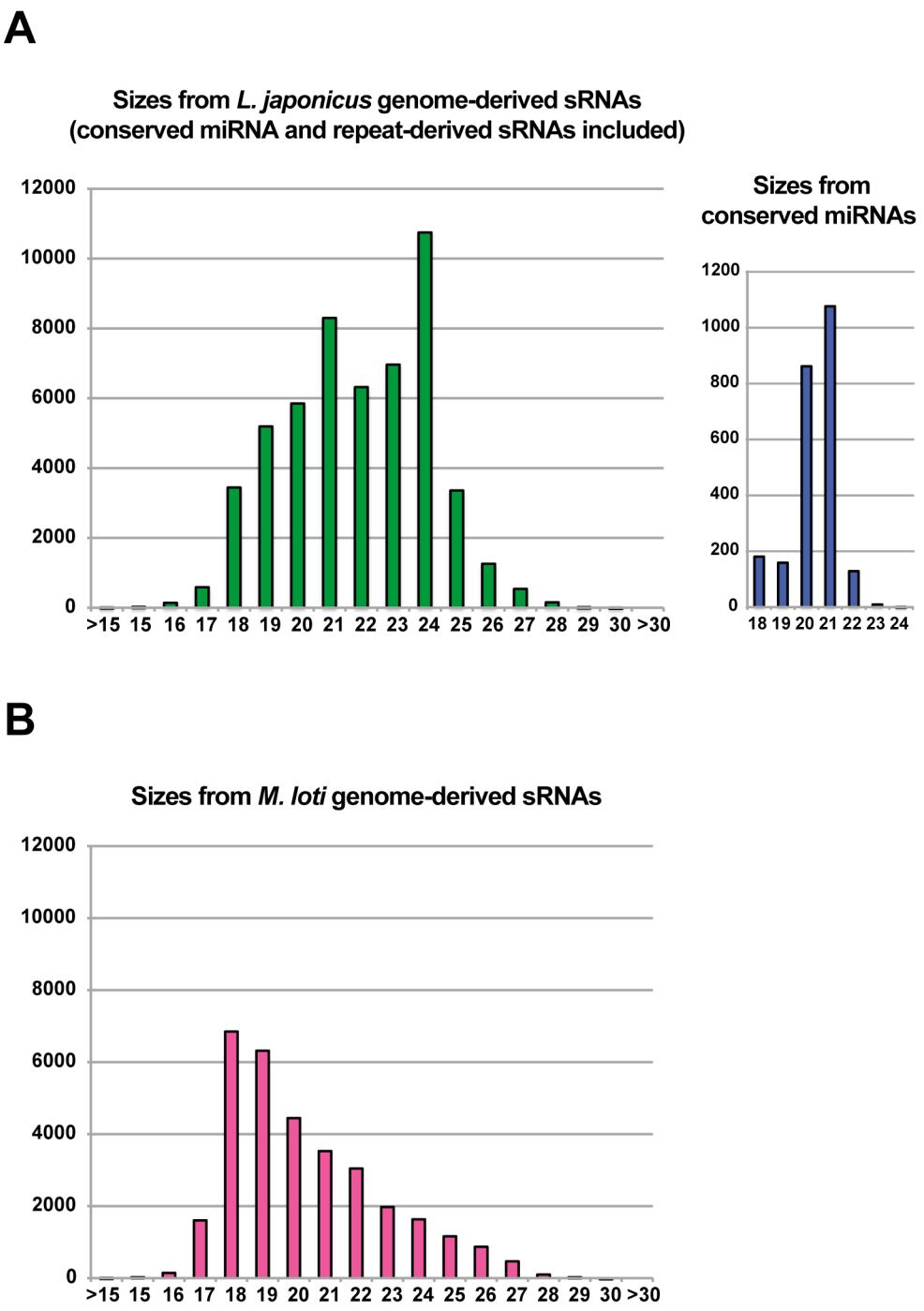


Figure R6. Sizes of cloned small RNAs. A. Sizes from all *L. japonicus* genome-matching small RNAs from the 5 libraries combined, not including small RNAs annotated as "miscRNA" (left panel) and sizes for conserved miRNA-matching sRNAs (right panel). **B.** *M. loti* genome-matching small RNAs cloned from 3wpi nodules. In all cases, non-unique (redundant) reads are represented.

novel miRNAs in *L. japonicus*, encoded by 32 *MIRNA* loci. Besides, this algorithm also recognized from our libraries members from 16 conserved miRNA families and their corresponding 44 genomic loci (Figure R7-A).

2.1.2.2. Identification of conserved miRNA families in *L. japonicus.*

Analysis of sequence homology between raw sequencing reads and available plant miRNA databases (miRBase v13; ≤ 2 mm) revealed that our *L. japonicus* libraries, all combined, contained miRNAs corresponding to 24 conserved miRNA families, plus 3 recently identified soybean novel miRNA families: gma-1509, gma 1510 and gma-1511 (Table R4). We use the term 'conserved' for miRNAs present throughout one major plant clade, as angiosperms (Table R4; Axtell and Bowman, 2008). The 27 known miRNA families that we cloned represent a 2-fold increase when compared with those that successfully passed the miRCat prediction pipeline, which evaluates both the *MIRNA* and precursor genes, reflecting the high stringency of this tool. The members of 32 conserved miRNA families actually present in our libraries are listed in Table R5. In addition, out of the 32 families, corresponding miRNA* species were cloned at low frequencies for miR156, miR166, miR168, miR390, miR393, miR396 and miR408. These data complement the 19 conserved families recently identified *in silico* from *L. japonicus* genome (Sato et al., 2008) and indicate that cloning-based identification of miRNA loci is an important complement to bioinformatic analysis from sequenced genomes. Further attention to *L. japonicus* conserved miRNAs is given in Chapter II.

Cloned miRNA family[1]	Cloned miRNA*	Passed miRCat[2]	Plant species where it is conserved[3]
miR156[4]	yes	yes	ath,osa,zma,sbi,mtr, sof,gma,ptc,ppt,ghr, bna,pta,smo,wi,sly,bdi,aqc
miR157[4]	-	-	ath,gra,bra,bol,sly
miR159[4]	-	yes	ath,osa,sbi,sof,gma, zma,ptc,pta,smo,tae, bna,wi,sly,bra,mtr,pvu,aqc
miR160	-	-	ath,osa,zma,sbi,mtr, gma,ptc,ppt,smo,tae, wi,sly,bra,bdi,aqc
miR162	-	-	ath,osa,cpa,zma,mtr, ptc,ghr,wi,gma,sly,pta
miR164	-	yes	ath,osa,zma,sbi,ptc, mtr,tae,bna,wi,gma, bra
miR165[4]	-	-	ath
miR166[4]	yes	yes	ath,osa,zma,sbi,mtr, gma,ptc,ghr,ppt,pta, smo,bna,wi,sly,pvu, bdi,aqc
miR167	-	yes	ath,osa,zma,sbi,sof, gma,ptc,mtr,tae,bna, wi,sly,lja,bra,bdi,aqc
miR168	yes	yes	ath,osa,sbi,sof,gma, zma,ptc,mtr,bna,wi, aqc
miR169	-	yes	ath,osa,zma,sbi,mtr, gma,ptc,ghb,bna,wi, sly,bdi,aqc
miR171	-	yes	ath,osa,zma,sbi,mtr, ptc,ppt,bna,pta,smo, tae,wi,gma,sly,bra,bol,bdi,aqc
miR172	-	yes	ath,osa,sly,zma,sbi, gma,ptc,mtr,wi,sly,bra,bol,bdi,aqc
miR319[4]	-	yes	ath,pta,osa,sbi,mtr, gma,zma,ptc,ppt,mtr, smo,wi,sly,pvu,aqc
miR390	yes	yes	ath,osa,ptc,ppt,mtr, ghr,pta,bna,wi,gma, sbi
miR393	yes	yes	ath,osa,sbi,mtr,zma, ptc,bna,wi,gma
miR394	-	-	ath,osa,sbi,zma,ptc,wi
miR396	yes	yes	ath,osa,sbi,sof,gma, zma,ptc,mtr,ghr,bna, pta,smo,wi,lja,aqc
miR397	-	yes	ath,osa,ptc,bna,wi,sly,sbi,bdi
miR398	-	-	ath,osa,gma,ptc,mtr, pta,wi,bol,aqc
miR408	yes	yes	ath,osa,sof,zma,ptc, ppt,pta,smo,tae,wi,sbi,bdi,aqc
miR414	-	-	ath,osa,ppt
miR419	-	-	ath,osa,ppt
miR482	-	-	ptc, pta, gma, vvi, mdm, pvu, aqc
miR1310	-	-	pta
miR1426	-	-	osa
miR1509	-	-	gma,mtr[5]
miR1510	-	-	gma,mtr[5]
miR1511	-	yes	gma[5]
miR1875	-	-	osa

Table R4. Known miRNA families cloned from our libraries.

2.1.2.3. Characterization of potential *L. japonicus* novel miRNAs.

Using the miRCat prediction algorithm we identified 28 potential novel miRNAs, encoded by 32 miRNA loci (Table R5). While their precursor hairpins had structural and free energy found in most plant *MIRNA* genes (Figure R7-A; Moxon et al., 2008), additional data were needed for a reliable annotation of this set of sequences as *bona fide* miRNAs. First, we studied the cloning rates of these potential novel miRNAs. Young miRNAs are generally less expressed than conserved miRNAs, presumably to reduce off-targeting effects (Voinnet, 2009). This has also been observed in *M. truncatula* (Szittya et al., 2008). Accordingly, we observed that the most highly expressed, newly-identified miRNA, lja-10, was sequenced 10 times less than the most highly expressed conserved miRNA, miR159, with a total of 732 reads (Table R5). Secondly, *Arabidopsis* miRNAs start preferentially with a 5' U nucleotide, though this feature is not essential for miRNA function, as 15% diverge from this characteristic (Mi et al., 2008). Following this criterion, *L. japonicus* sequences from conserved and novel miRNAs obtained with miRCat analysis were scanned using WebLogo (weblogo.berkeley.edu/logo.cgi) (Figure R7-B and C). Sequence alignment revelead a 5'U bias in both groups, though it was moderate in the novel miRNA dataset. Thus, this set might include *bona fide* miRNAs but also some false positives.

Table R4. Known miRNA families cloned from our libraries.
(1) miRNA families retrieved through sequence homology (≤ 2mm) to miRBase v13. All libraries combined.
(2) Both miRNA sequence and miRNA precursor passed miRCat stringent criteria.
(3) Information retrieved from miRBase v14
(4) miR156/miR157, miR159/miR319 and miR165/miR166 form 3 independent miRNA families containing miRNA genes annotated with more than one number (Jones-Rhoades et al., 2006). However, annotation per organism in miRBase is listed separately for each of the numbers, and so was represented here.
(5) Legume-specific microRNAs
Aqc, Aquilegia coerulea; Ath, Arabidopsis thaliana; Bdi, Brachypodium distachyon; Bna, Brassica napus; Bol, Brassica oleracea; Bra, Brassica rapa; Cpa, Carica papaya; Ghb, Gossypium herbecium; Ghr, Gossypium hirsutum; Gma, Glycine max; Gra, Gossypium rammindii; Lja, Lotus japonicus; Mdm, Malus domestica; Mtr, Medicago truncatula; Osa, Oryza sativa; Ppt, Physcomitrella patens; Pta, Pinus taeda; Ptc, Populus trichocarpa; Pvu, Phaseolus vulgaris; Sbi, Sorghum bicolor; SlySolanum lycopersicum; Sof, Saccharum officinarum; Tae, Triticum aestivum; Vvi, Vitis vinifera; Zma, Zea mays.

The main criterion for the definition of a true plant miRNA is the precise excision from the stem-loop precursor in form of a discrete miRNA/miRNA* duplex; by contrast, siRNAs occur as populations along their dsRNA precursors (Meyers et al., 2008). Thus, we examined the abundance of miRNA* species in our libraries and found those corresponding to 5 of the predicted novel miRNAs (Table R5), including lja-10 and lja-12 (Figure R8-A). We then carried out Northern analysis to assess the accumulation of potential miRNAs. Although all of the highly abundant putative young miRNAs (lja-4 to lja-13) were systematically analyzed in different blotting experiments, we could only detect lja-4, lja-10 and lja-12 (Figure R8-B). These small RNAs accumulate as discrete species and are absent in RNA samples from *A. thaliana*. The use of highly sensitive detection methods such as LNA probes might be required for the study of the other potential miRNAs, as well as for miRNA* species, which could neither be detected by conventional Northern approach. Following these analyses, we studied the degree of conservation of lja-4, lja-10 and lja-12 in other plant species. We could not identify any plant genomic or EST sequences that matched these potential novel miRNAs (ncbi; ≤ 3 mm). Combined with their lack of detection in Arabidopsis RNA samples, this result suggest that these miRNAs are indeed *L. japonicus*-specific. Lastly, we carried out a prediction of miRNA targets using the Target Finder algorithm (Allen et al., 2005; Fahlgren et al., 2007) and the set of predicted *L. japonicus* CDS from Kazusa Institute (www.kazusa.org.jp/lotus) (Table R5). While annotation of predicted targets was rather diverse, there was a clear overrepresentation of disease resistance genes (lja-4, lja-10, lja-22) and pentatricopeptide repeat (PPR)-containing proteins (lja-12, lja-30), as previously observed for young miRNAs from legumes and other species (He et al., 2008; Szittya et al., 2008; Lelandais-Briere et al., 2009).

Taken together, these data give positive indications for lja-10 and lja-12 being genuine *L. japonicus* young miRNAs. In the case of lja-4, were the miRNA* was absent from our libraries, the accumulation of species from regions surrounding the miRNA itself in the hairpin should be ruled out before conclusive annotation as miRNA (Meyers et al., 2008). A further issue regarding the precursor from which these potential novel miRNAs originate is

that the *L. japonicus* genome was sequenced via a double strategy, i.e. clone-by-clone and shotgun sequencing, that in some cases yield redundant sequence information (Sato et al., 2008). Thus, it is possible that some of the newly identified *MIRNA* loci are redundant (Table R5).

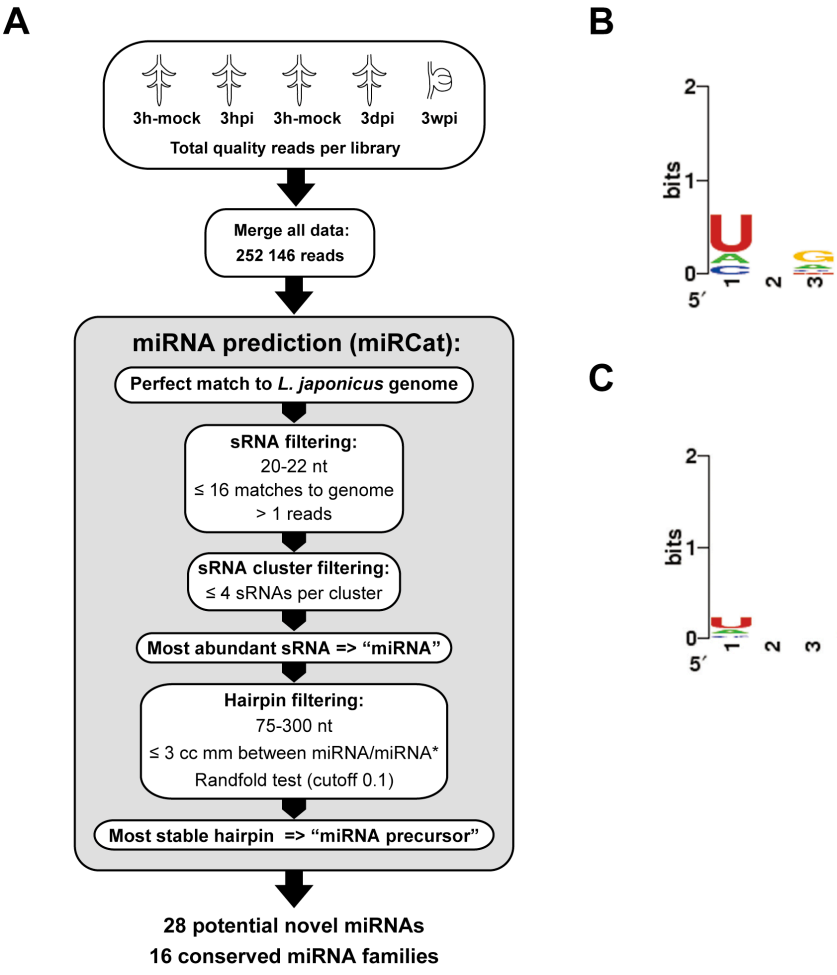


Figure R7. Prediction of novel miRNAs in *L. japonicus*. A. Flowchart of miRCat miRNA prediction tool (Moxon et al., 2008). nt, nucleotide; cc, consecutive; mm, mismatch. **B, C.** Nucleotide pattern analysis of miRCat-predicted *Lotus* conserved (B) and potential novel miRNAs (C) using WebLogo (weblogo.berkeley.edu/logo.cgi). Non-redundant 20-22 nt sequences were used. The first three nucleotides are shown. mm, mismatch; cc, consecutive; nt, nucleotide.

ID	Sequence	nt	miRNA loci[(1)]	Reads[(2)]	miRNA*[(3)]	Detect[(3)]	Conserved[(4)]	Predicted target[(5)]	Annotation [(6)]
lja-4	UCUUCCAUCCAUACAUCAUCU	21	CM0037 (2), LjSGA_016001	53	no	N	no	chr3.CM0106.560.nd (2,5)	NBS-LRR drp
lja-8	UUACCGUUACGCCGAGGCUCG	21	LjSGA_011321	13	yes (6)	-	no	LjSGA_051161.1 (2,5)	-
lja-9	CAUUUUGUCAUCUUGAUUUGG	21	LjSGA_120332	25	no	-	no	LjT01K14.30.nd (3,5)	-
lja-10	UAGCGGGUGUCUUCGCCUCUGA	22	LjSGA_009943, LjT10I14	55	yes (5)	N	no	chr3.CM0241.210.nd (3,5)	TIR-NBS-LRR drp
lja-11	UUGCGCACUGAGCAAGGACAGG	22	LjSGA_063310	16	no	-	no	-	-
lja-12	UCAUCAUGAAAUUGCAGUGCU	21	LjSGA_107863	7	yes (7)	N	no	LjT03E06.70.nd (0)	PPR- containing
lja-13	AGAGGCUUAUGGAGUGAGAGA	21	LjSGA_029065, LjSGA_126006	6	yes (1)	-	no	chr3.CM0406.320.nd (3)	AFH1
lja-14	ACUUGAGAGGAUUUCACUGGA	21	CM0155	3	no	-	no	chr1.CM0113.240.nc (4)	cytochrome b
lja-15	GAGGGGGAAGGUGAUGACAUC	21	LjSGA_014283	3	no	-	no	LjSGA_052661.1 (3,5)	presenilin family
lja-16	ACUCUCCCUCAAGGGCUUCUA	21	CM0060	2	no	-	acq(1mm)	LjSGA_013995.1 (4)	RCC
lja-17	UUAGAAAGAAAAUGUUGUUAGC	22	CM0124	2	no	-	-	LjSGA_055678.1 (3)	-
lja-18	UCGAGAGAGAGAGCGACGAGG	21	CM0127	2	no	-	-	LjT34K24.150.nd (3)	-
lja-19	UAGGAAAUACGCCUGCGGUUCC	22	CM0147	2	no	-	-	chr4.CM0170.10.nd (4)	ATP-ase
lja-20	UCUGUCGCGGGAGAGAUGACAC	22	CM0225	2	no	-	-	LjSGA_008791.0.1 (3)	-
lja-21	AUCAAGGUAGCUGUAACUUCC	21	CM0361	2	no	-	-	LjT43M12.180.nd (1)	-
lja-22	UGUUGUUUUACCUAUUCCUCC	21	CM0437	2	yes (2)	-	mtr (1mm); gma (3mm)	LjSGA_026865.1 (1,5)	TIR-NBS-LRR drp
lja-23	ACCACCGGGCUCGAGGAUCAGC	22	CM0442	2	no	-	-	-	-
lja-24	UAUAAUAGUGAUCGGAGGGU	20	LjSGA_009094	2	no	-	mtr (0 mm)	LjSGA_030850.0.1 (3)	-
lja-25	CGGUUUCGGACGGUUUCAGGC	21	LjSGA_018941	2	no	-	-	-	-
lja-26	UGUUAGAAGGUAUGAAAACGG	21	LjSGA_027672	2	no	-	-	LjSGA_030850.0.1 (3,5)	-
lja-27	UGGGGGCUUCUAAUGGCUAGC	22	LjSGA_033384	2	no	-	-	-	-
lja-28	UUAAAUUUCCAAAUCAAUGGCU	22	LjSGA_033997	2	no	-	-	-	-
lja-29	GCGGUGAAUAGUAGAGUAGCGG	22	LjSGA_038711	2	no	-	-	chr6.CM0037.570.nd (4)	endomembrane
lja-30	CACGUUCAAGCAUUUCGAGC	20	LjSGA_118482	2	no	-	-	chr1.CM1255.10.nc (3,5)	PPR-containing
lja-31	ACCACCGGGCUCGAGGAUCAGC	22	LjT01F24	2	no	-	-	-	-
lja-32	CCGUAGCAUCAAUUUAUCCGA	21	LjT15D05	2	no	-	-	chr3.CM0282.170.nd (4)	Sugar transporter
lja-33	CAUGGCGUGCAAAACCCCACGC	22	LjT16L01	2	no	-	-	-	-
lja-34	UAUAAUAGUGAUCGGAGGGU	20	LjT46I09	2	no	-	-	LjSGA_014949.2 (4)	damaged DNA bind

Table R5. Potential novel miRNAs predicted through miRCat (Moxon et al., 2008

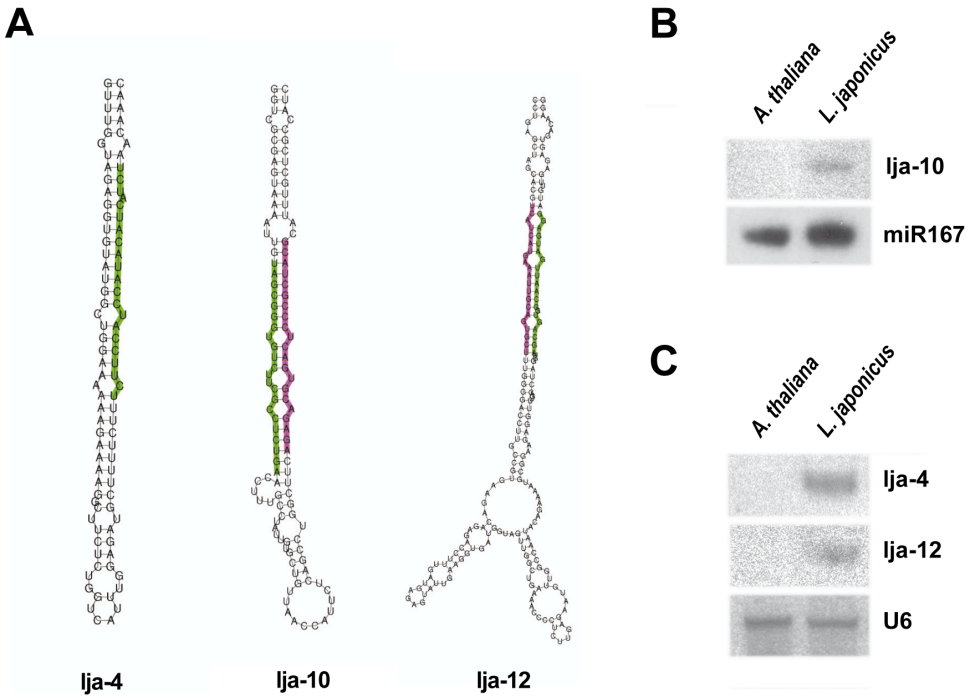


Figure R8. *L. japonicus* potential novel miRNAs. A. miRNA precursors as folded with RNAfold. Green and pink depict mature miRNA and miRNA* sequences, respectively. **B, C.** Detection of several potential novel miRNAs by northern blot. Leaf tissue from *A.thaliana* and *L. japonicus* was used. B and C correspond to independent assays.

Table R5. Potential novel miRNAs predicted through miRCat (Moxon et al., 2008).

(1) Some of the loci for each potential novel miRNA could be redundant. For lja-4, 2 precursors were detected in the genomic clone CM0037.

(2) Total read number from all libraries combined.

(3) The number in parentheses shows the total reads obtained from miRNA* species.

(3) Detection by northern blot.

(4) Conservation in other plant species was assessed by sequence homology to NCBI genomic and EST databases (≤ 3 mm) and subsequent folding of genomic sequences; In some cases, homolog genomic stretches were found but the stem-loop folding of miRNA precursor was negative (RNAfold; Hofacker et al., 2003). acq, *Aquilegia coerulea*; gma; *G. max*; mtr, *M. truncatula*; (-), not studied.

(5) Highest scoring predicted target through TargetFinder (Allen et al., 2005). Score is indicated between parentheses.

(6) Annotation is shown as indicated at Kazusa Institute (Blast2). Key: drp, disease resistance protein; PPR, pentatricopeptide repeat-containing protein; RCC, regulator of chromosome condensation; AFH1, Formin homology 1; PLP, Poly-L-proline-containing protein involved in actin cytoskeleton rearrangement.

2.2. DISCUSSION

2.2.1. Deep sequencing of *L. japonicus* nodule and root libraries.

Using the 454 sequencing technology, we obtained 250.000 total reads from inoculated/mock roots and mature nodules from *L. japonicus*, the first medium-scale small RNA analysis undertaken in this model legume. Bioinformatic analysis revealed that a high percentage of rRNA/tRNA was present in our whole root small RNA libraries, which resulted in a lower rate of sequences available for silencing-related small RNA analysis. Considering that this has also been observed in other studies (Hewezi et al., 2008; Subramanian et al., 2008), whole root studies might require higher sequencing depths, such as those provided by the Illumina technology, for example. Besides, we were also surprised by the high content of *M. loti*-derived sequences found in mature nodules and by the 18-19nt peaks that they formed. Indeed, this size range could reflect *in planta* Dicer-mediated processing of longer bacterial dsRNA possibly translocated into the host. Several of the most abundant of these reads were probed by Northern blot and discrete bands corresponding to approx. 100nt were observed, suggesting that they actually derive from degradation of longer bacterial RNAs, despite the fact that only 4,85% correspond to annotated *M. loti* r/tRNAs. It has been speculated that *Rhizobial* noncoding RNAs could regulate plant gene expression (Simon et al., 2009), but the present analysis does not provide experimental support to this hypothesis. We note, however, that several *M. loti*-derived sequences found in our libraries match *L. japonicus* CDS (≤ 1 mm), and this might deserve further attention.

2.2.2. Novel miRNAs

Due to the agronomical importance of legume-*Rhizobia* symbiosis, several small RNA sequencing analyses on symbiotic interactions have been initiated over the recent years. Together, a considerable set of miRNA genes from legumes have been identified, including legume-specific miRNAs. In this study, we have predicted 28 novel miRNA families and so far confirmed 2 *Lotus*-specific novel

miRNAs, lja-10 and lja-12. Additional analysis will most probably increase the set of confirmed Lotus young miRNAs. This rate of new miRNA discovery might appear very modest as compared to the total amount of reads. However, precise rules for plant miRNA annotation have recently been agreed, giving special importance to the identification of miRNA* species (Meyers et al., 2008). In fact, from *M. truncatula* studies were the same miRNA criteria and prediction tools were used, 26 novel miRNAs were predicted and only 8 validated from a total of nearly 4.000.000 sequences (Szittya et al., 2008). Though it is necessary to avoid mis-annotations, cloning of miRNA* is a rather stringent criteria that implies the use of the deepest sequencing technologies, as reflected by the fact that we cloned miRNA* from only 1/4 of all conserved miRNAs. Furthermore, independent studies in Medicago have found very little overlap in the novel miRNA set identified, even when the Illumina technology was used (Szittya et al., 2008; Jagadeeswaran et al., 2009). This supports the idea that identification of novel miRNA in legumes is far from saturation, as is also the case in *A. thaliana*, notably because most libraries generated so far involved plants grown in ideal or highly controlled laboratory conditions (Voinnet et al., 2009).

Following miRNA prediction and validation, we carried out an *in silico* target analysis. Target prediction should be considered carefully in two main aspects, first, because "canonical" rules for miRNA targeting are not always followed *in vivo* (Brodersen and Voinnet, 2009) and, secondly, because the annotation of those *Lotus* predicted targets has been inferred bioinformatically and lack a functional validation. However, it is striking that distinct studies of young miRNA in different plant species have all predicted or even confirmed the targeting of NBS-LRR disease resistance genes by these miRNAs (He et al., 2008; Szittya et al., 2008; Jagadeeswaran et al., 2009; Lelandais-Briere et al., 2009). In our case, we have found that pentatricopeptide repeat (PPR)-containing genes could also be targeted by several of the predicted novel miRNAs, as previously found for *A. thaliana* young miRNA miR161 (Allen et al., 2004). This is not surprising, though, as both *NBS-LRR* and *PPR* represent extremely large gene families that are rapidly evolving through duplications, sometimes under selective

pressures (Geddy et al., 2007). This makes them prone to miRNA-mediated regulation, as one of the means for *de novo* generation of *MIRNA* genes is the initial inverted duplication of the corresponding target gene (Allen et al., 2004; Voinnet et al., 2009). Accordingly, we found extended sequence homology between lja-12 precursor and the PPR-containing predicted target gene.

CHAPTER 3. Investigation of symbiosis-related microRNAs

3.1. RESULTS

Following the analysis of the microRNAome in *L. japonicus*, we focused on its potential contribution to the legume-*Rhizobia* interactions. As previously mentioned, small RNA cloning/sequencing encompassed several key stages of the establishment of symbiosis, i.e., initial recognition events (3hpi roots), meristem growth and bacterial infection (3dpi roots) and, finally, maintenance of 3wpi colonized, mature nodules. In addition, two mock-inoculated control root libraries for each inoculation timepoint were also constructed, named respectively 3h-mock (3hm) and 3d-mock (3dm). Comparison of miRNA cloning frequency rates between different libraries was the first strategy chosen for the detection of candidate symbiosis-related miRNAs. Quantitative miRNA expression profiles were obtained by analyzing small RNA cloning frequencies with the algorithm miRProf (Moxon et al., 2008), where reads matching to a fixed conserved miRNA family (miRBase database v13) are summed up and an unambiguous match signature assigned (Figure R9-A). Data were normalized to total reads excluding those matching ribosomal or transfer RNA (r/tRNA), so that a miRNA expression profile was obtained for each small RNA library (Table R6). Nonetheless, the small RNA cloning data from each sample was also manually studied. This included the search for interesting candidates whose low cloning frequencies would have hampered quantitative comparison, the analysis of potentially relevant miRNA isoforms, and the assessment of novel miRNA expression patterns not included in the miRProf pipeline.

miRProf signature	3h-mock	3hpi	3d-mock	3dpi	3wpi nod
miR159	208,9	304,9	748,7	492,4	158,2
miR172	0,0	0,0	0,0	1,9	111,1
miR166	331,8	457,3	555,7	726,3	99,3
miR397	12,3	32,1	9,3	7,5	48,8
miR167	12,3	8,0	7,0	5,7	35,4
miR390	0,0	0,0	0,0	1,9	26,9
miR156; miR157	16,0	0,0	9,3	11,3	21,9
miR396	98,3	96,3	123,2	103,8	13,5
miR156	36,9	24,1	23,3	24,5	11,8
miR393	73,7	96,3	39,5	64,1	10,1
miR171[(1)]	0,0	24,1	18,6	7,5	10,1
miR319	49,2	80,2	102,3	81,1	5,1
miR408	0,0	8,0	2,3	1,9	5,1
miR1511	49,2	32,1	37,2	58,5	3,4
miR159; miR319	0,0	0,0	18,6	1,9	3,4
miR162	24,6	0,0	9,3	9,4	3,4
miR164	12,3	16,0	25,6	15,1	3,4
miR482	0,0	0,0	9,3	13,2	1,7
miR1426	0,0	0,0	0,0	0,0	1,7
miR1509	0,0	0,0	0,0	0,0	1,7
miR1875	0,0	0,0	0,0	0,0	1,7
miR419	0,0	0,0	0,0	0,0	1,7
miR168	12,3	40,1	55,8	28,3	0,0
miR1510	0,0	0,0	4,7	1,9	0,0
miR398	0,0	0,0	4,7	5,7	0,0
miR1310	0,0	0,0	2,3	0,0	0,0
miR160	12,3	8,0	2,3	0,0	0,0
miR165	0,0	8,0	2,3	0,0	0,0
miR169	0,0	0,0	2,3	1,9	0,0
miR414	12,3	0,0	0,0	0,0	0,0
miR165; miR166	0,0	8,0	0,0	0,0	0,0
miR394	0,0	8,0	0,0	1,9	0,0
Total non-t/rRNA reads	8138	12464	43007	53006	59406[(2)]

Table R6. Conserved miRNA profiles: quantitative comparison of cloning data. Normalized counts per 100.000 non-t/rRNA reads are shown, as obtained by miRProf (Moxon et al., 2008); *nod*, nodules.
(1) Fairly distinct miR171 isoforms were observed and the miRProf grouping data from this miRNA family should not be considered further.
(2) *Mesorhizobium loti*-derived small RNAs are included in this set.

3.1.1. miRNA expression analysis at early *Rhizobia-Lotus* interactions

Comparison of miRNA profiles from each of the four root libraries (Figure R9-B) showed that 4 miRNAs represent 3/4 of the conserved miRNA population in *Lotus japonicus* roots: miR159, miR166, miR396, and miR319. Pairwise comparison of miRNA profiles from 3hpi and 3dpi libraries with their respective mock-inoculated controls revealed some differences in the accumulation of several miRNAs. Likewise, differential cloning frequency rates were observed when libraries from the two different root inoculation timepoints (3hpi, 3dpi) were contrasted, though the 4-fold difference in global sequence throughput can introduce biases in this comparison (Table R6; Figure R9-B). We then proceeded to the validation of the differences in revealed by cloning/sequencing using Northern analysis (Figure R10-A and B). Contrary to our expectations, miR166, miR159 and miR393, which showed markedly contrasted cloning rates, followed the same constant expression pattern as miR396 and miR167, which were cloned uniformly among the different root libraries. Conserved miRNAs with low cloning frequencies (Table R6; miR156 to miR1427) were not analyzed further as their comparative analysis could not be considered significant. Besides, expression patterns of the potential novel miRNAs lja-4 and lja-10 were assessed by Northern analysis and, in accordance to cloning data, no differences were found between inoculated and mock-inoculated tissues (Table R6; Figure R10-A). Thus, we could not identify *Rhizobia*-induced miRNAs at the onset of symbiosis using these approaches. Yet, our molecular analysis was not exhaustive and potentially interesting miRNAs such as the Lotus ortholog of soybean gma-miR1511, which shows a two-fold increase at 3dpi (Table R6), might deserve a closer study.

Potential novel miRNAs	3h-mock	3hpi	3d-mock	3dpi	3wpi nod	sRNA total reads[(1)]
lja-4	36,9	8,0	51,2	28,3	13,5	49
lja-8	12,3	0,0	4,7	13,2	1,7	11
lja-9	12,3	16,0	16,3	13,2	5,0	20
lja-10	49,2	24,1	25,6	30,2	3,4	36
lja-11	12,3	16,0	4,7	5,7	0,0	8
lja-12	0,0	0,0	7,0	0,0	0,0	3
lja-13	36,9	0,0	4,7	0,0	0,0	5
lja-15	0,0	0,0	2,3	0,0	3,4	3
Total non-t/rRNA reads	8138	12464	43007	53006	59406[(2)]	-

Table R7. Potential novel miRNA profiles: quantitative comparison of cloning data. Normalized counts per 100.000 non-t/rRNA reads are shown.
(1) Only frequencies for novel miRNAs cloned more than twice in all libraries combined are shown; *nod*, nodules.
(2) *Mesorhizobium loti*-derived small RNAs are included in this set.

3.1.2. miRNA expression profiles in mature *L. japonicus* nodules

Determinate nodules are newly formed organs that derive from *Rhizobia*-induced root cortex meristems, which cease growth once the complete structure is formed. *Lotus japonicus* nodules undergo a series of developmental steps leading to a mature stage at 2-3wpi, when they host the differentiated *Mesorhizobium loti* N_2-fixing bacteroids. After 6wpi, a senescence process is started, together with the parallel formation of new nodules along the root (Hirsch, 1992; Puppo et al., 2005). Unlike in indeterminate nodules, *Lotus* nodule developmental stages are thus spatially and temporally separated, enabling a detailed and straightforward analysis of each of the steps. In the present study, 3wpi nodules were selected for small RNA cloning a stage at which symbiosome differentiation has been completed and fully N_2-fixing ability attained (Hogslund et al., 2009).

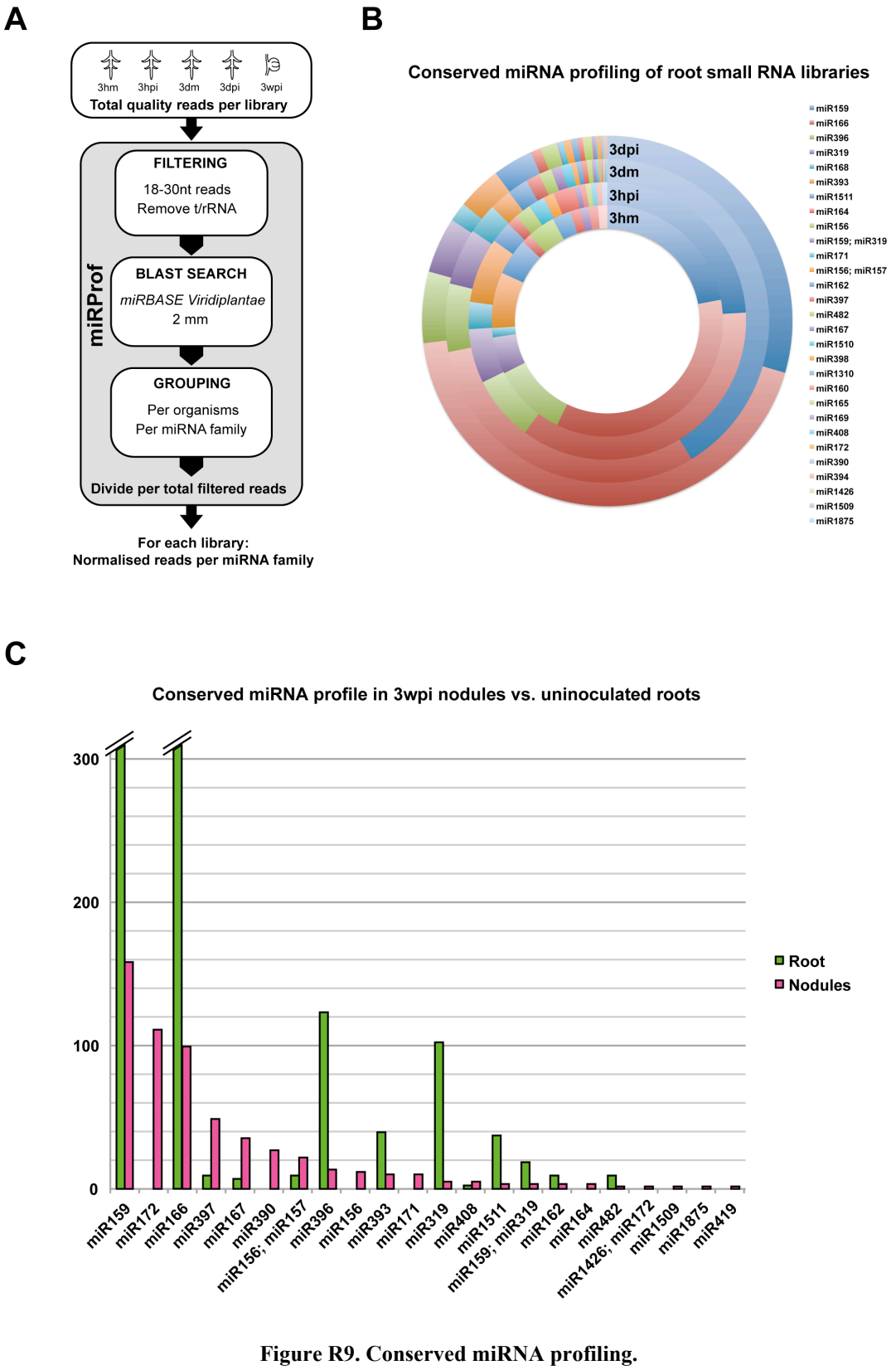


Figure R9. Conserved miRNA profiling.

3.1.2.1. Quantitative comparison and validation of nodule and root miRNA cloning profiles

Comparative expression analysis from root and nodule samples has successfully been used in the past for the identification of symbiosis-related genes (Colebatch et al., 2004; El Yahyaoui et al., 2004; Lelandais-Briere et al., 2009). Following this line, conserved miRNA expression profiles from the 3wpi nodule sRNA libraries were obtained and compared to cloning rate data from the 3dm (3-day mock-inoculated) library, corresponding to 3 week-old, developed root systems grown in the absence of *M. loti*. Several of the miRNAs that are highly expressed in roots were also cloned at high frequency from nodule tissue, such as miR159, miR166 and miR396). However, a distinct set of miRNAs was clearly more abundant in nodules, including miR167, miR172, miR390 and miR397 (Figure R9-C). As mentioned above (Table R2), nodule small RNA libraries contained almost 50% of *Mesorhizobium loti*-derived small RNAs. With the bioinformatic tools available at the time of our analysis, these sequences could not be removed for miRNA expression normalization. Hence, the actual nodule/root comparative expression levels of this set of conserved miRNAs might be greater than those depicted in Figure R9-C. Besides, analysis of the potential novel miRNAs showed that several of them are expressed in nodules, yet at considerably lower levels that in root tissue (Table R7).

Figure R9. Conserved miRNA profiling. A. miRProf pipeline for miRNA grouping used for profiling. Chosen parameters are indicated. **B.** Comparison of conserved miRNA profiles from Rhizobia-inoculated (3hpi; 3dpi) and mock-inoculated root samples (3hm; 3dm). Each cloned small RNA is unambiguously assigned to one match signature, listed on the right panel. **C.** Chart showing cloning frequencies of reads matching known miRNAs from 3wpi nodules compared to 3-day mock-inoculated roots. Normalised counts per 100.000 non r/tRNA reads are shown. Only the set of miRNAs sequenced from nodule libraries are shown. nt; nucleotide; mm, mismatches; hpi, hours post-inoculation; hm, hours after mock-inoculation; dpi, days post-inoculation; dm, days after mock-inoculation.

We then proceeded to the validation of the mature nodule-specific miRNA profiles obtained by 454 sequencing. The expression in nodules of the abundantly cloned miR166, miR167, miR393, miR159 and miR396 was confirmed by hybridization assays, as well as the accumulation of lja-4 and lja-10 (Figure R10-A and B). We then focused on the miRNAs uncovered by the comparative analysis between 3wpi nodules and uninoculated roots, i.e. miR167, miR172, miR390 and miR397 (Figure R9-C). Beyond the molecular validation of their possibly higher abundance in nodules, we first ask whether expression of these miRNAs would be specific to the 21dpi nodule stage chosen for library construction. Thus, nodules were harvested and RNA extracted at different times post-inoculation, ranging from laeghemoglobin-free white 10dpi nodules to mature N_2-fixing 14, 21 and 28dpi nodules (Figure R10-C). RNA from the root tissue remaining after nodule excision was analyzed separately on the same membrane. The high accumulation in nodules of miR167, miR172, miR390 and miR397 was confirmed by Northern analysis (Figure R10-D), though different patterns of expression were observed. Both miR167 and miR172 were easily detectable at all nodule stages assayed, indicating their independence from the establishment of full N_2-fixation ability. By contrast, miR390 and miR397 were fairly low expressed at earlier developmental phases. In addition, a slight tendency toward increased accumulation at later nodule stages was observed for most miRNAs, but this is yet to be confirmed in further experiments. In addition, weaker, albeit more variable expression in the remaining root tissues (as compared to nodules) was also observed for all miRNAs except miR390, which showed an unexpected high signal from 14dpi roots in two indepednet hybridization assays (Figure R10-D). Thus, comparative analysis of small RNA sequences from 3wpi nodules and roots as well as subsequent Northern validation show that miR167, miR172, miR390 and miR397 are specifically highly expressed in determinate *L.japonicus* nodules.

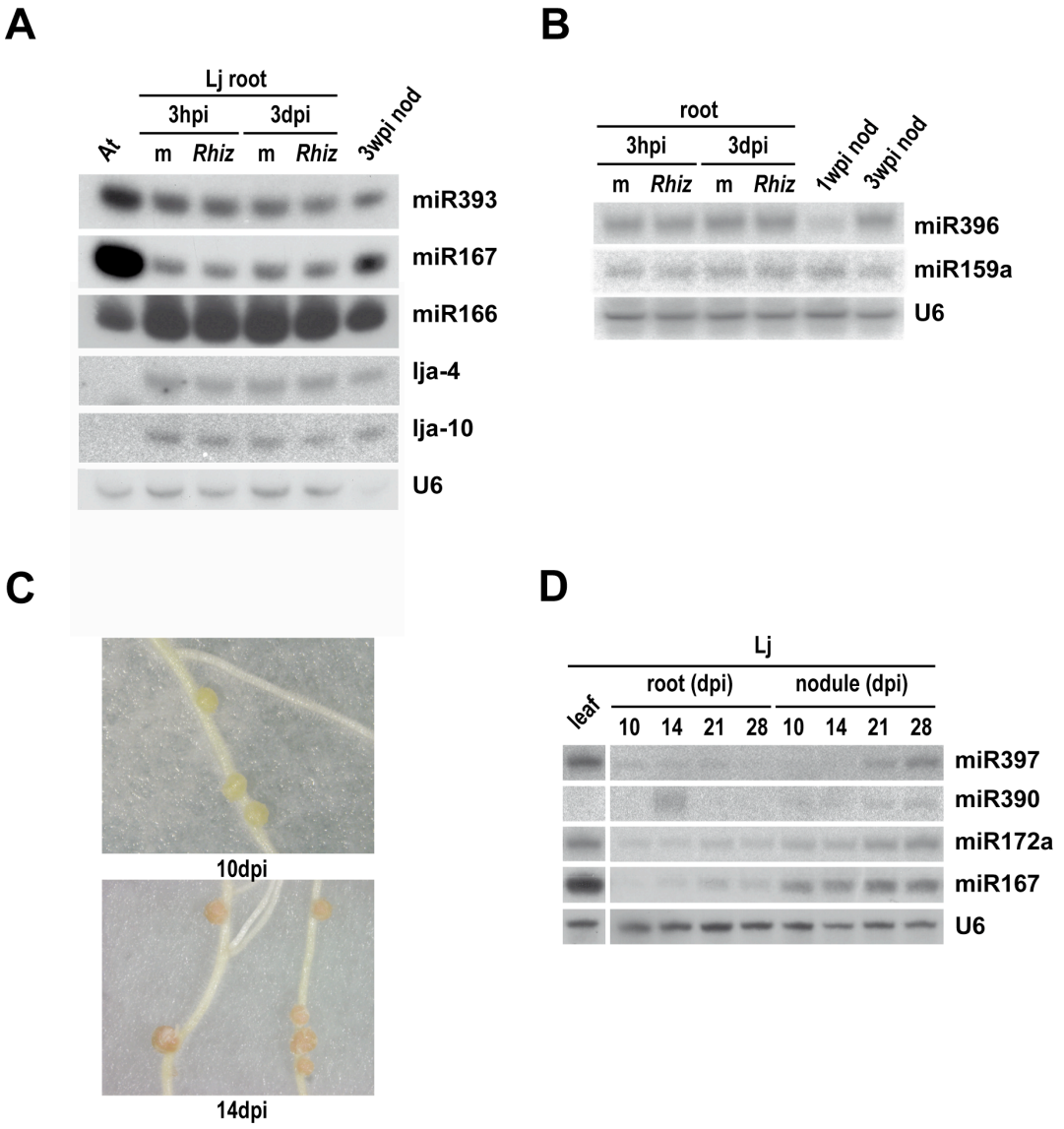


Figure R10. Analysis of miRNA expression in roots and nodules. A. miRNA accumulation at different times post-inoculation. RNA samples correspond to those used for small RNA library construction. **B.** miRNA accumulation in roots at different times post-inoculation. Two different nodule stages were included in the assay. **C.** Young 10 dpi and mature 14 dpi nodules. **D.** miRNA accumulation at different nodule developmental stages. Nodule tissue illustrated in (B) was used. Root samples comprise remaining tissue after nodule excision. Leaf samples from an independent experiment were used in the same RNA membrane as a hybridization control. *At, A. thaliana; Lj, L. japonicus*; hpi, hours post-inoculation; dpi, days post-inoculation; m, mock; Rhiz, *Rhizobia*-inoculated; nod; nodules.

3.1.2.2. Insight into miRNA isoforms cloned from nodules

The sequencing depth of our libraries restricted the quantitative expression analysis to the most abundant miRNAs, making it easy to overlook small RNAs with low cloning frequencies. Moreover, the miRProf comparative pipeline (Moxon et al., 2008) fuses different miRNA isoforms and often matches them to the same *MIRNA* locus. A recent work in *M. truncatula*, however, has revealed that distinct miRNA isoforms are differentially expressed in root and nodule tissue (Lelandais-Briere et al., 2009). We thus carried out a manual sequence analysis of root and nodule-derived sequenced reads. Several sequence variants harboring 5' or 3'-end 1-2 nucleotide mismatches or overhangs were found for most miRNAs cloned (data not shown) and were not considered further. However, and strikingly, different sequence isoforms were found in the miR171 family. In particular, a sequence variant harboring a 4-nucleotide 3'-overhang and a core sequence diverging from the canonical, conserved miR171a, was cloned from nodule tissues (3 reads/library) and several of the root libraries (1-2 reads/library). A single precursor gene was found for this miRNA variant and, following the identification of 4 other *MIR171* genes in *Lotus,* it was named lja-miR171c (Figure R11-A). Comparative analysis of the reads obtained for the miR171 family indicates that members of this specific miRNA family are low abundant, both in nodule and root tissues (Figure R11-A).

In order to study the *in vivo* expression pattern of the nodule-specific miR171c, Northern analysis from root and nodule tissues was performed (Figure R11-B). Surprisingly, 3wpi nodules showed a strong accumulation of lja-miR171c, whereas its expression in roots remained below detection levels. It is noteworthy that the low cloning frequency of miR171c correlated with a general difficulty in detecting this miRNA by Northern analysis. Nonetheless, sequence alignment of miR171 family members (miRBase v13), revealed that a near-identical mature miR171 species was identified in large-scale small RNA sequence analyses carried out both in soybean (Subramanian et al., 2008) and *M. truncatula* (Lelandais-Briere et al., 2009) (Figure R11-C). Thus, miR171c is conserved in legume species,

though its nodule-specific expression was not reported. Moreover, lja-miR171c is absent in *Arabidopsis*, but close sequences have been cloned in rice and stressed poplar (Lu et al., 2008). Taken together, these results show that the low abundant lja-miR171c is specifically expressed in nodules, and its conservation across legume species argues for a relevant biological role in RNS.

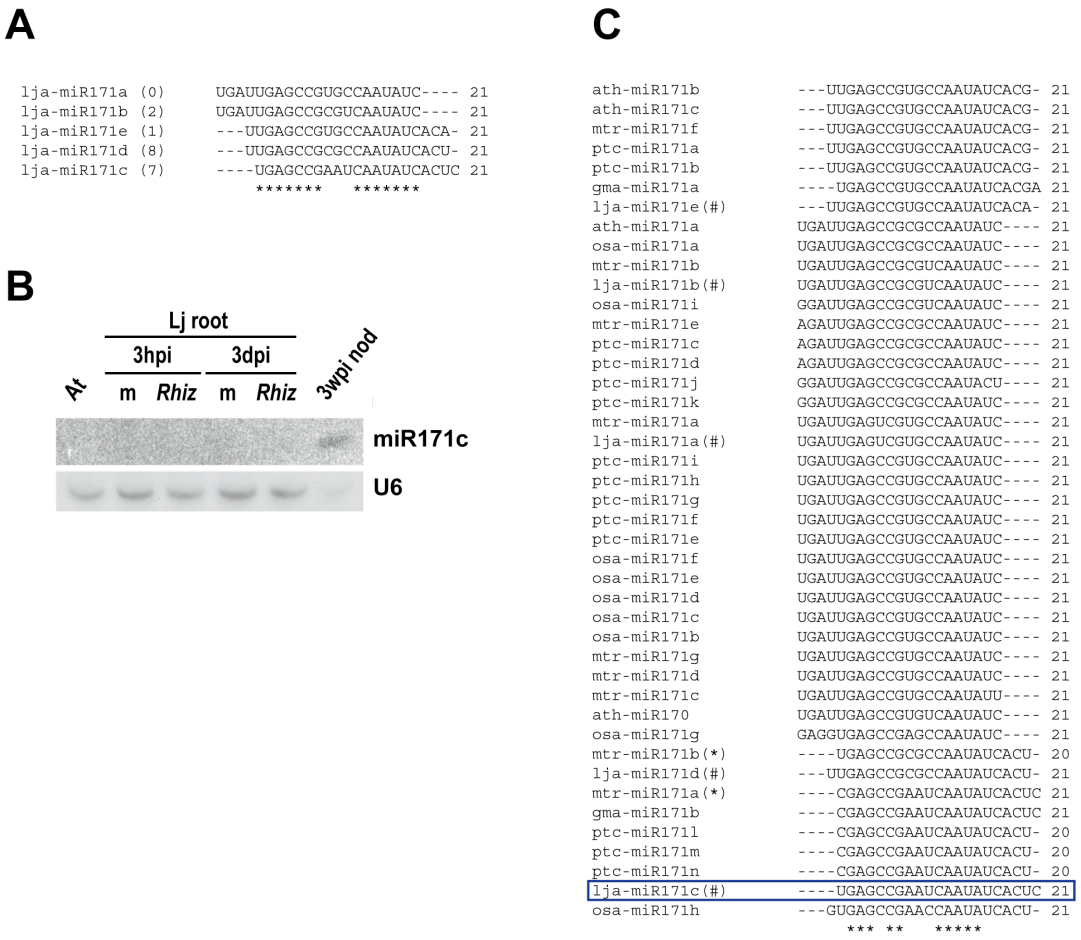


Figure R11. Analysis of a miR171 isoform cloned from nodule small RNA libraries. A. Sequence alignment of the *L. japonicus* miR171 family members identified in this study. Cloning frequencies from all libraries combined are indicated in parenthesis. **B.** Northern blot detection of lja-miR171c in roots and nodules. **C.** Sequence alignment of mature miR171 from different plant species. miRNA sequences identified in this study (#) and Lelandais-Briere et al., (2009) (*), were added to miRBase data from rice, poplar, *Arabidopsis* and legume species (miRBase v14 Sept 2009) and subsequently aligned using Clustal 2.0.11. osa, rice; ppt, poplar; ath, *Arabidopsis*; mtr, *M. truncatula*; gma, soybean; lja, *L. japonicus*.

3.1.2.3. Validation of nodule miRNA profiles using the *Lotus Affymetrix GeneChip®*

Through our small RNA sequencing strategy, we have defined the set of the most abundant miRNAs in mature, determinate *L. japonicus* nodules. MiRNAs are processed from stem-loop structures (pre-miRNAs) that in turn derive from poly-A containing precursor transcripts (pri-miRNAs) (Voinnet, 2009). Pri-miRNA expression has been effectively assessed by conventional microarray analysis strategies (Lin et al, 2008). Though pri-miRNA expression and miRNA activity in a specific tissue or cell lineage are not necessarily coincident owing to reported systemic miRNA mobility (Lin et al., 2008; Pant et al, 2008), this approach can indeed provide useful information. Taking advantage of the presence of *MIRNA* precursor genes in the *Lotus* Affymetrix GeneChip® (Hogslund et al., 2009), we performed a comparative expression analysis of 3wpi nodules versus non-inoculated roots. Among all *MIRNA* genes spotted on the array, a group was found upregulated in nodules at a statistically significant level (three biological replicates; False Discovery Rate p-value≤0,05) (Figure R12). Strikingly, this set of pri-miRNAs comprised the nodule miRNAs previously discovered by quantitative comparison of miRNA profiles, i.e, miR167, miR172, miR390, miR397, but also included the low abundant miR171c and miR408. The most significant change in nodules/root expression signal was observed for the pri-miR172, which showed a near 16-fold increase in nodules, in accordance with our small RNA cloning approach. A remarkable nodule-specific induction of the lja-MIR171c was also observed, confirming our Northern analysis data. It is noteworthy that the levels from the other 4 *Lotus* pri-miR171 spotted on the array remained statistically unchanged, including those of pri-miR171b and pri-miR171d, which were also cloned from nodule tissues. This argues for a specific role for miR171c in nodules, despite the low cloning frequency that we observed. Besides, *MIR167*, *MIR390*, *MIR397* and *MIR408* presented a fairly modest increased expression signals in 3wpi nodules as compared to uninoculated roots. Lastly, *MIR164* was the single gene whose primary transcript was upregulated in roots, whereas expression ratios from the rest of the pri-miRNAs spotted on the chip did not show statistically

significant changes between nodules and roots (Figure R12). Hence, Affymetrix analyses validated our previous observations made from small RNA cloning. They also reinforced the idea that miR171c and miR408 are potentially interesting nodule-specific miRNAs, despite their low abundance in nodules.

Taken together, these results show that mature determinate *Lotus japonicus* nodules, as compared to roots, accumulate a specific set of miRNAs, including miR167, miR171c, miR172, miR390, miR397 and miR408, which are either absent or considerably less abundant in root tissue. These miRNAs thus constitute possible candidates as riboregulators of various steps of symbiosis.

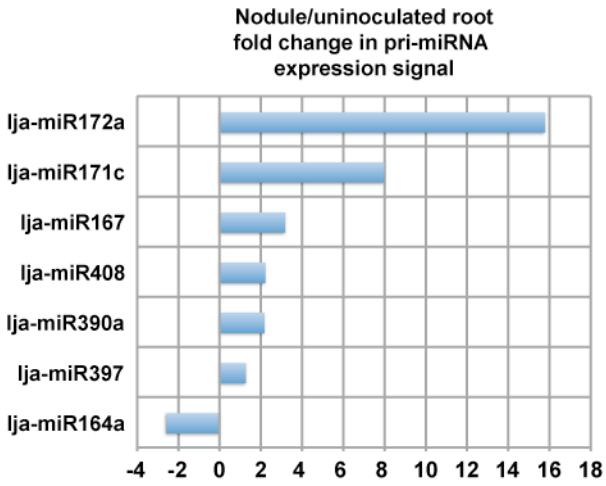


Figure R12. Comparative expression analysis of the pri-miRNAs spotted on the *Lotus Affymetrix GeneChip*®. Fold change was calculated as the ratio of normalised average expression values from wild-type 3wpi nodules vs. uninoculated roots (three independent biological replicates). From all pri-miRNA spotted on the chip, only those with a statistically significant differential expression are shown (False Discovery Rate (FDR) corrected p-value ≤ 0,05) (Hogslund et al., 2009).

miRNA	sequence	*L. japonicus* target ID [1]	score	*Arabidopsis* annotation	*M. truncatula* annotation	(2)
miR156	TTGACAGAAGATAGAGAGCA	LjB09A03.80.nc	3,5	ATFH8 (FORMIN 8); actin binding	Actin-binding	
		LjT10E17.60.nd	3,5	OPR3 (OPDA-REDUCTASE 3)	NADH:flavin oxidoreductase	
		LjSGA_040667.2	3,5	Protein phosphatase 2C	Protein phosphatase 2C	
		LjSGA_056737.2	3,5	Pentatricopeptide (PPR) repeat-contain	Tetratricopeptide-like helical	
miR157	CTGACAGAAGATAGAGAGCA	LjSGA_056737.2	3	Pentatricopeptide (PPR) repeat-contain	Tetratricopeptide-like helical	
		chr1.CM0591.130.nd	3	OPR1 (12-oxophytodienoate reductas	NADH:flavin oxidoreductase	
		LjB06N21.140.nd	3,5	rcd1-like cell differentiation protein	Cell differentiation protein	
		LjB06N21.160.nd	3,5	rcd1-like cell differentiation protein	Cell differentiation protein	
miR159	ATTGGAGTGAAGGGAGCTCC	LjSGA_035803.1	0	AtMYB81 (myb-domain protein 81)	Homeodomain-related	
		chr3.CM0005.370.nc	0	AtMYB65	Homeodomain-related	
		LjSGA_028341.1	1	Proton-dependent oligopeptide transporter	TGF-beta receptor	
		chr5.CM0200.1410.nd	3	-	-	
miR166	TCGGACCAGGCTTCATTCCCC	LjSGA_035832.1	2,5	ATHB-8 (HOMEOBOX GENE 8)	Homeodomain- elated	*
		LjSGA_101583.1	2,5	ATHB15	Homeodomain-related	
		LjSGA_014333.2	3	REV (REVOLUTA)	Homeodomain-related	
		LjSGA_029320.1	3	ATHB9 | PHV (PHAVOLUTA)	Homeodomain-related	
miR167	TGAAGCTGCCAGCATGATCT(GA)[4]	chr1.CM0295.550.nd	3,5	-	-	
		LjSGA_039126.1	4	Auxin Response Factor (ARF8)	Aux/IAA_ARF_dimerisation	*
		chr2.LjT43I20.180.nd	4	Auxin Response Factor (ARF6)	Aux/IAA_ARF_dimerisation	*
		chr4.LjT16L01.60.nd	4	Auxin Response Factor (ARF6)	Aux/IAA_ARF_dimerisation	*
miR171c	TGAGCCGAATCAATATCACTC	LjT26J05.10.nc (LjNSP2)	1	Scarecrow-like GRAS-domain (AtSCL26)	GRAS-domain transcription factor	
		LjT15C06.70.nc (LjNSP3)	1	Scarecrow-like GRAS-domain (AtSCL26)	GRAS-domain transcription factor	
		chr3.LjT02C24.90.nd	3	EMBRYO DEFECTIVE 2745	Tetratricopeptide-like helical protein	
		LjSGA_016315.1	4	RING-H2 finger C2A	Zinc finger, RING-type	
miR172	AGAATCTTGATGATGCTGCAG	LjT39B13.80.nc	1	RAP2.7/TOE1 (AP2-likeTF)	PR and ERF (AP2-like)	*
		LjSGA_005424.1	2	Apetala 2 (AP2) transcription factor	PR and ERF (AP2-like)	*
		chr1.CM0033.300.nc	2	Apetala 2-like transcription factor	PR and ERF (AP2-like)	*
		LjSGA_142418.1	2	-	-	

Table R8. Target prediction of miRNAs cloned from *L. japonicus* 3wpi nodules (cont.)

miRNA	sequence	*L. japonicus* target ID[(1)]	score	*Arabidopsis* annotation	*M. truncatula* annotation	(2)
miR319	TTGGAGTGAAGGGAGCTCC	LjSGA_035803.1	0	AtMYB81 (myb-domain protein 81)	Homeodomain-related	
		chr3.CM0005.370.nc	0	ATMYB65 (myb-domain protein 65)	Homeodomain-related	
		LjSGA_028341.1	1	proton-dependent oligopeptide transporter	TGF-beta receptor	
		chr5.CM0200.1410.nd	2	-	-	
miR390	AAGCTCAGGAGGGATAGCGCC	BP047016	- [(3)]	TAS3 precursor	TAS3 precursor	*
		LjSGA_051075.1	4	TMKL1 (transmembrane kinase-like)	Protein kinase	
miR393	TCCAAAGGGATCGCATTGATCC	LjSGA_042978.1	2	AtAFB3 (Auxin signalling F-box 3)	F-box protein	
		LjSGA_023031.1	2,5	AtTIR1 (Transport Inhibitor Response 1)	F-box protein	*
		LjSGA_087408.1	2,5	AtTIR1 (Transport Inhibitor Response 1)	F-box protein	*
miR396	TTCCACAGCTTTCTTGAACTG	LjSGA_054039.1	3	LjSGA_054039.1	Growth-regulating factor	
		LjT10E17.120.nd	3	AtGRF5 (Growth-regulating factor)	Growth-regulating factor	
		LjSGA_004481.1	3	AtGRF5 (Growth-regulating factor)	Growth-regulating factor	
		LjSGA_066892.1	3	AtGRF5 (Growth-regulating factor)	Growth-regulating factor	
miR397	TATTGAGTGCAGCGTTGATGA	LjSGA_075846.2	1	AtLAC10 (Laccase; copper ion binding/oxidoreductase)	Copper-resistance protein CopA	*
		chr1.CM0147.300.nc	2	AtLAC17 (Laccase; copper ion binding/oxidoreductase)	Copper-resistance protein CopA	*
		chr2.CM0249.1200.nc	2,5	AtLAC11 (Laccase; copper ion binding/oxidoreductase)	Copper-resistance protein CopA	*
		chr2.CM0249.1210.nc	3	AtLAC11 (Laccase; copper ion binding/oxidoreductase)	Copper-resistance protein CopA	*
miR408	ATGCACTGCCTCTTCCCTGGC	LjSGA_034638.1	2,5	AtARPN (PLANTACYANIN)	Blue copper domain protein	*
		LjSGA_142734.0.1	3	-	-	
		LjSGA_111866.1	4	AtARPN (PLANTACYANIN)	Blue copper domain protein	*
		chr1.CM0133.680.nc	4	AtHMA8 (Cu transmebrane transporter)	Cu-binding; Cu translocator	

Table R8. Target prediction of miRNAs cloned from *L. japonicus* 3wpi nodules (cont.)

Table R8. Target prediction of miRNAs cloned from *L. japonicus* 3wpi nodules. miRNAs cloned more than twice are shown, while the most frequent miRNA sequence was chosen for target search.. Predictions were carried out with Target Finder algorithm (Allen et al., 2005; Fahlgren et al., 2007). A maximum of 4 predicted targets is shown per miRNA. Annotation is shown as in Kazusa Institute *L. japonicus* CDS annotation, carried out through sequence similarity to *Arabidopsis* and *M. truncatula* genes (Blast2).
(1) Some of these CDS sequences can be redundant to an actual unique gene, due to redundancy in *L. japonicus* sequencing.
(2) Similar genes have been confirmed as targeted by corresponding miRNA in legume species (see Table I-2).
(3) *Lotus* TAS3 non-coding RNA precursor was recently identified and validated (Allen et al., 2005; Yan et al., 2009).
(4) The most frequent miRNA sequence in nodules was the 22-nt, though no predicted targets were obtained.
Key: PR, Pathogenesis-related; ERF, Ethylene-Response Factor.

3.1.3. Target prediction for miRNAs cloned from mature nodules

The sequences from nodule-induced miRNAs were used to perform a bioinformatic miRNA:target analysis using the algorithm Target Finder (Allen et al., 2005; Fahlgren et al., 2007) (Table R8). Genes identified as predicted targets were manually annotated by sequence homology to *Arabidopsis* and *M. truncatula* annotated genes. Interestingly, predicted targets for miR167, miR172, miR390, miR397 and miR408 were highly similar to those confirmed by 5'RACE in *Arabidopsis* (Jones-Rhoades et al., 2006) but also in recent studies from the legumes *M. truncatula* (Jagadeeswaran et al., 2009) and *P. vulgaris* (Arenas-Huertero et al., 2009). In both *Arabidopsis* and legumes, miR167 targets Auxin Response Factors similar to ARF6 and ARF8, while miR172 targets APETALA-2-like transcription factors. Similarly, miR390 targets in *M. truncatula* and *L. japonicus* the ta-siRNA precursor 3 (TAS3), which in turn generates ta-siRNAs targeting several Auxin Response Factor genes (Allen et al., 2005; Jagadeeswaran et al., 2009; Yan et al., 2009). Besides, a protein kinase was predicted as an additional miR390 target in *Lotus* (Table R8), in accordance to what has been found in rice (Sunkar et al., 2005b). Laccase family genes and plantacyanin are the respective targets for miR397 and miR408 in *Arabidopsis* and *Medicago* and are also predicted targets in *L. japonicus*. Perhaps most interestingly, our analysis revealed that lja-miR171c, but not other members of the lj-miR171 family, is likely potential regulator of the symbiosis-related GRAS-domain LjNSP2 transcription factor (heckmann et al., 2006), as well as the recently identified, closely related gene LjNSP3 (Hogslund et al., 2009). Sequence alignments for several of these miRNA:target pairs and their corresponding scores are depicted in figure R13. Although target confirmation in *Lotus* is ongoing, this information can provide hints for the regulatory role of these conserved miRNAs.

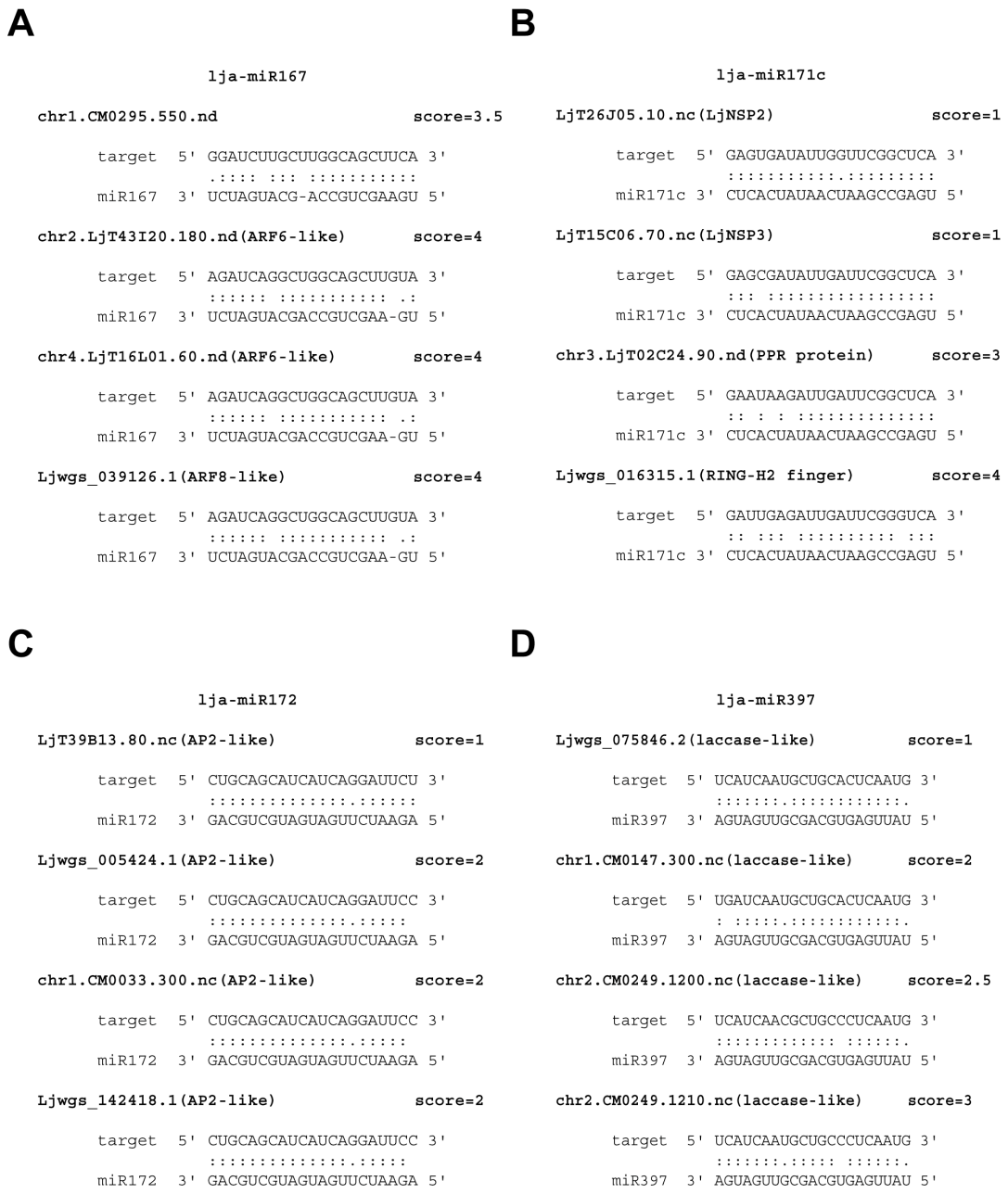


Figure R13. miRNA:predicted target alignments for several nodule miRNAs. A. miR167. **B.** miR171c. **C.** miR172. **D.** miR397. Target prediction was carried out using Target Finder (Allen et al., 2005; Fahlgren et al., 2007). The 4 predicted targets with best scores are shown for each miRNA. Key: (:) paired bases; (.) G:U pairs.

3.1.4. Analysis of *L. japonicus* spontaneous nodulation mutants

The observation that mature nodules express a specific set of miRNAs led us to ask whether they would play a role in the bacterial infection and N_2 fixation processes or, alternatively, in differentiation or maintenance of the nodule as an independent organ. To address this question, we exploited the availability of *Lotus japonicus* spontaneous nodulation mutants (*snf*), which form in the absence of bacteria (Tirichine et al., 2006). Spontaneous nodules show various phenotypical features that indicate their genuine nodule nature: they harbor peripheral vascular bundles, accumulate intracellular starch granules, similar to Fix- mutants, and present nearly identical ultrastructural features to those found in wild-type nodules (Tirichine et al., 2006). Three independent genes responsible for spontaneous nodulation have been isolated, and two of them, *snf1-1* and *snf2*, have recently been characterized (Tirichine et al., 2006; Tirichine et al., 2007). We used the *snf1-1* mutant line for our analysis and simultaneously grew mock-inoculated mutant and *Rhizobia*-inoculated wild-type *Gifu* plants to perform a side-by-side analysis of both nodule types at two times post-inoculation (Figure R14-A). The aim was to assess the expression of the previously identified nodule-specific miRNA, though the low detection levels of several of these miRNAs restricted our analysis to the abundant miR167, miR172 and miR397. Surprisingly, Northern analysis showed that miR397 was drastically depleted in *snf1-1* bacteria-free nodules as compared to *Gifu*-infected nodules, while miR167 and miR172 were equally detectable in both nodule types. No differences were found between the two post-inoculation timepoints analyzed, i.e. 3wpi and 5wpi (Figure R14-A).

Besides the ability to form nodules in the absence of *Rhizobia*, bacterial inoculated *snf* mutants are indeed able to establish successful symbiotic interactions, so that infected and spontaneous nodules can occasionally coexist in a given root system (Tirichine et al., 2006). Thus, specific miRNA expression patterns in spontaneous nodules could be due either to the absence of interaction with the bacterial symbiont, or to the *snf1-1* mutation itself, which results in a dominant CaCMK allele (Tirichine et al, 2006b). To investigate this question

further, *snf1-1* mutant plants were divided into two distinct sets, *Mesorhizobium loti* or mock-inoculated, and were compared to inoculated wild-type *Gifu* plants. Infected nodules from both *snf1-1* and wild-type *Gifu* are indistinguishable, as compared to laeghemoglobin-free, white spontaneous nodules (Figure R14-B). We thus performed small RNA analysis from each plant set using excised nodules as well as the root tissue remaining upon nodule excision (Figure R14-C). the analysis also included systemic leaf tissues (Figure R14-D). Confirming our previous analyses, miR167 and miR172 were detected in both spontaneous and infected nodules derived from the *snf1-1* mutant, though a slight decrease as compared to *Gifu* wild-type nodules was observed (Figure R14-C). While miR167 levels remained similar, inoculated roots displayed a remarkable upregulation of miR172 (Figure R14-C). When analyzing aerial tissue, both miR167 and miR172 levels remained unchanged regardless of the development of spontaneous or infected nodules (Figure R14-C). These results indicate that the expression of both miR167 and miR172 in nodules is independent of the establishment of symbiosis.

Considerably distinct patterns of expression were observed for miR397. Indeed, in the *snf1-1* background, miR397 was upregulated in infected nodules as compared to spontaneous nodules, but its accumulation was lower than in *Gifu* wild-type nodules (Figure R14-C). At the systemic level, miR397 was largely upregulated in wild-type *Gifu* nodulated plants as compared to mock-inoculated plants. Likewise, miR397 levels were increased in *snf1-1* inoculated plants, as compared to plants that had developed spontaneous nodules. Nonetheless, systemic miR397 levels in infected *snf1-1* mutant lines were lower than in infected wild-type lines, as previously seen for nodule tissue (Figure R14 -D). This pattern of modest but clear induction of miR397 in *snf1-1* inoculated plants was also observed in preliminary analysis of infected nodules from the *snf2* mutant line (data not shown), whose spontaneous nodulation phenotype is accounted for by a dominant mutation in the cytokinin receptor LHK1 (Tirichine et al., 2007). Collectively, these data show that miR397 is specifically up-regulated in nodules and leaves in an infection-specific manner. This result is significant, since it

identifies the first miRNA to be linked to processes related to the presence of the bacterial symbiont, rather than nodule development *per se*.

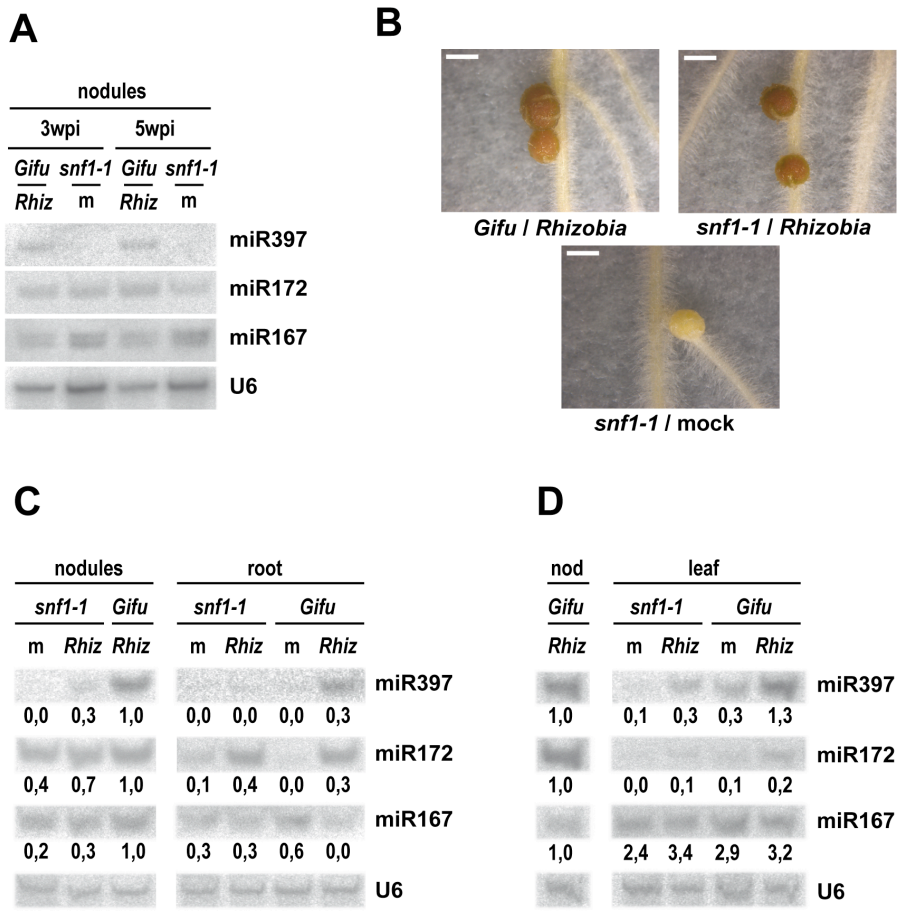


Figure R14. miRNA expression *Rhizobia*-free spontaneous nodules. A. Accumulation of nodule miRNAs in spontaneous nodules from *snf1-1* mutant, as compared to *Gifu* wild-type infected nodules. Tissue harvesting of spontaneous and infected nodules was carried out simultaneously at the two indicated times post-inoculation. **B.** Phenotypes of *Rhizobia*-inoculated nodules from wild-type *Gifu* and *snf1-1* mutant lines, as compared to mock-inoculated spontaneous *snf1-1* nodules. Scale bar =1mm. **C.** Comparison of miRNA levels in nodules and excised roots from mock-inoculated *snfl-1* mutant plants and *Rhizobia*-inoculated wild-type *Gifu* nodules. Tissue harvesting was carried out simultaneously. Left (nodules) and right (roots) panels correspond to a same hybridization experiment. All miRNA levels were quantified with U6 expression and normalized to Gifu wild-type nodules. **D.** Analysis of miRNA accumulation in systemic leaf tissues corresponding to (C). For comparison, miRNA levels were quantified and normalized to *Gifu* wild-type nodules, blotted on the same hybridization experiment. Quantification and normalization of northern blot signals was carried out with ImageGauge v4.1. Scale bar =1mm. wpi, weeks post-inoculation; m, mock-inoculated; Rhiz, *Rhizobia*-inoculated.

3.1.5 Analysis of miR397 expression in N_2 fixation-deficient symbiotic interactions

The specific induction of miR397 expression in *Rhizobia*-infected nodules led us to ask whether this miRNA is implicated in maintenance of the infection or, alternatively, in N_2 fixation-related processes. In order to address this question, we decided to analyze distinct Fix- symbiotic interactions that lead to N_2 fixation-defective endosymbiosis. First, we exploited the *Bradyrhizobium* sp. (*Lotus*) strain NZP2309, which establishes only partial symbiosis with wild-type *L. japonicus*. These bacteriae are effectively recognized by *L. japonicus* roots and nodule infection is allowed. However, nodule development is delayed, early senescence is observed and N_2 fixation is strongly reduced, leading to an obvious N_2-starvation phenotype and limited plant growth (Scott et al., 1995; Bek et al., 2010). Besides, we analyzed the *L. japonicus sen1* and *sst1* mutants which, when inoculated with wild-type *Mesorhizobium loti*, develop nodules where N_2 fixation is ineffective and display early senescence. Whereas N_2 fixation is totally abolished in *sen1*, it is reduced at 90% in *sst1* (Suganuma et al., 2003; Krussel et al., 2005). Thus, incomplete symbioses resulting from both the bacterial or the plant side were studied.

Wild-type *Gifu* or mutant plants were inoculated either with *M. loti* or inefficient *B. loti* NZP2309 and nodules as well as leaf tissues were harvested at 4wpi. Nodule phenotyping indeed confirmed the above-mentioned features of the mutants, as early senesced nodules were detected in inoculated roots (Figure R15-B). Besides, plant growth was severely limited in all these three incomplete symbiotic interactions and plants displayed nitrogen deprivation symptoms, being stronger in *sen1* and *sst1* mutants (data not shown). Subsequent Northern analysis of extracted RNA revealed that miR397 is slightly more abundant in the fully competent nodules resulting from *Gifu/M. loti* interactions, while it is still expressed in the other types of nodules (Figure R15-C). Strikingly, an analysis of leaf samples revealed a near 10-fold increase in miR397 expression in *M. loti* inoculated *Gifu* plants, as opposed to the Fix- symbiotic interactions assayed.

These results show that symbiosis, even if poorly effective, still induces miR397 expression in nodules, though at a lower level that in wild-type symbiotic interactions. By contrast, it is the occurrence of fully active N_2 fixation, rather than the presence of bacteriae, which specifically causes an increase of miR397 abundance in leaves. Therefore, activation of the N_2 metabolic pathway in nodules, which possibly involves miR397, is somehow sensed systemically throughout the plant.

In *Arabidopsis*, miR397 downregulates copper-containing enzymes, such as laccases, as a means of control of whole-plant copper levels (Yamasaki et al., 2009). Considering the predicted targets of this miRNA in *Lotus*, all belonging to the laccase copper-containing protein family (Table R8), we searched for the presence nutrient deficiency-related or other cis-regulatory elements in the *lja-MIR397* gene. While *Arabidopsis*, rice and poplar contain up to 3 *MIR397* family members (miRBase v13), a single locus was found in *L. japonicus*, though it is conceivable that additional genes will be uncovered with the completion of genome sequencing. *In silico* analysis of the *lja-MIR397* upstream sequence revealed an overrepresentation of GTAC motifs, essential for transcriptional activation upon Cu^{2+} shortage (Quinn et al., 2000; Yamasaki et al., 2009). Indeed, 9 of these motifs were present in the 3' 500-nt fragment. For comparison, the upstream sequence of *lja-MIR159*, also abundantly expressed in nodules, did not contain any GTAC motifs. Hence, this data provide a preliminary hint for a link between symbiotic-induced miR397 expression and Cu^{2+} levels in *Lotus*.

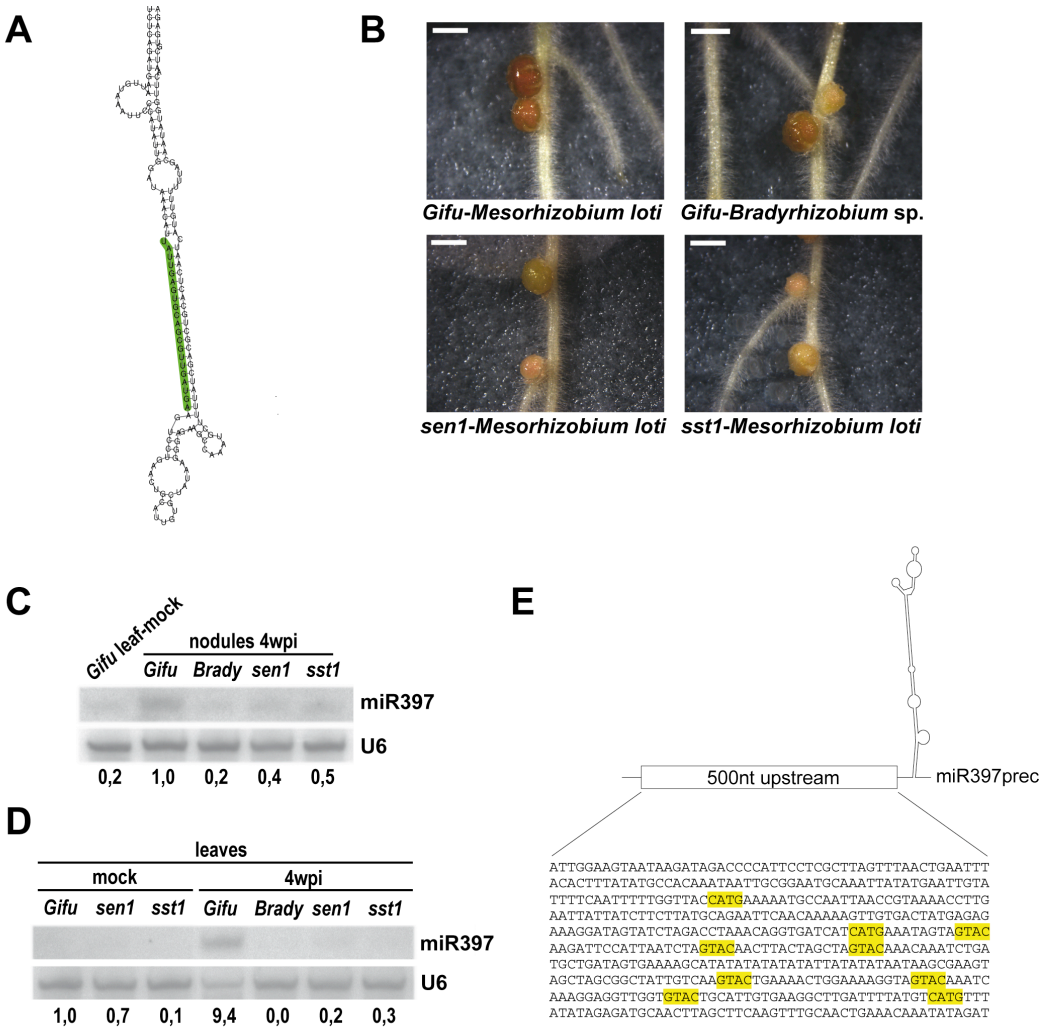


Figure R15. miR397 expression in N_2 fixation-deficient mutants. A. The single miR397 precursor found in *L. japonicus* genome. **B.** Pictures from the different nodules formed throughout the following *Lotus-Rhizobia* interactions: wild-type *Gifu* and *Mesorhizobium loti* (upper left); wild-type *Gifu* and *Bradyrhizobium* sp. (*Lotus*) strain NZP2309 (upper right); *sen1* mutant and *M. loti* (lower left); *sst1* mutant and *M. loti* (lower right). *M. loti* strain MAFF303099 was used. Scale bar =1mm. **C, D.** Analysis of miR397 expression in nodules (C) and leaves (D) from the Lotus-*Rhizobia* interactions depicted in (B). *Brady*, *Bradyrhizobium* sp. (*Lotus*). First to fourth leaves were harvested, including shoot tissue. Mock-inoculated *Gifu* leaves serve as reference among both blots. 16 ug total RNA were loaded. Quantification and normalization of northern blot signals was carried out with ImageGauge v4.1. **E.** Copper-responsive transcriptional activation motifs found in the 500nt upstream genomic sequence of lja-miR397 precursor. In silico analysis were carried out at PLACE (www.dna.affrc.go.jp/PLACE/).

3.2. DISCUSSION

3.2.1. Investigation of early symbiotic interactions through whole-root small RNA profiling.

Despite the considerable care given to RNA quantification prior to small RNA library construction, some cloning/sequencing differences found between early inoculation time points could not be further confirmed. As a result, differentially regulated miRNA genes at 3hpi or 3dpi stages of *Rhizobia-Lotus* interactions were not found. One possible explanation is that the high rRNA content in root libraries, also observed in *Arabidopsis* (Hewezi et al., 2008), strongly dilutes the small RNA populations available for analysis (Table R3). Hence, slight differences in miRNA accumulation can be amplified. A higher sequencing depth could possibly overcome this problem. A previous miRNA expression study from early *Rhizobia*-soybean interactions (Subramanian et al., 2008) also showed a high degree of divergence between cloned and detected miRNA, even when around 150.000 reads were obtained per library. In this work, Subramanian and collaborators described an induction of expression of miR393, miR159 and miR168 at 3hpi. Upregulation of these genes could not be confirmed in our studies, yet this could be due to specific differences between soybean and Lotus symbiotic interactions.

The *Rhizobia*-root symbiosis is a process that originates in discrete infection entry points along the plant root. Therefore in a whole-root cloning approach like ours, the expression of certain miRNAs in specific cell lineages or infected cells could be masked or diluted by gene expression in nonresponsive root regions, including the highly active root tip regions. Some examples can be found in the literature. For instance, transcription profiling of the *B. japonicum*-soybean interactions from 0 to 72hpi could not identify differentially expressed genes other than ENOD40, even when the highly infected supernodulation mutant was employed (Brechenmacher et al., 2008). Nevertheless, other studies have shown that the effect of the nonresponsive root region in large-scale transcriptome analysis is minimal (Hogslund et al., 2009). Taking this into account, large-scale

small RNA analysis at the root level would be most likely improved by saturating the sequencing depth instead of restricting the analyses to the infected zones.

3.2.2. A discrete set of miRNAs is differentially induced in determinate nodules

Nodules are independent organs morphologically and functionally different from the rest of the plant (Hirsch, 1992; Oldroyd and Downie, 2008). Quantitative comparative analysis of small RNA data from 3wpi nodules and uninoculated root tissue outlined a group of miRNAs (miR167, miR172, miR390 and miR397) as strong candidates for their involvement in nodule biology and a closer analysis further enriched our nodule-specific set of miRNAs through the identification of lja-miR171c. Two parallel studies of small RNA expression in nodules have recently been published. A small RNA analysis in *M. truncatula* (Lelandais-Briere et al., 2009) describes an upregulation of miR160, miR167, miR172 and miR398 in mature nodules versus root tip tissue, while miR167, miR172, miR396 and miR399 were cloned from soybean in 28dpi nodule libraries. Moreover, Northern analysis revealed that miR172 is hyperinduced in these symbiotic organs as compared to all other tissues assayed (Wang et al., 2009). While differences in sampling and sequencing depth preclude side-by-side comparisons, it is interesting that miR167 and miR172 have been uncovered in both studies, suggesting an actual role in nodule biology for these miRNAs, regardless of the determinate or indeterminate nature of the nodulation process (see below for further discussion). We also cloned miR396 from nodules, in agreement to what was found in soybean determinate nodules, although in *Lotus,* miR396 is also highly expressed in inoculated and uninoculated root tissue. However, preliminary analysis indicated a higher expression in mature nodules as compared to young nodules (Figure R10-B). This certainly points it out as a potentially interesting miRNA, especially if considered that one miR396 isoform was recently found as expressed in the N_2-fixing zone of M. truncatula nodules (Lelandais-Brière et al, 2009). Neither miR398 nor miR399 were cloned from our *L. japonicus* nodule libraries or detected by Northern blot,

which is in contrast to their accumulation reported in the above-mentioned nodule studies. While highly sensitive detection methods might be needed to uncover the potential involvement of miR399 in *Lotus* nodule biology, it is tempting to speculate that specific expression of some miRNAs might contribute to the distinct genetic networks underlying the developmental differences in determinate and indeterminate nodules. In any cases, the potential biological role of miRNA species cloned at similar or lower levels in nodules as opposed to root tissues should not be overlooked, nor should *MIRNA* genes whose low expression hampers their cloning and molecular detection. Indeed, miR169 was recently found as crucial for nodule development in *M. truncatula* (Combier et al., 2006), whereas it was absent from nodule libraries from this organism and remained below detection levels in nodule Northern analysis (Lelandais-Briere et al., 2009). Increasing the sequencing depth might therefore help to uncover the full miRNA landscape of each type of nodule tissues.

3.2.3. A miR171 variant is specifically expressed in *L. japonicus* nodules.

Enrichment of certain miRNA isoforms in nodules was recently reported in *M. truncatula* studies (Lelandais-Brière et al., 2009). In our case, we identified a miR171 variant, hereby called lja-miR171c, which displays 3 single nucleotide polymorphisms from the canonical miR171 sequences. This miRNA was found as specifically expressed in nodules, yet its low detection levels hindered further genetic analyses. Interestingly, lja-miR171c is conserved in other legume species, such as *M. truncatula* (Lelandais-Brière et al., 2009) or soybean (Subramanian et al., 2008), but also in a wider group of plant species that does not include *A. thaliana*. However, none of the previous legume studies reported an infection-specific or nodule expression for miR171c. In fact, a 20-fold expression increase in root tips was instead described in *M. truncatula* (Lelandais-Brière et al., 2009).

The sequence mismatches observed in lja-miR171c would hamper genetic regulation of the conserved targets predicted for the canonical miR171 family in

plants, namely, the SCARECROW-LIKE (SCL-LIKE) genes, which belong to the GRAS-domain transcription factor family (Jones-Rhoades et al., 2006). In fact, we predicted a divergent GRAS-domain transcription regulator, NSP2, as a potential target for miR171c in Lotus. NSP2 is crucial for Nod-factor-induced gene expression and, thus, for the subsequent establishment of symbiosis in both *L. japonicus* and *M. truncatula* (Kalo et al., 2005; Heckmann et al., 2006; Hirsch et al., 2009). It is noteworthy that NSP2 is strongly downregulated in *L. japonicus* mature nodules (Murakami et al., 2006; Hogslund et al., 2009), a pattern inversely correlated to that of miR171c. Besides, we could identify 2 miR171c precursor backbones in both *M. truncatula* and soybean, supporting a biological role for this miRNA in symbiosis. Therefore, analysis of the lja-miR171c targets will be given highly priority in future research.

3.2.4. miR167 and miR172 contribute to nodule biology independently of the presence of *Rhizobia*.

One of the most important recent advances in the understanding of legume-*Rhizobia* symbiosis has been the genetic uncoupling of the root infection and nodule organogenesis processes. In this regard, spontaneous nodulation mutants represent an invaluable tool to dissect the individual contribution of miRNA or other genes to the different aspects of root nodule symbiosis. In this study, we have shown that miR167 and miR172 are highly accumulated in the determinate *L. japonicus* nodules and that the presence of *Mesorhizobium loti* is not required for this phenomenon. This result suggests a regulatory function for both miRNAs in nodule development or maintenance trhough a genetic program that is solely contributed by the plant. Besides, two other recent independent studies have shown that both miR167 and miR172 are involved in nodule biology (Wang et al., 2009; Lelandais-Brière et al., 2009). The fact that this observation was made in different model legumes, i.e. soybean, *M. truncatula* and *L. japonicus*, suggests a role for these miRNAs in developmental steps shared by both determinate and indeterminate nodules.

Besides a strong expression of miR172 in spontaneous nodules, we have observed that it is also induced in root tissues isolated upon excision of infected nodules in both wild-type and *snf1-1* mutant, as opposed to uninoculated roots. These observations suggest that miR172 may not only be involved in nodule structural changes before bacterial colonization, but also in infection-triggered processes taking place at both the nodule and root levels. Accordingly, miR172 is expressed in both the meristematic and the infection zone in indeterminate *M. truncatula* nodules (Lelandais-Brière et al., 2009) and, furthermore, it is up-regulated at 1hpi in soybean roots (Subramanian et al., 2008). In many plant species, including *M. truncatula* and *P. vulgaris,* miR172 targets APETALA2-like (AP2-like) transcription factors (Aukerman and Sakai, 2003; Arenas-Huertero et al., 2009; Jagadeeswaran et al., 2009). The drastic upregulation of miR172 in all nodule types studied so far makes it, therefore, interesting to further study the role of the miR172/AP2 interaction in symbiosis. In fact, the widespread role of AP2-containing transcription factors in nodulation has been recently suggested by the involvement of several AP2/EREBPs (AP2/Ethylene Response Element Binding Proteins) genes in different steps of the legume-*Rhizobia* symbiosis (Andriankaja et al., 2007; Middleton et al., 2007; Asamizu et al., 2008; Vernié et al., 2008).

Similar to miR172, miR167 accumulates strongly in the different nodule types assessed so far, including in *L. japonicus* spontaneous nodules (Wang et al., 2009; Lelandais-Brière et al., 2009; this work). Moreover, we have shown that it is expressed in all nodule stages assayed in *Lotus*. Besides, it accumulates in *M. truncatula* nodule meristem and vascular bundles (Lelandais-Brière et al., 2009). Together, these data strongly point to a role in nodule development or maintenance. miR167 targets Auxin Response Factors (ARF) in many different plant species, including legumes (Jones-Rhoades et al., 2006; Jagadeeswaran et al., 2009). We have also shown that *L. japonicus* nodules express another conserved miRNA involved in auxin homeostasis in plants, the tasiRNA-generating miR390 (Yan et al., 2009). Evidence supporting the involvement of auxin transport or homeostasis in nodulation is available, though many aspects remain unknown (Grunewald et al., 2009). Thus, recent studies have revealed the specific expression

of auxin-related miRNAs, such as miR160, miR164, miR167 and miR393 upon *Rhizobi*a inoculation or in infected nodule tissue (Subramanian et al., 2008; Wang et al., 2009; Lelandais-Brière et al., 2009). Therefore, our observations provide a further hint for the involvement of auxin in determinate nodulation.

3.2.5. N_2-fixation as opposed to symbiosis induces lja-miR397 expression at the local and systemic level.

Our study has unraveled that miR397 is upregulated in *L. japonicus* nodules, especially at mature stages, and also displays an increased systemic accumulation, notably in leaves, as compared to uninoculated plants. In addition, this miRNA is depleted in *Rhizobia*-free spontaneous nodules but, when infected, *snf1-1* mutants revert this phenotype and show an induction in both nodules and leaves, although this induction is moderate as compared to wild-type plants. Thus, miR397 up-regulation is linked to *L. japonicus/M. loti* symbiosis. In addition, we have shown that nodules derived from *L. japonicus* symbiotic interactions with inefficient or absent N_2 fixation show a reduction in miR397 levels. Moreover, these Fix^- interactions lack the dramatic miR397 systemic up-regulation observed during wild-type symbiosis. Taken together, these results suggest an involvement of miR397 in N_2 fixation-derived processes, as opposed to nodule infection *per se*, at both the local and systemic levels. Therefore, miR397 constitutes the first example of a miRNA whose accumulation correlates with nodule functioning, which contrasts with the previously described miRNAs, all involved in nodule development (Combier et al, 2006; Boualem et al., 2008; Lelandais-Briere et al., 2009). However, it is important to note that miR397 expression has not been reported in an exhaustive analysis of indeterminate nodule miRNAs (Lelandais-Briere et al., 2009). Further work is therefore require to elucidate whether miR397 induction is directly or indirectly associated with determinate symbiosis.

MiR397 targets several laccase genes in both *Arabidopsis* (Jones-Rhoades et al., 2006) and *M. truncatula* (Jagadeeswaran et al., 2009), as revealed by

5'RACE analysis. Target prediction in *L. japonicus* likewise identified 6 distinct laccase-like genes as putative lja-miR397 targets. Though this requires molecular validation, it is likely that the conservation of this miRNA-target pair in the *Papilionaceae* family is extended to *L. japonicus*. Laccases are weakly characterized Cu^{2+}-containing polyphenol oxydoreductases associated with cell wall lignification and metabolism of phenolic compounds, among other cellular processes (Constabel et al., 2000; Cai et al., 2006; Burkhead et al., 2009). Interestingly, nodule differentiation involves the formation of a lignified cell-scleroid layer in determinate nodules or a suberin/lignin-enriched nodule endodermis in indeterminate nodules. These layers are thought to act as water, solute or gas barriers within the nodule (Brewin, 1991; Hirsch, 1992; Hartmann et al., 2002; Minchin et al., 2008). Consistent with this idea, several genes involved in lignin biosynthesis, including one laccase-like EST, were greatly repressed in nitrogen-fixing deficient laeghemoglobin RNAi nodules. These observations suggest that accumulation of lignin or other phenolic compounds is correlated to effective nitrogen fixation (Ott et al., 2009). In this context, it is conceivable that miR397 contributes in wild-type competent nodules to the spatio-temporal regulation of lignin or other phenolic compound metabolism by controlling laccase activity in specific cell lineages. Nonetheless, although laccases are clearly associated with lignin synthesis, additional functions in salt stress response and flowering time have been uncovered through mutation analyses of distinct *Arabidopsis* laccase genes (Cai et al., 2006). This suggests that miRNA-mediated regulation of distinct laccase genes inside the nodule might have implications beyond mere lignin synthesis.

The systemic induction of miR397 expression suggests that, in *L. japonicus*, this miRNA contributes to regulatory processes at the whole plant level that might be activated upon the onset of RNS. *Arabidopsis* miR397, together with miR398, miR408 and miR857, belong to a copper homeostasis regulatory mechanism coordinated by SQUAMOSA promoter binding protein-like 7 (SPL7). Upon copper deficiency, SPL7-responsive miRNAs are systemically activated. As as a result, their Cu^{2+}-containing targets including laccases, plantacyanin or

superoxide dismutates (Table I-2) are down-regulated so that copper is made available for essential, copper-demanding cellular processes (Abdel-Ghany and Pilon, 2008; Yamasaki et al., 2009; Pilon et al., 2009), including the electron transport chain in mitochondria and chloroplasts (Hänsch and Mendel, 2009; Palmer and Guerinot, 2009). Hence, in copper limiting conditions, vital plant functions are prioritized and ensured through miRNA action (Abdel-Ghany and Pilon, 2008). According to these results, it is thus possible that miR397 induction is coupled to Cu^{2+} deficiency and subsequent nutrient redistribution. This idea is further supported by the overrepresentation of the copper-responsive GTAC motifs in the lja-miR397 precursor upstream sequence (Yamasaki et al., 2009). Interstingly, another Cu-related miRNA, lja-miR408, was cloned from 3wpi nodules, although low detection levels on Northern analysis precluded further analysis of this miRNA. Additional evidences can be drawn from expression of miR398, which was absent in our nodule libraries and below detection levels in nodules. However, preliminary analysis showed that it was induced in roots and leaves upon inoculation with *Rhizobia*, in parallel to miR397 induction (data not shown). Intriguingly, miR398 is highly accumulated in *M. truncatula* nodules (Lelandais-Briere et al., 2009). Although extensive analyses are now required, these data support an actual involvement of copper in the nodule-triggered systemic induction of miR397.

Due to its biochemical role as an essential metalloprotein cofactor, Cu^{2+} is an important micronutrient for general plant metabolism and growth, and thus it is likely to participate also in nodulation-related processes. Though literature on this subject is scarce, it has been observed that Cu^{2+} deficiency has a negative effect on nodulation in subterraneum clover (*Trifolium subterranean L.*), lupin (*Lupinus luteus*) or faba bean (*Vicia faba*) (Cartwright and Hallsworth, 1970; O'hara et al., 1988; Seliga, 1998). Furthermore, an optimal copper supply increased nodulation, whole-plant N content and pod yield in various legume species. Importantly, upon external delivery, Cu^{2+} concentration in nodules was markedly increased, indicating that, at the whole plant level, nodules are relevant copper sinks. These

observations pointed to a direct role of Cu^{2+} in nodulation, yet its molecular basis remained unknown at the time of this study (Seliga, 1998).

At the biochemical level, nitrogen fixation requires (i) a low concentration of free oxygen, which could otherwise inactivate the O_2-sensitive nitrogenase enzymatic complex, as well as (ii) a high respiratory rate needed to produce the high demands of ATP (Horton et al., 2006). Nodule biochemistry is adapted to satisfy both requirements through different strategies, including the action of the oxygen carrier laeghemoglobin and the use of a specialized respiratory electron transport chain terminated by a Cu^{2+}-containing bb3-type cytochrome oxidase with high-affinity for O_2 (Preisig et al., 1996; Delgado et al., 1998; Ott et al, 2005). This heme-copper oxidase complex is encoded by the bacterial *fix*NOQP operon, conserved in *M. loti* and *S. meliloti* (Uchiumi et al., 2004). Cu^{2+} availability seems crucial for the function of this complex, as deletion of the adjacent *fix*GHIS genes, encoding for the copper-transporting ATPase FixI, resulted in defective nitrogen fixation in the soybean symbiont *B. japonicum* (Preisig et al., 1996b). It is therefore conceivable that, at the onset of bacterial N_2 fixation, nodules become a major copper sink in the plant. This is further supported by the specific expression of a plant copper transporter in *Medicago* nodules (Fedorova et al., 2002). However, Cu^{2+} may not only be required for specialized respiration in the nodule. Indeed, reactive oxygen species (ROS) are generated in nodules and superoxide dismutases (SOD), including Cu,Zn-binding SODs, constitute one of the various protective mechanisms from oxidative damage (Santos et al., 2001; Michin et al., 2008).

This non-exhaustive review highlights the importance of copper as a micronutrient, which might be required for different steps of RNS. In this scenario, plants might sense sudden demands of Cu^{2+} at the onset of N_2 fixation and thus activate systemic nutrient redistribution processes, including those in involving miRNAs. Whether miR397 transcription is induced systemically or a result of systemic translocation will need to be tested. In addition, it would be important to study whether an enhanced copper supply reverts miR397 accumulation systemically or even at the nodule level. It is not unreasonable to conceive

miRNAs as sentinels or actors in the adjustment of optimal nutritional ranges during N_2-fixing symbiotic interactions.

IV. CONCLUDING REMARKS AND PERSPECTIVES

The study of small RNAs and, more specifically, microRNAs in legumes has recently received great attention due to the key roles for these molecules in virtually all aspects of plant biology. Beyond the interest of mere legume miRNA discovery, our and other studies allow old questions to be re-assessed under a new light. RNS being such a complex biological process, the use of distinct but complementary model legumes has greatly multiplied the scientific breakthroughs in the recent years, especially at the molecular level. Following these lines of thought, we have investigated both the RNA silencing machinery and the small RNA populations from the determinate model legume *Lotus japonicus*,

The first outcome of this study is the identification of the RNA silencing machinery in *L. japonicus*, which paves the way for the generation of mutants required for the genetic dissection of small RNAs in legumes. Our experience highlights the pressing need of knock-out technologies such as T-DNA insertional mutant collections. In addition, we have obtained and characterized *L. japonicus* P19-expressing transgenic lines, as a complement to novel miRNA annotation in studies in this organism but also for other legume-specific miRNAs. In addition, using a high-throughput small RNA sequencing approach we have identified 27 known miRNA families, including 3 legume-specific, and predicted 28 potential novel miRNAs in Lotus. Molecular analyses have confirmed 2 Lotus-specific miRNAs from the latter group, while highly sensitive detection methods are now needed for the confirmation of the rest as *bona-fide* miRNAs.

A set of important findings that emerges from this work pertains to miRNA expression in nodules. Differential analysis of cloned small RNAs from root and mature nodule samples and transcriptome analysis of pri-miRNA have revealed that *L. japonicu*s mature nodules express a set of conserved miRNAs that is either absent or less expressed in root tissues: miR167, miR171c, miR172, miR390, miR397 and miR408. These, therefore, represent strong candidates for their

involvement in determinate nodulation. Both miR167 and miR172 are highly accumulated in indeterminate *M. truncatula* and determinate soybean nodules, and our assessment of spontaneous *Rhizobia*-free nodules further shows that their expression in mature nodules is independent, not only of the nodule type, but also of bacterial infection. This fact places these miRNAs at the nodule developmental steps and paves the way for in-depth studies. Besides, we have shown that RNS induces miR397 expression at both local and systemic levels, and genetic analyses have revealed that this induction is largely dependant on active N_2 fixation rather than on bacterial colonization. MiR397, thus, comes out as the first miRNA involved in nodule functioning, and future studies will aim at identifying the basis of such induction and its biological function(s).

Despite these results, our research has opened several lines of work. Beyond the miRNA pathway, we have indications for other small RNA being implicated in nodule biology, notably AGO5-dependent heterochromatic siRNAs. Much progress could be achieved in this regard by coupling genetic and biochemical approaches, i.e. immunoprecipitation of AGO5-like genes from mature nodule and spontaneous nodule tissues. This could indeed be instrumental to narrow down the spectrum of cloned siRNAs and identify those involved in infection as opposed to the other processes of nodulation. Nonetheless combining our work with previous studies, a large set of conserved miRNAs already emerge as attractive candidates for their roles in symbiosis. Expression analysis of their target genes will now be required in order to unravel the contribution of each of these miRNAs to the complex regulatory networks taking place in RNS. More compelling is our discovery of a specialized *MIR171* gene, not conserved in *Arabidopsis*, which is encoded by legumes and expressed in a nodule-specific manner in *L. japonicus*. miR171c is a strong candidate as an expression modulator of NSP2, a crucial regulator of the early nodulation signaling pathway. This is indeed an exciting issue that might give new insight into both early and late symbiotic events. Finally, it would be of upmost interest to test whether miRNAs can be used as an entry door to investigate the complex and yet poorly understood hormonal regulatory system that underlies Root Nodule Symbiosis.

V. MATERIALS AND METHODS

1. Biological material

1.1. Plant material

Lotus japonicus accession 'Gifu' B-129 (Handberg and Stougaard, 1992) and *Lotus japonicus* accession 'Miyakojima" MG-20 (Kawaguchi et al., 2001) were used in this study. In addition, this work included the following mutants, all in the *Gifu* background: *snf1-1* (Tirichine et al., 2006), *snf2* (Tirichine et al., 2007), *sen1* (Suganuma et al., 2003), *sst1* (Krusell et al., 2005).

1.2. Bacterial strains

Mesorhizobium loti strain MAFF303099 and *Bradyrhizobium* sp. (*Lotus*) strain NZP2309 (Scott et al., 1995; Bek et al., 2010) were used for symbiotic assays. Plant transformation was carried out with *Agrobacterium tumefaciens* strain GV3101. DNA cloning was performed using *Escherichia coli* strain DH5α.

2. Plant growth & culture

Lotus japonicus seeds were placed in reaction tubes and 2 volumes of H_2SO_4 were added. After a 10-minute incubation, seeds were washed 4 times with 1 ml sterile water and then incubated for 10 minutes with a bleach sterilization solution (5% sodium hypochlorite, 0,1% SDS) on a rotary shaker. Seeds were then washed 4 times with sterile water and kept imbibed at 4ºC overnight. Swollen seeds were placed on 0,8% agar GREINER 13x13 cm plates and kept overnight at 4ºC to synchronize germination. Plates were then placed vertically in a SANYO growth chamber at 23ºC, germinated in the dark for 3 days and then kept for 3 days at a 16/8h day/night cycle and daylight intensity of 4000 lux. For *in vitro* culture**,** seedlings were then transferred to B5 agar plates (Gamborg B5 with vitamins (Duchefa) 0,316%, sucrose 2%, agar 0,8%, pH 7,0) where the media was covered

by a sterile wet filter paper. For a proper root system development, plates were covered by dark plastic stripes at the root area level and kept vertical. Plants were grown at 16/8h day/night cycle and 23ºC. For greenhouse culture, plants were transferred to pots containing either SERAMIS® substrate or a mixture of 1/3 perlite and 2/3 N-rich soil (Fruhstorfer Erde, Hawita Gruppe GmbH, Germany) and grown at 16/8h day/night cycle, 22ºC/18ºC and 15.000 lux. *Arabidopsis thaliana* plants used as controls in this study were grown in greenhouse under the conditions above.

3. Bacterial culture

M. loti were grown on Triptone Yeast (TY) medium (bactotryptone 0,5%, yeast extract 0,3%, $CaCl_2$ 0,13%, pH 7,0) for 48h at 28ºC. For the assay of N_2 fixation-deficient nodules (see chapter 3.1.5.), *M. loti* and *Bradyrhizobium* sp. (Lotus) were grown on Yeast Mannitol Broth (YMB) medium (mannitol 1%, yeast extract 0,04%, K_2HPO_4 0,05%, 0.2 g $MgSO_4.7H_2O$, NaCl 0,01%, phoshomycin 100μg/ml, pH 6.8) for 72h and 96h, respectively. Tetracyclin (2,5 μg/ml) was added for *Bradyrhizobium* sp. (Lotus) selection. *A. tumefaciens* and *E. coli* were cultured in Luria Bertoni (LB) medium (bactotryptone 1%, yeast extract 0,5%, NaCl 0,5%, pH 7,4), adding the corresponding antibiotics, and incubated overnight in shakers at 28°C and 37°C, respectively.

4. Inoculation methods

4.1. Growth media for inoculation assays

Fahraeus (FP):		**Broughton and Dilworth (B&D):**	
$CaCl_2 \cdot 2H_20$	0,7 mM	$CaCl_2 \cdot 2H_2O$	1.0 mM
$MgSO_4$	1 mM	KH_2PO_4	0.5 mM
KH_2PO_4	0,7 mM	Fe-citrate	10 μM
Na_2HPO_4	1 mM	$MgSO_4 \cdot 7H_2O$	0.25 mM
Fe-citrate	20 μM	K_2SO_4	0.25 mM

H_3BO_3	46 μM	$MnSO_4 \cdot H_2O$	1.0 μM
$MnSO_4 \cdot H_20$	12 μM	H_3BO_3	2.0 μM
$ZnSO_4 \cdot 7H_20$	0,76 μM	$ZnSO_4 \cdot H_2O$	0.5 μM
$CuSO_4 \cdot 5H_20$	0,32 μM	$CuSO_4 \cdot 5H_2O$	0.2 μM
Na_2MoO_4	0,4 μM	$CoSO_4 \cdot H_2O$	0.1 μM
Optional: agar	0,8%	$Na_2MoO_2^4 \cdot 2H_2O$	0.1 μM
(Fahraeus, 1957)		Agar noble	1%
		(Broughton and Dilworth, 1971)	

4.2. *In vitro* inoculation

Two week-old *Lotus japonicus* seedlings were transferred to GREINER plates containing FP medium (see above) under a layer of sterile filter paper. *Mesorhizobium loti* were previously grown on TY medium (see above). Bacterial culture was centrifuged twice at 9.000 rpm, resuspended in liquid FP medium (OD_{600} at 0,1) and 50μl of this suspension was applied directly to each root. Plates were covered by plastic stripes at the root level and placed in the culture chamber, 23ºC, 16h light/8h dark regime. Different plant tissues were dissected with a scalpel and harvested at the established inoculation points.

Specifically for the assay of N_2 fixation-deficient nodules (see chapter 3.1.5.), kindly carried out by K. Markmann (Department of Molecular Biology, Aarhus, Denmark), germinated *Lotus* seedlings were transferred to 1/4xB&D medium plates (see above). *M. loti* and *Bradyrhizobium* sp. (*Lotus*) were grown on YMB medium (see above), pelleted and washed twice with 1/4xB&D medium and then resuspended in liquid 1/4xB&D medium (OD_{600} at 0,01). 1 ml of inoculum was used per plate, each containing 15 plants. Plates were covered by an aluminum wedge and grown at 21ºC constant, 16h light/8h dark regime. Plant tissues were collected at 4wpi using a scalpel and stored at -80ºC.

4.3. Flood-inoculation

Sterile culture boxes or open culture trays were filled with autoclaved Seramis® substrate and soaked with sterile FP medium (see above). 1-week old agar-germinated *Lotus japonicus* seedlings were subsequently transferred and grown for 2 more weeks. *Mesorhizobium loti* were grown at 28ºC for 48h on TY medium and the resulting culture was centrifuged at 9.000 rpm, resuspended in sterile FP medium to an OD_{600} of 0.01 and applied to the culture boxes by flooding. Control plants were flooded with sterile FP medium alone and excess media was removed by pipetting. Plant material was collected with N_2 at established timepoints and kept at -80ºC until processing.

5. Vectors and cloning procedures

The β-Glucuronidase (GUS) and the tomato bushy stunt virus (TBSV) P19 genes were cloned into the pDONR221 vector (Invitrogen) and then into the pK7WG2D vector (Karimi et al., 2002) using the Gateway® technology and following the manufacturer's instructions (Invitrogen). Plasmid DNA was isolated from bacteria by using a DNA 'mini-prep' kit (Macherey Nagel) and sent to the sequencing service using specific primers.

6. Plant transformation

Generation of *L. japonicus* stable transformants was carried out starting from root explant tissue and strictly following the protocol established by Lombari and collaborators (Lombari et al., 2003) except that transformed calli were selected using the antibiotic G418 (7 μg/ml) instead of kanamycin, as it is more effective in *Lotus* (Tirichine et al., 2005).

7. GUS staining

Plant tissues were vacuum-infiltrated three times with the GUS staining solution (50 mM Na_2HPO_4 pH 7.0, 0.5% Triton X-100, 10 mM EDTA sodium salt, 0.5 mM $K_3Fe(CN)_6$, 0.5 mM $K_4Fe(CN)_6$.3H2O and 2 mM X-Gluc), incubated 37ºC overnight and subsequently washed for 1h with increasing concentrations of ethanol (20%, 50% and 70%), following by a last 70% ethanol wash overnight.

8. Plant imaging

Pictures were taken using a SMZ1500 dissecting binocular (Nikon). GFP observations were carried out using a 100W epi-fluorescence module. Measure of plant growth was done either manually or with the assistance of ImageJ 1.43g (National Institutes of Health, USA). Mutant nodule pictures were obtained using a Leica M165FC stereomicroscope and are courtesy of K. Markmann (Department of Molecular Biology, Aarhus, Denmark).

9. DNA genotyping

Plant genomic DNA from TILLING mutant lines (Sainsbury Laboratory) was extracted and PCR-amplified with the Extract-N-Amp PCR Kit (Sigma) for genotyping and sequencing applications. DNA from *Ljdcl1-1* (SL-693) plants was amplified using specific primers (forward: CAGAGATACCATGGAGA CACAAA; reverse: GGCTTGGCTGGTTGATAA AA) and digested with XmaJI (Fermentas). DNA from *Ljdcl1-2* (SL-2749) mutants was amplified by PCR (forward: TGAAGAGGATCATGACAAGACTTA; reverse: AACAGTAGAAGT TAAACTACCTTGAGC) and digested with HphI (Fermentas). Restriction fragments were separated by electrophoresis in horizontal agarose gels (1% agarose, 0.5X TBE buffer) at 100V. Analytical gels were photographed on a short wavelength UV trans-illuminator (BioRad).

10. RNA protocols

10.1. RNA extraction protocols

10.1.1. *Lotus japonicus* RNA extraction

10.1.1.1 Adapted *Hot Borate Buffer* protocol

This protocol was adapted from Wan and Wilkins (1994) and set up for the extraction of large quantities of RNA necessary for Northern analysis and small RNA cloning. 100-150mg of plant tissue was ground under liquid nitrogen and the resulting powder was homogenized in 0,5ml of 95ºC-heated extraction buffer (2000mM Sodium Tetraborate Decahydrate, 30mM EGTA, 5mM EDTA, 1% SDS, 1% Sodium Deoxycholate; adding before use 100µl DTT, 35µl ß-Mercaptoethanol and 0,1g PVP-40 per 5ml buffer). For protein digestion, 0,5mg of Proteinase K was added and samples were incubated 45 min at 37ºC on a shaker. The reaction was stopped by adding 39µl of 2M KCl and placing the samples on ice for 30 min, followed by a 10 min centrifugation at 12.000 rpm at 4ºC. Supernatant was transferred to a new tube and TRIzol-LS® Reagent (Invitrogen) was added to a ratio of 1,2ml per 400µl supernatant. Following homogeneization and incubation at RT for 5min, samples were centrifuged for 10 min at 12.000rpm (4ºC) for polysaccharide removal. Samples were subsequently extracted with 320µl Chlorophorm and incubated at RT for 10 min. After a further centrifugation step of 15 min at 13.000 rpm (4ºC), the upper phase was transferred to a new tube, 600µl isopropanol were added and the mixture incubated at RT for 20 min. A last centrifugation step (20 min, 13.000 rpm, 4ºC) was carried out to collect RNA precipitate. The supernatant was discarded and the pellet washed with 1600µl of 80% ethanol, then allowed to air-dry at RT for 5-10 min. The pellet was resuspended in 20µl of 50% formamide by repeated pipetting and heating at 65ºC for 2 min on a shaker. OD was measured at a 1/100 dilution in sterile water. Quality of the total RNA extract and equal loading of the samples were assessed by loading 0,3µg from a 1/20 sample dilution on an Et-Br-stained 1% agarose gel.

10.1.1.2 RNA extraction protocol for qRT-PCR

Extraction of RNA of high purity was performed by using the PureLink® Plant RNA Reagent (Invitrogen) following the manufacturer's instructions. RNA samples were resuspended in 100µl RNase-free water and, to increase quality, purified following the *RNA Cleanup* protocol from the RNeasy Mini Kit (Qiagen). For quantification, OD was measured on a spectophotometer and RNA quality was checked by loading 1µl on a EtBr-stained 1% agarose gel. Integrity of mRNA was checked by amplification of Ubiquitin by RT-PCR (see below).

10.1.2. *Arabidopsis thaliana* RNA extraction

100 mg of plant tissue was ground under liquid nitrogen. The resulting powder was homogenized in 100 mL of TRIZOL® Reagent (Invitrogen), a mono-phasic solution of phenol and guanidine isothiocyanate. 200 µL of chloroform was added, the mixture thoroughly mixed and left at room temperature (RT) for 2 minutes (min). The aqueous and organic phase were separated by centrifugation at 4°C and at maximum speed (13, 000 g) for 15 min. The upper aqueous phase (approximately 600 µL) was transferred to a new tube and mixed with 1 volume of isopropanol. The mixture was left at RT for a maximum of 10 minutes. A second centrifugation step (4°C, 13,000g, 15 min) was carried out to collect the RNA precipitate. The supernatant was discarded and the pellet washed with 80% ethanol, then allowed to air-dry at RT for 5-10 min. The pellet was resuspended in 20-40 µL of 50% formamide by repeated pipetting. OD was measured at a 1/100 dilution in sterile water. Quality of the total RNA extract and equal loading of the samples was assessed by loading 0.5 µg on an agarose gel.

10.2. Northern analysis

10.2.1. High Molecular Weight (HMW) RNA blots

A denaturing agarose gel (1%) was prepared by melting 2.5g of agarose in 250ml of deionised water. 25ml of 10X HEPES buffer (HEPES 200mM; EDTA 10mM; KOH 9.5g/l; pH 7,2) was added and the solution was allowed to cool down

at RT. 40,5ml of formaldehyde was added, the solution was mixed and poured to set the gel. 3 volumes of RNA loading buffer (10X HEPES: 200µl, deionised formamide: 1000µl, formaldehyde: 320µl, bromophenol blue 0.1% in 50% glycerol: 120µl, 0,5% EtBr: 8µl, for a 1,650ml stock) were added to one volume of each RNA sample to test (5-10 µg of total RNA), previously extracted through the Hot Borate Buffer protocol (see above). Samples were denatured at 65°C for 10 min and then chilled on ice for 5 min. After loading of the samples, the gel was run at 80-100V for 5-6h in 1X HEPES and photographed under UV. RNA was transferred overnight by capillarity in 20XSSC, on a neutral nylon membrane Hybond™-NX (Amersham). The membrane was UV cross-linked in a Stratalinker® (1200 Joules x100) and looked under UV to check for efficient RNA transfer.

10.2.2. Low Molecular Weight (LMW) RNA blots

17,5% denaturing polyacrylamide gels were prepared by mixing 12.6g of urea, 13.1ml of 40% acrylamide/bis-acrylamide (ratio 19:1) and 1.5ml of 10X TBE buffer in a 50ml Falcon® tube. After urea dissolution at 37°C, water was added to a final volume of 30 mL. 100µL of 25% APS and 11µl of TEMED were added just before pouring the mixture into a BioRad Protein-III apparatus.

RNA samples previously obtained through the Hot Borate Buffer protocol (see above), were denatured at 95°C for 4 min, chilled on ice for 5-10 min and then mixed to 4X loading buffer (50% glycerol, 50mM Tris pH 7.7, 5mM EDTA, 0.03% bromophenol blue). After a 30 min pre-run of the gels in 0.5X TBE buffer, wells were washed twice with 0.5X TBE buffer, RNA samples were loaded and the gels were run at 80V for 4h. RNAs were then transferred on a neutral nylon membrane Hybond™-NX (Amersham) in 0.5X TBE buffer in a BioRad electroblot apparatus for 1 h at 80V (300mA) and at 4°C. For detection of highly expressed microRNAs, the membrane was equilibrated on Whatmann paper soaked with 2X SSC for 10-20 min and UV cross-linked twice in Stratalinker® (1200 Joules x100). For detection of low expressed miRNAs, RNAs were chemically cross-linked to the membrane with a fixing solution (0,16M *N*-ethyl-*N'*-3-dimethylaminopropyl

carbodiimide hydrochloride (EDC); 127 mM 1-methylimidazole; pH 8,0), and incubation at 60ºC for 3h (Pall and Hamilton, 2008).

10.2.3. Hybridization and washes

10.2.3.1. Probe labelling

HMV RNAs northern blots were hybridized with random-priming-labelled DNA probes. 50-100 ng of purified DNA fragment were denaturated at 95°C for 5 min then put on ice for 5 min. Random Priming Kit reactions were prepared following the manufacturer's instructions (Promega) and using 25 μCi of ^{32}P-CTP. Non-incorporated nucleotides were removed by loading the reaction onto a G50 sephadex column (prepared by centrifugation of 600 μL G-50 sephadex at 2.000 g for 2 min) and carrying out a centrifugation step at 2.000 g for 2 min. The purified probe was denatured at 95°C for 5 min and incubated on ice for 10 min.

LMW RNAs northern blots were hybridized with end-labelled DNA oligonucleotides for the detection of miRNAs. 20pmol of DNA oligonucleotide were radiolabelled using T4-Polynucleotide Kinase (PNK) enzyme and buffer (New England Biolabs) and 25 μCi of ^{32}P-γATP. The reaction was incubated at 37°C for 45 min. Non-incorporated nucleotides were removed as described above using a G25 Sephadex column (Amersham). Sequences for DNA oligonucleotides used as probes in this study are listed in Table M1.

Name	Oligonucletide sequence
lja-4	AGATGATGTATGGATGGAAGA
lja-10	TCAGAGGCGAAGACACCCGCTA
lja-14	AGCACTGCAATTTCATGATGA
miR156	GTGCTCACTCTCTTCTGTCA
miR159	GGAGCTCCCTTCACTCCAAT
miR160	GCGTATGAGGAGCCAAGCA
miR160*	TGCTTGGCTCCTCATACGC
miR167	AGATCATGCTGGCAGCTTCA
miR171c	GAGTGATATTGATTCGGCTCA
miR172	CTGCAGCATCATCAAGATTCT
miR390	GGCGCTATCCCTCCTGAGCTT
miR393	GGATCAATGCGATCCCTTTGGA
miR396	CAGTTCAAGAAAGCTGTGGAA
miR397	TCATCAACGCTGCACTCAATA
miR398	AAGGGGTGACCTGAGAACACA

Table M1. DNA oligonucleotides used for γATP radiolabelling.

10.2.3.2. Hybridization

The membranes were pre-hybridized for 30 min in PerfectHyb Plus buffer (Sigma) and, after probe addition, incubated overnight at 65°C for HMV RNAs and at 42°C for LMW RNAs.

10.2.3.3. Washes

HMW membranes were washed twice for 20 min at 65°C in a low stringency washing buffer (2X SSC, 0.1% SDS) and, if necessary, another 20 min in a high stringency buffer (0.5X SSC, 0.1% SDS). LMW RNA membranes were washed twice for 20 min in a low stringency washing buffer (2X SSC, 2% SDS) and, if necessary, another 10 min in a high stringency buffer (2 X SSC, 1% SDS).

10.3.3.4. Signal visualization

After hybridization and washes, the membranes were analyzed using a PhosphorImager (FLA-7000, Fujifilm). In LMW northerns, bands were quantified using ImageJ 1.43g (National Institutes of Health, USA) or ImageGauge version

4.1. The U6 snRNA signal was used as internal loading control for the quantification step. Data were normalized against the reference sample.

10.3. Quantitative RT-PCR

1 μg of total RNA was incubated with DNaseI (1u/μg RNA) for 50 min at 37ºC (Promega) and reverse-transcribed into polydT cDNAs using the SuperScript III reverse transcriptase (Invitrogen) in accordance with the manufacturer's recommendations. The cDNAs were quantified using SYBR Green Q-PCR kit (Eurogentech) and gene-specific primers on a LightCycler 480 II apparatus (Roche) in accordance with the manufacturer's recommendations. PCR were carried out in 384-well optical reaction plates heated to 95°C for 10 min to activate hot start Taq DNA polymerase, followed by 42 cycles of denaturation at 95°C for 15s and an annealing-extension at 62°C for 30s. Ubiquitin-conjugating enzyme 10 (LjUBC10), GPI-anchored protein (LjGPI) and Ubiquitin (LjUBIQ) were selected as housekeeping genes for cDNA quantification (Czechowski et al., 2005; Sanchez et al., 2008). Primer sequences are listed in Table M2.

Name	Primer sequence	CDS sequence	Reference
LjGPI-F	AGGTTGTTCCGTGAATTTCG	chr3.CM0047.37	Sanchez et al., 2008
LjGPI-R	GGTCCTTTGCATTTGCTTGT	chr3.CM0047.37	Sanchez et al., 2008
LjUBC10-F	GCTCTTATCAAGGGACCATCAG	chr1.TM0487.4	Sanchez et al., 2008
LjUBC10-R	ACTGCTCTGGAACAGAAAAAGC	chr1.TM0487.4	Sanchez et al., 2008
LjUBQ-F	TTCACCTTGTGCTCCGTCTTC	chr5.CM0956.27	Sanchez et al., 2008
LjUBQ-R	AACAACAGCACACACAGACAATCC	chr5.CM0956.27	Sanchez et al., 2008
LjARF6-like-F	AGTATTGGTGAATCAGGCTTCC	chr4.LjT16L01.06.nd	-
LjARF6-like-F	CTGGAAGGTGAGTTTCCTTGA	chr4.LjT16L01.06.nd	-
LjNAC-like1-F	CTCCGATCCACACAGATGAA	chr3.CM0279.430.n	-
LjNAC-like1-R	GGGTTGGTTTGGTTTTGAGA	chr3.CM0279.430.n	-
LjNAC-like2-F	GGGTTATTTGCAGGGTGTTTCA	chr4.CM0087.8	-
LjNAC-like2-R	GAATCCAACCTCATTATCCCAGAA	chr4.CM0087.8	-

Table M2. qRT-PCR primer sequences used in this study

10.4. Cloning of small RNAs

Isolation and cloning of small RNAs from *Lotus japonicus* roots and nodules were carried out as previously described (Pfeffer, 2007). Briefly, 150 μg of total RNA were separated by denaturing PAGE and the 19-24 nt fraction isolated with the help of radiolabelled size markers. Small RNAs were ligated to 3' RNA adaptors (5'-TTTAACCGCGAATTCCAG-L-3'; L: blocking C7 amino group) and subsequently to 5' RNA adaptors (5'-ACGGAATTCCTCACTrArArA-3'). Reverse transcription was performed and cDNAs were amplified by 20 PCR cycles (RT/Reverse primer: 5'-GACTAGCTGGAATTCGCGGTTAAA-3'; Forward primer: 5'-CAGCCAACGGAATTCCTCACTAAA-3'). A fraction of cloned small RNAs were concatamerized, subcloned through the TOPO reaction (Invitrogen) and sequenced through the Sanger method to assess library quality. Then, individually cloned small RNAs were amplified through a 2nd PCR round using primers carrying distinct 4-nt tags for each of the five libraries. Samples were pooled and underwent two rounds of 454 pyrosequencing.

11. Bioinformatics

11.1. Small RNA analysis

11.1.1. Library annotation and analysis

Following 454 sequencing, adaptor sequences were bioinformatically removed. High quality sequences were annotated with blast (word size = 7 / no filter) using in-house analysis platform (bioimage.u-strasbg.fr) and the following databases as references: plant microRNA sequences were analyzed using miRbase v13; tRNA, rRNA, and other non-coding RNA sequences were extracted from Genbank (release 169); *M. loti* genomic sequence was retrieved at Rhizobase (14/03/2008); DNA repeat sequences were obtained from Repbase v14.02; *L. japonicus* genomic sequences were retrieved at Kazusa Institute (2008). The results

were filtered to authorize 1 mm to *M. loti* and *L. japonicus* genomic databases and 2 mm for the rest of databases.

11.1.2. miRNA prediction, profiling and analysis

Novel miRNA identification was carried out using miRCat, selecting default parameters except for read abundance (>1) (Moxon et al., 2008). MiRNA abundance profiling was done using miRProf (miRBASE v13), allowing 2 mm and/or overhangs for the sequence search and selecting default values for output grouping (Moxon et al., 2008). MiRNA nucleotide bias was analyzed through the WebLogo interface (weblogo.berkeley.edu/logo.cgi).

11.1.3. Target prediction

MiRNA target predictions were carried out with the Target Finder algorithm (Allen et al., 2005; Falhlgren et al., 2007).

11.2. Retrieval and prediction of *Lotus* CDS sequences

Search for gene ortholog sequence in *L. japonicus* was carried out through the NCBI database, the Kazusa Institute's *Lotus japonicus* Genome Database (www.kazusa.or.jp/lotus/blast.html), as well as through personal communication with S. Sato and authorized access to the Lotus Genome Interface (www.brics.dk/~compbio/LotusBlast/). CDS gene prediction was carried out through GeneMark (http://exon.biology.gatech.edu/eukhmm.cgi) (Lomsadze et al., 2005), and Twinscan Gene Predictor (http://mblab.wustl.edu/nscan/submit/) (Gross and Brent, 2006). Protein domains were analyzed by scanning protein sequences against the InterPro signature database (www.ebi.ac.uk/InterProScan).

11.3. Sequence alignment and phylogenetic analysis

Multiple alignments were done using the ClustalW2, Align or MUSCLE (Edgar et al., 2004) algorithms at EMBL-EBI (www.ebi.ac.uk). Phylogenetic analyses were conducted using the Phylogeny.fr interface (Dereeper et al., 2008). Trees were constructed using the maximum likelihood-based algorithm PhyML 3.0

(Guindon et al., 2003) choosing the Dayhoff matrix model for aminoacid substitution and bootstrap with 100 replications. Trees were visualized using FigTree v1.2.3.

11.4. Promoter analysis

Promoter analyses were carried out through the Plant Cis-acting Regulatory DNA Elements (PLACE; www.dna.affrc.go.jp/PLACE/) (Higo et al., 1998).

12. Analysis of Affymetrix transcriptome data

Normalized and pairwise-compared expression data from *L. japonicus* Affymetrix GeneChip® were kindly provided by K. Markmann and S. Radutoiu (Department of Molecular Biology, Aarhus, Denmark) (M, FDR; where M is the log2 ratio of average signal intensity values from any two conditions and FDR is the False Discovery Rate-corrected p-value) (Hogslund et al., 2009). Probe Sets (IDs) for each predicted CDS were obtained through sequence alignment. Nodule expression values as compared to those from uninoculated root samples of the same genotype were analyzed. A significance criterion of FDR-corrected p-value ≤ 0,05 was applied (Hogslund et al., 2009). Fold change values were calculated as $|M|^2$, being M the log2 ratio of nodule/root expression.

13. Statistical analysis

Chi-square and unpaired *t*-test analysis were carried out using GraphPad. Distribution analysis and ANOVA tests were done using StatGraphics.

VI. ACKNOWLEDGEMENTS

We would like to thank M. Charpentier and M. Parniske for initial training on *L. japonicus* experimental protocols, plant space and useful discussions. We acknowledge S. Sato and S. Tabata for genome access ahead of publication. We also thank J. Perry and T. Wang for TILLING mutant seeds. We appreciate the help of C. Himber, who contributed to the generation of transgenic lines, V. Cognat, who enthusiastically assisted the bioinformatic analysis all throughout the project, and C. Ciaudo, who represented an important help to the small RNA library construction and sequencing. The greenhouse team at IBMP is also acknowledged for excellent plant care. We are also grateful to F. Schwach and S. Moxon for running the UEA Small RNA Snalysis Toolkit on a private basis. Lastly, we thank K. Markmann and J. Stougaard for the experimental set up of the nodulation mutant study and useful feedback, together with S. Radutoui, for access to full *Affymetrix* data. This work was partly supported by the Marie Curie MRTN-INTEGRAL network.

VII. REFERENCES

Abdel-Ghany, S.E., and Pilon, M. (2008). MicroRNA-mediated systemic down-regulation of copper protein expression in response to low copper availability in Arabidopsis. J Biol Chem **283**: 15932–15945.

Addo-Quaye, C., Eshoo, T.W., Bartel, D.P., and Axtell, M.J. (2008). Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. Curr Biol **18**: 758–762.

Allen, E., Xie, Z., Gustafson, A.M., and Carrington, J.C. (2005). microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell **121**: 207–221.

Allen, E. et al. (2004). Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana. Nat Genet **36**: 1282–1290.

Anandalakshmi, R. et al. (1998). A viral suppressor of gene silencing in plants. Proc Natl Acad Sci U S A **95**: 13079–13084.

Andriankaja, A. et al. (2007). AP2-ERF transcription factors mediate Nod factor dependent Mt ENOD11 activation in root hairs via a novel cis-regulatory motif. Plant Cell **19**: 2866–2885.

Arenas-Huertero, C. et al. (2009). Conserved and novel miRNAs in the legume Phaseolus vulgaris in response to stress. Plant Mol Biol **70**: 385–401.

Asamizu, E., Shimoda, Y., Kouchi, H., Tabata, S., and Sato, S. (2008). A positive regulatory role for LjERF1 in the nodulation process is revealed by systematic analysis of nodule-associated transcription factors of Lotus japonicus. Plant Physiol **147**: 2030–2040.

Aukerman, M.J., and Sakai, H. (2003). Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. Plant Cell **15**: 2730–2741.

Axtell, M.J., and Bowman, J.L. (2008). Evolution of plant microRNAs and their targets. Trends Plant Sci **13**: 343–349.

Barker DG, et al. (1990). Medicago truncatula, a model plant for studying the molecular genetics of the Rhizobium-legume symbiosis. Plant Molecular Biology Reports **8**: 40–49.

Baumberger, N., and Baulcombe, D.C. (2005). Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proc Natl Acad Sci U S A **102**: 11928–11933.

Bek, A.S. et al. (2010). Improved Characterization of Nod Factors and Genetically Based Variation in LysM Receptor Domains Identify Amino Acids Expendable for Nod Factor Recognition in Lotus spp. Mol Plant Microbe Interact **23**: 58–66.

Benedito, V.A. et al. (2008). A gene expression atlas of the model legume Medicago truncatula. Plant J **55**: 504–513.

Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., and Zhu, J.K. (2005). Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. Cell **123**: 1279–1291.

Bortolamiol, D., Pazhouhandeh, M., Marrocco, K., Genschik, P., and Ziegler-Graff, V. (2007). The Polerovirus F box protein P0 targets ARGONAUTE1 to suppress RNA silencing. Curr Biol **17**: 1615–1621.

Boualem, A. et al. (2008). MicroRNA166 controls root and nodule development in Medicago truncatula. Plant J **54**: 876–887.

Brechenmacher, L. et al. (2008). Transcription profiling of soybean nodulation by Bradyrhizobium japonicum. Mol Plant Microbe Interact **21**: 631–645.

Brewin, N.J. (1991). Development of the legume root nodule. Annu Rev Cell Biol **7**: 191–226.

Brigneti, G. et al. (1998). Viral pathogenicity determinants are suppressors of transgene silencing in Nicotiana benthamiana. EMBO J **17**: 6739–6746.

Brodersen, P. et al. (2008). Widespread translational inhibition by plant miRNAs and siRNAs. Science **320**: 1185–1190.

Brodersen, P., and Voinnet, O. (2006). The diversity of RNA silencing pathways in plants. Trends Genet **22**: 268–280.

Brodersen, P., and Voinnet, O. (2009). Revisiting the principles of microRNA target recognition and mode of action. Nat Rev Mol Cell Biol **10**: 141–148.

Broughton, W.J., and Dilworth, M.J. (1971). Control of leghaemoglobin synthesis in snake beans. Biochem J **125**: 1075–1080.

Burkhead, J.L., Reynolds, K.A., Abdel-Ghany, S.E., Cohu, C.M., and Pilon, M. (2009). Copper homeostasis. New Phytol **182**: 799–816

Cai, X. et al. (2006). Mutant identification and characterization of the laccase gene family in Arabidopsis. J Exp Bot **57**: 2563–2569.

Cannon, S.B., May, G.D., and Jackson, S.A. (2009). Comparative Genomics of Legumes Three Sequenced Legume Genomes and Many Crop Species: Rich Opportunities for Translational Genomics. Plant Physiol **151**: 970–977

Cartwright, B., and Hallsworth, E.G. (1970). Effects of copper deficiency on root nodules of subterranean clover. Plant and Soil **33**: 685–698.

Chapman, E.J., and Carrington, J.C. (2007). Specialization and evolution of endogenous small RNA pathways. Nat Rev Genet **8**: 884–896.

Chapman, E.J., Prokhnevsky, A.I., Gopinath, K., Dolja, V.V., and Carrington, J.C. (2004). Viral RNA silencing suppressors inhibit the microRNA pathway at an intermediate step. Genes Dev **18**: 1179–1186.

Charpentier, M. et al. (2008). Lotus japonicus CASTOR and POLLUX Are Ion Channels Essential for Perinuclear Calcium Spiking in Legume Root Endosymbiosis. Plant Cell **20**(12):3467-79

Chen, X. (2004). A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. Science **303**: 2022–2025.

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

Colebatch, G. et al. (2004). Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in Lotus japonicus. Plant J **39**: 487–512.

Combier, J.P. et al. (2006). MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in Medicago truncatula. Genes Dev **20**: 3084–3088.

Constabel, C.P., Yip, L., Patton, J.J., and Christopher, M.E. (2000). Polyphenol oxidase from hybrid poplar. Cloning and expression in response to wounding and herbivory. Plant Physiol **124**: 285–295.

Crespi, M., and Frugier, F. (2008). De novo organ formation from differentiated cells: root nodule organogenesis. Sci Signal **1(49)**: re11.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K., and Scheible, W.R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. Plant Physiol **139**: 5–17.

de Felippes, F.F., Schneeberger, K., Dezulian, T., Huson, D.H., and Weigel, D. (2008). Evolution of Arabidopsis thaliana microRNAs from random sequences. RNA **14**: 2455–2459.

Deleris, A. et al. (2006). Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. Science **313**: 68–71.

Delgado, M.J., Bedmar, E.J., and Downie, J.A. (1998). Genes involved in the formation and assembly of rhizobial cytochromes and their role in symbiotic nitrogen fixation. Adv Microb Physiol **40**: 191–231.

Den Herder, G., and Parniske, M. (2009). The unbearable naivety of legumes in symbiosis. Curr Opin Plant Biol **12**: 491–499.

Dezulian, T., Remmert, M., Palatnik, J.F., Weigel, D., and Huson, D.H. (2006). Identification of plant microRNA homologs. Bioinformatics **22**: 359–360.

Ding, Y., and Oldroyd, G.E. (2009). Positioning the nodule, the hormone dictum. Plant Signal Behav **4**: 89–93.

Ding, Y.F., and Zhu, C. (2009). The role of microRNAs in copper and cadmium homeostasis. Biochem Biophys Res Commun **386**: 6–10.

Dolgosheina, E.V. et al. (2008). Conifers have a unique small RNA silencing signature. RNA **14**: 1508–1515.

Dunoyer, P., Himber, C., Ruiz-Ferrer, V., Alioua, A., and Voinnet, O. (2007). Intra- and intercellular RNA interference in Arabidopsis thaliana requires components of the microRNA and heterochromatic silencing pathways. Nat Genet **39**: 848–856.

Dunoyer, P., Lecellier, C.H., Parizotto, E.A., Himber, C., and Voinnet, O. (2004). Probing the microRNA and small interfering RNA pathways with virus-encoded suppressors of RNA silencing. Plant Cell **16**: 1235–1250.

El Yahyaoui, F. et al. (2004). Expression profiling in Medicago truncatula identifies more than 750 genes differentially expressed during nodulation, including many potential regulators of the symbiotic program. Plant Physiol **136**: 3159–3176.

Fahlgren, N. et al. (2007). High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. PLoS ONE **2**: e219.

Fahraeus, G. (1957). The infection of clover root hairs by nodule bacteria studied by a simple glass slide technique. J Gen Microbiol **16**: 374–381.

Fedorova, M. et al. (2002). Genome-wide identification of nodule-specific transcripts in the model legume Medicago truncatula. Plant Physiol **130**: 519–537.

Franco-Zorrilla, J.M. et al. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. Nat Genet **39**: 1033–1037.

Frugier, F., Kosuta, S., Murray, J.D., Crespi, M., and Szczyglowski, K. (2008). Cytokinin: secret agent of symbiosis. Trends Plant Sci **13**: 115–120.

Garcia, D. (2008). A miRacle in plant development: role of microRNAs in cell differentiation and patterning. Semin Cell Dev Biol **19**: 586–595.

Geddy, R., and Brown, G.G. (2007). Genes encoding pentatricopeptide repeat (PPR) proteins are not conserved in location in plant genomes and may be subject to diversifying selection. BMC Genomics **8**: 130.

German, M.A. et al. (2008). Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. Nat Biotechnol **26**: 941–946.

Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. Nat Rev Genet **10**: 94–108.

Gleason, C. et al. (2006). Nodulation independent of rhizobia induced by a calcium-activated kinase lacking autoinhibition. Nature **441**: 1149–1152.

Gregory, B.D. et al. (2008). A link between RNA metabolism and silencing affecting Arabidopsis development. Dev Cell **14**: 854–866.

Gross, S.S., and Brent, M.R. (2006). Using multiple alignments to improve gene prediction. J Comput Biol **13**: 379–393.

Grunewald, W. et al. (2009). Manipulation of Auxin Transport in Plant Roots during Rhizobium Symbiosis and Nematode Parasitism. Plant Cell **182**(1): 200-12.

Guether, M. et al. (2009). Genome-wide reprogramming of regulatory networks, transport, cell wall and membrane biogenesis during arbuscular mycorrhizal symbiosis in Lotus japonicus. New Phytol **182**: 200–212.

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol **52**: 696–704.

Hakoyama, T. et al. (2009). Host plant genome overcomes the lack of a bacterial gene for symbiotic nitrogen fixation. Nature **462**: 514–517.

Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science **286**: 950–952.

Handberg, K., and Stougaard, J. (1992). Lotus japonicus, an autogamous, diploid legume species for classical and molecular-genetics. Plant Journal **2**: 487–496.

Hansch, R., and Mendel, R.R. (2009). Physiological functions of mineral micronutrients (Cu, Zn, Mn, Fe, Ni, Mo, B, Cl). Curr Opin Plant Biol **12**: 259–266.

Hartmann, K., Peiter, E., Koch, K., Schubert, S., and Schreiber, L. (2002). Chemical composition and ultrastructure of broad bean (Vicia faba L.) nodule endodermis in comparison to the root endodermis. Planta **215**: 14–25.

He, X.F., Fang, Y.Y., Feng, L., and Guo, H.S. (2008). Characterization of conserved and novel microRNAs and their targets, including a TuMV-induced

TIR-NBS-LRR class R gene-derived novel miRNA in Brassica. FEBS Lett **582**: 2445–2452.

Heckmann, A.B. et al. (2006). Lotus japonicus nodulation requires two GRAS domain regulators, one of which is functionally conserved in a non-legume. Plant Physiol **142**: 1739–1750.

Hewezi, T., Howe, P., Maier, T.R., and Baum, T.J. (2008). Arabidopsis small RNAs and their targets during cyst nematode parasitism. Mol Plant Microbe Interact **21**: 1622–1634.

Higo, K., Ugawa, Y., Iwamoto, M., and Higo, H. (1998). PLACE: a database of plant cis-acting regulatory DNA elements. Nucleic Acids Res **26**: 358–359.

Hirsch, A.M. (1992). Developmental biology of legume nodulation. New Phytologist **122**: 211–211.

Hirsch, S. et al. (2009). GRAS Proteins Form a DNA Binding Complex to Induce Gene Expression during Nodulation Signaling in Medicago truncatula. Plant Cell **21**(2):545-57

Hogslund, N. et al. (2009). Dissection of symbiosis and organ development by integrated transcriptome analysis of lotus japonicus mutant and wild-type plants. PLoS One **4**: e6556.

Horton, R., Moran, L.A., Scrimgeour, G., Perry, M., and Rawn, D. (2006). Principles of Biochemistry (4th Edition). (Prentice Hall, New Jersey).

Hossain, M.S., Umehara, Y., and Kouchi, H. (2006). A novel fix- symbiotic mutant of Lotus japonicus, Ljsym105, shows impaired development and premature deterioration of nodule infected cells and symbiosomes. Mol Plant Microbe Interact **19**: 780–788.

Huang, Y. et al. (2009). Structural insights into mechanisms of the small RNA methyltransferase HEN1. Nature **461**: 823–827.

Jagadeeswaran, G. et al. (2009). Cloning and characterization of small RNAs from Medicago truncatula reveals four novel legume-specific microRNA families. New Phytol

John, B. et al. (2004). Human MicroRNA targets. PLoS Biol **2**: e363.

Jones-Rhoades, M.W., and Bartel, D.P. (2004). Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol Cell **14**: 787–799.

Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. (2006). MicroRNAS and their regulatory roles in plants. Annu Rev Plant Biol **57**: 19–53.

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

Kalo, P. et al. (2005). Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. Science **308**: 1786–1789.

Kanamori, N. et al. (2006). A nucleoporin is required for induction of Ca2+ spiking in legume nodule development and essential for rhizobial and fungal symbiosis. Proc Natl Acad Sci U S A **103**: 359–364.

Kapoor, M. et al. (2008). Genome-wide identification, organization and phylogenetic analysis of Dicer-like, Argonaute and RNA-dependent RNA Polymerase gene families and their expression analysis during reproductive development and stress in rice. BMC Genomics **9**: 451.

Karimi, M., Inze, D., and Depicker, A. (2002). GATEWAY vectors for Agrobacterium-mediated plant transformation. Trends Plant Sci **7**: 193–195.

Kasschau, K.D., and Carrington, J.C. (1998). A counterdefensive strategy of plant viruses: suppression of posttranscriptional gene silencing. Cell **95**: 461–470.

Katiyar-Agarwal, S., Gao, S., Vivian-Smith, A., and Jin, H. (2007). A novel class of bacteria-induced small RNAs in Arabidopsis. Genes Dev **21**: 3123–3134.

Katiyar-Agarwal, S. et al. (2006). A pathogen-inducible endogenous siRNA in plant immunity. Proc Natl Acad Sci U S A **103**: 18002–18007.

Kawaguchi, M. et al. (2002). Root, root hair, and symbiotic mutants of the model legume Lotus japonicus. Mol Plant Microbe Interact **15**: 17–26.

Kawaguchi, M., Motomura, T., Imaizumi-Anraku, H., Akao, S., and Kawasaki, S. (2001). Providing the basis for genomics in Lotus japonicus: the

accessions Miyakojima and Gifu are appropriate crossing partners for genetic analyses. Mol Genet Genomics **266**: 157–166.

Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., and Segal, E. (2007). The role of site accessibility in microRNA target recognition. Nat Genet **39**: 1278–1284.

Kidner, C.A., and Martienssen, R.A. (2004). Spatially restricted microRNA directs leaf polarity through ARGONAUTE1. Nature **428**: 81–84.

Kidner, C.A., and Martienssen, R.A. (2005). The developmental role of microRNA in plants. Curr Opin Plant Biol **8**: 38–44.

Kouchi, H. et al. (2004). Large-scale analysis of gene expression profiles during early stages of root nodule formation in a model legume, Lotus japonicus. DNA Res **11**: 263–274.

Krusell, L. et al. (2005). The sulfate transporter SST1 is crucial for symbiotic nitrogen fixation in Lotus japonicus root nodules. Plant Cell **17**: 1625–1636.

Kumagai, H. et al. (2007). A novel ankyrin-repeat membrane protein, IGN1, is required for persistence of nitrogen-fixing symbiosis in root nodules of Lotus japonicus. Plant Physiol **143**: 1293–1305.

Kurihara, Y., Takashi, Y., and Watanabe, Y. (2006). The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. RNA **12**: 206–212.

Laubinger, S. et al. (2008). Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in Arabidopsis thaliana. Proc Natl Acad Sci U S A **105**: 8795–8800.

Lelandais-Brière, C. et al. (2009). Genome-Wide Medicago truncatula Small RNA Analysis Revealed Novel MicroRNAs and Isoforms Differentially Regulated in Roots and Nodules. Plant Cell **21**(9):2780-96

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

Lin, S.I. et al. (2008). Regulatory network of microRNA399 and PHO2 by systemic signaling. Plant Physiol **147**: 732–746.

Lingel, A., Simon, B., Izaurralde, E., and Sattler, M. (2003). Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. Nature **426**: 465–469.

Liu, J., He, Y., Amasino, R., and Chen, X. (2004). siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in Arabidopsis. Genes Dev **18**: 2873–2878.

Liu, Q. et al. (2009). Expression analysis of phytohormone-regulated microRNAs in rice, implying their regulation roles in plant hormone signaling. FEBS Lett **583:** 723:728

Lombari, P., Ercolano, E., El Alaoui, H., and Chiurazzi, M. (2003). A new transformation-regeneration procedure in the model legume Lotus japonicus: root explants as a source of large numbers of cells susceptible to Agrobacterium-mediated transformation. Plant Cell Rep **21**: 771–777.

Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y.O., and Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res **33**: 6494–6506.

Lu, C. et al. (2005). Elucidation of the small RNA component of the transcriptome. Science **309**: 1567–1569.

Lu, S., Sun, Y.H., and Chiang, V.L. (2008). Stress-responsive microRNAs in Populus. Plant J **55**: 131–151.

Madsen, E.B. et al. (2003). A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. Nature **425**: 637–640.

Maekawa, T. et al. (2008). Polyubiquitin promoter-based binary vectors for overexpression and gene silencing in Lotus japonicus. Mol Plant Microbe Interact **21**: 375–382.

Mallory, A.C. et al. (2009). Redundant and specific roles of the ARGONAUTE proteins AGO1 and ZLL in development and small RNA-directed gene silencing. PLoS Genet **5**: e1000646.

Mallory, A.C., and Vaucheret, H. (2009). ARGONAUTE 1 homeostasis invokes the coordinate action of the microRNA and siRNA pathways. EMBO Rep **10**: 521–526.

Margis, R. et al. (2006). The evolution and diversification of Dicers in plants. FEBS Lett **580**: 2442–2450.

Markmann, K., and Parniske, M. (2009). Evolution of root endosymbiosis with bacteria: how novel are nodules? Trends Plant Sci **14**: 77–86.

Meyers, B.C. et al. (2008). Criteria for Annotation of Plant MicroRNAs. Plant Cell **20**(12):3186-90

Mi, S. et al. (2008). Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. Cell **133**: 116–127.

Middleton, P.H. et al. (2007). An ERF transcription factor in Medicago truncatula that is essential for Nod factor signal transduction. Plant Cell **19**: 1221–1234.

Minchin, F.R., James, E.K., and Becana, M. (2008). Oxygen diffusion, production of reactive oxygen and nitrogen species, and antioxidants in legume nodules. Nitrogen Fixation-Origins Applications and Research Progress, Nitrogen-fixing leguminous symbioses.(eds) Dilworth, MJ James, EK Sprent, JI Newton, WE Springer, PO Box **17**: 321–362.

Mitra, R.M., Shaw, S.L., and Long, S.R. (2004). Six nonnodulating plant mutants defective for Nod factor-induced transcriptional changes associated with the legume-rhizobia symbiosis. Proc Natl Acad Sci U S A **101**: 10217–10222.

Montgomery, T.A. et al. (2008a). Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. Cell **133**: 128–141.

Montgomery, T.A. et al. (2008b). AGO1-miR173 complex initiates phased siRNA formation in plants. Proc Natl Acad Sci U S A **105**: 20055–20062.

Moxon, S. et al. (2008). A toolkit for analysing large-scale plant small RNA datasets. Bioinformatics **24**: 2252–2253.

Murakami, Y. et al. (2006). Positional cloning identifies Lotus japonicus NSP2, a putative transcription factor of the GRAS family, required for NIN and ENOD40 gene expression in nodule initiation. DNA Res **13**: 255–265.

Navarro, L. et al. (2006). A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. Science **312**: 436–439.

Nikovics, K. et al. (2006). The balance between the MIR164A and CUC2 genes controls leaf margin serration in Arabidopsis. Plant Cell **18**: 2929–2945.

Nonomura, K. et al. (2007). A germ cell specific gene of the ARGONAUTE family is essential for the progression of premeiotic mitosis and meiosis during sporogenesis in rice. Plant Cell **19**: 2583–2594.

O'hara, G.W., Boonkerd, N., and Dilworth, M.J. (1988). Mineral constraints to nitrogen fixation. Plant and Soil **108**: 93-110

Obbard, D.J., Jiggins, F.M., Halligan, D.L., and Little, T.J. (2006). Natural selection drives extremely rapid evolution in antiviral RNAi genes. Curr Biol **16**: 580–585.

Oldroyd, G.E., and Downie, J.A. (2008). Coordinating nodule morphogenesis with rhizobial infection in legumes. Annu Rev Plant Biol **59**: 519–546.

Ori, N. et al. (2007). Regulation of LANCEOLATE by miR319 is required for compound-leaf development in tomato. Nat Genet **39**: 787–791.

Ott, T. et al. (2009). Absence of symbiotic leghemoglobins alters bacteroid and plant cell differentiation during development of Lotus japonicus root nodules. Mol Plant Microbe Interact **22**: 800–808.

Ott, T. et al. (2005). Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development. Curr Biol **15**: 531–535.

Pall, G.S., and Hamilton, A.J. (2008). Improved northern blot method for enhanced detection of small RNA. Nat Protoc **3**: 1077–1084.

Palmer, C.M., and Guerinot, M.L. (2009). Facing the challenges of Cu, Fe and Zn homeostasis in plants. Nat Chem Biol **5**: 333–340.

Pant, B.D., Buhtz, A., Kehr, J., and Scheible, W.R. (2008). MicroRNA399 is a long-distance signal for the regulation of plant phosphate homeostasis. Plant J **53**: 731–738.

Park, M.Y., Wu, G., Gonzalez-Sulser, A., Vaucheret, H., and Poethig, R.S. (2005). Nuclear processing and export of microRNAs in Arabidopsis. Proc Natl Acad Sci U S A **102**: 3691–3696.

Parniske, M. (2008). Arbuscular mycorrhiza: the mother of plant root endosymbioses. Nat Rev Microbiol **6**: 763–775.

Perry, J.A. et al. (2003). A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume Lotus japonicus. Plant Physiol **131**: 866–871.

Pfeffer, S. (2007). Identification of virally encoded microRNAs. Methods Enzymol **427**: 51–63.

Pilon, M., Cohu, C.M., Ravet, K., Abdel-Ghany, S.E., and Gaymard, F. (2009). Essential transition metal homeostasis in plants. Curr Opin Plant Biol **12**: 347–357.

Piriyapongsa, J., and Jordan, I.K. (2008). Dual coding of siRNAs and miRNAs by plant transposable elements. RNA **14**: 814–821.

Preisig, O., Zufferey, R., Thony-Meyer, L., Appleby, C.A., and Hennecke, H. (1996). A high-affinity cbb3-type cytochrome oxidase terminates the symbiosis-specific respiratory chain of Bradyrhizobium japonicum. J Bacteriol **178**: 1532–1538.

Puppo, A. et al. (2005). Legume nodule senescence: roles for redox and hormone signalling in the orchestration of the natural aging process. New Phytol **165**: 683–701.

Qi, X., Bao, F.S., and Xie, Z. (2009). Small RNA deep sequencing reveals role for Arabidopsis thaliana RNA-dependent RNA polymerases in viral siRNA biogenesis. PLoS One **4**: e4971.

Quinn, J.M., Barraco, P., Eriksson, M., and Merchant, S. (2000). Coordinate copper- and oxygen-responsive Cyc6 and Cpx1 expression in Chlamydomonas is mediated by the same element. J Biol Chem **275**: 6080–6089.

Radutoiu, S. et al. (2003). Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. Nature **425**: 585–592.

Radutoiu, S. et al. (2007). LysM domains mediate lipochitin-oligosaccharide recognition and Nfr genes extend the symbiotic host range. EMBO J **26**: 3923–3935.

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

Ramachandran, V., and Chen, X. (2008). Degradation of microRNAs by a family of exoribonucleases in Arabidopsis. Science **321**: 1490–1492.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. (2002). MicroRNAs in plants. Genes Dev **16**: 1616–1626.

Ruiz-Ferrer, V. and Voinnet, O. (2009). Roles of plant small RNAs in biotic stress responses. Annu Rev Plant Biol **60**: 485–510.

Saito, K. et al. (2007). NUCLEOPORIN85 is required for calcium spiking, fungal and bacterial symbioses, and seed production in Lotus japonicus. Plant Cell **19**: 610–624.

Sanchez, D.H. et al. (2008). Integrative functional genomics of salt acclimatization in the model legume Lotus japonicus. Plant J **53**: 973–987.

Sandal, N. et al. (2006). Genetics of symbiosis in Lotus japonicus: recombinant inbred lines, comparative genetic maps, and map position of 35 symbiotic loci. Mol Plant Microbe Interact **19**: 80–91.

Santos, R., Herouart, D., Sigaud, S., Touati, D., and Puppo, A. (2001). Oxidative burst in alfalfa-Sinorhizobium meliloti symbiotic interaction. Mol Plant Microbe Interact **14**: 86–89.

Sato, S. et al. (2008). Genome structure of the legume, Lotus japonicus. DNA Res **15**: 227–239.

Schauer, S.E., Jacobsen, S.E., Meinke, D.W., and Ray, A. (2002). DICER-LIKE1: blind men and elephants in Arabidopsis development. Trends Plant Sci **7**: 487–491.

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

Scott, DB, Chua, K, Jarvis, BDW, and Pankhurst, C.E. (1995). Molecular cloning of a nodulation gene from fast- and slow-growing strains of Lotus rhizobia. Mol Gen Genet **201**: 43–50.

Seliga, H. (1998). Nitrogen fixation in several grain legume species with contrasting sensitivities to copper nutrition. Acta Physiologiae Plantarum **20**: 263–267.

Shen, D., et al. (2009). Molecular phylogeny of miR390-guided trans-acting siRNA genes(TAS3) in the grass family. Plant Systematics and Evolution **28**: 125–132.

Silhavy, D. et al. (2002). A viral protein suppresses RNA silencing and binds silencing-generated, 21- to 25-nucleotide double-stranded RNAs. EMBO J **21**: 3070–3080.

Simon, S.A., Meyers, B.C., and Sherrier, D.J. (2009). MicroRNAs in the Rhizobia-Legume Symbiosis. Plant Physiol **151**(3):1002-8

Soltis, P.S., Soltis, D.E., and Chase, M.W. (1999). Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. Nature **402**: 402–404.

Stracke, S. et al. (2002). A plant receptor-like kinase required for both bacterial and fungal symbiosis. Nature **417**: 959–962.

Subramanian, S. et al. (2008). Novel and nodulation-regulated microRNAs in soybean roots. BMC Genomics **9**: 160.

Suganuma, N. et al. (2003). The Lotus japonicus Sen1 gene controls rhizobial differentiation into nitrogen-fixing bacteroids in nodules. Mol Genet Genomics **269**: 312–320.

Sunkar, R., Chinnusamy, V., Zhu, J., and Zhu, J.K. (2007). Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. Trends Plant Sci **12**: 301–309.

Sunkar, R., and Jagadeeswaran, G. (2008). In silico identification of conserved microRNAs in large number of diverse plant species. BMC Plant Biol **8**: 37.

Szczyglowski, K. et al. (1998). Nodule organogenesis and symbiotic mutants of the model legume Lotus japonicus. Molecular Plant-Microbe Interactions **11**: 684–697.

Szittya, G. et al. (2008). High-throughput sequencing of Medicago truncatula short RNAs identifies eight new miRNA families. BMC Genomics **9**: 593.

Tagami, Y., Motose, H., and Watanabe, Y. (2009). A dominant mutation in DCL1 suppresses the hyl1 mutant phenotype by promoting the processing of miRNA. RNA **15**: 450–458.

Takeda, A., Iwasaki, S., Watanabe, T., Utsumi, M., and Watanabe, Y. (2008). The mechanism selecting the guide strand from small RNA duplexes is different among argonaute proteins. Plant Cell Physiol **49**: 493–500.

Tirichine, L., Herrera-Cervera, J.A., and Stougaard, J. (2005). in *Lotus japonicus* Handbook. (Springer, Dordrecht, The Netherlands).

Tirichine, L. et al. (2006a). Deregulation of a Ca2+/calmodulin-dependent kinase leads to spontaneous nodule development. Nature **441**: 1153–1156.

Tirichine, L., James, E.K., Sandal, N., and Stougaard, J. (2006b). Spontaneous root-nodule formation in the model legume Lotus japonicus: a novel class of mutants nodulates in the absence of rhizobia. Mol Plant Microbe Interact **19**: 373–382.

Tirichine, L. et al. (2007). A gain-of-function mutation in a cytokinin receptor triggers spontaneous root nodule organogenesis. Science **315**: 104–107.

Tolia, N.H., and Joshua-Tor, L. (2007). Slicer and the argonautes. Nat Chem Biol **3**: 36–43.

Uchiumi, T. et al. (2004). Expression islands clustered on the symbiosis island of the Mesorhizobium loti genome. J Bacteriol **186**: 2439–2448.

Udvardi, M.K., Tabata, S., Parniske, M., and Stougaard, J. (2005). Lotus japonicus: legume research in the fast lane. Trends Plant Sci **10**: 222–228.

Vargason, J.M., Szittya, G., Burgyan, J., and Hall, T.M. (2003). Size selective recognition of siRNA by an RNA silencing suppressor. Cell **115**: 799–811.

Vasse, J., de Billy, F., Camut, S., and Truchet, G. (1990). Correlation between ultrastructural differentiation of bacteroids and nitrogen fixation in alfalfa nodules. J Bacteriol **172**: 4295–4306.

Vaucheret, H. (2008). Plant ARGONAUTES. Trends Plant Sci **13**: 350–358.

Vazquez, F. (2006). Arabidopsis endogenous small RNAs: highways and byways. Trends Plant Sci **11**: 460–468.

Vernie, T. et al. (2008). EFD Is an ERF transcription factor involved in the control of nodule number and differentiation in Medicago truncatula. Plant Cell **20**: 2696–2713.

Voinnet, O. (2001). RNA silencing as a plant immune system against viruses. Trends Genet **17**: 449–459.

Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. Cell **136**: 669–687.

Voinnet, O., Pinto, Y.M., and Baulcombe, D.C. (1999). Suppression of gene silencing: a general strategy used by diverse DNA and RNA viruses of plants. Proc Natl Acad Sci U S A **96**: 14147–14152.

Wan, C.Y., and Wilkins, T.A. (1994). A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (Gossypium hirsutum L.). Anal Biochem **223**: 7–12.

Wang, X.J., Gaasterland, T., and Chua, N.H. (2005). Genome-wide prediction and identification of cis-natural antisense transcripts in Arabidopsis thaliana. Genome Biol **6**: R30.

Wang, Y. et al. (2009). Identification and expression analysis of miRNAs from nitrogen-fixing soybean nodules. Biochem Biophys Res Commun **378**: 799–803.

White, J., Prell, J., James, E.K., and Poole, P. (2007). Nutrient sharing between symbionts. Plant Physiol **144**: 604–614.

Yamasaki, H., Hayashi, M., Fukazawa, M., Kobayashi, Y., and Shikanai, T. (2009). SQUAMOSA Promoter Binding Protein-Like7 Is a Central Regulator for Copper Homeostasis in Arabidopsis. Plant Cell **21**(1):347-61

Yan, J. et al. (2009). The REDUCED LEAFLET genes encode key components of the trans-acting siRNA pathway and regulate compound leaf and flower development in Lotus japonicus. Plant Physiol (online version ahead of printing)

Yan, K.S. et al. (2003). Structure and conserved RNA binding of the PAZ domain. Nature **426**: 468–474.

Yang, L., Liu, Z., Lu, F., Dong, A., and Huang, H. (2006). SERRATE is a novel nuclear regulator in primary microRNA processing in Arabidopsis. Plant J **47**: 841–850.

Yano, K. et al. (2008). CYCLOPS, a mediator of symbiotic intracellular accommodation. Proc Natl Acad Sci U S A **105**: 20540–20545.

Ye, K., Malinina, L., and Patel, D.J. (2003). Recognition of small interfering RNA by a viral suppressor of RNA silencing. Nature **426**: 874–878.

Yokota, K. et al. (2009). Rearrangement of Actin Cytoskeleton Mediates Invasion of Lotus japonicus Roots by Mesorhizobium loti. Plant Cell **21**(1):267-84

Yoon, E.K. et al. (2009). Auxin regulation of the microRNA390-dependent transacting small interfering RNA pathway in Arabidopsis lateral root development. Nucleic Acids Res (Online version ahead of printing)

Yoshikawa, M., Peragine, A., Park, M.Y., and Poethig, R.S. (2005). A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. Genes Dev **19**: 2164–2175.

Yu, B. et al. (2008). The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. Proc Natl Acad Sci U S A **105**: 10073–10078.

Yu, B. et al. (2005). Methylation as a crucial step in plant microRNA biogenesis. Science **307**: 932–935.

Zhang, X. et al. (2006). Cucumber mosaic virus-encoded 2b suppressor inhibits Arabidopsis Argonaute1 cleavage activity to counter plant defense. Genes Dev **20**: 3255–3268.

Zhang, Y. (2005). miRU: an automated plant miRNA target prediction server. Nucleic Acids Res **33**: W701–4.

Zheng, X., Zhu, J., Kapoor, A., and Zhu, J.K. (2007). Role of Arabidopsis AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. EMBO J **26**: 1691–1701.

Zilberman, D., Cao, X., and Jacobsen, S.E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science **299**: 716–719.

Zong, J., Yao, X., Yin, J., Zhang, D., and Ma, H. (2009). Evolution of the RNA-dependent RNA polymerase (RdRP) genes: duplications and possible losses before and after the divergence of major eukaryotic groups. Gene **447**: 29–39.