The influence of the plant circadian clock on nodulation in *Medicago truncatula*

Proyash Roy

Thesis

Submitted to the University of Warwick
for the degree of

Doctor of Philosophy

School of Life Sciences
Supervisor: Miriam Gifford

April 2021

Table of Contents

List of abbreviation

AMPs – Anti-Microbial Peptides
ARF – AUXIN RESPONSE FACTOR
CaML – CALMODULIN-LIKE
CCA1 – CIRCADIAN CLOCK ASSOCIATED 1
CCaMK – CALCIUM-CALMODULIN DEPENDENT PROTEIN KINASE
DE – Differential Expression
DEFL – DEFensin-Like
DMI – DOESN'T MAKE INFECTION
DNA – Deoxyribose Nucleic Acid
DNF – DEFECTIVE IN NITROGEN FIXATION
EC – Evening Complex
EE – Evening Element
ELF – EARLY FLOWERING
EPS – ExoPolySaccharide
ERF – ETHYLENE RESPONSE FACTOR
ERN – ERF REQUIRED FOR NODULATION
FACS – Fluorescence Activated Cell Sorting
Fix- – Not Nitrogen fixing
Fix+ – Nitrogen fixing
FT – FLOWERING LOCUS T
GFP – Green Fluorescent Protein
GI – GIGANTEA
GO – Gene Ontology
GRPs – Glycine Rich Peptides
Grx – Glutaredoxin
GUS – β-GLUCURONIDASE
HrrP – HOST RANGE RESTRICTION PEPTIDASE
HrrP+ – HrrP expressing
IPD3 – INTERACTING PROTEIN OF DMI3
IRLC – Inverted Repeat Lacking Clade
KEGG – Kyoto Encyclopedia of Genes and Genomes
LB – Luria-Bertani
LHY – LATE-ELONGATED HYPOCOTYLE
LNK – NIGHT LIGHT-INDUCIBLE AND CLOCK REGULATED
LPS – LipoPolySaccharide
LS – Lomb Scargle
LUX – LUX ARRHYTHMO
LYK3 – LYSM DOMAIN-CONTAINING RECEPTOR-LIKE KINASE

MAMPs – Microbe Associated Molecular Patterns
MFM – Modified Fahreus Media
MsrA – METHIONINE SULFOXIDE REDUCTASE A
N – Nitrogen
NCR – Nodule-specific Cysteine-Rich
NFP – Nodulation Factor Perception
NFR – Nodulation Factor Receptor
NFS – Nitrogen Fixation Specificity
NIN – NODULE INCEPTION
Nod – Nodulation
NORK – NODULATION RECEPTOR KINASE
NSP – Nodulation Signaling Pathway
NSS – N-terminal Signal Sequence
PCR – Polymerase Chain Reaction
PD – Plasmodesmata
PDCBs – Plasmodesmata Callose-Binding proteins
PDLPs – Plasmodesmata-Located Proteins
p*I* – Isoelectric point
PRR – PSEUDO-RESPONSE REGULATOR
qPCR – quantitative Polymerase Chain Reaction
RIN – RNA Integrity Number
RLM – Rhythmic Leaf Movement
RNA – RiboNucleic Acid
ROS – Reactive Oxygen Species
RVE – ReVEille
SPC – Signal Peptidase Complex
SYMRK – SYMBIOSIS RECEPTOR KINASE
TBD – Terminal Bacteroid Differentiation
TF – Transcription Factor
TOC1 – TIMING OF CAB EXPRESSION 1
TPM – Transcripts Per Million
Trx – Thioredoxin
TY – Tryptone Yeast
UTR – Un-Translated Region

List of figures and tables

Acknowledgements

First and foremost, my heartfelt gratitude to my wonderful supervisor, Dr Miriam Gifford, not only because of her support and guidance during my PhD but also for her boundless encouragement during my PhD application period. Even after some failed attempts, she kept me inspired that finally enabled me to achieve the Commonwealth Scholarship to pursue a PhD degree in the UK. Throughout my PhD, whenever I struggled with my research and felt very low, she came up with an encouraging solution, kept me going, and at the same time took care of my mental and academic wellbeing. I can easily write a couple of pages of eulogy admiring her but to put it in one sentence, I have never dreamt of a better supervisor than Miriam in any related aspect.

I would like to especially thank two persons, Dr Beatriz Lagunas and Dr Liam Walker, who guided me in my early days at the lab; patiently responded to my many silly queries and helped me to adapt to a new research environment of a new country. I am also enormously grateful to Dr Emeka Okechukwu for those numerous tips and insightful discussions that eased many aspects of my research and social life. I wish to thank all the past and present Gifford lab members, but especially, Joseph Blackwell, Jack Simcox, Nadine Aschauer, Beth Richmond for directly helping me in some of my experiments; Cantug Bar, Helen Wilkinson for many thoughtful chats and help; Monique Rowson for her contagious exuberance (!) and finally, Dr Mingkee Achom whose PhD work led the foundation for my research.

I would like to record my thanks to my advisory panel members, Dr Graham Teakle and Dr Charlotte Allender for their careful reviews and valuable feedbacks throughout my project. I am thankful to Dr Laura Baxter, Dr Emma Picot, Dr Alonso Pardal and Luke Richards for helping me with different bioinformatic analyses. I am very grateful to Gary Grant who was always very helpful, even when I made any last-minute requests, that made my life easier in setting up different experiments at the Phytobiology facility. I will also keep my friend, Saadlee Shehreen, in this thanksgiving; not that she helped me a lot in my PhD, but without her generous help during my undergraduate I might have not been in a position to start a PhD!

I like to try but it is impossible to thank someone enough who left her country, parents, job, and all other belongings to accompany me in a new country with different culture and language. My loving wife, Mitaly Sikder, I acknowledge and thank you to the fullest sense of these words for your persistence and support; and be assured that you share a substantial amount of credit for whatever I have already (or going to) achieved in life.

Finally, but most importantly, I will thank and dedicate this work to my parents, Seema Roy and Pabitra Kumar Roy. It is my mother whom I believe is responsible for installing a logical mind inside me in my childhood, and my father taught me how to be circumspective and practical. My inherent reticence never allowed me to properly express my gratitude towards them, but deep-down in my heart, I believe they understand how I feel about them, and they must be very proud of my achievements!

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by me and has not been submitted in any previous application for any degree.

All the research and analyses were carried out by the author; however, the Rhythmic Leaf Movement (RLM) assay and the qPCR analysis (presented in Section 4.2.3 and 4.2.4) were done in conjunction with two master's student, Jack Simcox and Joseph Blackwell, respectively and had been used as part of their master's thesis.

A substantial part of 'Chapter 1: Introduction' has already been published as a review article (Roy P et al., 2020) in *Genes* (doi.org/10.3390/genes11040348). The findings presented in Chapter 3 and 4 have also been deposited in the form of a research manuscript in *BioRxiv* (doi.org/10.1101/2021.03.22.435813) that is also currently under review in a reputed journal.

Abstract

Leguminous plants (or legumes) are well-known for their symbiotic relationship with nitrogen (N) fixing soil bacteria (rhizobia). Through providing a niche to rhizobia in specialized root organelles called nodules and in exchange for plant sugars and nutrients, leguminous plants gain access to the nitrogen fixed by the bacteria. As a result, nodulating plant species can grow in N-depleted soil while most non-legume crop plants require N-fertilizers to reach significant yields. A family of peptides, called Nodule-specific Cysteine-Rich (NCR) peptides, play important roles in rhizobial differentiation and nodulation in certain legumes (e.g., *Medicago truncatula*) belonging to the Inverted-Repeat Lacking Clade (IRLC). Interestingly, a subset of these NCR gene promoters contain the Evening Element (EE), a *cis*-regulatory motif which acts as a binding site for a number of plant circadian transcription factors including LATE-ELONGATED HYPOCOTYL (LHY). In order to characterize a potential LHY-NCR-nodulation mechanism, a free-running time-course experiment was carried out where *M. truncatula* A17 nodules, roots and leaves were sampled over 48 hours. Gene expression analysis through RNA sequencing revealed around 5% of all *M. truncatula* genes were diurnally oscillating in nodules, including 45 NCRs genes. More than half of these NCRs were highly expressed in the evening period and a significant number of them contain EE motif in their promoters. This supports the hypothesis that these NCRs are negatively regulated by the morning expressed circadian transcription factor, LHY. Furthermore, we have shown that the loss of LHY function leads to reduced nodulation and plant growth. In line with a regulatory role for the clock in nodule metabolic activity, when we clustered different rhythmically expressed plant and rhizobial genes in nodules into four time-of-day groups, we found that modulation of metabolism occurs in waves. This elaborated our understanding of the function of known rhythmic processes (inc. photosynthesis, carbohydrate metabolism) and identified novel functional links. In parallel to the LHY-NCR gene expression study, two cortex-specific GFP expression vectors were prepared and analyzed that could be useful for cell-type specific gene expression in future since the cortex is a site of nodule development. Application of our data is likely to enable us to maximize symbiotic N-fixation and potentially help us to employ such features in non-legume crop plants, bringing considerable benefits.

1. Introduction

1.1. Legume-rhizobia interactions and biological nitrogen fixation

Nitrogen is one of the key limiting nutrients that affect plant growth and crop production (Vance, 2001). Most plants, except those in the Fabaceae/Leguminosae family and the actinorhizal species, are highly dependent on nitrogen (N) availability in the soil, typically supplied through fertilizers in an agricultural situation. Leguminous plants, however, employ a symbiotic strategy with nitrogen fixing soil bacteria (known as rhizobia; actinorhizal plants form similar symbiosis with *Frankia spp.*) that are housed in specialized root organs (nodules) (Figure 1.1). This enables them to grow in nitrogen-depleted soil without N-fertilizers. Most of the major crops used in agriculture (e.g. rice, wheat, corn) are non-legumes and non-actinorhizal, thus require N-fertilizers (in the form of ammonium nitrate, urea etc.). Application of N-fertilizers not only increases the cost of crop production but it also brings environmental issues due to the large amount of greenhouse gas emission

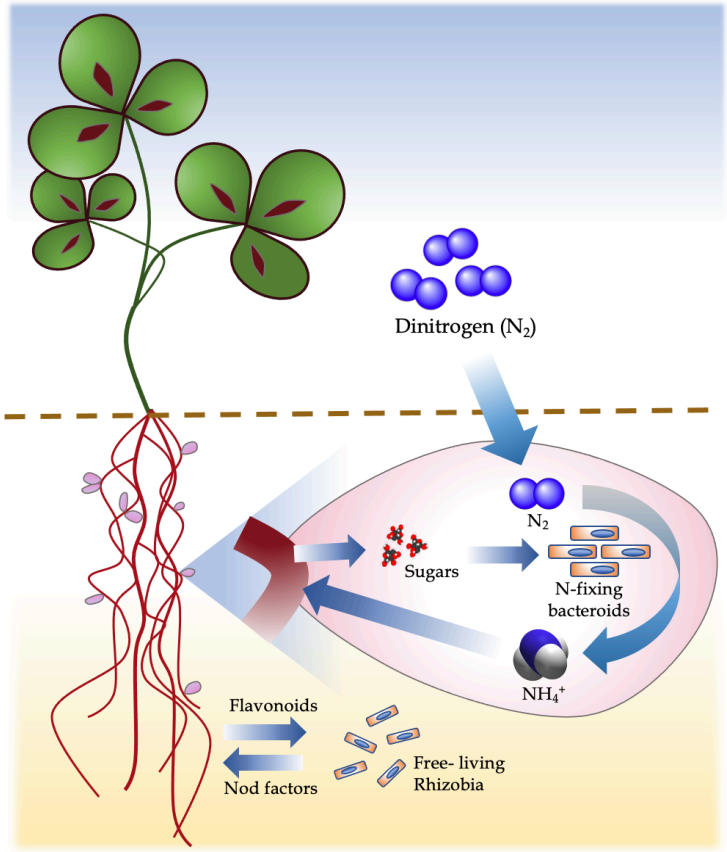


Figure 1.1: Overview of legume-rhizobia symbiosis. Free-living rhizobia in the soil can sense plant-derived flavonoids/isoflavonoids and in turn secret Nodulation (Nod) factors. Root hair cells recognize Nod factors and initiate a process that results in the formation of a root organ called nodules (pink) where rhizobia colonize. Within nodules, rhizobia go through a specific differentiation process and gain the ability to convert atmospheric gaseous nitrogen into a biologically accessible form (e.g., NH_4^+) that plant can utilise. In return, plant also provides shelter, sugars and other nutrients to the symbiont.

during their manufacture and use (Zhang W et al., 2013). Although nodulation is restricted to legumes and actinorhizal plants, some of the genes required for this process are conserved in many plant species (Mu and Luo, 2019). Thus, if we understand the molecular mechanism of nodulation we could transfer this process to non-legume crops using crop breeding or more advanced synthetic biology approaches, with considerable economic and environmental benefits (Vicente and Dean, 2017).

1.2. Nodule development requires specific molecular interactions and physiological changes in a spatial and temporal manner

In N-depleted soil, many legume plants release varieties of flavonoids and isoflavonoids through their roots into the soil. These serve as chemoattraction signals for the rhizobial species that, in response, move into close proximity of plant roots, and secrete signaling molecules known as Nod (Nodulation) factors (Long, 1996) (Figure 1.1 & 1.2a). Upon recognizing Nod factors through specific receptors (e.g., NOD FACTOR PERCEPTION (NFP) and LysM DOMAIN-CONTAINING RECEPTOR-LIKE KINASE (LYK3) in *Medicago truncatula*), legume plants trap rhizobia in a root hair curl (Figure 1.2b). The root hair curl is formed as a result of Nod factor-induced localized growth inhibition at root hair tips. This results in an asymmetric growth of the root hair, creating a pocket (hair curl) where rhizobia can be entrapped (Esseling et al., 2003). Rhizobia then enters the root hair cell through an opening generated, presumably, by a host mediated localized cell wall degradation (Xie et al., 2012). In the next step, an

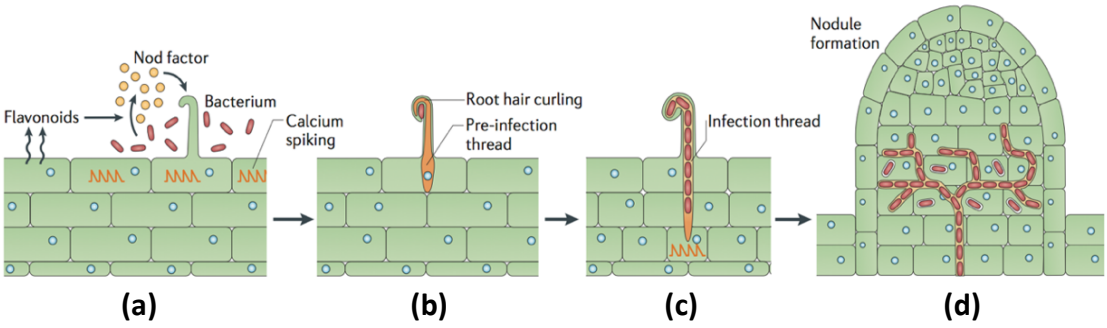


Figure 1.2: Steps in rhizobial colonization. (a) Rhizobia enter root tissue through the root hair but it is not yet clear whether rhizobia first attach to the root epidermis and then to the root hair or attach directly to the root hair (Poole et al., 2018). **(b)** Nod factors bind to specific receptors localised in root hairs, leading to asymmetric root hair growth which ultimately entraps the bacteria in a root hair curl. **(c)** An infection thread is initiated from the root hair curl that directs bacteria towards the cortex where a nodule primordia is forming. **(d)** The nodule primordia is ultimately colonized by rhizobia and develops into mature nodule. Figure adapted from Oldroyd (2013).

infection thread (Figure 1.2c) is formed in the root hair through cytoskeletal rearrangement where rhizobia actively divide and grow towards underlying cortical cells (Timmers et al., 1999, Gage, 2004). At the same time as when root hair curling and the infection thread are developing, the first two layers of underlying cortical cells start to divide and de-differentiate to give rise to a nodule primordia that is eventually infected and colonized by rhizobia from the infection thread (Figure 1.2d).

The structural differences between different rhizobial Nod factors and between host-specific receptors primarily determine host-symbiont compatibility and underpin development of species-specific interactions (Oldroyd, 2013). In *M. truncatula*, two receptors, NFP and LYK3, have been identified that interact with Nod factors secreted by compatible rhizobia (e.g., *Sinorhizobium meliloti* for *M. truncatula*) (Smit et al., 2007, Rey et al., 2013). The exact mechanism of how these Nod Factor Receptors (NFRs) work is yet to be fully understood but they are reported to co-localize and interact with some downstream signaling proteins (Moling et al., 2014) (Figure 1.3). One such protein is NORK (NODULATION RECEPTOR KINASE; also termed as DOESN'T MAKE INFECTION 2 (DMI2) that is homologous to SYMBIOSIS RECEPTOR KINASE (SYMRK) in *Lotus japonicus*) that is essential for regulation of nuclear Ca^{2+} oscillations upon Nod factor perception (Charpentier and Oldroyd, 2013). This Ca^{2+} oscillation is dependent on specific ion channels (e.g., DMI1 in *M. truncatula*) (Charpentier et al., 2008). The Ca^{2+} oscillation results in the phosphorylation of a Ca^{2+}/CALMODULIN DEPENDENT PROTEIN KINASE (CCaMK), known as DMI3 in *M. truncatula*, that in turn also phosphorylates IPD3 (INTERACTING PROTEIN OF DMI3) (Jin et al., 2018). DMI3 and IPD3 usually form a complex and when phosphorylated they induce the expression of a number of important nodulation related genes that includes *NIN* (*NODULE INCEPTION*), *ERN1* and *ERN2* (*ETHYLENE RESPONSE FACTOR REQUIRED FOR NODULATION* 1/2) etc. NIN is a transcription factor and is found to be highly conserved among nodulating plant species including the non-legume *Parasponia andersonii* (van Velzen et al., 2018). NIN plays an important role in initiating cytokinin signaling (discussed in Chapter 6) which is vital for nodule formation (Vernié et al., 2015). The ERN1/ERN2 complex also plays a vital role in the formation of the infection thread (Kawaharada et al., 2017). The induction of these genes together with other important known (e.g., *NODULATION SIGNALING PATHWAY 1* and *2* (*NSP1* and *2*)) and yet unknown genes enables epidermal cells to be infected by rhizobia

and then for subsequent formation of the infection thread and development of the nodule primordia in the cortex to occur (Figure 1.3).

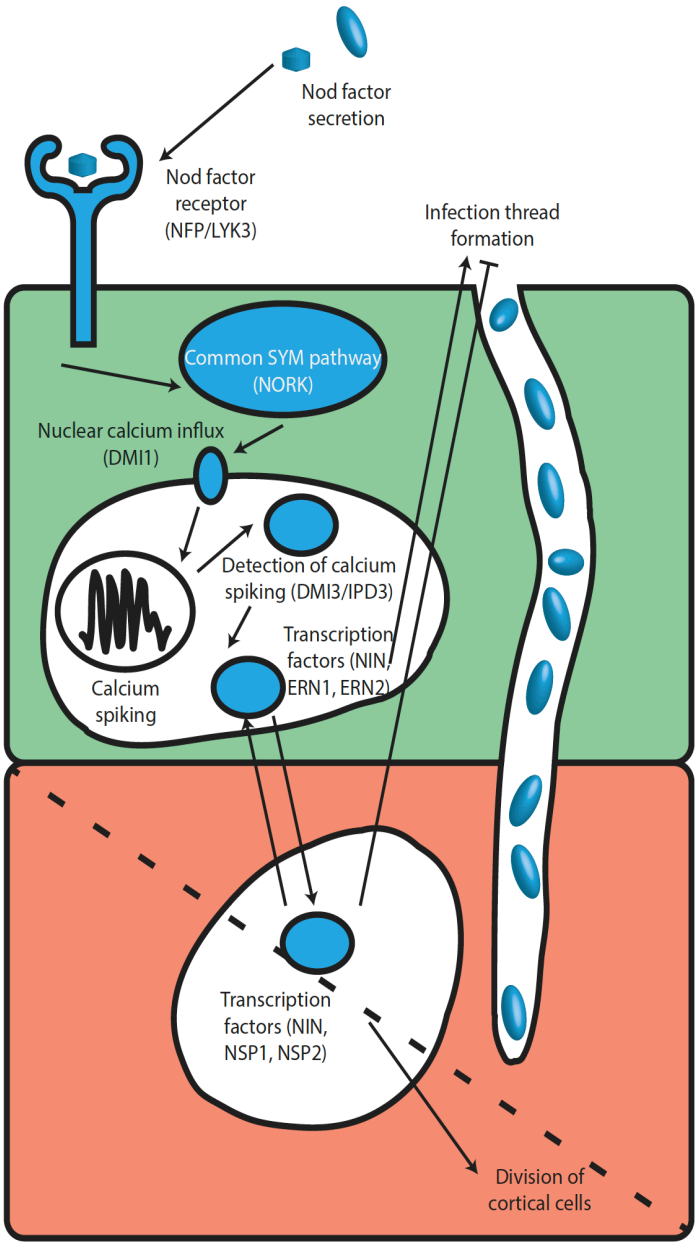


Figure 1.3: Simplified model of the molecular signaling in the nodulation pathway of *M. truncatula*. Nod factor perception by NFPs results in a cascade of phosphorylation reactions that results in a Ca^{2+} spiking in the nucleus. The oscillation in Ca^{2+} level leads to phosphorylation of the DMI3-IPD3 complex which then induces the expression of many nodulation related genes. The induction of these genes facilitates rhizobial infection and the formation of an infection thread. Concurrently, they induce specific cortical cells to divide and form a nodule primordium where other nodulation associated genes are expressed. These genes then act to inhibit multiple infection thread formation and function in nodule development. Green and orange colours represent epidermal and cortical cells, respectively. Figure adapted from Walker (2019)

In developing nodules, symbiotic host cells become enlarged as they undergo repeated endoreduplications (multiple rounds of genome duplication without cytokinesis) to accommodate thousands of rhizobial cells (Gonzalez-Sama et al., 2006, Kondorosi et al., 2000). Rhizobia that colonize the host cells become surrounded by a host-derived membrane, forming an organelle-like structure called the symbiosome. Within symbiosomes, rhizobia undergo metabolic changes, and in particular legumes they also undergo morphological changes, resulting in differentiation into their symbiotic forms

(bacteroids) that are able to fix atmospheric dinitrogen into a form that is biologically accessible for the plant (Prell and Poole, 2006, Udvardi and Poole, 2013).

1.3. Inverted Repeat Lacking Clade (IRLC) legumes enforce Terminal Bacteroid Differentiation (TBD)

Depending on the legume host, nodules can be of different types in terms of their structure and development (Franssen et al., 1992). Legume species belonging to the Inverted Repeat Lacking Clade (IRLC) (Figure 1.4), characterized by the lack of a 25-kb inverted DNA repeat in the plastid genome (e.g., *Pisum sativum, Medicago truncatula*), form

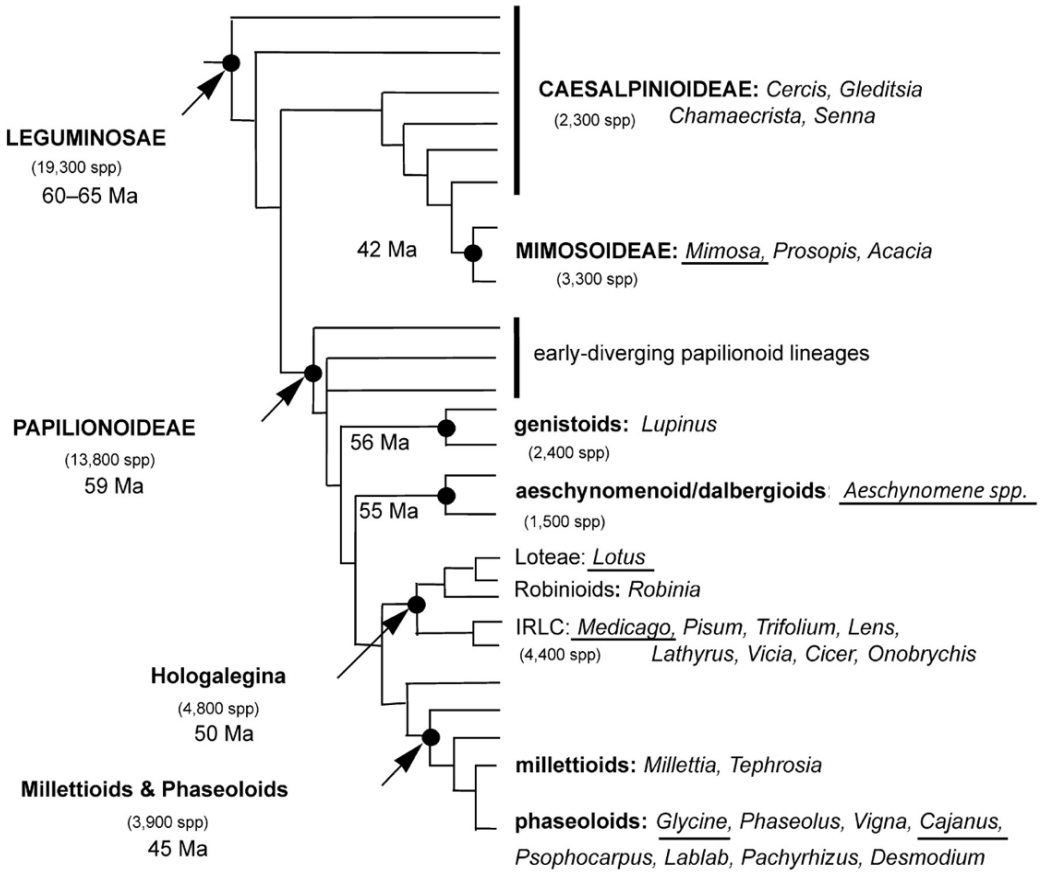


Figure 1.4: Simplified phylogenetic tree of the leguminosae family. Among the three well-studied legume subfamilies (Papilionoideae, Mimosoideae and Caesalpinioideae) nodulation is very common in Papilionoideae, then Mimosoideae, whereas only few members of Caesalpinioideae subfamily form nodules. Papilionoideae is the largest among the subfamilies (~13,800 spp.) and can be delineated into different evolutionary clades (e.g., IRLC). Different plants species that are mentioned in this thesis are underlined. The tree is adapted from Gepts et al. (2005); it was not exactly scaled but different important diverging nodes are indicated with black circles and corresponding historical time periods are mentioned.

indeterminate nodules with a persistent meristem (Lavin et al., 1990). As reviewed by Pan and Wang (2017), indeterminate nodules can be spatially divided into successive developmental zones (Figure 1.5): the uninfected meristematic zone (ZI), that actively produces new nodule cells; the infection zone (ZII), where rhizobia infect host cells; the interzone (IZ), where both host and bacterial cells undergo profound changes and become significantly enlarged; the fixation zone (ZIII), where fully differentiated bacteroids fix nitrogen; and the senescence zone (ZIV), where both bacteroids and host cells die.

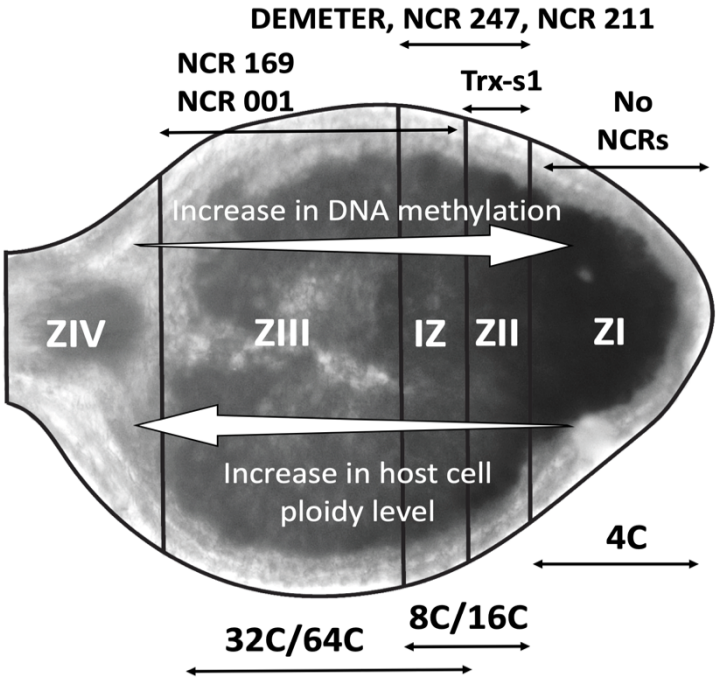


Figure 1.5: Schematic of the different zones in the nodule. From the meristem (ZI) there are four distinct zones in indeterminate nodules whose ploidy level increases: the infection zone (ZII), the interzone (IZ), the nitrogen fixation zone (ZIII) and the senescence zone (ZIV). Zonal changes are also marked by specific expression of different important enzymes and small peptides (e.g., NCRs) that are discussed in detail later in the chapter. The image in the drawing shows a nodule for illustrative purposes, with shading indicating levels of leghemoglobin. Figure from Roy P et al. (2020).

Within indeterminate nodules of IRLC legumes, rhizobia are subjected to tight host control that results in terminal bacteroid differentiation (TBD). This is an irreversible change characterized by genome amplification, inhibition of cell division, cell elongation, increased membrane permeability and almost complete loss of bacterial reproductive ability (Mergaert et al., 2006).

Nodule functioning is different in non-IRLC legumes, e.g., *Glycine max, Lotus japonicus* - most non-IRLC species form determinate nodules that are round shaped and have short-lived meristematic activity. Bacteroids in determinate nodules do not show features of TBD and are capable of multiplying as free-living bacteria when isolated from nodules (Franssen et al., 1992, Mergaert et al., 2006, Pan and Wang, 2017).

Endoreduplication occurs in both determinate and indeterminate nodules and results in symbiotic polyploid nodule cells (usually 64C or 128C, where C denotes haploid DNA content), however, rhizobial endoreduplication mostly occurs in the indeterminate nodules of IRLC legumes (Mergaert et al., 2006, Maroti and Kondorosi, 2014) and is comparatively rare or less significant outside the IRLC. For example, in *Medicago truncatula*, endoreduplication of *Sinorhizobium meliloti* results in an amplification of genome content up to 24C, compared with 1C/2C in free-living bacteria (Mergaert et al., 2006). Interestingly, whilst the non-IRLC legume *Mimosa pudica* forms indeterminate nodules with a persistent meristem when infected by *Cupriavidus taiwanensis*, bacteroids are not terminally differentiated (Marchetti et al., 2011). Therefore, TBD might not be related to nodule ontogeny or phenotype but rather to the genetic (e.g., IRLC vs non-IRLC) and/or nodule-specific molecular (e.g., expression of certain group of peptides) profile of the host.

1.4. IRLC legume genomes contain large numbers of Nodule-specific Cysteine-Rich (NCR) peptide-encoding genes

TBD in IRLC legumes is mediated, at least in part, by a family of small Nodule-specific Cysteine-Rich (NCR) peptides (Mergaert et al., 2003, Van de Velde et al., 2010), that are expressed by the host almost exclusively in nodule cells with the exception of only five (out of 334 that were analyzed) being reported to be expressed outside nodules (Guefrachi et al., 2014). All IRLC legumes analyzed so far express NCRs, although the number of genes varies significantly, from 7 to ~700, depending on the host species (Montiel et al., 2017). As reviewed in Roy P et al. (2020), NCRs are cysteine rich, defensin-like (DEFL) plant peptides that are derived from their non-functional precursors upon proteolytic cleavage of an N-terminal Signal Sequence (NSS). The NSS directs the peptide through the endoplasmic reticulum (ER) and once cleaved by Signal Peptidase Complex

(SPC), mature NCRs are secreted outside ER to interact with the bacteroids (Tavormina et al., 2015). NCR peptides are diverse and are only highly similar to each other in two regions - the NSS region (Nallu et al., 2014) and 4 or 6 regularly spaced cysteine residues within the coding sequence (Figure 1.6a). Mature NCRs consist of around 20-50 amino acids including the 4 or 6 cysteines that potentially form 2 or 3 disulphide bridges. The number of cysteines distinguishes them from plant defensins that are usually 45-70 amino acid long and have 8 cysteine residues (forming 4 disulphide bridges) (Figure 1.6b) (Farkas et al., 2018, Carvalho and Gomes, 2009, Mergaert et al., 2003). Other distinguishing features between these small peptides are that defensins are expressed ubiquitously, not being nodule-specific, and are all cationic (isoelectric point, p*I*, ~ 9.0) (Carvalho and Gomes, 2009). In contrast, based on their p*I*, NCRs can be divided into cationic, anionic and neutral groups (Mergaert et al., 2003). Like defensins, certain cationic NCRs (e.g., NCR247, NCR335) exhibit strong *in vitro* antimicrobial activity (Farkas et al., 2018, Van de Velde et al., 2010). However, among the NCRs detected inside bacteroids by Durgo et al. (2015), most are anionic or neutral. The specificity of expression location (nodules for NCRs; almost everywhere for defensins) and differing molecular properties (e.g., p*I*, charge) suggest that NCRs are probably involved in diverse functions relating to nodulation, whereas defensins are involved in general defense in a non-organ specific manner.

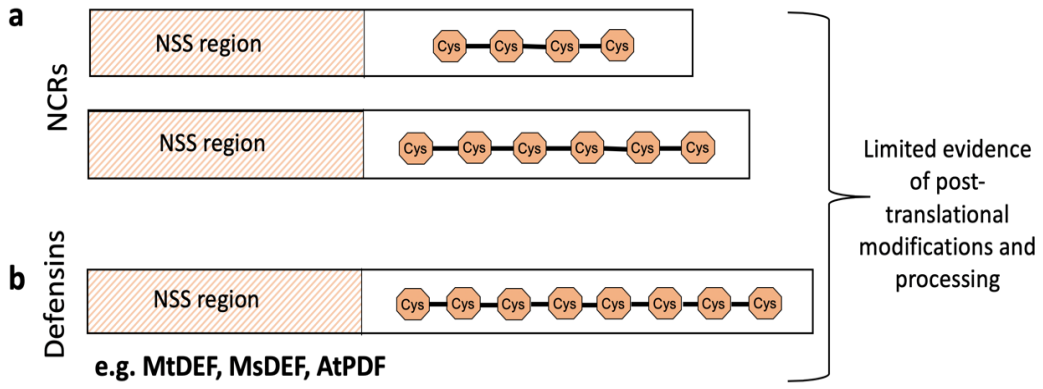


Figure 1.6. Simplified precursor structure of NCR peptides with defensins for comparison. (a) NCRs in indeterminate legumes such as *Medicago truncatula* and *Aeschynomene* are characterized by the presence of a signaling peptide at the 5' end and 4 or 6 cysteines at conserved positions. (b) Defensins similarly have a 5' signal peptide but have eight cysteine residues; NSS: N-terminal Signal Sequence. Figure from Roy P et al. (2020).

To date, NCR peptides have not been reported outside IRLC legumes, with one exception of *Aeschynomene spp*., belonging to comparatively ancient Dalbergoid lineage of legumes. The nodules of *Aeschynomene spp*. are neither determinate nor indeterminate; they have a central zone of infected tissue which originates from the consecutive divisions of one or multiple cortical cells, initially infected though a non-transcellular, infection thread-independent process (Bonaldi et al., 2011). Therefore these nodules do not have an uninfected primordium and both bacterial and host cell differentiate simultaneously from the start. However, they do express NCR-like peptides that are cysteine-rich and highly nodule specific, and show signs of TBD, as bacteroids are enlarged, polyploid and irreversibly differentiated. Interestingly, NCR-like peptides in *Aeschynomene spp*. are different (in terms of sequence similarity) from their counterparts in IRLC legumes, suggesting two independent evolutionary events with convergent and common roles in regulating bacteroids (Czernic et al., 2015). To clarify, NCRs in IRLC legumes and the NCR-like peptides from *Aeschynomene spp*. were most likely originated and then evolved independently as their respective harbouring species belong to different evolutionary lineages. However, since NCR-like peptides in *Aeschynomene spp*. were shown to be implicated in bacterial differentiation which is analogous to the function of NCRs, it can be hypothesised that the products of two independent evolutionary events converged into one common function.

TBD is likely to be beneficial for host legumes for a number of reasons - firstly, it has been shown by Oono and Denison (2010) that swollen bacteroids are better at conferring host benefits (enhanced N-fixation and better plant growth per unit of nodule mass) compared to their non-swollen counterparts in different host species. Therefore, terminally differentiated and elongated bacteroids could be considered to be more useful for their hosts. Secondly, the loss of bacterial reproductive capability could be a way to control bacterial number with the consequence of a better cost-gain trade-off for the host. Finally, as reviewed by Mergaert et al. (2017), activity of symbiotic AMPs such as NCRs can result in introduction of transient membrane pores or an increase of permeability of the bacterial membrane that might allow better metabolic exchange between partners. To what extent, if any, the terminal differentiation process offers an advantage over non-NCR expressing, non-IRLC legumes is not yet known, but if understood it would be highly relevant for understanding legume-rhizobial co-evolution. Such knowledge is likely

to be enormously useful in achieving our ultimate goal of introducing N-fixing capability in non-legume crop plants.

1.5. NCRs primarily, but not only, act as mediators of TBD

A number of findings have implicated NCRs as the key mediators of terminal or irreversible bacteroid differentiation and amplification of the bacterial genome. Van de Velde et al. (2010) ectopically expressed eight different *M. truncatula* NCR genes in *Lotus japonicus*, an NCR-lacking non-IRLC legume. They reported that expression of at least one of them, NCR035, resulted in bacteroids with enhanced membrane permeability, cell elongation and compromised reproductive ability; all reminiscent of TBD. Moreover, *in vitro* treatment of free-living *S. meliloti* with a sub-lethal concentration of certain NCR peptides inhibited cell division and induced bacteroid-like characteristics (Van de Velde et al., 2010, Penterman et al., 2014). In general, the extent of bacteroid differentiation in terms of changes in bacterial morphology depends on the size and complexity of the NCR gene family (Montiel et al., 2017, Montiel et al., 2016). A comparative study amongst ten IRLC legumes showed that the higher the number of expressed NCRs, the greater the degree of bacteroid elongation (Montiel et al., 2017). The extent of bacteroid elongation also correlates with their morphotype (shape), varying from swollen to elongated and elongated-branched, depending on the NCR profile of host species. For example, in *Glycyrrhiza uralensis* where only 7 NCR genes are known to be expressed, the infecting rhizobia differentiate into slightly swollen (compared to free-living bacteria) bacteroids, whereas with about 700 NCR genes, *M. truncatula* harbours elongated and branched bacteroids (Montiel et al., 2017, Montiel et al., 2016).

Particular NCRs have precise and non-redundant functions in different phases of rhizobial differentiation (Horvath et al., 2015, Kim et al., 2015b). Two *M. truncatula* mutants, *dnf7* and *dnf4* (*defective in nitrogen fixation* 7 and 4), both have a similar phenotype with small and white nodules incapable of fixing nitrogen. *dnf7* mutation disrupts the NCR169 gene, whereas *dnf4* abolishes NCR211 functions (Horvath et al., 2015, Kim et al., 2015b, Lang and Long, 2015). The transcriptomic profiles of both mutants are also very similar (Lang and Long, 2015). In wild type plants, both NCR169 and NCR211 share an overlapping expression pattern where NCR211 is largely expressed in the infection and interzone and NCR169 is exclusively expressed in the interzone and mature fixation zone

(Figure 1.5) (Horvath et al., 2015, Kim et al., 2015b). Moreover, both NCRs localize to the peribacteroid space, the region between symbiosome membrane and bacteroids (Pan and Wang, 2017). However, the extent of bacteroid differentiation is different in these mutants. While bacteroids in *dnf7* mutants (without NCR169) never reach terminal differentiation, bacteroids in *dnf4* mutants (without NCR211) die prematurely in the fixation zone just before or soon after reaching terminal differentiation (Kim et al., 2015b, Horvath et al., 2015). These observations suggest that NCR169 has a critical role in terminal differentiation while NCR211 might be important for survival of differentiated bacteroids in the symbiosome. In both *dnf7* and *dnf4* mutants, bacteria do differentiate to some extent, therefore it is unlikely that NCR169 and NCR211 interfere in the cell cycle. Also, their function might not be the elimination of incompatible bacteria because they are required for bacteroid survival and TBD. A pull-down assay suggests that NCR169 (as well as NCR28 and NCR290) might physically interact with NCR247, a cationic NCR with antimicrobial effect and unusually high p*I* (10.39) (Farkas et al., 2014). As discussed by Mergaert (2018), some NCRs could function counteractively by binding one another to compensate or reduce the harmful effects of other NCRs or to facilitate/modify their functions. For example, an anionic or neutral NCR might bind to cationic NCRs to antagonize their antimicrobial activity.

Proteins involved in the maturation process of NCR peptides are also important for effective symbiosis. This is illustrated in the case of *M. truncatula dnf1* mutants where a component of the nodule specific peptidase complex Defective in Nitrogen Fixation 1 (DNF1), that cleaves the NSS from pre-mature NCR peptides, is non-functional. In the absence of functional DNF1, a number of NCRs tested by Van de Velde et al. (2010) were found to be trapped inside the endoplasmic reticulum, effectively blocking bacteroid differentiation.

It has been found that different plant- and bacteria-derived thioredoxin (Trx) and glutaredoxin (Grx) systems are necessary for nitrogen-fixing symbiosis (reviewed in Alloing et al. (2018)). For example, one plant-derived thioredoxin (Trx-s1) has been found to be induced in the nodule infection zone and to interact with several NCR peptides. While inactivation of Trx-s1 in the host plant negatively affects TBD, the ectopic expression of this enzyme in the endosymbiont partially enhances TBD in Trx-s1 inactivated plants (Ribeiro et al., 2017). It is likely that these redox enzymes play a crucial

role in maintaining the appropriate cellular redox state including that of the multiple cysteine residues in NCR peptides, thus regulating the functions of NCRs. Therefore not only NCRs, but also other cellular components that regulate NCR activity, are essential for establishing and maintaining effective symbiosis.

1.6. Rapid gene family expansion and variation in spatio-temporal expression patterns enable NCR functional diversity

The genome structure of the Papilionoideae subfamily that contains most nodulating legume species was strongly shaped by a whole genome duplication event around 58 million years ago (MYA) (Young et al., 2011). The origin of the IRLC group can be dated back to ~39 MYA, but exactly when and how the NCR gene family appeared is still far from clear (Sprent, 2007, Young et al., 2011). One hypothesis suggests a recent origin of NCR genes that then rapidly expanded (possibly by gene duplications) and diversified in certain species (e.g., *M. truncatula*) (Montiel et al., 2017). This is supported by several observations - firstly, NCRs or genes that resemble NCRs from the IRLC and dalbergoid clades, respectively, are lineage specific (orphan) genes that lack orthologs, even in other closely related lineages (Mergaert et al., 2003, Czernic et al., 2015); this suggests an evolutionary recent origin. As NCRs are antimicrobial peptide (AMP)-like, and many AMPs are orphan genes like NCRs, it could be that NCRs evolved from a primitive AMP gene upon duplication and rearrangement. Secondly, NCR numbers vary highly (7 to ~700) among IRLC legumes, implying species-specific expansion. Mature NCR peptides from orthologous genes of different *M. truncatula* accessions contain signatures of both purifying (conservation) and diversifying selection, where other DEFLs are under purifying selection only (Nallu et al., 2014). NCR gene loci that are under purifying selection might share common functions, whereas those under diversifying selection could be rapidly evolving towards novel functions. Thus NCRs seem to be an example of subneofunctionalization (He and Zhang, 2005), where members of duplicated gene families acquire new functions simultaneously. Finally, *M. truncatula*, that encodes over 700 NCR peptides (the highest number currently known), harbours relatively more locally duplicated genes (e.g., 3.1 fold higher than soybean) and a high number of transposons, located close to many NCR genes (Young et al., 2011, Satge et al., 2016, Roy P et al., 2020). NCR genes are spread over the *M. truncatula* genome, found in all the chromosomes often in small clusters (Alunni et al., 2007) (Figure 1.7). Together, this

Figure 1.7: Schematic diagram showing genomic position of 111 NCR genes in chromosome 4 of *M. truncatula*. Each vertical line corresponds to an NCR gene. Some of the NCR gene clusters (several NCR genes in very close proximity) are marked with black arrows. Centromere position is marked by black circle. Figure generated using NCR gene IDs and ShinyGo v0.61 tool available online at: www.bioinformatics.sdstate.edu/go

genomic landscape suggests a process of transposon-mediated long-distance duplication, followed by local duplication and diversification of NCRs in this species. Such duplication events might have affected not only NCRs, but other symbiosis-related gene families (e.g., leghemoglobins, flavonoid signaling genes etc.) in *M. truncatula*, as their numbers are particularly high (Young et al., 2011). Among these symbiosis related gene families, some are also strongly linked to nodule function, e.g., Glycine Rich Peptides (GRPs) and Calmodulin-like proteins (CaMLs) (Pecrix et al., 2018). In general, all of these genes are found in clusters across the genome, similar to the case with NCRs, suggesting duplication events were associated with their evolution (Alunni et al., 2007, Young et al., 2011, Pecrix et al., 2018).

Like many other genes, the expression of many NCR genes is regulated by DNA demethylation and histone modification (Satge et al., 2016, Nagymihaly et al., 2017). However, one particular DNA demethylase gene, *DEMETER* (*DME*) was found to be expressed preferentially in differentiating nodule cells (Figure 1.5) (Satge et al., 2016), possibly involved in the regulation of many NCRs. Also, expression of some NCR genes is associated with reduction in the suppressive histone (H) mark H3K27me3 (trimethylation of H3 lysine-27) and increase in the activating H3K9ac (acetylation of H3 lysine-9) (Nagymihaly et al., 2017), which could form part of the same methylation regulatory mechanism. As mentioned earlier, many NCRs are neighbored by transposons that are usually silenced by DNA methylation. Interestingly, DNA demethylation that activates NCRs can also result in activation of these transposons (Satge et al., 2016). Therefore, the accumulation of these epigenetic modifications, transposon mediated movement or duplication events might have a role in the evolution or expansion in

expression and functional diversity of NCRs. Our understanding of these evolutionary events should increase with analysis of sequences from whole genome sequencing and molecular characterization of additional IRLC legumes.

The functional diversity of NCRs is further highlighted by the observation that not all NCRs are expressed at the same time and in the same place. In fact, different NCRs are activated in successive but overlapping locations, both spatially, in different nodule zones during development, and temporally (Roux et al., 2014, Marx et al., 2016, Guefrachi et al., 2014) (Figure 1.5). A study taking advantage of laser capture microdissection coupled to RNA sequencing showed that only a few NCRs were expressed in the distal part of the ZII (infection site) whereas others were expressed in the proximal part of the ZII as well as in the IZ, and the ZIII (Roux et al., 2014). Peak expression was observed at the IZ (with 411 expressed NCRs), where bacteroids undergo rapid differentiation. No NCRs were detected in ZI and NCR expression was minimal in the ZIII (fixation zone) (Roux et al., 2014, Guefrachi et al., 2014). These findings suggest that NCRs are mostly expressed in response to invading rhizobia and deactivated when bacterial differentiation is complete, supporting their role in TBD. Marx et al., (2016) examined the expression over time of NCRs. They collected nodule samples at 10, 14 and 28 days post rhizobial inoculation and upon proteomic analysis of these samples, they divided NCRs into early, intermediate and late, based on their relative temporal abundance (Marx et al., 2016). Early stage-expressed NCRs might be involved in enhancing rhizobial membrane permeability that allows other NCRs to enter the bacterial cytosol. Intermediate and late-expressed NCRs could target rhizobial metabolism to advance differentiation into N-fixing bacteroids (Marx et al., 2016). Although NCRs are generally thought to be very specific to nodules, at least two NCRs (NCR122 and 218) were reported to be expressed at a similar level in both nodules and uninfected roots, and three others (NCR077, NCR235 and NCR247) were found to show somewhat reduced nodule specificity as some lower levels of expression were also detected in uninfected roots (Guefrachi et al., 2014), implying they might play other roles (so far unknown) beyond that of the regulation of nodulation.

1.7. NCR activity targets the rhizobial cell cycle and other crucial rhizobial cellular pathways

Much of the information within the proposed mechanism by which NCRs regulate bacterial activity is informed by the study of NCR247 (Farkas et al., 2014, Penterman et al., 2014), the best characterized NCR so far (Figure 1.8). The general model is that NCR peptides alter rhizobial cell cycle progression by targeting cell cycle components (Mergaert, 2018). One such regulator is the CtrA protein that acts late in the cell cycle to repress the initiation of replication and promote cell division (Collier, 2012, Pini et al., 2015). The loss of CtrA function is associated with the formation of swollen, polyploid bacteria that resembles bacteroids inside nodules. Therefore, it is possible that NCR peptides exert control by limiting CtrA function (Pini et al., 2015). *In vitro* application of NCR247 downregulates *ctrA* gene expression and also most *CtrA*-regulated genes in *S. meliloti* (Penterman et al., 2014). Moreover, *CtrA* transcript levels drop steadily along with bacterial differentiation (Roux et al., 2014), and in mature bacteroids the CtrA protein is almost absent (Pini et al., 2013). In addition, mutations in bacterial genes that inhibit CtrA (e.g., *cbrA, cpd*R or *divJ*) result in abnormal bacteroid differentiation, leading to formation of non-functional nodules (Mergaert, 2018, Pini et al., 2013). While all of these findings implicate CtrA as a potential regulatory node for NCR mediated effects, the actual mode of action is still not clear. NCR247 not only affects CtrA but also inhibits the expression of two other important rhizobial cell cycle regulators, DnaA and GcrA. DnaA is important for the initiation of prokaryotic DNA replication and it activates GcrA. Once activated, GcrA represses DnaA and upregulates CtrA to facilitate cell division (Penterman et al., 2014, Collier, 2012). While inhibition of GcrA (an activator of CtrA) fits the proposed mechanism, forcing the bacteroid to stop cell division, the presumed downregulation of DnaA is not so straightforward to explain, as the continuation of DNA replication is required for TBD. However, a study in *Caulobacter crescentus*, a model bacterial species that is used to study the regulation of the cell cycle, suggests that CtrA and DnaA both compete for binding to the origin of replication to repress or initiate DNA replication, respectively (Taylor et al., 2011). Therefore, it could be that when CtrA levels are significantly reduced, even low levels of DnaA can still initiate DNA replication, enabling endoreduplication to take place. At the same time, DnaA activity is also regulated post-transcriptionally (Collier, 2012, Penterman et al., 2014), therefore a lower transcript level does not necessarily represent a lower protein level or reduced activity. NCR247 is

also known to interact with another conserved bacterial cell cycle protein, FtsZ (Farkas et al., 2014), which is involved in the formation of the Z-ring and subsequent cell division (Ma et al., 1996, Lan et al., 2007). NCR247 is thought to disrupt septum specific localization of FtsZ protein through inhibiting the polymerization of FtsZ by binding to its monomers (Farkas et al., 2014, Penterman et al., 2014).

A number of other bacteroid cellular components have been found to interact physically with NCR247, including the chaperonin GroEL, ribosomal proteins, ATP synthase subunits, members of the nitrogenase complex and several TCA-cycle related enzymes (Farkas et al., 2014, Mergaert, 2018). The NCR247-GroEL interaction is particularly interesting as one of the five *S. meliloti* groEL genes (groEL1) is essential for functional symbiosis (Bittner et al., 2007). In general, GroEL interacts with hundreds of proteins, facilitating their proper folding, thus binding might modify the NCR247 interaction network that can alter bacterial metabolism. The interactions of NCR247 with ribosomal proteins and metabolic enzymes could be a means to modulate bacterial protein synthesis and metabolic activity, respectively (Penterman et al., 2014). It can be predicted

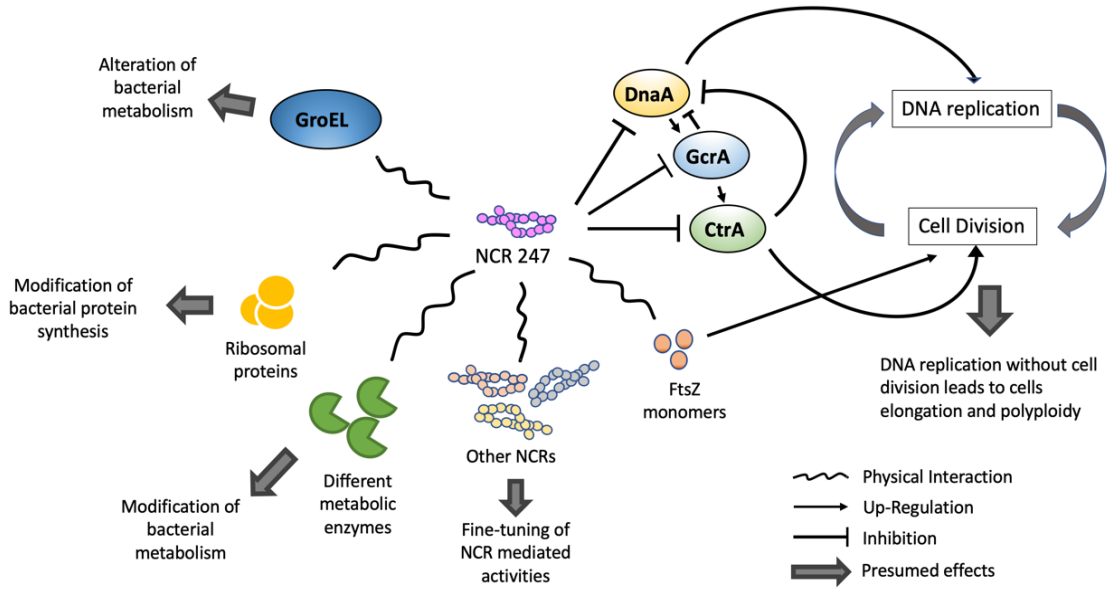


Figure 1.8. NCR247 modulates rhizobial activity via regulation of bacterial DNA replication, protein synthesis and metabolism. NCR247 appears to modulate a wide range of microbial pathways by repressing different cell cycle regulators (e.g., CtrA, DnaA, GcrA) or binding to a range of bacterial proteins (e.g., GroEL, FtsZ, ribosomal proteins) as well as other plant NCRs (e.g., NCR028, NCR169, NCR290). Figure adapted from Roy P et al. (2020).

that other NCRs also affect bacterial cellular pathways and have similarly broad effects. However, functional characterization of such other NCRs is essential in order to understand to what extent this is a widespread regulatory role, particularly since, unlike most other NCRs, NCR247 is not exclusively nodule-specific (mentioned earlier), having low levels of expression in roots and stems. Therefore, NCR247 could have functions outside nodulation (e.g., providing general immunity like defensins) and might not be fully representative of this family of genes.

1.8. The role of NCRs in the legume-rhizobia arms race

Legume-rhizobia symbiosis is typically depicted as an altruistic mutualism between two partners. However, a closer inspection of NCR-mediated TBD of rhizobia reveals a 'scuffling' situation where IRLC legumes subjugate their bacterial partner into a semi-viable, sterile state. To counter this, endosymbionts also upregulate different stress and defense related genes to minimize or escape from host-imposed controls. Potentially, the higher the number of NCRs and the more diverse the NCRs are in the host, the greater the control over symbionts. However, ultimately, the outcome of the nitrogen fixation symbiosis depends on the host legume-endosymbiont balance to control each other's functions (as reviewed by Roy P et al., 2020).

So far, one observation suggests that some rhizobial strains might have evolved mechanisms to avoid terminal differentiation entirely when interacting with IRLC legumes. *Glycyrrhiza uralensis* (liquorice), an IRLC legume able to trigger terminal differentiation with some rhizobial species (e.g., *Rhizobium galegae* bv. orientalis) (Montiel et al., 2017), fails to enforce terminal differentiation when infected by *Sinorhizobium fredii* strain HH103 (Crespo-Rivas et al., 2016). Although HH103 escapes terminal differentiation, it still fixes nitrogen for liquorice, resembling non-IRLC legume bacteroids. Remarkably, *S. fredii* HH103 also shows no or minimal sensitivity when treated *in vitro* with NCR247 and NCR335. HH103 bacteroids isolated from liquorice exhibit altered lipopolysaccharide (LPS) content on their outer membrane, a key molecule in legume-rhizobia interactions (Downie, 2010). However, when HH103 infects non-IRLC legumes, *Glycine max* or *Cajanus cajan*, such modification of LPS is not observed (Crespo-Rivas et al., 2016). LPS modifications, could therefore be a possible bacterial defense mechanism against NCR-mediated control. Furthermore, although *G. uralensis* expresses

NCR peptides, only 7 have been identified in the *G. uralensis* genome to date (Montiel et al., 2017), suggesting that it might not have the genomic potential to 'dominate' over a broad range of rhizobial species. In many cases (discussed in section 1.7.1), counter-active measures between host and rhizobia result in incompatibility, even after successful initiation of nodulation. Therefore, in the arms race between legume and rhizobia, a fine-tuned balance has to be met in order to both initiate and maintain functional N-fixing symbiosis.

1.8.1. NCRs can act as determinants of host-symbiont compatibility

As mentioned earlier, early interaction and symbiotic compatibility between legumes and rhizobia is first informed by the initial perception of rhizobial nodulation (Nod) factors by the host through its Nod-factor receptors (Oldroyd, 2013, Lagunas et al., 2015). Host legumes also recognize other bacterial effectors or surface proteins, e.g., LPS or microbe associated molecular patterns (MAMPs), and elicit immune responses against prospective symbionts. The success of symbiotic interactions depends on the extent of these immune responses and rhizobial susceptibility to them (Cao et al., 2017). Even if these initial checkpoints have been successfully passed, symbiotic incompatibility can still occur at the later phases of nodule development and NCR peptides might play an important role at those points. Insight into this came from the discovery of two NCR-encoding genes, *NFS1* (*NITROGEN FIXATION SPECIFICITY* 1; Medtr8g073380) and *NFS2* (Medtr8g465280), that determine the fate of rhizobial colonization depending on which allelic variants of those genes are present and the identity of bacterial strain that is infecting. This has been demonstrated in two *M. truncatula* accessions, Jemalong A17 and DZA315.16, that, respectively, produce non-N-fixing (Fix-) or N-fixing (Fix+) nodules, with particular *S. meliloti* strains (A145 and Rm41) (Wang et al., 2018). Further studies have revealed that *NFS1* and *NFS2* alleles from A17 cause early senescence of bacteroids, whereas alleles from DZA315.16 do not have such an effect. The A17 alleles also act dominantly over DZA315.16 alleles in heterozygous plants, suggesting that A17 alleles might encode for peptides that kill those particular rhizobial strains, resulting in an incompatible interaction. However, other *S. meliloti* strains are not susceptible to NFS1 and NFS2-mediated killing, regardless of which alleles are present. Other than this strain-specific antimicrobial effect, *NFS1* and *NFS2* might not be essential for N-fixing symbiosis since elimination of the genes altogether results in functional symbiosis (Wang

et al., 2018, Yang et al., 2017, Wang et al., 2017). The sensitivity of particular bacterial strains to NFS1 and NFS2 might be due to the chemical structure of the bacterial exopolysaccharides (EPS). Interestingly, the EPS of incompatible strains are significantly less succinylated than the EPS of the compatible strains. Moreover, expression of EPS-compatible strain biosynthesis genes in an incompatible strain resulted in compatibility (Simsek et al., 2007, Simsek et al., 2013). Therefore, the presence of abundant negatively charged succinate groups in the bacterial EPS of the compatible strains (along with alteration in LPS, as discussed earlier) could act as a protection mechanism against the antimicrobial effect of particular NCR peptides, potentially disrupting their recognition or binding. This idea is further supported by the finding that expression of exopolysaccharides in *S. meliloti* is associated with better protection from NCR247 (Arnold et al., 2018). Nonetheless, the example of *NFS1* and *NFS2* highlights one mechanism for how host NCRs dictate the outcome of symbiotic interactions (Figure 1.9).

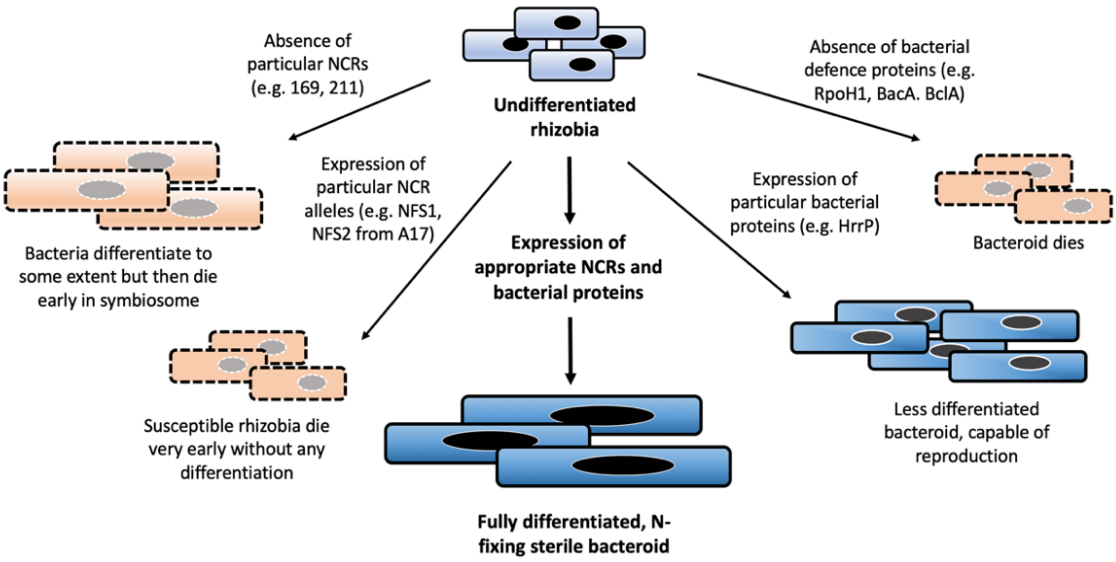


Figure 1.9. Expression of particular plant NCRs and bacterial proteins alters the fate of rhizobial bacteroids in the nodule. Research has identified the importance of NCRs not only that have defensin-like roles to kill rhizobia (e.g. NFS1 and NFS2 in *M. truncatula* A17) but that are required for bacterial survival and differentiation (e.g. NCR169, NCR211). Bacterial-expressed genes and defense proteins are similarly required for proper differentiation and bacterial survival, but also their capability to reproduce. When all required plant and bacterial proteins are expressed, rhizobia are able to fully differentiate and are capable of nitrogen fixation. Figure adapted from Roy P et al. (2020).

1.8.2. Bacterial defense against NCR actions

A number of rhizobial stress-related genes are upregulated when *S. meliloti* is treated with *in vitro* synthesised peptides NCR247 or NCR335. Most prominent among them are: an RNA polymerase σ-factor, RpoH1; a heat shock protein, IbpA and a methionine sulfoxide reductase A, MsrA1 (Tiricz et al., 2013). Upregulation of RpoH1 is particularly interesting as *rpoH1* mutants die prematurely in the symbiosome, reflecting that RpoH1-related responses are important for bacteroid survival (Mitsui et al., 2004). Since host plant cells infected by rhizobia introduce various oxidative/nitrosative stresses to bacteria (Syska et al., 2019), the function of an antioxidant enzyme such as MsrA (Fu et al., 2017) and other rhizobial catalases and superoxide-dismutases appears to be important for proper bacteroid differentiation (Jamet et al., 2003). The two-component regulatory systems ExoS-ChvI and FeuP-FeuQ, and their regulons that are involved in EPS and cyclic glucan production, respectively (Belanger et al., 2009, Griffitts et al., 2008), are also rapidly induced in rhizobia upon treatment with NCR247 (Penterman et al., 2014). Many pathogens also use two-component systems to sense AMP activity (Otto, 2009) and rhizobial mutants disrupted in ExoS-ChvI and FeuP-FeuQ functions fail to promote infection thread development and functional symbiosis (Belanger et al., 2009, Cheng and Walker, 1998). Such observations suggest that these two-component systems might be utilized by rhizobia to sense and promote resistance against NCR peptides in nodules.

Apart from these broad-spectrum stress-related responses, rhizobia have some unique strategies that seem to enable them to counteract the effects of NCRs. For example, two orthologous membrane transporter proteins, BacA in *S. meliloti* and BclA from *Bradyrhizobium* spp., are critical for symbiosis in NCR-producing legumes (Glazebrook et al., 1993, Haag et al., 2011, Guefrachi et al., 2015). *bacA* and *bclA* bacterial mutants initiate nodulation normally but die early in the symbiosome of NCR-producing legumes. It has been found that *bacA* mutants persist longer inside nodules of *M. truncatula* mutants deficient in NCR functions (e.g., *dnf1*) (Haag et al., 2011). BacA and BclA are similar to ABC transporters and are likely to involve in the import of a number of different peptides (LeVier and Walker, 2001). Therefore, it is hypothesised that these transporters act by importing NCR peptides into the cytosol, effectively removing them from the cell surface (the site of action for many AMPs). Since the homologs of *bacA* and *bclA* are widely present in all rhizobia, and even in some non-symbiotic bacteria (Glazebrook et

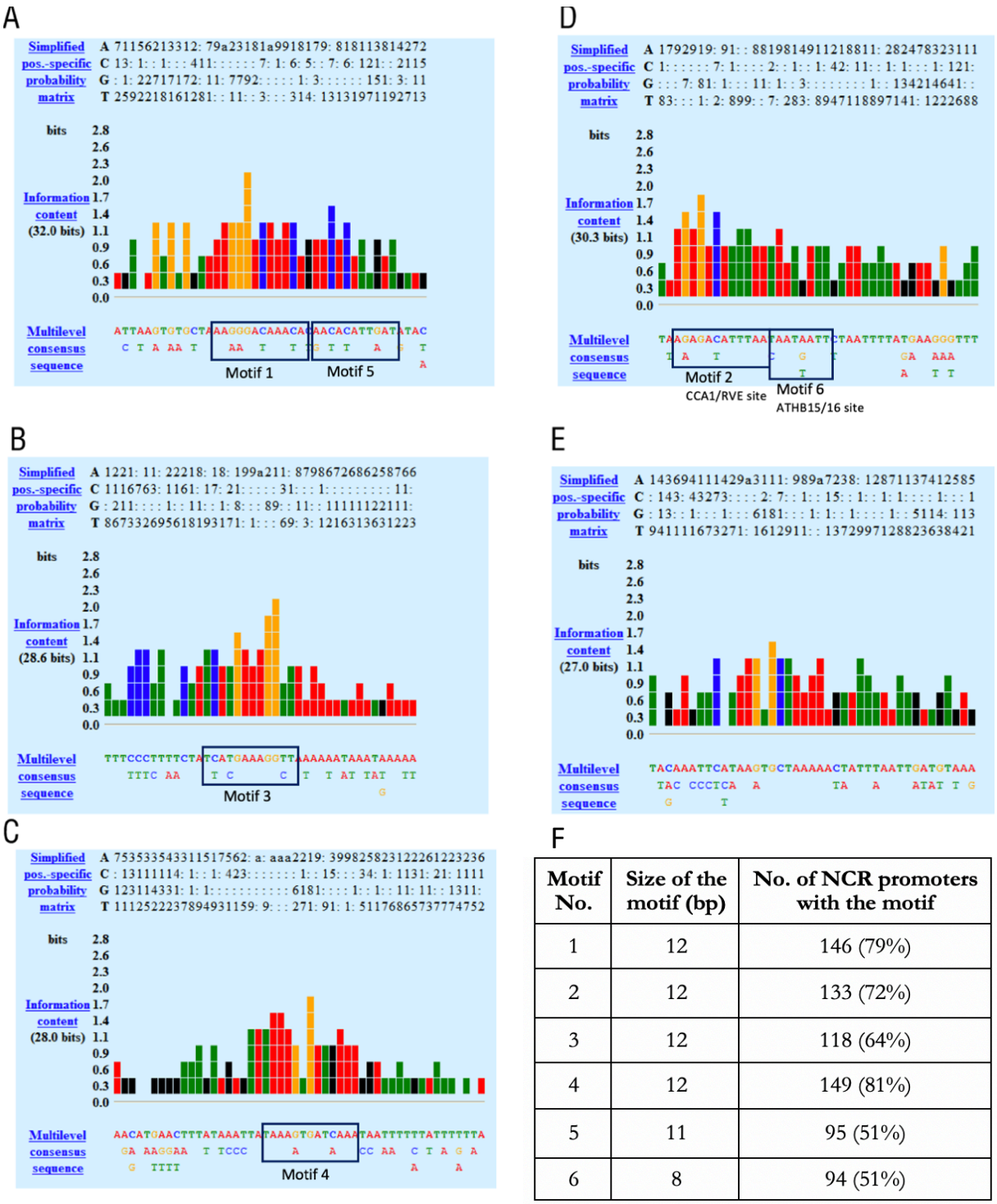
al., 1993, Guefrachi et al., 2015, Mergaert, 2018), it is possible that their role in symbiosis might be supplementary to their non-symbiotic (yet mostly unknown) functions. In addition, because the presence or absence of these transporters determines whether symbionts would be able to persist in certain legumes, these transporters might have played an important role in symbiotic co-evolution, in determination of legume-rhizobia partner specificities.

Some rhizobia might have evolved a more efficient way of resisting NCRs. *S. meliloti* strain B800, carrying an accessory plasmid (pHRB800), has been found to form Fix+ nodules in *M. truncatula* accession A17, but Fix- nodules in accession A20 (Crook et al., 2012). A closer inspection reveals that the plasmid contains an M16A family zinc-metallopeptidase-encoding gene that can degrade many NCRs of different p*I*, *in vitro*, although with varying efficiency (up to 15 folds among 10 NCRs tested) (Price et al., 2015). When this gene is present in rhizobia it affects N-fixation (e.g., formation of defective, non-N-fixing nodules) with several *M. truncatula* accessions (ecotypes), but the same bacterial strain without the gene has no effect on N-fixation with corresponding hosts. Therefore, the gene is designated as a *HOST RANGE RESTRICTION PEPTIDASE* (*hrrP*) as its presence limits the effective N-fixing host range of rhizobia. To clarify more, if this gene was not existent, more rhizobia would form proper N-fixing symbiosis, hence it restricts the host range. Interestingly, HrrP-expressing (HrrP+) bacteria proliferate better in both functional (A17 accession) and non-functional (A20 accession) *M. truncatula* nodules compared to *hrrP* mutants. This is likely because HrrP diminishes the host NCR arsenal completely or to some extent (Price et al., 2015). It is, however, not clear whether NCRs are the natural substrate of HrrP or if it simply degrades many NCRs, as they are analogous to the so far 'unknown' substrate(s). It is also not clear how some hosts (e.g., A17) induce HrrP+ bacteria to fix nitrogen to a normal level (Price et al., 2015). Variability of NCR genes among different *M. truncatula* accessions (Nallu et al., 2014) and possible differences in their expression patterns could be an answer to this. Nevertheless, from the perspective of an individual bacterium, having *hrrP* is an advantage, no matter whether they fix nitrogen or not, since their reproductive capability is less compromised compared to HrrP-lacking strains (Pan and Wang, 2017). Phylogenetic analysis has revealed that some distantly related *Sinorhizobium* isolates contain very similar *hrrP* alleles, while some closely related strains have comparatively divergent *hrrP* sequences. Interestingly, *hrrP* sequences are surrounded by transposable elements

that are likely to be from the genus *Rhizobium* in origin. Moreover, various *Rhizobium* species harbour *hrrP* homologs in their chromosomes. Therefore, it is proposed that *Sinorhizobium* might have acquired *hrrP* from Rhizobium through a relatively recent horizontal gene transfer event, and that then the gene has spread among various *Sinorhizobium* strains (Price et al., 2015), being selected within these strains for its positive benefits for the bacteria.

1.9. Plant regulation of NCRs via binding of unique and conserved motifs in NCR promoter regions

In order to ask what might regulate NCR expression, either temporally or spatially, Nallu et al. (2013) carried out analysis of NCR promoter regions. They identified five conserved motifs (Figure 1.10) ranging from 41 to 50bp in length in many NCR promoter regions that were densely clustered in the 400bp upstream region of putative translation start site. According to the authors, these motifs were not identified in the promoter region of the other 33,131 annotated genes (excluding NCRs) in the *M. truncatula* genome (vMt2.0) or in the promoters, introns or 3'-UTR region of 88 DEFL analyzed by them. Therefore, these motifs seem to be unique to NCR genes and contain binding sites for different transcription factors (TFs), such as, ID1 (a zinc finger TF), AUXIN RESPONSE FACTOR (ARF), DOF protein, MADS box proteins, NICE2 etc. A similar promoter analysis was carried out previously in the Gifford lab (Achom, 2019) with the upstream 500bp region of NCR genes using Mt4.0 genome. The analysis was carried out on 185 NCR gene promoters that were expressed in a microarray experiment conducted using NimbleGen array (only 185 NCRs were found to be expressed since the array did not contain probes for many NCR genes). Six overrepresented motifs (termed as motif 1 to 6; Figure 1.10) were found that overlapped with those mentioned by Nallu et al. (2013) but were comparatively shorter (8-12 bp) and did not match the binding sites for the transcription factors described in Nallu et al. (2013). Instead, binding sites for plant regulatory proteins, LATE ELONGATED HYPOCOTYL (LHY)/CIRCADIAN CLOCK ASSOCIATED 1 (CCA1)/REVEILLE (RVE) genes and Homeobox leucine-zipper protein ATHB15/16 were found within two of those six conserved motifs identified. LHY, CCA1 and RVEs are transcription factors that are homologous to each other and play central roles in the regulation of many genes involved in maintaining plant circadian rhythms.



F

Motif No.	Size of the motif (bp)	No. of NCR promoters with the motif
1	12	146 (79%)
2	12	133 (72%)
3	12	118 (64%)
4	12	149 (81%)
5	11	95 (51%)
6	8	94 (51%)

Figure 1.10: Conserved motifs in NCR promoters. (**A-E**) Five conserved motifs identified by Nallu et al. (2013). The motif analysis information was taken from the supplementary information provided in Nallu et al. (2013). Position specific probability matrix on the top of each figure shows the probability for any nucleotide to be in that position; the nucleotide with highest probability was shown with representative colour (A: red, C: blue, G: yellow, T: green) in the graph and in the top line of multilevel consensus sequence. The middle or bottom lines in the consensus sequence represent other possible nucleotide with respectively lower probability (**F**) Six shorter motifs that were identified by Achom (2019) analyzing 185 NCR promoters (that were expressed in a microarray experiment) are listed in the table and their position is marked by black boxes within the longer motifs found by Nallu et al.

The binding site for LHY/CCA1/RVEs, AGATATTT or its reverse complement, AAATATCT, is known as the Evening Element (EE) and is found in the promoter region of many evening expressed genes (Adams et al., 2018). A motif sequence (denoted as motif-2) that is very similar to EE, was found in 133 (72%) out of 185 promoter analyzed by Achom (2019) (Figure 1.10d & 1.11a). Motif-2 is very similar to EE, except that- firstly, it has two additional highly conserved 'AG' residues at the 5' end; and secondly, some diversification has been seen around a highly conserved 'T' residue which is replaced by a 'C' in about half of the promoters where motif-2 was found. The two flanking 'A' residues were also replaced by 'G' on many occasions (Figure 1.11a). The putative significance of these differences is further discussed in Chapter 7 (General Discussion). Nonetheless, the presence of EE-like or related motif in many NCR promoters suggests the regulation of NCR genes could be, at least in part, regulated by the circadian clock (Figure 1.11b). Since NCRs are almost exclusively expressed in nodules, it can be hypothesised that the plant circadian system plays an important role in regulating nodulation, as it does for many other physiological processes.

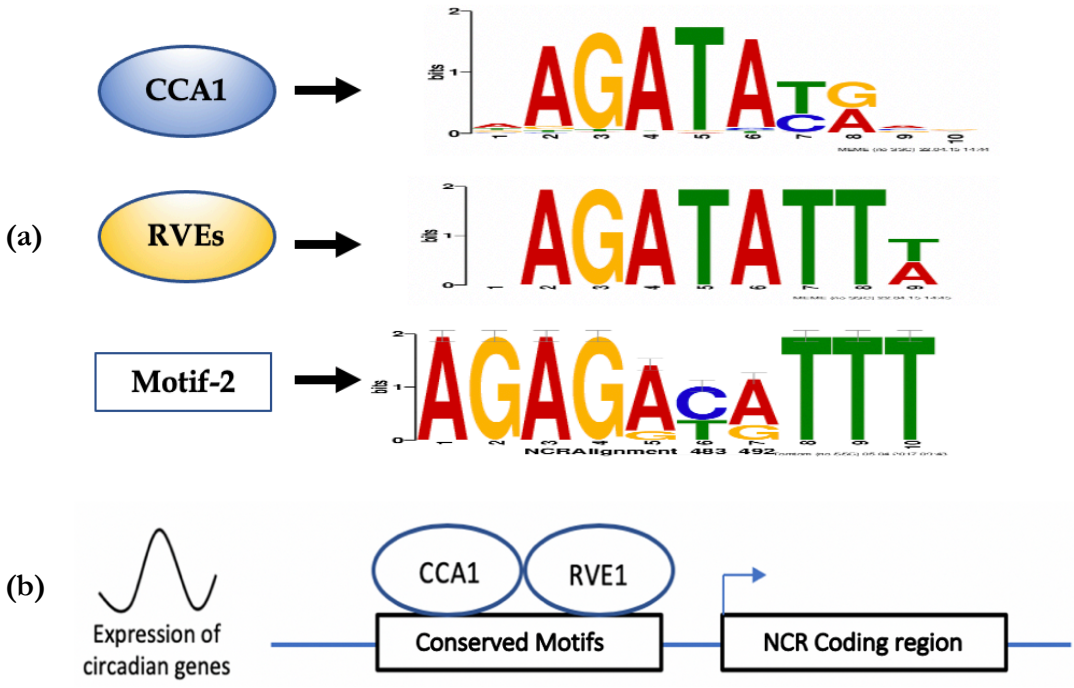


Figure 1.11: Circadian clock associated motifs. (a) The binding site for CCA1 and RVE proteins (retrieved from PlantPAN 2.0); similar motif sequence (Motif-2) was found in the upstream region of NCR genes (Achom, 2019). Motif-2 has two additional highly conserved residues, 'AG' at the 5' end, and some diversification at the middle as described in the text. **(b)** Presence of conserved clock motifs in the promoter region of NCR genes suggests circadian clock-associated or rhythmic regulation for many NCR genes.

1.10. The plant circadian clock

Circadian systems are common endogenous mechanisms that allow living organisms to anticipate and then coordinate themselves with regular, diurnal day-night cycles and related natural changes. The circadian system functions not to merely respond to different environmental changes but it can enable organisms to predict and prepare for them. As sessile organisms, plants need to anticipate and adapt with regular environmental changes more efficiently than animals, and the plant circadian system helps enable this regulation (Sanchez and Kay, 2016). In fact, a myriad of different known and unknown genes (about 30-40% in *A. thaliana*) are already identified that are associated with plant

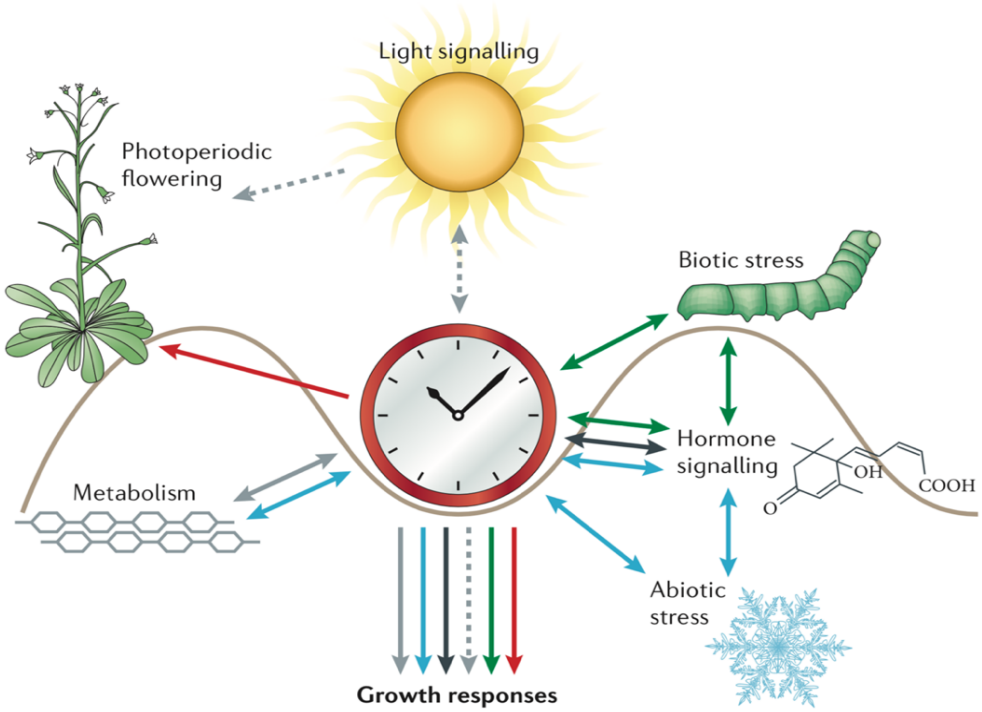


Figure 1.12: Overview of the main processes regulated (directly and indirectly) by the circadian system in plants. Traditionally, it has been thought that the plant clock becomes entrained by receiving environmental stimulus (e.g., light, temperature) and then regulates different output pathways in a unidirectional way. However, recent data suggests that the clock receives feedback from output pathways, and this modifies its sensitivity to entraining stimuli, for example, by regulating the expression of photoreceptors. Different metabolic and hormone signaling pathways, biotic and abiotic stress responses etc. communicate with the central oscillator in a dynamic and bidirectional way (represented by bidirectional arrows), and there is also crosstalk amongst these output pathways (coloured arrows). The circadian clock acts as a hub for all the environmental stimuli and internal feedback and enables coordination of suitable growth and developmental responses. Figure adapted from Greenham and McClung (2015)

circadian clock functions or its regulation (Covington et al., 2008). Together these genes process environmental stimulus or signal (e.g., light) as 'inputs' and modulate a number of plant responses as 'outputs'. The central oscillator also takes feedbacks from output pathways and adjust responses accordingly. In general, the plant circadian clock sits at the centre of plant regulatory network and associated with almost every physiological or metabolic processes, directly or indirectly (Figure 1.12) (Singh and Mas, 2018).

One of the first molecular models of the plant clock was proposed by Alabadi et al. (2001) based on the study on the model plant, *Arabidopsis thaliana* (Figure 1.13). At the centre of this model are two Myb transcription factors, CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) and LATE ELONGATED HYPOCOTYL (LHY) and another transcription factor named TIMING OF CAB EXPRESSION 1 (TOC1) (Lu et

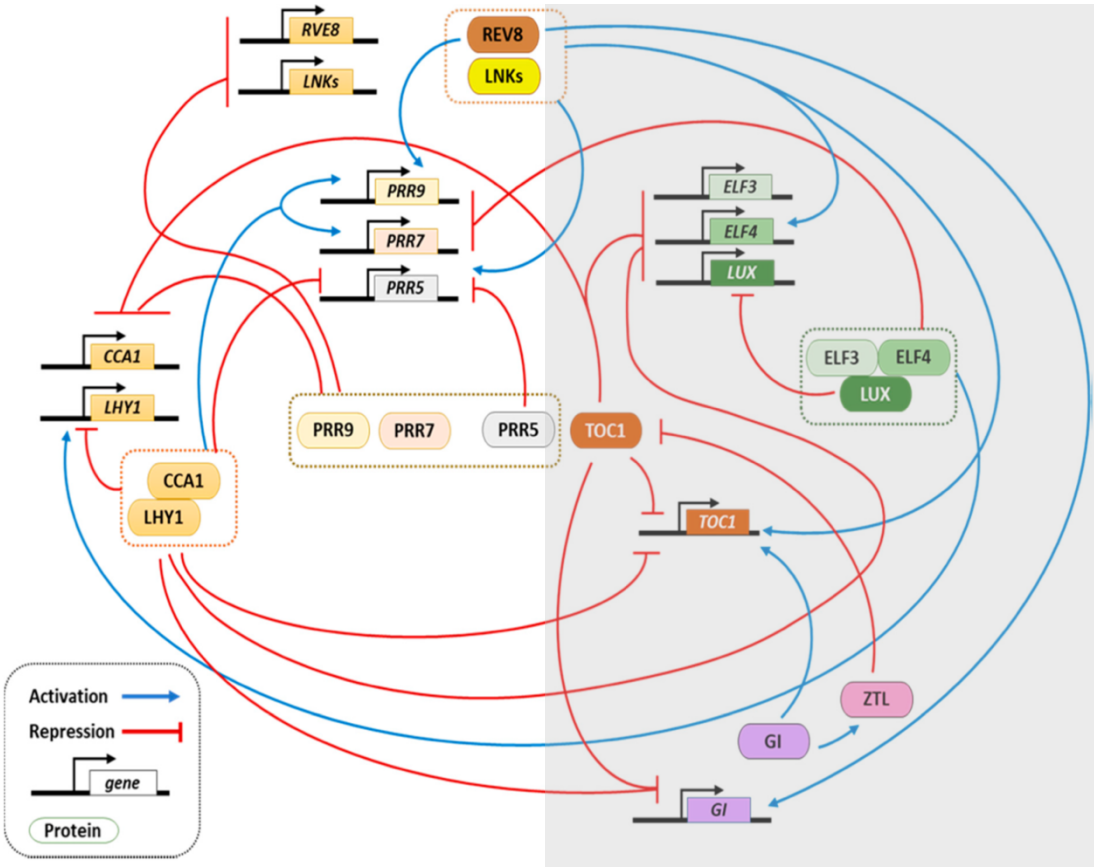


Figure 1.13: Simplistic model of the plant circadian clock in *Arabidopsis thaliana*. Different components are represented here from left to right accordingly based on their expression peak during the day. White and grey backgrounds represent day and night-time respectively. Arrows and bars denote up- and down-regulation, respectively. Figure adapted from Singh and Mas (2018)

.

al., 2009). The expression of *CCA1* and *LHY* peaks in the morning and they mutually repress TOC1. In the evening, when the CCA1 and LHY levels are at a minimum, *TOC1* expression is able to increase. *TOC1* then is able to negatively regulates the transcription of *CCA1* and *LHY* (Gendron et al., 2012, Huang et al., 2012). In addition, both CCA1 and LHY can bind to their own and each other's promoter and down-regulate their transcription (Adams et al., 2015). Together, these three transcription factors form a negative feedback loop. A number of other components are connected to this and form additional loops. Members of the PSEUDO-RESPONSE REGULATOR (PRR) protein family, PRR9, PRR7 and PRR5, are expressed sequentially during the day and downregulate *CCA1* and *LHY* transcription. These proteins are homologous to TOC1 which is also known as PRR1 (Nakamichi et al., 2010).

PRR9 expression is repressed by a protein complex known as the "evening complex" (EC) composed of a Myb-like transcription factor, LUX ARRHYTHMO (LUX) and two nuclear-localised proteins, EARLY FLOWERING 3 (ELF3) and ELF4. As the name suggests, EC expression peaks in the evening and is repressed by CCA1 and LHY. In an opposite fashion, the EC promotes the expression of *CCA1* and *LHY* in an indirect manner (Nusinow et al., 2011, Nagel and Kay, 2012).

Another important protein associated with the plant circadian clock, is a nuclear protein named GIGANTEA (GI) which is associated with diverse plant specific physiological processes, e.g., flowering time regulation, hypocotyl elongation, light signaling and many metabolic processes. *GI* expression peaks around dusk and is repressed by CCA1, LHY and TOC1, however, GI promotes the expression of *CCA1* and *LHY* by a currently unknown mechanism. *GI* might not be considered as a core component of circadian clock, but it possibly acts as a hub that connects the central oscillators to important biochemical and physiological processes (Mishra and Panigrahi, 2015). In recent years, additional clock components and loops have been identified. Among them, a group of morning-expressed genes, known as *REVEILLEs* (*RVEs*) might have important clock functions. As mentioned previously, *RVE* genes are homologous to *CCA1* and *LHY*. However, unlike CCA1 and LHY that are repressors, RVE8 induces some afternoon and evening expressed genes, including *PRR5, TOC1, LUX* and *ELF4* (Hsu et al., 2013, Hsu and Harmer, 2014). In addition, an F-box protein known as ZEITLUPE (ZTL) is essential for sustaining normal circadian period as it

controls the degradation of TOC1. ZTL mRNA is constitutively transcribed but ZTL protein levels oscillate through a mechanism where GI plays important roles (Kim et al., 2007). Another small family of genes were recently described, termed as '*NIGHT LIGHT-INDUCIBLE AND CLOCK REGULATED*' or *LNK*, that are morning-expressed and found to be involved in the control of circadian rhythms, photomorphogenic responses and flowering amongst other processes, possibly by controlling a subset of afternoon/evening peaking clock (e.g., *PRR5, ELF4*) and flowering time genes (e.g., *FLOWERING LOCUS T* (*FT*)) (Rugnone et al., 2013).

1.11. Aims and rationale of this project

Outside of the model plant *A. thaliana*, plant circadian clock functioning has been comparatively less studied. However, homologs of *Arabidopsis* clock genes have been found in other angiosperms, including legumes, e.g., soybean (*Glycine max*), pea (*Pisum sativum*), common bean (*Phaseolus vulgaris*) and *Medicago truncatula*, etc. and shown to have a similar expression pattern compared to *A. thaliana* (Song et al., 2010). Therefore, the function of the plant clock is likely to be widely present and well-conserved amongst legumes. Ueoka-Nakanishi et al. (2012) has shown that the promoter activity of *A. thaliana CCA1* and *PRR5* genes has been retained in heterologous *L. japonicus* tissues. This suggests that not only the coding regions but also the promoter regions of these clock genes might be functionally conserved.

In *A. thaliana*, circadian gene expression varies from organ to organ in terms of period and amplitude. The clock responsiveness to light quality is also different in different organs (Bordage et al., 2016). As far we understand, the leaf circadian system is much more sensitive to external stimuli (e.g., light) than other organs, and plays a governing role on the circadian regulation of other organs. Different plant microbe interactions (both mutualistic and pathogenic) and plant's immunogenic responses were reported to be regulated to some extent by the plant clock (Lu et al., 2017, Roden and Ingle, 2009, Hubbard et al., 2018) (discussed later in Chapter 3 & 5). However, since much of our understanding of the plant clock has been generated from the studies on *A. thaliana* rather than legumes, we have extremely limited understanding of whether the nodule development and functioning or the plant-rhizobia interactions, in general, are influenced by clock, and if that is the case, then, to what extent.

Since almost every physiological process in plants is directly or indirectly regulated to some extent by the clock, it is highly likely that the circadian system also plays a role in controlling nodulation and most likely the activity of the bacteria harboured in nodules. Preliminary supporting evidence in favour of the assumption that the plant clock controls nodule activity comes from the finding that many NCR genes contain clock-associated *cis*-regulatory motifs in their promoter (Achom, 2019).

In this thesis, firstly, a time-course experiment was carried out with a view to investigating the influence of the circadian clock on gene regulation within nodules of the model legume, *M. truncatula*. Secondly, *M. truncatula* lines where expression of the key clock component *LHY* was reduced/knocked out were compared to wild type plants to determine the extent of effect of clock disruption on nodulation, growth, and development. Finally, bacteroid gene expression within nodules (from time-course experiments) were analyzed to ask if there is any (or a set of) bacterial gene(s) whose expression is rhythmic and might be influenced by the circadian rhythm of the plant. In addition, the potential use of two cell-type specific promoters from *A. thaliana* for conducting cell-type specific gene expression studies in *M. truncatula* in future was determined, since studying cell type gene expression can help to elucidate developmental regulatory mechanisms. The experiments, results and findings are presented in four results chapters in this thesis as outlined below.

In Chapter 3, findings are presented and discussed from a time-course experiment where mature nodules were collected at different times during the day and analyzed to ask if there is nodule-localised rhythmic gene expression. Using RNA sequencing and functional characterization approaches, a set of genes including several NCRs that are rhythmically expressed, were determined. Analyzing the expression pattern of different rhythmic genes also provided valuable insights into which different physiological and metabolic processes are enhanced or repressed at specific periods during the day or night.

In Chapter 4, the phenotype (shoot growth, leaf movement, nodule mass) of the two *M. truncatula lhy* mutant lines was evaluated and discussed. These findings show that nodulation and growth-related processes in the nodule are impaired when clock functioning is disrupted.

Chapter 5 provides insights into the rhythmicity of rhizobial gene expression in the bacteroids inside nodules. A number of bacterial genes have been identified that are likely to be diurnally regulated. These findings open up future studies to better understand clock regulation in prokaryotic systems, particularly those that are in a mutualistic relationship with a more complex multicellular organism.

In Chapter 6, two *A. thaliana* promoters that confer cell type specific expression in *A. thaliana* were transformed into *M. truncatula* in order to generate tools to study cell-type specific gene expression. The results show that these Arabidopsis promoters can direct gene expression in the cortical cells of *M. truncatula.* Therefore, these constructs could be used to generate stable transgenic lines harbouring a cortex specific marker gene (e.g., fluorescent protein) expression, enabling gene expression to be studied in cortical cells- the location from which most nodule tissue arises.

Analysis of the combination of transcriptomic and phenotypic data that are presented in this thesis supports that, like many other physiological processes, nodulation and rhizobial activity is also regulated by the plant's endogenous circadian clock. This new data and related insights gained from it will enable further studies to better understand the influence of circadian clock in legume-rhizobia symbiosis.

2. Materials and Methods

2.1. Plant growth and sampling

2.1.1. Plant material

Medicago truncatula cv. Jemalong A17 seeds were obtained from the Genetic Resources Unit, IGER, Aberystwyth (www.igergru.ibers.aber.ac.uk). Wild type *M. truncatula* R108 and two Tnt1 retrotransposon insertion lines, NF17115 and NF16461 (Mt*lhy1* and Mt*lhy2*, respectively) in the R108 background were procured from the Noble Research Institute, Ardmore, Oklahoma, USA. NF17115 and NF16461 was chosen as they harbour Tnt1 insertion in the promoter and 5th exon of Mt*LHY* (along with insertions at other random positions), respectively.

2.1.2. Seed germination

All the different lines of *M. truncatula* seeds used (A17, R108 and the two transposon insertion lines mt*lhy1* and mt*lhy2*) were germinated in the same way. To begin, the seeds were extracted by crushing the seed pods carefully with a plasterer's hawk on a corrugated rubber mat. Seeds that came out of the pods without any apparent damage were picked by tweezers into a 50mL falcon tube. Seeds were then chemically scarified by treating with concentrated sulfuric acid (~1-2 mL per 100 seeds), usually for 10-20 minutes, until 2-3 dark spots appeared on the surface of most seeds. After removing excess sulfuric acid, seeds were thoroughly washed 3-5 times with sterile water. Scarified seeds were sterilized by treating with 7% sodium hypochlorite solution for 5 minutes with gentle shaking. After sterilisation, seeds were sown one by one on 1.5% phyto-agar in square (12 x 12cm) petri dishes and then the plates were sealed using 3M Micropore^{TM} tape. About 36 seeds (6 x 6) were sown per plate and a wet (with sterile dd H_2O) seed germination/growth pouch was kept inside the petri dish to maintain humidity. Petri dishes containing scarified and sterilised seeds were then wrapped with aluminium foil (to keep seeds in dark) and kept in 4°C. After 3-6 days of cold treatment, wrapped petri dishes were transferred into a Sanyo MLR-352 growth chamber set at 25°C. At this temperature, seeds are usually germinated and become ready for sowing or transformation within 2-3 days.

2.1.3. Growth conditions

Germinated seedlings that were healthy (actively growing with green cotyledons) and more than 2 cm long were planted on FP11 (11 x 11 cm) pots, filled with sterilised perlite with 1-2cm top layer of sterile vermiculite. For time series experiments, plants were grown for 40 days after potting in Sanyo 2279 growth cabinet in 12/12 hours light/dark photoperiod at an irradiance of 200 μmol m^{-2} s^{-1} and temperature of 24°C during day and 21°C at night. For mutant phenotyping experiments, FP9 (9 x 9 cm) pots were used, and plants were grown for 5 weeks at 24/21°C (16/8 hours day/night cycle) in a temperature-controlled glasshouse compartment. For rhythmic leaf movement (RLM) assays, plants were grown for 2 weeks only in a Sanyo MLR-352 cabinet in 12/12 hours day/night photoperiod at 24 °C before transferring them to constant light. In all the cases, pots were kept on suitable trays and watered 2-3 times per week when the surface of the pot medium was dry, with modified Broughton and Dilworth (1971) nutrient solution (1 mM $CaCl_2 \cdot 2H_2O$, 1 mM KH_2PO_4, 75 μM FeNa EDTA, 1 mM $MgSO_4.7H_2O$, 0.25 mM K_2SO_4, 6 μM $MnSO_4.H_2O$, 1 μM $ZnSO_4.7H_2O$, 0.5 μM $CuSO_4.5H_2O$, 0.05 μM $CoSO_4.7H_2O$, 20 μM H_3BO_3, and 0.1 μM $Na_2MoO_4.2H_2O$), adjusted to pH ~6.5 with KOH.

To prepare seedlings for 'hairy root' transformation, the *Agrobacterium rhizogenes* inoculated seedlings were grown in square petri dishes (12 x 12cm) on 1.5% phytoagar-solidified Modified Fahreus Media (MFM) (0.5 mM $MgSO_4.7H_20$, 0.7 mM KH_2PO_4, 0.8 mM $Na_2HPO_4.2H_2O$, 5 mM NH_4NO_3, 20 μM ferric citrate, 8 μM $MnSO_4.H_2O$, 4 μM $CuSO_4.5H_2O$, 7.34 μM $ZnSO_4.7H_2O$, 16 μM H_3BO_3, 4.13 μM Na_2MoO_4 and 1 mM $CaCl_2$; pH 6.5) for 7 days at 25°C. Ten seedlings were grown side by side in each plate. Plates were wrapped with black polythene so that the cotyledons were able to access light, but the radicles were not, mimicking the natural above- and below-ground light conditions.

2.1.4. Root cross section preparation

Fresh root parts of interest were cut (in about 5mm pieces) using a clean razor blade. At around same time, a 5% agar suspension was prepared and poured onto a petri dish. When the temperature of the agar suspension went down to ~58°C, cut root pieces were dipped vertically into the suspension. The suspension was then placed at 4°C to rapidly solidify it. Then, using a sharp scalpel, around 1 cm^2 agar blocks were cut around

each immobilized root section. It was ensured that the root sections remain at the centre of the blocks. Next, using a vibratome (Series 1000 classic), 100-200 μm thick sections of agar block were sliced. Root cross-sections within each agar slice were viewed under a light microscope to select the best cross-sections for confocal microscopy.

2.1.5. Time course sampling and plant phenotyping

In the time-course experiment, 6-7 plants were uprooted from pots for sample collection at each time-point. All the nodules that were fully grown (>3mm in length), actively fixing nitrogen (pink in colour) were picked from roots using fine-tipped tweezers and collected into 2mL sterile, RNase free tubes that were suspended in liquid-N. Leaves (4-5 trifoliates) and part of the root system without any nodules, were also collected in separate 2mL tubes in liquid-N. Samples were then kept in -80˚C for RNA extraction.

For mutant phenotyping experiments, plants were removed from pots and the fresh shoot weight of each plant was measured. Next, they were transferred to a 60˚C oven and left overnight; shoot dry weight was measured the next day. From the root system the nodules were picked using tweezers and fresh nodule weight was measured collectively per genotype/treatment.

2.2. Bacterial culture and plant inoculation

2.2.1. Rhizobial inoculation

The highly efficient N-fixing rhizobial symbiont for *M. truncatula*, *Sinorhizobium meliloti* 1022 was grown on agar with TY media (tryptone 5g/L, yeast extract 3g/L, $CaCl_2 \cdot 2H_2O$ 0.87g/L, pH adjusted to 6.8-7.0) at 28˚C for 2 days. From the solid culture, *S. meliloti* 1022 was inoculated into 10 mL liquid TY medium and grown for ~24 hours with gentle shaking at 28˚C. Rhizobial cells were then harvested by centrifuging at 3200g for 10 minutes. The supernatant was discarded, and the pellet resuspended in sterile water and diluted to an optical density (OD_{600}) of ~0.05. 500 μL of freshly prepared rhizobial solution was used to inoculate each *M. truncatula* seedling the day after potting by pipetting onto the vermiculite layer in close proximity to the seedlings.

2.2.2. Hairy root transformation with *A. rhizogenes*

First, *A. rhizogenes* strain ARqua1 was transformed with pre-prepared expression vectors (see 2.3.1) through the electroporation method (Andreason, 1993) using Bio-Rad MicroPulser electroporator and 0.1 cm-gap electroporation cuvettes. Then, transformed *A. rhizogenes* was grown in 10 mL of liquid LB media (Trypton 10 g/L, Yeast extract 5 g/L, NaCl 3 g/L and pH adjusted to 7.5, with 50 µg/mL streptomycin for selection) overnight at 28˚C. Next, 250 µL of overnight grown culture was spread on solid LB agar media with 50 µg/mL streptomycin and grown for 2 days at 28 ˚C, so that a homogenous layer of bacteria was formed. To transform plants with *A. rhizogenes*, ~3 mm of the root tips of 3 days old seedlings were cut with a sterile scalpel, and then the cut-roots were rubbed against the layer of bacteria so that a visible tiny ball of culture became attached to the cut-tips. *A. rhizogenes* treated seedling were then grown for a week as described earlier in section 2.1.3.

2.3. Molecular Methods

2.3.1. Cloning and preparation of expression vectors

For expression vector preparation, TOPO® and Gateway® cloning methods were followed. First, the desired promoter regions (pAtCO2 and pAtPEP) were amplified through Polymerase Chain Reaction (PCR) from the genomic DNA of *Arabidopsis thaliana*, using high fidelity Phusion® DNA polymerase and primers (listed in Table 2.1). Genomic DNA was extracted from *A. thaliana* ecotype Columbia-0 using Chelex® 100 resin (from Sigma-Aldrich) and following the protocol developed by HwangBo et al.

Primer Name	Sequences 5'-3'	Tm (˚C) **
AtCO2_F	caccTAACTCCATTATTTACGACTGTGC *	66.5
AtCO2_R	AACTCTTGTTGCATTATTGTCAAATC	62.1
AtPEP_F	cacCACTTCATTCGATGTTCACC *	64.8
AtPEP_R	GGTTTTGGCTAATGTGATTGTGTAG	63.5

* Bases in lowercase are 5' overhang added to the forward primers as per TOPO® cloning manual
** Tm are calculated for Phusion polymerase and 0.5 µM primer conc. using Tm calculator at Thermo Fisher's website

Table 2.1: List of PCR primers used to amplify promoter sequences (pAtCO2 and pAtPEP).

(2010). Amplified PCR products were further purified using the QIAquick PCR purification kit. The PCR products were then cloned into pENTR (entry vector) and transformed into One Shot® chemically competent *Escherichia coli*. Transformed *E. coli* were then grown in liquid LB media overnight at 37°C with 25 µg/mL kanamycin for selection. Next, the pENTR vector was extracted from 3 mL of overnight grown culture using the QIAprep spin miniprep kit as manufacturer's instructions. In the next step, an LR reaction was performed to incorporate the cloned promoter fragment into the Gateway-compatible pBGWFS7.0 destination vector that has both streptomycin and spectinomycin resistance and recombination insertion site immediately upstream of a eGFP-GUS coding sequence. For the LR reaction, 25 ng of entry vector was mixed with 50 ng of destination vector and ddH_2O added to make the final volume up to 3.5 µL. Then, 0.5 µL of LR Clonase™ II enzyme (Invitrogen™) was added and incubated for 1 hour at 25°C. After incubation, the total reaction mixture (~4 µL) was used to transform competent *E. coli* that were then grown on LB media with both streptomycin and spectinomycin at 50 µg/mL concentration for the selection of colonies. To confirm whether the promoter sequences were properly integrated into the vector in terms of orientation and position, the vectors were extracted from transformed *E. coli* and evaluated by analyzing Sanger sequencing reads (prepared through GATC sequencing service). For plant transformation, the *A. rhizogenes* ARqua1 strain was transformed with the vectors (as mentioned in section 2.2.2). Plasmid vectors were kept at -20°C, or for long time storage vectors containing *E. coli* were kept as glycerol stocks (150 µL overnight grown culture in 850 µL 15% v/v glycerol) at -80°C.

2.3.2. RNA extraction and sequencing

Frozen plant material was first ground using a mortar and pestle. Before use the mortar, pestle, and other tools (inc. spatula) were autoclaved and washed with RNase AWAY™ and placed in liquid-N during grinding to keep the temperature as low as possible. RNA extraction was carried out from ~100 mg of ground sample using the Monarch® Total RNA Miniprep Kit that included a DNase treatment step. Extracted RNA samples were then analyzed for their concentration and quality using readings obtained on a NanoDrop spectrophotometer. The minimal quantity and quality threshold considered for samples was 50 ng/µL of RNA with both A260/280 and A260/230 ratios above 2.0. Samples were also analyzed for their RNA integrity using a Bioanalyzer 2100

RNA 6000 Pico Total RNA kit (Agilent Technologies). A minimum RNA Integrity Number (RIN) of 8.5 were maintained for all the samples to confirm good RNA integrity. RNA samples were then sent to Novogene (UK) company ltd. for RNA library preparation and sequencing. First, the rRNA was removed using a Ribo-Zero kit, then the libraries prepared following an Illumina TruSeq^{TM} RNA library preparation protocol, both according to manufacturer's instructions. Once the library had been prepared and passed the quality control tests (to confirm appropriate library concentration, insert size), it was Ilumina-sequenced for unstranded, 150 bp-long, paired-end (PE) sequencing.

2.3.3. cDNA synthesis and quantitative PCR (qPCR)

cDNA was synthesized using ~600 ng of total RNA and ProtoScript II cDNA synthesis kit and random primer mix from New England Biolabs following the manufacturer's instructions. cDNA was stored at -20˚C until used for qPCR. Primers (listed in Table 2.2) for qPCR were designed using Primer3Plus (Steve Rozen, 2000) with the following parameters: primer melting temperature between 59 and 64 ℃ (60 ℃ optimum), product size 50- 200 bp and a GC content between 45-55% (50% optimum). Primer efficiency was evaluated by using linear calibration curve and melting curve (also known as dissociation curve) analysis. Linear calibration curves were obtained from serial dilutions (10^{-1} to 10^{-6}) of cDNAs in equal fraction. Linear regression analysis of Ct versus log(dilution) was used to derive the efficiency as, $m = -(1/\log E)$, where m is the slope of the regression line fit derived from standard curve and E is the efficiency of the primer (Schmittgen and Livak, 2008). The efficiency for all the qPCR primers (listed in Table 2.2) was over 1.85 and a single peak was seen in all the melt-curves, suggesting good efficiency and specificity of the primers.

qPCR was performed in 96 well plates using SYBR Green JumpStart Taq ReadyMix (Sigma-Aldrich) as manufacturer's instructions. A reaction mixture of 60 µL (containing 5 µL of cDNA) was prepared for each sample and then split and loaded into 3 separate wells (each containing 20 µl) to make three technical replicates to account for technical errors. Thermocycler conditions were as follows: 95 ˚C for the initial 10 min; 95 ˚C for 30 sec then 60 ˚C for 1 min for 40 cycles; and finally one cycle (to generate measurements to plot a melting curve) at 95 ˚C for 1 min, 55 ˚C for 30 sec and 95 ˚C for 30 sec. Expression of genes of interest was normalized against the reference gene, β-Tubulin

(Medtr7g089120) (Kakar et al., 2008) using the ΔC_t method, a derivation of the $\Delta\Delta C_t$ (Livak and Schmittgen, 2001).

Target Gene	Sequences 5'-3'	Tm (˚C) *
β Tubulin	TTTGCTCCTCTTACATCCCGTG	63.6
Medtr7g089120	GCAGCACACATCATGTTTTTGG	63.7
LHY (2nd & 3rd exon)	CACAAAACAAAGAGAACGATGG	59.6
Medtr7g118330	ATGGCTCCTGATTTGCACAG	61.2
LHY (5th & 6th exon)	ACCTGATGCAGACCATTTAGC	59.2
Medtr7g118330	AAGGCAAGTCGTCCCTCTTC	60.8
ELF4	AAGACACCAATCACCGTTCC	59.8
Medtr3g070490	TTAAGGGTCTGCCATGCTTC	60.2
PRR5c	CTGGCAGCTCATGCAAACAAG	60.7
Medtr8g024260	ACTGCCATCGTACCGAGAAAAG	60.7
PRR7	CAAACCTATTCGGAGGAACG	59.6
Medtr4g061360	CGGGTGTTTGTTCCACTTTC	60.4

Table 2.2: List of primers used for qPCR to quantify the relative abundance of different circadian gene transcripts.

2.3.4. Confocal microscopy

Root cross-sections were placed onto a microscope slide. Then, 2-3 drops of the cell wall-staining dye, propidium iodide (100 μg/mL) were applied onto the cross-sections and then covered with coverslips and taken to the microscope. Confocal microscopy was carried out using a Zeiss LSM 880 microscope. To detect GFP, the 488 nm laser was used, and fluorescence emission was visualised in the 500-550 nm range. For propidium iodide/cell wall visualisation, the same laser was used but fluorescence was visualized at the 600-700 nm range.

2.4. Bioinformatics and statistics

2.4.1. RNA sequence data analysis

Raw sequencing data was received from Novogene in the form of a pair of .fq.gz files for each sample. The depth of sequencing was at least 34 million reads per sample,

containing reads from both plant and rhizobial sources. The .fq.gz files were uploaded onto the galaxy EU server (usegalaxy.eu). Different freely accessible software packages, available on the galaxy server, were used for most of the data processing. First, the quality of raw sequencing data was analyzed by FastQC (Andrews, 2010). The replicate-2 for 21h time-point and replicate-3 for 15h time-point contained excessively high level of duplicated reads that were also inconsistent with other samples/replicates; suggesting something went wrong during library preparation. These data were not trustable and could not be repaired using available techniques, therefore, discarded from further processing and analysis. Next, contaminating adapter and poor quality sequences (based on phred33 quality scores) were removed by Trimmomatic v36.4 (Bolger et al., 2014) using a custom made fasta file containing Illumina adaptor sequences (adaptors.fa) and executing following settings: illuminaclip:adaptors.fa:2:30:10:2, slidingwindow:4:20 and minlen:40. Next, clean, trimmed and paired reads were used to generate raw transcript read-counts and TPM (transcripts per million) normalized read-counts using Salmon quant v0.14.1 (Patro et al., 2017) and using the *M. truncatula* reference transcript sequences (Mt4.0v1), downloaded from Phytozome database (phytozome.jgi.doe.gov). Read counts were further normalized as log2 transcripts per million (logTPM) using functions in Excel.

To determine rhizobial transcript read-counts, clean reads (trimmed to remove adaptor or poor-quality sequences) were mapped to the *S. meliloti* 1022 genome (downloaded from NCBI database; accession: PRJNA636618) using Hisat2 v2.1.0 (Kim et al., 2015a), using the default parameters to generate BAM alignment files. Raw gene read counts were generated from BAM files using FeatureCounts v1.6.4 (Liao et al., 2013) with gene annotations (also retrieved from NCBI assembly database; accession: ASM1331577v1) and the following settings: feature type (-t): gene, gene identifier (-g): gene_id and all other parameters set to default. Raw counts were normalized as log2 transcripts per million (logTPM) values using Excel functions.

To identify diurnally oscillating *M. truncatula* or *S. meliloti* transcripts, logTPM expression data were analyzed with the R MetaCycle v1.2.0 package in R (v3.6.1) (R-Core-Team, 2018), with following settings: minper: 20, maxper: 28, cycMethod: LS (Lomb-Scargle).

2.4.2. Clustering, gene ontology and promoter analysis

The expression levels of oscillating transcripts with a *p*-value less than 0.05 were clustered by first importing the mean logTPM expression values from three replicates into R. Then the expression values for each gene in different time-points were Z-normalized (mean substruction followed by division with standard deviation). For clustering and to generate cluster related heatmap and plots, the following R packages were used: "pheatmap", "factoextra", "ggplot2", "ggplots", "RColorBrewer".

For plant gene ontology (GO) analysis, both Phytomine (available online at: phytozome.jgi.doe.gov/phytomine/begin.do) and ShinyGO V0.61 tools (available at: bioinformatics.sdstate.edu/go/) were used and the outputs were compared. Clustering of top hit (top 10 or 20 hits based on FDR corrected p-values) GO terms was carried out using the ShinyGO 'Tree' tool. For rhizobial GO analysis, the R package "topGO" was used. The word frequency analysis was carried out using the R packages: "tm", "SnowballC", "wordcloud". For all R packages mentioned above, the latest available versions were used.

To identify specific motifs in the gene promoters of different cluster, first, the promoter sequences of corresponding genes were extracted from the reference genome (Mt4.0v1) using the following galaxy server bedtools: "FlankBed", GetFastaBed". Then, the presence of query motifs was identified using FIMO (Find Individual Motif Occurrence) tool from MEME suite (Bailey et al., 2009) (available online at: meme-suite.org/tools/fimo).

2.4.3. Reciprocal BLASTp and homology search

M. truncatula clock gene homologues were primarily identified by carrying out a reciprocal BLASTp search against the NCBI *M. truncatula* non-redundant (nr) protein sequence database using *A. thaliana* clock protein sequences. Hits were sorted using maximum bit-score followed by expect (E) value. This homologous gene list was then compared with the homology data in the KEGG (Kyoto Encyclopedia of Genes and Genomes) Orthology database and only those hits that were also listed in KEGG database were retained for further analysis. Homology was also analyzed using the Phytozome

database (available at: phytozome.jgi.doe.gov) to detect the level of similarity between homologous protein sequences using the Smith-Watermann alignment algorithm (Smith and Waterman, 1981).

2.4.4. Statistical analysis of phenotypic data

Phenotypic data were analyzed in excel using 'Data analysis' add-in options. Firstly, the descriptive statistics (e.g., mean, variances, SEM) of each group of samples were determined. Depending on the level of variance (determined by F-test; assuming equal variance when F-test result is insignificant and vice versa), two sample t-tests assuming equal or unequal variances were performed to determine if there was any statistical significance between samples. Data boxplots were also generated using excel chart functions.

2.4.5. Statistical analysis of rhythmic leaf movement assay data

For RLM assays, 2 weeks old plants were moved to constant light conditions at the same irradiance as their normal daylight, and then imaged periodically using time-series cameras every ten minutes over the course of a week. Using ImageJ, images were grey scaled, the areas around the full extent of growth of the second true leaves were defined and integrated density (product of area and mean grey value) of each leaf were calculated per frame. After normalising the values (to their mean over the time-points) they were analyzed evaluated for periodicity using BioDare2 (https://biodare2.ed.ac.uk) and rhythmic leaf movement visualised graphically.

3. Rhythmic gene expression within nodules supports the presence of a functional nodule clock involved in diurnal coordination of nodule metabolism

3.1. Introduction

The importance of an endogenous circadian clock in plant growth and development is now well-established and documented by the numerous studies based on the model plant *A. thaliana*. Many core clock genes that encode for transcription factors target numerous other genes with distinctive functions (Kim et al., 2017). In *A. thaliana*, as many as one-third of all expressed genes were reported to be under clock regulation (Michael and McClung, 2003, Covington et al., 2008), and as much as 89% of the transcripts were found to be rhythmically regulated under different environmental cycling conditions (e.g., light-dark or warm-cool) (Michael et al., 2008b). Clock-regulated genes were also reported to be over-represented in all major hormone signaling and several stress response pathways (Covington et al., 2008). Recent studies also link circadian regulation to plant health and innate immunity (reviewed by Lu et al. (2017)). Immuno-regulation plays an important role in controlling susceptibility to different pathogens and in allowing beneficial microbes (e.g., rhizobia) to colonize (Pel and Pieterse, 2013, Zamioudis and Pieterse, 2012). Therefore, it is likely that the circadian clock regulates nodulation in the legumes, at least to some extent, by modifying both the plant's immunogenic response towards infecting rhizobia and the extent to which plant root development can be modulated to activate nodule formation.

3.1.1. Regulation of plant immunity by the circadian clock

Light has long been known to influence different plant defense mechanisms including Systemic Acquired Resistance (SAR) and the Hypersensitive Reactions (HR). Several photoreceptors and photosynthetic genes are also found to be part of the plant immune system (Hua, 2013, Karpinski et al., 2003). Since light is the major zeitgeber (an external cue that can entrain the biological clock) for the circadian system and given that most defense assays were carried out under diurnal light-dark conditions, it can be difficult to differentiate among defense responses that were triggered not by light but by the endogenous clock. However, a growing body of evidence supports that not only light, but

the circadian clock independently coordinates different defense responses in a diurnally regulated manner, enabling the anticipation of regular attacks from different pest and pathogens. For instance, Bhardwaj et al. (2011) inoculated *A. thaliana* leaves with the gram-negative bacterium *Pseudomonas syringae* at different times during the days. The plants were initially grown in 16h/8h light/dark cycles for 3 weeks and then moved to continuous light conditions and kept there for a day and then infections were carried out at different time-points. The researchers observed that the susceptibility of the plant to *P. syringae* varies temporally during the day, being relatively highly resistant to infection at around the subjective morning and most susceptible around the subjective midnight. They also found that this temporal variation in resistance was disrupted in two different clock mutants (*cca1* and *elf3*) that display arrhythmic circadian outputs. Several other studies in different clock mutants also supports the requirement of proper clock functioning in plant immunity (Lu et al., 2017).

One of the most straightforward ways to restrict pathogenic invasion might be by regulating stomatal opening. Stomatal opening act as a gateway for some pathogens to gain access to plant tissues, nutrients, and space for colonization (Melotto et al., 2006). Plant stomata remain mostly open during the day and closed at night, a process which is tightly controlled by the circadian clock in anticipation of the daily changes in light and moisture content in the air. This is consistent with the finding that *A. thaliana* plants that are spray-inoculated (not through injury) with *P. syringae* exhibit greater resistance during night when stomata remain mostly closed (Korneli et al., 2014).

Apart from regulating physical barriers like stomata, the plant circadian clock can activate induced defense pathways (that is usually activated upon an injury or pathogenic attack) to restrict pathogens. A number of defense-related genes oscillate in a circadian manner and have been found to be regulatory targets of the core clock genes, e.g., *CCA1, TOC1, PRRs.* Many of these target genes are involved in two major plant immune response pathways, PTI (PAMP-Triggered Immunity) and ETI (Effector-Triggered Immunity) (Lu et al., 2017)). Increasing numbers of studies have shown that the major plant hormones (e.g., auxins, cytokinins, gibberellins, abscisic acids, ethylene) are all involved in the regulation of plant-pathogen interactions. There is crosstalk between all these hormonal signaling pathways and circadian regulation (Atamian and Harmer, 2016,

Robert-Seilaniantz et al., 2011). The biosynthesis and regulation of two important, well-known, counteracting plant defense hormones, SA (Salicylic Acid) and JA (Jasmonic Acid), are also influenced by the circadian clock (Goodspeed et al., 2012, Goodspeed et al., 2013). Furthermore, the production and homeostasis of ROS (Reactive Oxygen Species) are regulated in a diurnal manner (Lai et al., 2012). Considering all these, the influence of circadian regulation in plant immunity is clear, although, the detailed mechanism of defense-related regulation of the clock is yet far from understood.

3.1.2. The plant circadian clock could also be modified by invading pathogens

According to our understanding of pathogen-plant interactions, pathogen success involves regulation of plant pathways to enable their colonisation. One means to achieve successful infection could be for the pathogen to modulate or disrupt the circadian regulation in their favour. To support this, infection with *P. syringae* can shorten the circadian period, as measured by promoter activity of the *CCA1* gene (Zhang C et al., 2013). Infection by the oomycete *Hyaloperonospora arabidopsidis* (*Hpa*) can disrupt the rhythmicity of *CCA1* expression (Wang et al., 2011). While pathogen mediated changes in the clock activity are common, these changes are also associated with programmed cell death of the infected tissue. Therefore, it is not clear whether the altered clock activity is orchestrated by the pathogen, or it just an indirect effect of the molecular measures employed by the host to restrict infection spreading. So far there is no conclusive evidence for pathogen-mediated direct alteration of the circadian clock, however, some pathogens might have indirect means of modifying the clock. For example, in *A. thaliana*, *P. syringae* infection can activate the synthesis and/or signaling of auxin and abscisic acids. These hormones are tightly regulated by the clock but their activity can also reciprocally modify clock activity. In fact, as discussed in Chapter 1, the central oscillator takes feedback from output pathways to adjust itself. Therefore, any changes in hormonal activity or other physiological modifications caused by pathogens, can hypothetically alter circadian-related activities.

3.1.3. Nodulation in legumes could be under circadian influence

Circadian regulation is so multifaceted that every aspects of plant physiology is, directly or indirectly, under this influence. It is, therefore, very unlikely that nodule formation process in legume plants would not be affected or controlled to some extent

by the clock. Some support in this regard comes from the study by Dalla Via et al. (2015) where they found some clock-related genes including several *PRR* genes were differentially expressed in common bean (*Phaseolus vulgaris*) upon Nod Factor and EPS (Exopolysaccharide) perception. This suggests that a root circadian clock could be adjusted at the very early stages of symbiotic interaction. It has been reported long ago (Minchin and Pate, 1974) that the nitrogen fixation rate in the nodules of *Pisum sativum* is responsive to the daily changes in light and temperature. It usually increases throughout the day, reaching a maximum in the evening and then decreases at night. In soybean (*Glycine max*), an important transcription factor for nodule growth and fitness, known as bHLHm1 (basic Helix-Loop-Helix membrane-1) or previously known as SAT1 (Symbiotic Ammonium Transporter 1), has been found to be diurnally regulated with higher expression during the night (Chiasson et al., 2014). These observations indeed support a link between the circadian clock and the nodulation process. As described earlier, the clock regulates various aspects of the plant immune response towards pathogenic invaders. This is interesting in regard to nodulation since one hypothesis suggests that the symbiotic interactions originate from pathogenic interactions (reviewed in Lagunas et al. (2015)). Like pathogens, initial perception of beneficial microbes triggers plant immunity, but beneficial ones usually downregulate immunogenic responses at later stages (Pel and Pieterse, 2013, Zamioudis and Pieterse, 2012). Sometimes, specific defense response can also inhibit usual rhizobial symbiosis. For example, Lopez-Gomez et al. (2011) has demonstrated that administration of a Microbe-Associated Molecular Pattern (MAMP), known as 'flg22', can induce immune responses that inhibit rhizobial symbiosis in *L. japonicus*. The key point here is that both the pathogenic and symbiotic interactions can activate specific immunogenic responses and, in turn, both interactions can be shaped and modified by the host immune system. Now, if pathogenic interactions are adjusted by the endogenous clock (as discussed in previous section), then it is highly likely that the symbiotic responses are also conditioned by it.

3.1.4. Aims and objectives

Our understanding of circadian regulation is far from comprehensive, but outside *A. thaliana*, it is even less developed. However, important clock genes are conserved in both monocots and dicots including legumes and have been reported to be associated with critical agronomic traits such as growth and flowering time (Song et al., 2010, Bendix et al., 2015, Li et al., 2019). In particular, clock gene homologs (e.g., *MtLHY*, *MtPRR*s, *MtELF*s) were already identified in the model legume, *M. truncatula*, that could be very useful in studying the circadian regulation of legume-rhizobia symbiosis (Young et al., 2011, Song et al., 2010). As mentioned earlier in Chapter 1, not only the coding regions but the promoter activity (and thus regulation) of these circadian genes might also be conserved (Ueoka-Nakanishi et al., 2012). The rhythmic behaviour of the clock genes are slightly different in different tissues and cell layers in terms of their period, amplitude, responsiveness to light quality etc, therefore, the clock might have become adapted to the functions of different tissues (Bordage et al., 2016). It has been shown that root circadian rhythms are regulated by shoot or shoot-meristem derived signals (Takahashi et al., 2015). Since nodules are connected to roots, if there is a functioning clock in nodules then that could be, to some extent, regulated by root clock or distantly by shoot clock. Therefore, it would be interesting to look at the regulation of clock genes in nodules.

The clock influences almost all the physiological processes investigated to date and likely to play roles in nodulation as well. However, there is a lack of information about the role of circadian clock during legume-rhizobia symbiotic interactions. As discussed in Chapter 1, many nodule specific NCR genes contain Evening Element (EE), a clock-associated *cis*-regulatory motif, in their promoters which implies a possible link between nodulation and circadian regulation. Different roles of these NCR genes are now evident, but they were first characterized for their antibacterial activities (Mergaert et al., 2003). Since many defense related genes are regulated in a diel manner, some of these NCR genes could be rhythmically expressed to keep the infecting population of rhizobia or pathogens in balance, in tune with other plant growth and development processes that occur over the day and night.

To the best of our knowledge, clock functioning in the nodules of legumes has not been studied extensively and there is no known study that analyzes the gene expression

profile of legume nodules over a period of 24 hours. The paucity of nodule-specific clock related gene expression data, together with our observation that many NCR gene promoters contain the EE motif, led us to test for circadian regulation of gene expression in nodules. In the next section, the experimental design of a time-series experiment and the associated results are presented and discussed. The experiment enabled us to find out the genes that are expressed in a diurnal pattern under constant environmental conditions and to get a tentative idea on how the expression of different genes could be influenced by the circadian clock in mature nodules of *M. truncatula.* Special attention was given to determine, firstly, the expression pattern of the key circadian clock genes, secondly, the overall pattern of rhythmic gene expressions or in other words identifying processes that are rhythmically regulated, and finally, to find out whether any NCR genes were involved in rhythmic regulation.

3.2. Results and Discussion

3.2.1. Experimental overview of the time-course analysis

In order to determine how gene expression in nodules changes over time, *M. truncatula* plants were grown in diurnal light-dark cycles for 40 days in order to entrain the circadian clock (Figure 3.1). The plants were then transferred to constant light and temperature conditions to examine the persistence of rhythms. Nodules, leaves and roots were sampled every 3h for the first 24 hours then every 6h for another 24 hours (a total of 13 time-points) in constant light and temperature. RNA was extracted from samples of nodules, and RNA integrity was evaluated; RNA Integrity Number (RIN) for all the

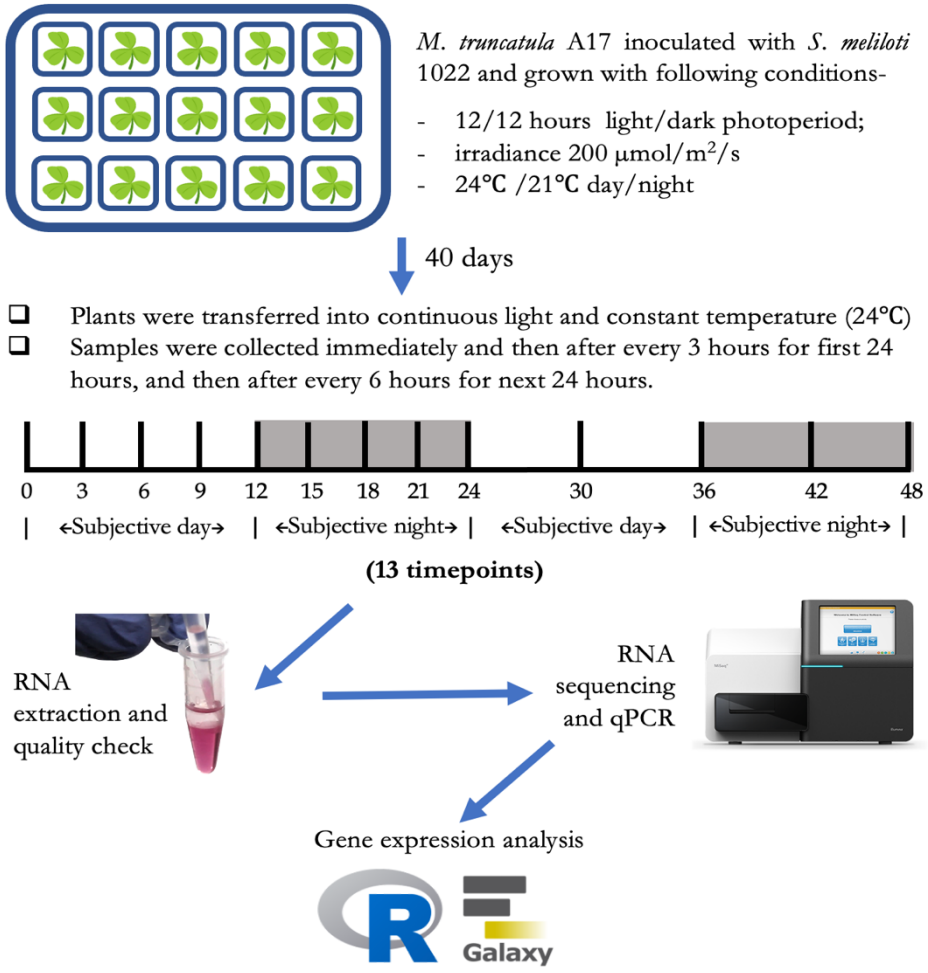
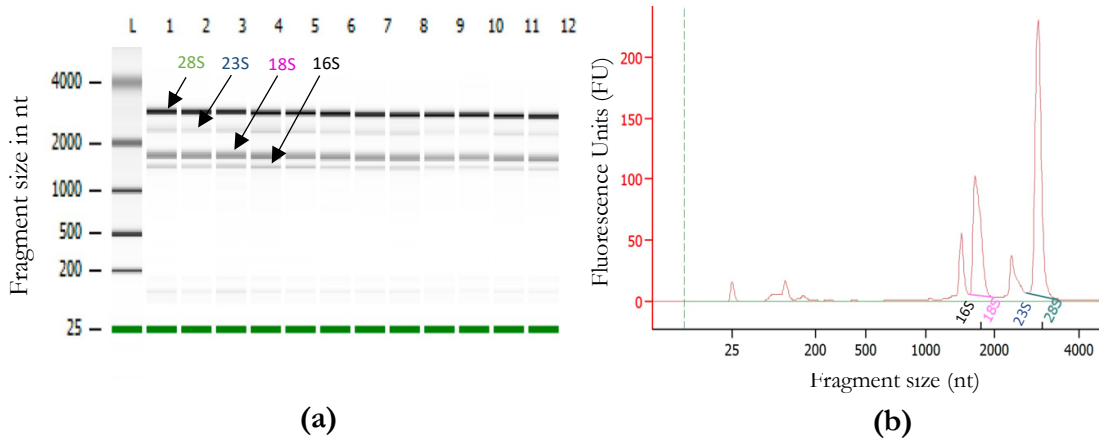


Figure 3.1: Overview of experimental design and gene expression analysis. Samples were collected in free-running condtions at 13 time-points over 48 hours. The periods between time-point 12 to 24 and 36 to 48 represent subjective nights, marked with a grey background.



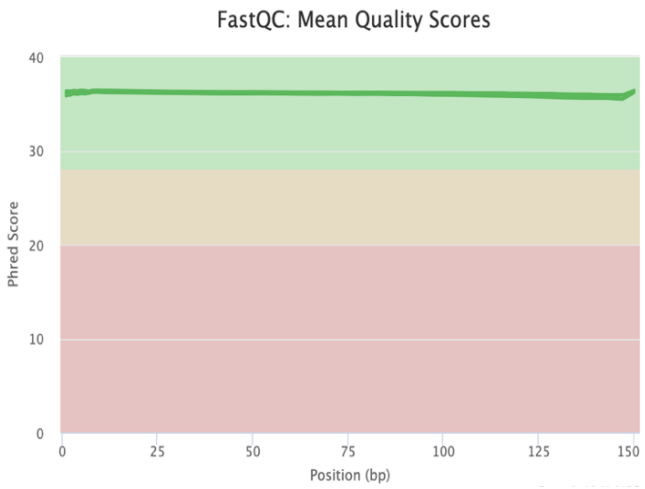
(a) **(b)**

Figure 3.2: Total RNA quality analysis on Bioanalyzer. (a) Electrophoresis (first 12 timepoints from replicate 1) shows four distinct bands; two for plant rRNA subunit 28S and 18S, another two for bacterial rRNA subunits 23S and 16S. Green bars indicate the lower marker (25 nt) **(b)** Electropherogram of a representative sample shows distinct peaks for the rRNA subunits and signals for RNAs smaller than 200 nt.

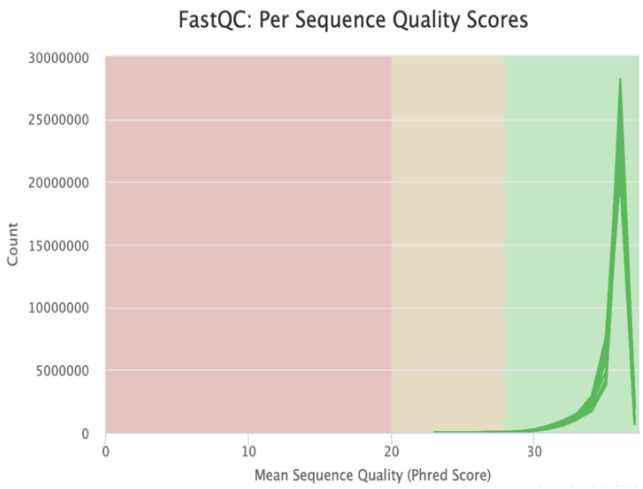
samples were above 8.5, suggesting good RNA quality and integrity as confirmed by electrophoresis and electropherogram analysis (Figure 3.2). The RNA samples were sequenced, generating 150bp, paired end RNA sequencing data (see Chapter 2 for methods). In total, three replicate experiments were set; therefore, there were three replicate samples for each timepoint (except for 15h and 21h timepoints for which two replicates were used as clarified in section 2.4.1). All the replicate samples were collected, processed and analyzed separately (using same protocol) to evaluate and confirm the reproducibility of our results. In general, the RNA expression dataset was analyzed to ask if there were any periodic changes in gene expression in the nodules. qPCR was carried out to check the expression of the important clock gene Mt*LHY* (the *M. truncatula* homolog for both *CCA1* and *LHY*; further characterised in Chapter 4), in all the samples from leaves, roots and nodules. This enabled comparison of the rhythmic behaviour of the Mt*LHY* gene in different tissues.

3.2.2. About two thirds of *M. truncatula* genes are expressed in nodules

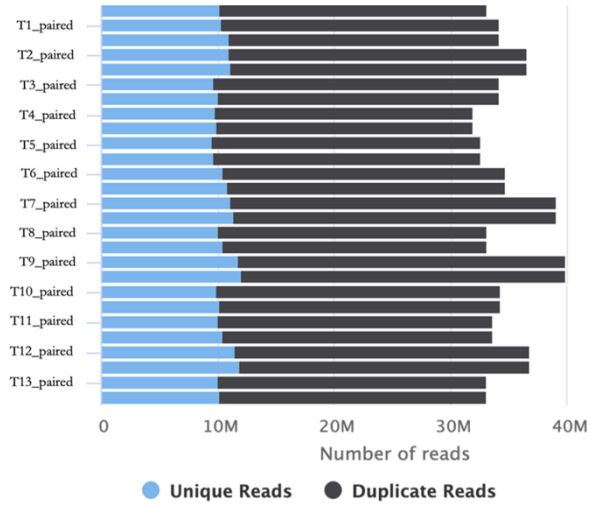
Upon receiving the raw sequencing data, the quality control information was analyzed and the raw data were processed to maintain high standards (as shown in Figure 3.3). Once the sequenced reads had been processed and normalized, they were mapped to the *M. truncatula* transcriptome (Mt4.0v1) (see Chapter 2 for methods). After this, read-counts and TPM normalized expression levels for 61,510 transcripts were generated. About one-third (33.5%) of these transcripts contain no read or less than one read on



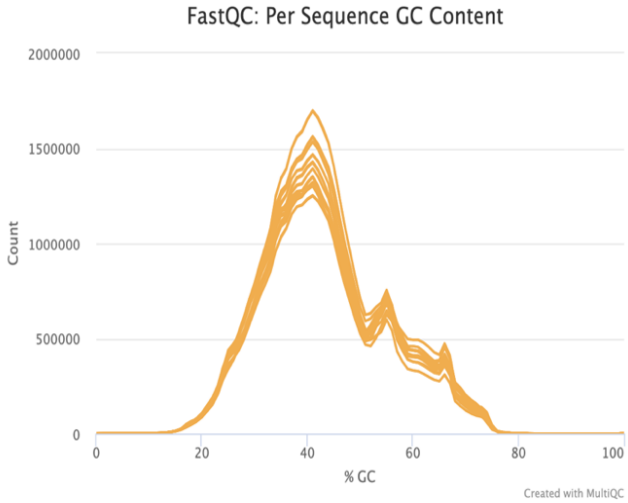
(a)



(b)



(c)



(d)

Figure 3.3: Quality control information of the raw reads (post trimming) for each time-point. Quality data presented here are from replicate-1, but are representative of reads from other replicates as well. For (a,b,d) each line denotes a sample. **(a)** The mean phred quality scores across each base position in the read. High quality scores (above 35) were observed throughout, even at the very end of 150bp long reads. **(b)** The number of reads with average phred scores show that most of the reads had high mean score (>35) with no reads with a phred score below 20. **(c)** Analysis of sequence count shows there are a high number of duplicated reads which is not uncommon in RNA-seq data. The consistency in the duplication rate in all the samples (T1-T13) was supportive of good data quality. T denotes time-point, e.g., T1_paired represents pair of fastq files from time-point 1 (0h). **(d)** The distributions of GC content exhibit a similar pattern across all the samples.

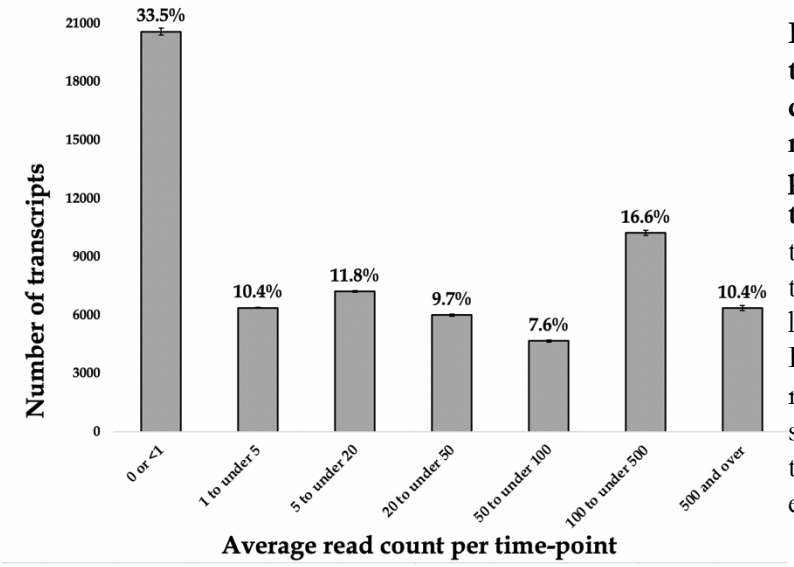


Figure 3.4: Number of transcripts with different ranges of reads count (per time-point averaged over 13 time-points) represents the percentage of transcripts with different levels of expression. Error bars are representative of the standard deviation across three biological replicate experiments.

average per time-point over the 13 time-points, suggesting that the corresponding genes had no or negligible expression. Another ~6,000 transcripts had very low expression levels (comparative to other expressed transcripts; less than 5 reads on average). Around 29% of genes (~18,000 transcripts) had relatively medium level expression (5 to less than 100 reads). The remainder of the genes (27%), have a moderate to high expression level of more than 100 reads per time-point on average. These percentages were very consistent over the three experimental replicates, with a very small standard deviation between values (Figure 3.4).

3.2.3. Core clock components exhibit characteristic rhythmicity in nodules

In order to ask if circadian clock oscillations are conserved in nodules, first, a number of different circadian clock components in *M. truncatula* that are orthologous to *A. thaliana* clock proteins were identified thorough reciprocal BLASTp (protein BLAST) search and homology analysis using Phytozome database and *A. thaliana* clock protein sequences as starting sequences (Table 3.1). The ortholog list was then also cross checked with KEGG Orthology (KO) database; the ones that were not founded in KO database were discarded. Interestingly, in terms of homology, *A. thaliana CCA1* (At2g46830) and *LHY* (At1g01060) are represented by a single gene, Medtr7g118330, in *M. truncatula*. This suggests that this single gene is likely to confer the functions of both CCA1 and LHY in *M. truncatula*. This is not unusual since some other plant species (e.g., rice, garden pea) also contain a single close homolog for both *CCA1* and *LHY* (Song et al., 2010). The homolog in *M. truncatula* can be considered as either *CCA1* or *LHY* but it is comparatively more similar to *A. thaliana LHY* (*AtLHY*) and is therefore annotated as *LHY (MtLHY)*.

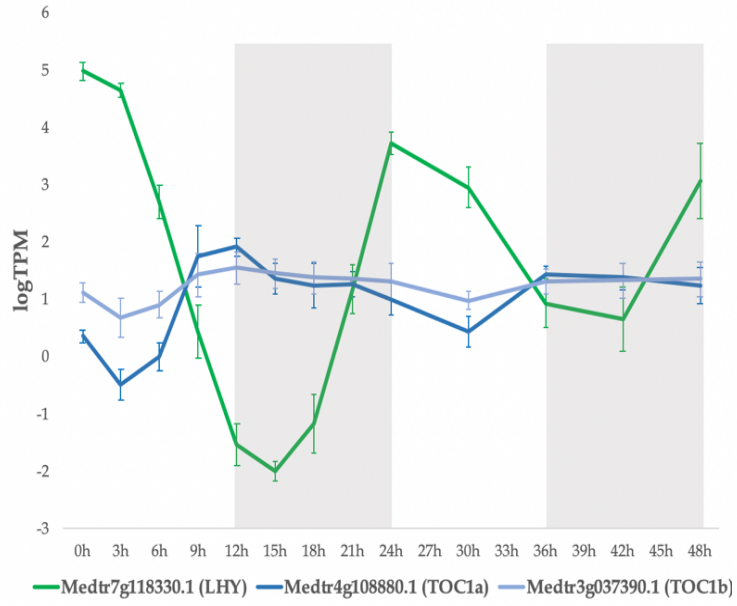
A. thaliana		*M. truncatula*		Reciprocal BLAST search						Phytozome homology analysis		
Clock gene	Locus ID	Locus ID	Clock gene	Total score	Coverage	E value	Recip. Score	Recip. Coverage	Recip. E	S-W Score	Similarity	Recip. Similarity
CCA1	At2g46830	Medtr7g118330	*MtLHY*	332	94%	2E-49	283	98%	4E-85	658	54.00%	44.00%
LHY	At1g01060	Medtr7g118330		393	99%	6E-127	385	98%	1E-123	986	66.60%	51.60%
TOC1	At5g61380	Medtr4g108880	*MtTOC1a*	517	97%	9E-178	523	96%	9E-180	1421	60.20%	63.60%
		Medtr3g037390	*MtTOC1b*	337	83%	5E-109	475	98%	2E-161	1247	55.80%	60.00%
ELF3	At2g25930	Medtr3g103970	*MtELF3a*	327	99%	8E-102	310	99%	3E-95	819	45.00%	45.50%
		Medtr1g016920	*MtELF3b*	276	99%	1E-82	256	99%	3E-75	674	43.70%	46.50%
		Medtr8g015480	*MtELF3c*	139	53%	3E-34	183	34%	1E-32	469	40.00%	38.10%
ELF4	At2g40080	Medtr3g070490	*MtELF4a*	116	73%	2E-34	115	61%	5E-34	335	63.10%	53.80%
		Medtr4g125590	*MtELF4b*	82.8	65%	2E-21	82.8	73%	3E-21	241	51.40%	49.60%
		Medtr2g041310	*MtELF4c*	80.1	76%	2E-20	80.9	80%	2E-20	231	50.50%	49.10%
PRR5	At5g24470	Medtr3g092780	*MtPRR5a*	412	79%	8E-92	388	43%	5E-92	1114	48.90%	48.00%
		Medtr7g118260	*MtPRR5b*	293	76%	1E-90	313	71%	6E-65	887	43.80%	49.50%
		Medtr8g024260	*MtPRR5c*	288	36%	2E-55	296	42%	2E-57	693	40.80%	42.50%
		Medtr4g061360	*MtPRR7b*	269	26%	2E-49	271	22%	1E-50	599	40.60%	34.00%
PRR7	At5g02810	Medtr1g067110	*MtPRR7a*	270	48%	9E-51	516	93%	3E-173	1171	54.50%	52.90%
		Medtr4g061360	*MtPRR7b*	276	47%	3E-53	407	91%	1E-130	1082	50.10%	45.70%
		Medtr3g092780	*MtPRR5a*	350	64%	8E-69	263	30%	7E-48	611	35.90%	38.10%
		Medtr7g118260	*MtPRR5b*	295	37%	2E-60	246	29%	6E-43	566	34.10%	45.80%
		Medtr8g024260	*MtPRR5c*	287	49%	9E-58	248	28%	3E-41	566	39.10%	44.40%
PRR9	At2g46790	Medtr3g092780	*MtPRR5a*	360	69%	2E-68	360	45%	3E-68	802	55.60%	38.00%
		Medtr7g118260	*MtPRR5b*	296	37%	2E-60	296	29%	4E-60	709	51.70%	38.40%
		Medtr4g061360	*MtPRR7b*	276	47%	3E-53	239	26%	9E-54	682	51.70%	30.40%
		Medtr8g024260	*MtPRR5c*	287	49%	9E-58	250	68%	8E-58	668	60.30%	44.10%
GI	At1g22770	Medtr1g098160	*MtGI*	734	99%	0	1712	99%	0	4973	80.60%	80.40%
LUX	At3g46640	Medtr4g064730	*MtLUX*	239	70%	6E-77	244	92%	6E-79	565	44.10%	45.80%
RVE1	At5g17300	Medtr5g076960	*MtRVE1*	196	75%	4E-58	185	78%	8E-54	415	18.90%	17.40%
RVE7	At1g18330	Medtr6g477860	*MtRVE7*	223	82%	2E-52	230	77%	3E-55	497	47.80%	35.50%

Table 3.1: Orthologous clock genes in *Medicago truncatula*. Orthologous *M. truncatula* clock genes were identified through reciprocal BLASTp (protein BLAST) searching and analysis using the Phytozome database using *A. thaliana* clock protein sequences as starting sequences. Reciprocal BLASTp was carried out with the NCBI blastp suite; bit scores, percentage query coverages and Expect (E) values are reported for forward and reciprocal queries. The Smith-Watermann (SW) score were retrieved from the Phytozome v12.1.6 database; SW scores, forward and reciprocal similarity are reported. In cases where a reciprocal high degree of similarity was identified, multiple potential homologues are listed. Multiple homologues are named as a,b,c, respectively, based on their level of similarity to Arabidopsis ortholog, with 'a' being the highest.

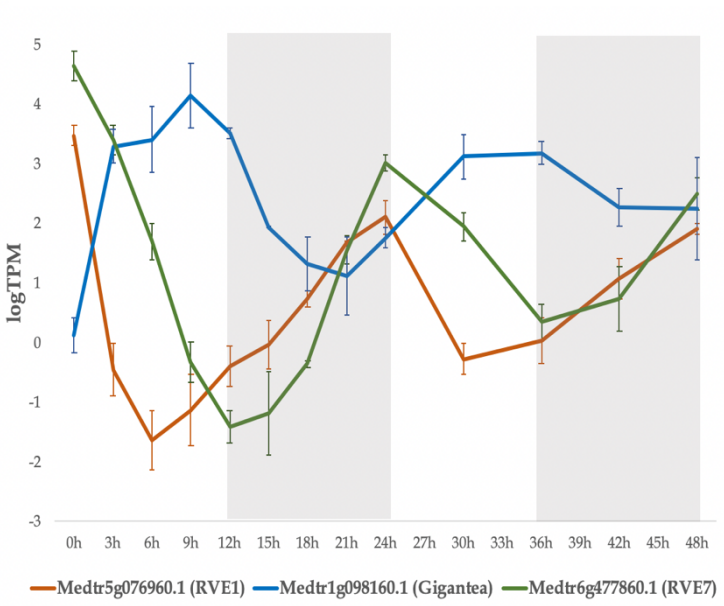
Some clock genes from *A. thaliana* (e.g., *TOC1*, ELF3, PRRs) have multiple close homologs in *M. truncatula* that are similar to each other. For example, *AtTOC1* shares about 60% and 56% similarity, respectively, at protein level with two *M. truncatula* genes, medtr4g108880 and medtr3g037390; both of them annotated as *MtTOC1* in *M. truncatula* genome.

The RNAseq expression data of the key circadian genes (listed in Table 3.1) was individually evaluated. In general, the expression of all studied clock components in nodules had shown a similar periodicity and expression pattern compared to *A. thaliana*. The expression of *LHY* was highest at beginning of the day, at the 0th and 24th hour (h) time-point, and then declined throughout the day. It was at its lowest level at subjective evening or early night (15th, 42nd h). For rest of the night, *LHY* expression had risen continuously to reach its peak in the next subjective morning (24th, 48th h) (Figure 3.5a). Two close paralogs of *MtLHY*, Medtr5g076960 and Medtr6g477860, that are MYB-transcription factors and orthologs of *AtRVE1* and *AtRVE7*, respectively, showed similar expression pattern as *MtLHY* (Figure 3.5b). However, *MtRVE1* expression declined much more quickly during the day and reached its lowest point at around mid-day (6th, 30th h). Exhibiting an almost opposite expression pattern, *TOC1* expressions were lowest in early morning (3rd, 30th h) and highest at the evening (12th, 36th h) (Figure 3.5a). There are two *TOC1* paralogs in *M. truncatula* as depicted by *TOC1a* and *TOC1b*. Both the *TOC1* paralogs had shown similar rhythmic pattern, but *TOC1a* oscillated with a much higher amplitude compared to *TOC1b*.

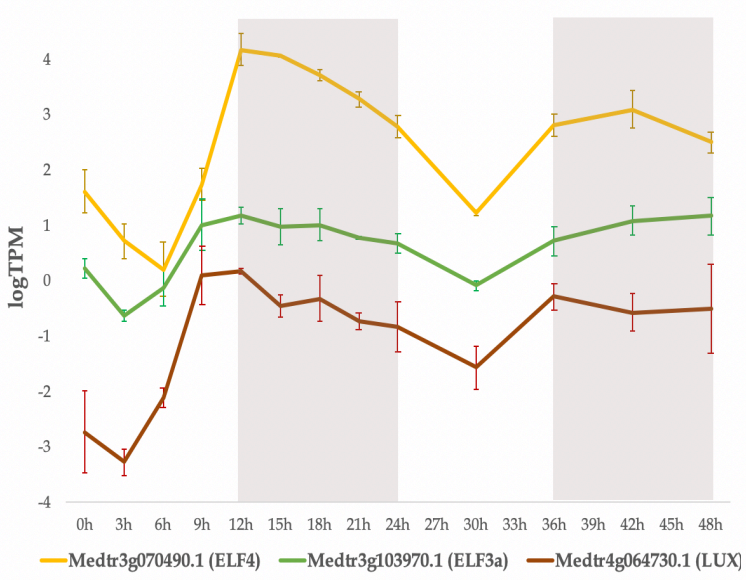
The expression of evening complex (EC) components *ELF4, ELF4* and *LUX* were similar to that of *TOC1s* (Figure 3.5c). In *A. thaliana*, *TOC1* peaks a little earlier than the EC components (Sanchez and Kay, 2016). This could be the case here as well, but since there is, at least a 3 hour gap between each sample (6h in the second cycle), it was not resolvable. Among EC components, *ELF4* had the highest expression magnitude and *LUX* was expressed in low levels. Among the five *PRR* homologs (listed in Table 3.1), two (*PRR5b, PRR7b*) had no or negligible expression. The other three peaked sequentially (*PRR5c, PRR5a, PRR7a*) at different times during the day which matched the expression of different PRR gene in *A. thaliana* (Figure 3.5d). While the expression of *PRR5a, PRR7a* peaked around the subjective afternoon, *PRR5c* peaked in the morning (3rd, 30th h). *MtPRR5c* is similar at the sequence level to *AtPRR9* which is also expressed earlier than other *PRRs* in *A. thaliana* (Nakamichi et al., 2010). The expression of another important



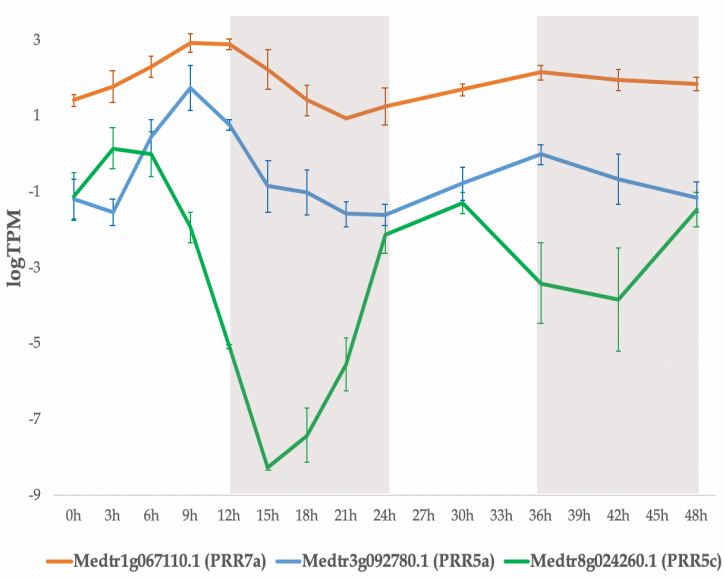
(a) The core oscillator



(b) RVEs and *GIGANTEA* (GI)



(c) Evening Complex (EC)



(d) Pseudo-response regulators (PRRs)

Figure 3.5: RNAseq expression data of key circadian clock genes in *M. truncatula* nodules over a time-course. (a) The core oscillator: *LHY* and *TOC1* form the core oscillatory loop via a negative feedback mechanism; **(b) RVEs and Gigantea (GI):** *RVE*s and *GI* are important clock components that synchronize the central clock with output pathways.; **(c) Evening Complex (EC):** The EC is formed of three proteins, ELF3, ELF4 and LUX, whose expression peaks in the evening**; (d) Pseudo-response regulators (PRRs):** *PRR* genes are homologous to *TOC1* and are expressed mainly during the day to form additional regulatory loops.

(a-d) White and grey background represents subjective day and night. Plants were grown in 12/12h day/night cycle, but samples were taken from plants grown in constant light conditions; 0, 24h and 12, 36h time-points mark the beginning of the day and night, respectively). Error bars indicate standard deviation of the mean (n=3 biological replicates, except for 15h and 21h where n=2).

clock gene, *GI*, peaked around the afternoon (9th h), in a similar manner to its expression in other plants.

In summary, the clock genes are expressed and oscillate in an expected way in *M. truncatula* nodules which is similar to their rhythmic expression (as discussed in Chapter 1) in other species (mainly *A. thaliana*, since most studies were carried out in it). The expression and persistence of typical rhythmic behaviour of these clock genes suggests the existence of a nodule clock that might regulate many rhythmic processes during nodulation and N-fixation.

3.2.4. *MtLHY* expression amplitude varies in different tissues

In order to determine whether clock genes are expressed in a similar manner in different organs, the expression of *MtLHY* in leaves, roots and nodules was quantified by quantitative PCR (qPCR) (Figure 3.6) using RNA samples from the same time-course

Figure 3.6: Expression of *MtLHY* in leaves, roots and nodules over a time-course. The expression values are generated through qPCR and are relative to beta-tubulin (Medtr7g089120). White and grey background represents subjective day and night. Error bars are representative of the standard deviation of the mean of three biological replicates.

RNAseq experiment mentioned above. The rhythmicity of *MtLHY* expressions was found to be very similar in these tissues, however, the amplitude differed. In leaves, the amplitude was highest and was comparatively decreased in roots and nodules. Interestingly, a slight phase-shift to a later peak was observed in leaves in the second cycle (24^{th}-48^{th} h). In the first cycle (0^{th}-24^{th} h), the expression of *MtLHY* was highest at the beginning of the day and declined steadily during the day, reaching a minimum in the evening (12^{th} h). However, the next peak was observed around midday (30^{th} h, a tentative phase shift of 6h). The expression minimum in the second cycle was reached at around midnight (42h), which was later than the evening minimum in the first cycle. Although, due to spacing of sampling time it is difficult to determine whether the phase shift was precisely 6h or different, but according to available data it is likely that the second peak was delayed in leaves. On the other hand, in root samples *LHY* expression reached its maximum and minimum point, respectively, at the beginning of the day and at around evening in both cycles. In nodules, the expression peak was also observed at the beginning of the day in both cycles, but the decline was comparatively slower, and the lowest expression was reached closer to midnight. These observations suggest that *MtLHY* rhythmicity is much higher in leaves and more sensitive to light as the phase-shift was potentially due to the greater responsiveness of leaf tissue to continuous light conditions. In both roots and nodules, the overall expression levels were much lower than leaves although the expression peaks at the same time at dawn. There was however a lag in decline, more prominently in nodules than roots. However, in general, both the diurnal expression pattern of *MtLHY* together with its expression variation across different organs is in line with known role and specificity of this gene (Endo et al., 2014, Bordage et al., 2016, Takahashi et al., 2015). The data, nonetheless, supports the existence of clock function in nodules.

3.2.5. Around 5% of all genes are rhythmically expressed in nodules

We used MetaCycle (see Chapter 2 for methods) to identify rhythmically expressed transcripts out of 61,510 transcripts (50,894 genes) for which the expression data were generated through RNA sequencing analysis. A total of 1,363 rhythmic transcripts (from 1,296 genes) were identified with p-value <0.01; p-values were found using a Lomb-Scargle periodogram. All of the key clock/circadian genes that (Table 3.1) were expressed (Table 3.1 & Figure 3.5) found within this range (p <0.01). A Gene Ontology (GO)

enrichment analysis was carried out on annotations from these 1,296 genes. The results showed that most of these genes are associated with 'transport and localization', 'carbohydrate metabolism', 'circadian rhythm', and some other biological processes (Figure 3.7). This seems reasonable as photosynthetic CO_2 fixation that occurs during the day drives the synthesis of sucrose and their subsequent transport to amyloplast for starch accumulation. During the night, starch molecules are hydrolyzed and resulting sugar molecules are transported to various sink organs and used for energy production and plant growth (Kim et al., 2017). Many rhythmic genes detected in our analysis are, therefore, likely to be activated in a diurnally regulated, timely manner to coordinate

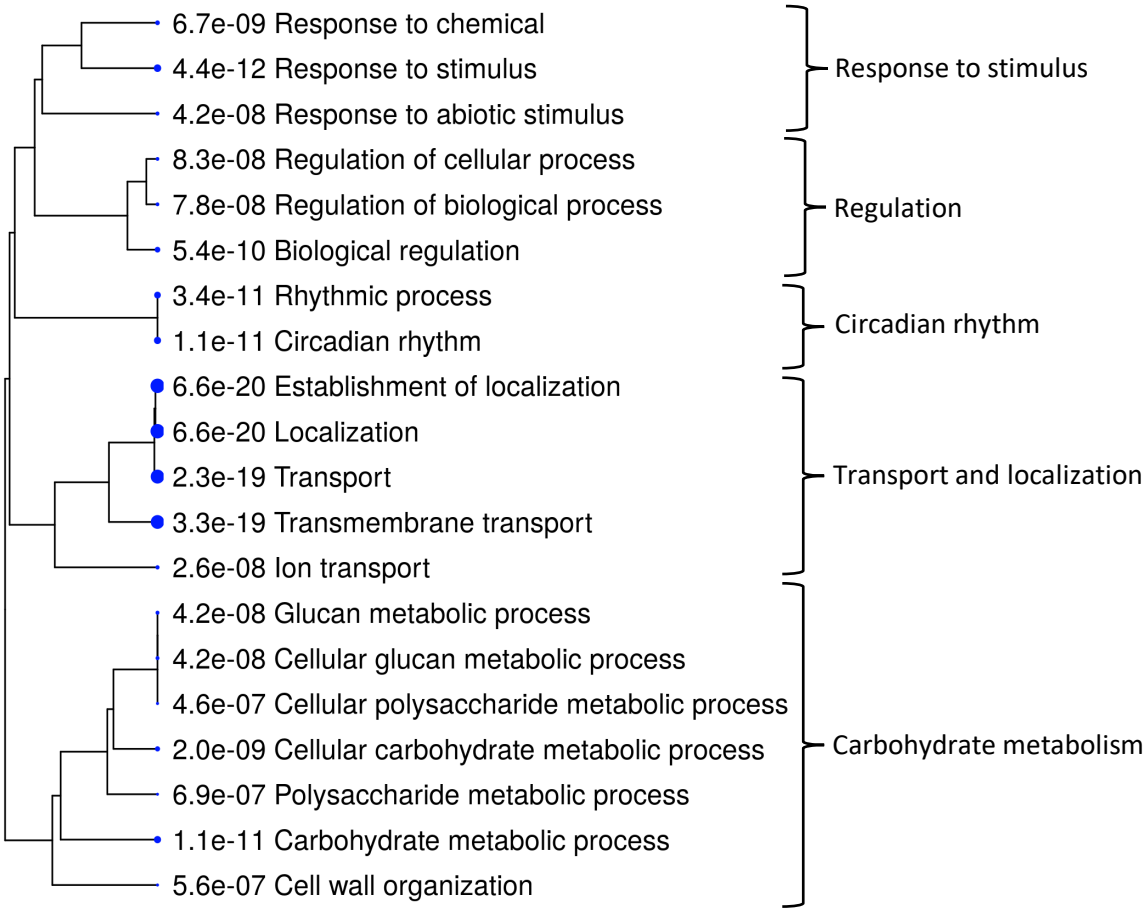


Figure 3.7: A hierarchical clustering tree representing the significantly enriched (top 20) biological processes in 1,363 rhythmic transcripts, grouped by similar functions in the GO hierarchy. Processes with many shared genes are clustered together. Enrichment *p*-values (FDR adjusted) are shown for each biological process. *p*-values are also presented by blue dots where larger dots represent more significant *p*-values. The clustering was generated using the ShinyGO v0.61 tool (Ge et al., 2019), available online at: http://bioinformatics.sdstate.edu/go.

different carbohydrate related catabolic, anabolic or other regulatory processes involving movement and localization of different biomolecules.

Many genes that are regulated by circadian rhythm in an indirect or distant manner (for example, the genes whose promoters are not bound by a core circadian TF (e.g., LHY, TOC1), but regulated by some other oscillating components), might oscillate but with smaller amplitudes. Such genes might not be detected with stringent cycling parameters, such as a *p*-value of less than 0.01. To increase the chance of identifying a wide range of levels of oscillation in nodules, all transcripts with oscillating $p < 0.05$ were considered and a total of 2,832 rhythmic transcripts (2,703 genes; 5.31% of all genes) were found (See supplementary Table S1 & S3). GO enrichment analysis with these transcripts also resulted in very similar enrichments (as with $p < 0.01$) with some additional but related ontologies that includes 'biosynthetic process', 'macromolecule modification', 'phosphorylation' and few others. Since this extended list of rhythmic transcripts ($p<0.05$) seemed to capture more biological processes, we decided to use these transcripts for further analysis.

To understand the rhythmic behaviour of these oscillating transcripts, the logTPM expression values for all transcripts were subjected to Z-score normalization that enables mean-centered expression values to be generated that are well suited for the comparison of different rhythmic transcripts with varying expression levels. Next, hierarchical clustering with the elbow method was used to tentatively organize these 2,832 transcripts into four major clusters to identify genes and related biological processes that seem to peak at specific times during the photoperiod, and thus might function together. Based on the timing of their expression peaks, the four clusters were name as 'dawn' (cluster 1), 'midday' (cluster 2), 'dusk' (cluster 3) and 'night' (cluster 4) (Figure 3.8).

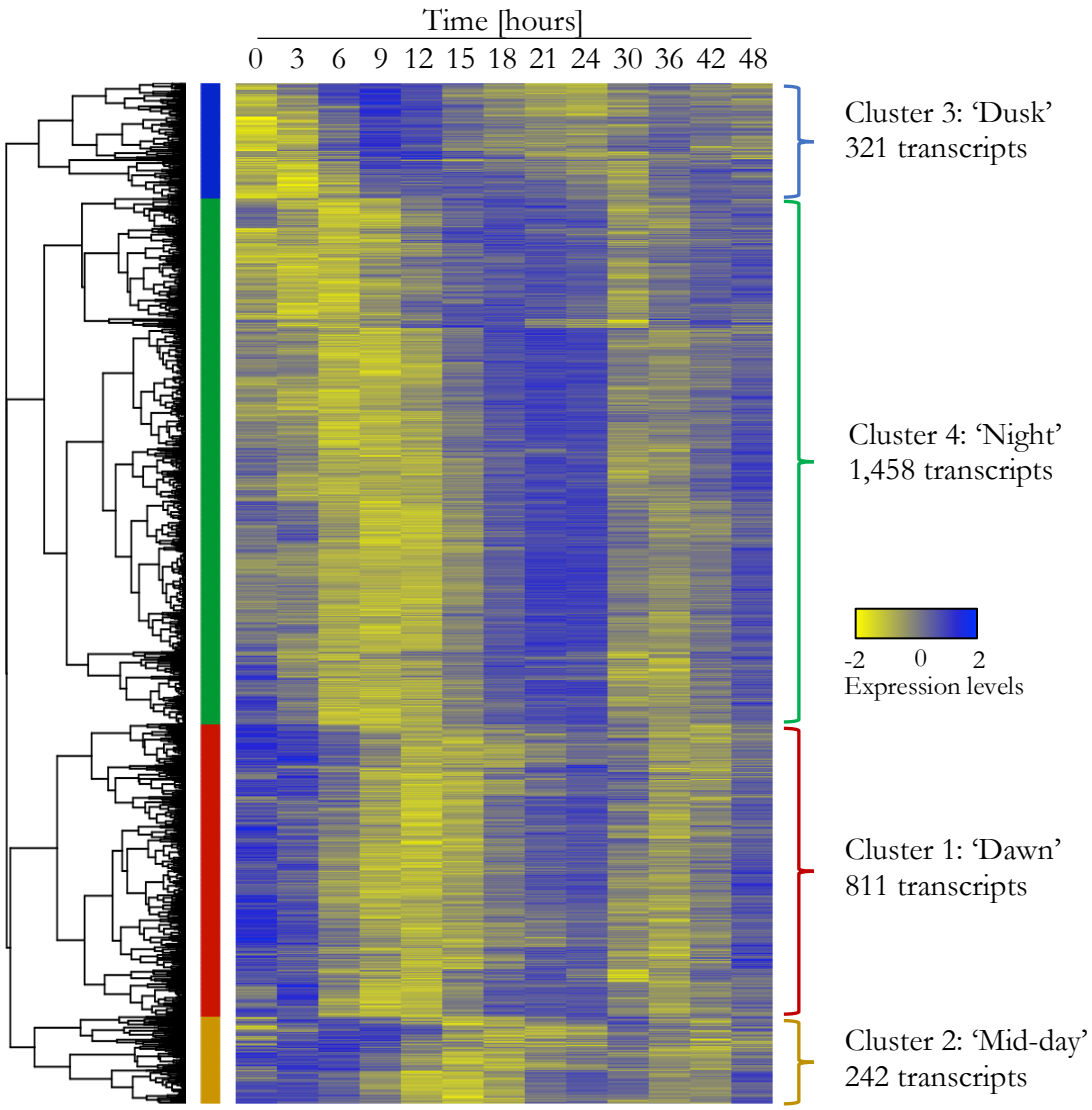
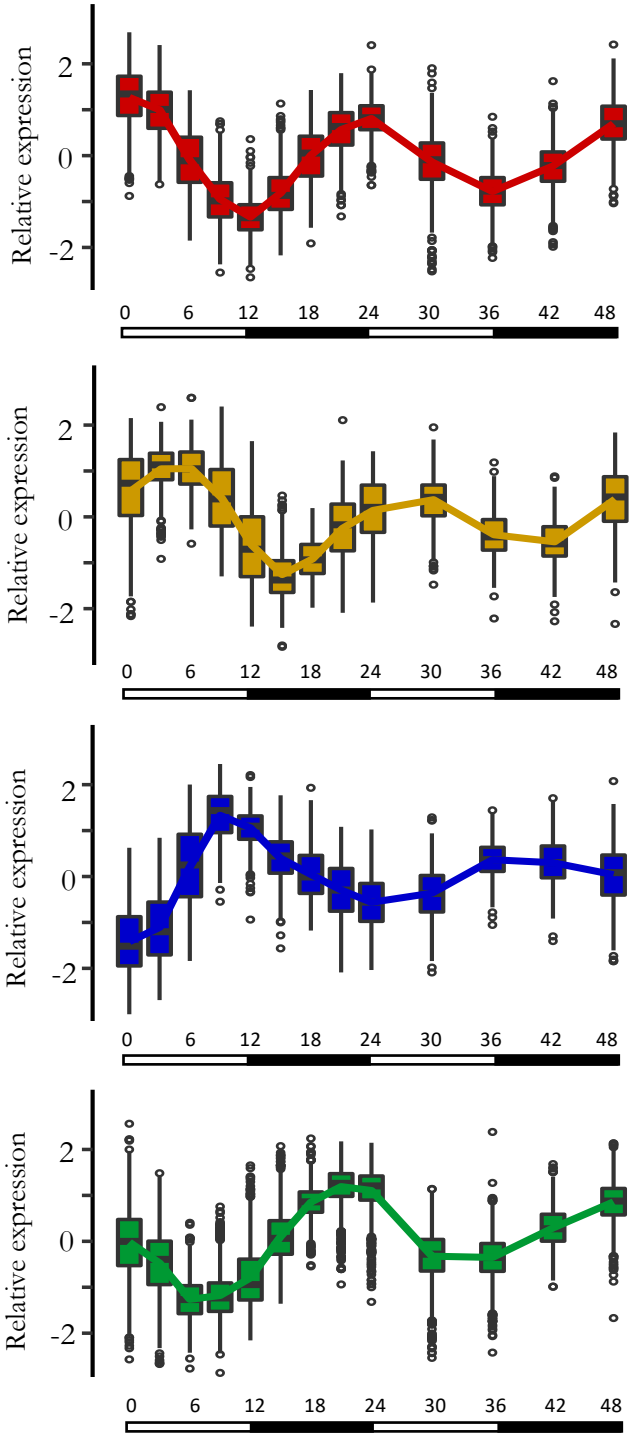


Figure 3.8: A heat map based on hierarchical clustering shows the expression of 2,832 rhythmic transcripts. Four major clusters (1 to 4) are represented respectively by red, yellow, blue and green colours/curly brackets. Based on the timing of their expression peak they are also named as 'Dawn', 'Midday', 'Dusk' and 'Night' clusters. In the heatmap, the intensity of blue and yellow colour represents higher and lower expression (Z-normalized log_2TPM expression values), respectively.

The vast majority of oscillating genes were in the night (1,458 genes) and dawn (811 genes) categories. On first inspection, the expression pattern of these two clusters might seem very similar, both peaking around the 21-24h time-points (Figure 3.9 a & d). However, at 0h time-point the expression of the night cluster genes was much lower than the dawn genes. In addition, the night genes peaked a little earlier (at around 21h) than dawn genes (at around 24h). In general, most genes within the night group were actually expressed most highly in the late night. Since a total of 2,269 transcripts (out of 2,832)

belong to the night-peaking and dawn-peaking cluster category, therefore, it seems that most oscillating genes accelerate their expression during the night period (Figure 3.9 a & d). The midday and dusk clusters contain 242 and 321 transcripts, respectively, and exhibit distinctive expression patterns (Figure 3.9 b & c). While analyzing RNA sequencing expression data, it should be always taken into consideration that the expression values for the first half (0-24h) were more accurate (timing-wise) than the second half (24-48h) due to the shorter sample spacing in the first day. In addition, by the second day, plants could be already trying to adapt to the new continuous light conditions and thus gene expression timing is altered, as seen in other studies (Webb et al., 2019).

All the major circadian genes analyzed, were in respective clusters that matched their expected expression peaks. The morning-expressed *LHY* and *RVE7* were found in cluster-1 (dawn). Evening complex components (*ELF3*, *ELF4* and *LUX*) and *TOC1* were found in either cluster 3 (dusk; *ELF3, LUX, TOC1*) or cluster 4 (night, *ELF4*). Different PRR genes were found in either cluster 2 (midday; *PRR5c*) or 3 (dusk; *PRR5a, PRR7a*), as they are expressed sequentially throughout the day (as described earlier in Section 3.2.3). One gene that did not appear in the expected cluster is *RVE1* that is in cluster 4 (night) which is predominantly a morning-expressed gene However, *RVE1* expression dropped sharply to reach its minimum earlier than other morning genes (as shown in Figure 3.5b). The timing of minimal expression resembled that of night-peaking genes which could be a reason for its unexpected clustering location. In general, the presence of most key circadian genes in expected clusters adds confidence to our experimental data and clustering analysis. This also suggests that although there might be some specific differences due to cell and organ type, in general the nodule clock operates in a similar way to the clock in roots or leaves, which is consistent with its role as a core regulator of growth and development.



(a) Cluster 1: 'Dawn'

811 transcripts, including-
LHY, *RVE2*
2 *NCRs*

(b) Cluster 2: 'Mid-day'

242 transcripts, including-
GI, *PRR5c*
5 *NCRs*

(c Cluster 3: 'Dusk'

321 transcripts, including-
ELF3, *PRRs*, *LUX*
24 *NCRs*

(d) Cluster 4: 'Night'

1,458 transcripts, including-
ELF4, *RVE1*
14 *NCRs*

Figure 3.9: Expression pattern of gene transcripts within each cluster, shown as a centroid boxplot for each cluster. (a) Cluster-1 (red) genes peak at 0 and 24h time-points that represent the beginning of the day (dawn); **(b) Cluster-2 (yellow)** genes peak around 6 and 30h (around midday); **(c) Cluster-3 (blue)** transcripts peak at 9h and around 36h (around dusk); **(d)** The **Cluster-4 (green)** gene expression pattern looks similar to cluster-1, but at 0h expression is much lower compared to cluster-1 and expression also peaks slightly earlier (21h) in the next experimental day, hence these are categorized as 'night' genes. (a-d) White and black bars underneath each plot indicate subjective day and night, respectively. Expression values for each gene are log2 normalised and relative to its mean expression over time. Box, line and dots at each time-point shows the distribution of expression values (dots are outliers) for corresponding time-point and cluster.

3.2.6. Rhythmic processes in the nodule are putatively coordinated in a manner similar to other known rhythmic processes in different plants and tissues

To identify different cluster specific biological processes and to investigate different aspect of nodule functions that might be under circadian regulation, a cluster specific GO enrichment analysis was carried out (Figure 3.10). Additionally, gene descriptions for all the rhythmic genes in four clusters were obtained from the Phytomine database and analyzed to look for pathway and protein domain enrichments.

The dawn cluster was found to be enriched for various ion transport and localization related GO terms (Figure 3.10a). Specially, this cluster contains more than a dozen of different anion transporters (e.g., five sulphate/bicarbonate/oxalate exchanger and transporter- Medtr7g022870, Medtr7g022870, Medtr5g061860, Medtr3g073730, Medtr6g086170). Plant anions transporters are well known for their role in cell signaling, osmoregulation, compartmentalization of metabolites, etc (Barbier-Brygoo et al., 2011). In general, this cluster contains about hundred genes that are associated with cellular localizations. Therefore, genes in this cluster could be associated with the movement and compartmentalization of compounds formed during the night. This cluster was also enriched for secondary metabolic pathways related to the biosynthesis of different flavonoids and flavonoid derivatives (e.g., formononetin derivatives like medicarpin, maackiain biosynthesis). Flavonoids and isoflavonoids are polycyclic, polyphenolic compounds that belong to a wider group of phytoalexins and are well known for their antibacterial activities (Cushnie and Lamb, 2005). Interestingly, as reviewed in chapter 1, different flavonoids and isoflavonoids induce rhizobia to secrete Nod factors and initiate symbiotic interactions. The exploitation of pre-existing secondary metabolites with a traditional role in defense into molecules with a role in beneficial plant-microbe interactions could be a reflection of the evolution of the mutualism continuum (Mierziak et al., 2014). However, the expression of flavonoid biosynthesis genes (e.g., two Isoflavone Reductase-Like Protein, Medtr5g020760 and Medtr5g020800) in the nodule suggests a role for these molecules beyond the initial manipulation of rhizobia to initiate nodulation. Since they have antimicrobial properties and were reported to be mostly produced in the infection zone (Chen et al., 2015b), they could be involved in maintaining a homogenous rhizobial population whilst also protecting nodules from potential

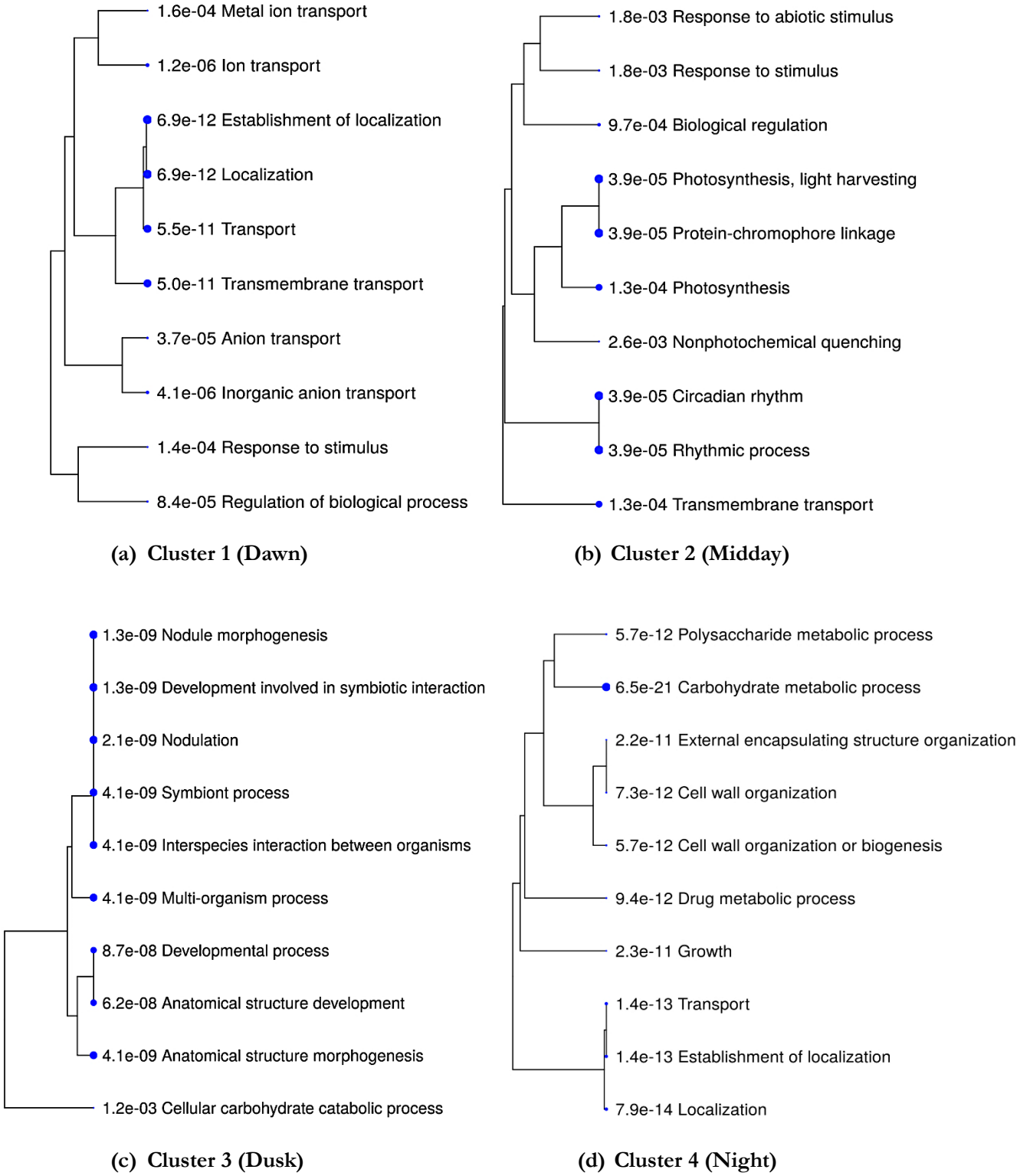


Figure 3.10: Cluster-specific enriched biological processes. (a) Cluster 1 (dawn) is enriched for ion transport and localization related genes; **(b)** Photosynthesis-related genes are enriched in the midday cluster; **(c)** Cluster 3 (dusk) contains genes that are important for developmental processes, especially nodulation; **(d)** Cluster 4 (night) is mostly enriched for carbohydrate related metabolic processes. The hierarchical clustering was generated using the ShinyGO v0.61 tool (Ge et al., 2019), available at: http://bioinformatics.sdstate.edu/go.

pathogens. Oscillation in flavonoid gene expression has also been reported in other plants, such as, in *A. thaliana* leaves (Deikman and Hammer, 1995) and ginkgo leaves which are a raw material for extraction of flavonoids for pharmaceutical use (Ni et al., 2018). In both plants, the expression of responsible genes was maximal around dawn. These observations and our findings are in line with the finding that *A. thaliana* leaves exhibit higher resistance to *P. syringae* infection in the morning (Bhardwaj et al., 2011). With the knowledge that flavonoid and isoflavonoid biosynthesis in nodules might follow a rhythmic pattern, it is tempting to hypothesize that the rhythmic pattern of leaves defense might be shared by nodules as well, potentially to stop pathogenic microbes from entering with rhizobia. Finally, the dawn cluster was also enriched for 'zinc finger' domain proteins, indicating the expression of nucleic acid binding proteins that might regulate transport and flavonoid metabolism.

The midday cluster that peaked around 3-6th h and then at 30th h was mostly highly enriched for light harvesting, photoperiodism and photosynthesis related genes (Figure 3.10b) including five light-harvesting complex I *CHLOROPHYLL A/B BINDING* genes (Medtr2g008610, Medtr4g015570, Medtr4g094605, Medtr5g097280 and Medtr6g060175) *PHYTOCHROME INTERACTING FACTOR* 45 (*PIF*45, Medtr3g449770) etc. This suggests a light signal sensing and transduction activity in nodules during the period of the day with the highest levels of light, although nodules are underground organ where light access is limited. The Mid-day cluster-2 also contains eight *WALLS ARE THIN (WAT)-1*-related genes that encode for different glycoside hydrolases and nodulin genes associated with cell wall biosynthesis. Arabidopsis WAT1 is a tonoplast located protein and a homolog of Medicago nodulin21 which is required for secondary cell wall deposition. This gene also affects flavonoid biosynthesis and chlorophyll a/b content (Ranocha et al., 2010), which could link it to the coordinated expression of other photosynthetic genes. Transmembrane transport genes were also enriched in this cluster that might be involved in the movement of carbon assimilated molecules related to nitrogen fixation activity. In general, the abundance of different plastid-localized, thylakoid membrane-associated and photosystem I or II related genes, in cluster-2 suggests enrichment of photosynthesis-related activities for genes peaking around midday.

The dusk cluster (cluster 3) was dominated by nodulation-related GO terms (Figure 3.10c). This cluster contains a large number of NCRs (n = 24) that are highly nodule-specific genes (discussed later). Besides, more than thirty anatomical development and morphogenesis associated genes were also found in this cluster, suggesting nodule growth might be enhanced around dusk which is in line with our understanding that plants grow comparatively faster in the evening and at night (McDonald, 2011). Carbohydrate catabolic processes, glycogen and starch degradation pathways (including genes like β-*AMYLASE*, Medtr5g013620; β-*GLUCOSIDASE*, Medtr4g122980; etc.) were also enriched in this cluster. Starch degradation in particular has been reported to be regulated by the circadian clock (Graf et al., 2010) and our analysis suggests that the starch degradation peaks in the evening or during the early part of the night in nodules, and is possibly used for nodule growth.

The night cluster (cluster 4) contains about half of all oscillating genes (1,458 transcripts) and they mostly peaked few hours before dawn. This cluster was also enriched for carbohydrate metabolism, transport and growth-related processes and genes involved in the glycogen degradation pathway (including six different glucosidase and several amylases). This suggests that starch degradation and growth continue during the night as well. GO-terms associated with phosphate metabolic processes (over 100 genes) and pathways related to nitrogen-metabolism were all enriched in this cluster. Orthologs of *A. thaliana* plasmodesmata (PD) callose-binding proteins (PDCBs; Medtr4g109490, Medtr5g044530) and plasmodesmata-located proteins (PDLPs; Medtr8g104990) were found in this cluster. This is in line with our current understanding that PD transport is to some extent circadian regulated, usually enhanced during day and declined at night. Callose (β-1,3-glucans) deposition around plasmodesmata is one of the mechanisms by which plant block PD transport (Brunkard and Zambryski, 2019). The PDCBs and PDLP found in this cluster might be associated with this callose accumulation and regulation of PD transport at night. Pathway enrichment showed that the late-night peaking cluster contained ureide biosynthesis genes (e.g., *URICASE/URATE OXIDASE*, Medtr1g048370). Ureides are synthesised in nodules then transported through the xylem vessels to the leaf tissue where they are typically used up for ammonia metabolism (Gibon et al., 2004). Although classically, indeterminate nodule producing IRLC legumes like *M. truncatula* have been classified as amide-type (rather than ureide-type), ureide compounds have also been identified in this plant species, suggesting that this part of the metabolism

still occurs in these species during or related to N fixation activity (Sprent, 2007). The presence of the PHENYLALANINE AMMONIA-LYASE LIKE protein (PAL-like, Medtr1g094780), involved in the phenylpropanoid pathway, could be related to the synthesis of flavonoids, whose biosynthesis peaks slightly later, in cluster 1 as described above. Nitrogen metabolism genes in this cluster include many glutamate metabolic genes, nitrate/nitrite transporters, nitrate-transporting ATPases, and amino acid transporters. Many of these gene including *GLUTAMINE SYNTHETASE* (Medtr3g065250) were also found in the dawn cluster. Altogether, these data indicate that the nitrogen fixation that occurs throughout the day and that peaks around the evening (Minchin and Pate, 1974) enhances nitrogen metabolism, and production and transport of N-containing compounds (e.g., glutamine, asparagine) mostly during the night and possibly until dawn. *NODULE INCEPTION* (*NIN;* Medtr5g099060), an essential nodulation regulatory gene, was also relatively highly expressed during the night in this cluster, which again highlights the night phase as a period of enhanced nodule activity.

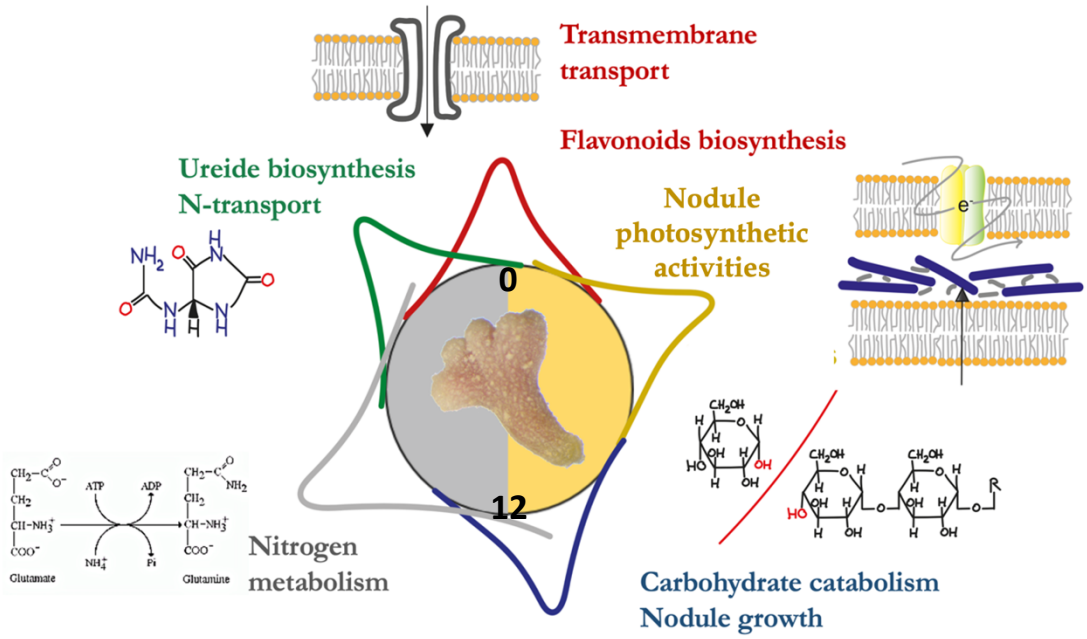


Figure 3.11: Overview of the processes that are regulated at the transcriptional level in nodules over time. Different major events that correspond to dawn, midday, dusk and night clusters are coloured red, yellow, blue and green, respectively (as Figure 3.8). Nitrogen metabolism (in grey) probably starts in the evening and occurs throughout the night; shown separately in between processes that are active between dusk and night. Hours 0 and 12 are indicated.

3.2.7. The Evening Element (EE) motif is enriched in dusk-peaking genes

As discussed in Chapter 1, a motif analysis previously carried out in our lab (Achom, 2019) revealed that a CCA1 binding motif [AGATATTT; known as Evening Element (EE)] or its close variant (AGACATTT; EE-like) is conserved in the promoters of many NCR genes. EE motif is common in rhythmically expressed gene promoters, specifically in the genes that are highly expressed in the evening (Adams et al., 2018). In order to ask whether this is also the case in our data, the promoter region of all the rhythmic genes (n = 2,703) were scanned for EE or EE-like motif occurrence using the FIMO tool, available in MEME suite (see Chapter 2 for methods).

The EE motif was observed 168 times in 157 rhythmic gene promoters whereas the EE-like motif was found only 59 times in 56 genes. Consistent with existing literature (Michael et al., 2008b, Adams et al., 2018), the EE motif was enriched in the dusk cluster with a prevalence of about 21% compared to equal or less than 5% in all other clusters (Table 3.2). This represents an enrichment of about 4 folds (hypergeometric test; p = 4.71e-25) compared to the overall presence of EE motif in all the rhythmic genes. No such enrichment was observed for the EE-like motif as it was found to be almost evenly distributed (1.5-2.5%) among all the clusters. Therefore, although the EE-like motif is very similar to the EE motif, it might not be a CCA1 or LHY binding site and might not be specifically related to the circadian clock.

Motif	#genes in dawn cluster	#genes in midday cluster	#genes in dusk cluster	#genes in night cluster
AGATATTT	28 (3.5%)	12 (5%)	68 (21.2%)	49 (3.7%)
AGACATTT	11 (1.6%)	6 (2.5%)	5 (1.5%)	34 (2.3%)

Table 3.2: Frequency of the EE (AGATATTT) or EE-like (AGACATTT) motifs in different clusters. EE motif is highly enriched in the dusk cluster. Such enrichment was not observed for the EE-like motif. The percentage values represent the prevalence of each motif among the genes belong to particular cluster.

3.2.8. NCRs that oscillate mostly belong to the dusk-peaking cluster

A list of 701 NCR genes (with gene locus ID) has been compiled from the supplementary information provided by Montiel et al. (2017) and de Bang et al. (2017) (Supplementary Table S3). Among the 61,510 transcripts analyzed in our study, 743 transcripts were found to be potentially expressed from those 701 NCR genes. Most of these NCR genes (> 600) were highly expressed with more than 100 reads or logTPM expression values over 3.0 on an average per time-point. 45 NCR transcripts were found within the 2,832 oscillating transcripts ($p < 0.05$) with 20 of them having a $p < 0.01$.

More than half (n=24) of the 45 oscillating NCRs belong to cluster-3 (dusk-peaking). As outlined above, cluster-3 contains only 321 transcripts in total which is much smaller cluster than cluster 1 or 4, and therefore in terms of percentage, cluster-3 contains more NCRs (7.5%) than any other clusters (0.25%, 2%, 1% respectively for cluster 1, 2, 4) (Figure 3.12). This represents a 4.7-fold over and significant enrichment (hypergeometric test; p=3.67e-12) of these genes in this cluster. This corroborates with our hypothesis that Mt*LHY* negatively regulates the expression of NCRs by binding to their promoters, thus their expression pattern was opposite to that of Mt*LHY* over time. Similarly, other genes that are known to be regulated by CCA1/LHY (e.g., *ELF3, TOC1, LUX* etc.) are expressed highly at dusk when Mt*LHY* expression is at a minimum. Cluster-4 contains fourteen NCRs which is similarly in line with our hypothesis of MtLHY-NCR regulation since *MtLHY* levels are still low for most of the night. Cluster-1 and -2 contains only five and two oscillating NCRs, respectively. The regulation for these seven NCR transcripts cannot be explained by a MtLHY-NCR regulation hypothesis. Nonetheless, the findings that some NCRs did oscillate, and most oscillating NCRs were expressed highly at the evening or night support that NCR-mediated processes in nodules are regulated by Mt*LHY*, at least to some extent.

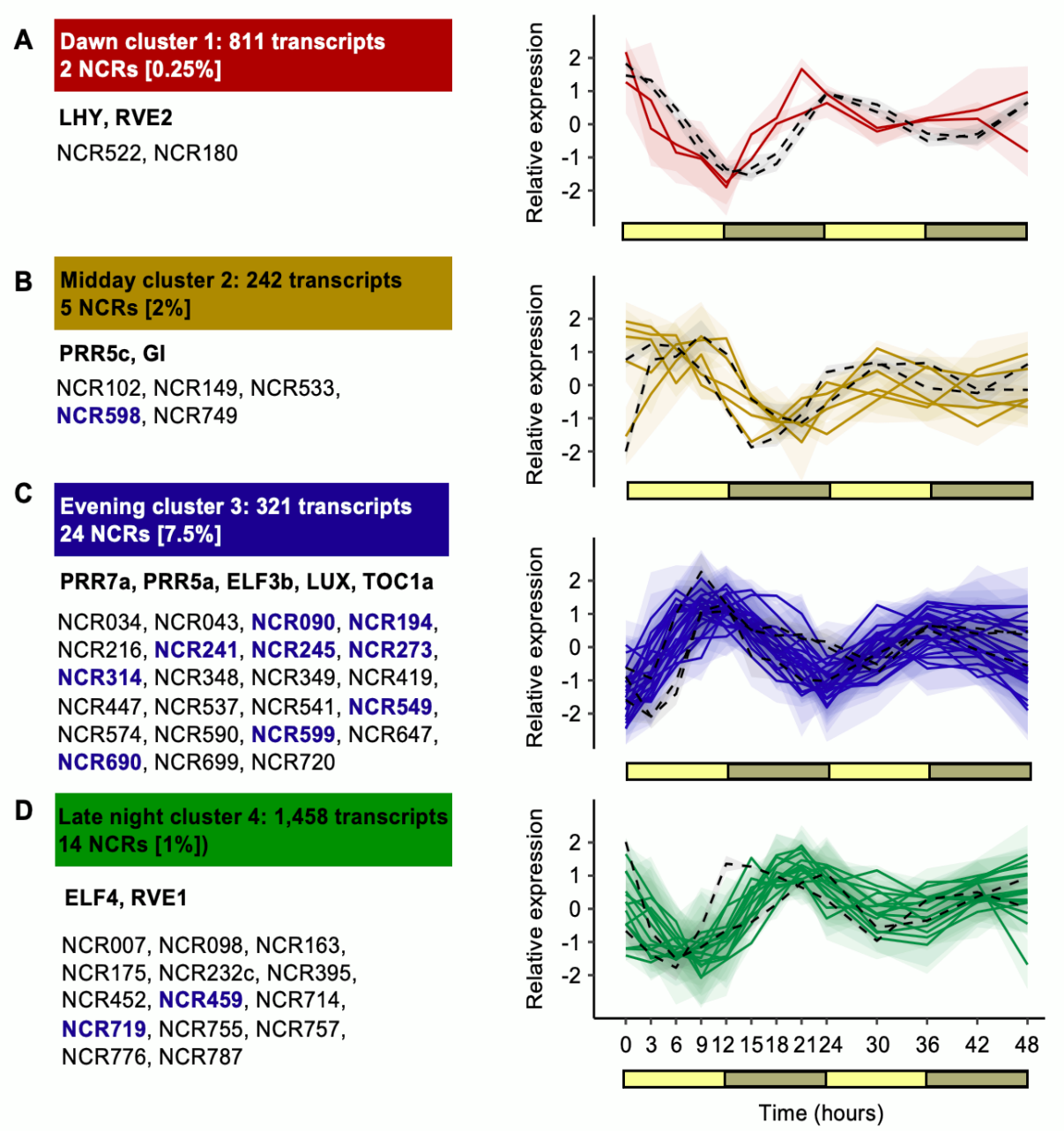


Figure 3.12: Rhythmic expression of NCR genes. (A-D) Expression of NCR genes in each cluster is shown by respective coloured lines. Dashed lines in the graphs show the expression of representative clock genes (bold black in text) belonging to the corresponding cluster. Thirteen rhythmic NCRs with EE motif are indicated with blue bold text. The expression range of each group of genes are shown with the coloured cloud. Expression values for each gene are log2 normalised and relative to its mean expression over time. Modified from Achom, Roy *et al* 2020.

Previous analysis in our lab (unpublished data) revealed that about 10.7% of NCR gene promoters contain the EE motif (p= 0.0047 for enrichment against a randomized set of non-NCR gene promoterss) while about 21.8% of them have the EE-like motif (p=0.00004 for enrichment). Among the 45 rhythmic NCR promoters, 13 (~29%) harbor

the EE motif which represent a significant enrichment of 2.7-fold (p = 0.0004) compared to the overall presence of this motif in NCR gene promoters. Interestingly, 10 out of those 13 rhythmic NCRs with an EE-motif in their promoter belong to dusk cluster. The EE-like motif is only present in 6 (~13%) oscillating NCR promoters which is lower and not significant compared to their relative high abundance in NCR gene promoters. Although many NCR gene promoters contain the EE-like motif, this motif is neither enriched in oscillating NCRs (Table 3.3) nor in other oscillating gene promoters (Table 3.2). Therefore, the EE-like motif might be involved with some other unknown regulation or could be involved in the evolution of EE motif in NCR promoters, but is unlikely to be involved directly in CCA1/LHY-mediated circadian regulation. Nonetheless, the presence of EE and EE-like motifs (which might share same evolutionary history) in NCR promoters and the enrichment of EE motif in rhythmic NCRs further link *MtLHY* to the regulation of at least some NCR genes.

Motif	#out of 45 oscillating NCR promoters	#out of 701 NCR gene promoters
AGATATTT	13 (28.9%)	75 (10.7%)
AGACATTT	6 (13.3%)	153 (21.8%)

Table 3.3: Frequency of the EE (AGATATTT) or EE-like (AGACATTT) motifs in NCR gene promoters. The EE motif is enriched in the oscillating NCR genes relative to their overall percentage among all NCR genes. No such enrichment was observed for the EE-like motif, although the EE-like motif is abundant in NCR promoters.

3.3. Discussion

The findings from the analysis of the time-course experiment in nodules has provided new insights regarding circadian regulation in mature nodules. Firstly, it has shown that the circadian clock operates within nodules in a similar (but not exact) manner to its regulation in other plants and other organs. Earlier studies in *A. thaliana* using genome-wide microarray reported a varying percentage of rhythmic transcripts under circadian control; ~6% by Harmer et al. (2000), ~10% by Covington and Harmer (2007) and about 16% by Edwards et al. (2006). However, after combining and reanalyzing these three datasets, Covington et al. (2008) suggested that about 30-40% genes might be under circadian control in *A. thaliana* seedlings. About 2,700 genes (~5.3%) were found to be under circadian regulation in *M. truncatula* nodules in our study which is much lower compared to *A. thaliana*. However, our data was from nodules and the number of oscillating genes could be much higher in other tissues, for example, in leaves where *LHY* expression was more pronounced and sensitive in response to external stimuli. It should also be kept in mind that the identification of rhythmic genes largely depends on the algorithms and parameters used for the detection. The above-mentioned studies in *A. thaliana* mainly used a method known as COSOPT, which is not very popular among researchers now-a-days for a number of reasons. It does not allow unevenly sampled data to be used, and it uses a curve-fitting approach that has problems in detecting rhythms that are not perfectly sinusoidal (Laloum and Robinson-Rechavi, 2020, Yang and Su, 2010). In our study, the Lomb-Scargle (LS) algorithm was used since it can deal with unevenly spaced sampling time-points, the use of replicates and missing data (as mentioned in Chapter 2, replicate 2 for 21h and replicate 3 for 15h time-points were discarded, therefore, missing in our analysis). Also, the LS method is conservative in detecting rhythmic genes and it usually outputs a comparatively smaller number of rhythmic genes but with low false positive results (Glynn et al., 2006, Laloum and Robinson-Rechavi, 2020). This could be another reason for seeing a comparatively small number of rhythmic genes. Nonetheless, while the analysis method used in our study could potentially fail to detect some other rhythmic genes, nonetheless a stringent small number of genes with a low false-positive rate, adds confidence to the data presented here.

One or multiple homologs of all key clock genes (e.g., *LHY, TOC1, PRRs, ELFs, GI*) has been found to be expressed in a rhythmic manner, suggesting these genes are functionally active in nodules. However, the comparison of *MtLHY* expression in different tissues suggests that the clock functioning might be slightly different in nodules compared to that of leaves and roots. The expression pattern of *LHY* in nodules was similar to the root expression in terms of its maximum and minimum expression levels (Figure 3.6). This is reasonable since nodules are root organs and some of the nodule developmental processes overlap with that of lateral roots (Bishopp and Bennett, 2019) (discussed later in Chapter 6). However, compared to the roots, *LHY* levels in nodules went down slowly (from its maximum at the dawn) during the day and reached its lowest point slightly after the root minimum. This suggests that nodules maintain a slightly higher level of *LHY* during the day relative to the root *LHY* level. The significance of this is not clear at this moment but further comparative studies (nodules vs roots or other organs) including analysis of LHY targets could be used to further investigate clock function in nodules. The overall level of *LHY* expression in leaves was much higher and the rhythmicity was more responsive to changing entraining signals (e.g., transfer from diurnal light/dark to constant light condition), compared to roots or nodules. These observations are consistent with the organ-specific variations in the clock functioning in *A. thaliana* (Bordage et al., 2016, Takahashi et al., 2015). The comparison of the expression of other clock and clock-regulated genes in different tissues of *M. truncatula* using qPCR or through generating different tissue-specific RNAseq gene expression datasets are likely to shed more light in this regard.

The oscillation patterns of rhythmic genes in nodules corroborates our common understanding about different processes that are regulated by the clock. For example, photosynthesis-associated genes were enriched amongst genes peaking at midday (the midday cluster) and starch degradation related genes were mostly found in the dusk-peaking cluster. As a nodule is an underground organ that has been evolved for a very specific purpose – housing rhizobial bacteria that enable N-fixation, it is unlikely that rhythmic expression of photosynthesis or carbohydrate metabolism plays a direct nodule-specific role. However, nodules are connected to the plant and must send/receive many molecules, and thus the detection of these commonly rhythmic processes in nodules is in line with their role as a metabolic factory for the plant. Similarly, almost all plant organs (flowers, fruits, roots, seeds; not only leaves) are found to have some degree of

photosynthetic activity (Brazel and Ó'Maoiléidigh, 2019). Therefore, appropriate detection of these universal plant rhythms in nodules is in line with our understanding of plant physiology.

Many nodulation-related GO terms were found to be enriched amongst rhythmic genes and a high percentage of rhythmic NCRs were found in the dusk cluster. These findings highlight the evening period as an important time for nodule-specific activities, which is in line with the fact that the nitrogen fixation rate is highest at the evening (Minchin and Pate, 1974). In addition, the amount of different N-containing compounds, specifically the concentrations of asparagine and glutamine that are used in N-transport, continuously increases during the night (Minchin and Pate, 1974). This again corroborates our finding that many nitrogen metabolism and transport related genes (e.g., nitrate transporter, amino acid transporter, glutamine synthetase) were found in the night and dawn clusters, meaning their expression increased during night and peaked around late night or dawn. Overall, our cluster-specific gene enrichment analysis in the context of existing literature suggest that nodules become highly active in fixing nitrogen around dawn. The fixed nitrogen is then transported and metabolised into other N-containing compounds mostly throughout the night.

Finally, and most remarkably, the identification of several rhythmic NCRs and their enrichment in the dusk cluster supports our hypothesis that *MtLHY* could be involved in the regulation of NCRs and in the nodulation process in general. This was further supported by the finding of enrichment of the EE motif (LHY binding site) in oscillating NCRs, particularly those that peaked at dusk. As discussed earlier in the chapter, plant immunity to pathogenic invaders varies throughout the day in a circadian manner. NCRs are known for their antimicrobial activities; therefore, they could be expressed in a high amount to provide an extra layer of defense in the nodule at times when plant general immunity is at a minimum; for example, Bhardwaj et al. (2011) found that *A. thaliana* leaves are most susceptible to infection at midnight. More insights could be gained through studies focused on these rhythmic NCRs to understand their functional role in nodulation. In regard to the MtLHY-NCR hypothesis, it is important to keep in mind that the EE motif serves as a binding site for some other proteins, including RVEs. *RVE*s are homologous to *LHY* and have a similar expression pattern to *LHY* (as shown in Figure 3.5). Therefore, the hypothesized link of MtLHY to NCRs could be only one layer of

interaction and RVEs might play important roles in this regulation. Therefore, studies involving the use of specific technologies to characterize protein-DNA binding (e.g., Chromatin Immunoprecipitation with massively parallel DNA-sequencing or ChIP-Seq) are needed to confirm if there is a true protein-DNA interaction between *Mt*LHY and NCRs, and then (or if not), what other proteins interact with this conserved EE motif in NCR promoters.

To conclude, our data provide a wealth of information and insights into nodule-specific circadian regulation of molecular processes. Similar experiments with other tissue or nodule samples from different developmental stages to generate comparative data would improve our understanding of the links between *LHY* and NCRs, as well as the extent of tissue specific and development-related spatial and temporal regulation related to the circadian clock.

4. The effect of *LHY* disruption on nodulation, growth and root development of *M. truncatula*

4.1. Introduction

The findings presented and discussed in Chapter 3 showed that many genes (about 5%) are under circadian regulation in nodules. Many metabolic and physiological processes within nodules are therefore likely to be regulated in a time-dependent manner. The comparison of rhythmic behaviour of the key circadian gene *MtLHY* in different organs (Figure 3.6) provides further evidence in support of the existence of a robust nodule clock which could be slightly different (in terms of rhythmic amplitudes, periods and sensitivity to entraining signals) in the nodule compared to other organs. It was suggested that the root clock is possibly entrained by shoot-derived signals (Bordage et al., 2016). Takahashi et al. (2015) demonstrated that the rhythms in roots can be altered by shoot apex excision or micrografting. In a similar way, the nodule clock can also be entrained by shoot or root clock and, in general, any disruption in clock functioning might affect nodulation (e.g., increased or decreased nodule number) and other plant developmental processes. Disrupting key clock genes is one way to characterize this.

In *A. thaliana,* mutations in either *CCA1* or *LHY* genes were reported to be associated with short-period rhythms, measured by the promoter activity of CHLOROPHYLL A/B BINDING (CAB) protein. When both *CCA1* and *LHY* were defective, the period became even shorter (19.73 ± 0.12 h, compared to 24.77 ± 0.07 h in *cca1* mutant and 26.41 ± 0.16 h in wild type) (Lu et al., 2009). Plants harbouring defective *lhy* genes had shown an increased expression of the important flowering gene, *FLOWERING LOCUS T* (*FT*), and an accelerated flowering phenotype (Park et al., 2016). Loss of *CCA1* or *LHY* gene expression also results in reduced iron (Fe) uptake and photosynthetic efficiency, and reciprocally, overexpression of *CCA1* results in opposite effects, with better growth in Fe-deficient conditions (Xu et al., 2019).

It has been found that clock genes also play important roles in the control of seed dormancy. Interestingly, *cca1* and *lhy* mutants were associated with reduced seed dormancy (high germination rate) whereas the *gi* (*gigantea*) mutant showed strong seed dormancy (low germination) (Penfield and Hall, 2009). It had been demonstrated that the

transcriptional clock is halted in an evening-like state in dry *Arabidopsis* seeds where different evening-expressed genes, e.g., *TOC1*, *GI*, *LUX*, were found in high levels and *CCA1* and *LHY* levels were negligible. Upon imbibition, the seed clock restarts and potentially plays critical role in breaking seed dormancy (Penfield and Hall, 2009). This evidence clearly suggests that proper clock functioning plays an important role in different stages of plant development, from seed germination to flowering. Clock mutant *A. thaliana* plants were also reported to have an altered microbiome population structure in their rhizosphere (discussed later in Chapter 5) (Hubbard et al., 2018). Therefore, the plant clock might play roles in shaping different types of plant-microbe interactions (e.g., mutualistic or pathogenic) too.

The effect of clock gene mutations has not been extensively studied in legumes. Some insights have come from a few studies carried out in legume crop plants (e.g., peas, soybeans). In garden peas (*Pisum sativum*), a *gi* mutant allele was shown to alter the expression of different other clock genes, prevent the induction of *FT* and confer a late-flowering phenotype (Hecht et al., 2007). On the other hand, *elf4* mutant pea plants exhibit early flowering (Liew et al., 2009). Comparison among wild type (WT) and domesticated varieties of soybean suggests that different clock components are associated with different important agronomic traits like sensitivity to day/night cycle, maturation time and yield (Li and Lam, 2020). For example, mutations in *PRR3* genes have been associated with early maturation and flowering of soybean plants (Li and Lam, 2020, Lu et al., 2020).

Apart from some of these observations in legume crops, we know much less about clock functioning in legumes. In particular, there have not been many studies in the legume models, e.g., *M. truncatula*, *L. japonicus*. In this study, we evaluated the phenotypes (inc. shoot growth, nodule mass, leaf movement) of two *lhy* mutant lines obtained from a collection of transposon insertion lines from the Noble Research Institute, USA and compared them with WT plants. We also asked if the expression of some key clock genes was rhythmic in constant environmental conditions in *lhy* mutant plants in order to evaluate whether *lhy* mutation alters rhythmicity of other clock components. The results presented in the next sections offer valuable insights into the effect of clock malfunctioning on the development of the legume-rhizobial interaction and plant growth, and open-up new research possibilities for understanding the impact of the plant clock on symbiosis.

4.2. Results

4.2.1. Analysis of gene expression levels using quantitative PCR confirms reduced expression of *LHY* gene in two *lhy* mutants

In order to disrupt *LHY* expression, two mutant lines were obtained from the Noble Research Institute (Tadege et al., 2008b). The lines contain Tnt1 retrotransposon insertions, respectively, in the promoter region (*lhy-1*) and at the end of 5th exon (*lhy-2*) of Mt*LHY* gene (Figure 4.1a). The obtained mutant lines were genotyped in our lab previously and only the plants that were homozygous for Tnt1 insertions at *LHY* locus were propagated for seed stock generation (Achom, 2019). To determine the level of *lhy* expression in mutants, qPCR was carried out using leaf samples from each mutant, compared to the parental R108 line. Leaf samples were collected from 4-5 plants at two different times, early in the morning (~3h past dawn) and at the evening when *LHY* expression is at its highest and lowest, respectively. Two sets of primers were used - the

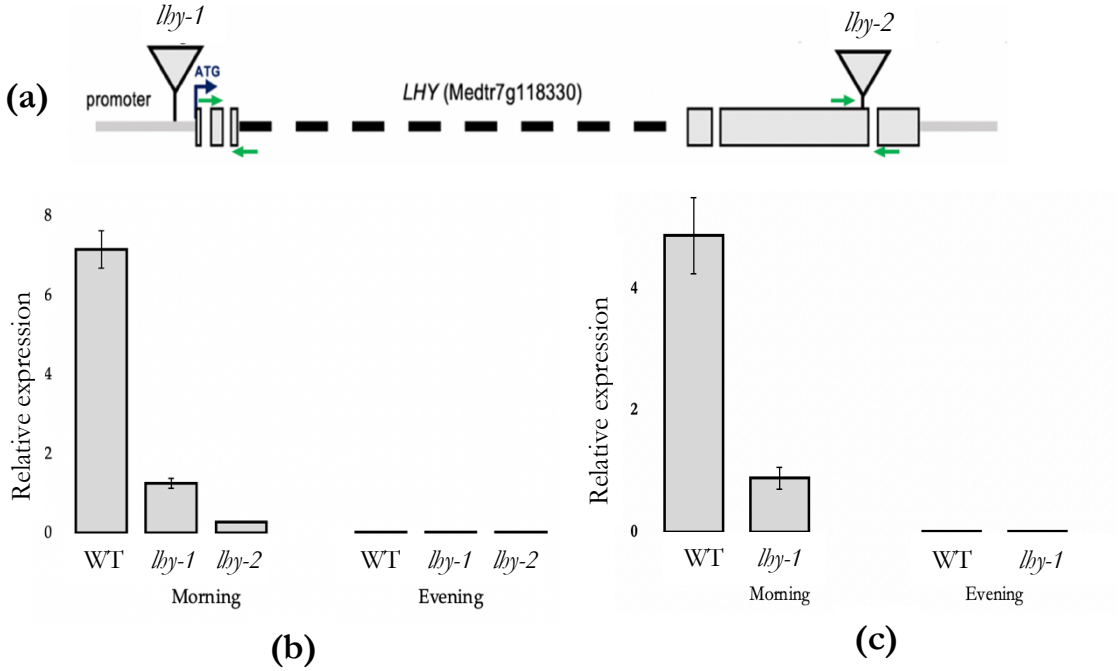


Figure 4.1: (a) Schematic representation of *LHY* gene (not exactly scaled) and the approximate location of Tnt1 insertions (triangles) at *lhy-1* and *lhy-2* lines. Exons are presented as grey rectangles and introns as spaces/dashed line. qPCR primer positions are shown as green arrows. (**b) Quantification of *LHY* expression using primers based on sequence between 2nd and 3rd exons. (c) Quantification of *lhy* expression using primers based on sequence between 5th and 6th exons**. qPCR did not work in *lhy-2* with this primer pair since the binding site was disrupted by Tnt1 insertion. The expression values are relative to beta-tubulin (Medtr7g089120). Error bars represent the standard deviation of the mean between three technical replicates.

first pair designed based on the sequence between the 2nd and 3rd exon boundaries (Figure 4.1a & b) and also used for the quantification of *LHY* expression in WT leaves, roots and nodules (Chapter 3) and the second pair designed between the 5th and 6th exon boundaries (Figure 4.1b) based on the Tnt1 insertion site (*lhy-2*).

As expected, the qPCR data showed high expression of *LHY* in the morning and negligible expression at the evening (Figure 4.1b & c). *LHY* expression was several fold higher in WT plants compared to either mutants. Between the mutants, *lhy-1* plants had higher *LHY* expression than *lhy-2* (Figure 4.1b). Since the *lhy-1* has a Tnt1 insertion in the promoter region, in theory, it could express functional and intact *LHY* transcript, but the promoter disruption affected the expression which was significantly lower than WT (Figure 4.1b & c). In the *lhy-2* line, a partially or non-functional transcript is likely to be produced since the Tnt1 insertion disrupts the coding region. Since LHY binds to its own promoter and is mainly regulated by feedback mechanism (as discussed in Chapter-1), functional LHY is essential for its proper expression. This could be the reason for further lower *lhy* expression in *lhy-2* compared to *lhy-1*. When the second primer pair (designed based on 5th and 6th exons) was used (Figure 4.1c), no expression could be detected in *lhy-2*, this confirms that this part of coding region was disrupted by Tnt1 insertion.

4.2.2. *lhy* mutants show reduced nodulation and reduced vegetative growth

To evaluate the effect of *lhy* mutation on nodulation, the *lhy* mutant seedlings along with WT (R108) were inoculated with *Sinorhizobium meliloti* 1022 and grown for five weeks in low nitrogen conditions. At the end of the five weeks, plants were uprooted and fresh shoot, root and nodule weights were measured. Rhizobial-inoculated WT plants had a significantly greater shoot and root compared to both mutant lines. The nodule size and overall weight are also significantly higher for the WT plants (Figure 4.2, 4.3 and Table 4.1). Such a difference was not however observed in mock-inoculated plants that had not been inoculated with rhizobia. All mock-inoculated plants (WT vs mutants) had shown similar shoot sizes and had not exhibited much growth since they were in low nitrogen conditions, without beneficial rhizobial inoculation, therefore, no nodulation. Interestingly, a higher percentage (Table-4.1) of *lhy-2* plants had died in both treatments (significantly higher in mock treatment; Fisher Exact test, *p*-value: 0.0252;). The death rate of *lhy-1* is also higher in mock treatment but not significant. Therefore, mutant plants

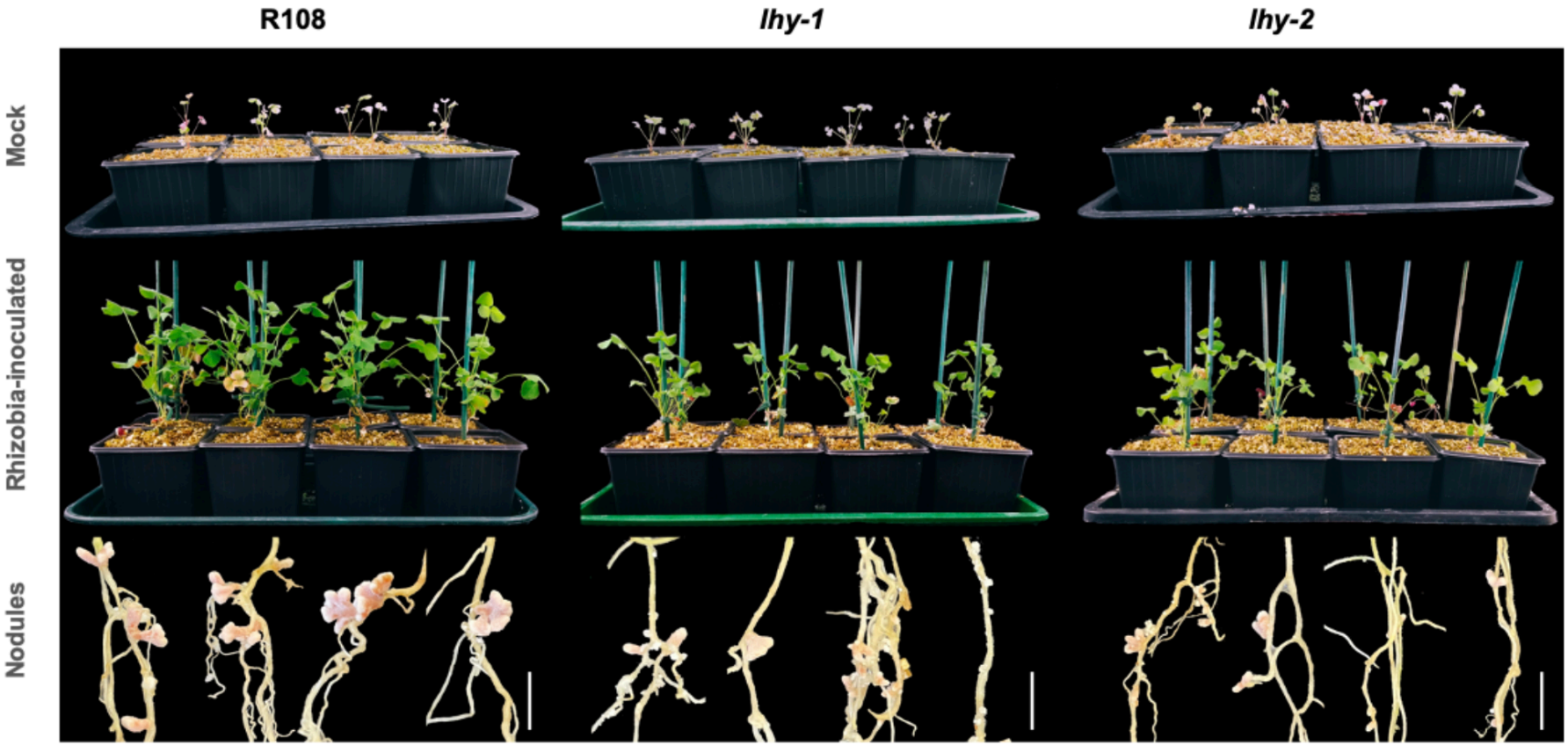


Figure 4.2: Loss of *LHY* expression affects plant growth and nodulation. Plants grown on perlite-vermiculite pots have a similar phenotype in the absence of rhizobial inoculation (top row). *lhy* mutant lines have reduced growth compared to their WT counterpart when grown with rhizobial inoculation (middle row). Nodules in mutant plants were comparatively smaller in size but similar in appearance (bottom row). N-fixation efficiency of nodules was not evaluated in the study.

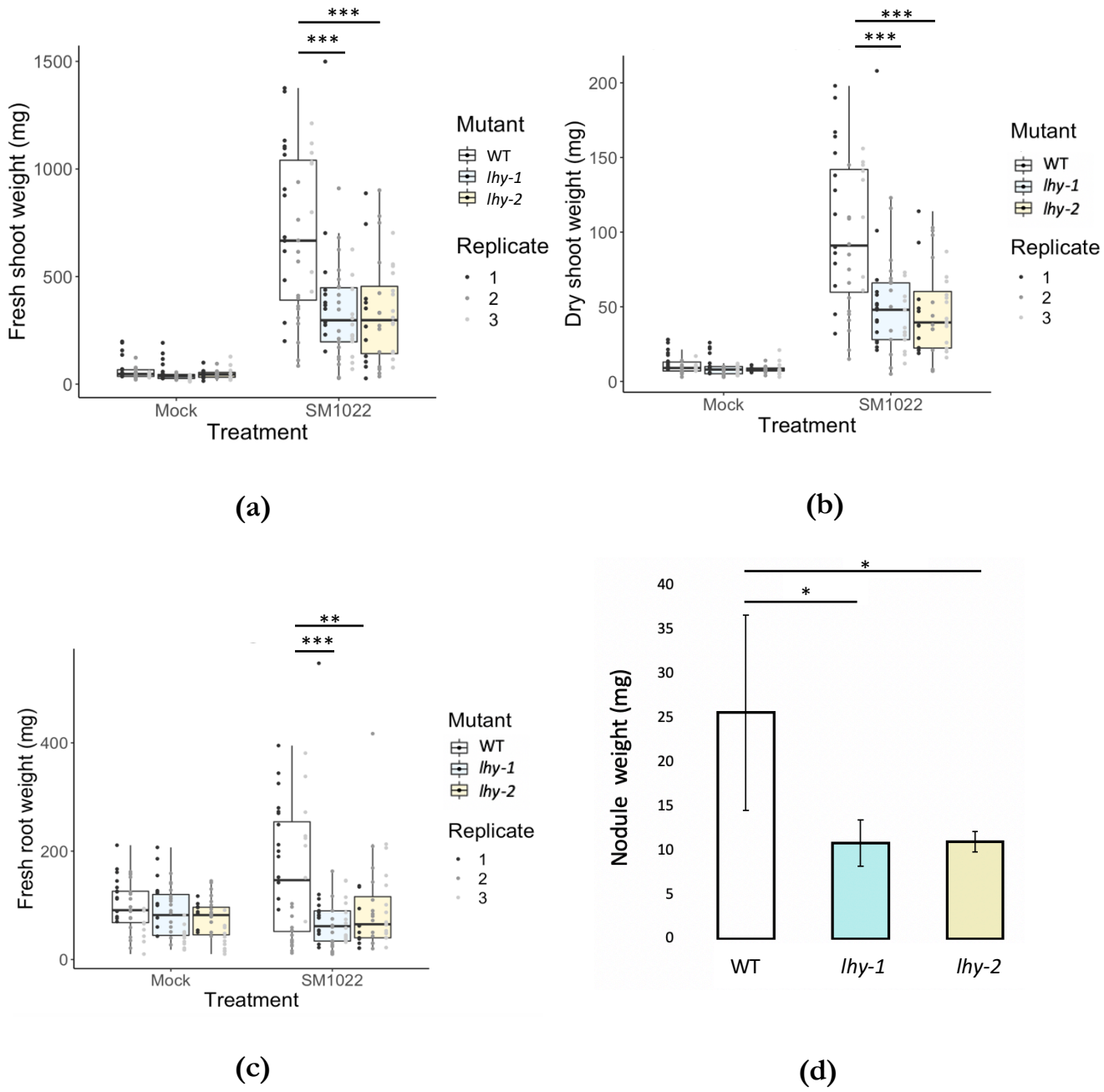


Figure 4.3: Reduced *LHY* expression affects plant growth and nodulation. **(a) & (b)** Fresh and dry shoot weights from *S. meliloti* 1022 (SM1022) and mock-inoculated plants. Data from three replicate experiments are shown separately in each representative box. Compared to mock-inoculated plants, rhizobial-inoculated plants were significantly larger. Amongst them, WT plants were significantly larger than the mutants. **(c)** Fresh root weights were also significantly higher in rhizobial-inoculated WT plants compared to other plants, however no significant difference was found between any other conditions or lines, suggesting that *lhy* root growth is similar between conditions. **(d)** Average nodule weight per plant suggests that nodulation was also affected in *lhy* mutant plants. Error bar represent 95% CI from three replicate experiments. One asterisk (*) indicates $p < 0.05$; two (**) or three (***) indicates $p < 0.01$ and 0.001, respectively.

	SM1022 inoculated		
	Wild type	***lhy-1***	***lhy-2***
Seedlings sown	40	48	48
Plants sampled	36 (90%)	46 (96%)	40 (83%)
Plants died (before sampling)	4 (10%)	2 (4%)	8 (17%)
Fresh shoot weight (g)	0.698 ± 0.121	0.348 ± 0.076	0.343 ± 0.076
Dry shoot weight (g)	0.099 ± 0.016	0.051 ± 0.010	0.047 ± 0.009
Fresh root weight (g)	0.16 ± 0.038	0.076 ± 0.023	0.090 ± 0.023
Fresh nodule weight (g)	0.026 ± 0.011	0.011 ± 0.002	0.011 ± 0.001
	Mock treated		
	Wild type	***lhy-1***	***lhy-2***
Seedlings sown	40	48	48
Plants sampled	37 (93%)	39 (81%)	35 (73%)
Plants died (before sampling)	3 (7%)	9 (19%)	13 (27%)
Fresh shoot weight (g)	0.062 ± 0.119	0.044 ± 0.082	0.047 ± 0.081
Dry shoot weight	0.012 ± 0.002	0.009 ± 0.002	0.008 ± 0.001
Fresh root weight	0.098 ± 0.015	0.086 ± 0.015	0.073 ± 0.012
Fresh nodule weight	n/a	n/a	n/a

Table 4.1: Summary of all data collected in mutant phenotyping experiment.

(specially, *lhy-2* line) might have some growth associated impairment which is relatively more expressive (considering less survival rate) in mock treatment

In terms of root growth, the difference between mock and rhizobial-inoculated plants was less pronounced (in comparison to the difference in shoot growth) (Figure 4.3c). There was no significant difference in root growth for either of the mutant plants between mock and rhizobial inoculated, however, WT plants had a significantly larger root system in the presence of rhizobia compared to without rhizobia, or compared to the *lhy* mutants in both inoculated/mock-inoculated conditions. Therefore, it can be concluded that proper *LHY* functioning is important for both shoot and root growth. These observations potentially indicate that reduced or disrupted *lhy* expression affects nodule formation, resulting in reduced nodule mass which ultimately causes reduced plant growth.

4.2.3. Plant rhythmic leaf movements are affected by mutation in *lhy*

Like most other plants *M. truncatula* leaves display rhythmic nastic movement in a circadian manner where leaflets are in a horizontal (open) position under light and are in a vertical (closed) position in the dark (Chen et al., 2012). To determine whether Rhythmic Leaf Movements (RLM) are affected by *lhy* mutation, the rhythmicity of the *lhy* mutants were examined by measuring 'visible leaf area' as a measure of opening and closing of leaves (Figure 4.4). This was measured for 2 week old *lhy-1*, *lhy-2* and WT (R108) plants (when the second true leaf has fully opened) over a 120 hour period. True rhythmicity can be seen in plants if they maintain RLM in constant light ('free-running' conditions), thus plants were initially grown in diurnal light-dark cycles, then transferred to constant light conditions before measurements were taken. The leaf area of the second true leaf was measured for consistency across plants (See Chapter 2, Section 2.5.5 for methodological details).

Analysis of RLM data suggests, for the WT plants (n=7) there was a mean periodicity of ~25h over 120 hours. However, in the case of *lhy-1* (n=14) and *lhy-2* (n=12) plants there was a periodicity of about 18h and 12h, respectively, for first 48 hours, then both *lhy* mutants became more or less arrhythmic over the latter 48 hours. These findings suggest that *LHY* plays an important role in maintaining regular leaf movement, at least in free-running conditions.

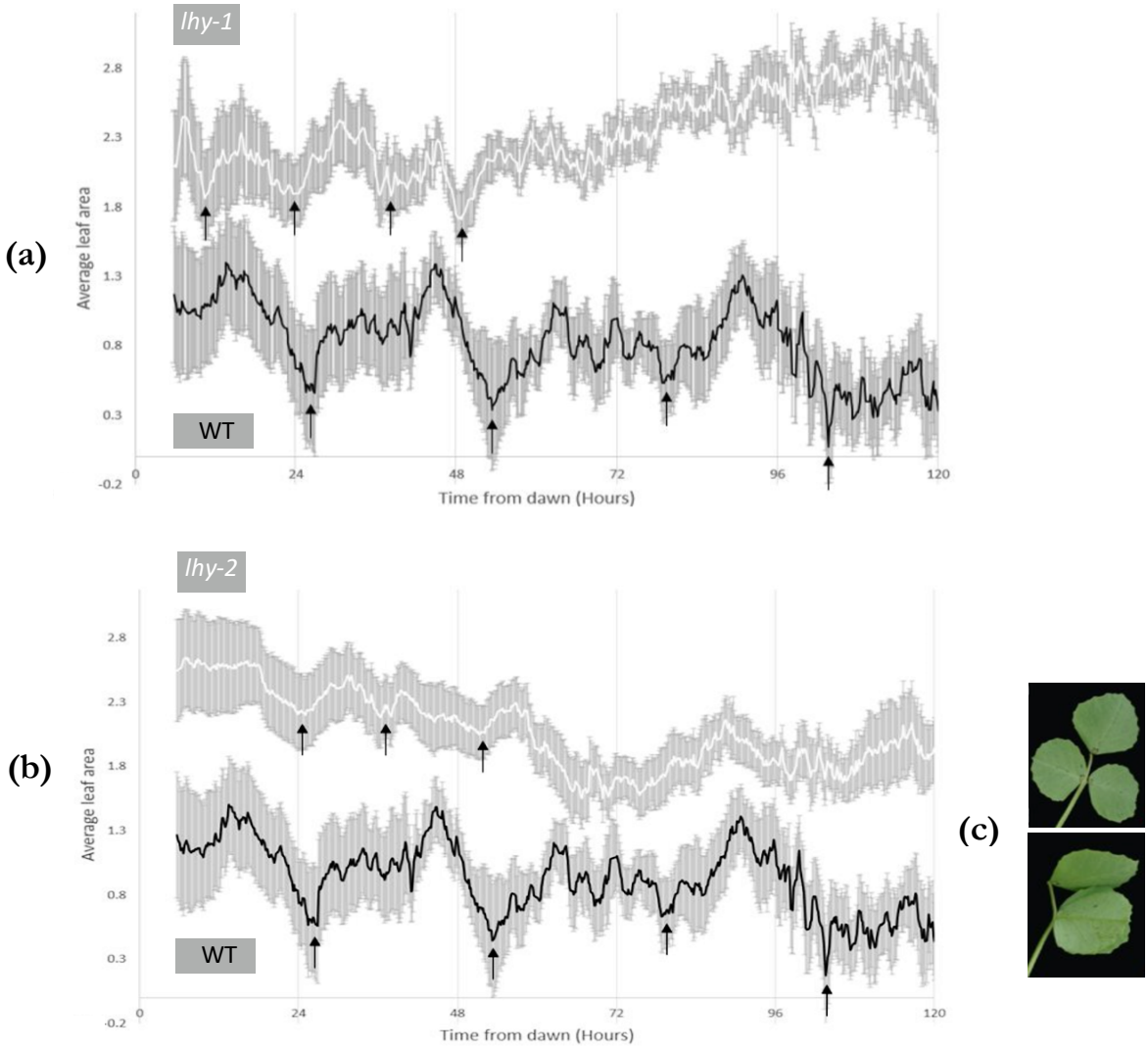


Figure 4.4: *lhy* mutation affects circadian leaf movements. Leaf opening and closing was monitored by measuring the visible leaf area from the top of the rosette of each plant and measurements over different time-points were normalized to the mean. The rhythmicity of *lhy-1* **(a)** & *lhy-2* **(b)** (white line) leaves initially has a shorter period and lower amplitude; black arrows (placed tentatively based on the visual observation of the data) indicate the time when average leaf area is its lowest (closed) at a particular rhythmic cycle. **(c)** Leaves of WT plants in open (top) and closed (bottom) position. After 48 hours the *lhy* mutants (**a & b**) became generally arrhythmic. WT R108 (black line) leaves, in general, maintained a rhythmic movement in a circadian manner over 120 hours in constant light.

4.2.4. Expression patterns of circadian clock genes suggest non-persistent rhythmicity at molecular level in *lhy* mutants

To evaluate the effect of *LHY* expression on the expression of other key clock genes, a time series experiment similar to that used earlier in this thesis (discussed in Chapter 3) was carried out with the *lhy-1* mutant line. Nodule samples were collected at different time-points (after every 3 hours) over 48 hours after transfer to constant light and temperature conditions (after being grown in light/dark cycling conditions); three biological replicates were sampled. However, instead of carrying out RNA sequencing, qPCR was used to analyse the expression of *LHY, ELF4* and *PRR5c* in the nodules of *lhy-1* mutant plants over 48 hours in free running conditions. The qPCR values were calculated relative to β-tubulin and therefore not a measure of absolute expression of the corresponding genes. Instead, these values were compared with corresponding WT RNA sequencing data values (from the previous time series experiment presented in Chapter 3) to determine if there is any difference in the rhythmic behaviour of these genes in the *lhy-1* mutant compared to WT (R108) (Figure 4.5). It would be better if WT *M. truncatula* R108 plants could be grown alongside the mutants as control for comparison, but we did not have enough WT R108 seeds then to set an entire time-course.

In constant conditions, the clock genes lost their rhythmic characteristics after the first cycle (the first 24 hours) in *lhy-1* mutant plants. In the first 12 hours, the expression of all three tested genes (*LHY, PRR5c, ELF4*) in mutants roughly followed the expression patterns in WT plants. However, with the passage of time, clock gene expression in *lhy-1* started to deviate more from WT. In the second subjective cycle (second 24 hours), the three clock genes in *lhy-1* plants had comparatively flatter and non-rhythmic expression trends. These observations suggest that the Tnt insertion in the *LHY* gene affects its own expression, and that reduced *LHY* expression dampens the rhythmicity of other clock genes. It is therefore also likely to disturb the expression of other rhythmic components at the gene expression and thus cellular and physiological level.

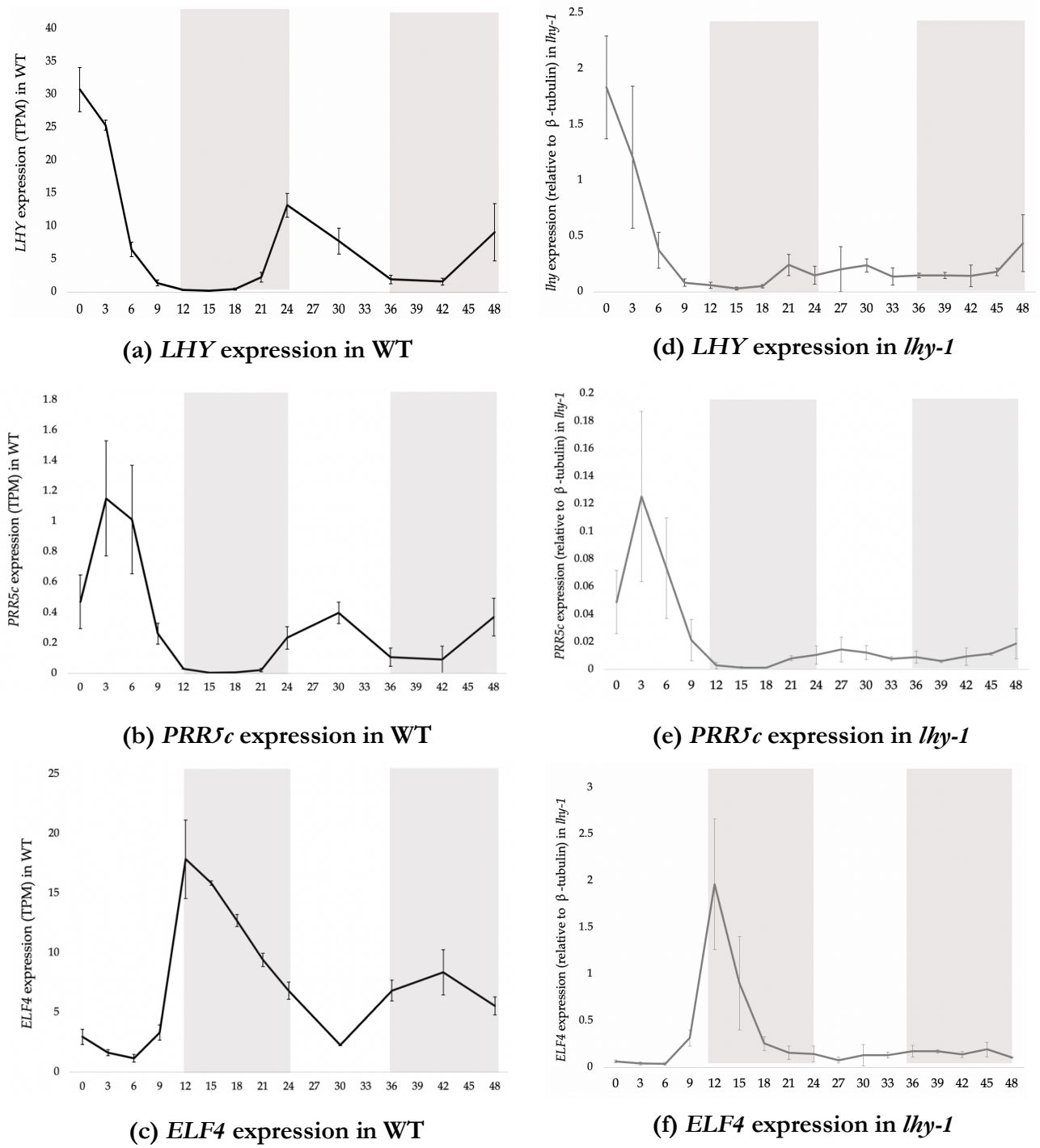


Figure 4.5: Comparison of *LHY* (Medtr7g118330)**, *PRR5c*** (Medtr8g024260) **and *ELF4*** (Medtr3g070490) **expressions in nodules of WT (a,b,c) vs *lhy-1* mutant (d,e,f) plants**. Under free running conditions the expression of *lhy* **(d)**, *PRR5c* **(e)** and *ELF4* **(f)** became almost arrhythmic after the first 24 hour period in *lhy-1* plants where in WT the genes maintained characteristic dirunal rhythms **(a,b,c),** though with lower amplitude in the second 24 hour subjective cycle. The expression values for WT (A17 ecotype) are from RNAseq experiment and are presented in Transcripts Per Million (TPM) units, whereas expression in *lhy-1* (generated on R108 background) is from qPCR data relative to β tubulin levels. These graphs are, therefore, not a comparison of absolute expression or level but they can be used to roughly compare the rhythmic behaviours (e.g., periods, phases) of these key clock genes between *lhy-1* and WT A17 plants. Grey shading designates subjective night. Error bars represent standard deviation of the mean from the three replicate experiments.

4.3. Discussion and future directions

The phenotypic observations and molecular data that are presented in this chapter strongly suggest that both the reduction (*lhy-1*) or loss (*lhy-2*) of fully functional MtLHY affects nodulation, plant growth and many rhythmic processes in *M. truncatula*. These range from the molecular (e.g., expression of other clock genes) to the physiological/developmental level (e.g., nyctinastic leaf movements). Since many genes are rhythmically expressed in nodules and the circadian clock is assumed to play an important role in those rhythmic processes (as discussed in Chapter 3), diminished *LHY* activity resulting in reduced nodulation makes sense. The retarded growth phenotype of *lhy* plants could simply be a consequence of decreased nitrogen assimilation from reduced nodulation. However, there might be other factors that could directly or indirectly affect plant growth when a properly functioning circadian clock is absent. For example, circadian clock-regulated nyctinastic leaf movements are also important for plant growth and development. It was reported that the overall plant growth was reduced in *M. truncatula* plants where leaf movements were disrupted due to mutations in *PETIOLULE-LIKE PULVINUS* (*PLP*) genes (Zhou et al., 2012). This is likely because plants can optimize light intensity on the leaf surface to maximise energy capture through nastic leaf movements. Therefore, the growth impairment in our experiment could be partly because of the altered leaf movements in the *lhy* mutant plants. There were no apparent or visual differences between leaf or shoot morphology (e.g., size, colour, branching) that could affect growth and development in *lhy* mutants, however, this could be investigated in more detail in future.

In general, the molecular and phenotypic alterations that were identified in our study corroborate existing literature on clock effects on plant growth and development. Our data is also consistent with a recently published study (Kong et al., 2020) that also reported that the loss of function of the Mt*LHY* resulted in decreased nodule number, irregular leaf movements and overall reduction in plant biomass. In addition, Kong et al. (2020) reported that *lhy* mutants showed an early flowering phenotype whereas the constitutive overexpression of functional *LHY* gene (through introducing a p35S:Mt*LHY* construct in WT plants) caused delayed flowering and longer hypocotyl growth. These findings are in line with the data from *A. thaliana* that suggest Mt*LHY* plays a role similar to *CCA1/LHY* of *A. thaliana* in terms of flowering and growth (Lu et al., 2009, Wang and Tobin, 1998). Constitutive overexpression of LHY, however, did not result in any

significant changes in nodule development or their number (Kong et al., 2020). Although flowering time and the effect of *LHY* overexpression were not evaluated in our study, one can expect to see similar findings as reported by Kong et al. (2020) due to the similar nature of the conditions used.

Having said this, the mutant lines that Kong et al. used were not same as ours, as far as we can tell based on the information that they report. They used two Tnt1 insertion lines, named NF6569 and NF16126, that were obtained (similarly to us) from the Noble Research Institute, USA. According to their paper, both these lines harbour Tnt1 insertion in the fifth exon of Mt*LHY* (Medtr7g118330). However, when we assessed the Tnt1 insertion sites using the respective Noble Research Institute database (https://medicago-mutant.noble.org/mutant/database.php), we did not find any insertion at any position in the entire Mt*LHY* gene for those two Tnt1 lines. This has created confusion about whether Kong et al. used the appropriate mutants, although their genotyping data and related PCR primer sequences were consistent with the data that we have presented for NF17115 (*lhy-1*) and NF16461 (*lhy-2*). There could therefore be a typographical error in their paper since our major findings are similar in general, but this is an unresolved issue.

Our findings together with those of Kong et al. (2020) open up possibilities for new and exciting research on the effect of the plant clock on symbiosis. Since we have confirmed that nodulation is affected by the loss of *LHY* expression, it would be interesting to monitor nodule initiation and further development in WT vs. *lhy* mutants using microscopy and study gene expression at different stages of nodule growth in order to determine when and where *LHY* disruption affects nodulation. Also, based on our hypothesis on Chapter 3, the regulation of different NCRs should be evaluated in the *lhy* mutants to determine whether *LHY* directly regulates a set of NCRs or if the rhythmic NCRs are regulated by some other clock components. Phenotyping experiments similar to those discussed in this chapter, should be set up with other *M. truncatula* clock gene mutants (e.g., *TOC, PRRs, GI*) to determine whether the effects we have identified in this study relate solely to the function of Mt*LHY* or relate to the clock in general. To conclude, this chapter presents novel and exciting groundwork including some new insights into the role of Mt*LHY,* as (potentially) corroborated by published data, which could be used for further research to understand more about the circadian regulation of nodulation in legumes.

5. The rhythmic regulation of rhizobial genes in nodules.

5.1. Introduction

In a successful legume-rhizobia symbiosis, rhizobia undergo different metabolic and physiological changes within host symbiosomes and differentiate into N-fixing bacteroids (as discussed in Chapter 1). Therefore, it can be assumed that the gene expression profile of fully differentiated bacteroid is different than that of undifferentiated bacteroids or free-living rhizobia. This is supported by the study in *Aeschynomene spp.* (Lamouche et al., 2019b, Lamouche et al., 2019a), where several transcriptomic and metabolomic changes were observed between differentiated and undifferentiated bacteroids. The transcriptomic analysis further suggests that oxygenic and redox regulations play important role in bacterial gene expression (Lamouche et al., 2019a). It is likely that the expression of many bacteroid genes are influenced by the host. As shown in Chapter 3, many plant genes are likely to be under circadian regulation and are rhythmically expressed in nodules. It is plausible that some rhizobial genes are also rhythmically regulated inside nodules. In indeterminate nodules where bacteroids are terminally differentiated and comparatively tightly regulated by the host, plant circadian control may have a particularly important role in regulation of bacterial gene expression.

5.1.1. A circadian clock is present in certain photosynthetic bacteria

The existence of circadian clocks is well described in plants, animals and other eukaryotic systems. However, although circadian systems are apparently ubiquitous, they have rarely been reported in prokaryotes except for some photosynthetic bacteria (reviewed in Sartor et al. (2019)). It has been previously thought that a circadian system might not be present in prokaryotic organisms that have reproductive times shorter than the diurnal cycle. This idea was referred to as the "circadian-infradian rule" which presumed that a cell can incorporate circadian regulation only if it reproduces in a circadian (~24 hours cycle) or infradian (more than 24 hours cycle) fashion (Hillman, 1976). However, several studies on photosynthetic cyanobacteria, in particular on *Synechococcus elongatus*, provided strong evidence of a robust circadian clock existing within prokaryotic systems with a generation times of 5-10 hours (Kondo et al., 1994, Liu et al., 1995, Nakajima et al., 2005, Sartor et al., 2019), as outlined below.

In *Synechococcus spp.*, three essential circadian clock proteins, namely- KaiA, KaiB and KaiC, have been described (Figure 5.1) that can directly or indirectly regulate the expression of almost entire bacterial genome (Liu et al., 1995, Nakahira et al., 2004). This is much higher in terms of the proportion of genes under circadian control than any other eukaryotic model system. Among the Kai proteins, KaiC is considered to be the most important one and it is phosphorylated and dephosphorylated in a circadian manner (Tomita et al., 2005). It was first proposed that the core circadian loop is maintained by the negative-feedback regulation of the *kaiBC* operon by Kai proteins (Nakahira et al., 2004). However, later it was found that the temperature compensated, free-running, circadian cycling of the phosphorylation state of KaiC is tightly maintained even when both transcription and translation of *kaiBC* operon were inhibited (Tomita et al., 2005). This suggest that the core oscillatory loop is not formed by transcription-translation feedback regulation of the *kaiBC* operon but by the phosphorylation cycle of KaiC itself. Interestingly, KaiC is capable of both autophosphorylation and autodephosphorylation where KaiA promotes the autophosphorylation of KaiC (Iwasaki et al., 2002) during the day and KaiB antagonizes the effect of KaiA at night (Figure 5.1) (Kitayama et al., 2003, Xu et al., 2003). These findings suggest that an autonomous, self-sustaining oscillation of KaiC phosphorylation could be generated in the presence of KaiA and KaiB only, without requiring other kinases or phosphatases. In fact, the oscillatory phosphorylation of KaiC was retained when all of these proteins were ectopically expressed in *E. coli*. In *E. coli*, the oscillation of KaiC phosphorylation was observed only when both KaiA and KaiC were

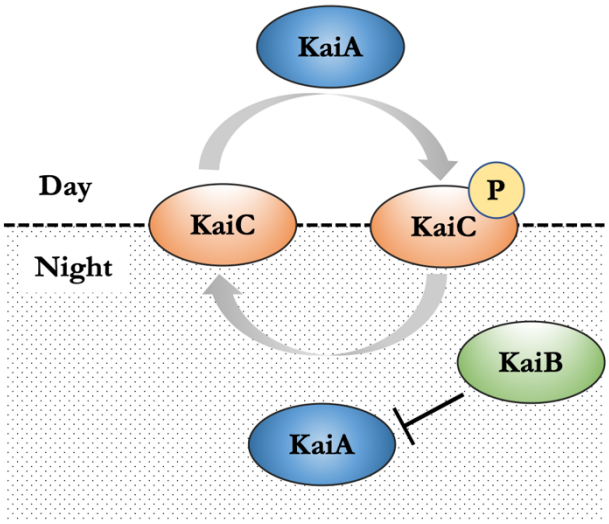


Figure 5.1: KaiA and KaiB antagonize each other to control the KaiC phosphorylation state. During the day, KaiA enhances the autophosphorylation of KaiC and the concentration of phosphorylated KaiC reaches its maximum at dusk. At night, KaiB antagonizes KaiA, the dephosphorylation of KaiC takes place rapidly and the concentration of dephosohorylated KaiC become maximum at dawn.

present and it was not dependent on KaiC expression levels, as the total KaiC expression did not show any rhythmic pattern over time (Chen et al., 2015a).

Rhythmic processes have been identified in photosynthetic purple bacteria too. For example, in *Rhodobacter sphaeroides*, an acyl-CoA dehydrogenase encoding gene (*aidB*) oscillates transcriptionally with a period of around 20.5h under aerobic conditions or with a much shorter period of 10-12h under anaerobic conditions (Min et al., 2005). However, these rhythms do not meet the criteria of a circadian rhythm as they are much shorter than 24h. Nonetheless, the *R. sphaeroides* genome contains orthologs of *kaiB* and *kaiC* genes (lacking *kaiA*). Furthermore, the *kaiBC* operon in *R. sphaeroides* contains a PAS (Per ARNT Sim) domain protein. PAS domain-containing proteins are often found to be associated with circadian clocks in different species. For example, in *A. thaliana* three clock associated proteins, ZEITLUPE (ZTL), FLAVIN-KELCH-FBOX-1 (FKF1), and LOV-KELCH-PROTEIN-2 (LKP2), contain an N-terminal PAS domain (Vogt and Schippers, 2015). The PAS domain acts as a versatile sensor and can detect different chemical and physical signals that are essential in circadian regulations (Möglich et al., 2009). Based on these observations, it can be assumed that even if there is no typical circadian program in *R. sphaeroides,* it partially possesses some of the clock components and this provides evidence for the existence of endogenous temporal programs in prokaryotes.

All of the findings mentioned above demonstrate a number of facts. Firstly, circadian regulation can persist in single-cell organisms that reproduce continuously in less than 24 hours. Secondly, circadian clocks are not only present in certain prokaryotes but they can also be transplanted successfully into other prokaryotic systems that lack circadian activity as described by Chen et al. (2015a). Finally, key clock components are not exclusively present in species with known circadian regulation mechanisms, but also found in other species (discussed further, later). This suggests that these genes are evolutionarily significant and might be involved in other known or as yet unknown rhythmic processes.

5.1.2. Presence of circadian rhythms in non-photosynthetic bacteria

As mentioned earlier, the existence of a robust circadian regulation has not yet been reported in non-photosynthetic bacteria. However, rhythmicity in growth characteristics

has been reported a long time ago in *E. coli* (Halberg and Conner, 1961, Sturtevant, 1973a) and *Klebsiella pneumoniae* (Sturtevant, 1973b). Despite this, these rhythms failed to persist in a temperature-compensated or free-running manner, therefore, could not be treated as a true circadian rhythm.

Recently it has been demonstrated that *Enterobactor aerogenes*, a commensal from human gastrointestinal track, exhibits rhythmic patterns of swarming and motility (Paulose et al., 2016). In the presence of melatonin, an indole hormone abundantly present in the human gut, a synchronized, temperature-compensated rhythm of luciferase (fused to a flagellar motor protein) activity with a period of 25h was observed. This is, so far, the only reported rhythmic activity in non-photosynthetic prokaryotic species that meets the criteria of a classically defined circadian rhythm. However, the key components behind this rhythmicity and other related regulation (if any) are yet to be understood. Also, in this study (Paulose et al., 2016) it was noted that only 31-44% of the bacterial culture harbouring the luciferase reporter construct had shown rhythmicity, implicating conditional characteristics of the rhythm. Nonetheless, this observation presents, at least one example of how bacterial time-keeping machinery could be conditioned by a host derived signal, in this case, melatonin.

Interestingly, *kai* gene homologs have been found in non-photosynthetic bacteria (Dvornyk et al., 2003, Loza-Correa et al., 2010). Of particular note, the homologs of *kaiC* have been identified in three other major bacterial taxa outside cyanobacteria, namely- Proteobacteria, Thermotogae and Chloroflexi. *kaiB* genes was found in almost all cyanobacteria and a few proteobacteria while the *kaiA* gene was only found in cyanobacteria. Therefore, it was hypothesized that the *kaiC* gene is comparatively primitive in origin and kaiA is originated in cyanobacteria most recently. kaiB might also have originated in cyanobacteria and then been laterally transferred to other bacterial species (Dvornyk et al., 2003). The presence of *kai* gene homologs in different bacterial species other than cyanobacteria raises the possibility that these bacteria may have some kind of endogenous rhythm that is yet to be detected. Also, the observation of *kaiC* homologs in different species that do not contain homologs for *kaiA* and *kaiB* suggests that *kaiC* gene function might not be solely controlled by those two *kai* genes. In fact, a very recent study (Kawamoto et al., 2020) reported that *kaiBC* promoter activity can

oscillate in a dampened manner with a low-amplitude and a period of about 24-26h, even in the absence of the KaiA protein in *Synechococcus elongatus*. This dampened transcriptional rhythm is regulated by a transcription-translational feedback loop and requires the formation of KaiB-KaiC complex but is not dependent on the phosphorylation cycle of KaiC. This finding provides some insight helping to explain the observation of rhythmic behaviour in some bacterial species (e.g., different purple bacteria) that have *kaiB* and *kaiC* but lack a kaiA homolog. A closer inspection into other species that contain *kaiC* homologs only could shed light onto the regulation of *kaiC* and on other potential rhythmic activities in non-photosynthetic prokaryotes.

5.1.3. The influence of the host circadian clock on the colonizing microbial community

As mentioned earlier in the case of the gut commensal *E. serogenes*, the host derived signal can modify or introduce rhythmic activities in symbiotic prokaryotes (Paulose et al., 2016). There is multiple evidence that alteration in the microbiome population can also change the clock activity of the host (reviewed in Sartor et al. (2019)). Such interactions have been seen in the plant kingdom as well where the plant circadian clock was shown to influence the interacting microbial community. For example, Hubbard et al. (2018) reported that the cultivation of different *A. thaliana* clock mutants can lead to alteration in the rhizosphere bacterial community compared to the rhizosphere microbial composition of wild type plants. Furthermore, when wild type plants were grown in soil that mimic the microbial composition of the soil used to cultivate clock mutants, they show growth related impairments (e.g., smaller leaf rosette, delayed germination) compared to the controls. These observations suggest that the plant circadian clock not only influences the rhizosphere microbial community but also that proper clock functioning is needed for selection of a rhizosphere microbiome that is optimally composed for plant growth and development. These findings are interesting but not unexpected since different rhythmic processes in plants (e.g., photosynthesis, carbohydrate metabolism and partitioning) are likely to provide a rhythmic host microenvironment for the colonizing microorganisms. Therefore, the host together with microorganisms that are better adapted to that rhythmic host microenvironment play important role in shaping plant-microbial communities, bringing mutual benefits.

5.1.4. Aims and objectives

In the case of symbiotic interactions like legume-rhizobia associations, optimizing synchronization between the carbon metabolic cycle of plants and the nitrogen metabolism of the intracellular bacteroids and other processes is crucial. Rhizobia are members of proteobacteria and contain a *kaiC* homolog. However, no notable circadian or rhythmic processes have so far been identified in free-living rhizobia. This is an underexplored area. More studies on rhizobial gene expression and regulation, specifically in bacteroids when they are under host control, might help to determine if there is a free-running rhythm in bacteroids and how this is influenced by host circadian clock.

As described in Chapter 3, to examine the plant host transcriptome, time-course experiment samples were harvested. As well as enabling plant RNA expression (in nodules) to be quantified, the samples also contained rhizobial mRNAs. The library preparation protocol used allowed these rhizobial mRNAs to be sequenced and quantified. Therefore, the data from the time-course RNA-sequencing experiment enabled us to determine the changes in rhizobial gene expression inside nodules over time in free-running environmental conditions. In this chapter the rhizobial gene expression data was analyzed with a view to identifying genes (if any) that are rhythmically expressed in a circadian manner. Regulation or modification at post-transcriptional or protein level (e.g., changes in phosphorylation state of KaiC) could not be assessed, but this data provides a new understanding of bacteroid gene expression at the transcriptional level throughout the day within mature nodules of *M. truncatula.*

5.2. Results

5.2.1. Almost the entire rhizobial genome is expressed in bacteroids

In all our experiments, we used *Sinorhizobium meliloti* SM1022 (also known as *Ensifer meliloti* SM1022) strain to inoculate *M. truncatula* roots to form nodules. The rationale behind using this particular strain was that *S. meliloti* SM1022 is a comparatively highly efficient N-fixing strain and usually results in a high number of nodules (Terpolilli et al. (2013) and data from within the Gifford lab, Lagunas *et al.* In Prep). High nodule number allowed us to grow relatively fewer number of plants per time-points to have enough samples for downstream processing and analysis, enabling an efficient experiment to be designed and carried out. As well as measuring plant gene expression over a time-course (Chapter 3), the expression of bacterial genes was also quantified.

To analyse rhizobial gene expression from the obtained sequence data, the recently sequenced genome and annotations of *S. meliloti* SM1022 were downloaded from the NCBI database as a reference (accession: PRJNA636618). The SM1022 genome consists of a 3.7 million bp chromosome and two plasmids, pA and pB, that are around 1.4 and 1.6 million bp long, respectively. In total, the genome contains 6,328 gene loci and 6,058 protein coding sequences. Using the RNA-seq data analysis pipeline described in Chapter 2, transcriptomic expression data values for 6,328 rhizobial genes was generated. Unlike plant genes, for which around one-third of was not expressed in nodules, almost all of the rhizobial genes show some level of expression in bacteroids (Figure 5.2). Only 0.44% of genes (~28 on average from 3 replicates) had average read counts (over 13 time-points) of less than 5. Over 88% genes had average read counts above 50, representing moderate to high level of expression.

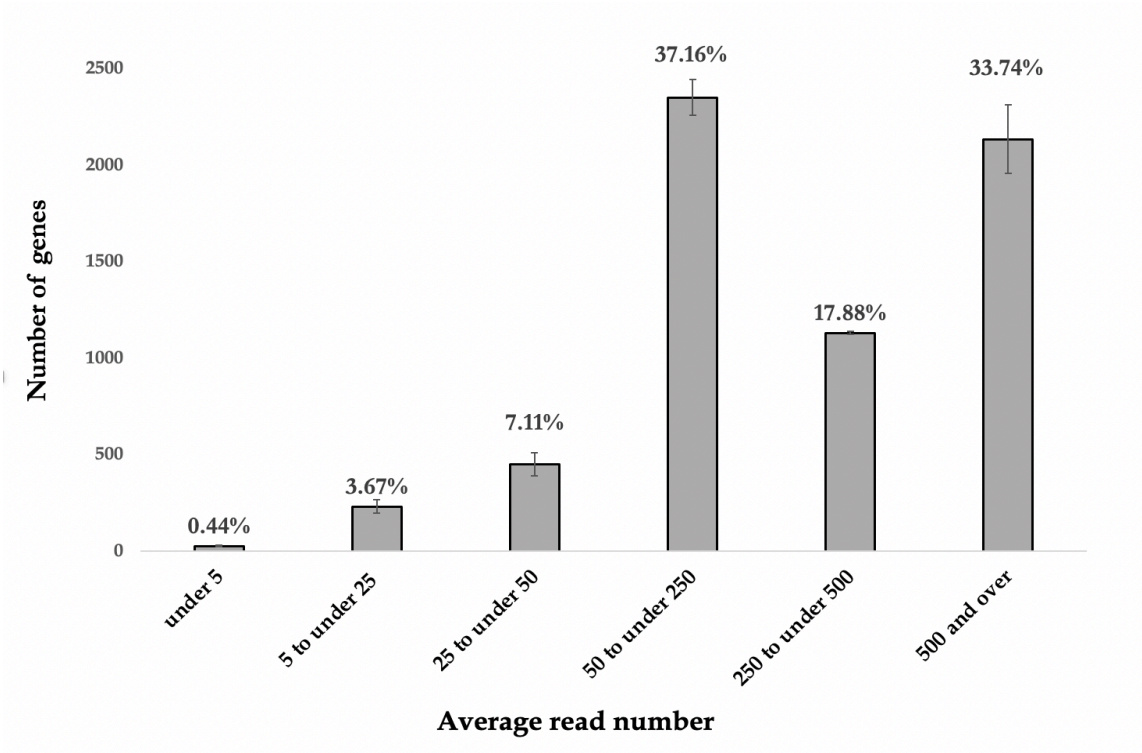


Figure 5.2: Number of SM1022 rhizobial genes with different ranges of read counts per time-point, averaged over 13 time-points. Graph showing that most rhizobial genes were highly expressed. Error bars represent standard deviation from three replicate experiments.

The observations that almost all the bacterial genes were expressed and many of them were highly expressed, were not unexpected for two reasons. Firstly, bacterial genomes are usually very compact and are not subjected to many levels of epigenetic control and silencing (e.g., chromatin remodelling). Secondly, there is a highly concentrated bacterial population inside nodules and they go through several folds of endoreduplication cycle to become mature bacteroids. Therefore, the gene copy number in bacteroids is very high, which can be responsible for comparatively more mRNAs (or read counts) even from a lowly expressed gene. However, since we have not compared gene expression between bacteroids and free-living rhizobia in this study it was not possible to determine whether most genes are generally highly expressed in bacteroids. This could be carried out as part of future work.

5.2.2. Several rhizobial genes show rhythmic expression in bacteroids

In order to identify rhythmic genes, we used metacycle and Lomb-Scargle periodogram, as used to identify rhythmic plant genes. We identified 58 rhythmic genes with a *p*-value <0.05 where 7 of them had *p*-value <0.01 (See supplementary Table S2 & S4). Use of a gene ontology analysis did not identify any particular ontology enrichment, possibly because of a small set of query genes (n=58) and the fact that the *S. meliloti* 1022 genome has been recently sequenced and many ontology terms are yet to be annotated to their respective genes. However, a word frequency analysis using the brief gene definitions of 58 genes uncovered a high frequency of 'transporter' terms, which (despite this being a non-systematic analysis) could be informative (Figure 5.3). Several of the genes annotated with 'transporter' in their term were different ABC (ATP-binding cassette) transporters and permeases. In addition, there were several synthase, dehydrogenase and transferase genes that are likely to be involved in different metabolic processes. This observation suggests that, similar to plants, some bacterial genes might be regulated in a rhythmic manner to facilitate the transport of different metabolites at specific times during the day.

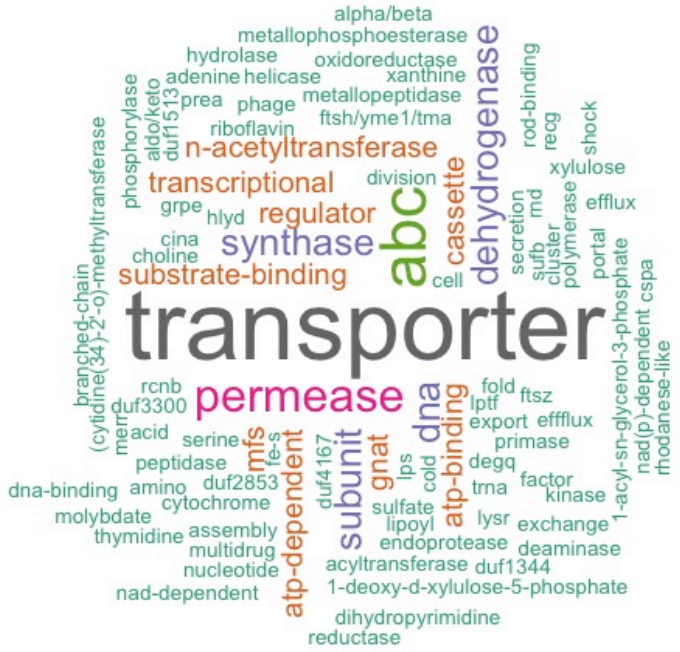


Figure 5.3: Overview of rhythmic rhizobial gene functions. Word-cloud showing the most frequent words in the brief gene definition of the oscillating rhizobial genes.

Next, oscillating genes were tentatively organized into different groups using a hierarchical clustering method, similar to that used in Chapter 3 (Figure 5.4).

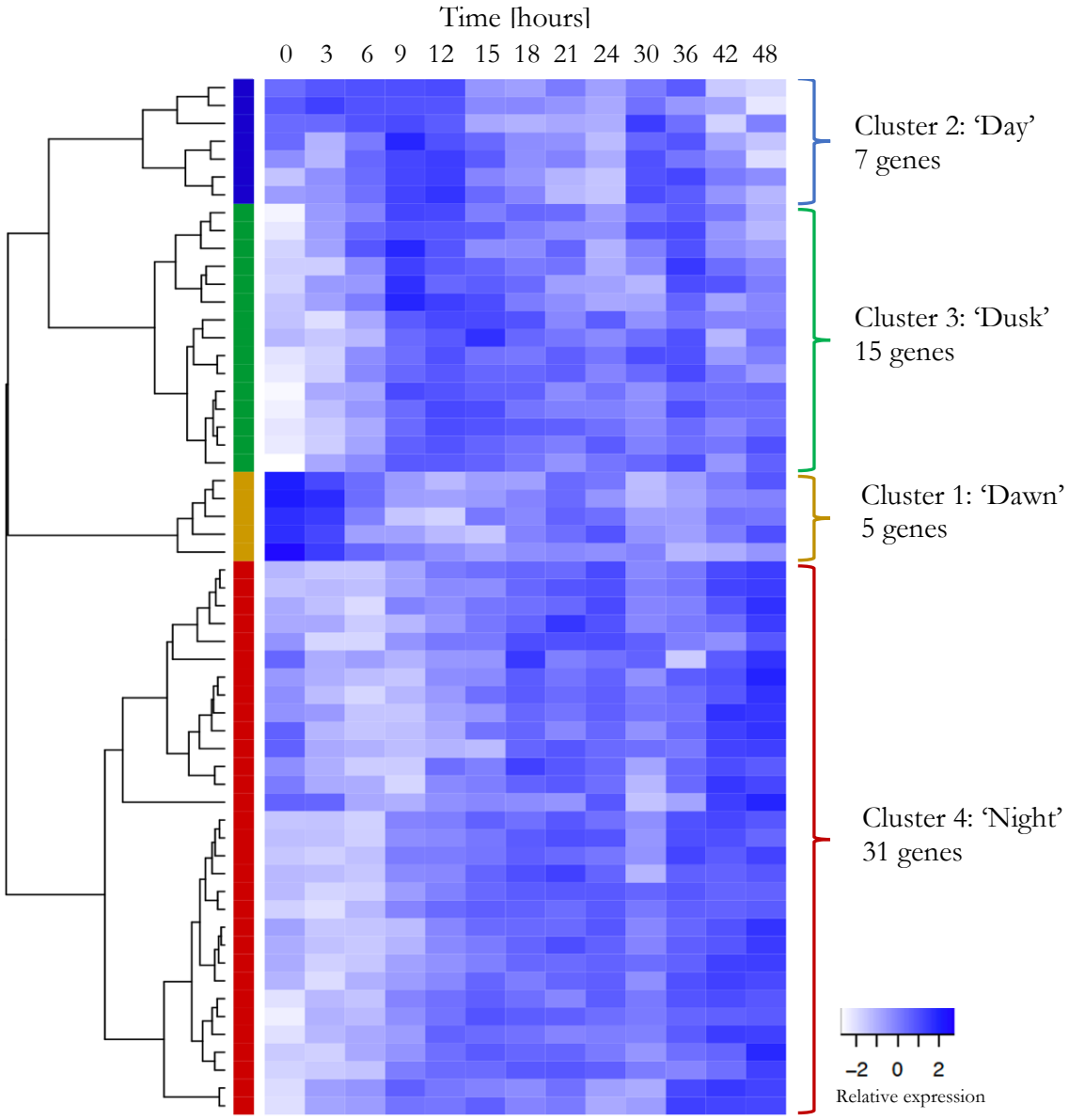
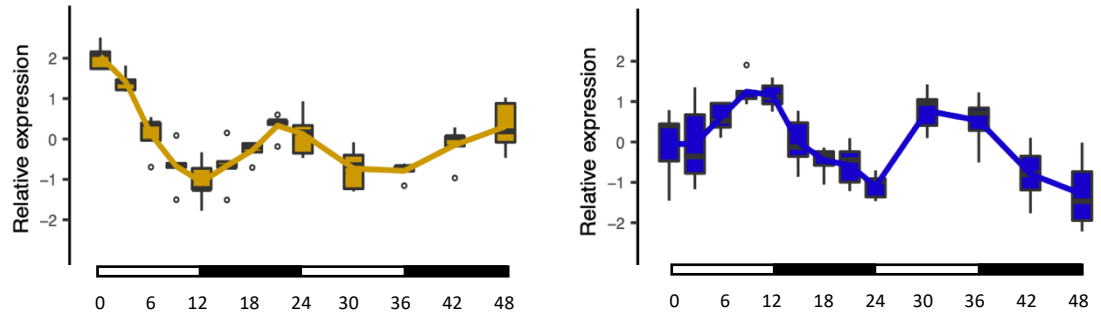
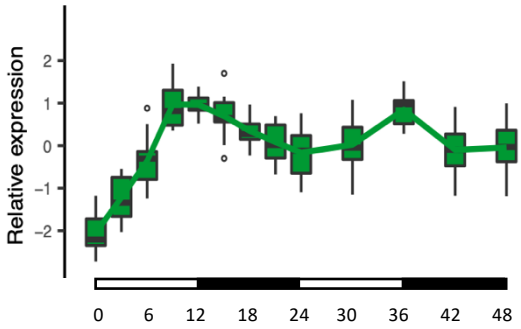


Figure 5.4: A heat map based on hierarchical clustering illustrates the expression of 58 rhythmic rhizobial genes. Four major clusters (1 to 4) are represented respectively by yellow, blue, green and red colours on the left of the heat map. Based on the timing of their expression peak they are also named as 'Dawn', 'Day', 'Dusk' and 'Night' clusters, respectively. In the heatmap, the intensity of blue and white colour represents higher and lower expression, respectively.

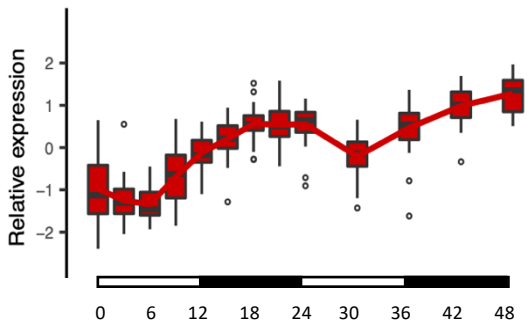


(a) Cluster 1: 'Dawn'

(b) Cluster 2: 'Day'



(c) Cluster 3: 'Dusk'



(d) Cluster 4: 'Night'

Figure 5.5: Expression pattern of genes within each of four timed expression clusters. A centroid plot is shown for each cluster. **(a) Cluster-1 (yellow)** genes peak at 0h and just before 24h, representing the beginning of the day (dawn); **(b) Cluster-2 (blue)** genes peak around 9h and 30h time-points and are comparatively more highly expressed during the subjective day period, hence are categorized as 'day' genes; **(c) Cluster-3 (green)** gene expression is similar to cluster-2 as the genes peak around 9-12h in the first cycle but then they peak distinctively around 36h in the second cycle. In general, the expression of these genes peaks slightly later than cluster-2 (day) genes (as seen in the heatmap, Figure 5.4) and they are therefore termed as 'dusk' genes; **(d) Cluster-4 (red)** genes are highly expressed during the subjective night and expression reaches its minimum at subjective mid-day (6h, 30h), hence these are categorized as 'night' genes. White and black bars underneath each plot indicate subjective day and night, respectively. Expression values for each gene are log2 normalised and relative to its mean expression over time.

Remarkably, amongst this small group of 58 genes, four clear patterns of expression could be differentiated, showing that there were four groups that were expressed and peaked at different times during the 24h cycle (Figure 5.5). More than half of the genes (n=31) were highly expressed during the subjective night (Cluster 4) (Figure 5.5d). This was also the case for rhythmic plant genes, of which around half of were highly expressed at night. Therefore there could be a connection between plant and bacterial rhythmic activity to coordinate related metabolic activities and exchange of different metabolites.

There are 5 genes (Cluster 1) that were highly expressed at the beginning of the day (Figure 5.5a). For the rest of the genes, expression continually increases during the day and decreases during the night. These genes formed two clusters: Cluster 2 and 3. Cluster-2 genes were comparatively more highly expressed for most of the day and peaked slightly before the evening whereas Cluster-3 genes peaked around the evening period; therefore, they were termed 'Day' and 'Dusk' genes, respectively (Figure 5.4 & 5.5b-c).

In Chapter 3, different cluster or time-of-day specific physiological processes (e.g., photosynthetic activity at mid-day period) could be attributed to different rhythmic plant gene clusters. However, this is difficult to do for rhythmic bacterial genes as they are very small in numbers and none of the four clusters was found to be enriched in any particular type of genes. Nonetheless, the rhythmic genes identified in this study could be candidates for some further studies to better understand the prokaryotic clocks outside Cyanobacteria enabling comparison with them.

As mentioned above, many transporter or transporter-associated genes (more than a dozen) were found to be rhythmically expressed. There were six ABC transporters or ABC transporter-associated proteins (one each in midday and dusk clusters, four in night cluster) among those transporter genes. ABC transporters are active transporters that can move molecules against a gradient using ATP energy. Bacterial ABC transporter plays important roles in bacterial survival by actively taking up or exporting different molecules to maintain osmotic pressure, virulence and antagonise toxins or antibiotics, etc (Davidson et al., 2008). As discussed in Chapter 1, some rhizobial species might use proteins similar to ABC transporters (e.g., BacA, BclA) to counteract NCR mediated toxicity on cell surface by importing NCR peptides inside the cell. The ABC transporter related proteins that were highly expressed at dusk and night might have some role against NCR regulation, given that most rhythmic NCRs were also found in dusk and night clusters.

In addition, several metabolic enzymes, for example, three acyltransferases, three dehydrogenases, two reductases, two phosphotransferases etc., were found to be rhythmically regulated. These enzymes are found across the different clusters, with no enrichment for a particular function in any one cluster. It could be assumed that the different metabolic processes within bacteroids, as well as the transport of raw materials

and synthesised products of rhizobial metabolism are controlled in a circadian time-dependent manner. It is also likely that different rhythmic processes in bacteroids are in sync with host rhythmic processes. For example, as discussed in Chapter 3, nitrogen fixation rate in nodules peaks at evening and different nitrogen metabolic and transport related activities are enhanced during night. A higher number of dusk (n=15) and night (n=31) peaking genes among the rhythmic rhizobial genes (n=58) indicates a possible link between these bacterial genes and different nitrogen metabolic processes in nodules.

The cell division gene, *ftsZ* (QKN14772), was found to be rhythmic, peaking at dusk. FtsZ is important in the formation of the Z-ring at the prospective site of cell division (Margolin, 2005). Furthermore, a DNA primase (dusk peaking), a Y-family DNA polymerase (night peaking) and a DNA helicase (night peaking) were found among the diurnally rhythmic genes. These findings raise the interesting prospect that bacterial cell division or endoreduplication related processes could be regulated in a circadian manner and enhanced upon the advent of night period.

Five DUF (Domain of Unknown Function) containing and several 'hypothetical' proteins were also found in different clusters among the rhythmic genes (with so specific cluster enrichment). These proteins might also be involved in important processes that could be investigated in future when more functional prediction data is available.

5.3. Discussion and future directions

Circadian rhythms might not be very common in prokaryotic organisms but different rhythmic activities relating to cell division, growth and external stimuli are abundant in prokaryotes (Sartor et al., 2019). In the rhizosphere and particularly during symbiotic interactions, these rhythmic activities need to be synchronised with host rhythms for optimal benefits to both microbe and plant. For example, in legume-rhizobia symbiosis, rhizobia utilise different photosynthates that are produced and consumed in a time-dependent manner by the plant. Our data in Chapter 3 suggest that N-fixation and N-transport are also enhanced at particular period. Therefore, rhizobia are likely to adapt with/to these rhythmically changing conditions, in particular, when they are inside plant cells. It is also likely that legumes impose tight control on different bacterial metabolic processes by regulating gene expression or by regulating the activity of different enzymes post-translationally as reviewed in Walker et al. (2020). Like the oscillation of KaiC phosphorylation state in cyanobacteria, the active state of other proteins could be rhythmically regulated at posttranslational level in bacteroids. To partially support this, several rhythmic bacterial genes identified in our analysis encode different redox and transferase enzyme that are usually associated with the regulation of other key enzymes/transporters or different metabolites, etc. Even though we could not detect any rhythmic process in bacteria, partly because of small number of rhythmic genes, the presence of these redox enzymes supports the likelihood of a metabolic cycle in bacteroids that coordinate with plant rhythms. As found in Chapter 3, over fifty percent of rhythmic plant genes are highly expressed at night, which could be related to the high metabolic activity and rapid growth of plants during night (Michael et al., 2008a). Around half of all rhythmic rhizobial genes identified here were also highly expressed during night. This might be an indication that different rhythmic processes between host and symbionts are synchronised.

The important nodulation gene in plant, *NIN*, has been found to rhythmically expressed in nodule. However, there was no rhizobial genes that is directly associated with or known to be important for nodulation (e.g., *nod, nif, fix* genes) among the rhythmically expressed genes. These genes were indeed highly expressed (compared to other expressed genes in bacteroids) but not rhythmically regulated at transcriptional level. However, like

KaiC in cyanobacteria, the product of those nodulation related genes can be regulated at post-transcriptional level by different bacterial or plant derived enzymes as mentioned previously. There is no evidence yet that the *kaiC* homolog in *S. meliloti* is rhythmically regulated. In line with this we found that the *S. meliloti kaiC* homolog was also not amongst the rhythmic genes. Therefore, *kaiC* is unlikely to be rhythmically regulated in bacteroids, at least, not at the transcriptional level.

Several DNA replication and cell division associated genes were rhythmically expressed in bacteroids. These genes are likely to be manipulated by certain legume hosts (e.g., *Medicago*) that 'force' rhizobia to undergo several cycle of endoreduplication, a process that keeps the rhizobial population in check while maximising N-fixation (Roy P et al., 2020). Along with FtsZ (as mentioned previously), an ATP-dependent metallopeptidase FtsH family protein (QKN15441.1) was found in the dusk cluster. *FtsH* is also very important for proper bacterial cell division and growth (Jayasekera et al., 2000). The existence of these cell division related genes among the circadially oscillating genes suggest that part of the bacterial reproductive machinery could be under host circadian control. However, it should be considered that the generation time of free-living *S. meliloti* is usually 2h but could be up to 6h depending on different conditions (Dai et al., 2018). In line with an oscillation in different DNA replication and cell division related genes with a period of 2-6 hours, regulated based on actual generation time, would make sense. These oscillations are therefore not likely to be under circadian control, but they were identified as circadian in our analysis. It could be that some reproductive genes are indeed diurnally regulated in mature bacteroid and were appropriately detected in our study. Alternatively, there is a possibility that these genes oscillate with a shorter period but identified with incorrect periodicity because of the comparatively longer sampling intervals of 3h (in the first 24h) or 6h (in the last 24h) in our experiment. To clarify this further, a sample interval of 3-6h is sufficient to detect circadian rhythms but to appropriately identify a rhythm with shorter period, sampling intervals need to shorten as well. Any rhythm with a small period is likely to be undetected or detected with incorrect periodicity due to the lack of data in between sampling intervals. As mentioned, the data presented in this chapter was generated from the time-course experiment broadly described in Chapter 3. The time-course experiment was primarily designed to identify rhythmic plant genes that are under endogenous circadian clock regulation in mature nodules. The metagenomic RNA sequencing protocol allowed us to gain some valuable insights about the rhizobial gene

expression too. However, while analyzing this data it should be kept in mind that for better detection of bacterial rhythmic activity, a smaller sampling interval would be even more valuable.

The gene expression in a small proportion of rhizobia growing in the infection zone and interzone of nodules that are not fully differentiated would be expected to be different from that of mature bacteroids. However, in our study the whole nodules were collected and ground to extract RNA, therefore, gene expression difference in different stages of bacteroids could not be determined. In general, our data shows bacterial gene expression in mature nodules as a whole but is mostly representative of fully differentiated bacteroids because of their high number and greater gene copies. Future studies involving microdissection of different nodule zones to extract bacteroid RNAs from different stages of nodule development would likely to provide even more insights into rhizobial gene regulation within nodules. It would be interesting to determine the expression of rhythmic genes in free-living rhizobia. In addition, some of the rhythmic genes could be knocked-out to determine how that impacts rhizobial colonization, differentiation and N-fixation etc. In conclusion, the data generated in our study is very preliminary, but it provides exciting grounds for some future studies.

6. Utilization of cell-type specific promoters to study cell type specific gene expression in *Medicago truncatula*.

6.1. Introduction

In a similar way to the development of lateral root formation, the onset of nodule organogenesis and further development are coordinated in a highly cell-type specific manner. As discussed in Chapter 1, nodulation can be considered to begin with the cross-talk between flavonoid-secreting plant root hair cells and soil rhizobia secreting Nod factors. The initial perception of Nod factors leads to a calcium ion spike in epidermal cells that results in root hair curling and *de novo* cell divisions in the underlying inner-cortical cells, giving rise to a nodule-primordium (Oldroyd, 2013). When the inner cortical cells become mitotically activated after Nod factor recognition, rhizobia are still at the epidermis. Interestingly, Nod factors are immobile signal molecules (Goedhart et al., 2000) but their perception triggers cell divisions in the inner-most cortical cell layer, however the outer cortical cells which are closest to the perceived signal respond relatively later, showing complexity even amongst cells of the same type (cortical cells) (Xiao et al., 2014). However, there must be coordination of cell division in underlying cell layers, combined with the development of an infection thread to enable bacterial entry. Precisely how initial perception of rhizobia in the epidermis and related early signaling results in coordinated cell divisions and nodule organogenesis in a temporally and spatially restricted manner is yet to be fully understood.

Root nodules are often considered to be modified lateral roots as the development of both organs share overlapping developmental processes (Schiessl et al., 2019). However, there are key differences between these two processes. Unlike nodules, lateral roots are initiated from the cell divisions of predefined pericycle cells termed as founder cells. These founder cells are assumed to be primed by periodically oscillating auxin levels. However, no evidence has been found so far that nodules are originated from predefined cortical founder cells. Nodule organogenesis is initiated by CRE-1 (CYTOKININ RESPONSE-1) mediated cytokinin signaling that induces the expression of the key transcription factor NIN (NODULE INCEPTION). Both CRE-1 and NIN are indispensable for the induction of nodulation (Vernié et al., 2015). While cytokinin signaling is important for nodulation, it suppresses lateral root formation. This is also

supported by the observation that *cre1* mutants produce fewer nodules with an increased number of lateral roots (Gonzalez-Rizzo et al., 2006). Auxin accumulation has been observed at the sites of both nodule and lateral root primordia, therefore it is an important regulator of both the processes, but initial cytokinin signaling is essential specifically for nodule initiation. The initial cytokinin signaling possibly blocks the polar transport of auxin underneath the site of rhizobial perception causing a localized auxin accumulation that directs nodule development (Kohlen et al., 2017).

Despite their different modes of initiation, both lateral root development and nodulation have some commonalities and shared regulatory components (Schiessl et al., 2019) (Figure 6.1). In fact, observations from two recent studies supports the idea that the nodule organogenesis in legumes might have evolved by hijacking some key regulatory components from lateral root development (Soyano et al., 2019, Schiessl et al., 2019).

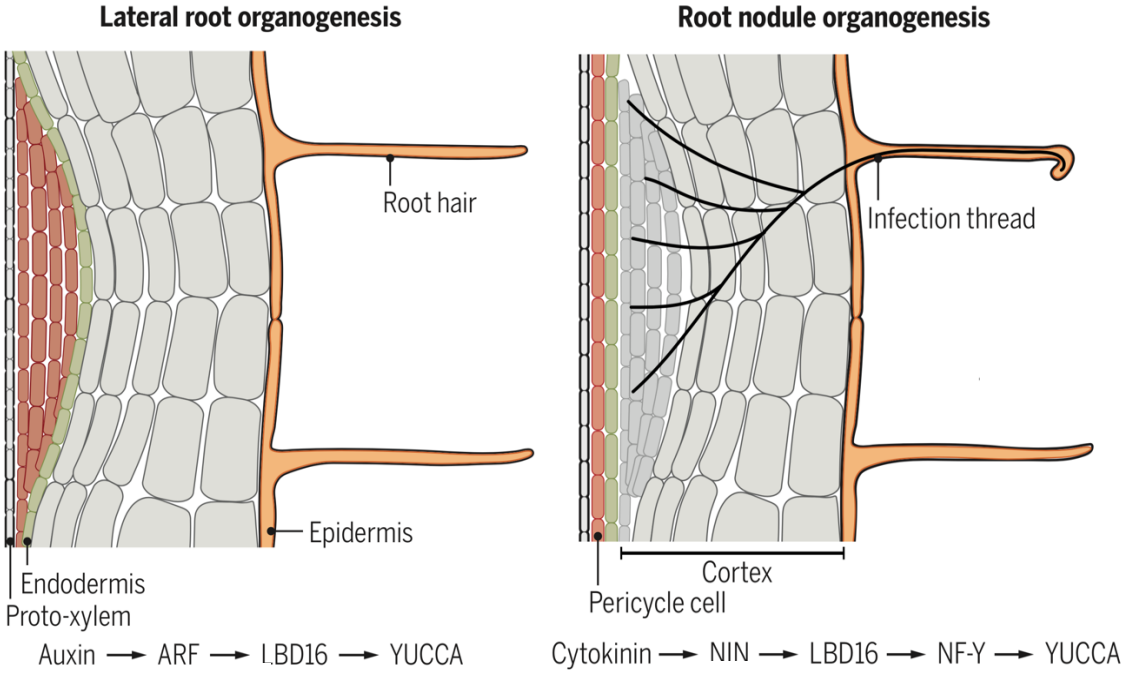


Figure 6.1: Lateral root and nodule organogenesis are initiated differently but share overlapping pathways. (a) Lateral root development begins with pericycle cell division as a result of signaling that varies due to periodic auxin maxima. (b) In contrast, a nodule primordia is formed via cortical cell divisions resulting from the perception of altered cytokinin signaling. For both (a) and (b), initial signaling converges on the expression of LBD16 that induces different auxin biosynthesis genes (e.g., YUCCAs) and further cell divisions enabling organ development. ARF, Auxin Response Factor; NF-Y, Nuclear Factor-Y (although they also have many distinct components of regulation); adapted from Bishopp and Bennett (2019).

A transcription factor known as LBD16 (Lateral organ Boundaries Domain 16) has been reported to be required for both nodule and lateral root formation (Figure 6.1). In *Arabidopsis*, *LBD16* is expressed in pericycle cells in response to auxin and leads to lateral root initiation. Soyano et al. (2019) reported the expression of *LBD16* in dividing cortical cells of nodule primordia in *Lotus japonicus*. Interestingly, they identified a NIN-binding site (NBS) in the intron of *LBD16* which is conserved along with its flanking region in many legume species; but not observed in non-legume *LBD16* orthologs. Furthermore, this NBS containing *LBD16* intron is sufficient for its cortex-specific expression when minimal promoter elements are available (Soyano et al., 2019). However, an intact, full-length *LBD16* promoter is required for expression in pericycle. It is not yet clear at this point how the binding of NIN at an intronic region alter the expression specificity of associated gene. Nonetheless, these observations suggest the possibility that an evolutionary origin of NBS in the intronic region of the *LBD16* gene of an ancestral legume somehow enabled it to ectopically express LBD16 in cortical cells. Such an event would have allowed a pericycle specific pathway to be introduced in the cortex that could be initiated by cytokinin/NIN instead of auxin and lead to a new organ development.

These observations and findings inspire further studies to understand how different regulatory components are coordinated in different types or layers of cells at different phases of nodule/lateral root organogenesis. This would be useful so we can find out which set of genes are differentially expressed in particular cells or tissues (e.g., epidermis, cortex, pericycle) in response to the perception of rhizobial inoculation. Most transcriptomic profiling studies so far have been based on whole organ study (e.g., entire root or root sections) where signals from particular cells or tissue types are usually abrogated by signals from different part of the organ. However, modern technologies enable us to generate cell-type specific data. One of the well-known ways of isolating subpopulations of particular tissue cells is laser-capture microdissection (LCM) (Casson et al., 2005, Portillo et al., 2009, Casson, 2008). Another alternative is the use of Fluorescence Activated Cell Sorting (FACS) where fluorescently labelled target cells can be isolated in a high-throughput manner (Birnbaum et al., 2005, Carter et al., 2013, Gifford et al., 2008).

For FACS there is a requirement to label cell types for sorting and it is possible to express different fluorescent proteins such as green fluorescent protein (GFP) or red fluorescent protein (RFP/dsRed) in target cells by means of cell-type specific promoters. A number of cell and tissue specific transgenic stable lines are already established for the model plant *A. thaliana*. However, such lines are not available or rare in the case of model legumes. Generation of cell-type specific fluorescent protein expression lines in any model legume plant would therefore enable the study of cortical, epidermal and/or other localized responses in legumes. Furthermore, this data could enable comparison with whole root expression data. Such transgenic lines would also be very useful for microscopic analysis, specially, in confocal microscopy to study developmental changes in different cell types over time (Genre, 2008).

Several root cell-type specific promoters have been identified in the model plant *A. thaliana* (Lee et al., 2006) and some other species (e.g., tomato (Bucher et al., 2002)). Some of these promoters that are specific for different root cell layers (e.g., epidermis, cortex, endodermis, pericycle, vascular tissues) have been ectopically tested for their efficiency in model legumes, *Lotus japonicus* (Gavrilovic et al., 2016) and *Medicago truncatula* (Sevin-Pujol et al., 2017) (Figure 6.2). In both those studies, β-*GLUCURONIDASE* (GUS) gene coding sequence was cloned under a promoter of interest and then plants transformed

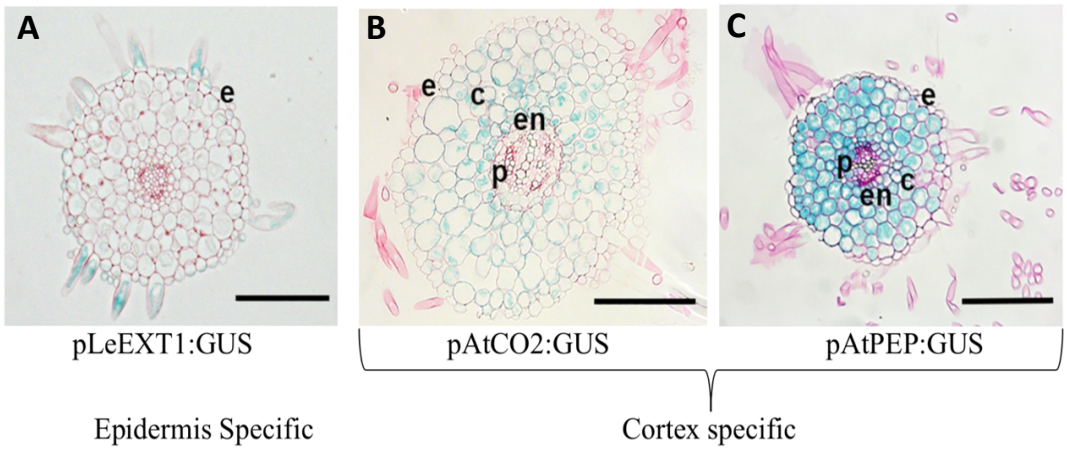


Figure 6.2: Expression of GUS under cell-type specific promoters. (A) Epidermis-specific expression directed by the pLeEXT1 promoter derived from an extensin-like gene in tomato; **(B) & (C)** Cortex-specific expression under two promoters, pAtCO2 and pAtPEP respectively, derived from *A. thaliana;* e: epidermis, c: cortex, en: endodermis, p: pericycle; Scale bar: 100 µm; figure adapted from Sevin-Pujol et al. (2017).

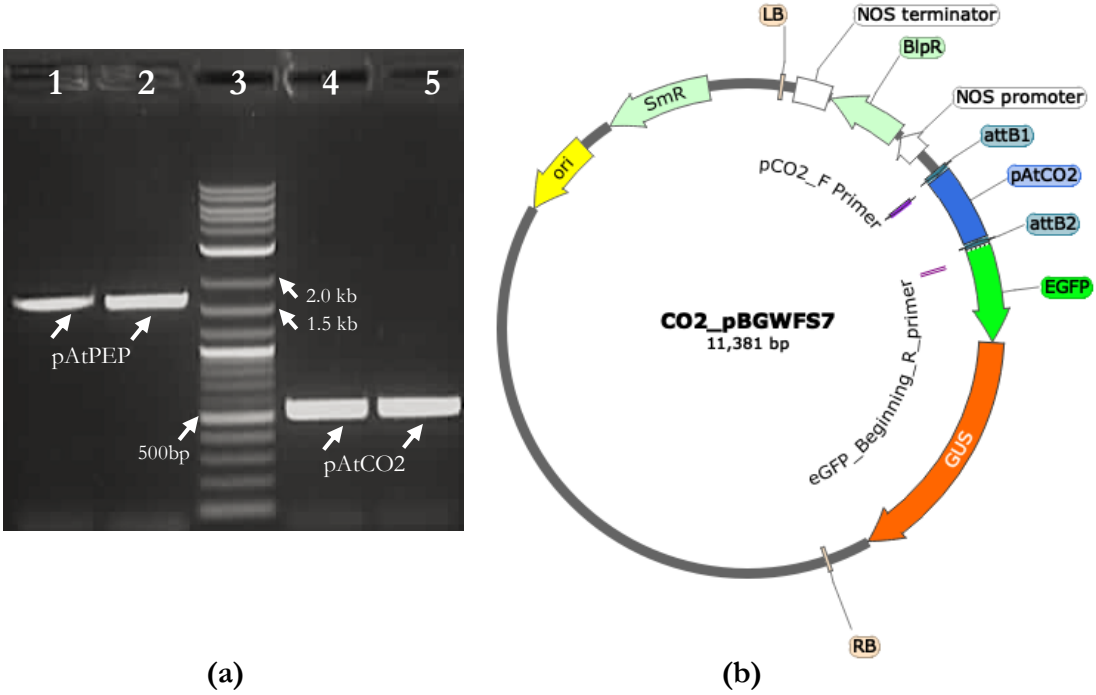
transiently with the hairy root transformation protocol (Ron et al., 2014, Boisson-Dernier et al., 2001, Stougaard et al., 1987). Findings from these two particular studies (Sevin-Pujol et al., 2017, Gavrilovic et al., 2016), partly presented in Figure 6.2, provide valuable insights about some prospective promoters that could be considered for generating cell specific stable fluorescent lines.

In the work presented here, the specificity of two cortex associated promoters from *A. thaliana*, pAtCO2 and pAtPEP, was evaluated for use in *M. truncatula* using GFP as reporter gene instead of GUS. This is because GFP is relatively convenient to detect under confocal or fluorescent microscope as no further pre-processing such as staining is usually required. In addition, cells expressing GFP are easier to sort and isolate through FACS. If the specificity of these promoters is successfully reproduced in transient transformation, they could be used to generate cortex specific GFP-expressing stable transgenic lines. We were more interested to generate cortex specific lines since nodule primordia begin from this cell-type and therefore cortex specific gene expression analysis might shed some light on the initiation of the nodulation process. Nonetheless, if we succeed that will also inspire generating other cell-type specific lines and comparative gene expression analysis.

6.2. Results

6.2.1. Generation of a cortex specific expression vector for use in *M. truncatula*

In order to generate cortex specific GFP expression lines, two promoters originating from Arabidopsis, known as AtCO2 and AtPEP, were chosen as their cortex specificity were ectopically checked in other species (Ron et al., 2014, Gavrilovic et al., 2016, Sevin-Pujol et al., 2017). AtCO2 and AtPEP corresponds to two Arabidopsis genes, At1g62500 and At1g09750, respectively. The entire upstream intergenic regions of these genes were amplified through PCR from *A. thaliana* (ecotype: Columbia-0) genomic DNA. The amplified products (Figure 6.3a) were then cloned into the pENTR™ entry vector primarily and then finally to pBGWFS7.0 expression plasmid (Figure 6.3b) using



(a) **(b)**

Figure 6.3: Preparation of promoter:reporter construct. (a) Gel electrophoresis of PCR-amplified promoter regions of AtPEP and AtCO2 genes. pAtPEP and pAtCO2 are 1,679bp and 552bp long, respectively, as confirmed by their respective positions in the gel. 100bp plus DNA ladder was used in lane-3. **(b)** Schematic map (generated with SnapGene v4.3) of expression vector pAtCO2:GFP-GUS, showing the respective position of the pAtCO2 promotor (in blue), EGFP (in bright green) and GUS (in orange) coding sequences. The streptomycin (SmR) and basta resistance (BlpR) genes are shown in light green. BlpR (under nopaline synthase promoter (NOS) and within T-DNA borders) enables the selection of transformed plant material, and SmR enables bacterial selection in selective medium; ori: origin of replication, LB/RB: left/right border of T-DNA repeat, attB1/2: recombination sites 1 and 2 for Gateway® BP reaction.

Gateway™ cloning method (see Chapter 2). The pBGWFS7.0 plasmid vector is specially designed for promoter analysis, where a GFP-GUS fusion protein could be expressed under promoter of interest (Karimi et al., 2002). The fusion protein could be detected by either GFP fluorescence or β-glucuronidase enzyme activity. The vector confers both streptomycin and spectinomycin resistance is bacteria and basta (glufosinate) resistance in host plants. Once the vector constructs were successfully produced, they were incorporated into *Agrobacterium rhizogenes* ARqua1 strain through electroporation and later used for hairy root transformation.

6.2.2. pAtCO2 and pAtPEP exhibit cortex specific activity in *M. truncatula*

To examine promoter activity in a time-efficient manner, young *M. truncatula* seedlings were subjected to hairy root transformation which is an easy and rapid way to evaluate the efficiency of vector constructs (Boisson-Dernier et al., 2001, Ron et al., 2014). In this procedure, a high percentage of newly formed, highly branched roots (hence, named as 'hairy'; Figure 6.4a) become transformed that can be analyzed for reporter gene (or other transformant) activity. Since the entire transformation procedure does not take much time (~2 weeks after inoculation), it is reasonable to use it before proceeding with stable transformation which could take up to a year for *Medicago spp.*

To carry out hairy root transformation, the root-tip of 3 days old seedlings were cut and inoculated with *A. rhizogenes* harbouring the vector containing the promoter:reporter of interest, then grown for 7 days. To facilitate root growth and branching and to improve the probability of transformation, no selective reagent (e.g., basta) was applied. After a week, seedlings were examined under a stereomicroscope fitted with a GFP filter. Transformed roots were identified based on the fluorescence conferred by the EGFP reporter gene (Figure 6.4).

From a total of 43 plants that were tested for pAtCO2 harbouring constructs, 35 (81%) plants had, at least, one branch/section of root showing GFP fluorescence. For pAtPEP, 34 out of 40 (85%) plants were fluorescent. Although the percentage of successful transformations were high, it could have been even higher since the AtCO2/AtPEP promoters might not have all be active in all transformed roots/root parts. Fluorescence was observed in both the meristematic and the older part of the

transformed roots for both the promoter transformations. Therefore, the promoters might not be specifically expressed in any one part of the root, e.g., growing or mature, as they were seen to direct the expression of GFP all over the roots in an inconsistent and sparse manner (Figure 6.4b).

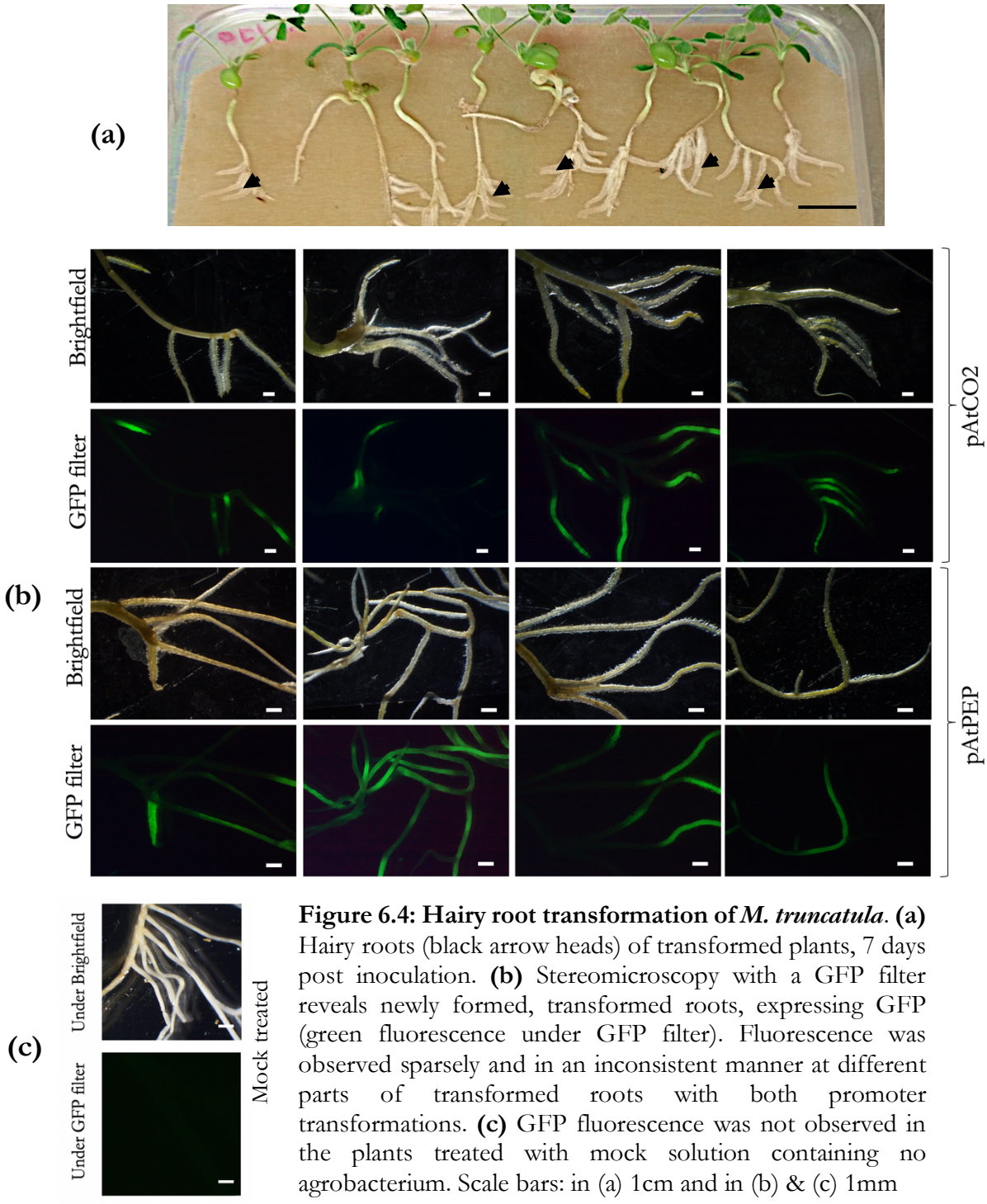


Figure 6.4: Hairy root transformation of *M. truncatula*. (a) Hairy roots (black arrow heads) of transformed plants, 7 days post inoculation. **(b)** Stereomicroscopy with a GFP filter reveals newly formed, transformed roots, expressing GFP (green fluorescence under GFP filter). Fluorescence was observed sparsely and in an inconsistent manner at different parts of transformed roots with both promoter transformations. **(c)** GFP fluorescence was not observed in the plants treated with mock solution containing no agrobacterium. Scale bars: in (a) 1cm and in (b) & (c) 1mm

In order to evaluate the cell-type specificity of the promoters, GFP-expressing root branches were identified and cut under a stereomicroscope with GFP filter on. To be specific, five pieces (~0.5cm long) of transformed branches from five different plants for each promoter were taken and cross-sectioned using a vibratome. Finely and evenly cut cross-sections (thickness of ~200 microns) were stained with propidium iodide (PI) and then examined through confocal microscopy. The confocal microscopy images showed high levels of cortex-specific expression of the EGFP reporter gene for both the promoters (Figure 6.5). No significant visual difference was observed between the activity of two promoters in terms of the number of cortical cells expressing GFP or their fluorescence intensity (Figure 6.5).

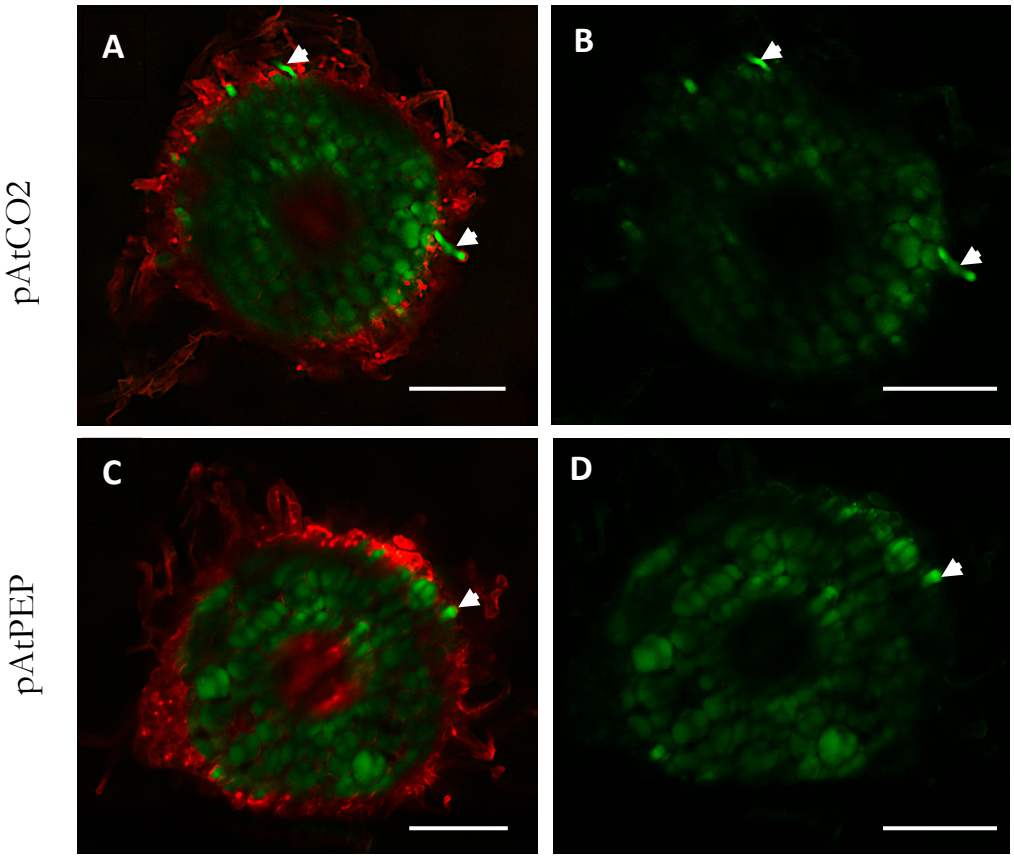


Figure 6.5: Confocal microscopy of cross-sections of transformed root shows cortex-specific expression for AtCO2 and AtPEP promoters. (A) & **(C)** Overlay of propidium iodide fluorescence (red) and GFP emission (green). **(B)** & **(D)** GFP fluorescence image only. At 7th day post inoculation (10 days old seedling), GFP fluorescence was observed almost exclusively in the cortical cells and in a very few epidermal cells, marked with white arrowheads. Scale bar: 100µm.

6.3. Discussion and future directions

While this study and published literature support the cortex specificity of the AtCO2 and AtPEP promoters, several questions are yet to be properly addressed. First, at which developmental stages are these promoters more active? Does their expression differ between growing and mature parts of root? Sevin-Pujol et al. (2017) reported that pAtCO2 activity was stronger in meristem or diving cortical cells than in any other parts of the root, whereas the pAtPEP activity was stronger in older parts of roots, but absent in meristem. Such conclusion could not be made in our study since the GFP expressions were observed all over the roots for both promoters. Second, whether the cell type specificity changes over time as the plant grows? Lagunas B. (2018) had reported that the *Arabidopsis* expansin gene promoter, pAtExp7, showed high epidermal activity at 4 days of root development but by the 7th day the cell-type specificity was lost. This could be the case for these cortex specific promoters as well which needs to be checked in future works. Finally, fluorescent proteins like GFP can diffuse from cell to cell to some extent by different means (e.g., through plasmodesmata). How would that affect downstream applications like FACS? This problem could be resolved by designing vector constructs that confine reporter gene expression to particular cellular location or organelle (e.g., Golgi body, endoplasmic reticulum). This could be achieved by adding a short N-terminal signal peptide sequence to the reporter gene (Katja Kapp, 2009). The vector constructs could be further improved by incorporating another fluorescent reporter (e.g., dsRED) under a constitutive promoter (e.g., pAtUbi) within the vector construct (e.g., pCambia vector used by Sevin-Pujol et al.) That would be useful to detect root parts that were transformed but showing no activity in regard to the promoters of interest, e.g., AtCO2 and AtPEP in our case. Although several questions remain unanswered, the works presented here, in line with existing literature, reconfirm the potential of these promoters for preparing cell-type specific GFP expressing stable transgenic lines as a tool for generating cell-type specific omics data and subsequent analysis.

7. General Discussion

7.1. Background to the research questions and experimental design

The NCR family has received huge attention in the legume research community because of their unique presence in IRLC legumes, exclusive expression in nodules and their critical roles in enforcing terminal bacteroid differentiation (as reviewed by Mergaert (2018), Roy P et al. (2020), Pan and Wang (2017)). The presence of NCR like peptides in non-indeterminate, nodule-forming, non-IRLC legumes (e.g., *Aeschynomene spp.*) was also correlated with relatively higher bacteroid differentiation and elongation (Czernic et al., 2015). Comparative studies and examples (Montiel et al., 2016, Montiel et al., 2017) from different IRLC legumes suggest that legume hosts with a greater number of NCRs apply tighter control over bacterial metabolism, becoming more dominant over them. The employment of NCRs could be an early evolutionary step towards a future, independent N-fixing plant organ (Coba de la Peña et al., 2017, Roy P et al., 2020). Some support in favour of this hypothesis comes from studying an amoeba, *Paulinella chromatophora*, which harbours an early-stage photosynthetic organelle called a chromatophore. The chromatophore is thought to be derived from cyanobacterium about one hundred million years ago. It has been found that *P. chromatophora* guides several small peptides that are less than 90 amino acid long to its chromatophores. Most of these peptides came from the ancestral amoebal host, not from cyanobacteria; therefore, are likely to be involved in the early stages of the integration of chromatophore (Singer et al., 2017). This highlights that the small symbiotic peptides like NCRs that are expressed by the host can play important functions in developing a permanent endosymbiosis. Not to mention that such incorporation of endosymbionts into a permanent organelle would be highly complicated in multicellular organisms compared to single cell organisms like amoeba. Interestingly, symbiotic peptides like NCRs have also been found outside plants, for instance, in the interaction between aphids and their bacterial endosymbiont *Buchnera aphidicola*. The aphid-*Buchnera* endosymbiosis was originated at least 200 MYA, and has now arrived at a stage of obligate mutualism where *Buchnera* no longer lives freely outside the host. These γ-proteobacteria are maternally transferred through host generations (Shigenobu and Stern, 2013). The aphids restrict endosymbionts to specialized symbiotic cells called bacteriocytes (comparable to symbiosomes) where at least seven cysteine-rich peptides

have been found to be expressed (Shigenobu and Stern, 2013, Uchi et al., 2019). Though, the precise functions of these bacteriocyte specific cysteine-rich (BCR) peptides are yet to be determined, their bacteriocyte-specific expression suggests a role in regulating the symbiosis. These insights and related evidence suggest a potential developing role for the NCR peptides in governing rhizobial differentiation, dominating over the bacterial partners and maximising their benefits. Therefore, understanding the regulation of these NCR genes has become very interesting from a legume-rhizobia research perspective.

Preliminary data to support the research questions and the experimental plans in this thesis came from a previous PhD work in our laboratory (Achom, 2019). In that study, the promoter regions of 166 NCR genes that were differentially expressed at 2hr or 6hr after nitrogen treatment in nodulating roots vs non-nodulating roots were analyzed to understand the basis of their regulation. Their regulation was of interesting since differential expression for NCRs in temporally spaced samples was not typically expected. As briefly described in Chapter 1, six conserved motifs were found in NCR promoters, and one of them contains the EE or EE-like motif (AGA[T/C]ATTT). The EE motif can be bound by several circadian regulators, including CCA1, LHY, RVE7, RVE8 in *A. thaliana*, and their homologs in other species (Chow et al., 2019).

The finding of the EE element in the promoters of a subset of NCR genes whose expression varied over time-points established a hypothetical link between the regulation of NCRs and the plant circadian clock. As discussed in Chapter 3, the circadian clock is also known to modulate the response of the plant to pathogens (Lu et al., 2017) and itself can be modified by the invaders (Wang et al., 2011). Therefore, it seemed reasonable that the circadian clock might also be involved in plant-rhizobial interactions and nodulation; hence, this thesis set out to study the circadian gene regulation specifically in nodules, through the use of and analysis of a time-series RNAseq experiment. We used a metagenomic approach for RNAseq library preparation that allowed us to not only quantify the expression of plant genes (Chapter 3) but also the bacteroid gene expression (Chapter 5) within nodules.

The robustness of our analysis and the authenticity of our data are supported by, firstly, the identification of all major clock genes (listed in Table 3.1) within a $p<0.01$ threshold; secondly, the appropriate presence of the major clock genes in specific time-

of-day clusters where they were expected to be found; finally, genes that are associated with common rhythmic processes (e.g., photosynthesis) are also found within the clusters that peaked as expected. Metagenomic RNA sequencing is comparatively expensive and considering our budget constraint we decided to sample and sequence fewer time-points for the second subjective day (24h-28h) (see Figure 3.1 for experimental overview). Therefore, we had 13 unevenly spaced time-points (9 time-points with 3h interval and 4 time-points with 6h interval for first and second day, respectively). A better experimental design would employ 17 evenly spaced time-points with 3h interval over 48 hours. However, there are algorithms that can detect rhythms in unevenly sampled data (LS method; discussed in Section 3.3), making this experimental design feasible for both budget and analysis. More time-points would likely to have enabled us to detect more rhythmic genes (from both plant and bacteroid) that potentially remained undetected due to their relatively smaller amplitudes. However, overall our new data enabled us to gain significant new insights, ranging from specifically detecting a set of oscillating NCR genes to generating a general understanding of the rhythmic behaviour of key clock genes and other rhythmic processes in nodules.

In order to evaluate whether clock disruption has any effect on legume-rhizobia interaction or nodulation, we conducted a comparative phenotyping study with two independent mutants with transposon insertions in *MtLHY* (Chapter 4). *CCA1* and *LHY* are the most well-studied clock genes in *A. thaliana* to date and considered as the core component of central oscillator (Adams et al., 2018, Park et al., 2016). However they had not been studied in legumes when this work started. In *M. truncatula*, a single homolog of both *CCA1* and *LHY* was found which was annotated as *MtLHY*. In theory, *mtlhy* mutation was likely to affect the expression of other clock genes, which later became evident after analysis of transcriptome data. We obtained the two *mtlhy* mutant lines, NF17115 and NF16461, from the Noble Research Institute harbouring a Tnt1 retrotransposon insertion in the promoter or 5th exon of *LHY* gene, respectively.

In addition to harbouring a Tnt1 insertion at *LHY* locus, these mutants also harbour Tnt1 insertion at different positions across the genome. We selected and propagated the plants that contain homozygous Tnt1 insertions at *LHY* locus, and used the two independent lines to confirm the *LHY* characterization. However, other insertions at different positions randomly segregate amongst the plants. In future work,

the *lhy* mutant lines will be backcrossed to the parental R108 line and re-selected to reduce the presence of other Tnt1 insertions. In general, Tnt1 is considered as stable and inheritable and has been used for developing insertional mutagenesis libraries in different plant species, e.g., soybean, potato (Cui et al., 2013, Tadege et al., 2008a). Recently, some evidence has been found that the retrotransposon can be activated and create new insertions during tissue culture procedure (Zhang et al., 2018). Although we have not done any tissue culture with the mutant lines (thus there is a low chance of new insertions), the pre-existing insertions at random locations (originated during the mutant line development) are undesirable. However, we had to accept these limitations since there is currently no other *lhy* mutant lines available for *M. truncatula*. The effect of these other insertions was not characterized; they might have no or negligible effect unless they disrupt some other vital genes. These insertions in the mutant genome might account for some variability seen in the phenotyping experiment (Chapter 4, Figure 4.3). However, we tried to address the issues with biological variability by setting up three replicate experiments using both independent *lhy* lines and using more than ten plants per mutant line in each replicate experiment. This strengthened our data and enabled us to make conclusions based on robust evidence. Since we now know that *LHY* plays a role in nodulation, implicating a role for the clock in controlling symbiosis, it would be interesting to examine the roles of other clock genes in nodules, such as *TOC1* and *ELF4*.

In parallel to the main research project, the development of two cortex specific reporter gene expression vectors was carried out (chapter 6), to develop tools for conducting future cell-type specific gene expression study. Such vectors were already available driving expression of the GUS reporter gene (Sevin-Pujol et al., 2017, Gavrilovic et al., 2016). Based on the available knowledge, we aimed to prepare cortex specific GFP reporter line as GFP-expressing cells are usually easier to isolate through FACS and it is from the cortex that most nodule tissue forms. This technical work could be very useful in future to study cortex-specific gene expression at different experimental conditions in *M. truncatula.*

7.2. Circadian regulation of NCR genes in nodules via the Evening Element

Probably the most remarkable finding presented in this thesis was the identification of diurnally regulated NCR genes, most of which were highly expressed during the

evening period. The EE motif was enriched in the rhythmic NCRs compared to non-rhythmic NCRs, offering a mechanism to explain this regulation. Furthermore, among the 13 rhythmic NCRs that contain EE in their promoter, 10 of them belong to the dusk-peaking cluster (Section 3.2.8). Put together, this supports our hypothesis that some NCRs are regulated by the plant clock, specifically, by LHY (as discussed in Chapter 3). However, this must be followed up with functional characterization of these rhythmic NCRs in the future in order to elaborate on the importance of their rhythmic expression.

We have found 45 rhythmic NCR genes out of 701 putative NCR genes studied. Therefore, about 6.5% of all NCR were rhythmic whereas about 5% of all genes were found to be rhythmic in nodules. In this regard, it cannot be said that NCRs are, in general, rhythmic in nature. However, in *M. truncatula* NCRs represent a moderately large family of peptides that are known or hypothesized to play diverse roles in the nodulation process (Roy P et al., 2020). It could be that those rhythmic NCRs are involved in a particular function. One hypothesis could be that they play roles in defense response. It had been demonstrated that the plant's general resistance to pathogenic invaders varies in a circadian manner and usually is low around midnight (Bhardwaj et al., 2011). As discussed in Chapter 1, NCR peptides are well-known for their antimicrobial actions. The set of NCRs that are highly expressed at the evening period could be associated with providing an extra layer of protection to the nodules, when plant's general immunity is in decline. Further analysis of these rhythmic NCRs, taking account of their peptide and promoter sequences might shed some light on their roles and regulations.

As reviewed in Chapter 1, Roux et al. (2014) determined the spatial expression of NCR genes in different nodule zones. Among the 521 NCR genes they analyzed (out of 701 NCR gene in our list), 53% were highly expressed in the nodule interzone (IZ). Interestingly, the interzone was also increasingly enriched for rhythmic NCRs (24 out of 45, ~69%) and for rhythmic NCRs with an EE (10 out of 13; 77%). Therefore, the interzone area of the nodule, where bacteroids start to express N-fixation genes and go through an extensive differentiation process, seemed to be particularly important for the rhythmic NCRs along with the expression of several important late nodulins by host cells (Mergaert et al., 2020). It is important to remember that our data is representative of the whole nodule, but the NCR dynamics and expression pattern can vary in different sections of the nodule. If any NCRs are particularly rhythmic in a specific nodule zone but not in

other zones, the overall expression of that NCR could be diluted down and remain undetected when studied at the whole nodule level, which is the rationale for carrying out cell type analysis to understand plant responses (Walker et al., 2017). In addition, our data was from mature nodules, but NCR functioning and expression dynamics could be different in the early stage of nodule development. Therefore, the NCRs that were identified in our studied might not be the only ones that are rhythmic, and there could be some other NCRs that are rhythmically regulated in different spatial and temporal conditions.

The presence of an EE-like motif (AGACATTT) in many NCR gene promoters was an interesting but loosely understood finding. The 'T' residue in the 4th position of the EE motif, AGATATTT, is the most conserved position and highly required for CCA1/LHY and RVE1 binding in *A. thaliana* (Franco-Zorrilla et al., 2014). The 'T' to 'C' transition is, therefore, likely to disrupt or weaken LHY binding. This also disrupts the GATA sequence within the EE motif. The GATA motif has been reported to be associated with different light and nitrate dependent processes (Reyes et al., 2004). It could be that the NCR promoters containing an EE-like motif diversified from an ancestral promoter that had the EE motif and was involved in important circadian regulation. It is possible that the circadian regulation of NCRs became less required in recent time, and thus EE motif lost its conserveness. Alternatively, the reverse could be the case where the EE containing NCR promoters originated from an EE-like ancestor to enable new circadian regulation of nodulation. Also, the motif that Achom (2019) originally found (Motif-2; Figure 1.11) contains two additional highly conserved 'AG' residues at the 5' end of the motif. The significance of this additional bases could not be explained by our current knowledge on different motifs and associated binding factors. However, it should be in our consideration that our understanding on different motifs and TFs are largely based on the studies on *A. thaliana.* Since there are differences among the homologous protein sequences among different species, it is likely that there would be some differences among their binding sites as well. This additional 'AG' could be a *M truncatula* associated modification in evening element, that could be important for the binding of Mt*LHY* or Mt*RVEs*. These are, of course, initial hypotheses and need to be examined by extensive analysis of different protein-DNA interactions through specialized techniques (e.g., ChIP-seq) incorporating different *M. truncatula* accessions, other IRLC species and legumes.

7.3. *LHY* could be redundant for growth regulation but important for nodulation

Similarly to the case in *M. truncatula*, a single homolog for both *CCA1* and *LHY* was found in several plant species (e.g., rice, garden pea, papaya) (Song et al., 2010, Takata et al., 2010). In some other species, multiple homologs for *LHY* gene have been found (e.g., soybean, duckweed, cottonwood) (Wang et al., 2021, Song et al., 2010, Takata et al., 2009). A brief phylogenetic analysis done in our lab (Achom et al., 2021) has demonstrated that the plants in Brassicaceae family (e.g., rapeseed, arabidopsis) usually contain both *CCA1* and *LHY* homologs, whereas legumes mostly contain a single *LHY* gene (e.g., *M. truncatula*) or several very closely related *LHY* paralogs (e.g., soybean has four *LHY* homologs) (Wang et al., 2021). The divergence of *CCA*1 and *LHY* from an ancestral *CCA1/LHY* genes assumed to happen in Brassicaceae through an angiosperm polyploidy event known as 'β-polyploidy'. Another polyploidy event known as 'salicoid polyploidy' probably gave rise to multiple *LHY* homologs in many plants belong to eurosids-I angiosperm lineage (Takata et al., 2010) Among other plants, this lineage includes legumes, other nodulating plant species (e.g., actinorhizal plants) and non-nodulating but symbiotic N-fixing plants, for example, *Populus trichocarpa* (Takata et al., 2010, Takata et al., 2009, Doty et al., 2016). Now, our understanding on these different orthologs of *CCA1/LHY* is far from comprehensive, therefore, at this point it is very difficult to interpret whether these duplications and diversifications of *CCA1* and *LHY* genes had any association with nodulation. However, it could be roughly assumed that the *CCA1* gene might not be involved in the evolution of nodulation since its emergence in Brassicaceae did not result in nodulation related phenomenon. On the other hand, *LHY* could have some distant roles in establishment of N-fixing symbiosis (too early to say though) since it had been duplicated and diversified in plants, including those that form N-fixing symbiosis. However, even if there is a connection, it must not be very direct since many plants have *LHY* orthologs regardless of whether they fix N or not. It is also important to consider that there are other closely related homologous clock components (e.g., different *RVE*s) that share sequence similarity and overlapping expression profiles with *CCA1/LHY*. All these genes encode proteins that contain a MYB/SANT domain with a highly conserved motif, SHAQKFF, that binds to the EE motif in the promoter of many evening expressed genes (Linde et al., 2017). All these different but related genes might share some level of functional redundancy among them. Therefore, it is difficult to assign a particular function to a specific gene. Further large-scale phylogenetic studies

incorporating all these related clock genes across many species from different plant families might shed light on whether the evolution of *LHY* had any connection with N-fixing symbiosis.

As mentioned in Chapter 4, CCA1 and LHY have partial redundancy and the ability to compensate each other's functions (Lu et al., 2009). In *A. thaliana*, loss of function of either CCA1 or LHY resulted in shorter rhythm and early flowering phenotypes. Interestingly, loss of functional CCA1 was shown to be associated with shorter hypocotyl growth. On the other hand, although increased expression of LHY resulted in longer hypocotyl, loss of LHY was not associated with growth impairment (Mizoguchi et al., 2002). This is in line with our the data presented in Chapter 4, where we found no obvious difference between *lhy* mutants and WT plants in control conditions (mock treatment). Another study in our lab (unpublished; not presented in this thesis), showed that there was no significant difference in *lhy* mutants vs WT plants in terms of hypocotyl growth when grown in soils containing N and all the essential nutrients. The *lhy* mutants have shown growth impairment only when they were forced to form nodules in low N conditions. These observations and related literature suggest that *MtLHY* might not be critical or indispensable for growth and development in favourable conditions. Therefore, it is likely that the loss of *LHY* affects nodule development primarily rather than impacting growth. Having said so, LHY functions are also likely to be partially redundant; other homologous clock components (e.g., RVEs) might compensate for some of its functions as *lhy* mutants did form functional nodules (although low in number). In general, the importance of this gene cannot be overlooked since it is a part of the central circadian oscillating loop and the loss of *LHY* functions disrupted the rhythmicity of other clock components in free-running conditions (as shown in Chapter 4), therefore, likely to disrupt many other physiological processes as well.

Clock disruption was reported to be associated with altered rhizophere microbiome community that was implicated in reduced plant growth (Hubbard et al., 2018). It would be interesting to ask if our *lhy* mutants are altered in their responses to pathogen attack in future studies. It is relevant to mention that both loss of function and overexpressing *A. thaliana lhy* mutants had shown increased susceptibility to pathogens like *Pseudomonas syringae* and *Hyaloperonospora arabidopsidis* (Zhang C et al., 2013). In this thesis, we have shown that *lhy*-mediated clock disruption was associated with reduced nodule growth (in

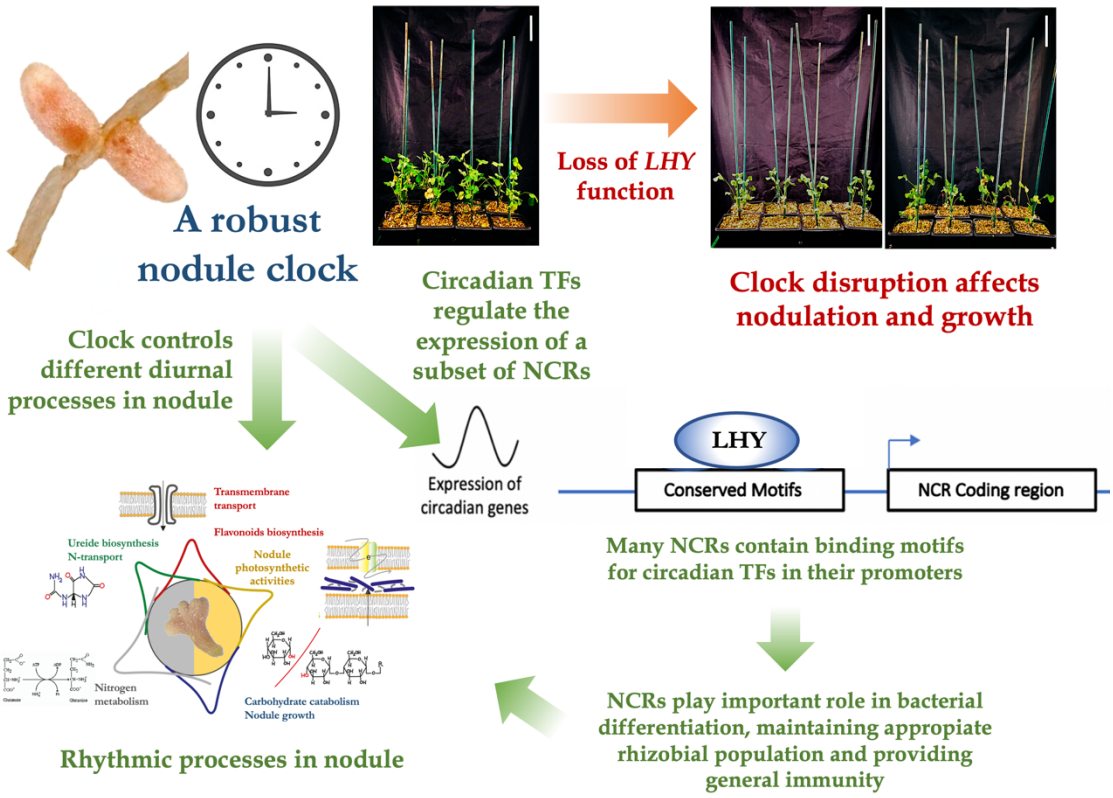


Figure 7.1: Graphical summary of the results and insights presented in this thesis. The characteristic rhythmic expression of the clock genes in nodules support the existence of a robustly functioning nodule clock. The clock is likely to regulate various diurnally regulated rhythmic processes within nodule. The disruption in clock through loss of *LHY* functioning results in reduced nodulation and growth. Most strikingly, our results suggest that the circadian clock regulates the expression of several NCRs, thus play important role in NCR mediated activities and the overall nodulation process.

terms of mass) which is supported by a recent study as well (Kong et al., 2020). Since it has been found that the plant clock is involved in the immunogenic response against pathogens (as discussed in Section 3.1.1), it can be predicted that it also plays some roles in symbiotic interaction. The reduced nodulation phenotype in *lhy* mutants could, therefore, be potentially because of disturbed clock functioning that impairs the communication between the symbiotic partners. Although the exact mode of action is not yet clear, our results confirm that the loss of *LHY* functions negatively impact nodule development and as a result, also indirectly affects overall plant growth and development.

7.4. Scope for further studies

In this work, RNAseq was used to generate a large transcriptomic dataset that enabled the gene expression profile of mature nodules to be studied; that included determination of the expression of both plant and rhizobial genes. Similarly to in many other species, most gene expression analysis in *M. truncatula* so far as been carried out using microarrays and many of these datasets were deposited in the *M. truncatula* Gene Expression Atlas (MtGEA) (He et al., 2009). Since we used RNAseq, our data are more comprehensive and likely to provide new insights compared with pre-existing data. During our analysis we mainly focused on the rhythmically expressed genes. However, the data can be further analyzed and compared with similar datasets from other tissues or different experimental conditions to identify genes that are differentially expressed in nodules in different environmental conditions or different rhizobial inoculation. For example, in our work we used a high efficiency rhizobial strain *S. meliloti* 1022 but almost all other studies used comparatively less efficient *S. meliloti* strains, e.g., 1021 or 2011 (Terpolilli et al., 2008). Therefore, our data can also be compared with those that used less efficient strains to identify genes that are responsible for developing an efficient N-fixing partnership between the host and the microsymbiont.

Apart from identifying several rhythmic NCRs, our dataset also provides gene expression information for all other known and putative NCR genes. It would be interesting to categorise the NCRs in different groups based on their level of expression, isoelectric point, size and evaluate if there is any association among the groups, since several of these features are known to be important for NCR function (Roy P et al., 2020). Based on these analyses, some candidate NCRs genes (e.g., the ones with high expression or rhythmicity) could be selected for further analysis. For example, corresponding mutant lines can be obtained to study the phenotypic and molecular effects of the disruption of NCR function.

As discussed in Chapter 5, circadian rhythms are not very common in prokaryotic species except cyanobacteria. However, in this study we have detected some rhizobial genes with circadian-like rhythmicity. It would be interesting to evaluate whether these genes are rhythmically expressed in free-living bacteria or they are only rhythmic at the

bacteroid stage. The influence of host gene on bacterial gene expression is a vastly unexplored area and so analyzing this gene expression dataset could be a starting point for some future studies after selecting candidate genes for functional characterization or studying reporter gene expression lines under candidate gene promoters.

7.5. Concluding Remarks

Nodulation is a process that requires different cells to work in coordination to turn on/off different cellular and developmental programmes in a spatial and temporal manner. It is a complex process that involved numerous known and yet poorly characterized genes, proteins, and other cellular components. Realizing its potential in achieving sustainable development goals, different research groups across the world have made a lot of discoveries in the last two decades that have advanced our understanding of legume-rhizobia interaction, nodulation and symbiotic nitrogen fixation (Roy S et al., 2020). To best of our knowledge, our study is one of the first (after Kong et al., 2020) to propose a direct molecular link between the circadian system and nodule metabolism and first to show some evidence for a possible LHY-NCR link. The preliminary data, insights, and the tools (cortex-specific vectors) developed in this study set the platform for some exciting, impactful, future-studies, that might further help to improve our understanding on beneficial plant-microbe symbiosis to achieve environment friendly, sustainable agriculture.

8. Supplementary Data

The supplementary information provided in the following tables can be accessed through this link,
https://drive.google.com/drive/folders/1Y5TzUjXOYIjAMys97s-V7WZWReTHxIio?usp=sharing

(Please, carefully copy and paste the entire link in the address bar of your web browser and remove any space that could be introduced during copying. If there is any problem in accessing the files, please email at p.roy@warwick.ac.uk)

Table S1: List of rhythmic plant genes.
Table S2: List of rhythmic rhizobial genes.
Table S3: List of NCR genes.
Table S4: Metacycle analysis and expression values (logTPM) for plant genes.
Table S5: Metacycle analysis and expression values (logTPM) for rhizobial genes.

9. References

ACHOM, M. 2019. The role of Nodule Cysteine Rich secreted peptides in controlling nodulation. *PhD Thesis,* University of Warwick.

ACHOM, M., ROY, P., LAGUNAS, B., BONYADI-POUR, R., PARDAL, A. J., BAXTER, L., ... & GIFFORD, M. L. (2021). The plant circadian clock gene *LHY* influences Medicago truncatula nodulation. *bioRxiv*.

ADAMS, S., GRUNDY, J., VEFLINGSTAD, S. R., DYER, N. P., HANNAH, M. A., OTT, S. & CARRÉ, I. A. 2018. Circadian control of abscisic acid biosynthesis and signaling pathways revealed by genome-wide analysis of LHY binding targets. *New Phytologist,* 220**,** 893-907.

ADAMS, S., MANFIELD, I., STOCKLEY, P. & CARRÉ, I. A. 2015. Revised Morning Loops of the Arabidopsis Circadian Clock Based on Analyses of Direct Regulatory Interactions. *PLoS One,* 10**,** e0143943.

ALABADI, D., OYAMA, T., YANOVSKY, M. J., HARMON, F. G., MAS, P. & KAY, S. A. 2001. Reciprocal regulation between TOC1 and LHY/CCA1 within the Arabidopsis circadian clock. *Science,* 293**,** 880-3.

ALLOING, G., MANDON, K., BONCOMPAGNI, E., MONTRICHARD, F. & FRENDO, P. 2018. Involvement of Glutaredoxin and Thioredoxin Systems in the Nitrogen-Fixing Symbiosis between Legumes and Rhizobia. *Antioxidants (Basel),* 7.

ALUNNI, B., KEVEI, Z., REDONDO-NIETO, M., KONDOROSI, A., MERGAERT, P. & KONDOROSI, E. 2007. Genomic organization and evolutionary insights on GRP and NCR genes, two large nodule-specific gene families in *Medicago truncatula*. *Mol Plant Microbe Interact,* 20**,** 1138-48.

ANDREASON, G. L. 1993. Electroporation as a technique for the transfer of macromolecules into mammalian cell lines. *Journal of tissue culture methods,* 15**,** 56-62.

ANDREWS, S. 2010. FastQC: A quality control tool for high throughput sequence data.

ARNOLD, M. F. F., PENTERMAN, J., SHABAB, M., CHEN, E. J. & WALKER, G. C. 2018. Important Late-Stage Symbiotic Role of the *Sinorhizobium meliloti* Exopolysaccharide Succinoglycan. *J Bacteriol,* 200.

ATAMIAN, H. S. & HARMER, S. L. 2016. Circadian regulation of hormone signaling and plant physiology. *Plant Molecular Biology,* 91**,** 691-702.

BAILEY, T. L., BODEN, M., BUSKE, F. A., FRITH, M., GRANT, C. E., CLEMENTI, L., REN, J., LI, W. W. & NOBLE, W. S. 2009. MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research,* 37**,** W202-W208.

BARBIER-BRYGOO, H., DE ANGELI, A., FILLEUR, S., FRACHISSE, J.M., GAMBALE, F., THOMINE, S. AND WEGE, S., 2011. Anion channels/transporters in plants: from molecular bases to regulatory networks. *Annual review of plant biology*, 62, 25-51.

BELANGER, L., DIMMICK, K. A., FLEMING, J. S. & CHARLES, T. C. 2009. Null mutations in *Sinorhizobium meliloti* exoS and chvI demonstrate the importance of this two-component regulatory system for symbiosis. *Mol Microbiol,* 74**,** 1223-37.

BENDIX, C., MARSHALL, CARINE M. & HARMON, FRANK G. 2015. Circadian Clock Genes Universally Control Key Agricultural Traits. *Molecular Plant,* 8**,** 1135-1152.

BHARDWAJ, V., MEIER, S., PETERSEN, L. N., INGLE, R. A. & RODEN, L. C. 2011. Defence Responses of *Arabidopsis thaliana* to Infection by *Pseudomonas syringae* Are Regulated by the Circadian Clock. *PLOS ONE,* 6**,** e26968.

BIRNBAUM, K., JUNG, J. W., WANG, J. Y., LAMBERT, G. M., HIRST, J. A., GALBRAITH, D. W. & BENFEY, P. N. 2005. Cell type-specific expression profiling in plants via cell sorting of protoplasts from fluorescent reporter lines. *Nat Methods,* 2**,** 615-9.

BISHOPP, A. & BENNETT, M. J. 2019. Turning lateral roots into nodules. *Science,* 366**,** 953-954.

BITTNER, A. N., FOLTZ, A. & OKE, V. 2007. Only one of five groEL genes is required for viability and successful symbiosis in *Sinorhizobium meliloti. J Bacteriol,* 189**,** 1884-9.

BOISSON-DERNIER, A., CHABAUD, M., GARCIA, F., BECARD, G., ROSENBERG, C. & BARKER, D. G. 2001. *Agrobacterium rhizogenes*-transformed roots of *Medicago truncatula* for the study of nitrogen-fixing and endomycorrhizal symbiotic associations. *Mol Plant Microbe Interact,* 14**,** 695-700.

BOLGER, A. M., LOHSE, M. & USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* 30**,** 2114-2120.

BONALDI, K., GARGANI, D., PRIN, Y., FARDOUX, J., GULLY, D., NOUWEN, N., GOORMACHTIG, S. & GIRAUD, E. 2011. Nodulation of *Aeschynomene afraspera* and *A. indica* by Photosynthetic *Bradyrhizobium Sp*. Strain ORS285: The Nod-Dependent Versus the Nod-Independent Symbiotic Interaction. *Molecular Plant-Microbe Interactions®,* 24**,** 1359-1371.

BORDAGE, S., SULLIVAN, S., LAIRD, J., MILLAR, A. J. & NIMMO, H. G. 2016. Organ specificity in the plant circadian system is explained by different light inputs to the shoot and root clocks. *New Phytologist,* 212**,** 136-149.

BRAZEL, A. J. & Ó'MAOILÉIDIGH, D. S. 2019. Photosynthetic activity of reproductive organs. *Journal of Experimental Botany,* 70**,** 1737-1754.

BRUNKARD, J.O. & ZAMBRYSKI, P., 2019. Plant cell-cell transport via plasmodesmata is regulated by light and the circadian clock. *Plant physiology*, 181(4), 1459-1467.

BUCHER, M., BRUNNER, S., ZIMMERMANN, P., ZARDI, G. I., AMRHEIN, N., WILLMITZER, L. & RIESMEIER, J. W. 2002. The Expression of an Extensin-Like Protein Correlates with Cellular Tip Growth in Tomato. *Plant Physiology,* 128**,** 911-923.

CAO, Y., HALANE, M. K., GASSMANN, W. & STACEY, G. 2017. The Role of Plant Innate Immunity in the Legume-Rhizobium Symbiosis. *Annu Rev Plant Biol,* 68**,** 535-561.

CARTER, A. D., BONYADI, R. & GIFFORD, M. L. 2013. The use of fluorescence-activated cell sorting in studying plant development and environmental responses. *Int J Dev Biol,* 57**,** 545-52.

CARVALHO, A. D. O. & GOMES, V. M. 2009. Plant defensins—Prospects for the biological functions and biotechnological properties. *Peptides,* 30**,** 1007-1020.

CASSON, S., SPENCER, M., WALKER, K. & LINDSEY, K. 2005. Laser capture microdissection for the analysis of gene expression during embryogenesis of Arabidopsis. *Plant J,* 42**,** 111-23.

CASSON, S. A., SPENCER, M.W., LINDSEY, K. 2008. Laser-Capture Microdissection to Study Global Transcriptional Changes During Plant Embryogenesis. *In:* SUAREZ, M. F., BOZHKOV, P.V. (ed.) *Plant Embryogenesis. Methods in Molecular Biology.* Humana Press.

CHARPENTIER, M., BREDEMEIER, R., WANNER, G., TAKEDA, N., SCHLEIFF, E. & PARNISKE, M. 2008. *Lotus japonicus* CASTOR and POLLUX Are Ion

Channels Essential for Perinuclear Calcium Spiking in Legume Root Endosymbiosis. *The Plant Cell,* 20**,** 3467-3479.
CHARPENTIER, M. & OLDROYD, G. E. D. 2013. Nuclear Calcium Signaling in Plants. *Plant Physiology,* 163**,** 496-503.
CHEN, A. H., LUBKOWICZ, D., YEONG, V., CHANG, R. L. & SILVER, P. A. 2015a. Transplantability of a circadian clock to a noncircadian organism. *Sci Adv,* 1.
CHEN, D. S., LIU, C. W., ROY, S., COUSINS, D., STACEY, N. & MURRAY, J. D. 2015b. Identification of a core set of rhizobial infection genes using data from single cell-types. *Front Plant Sci,* 6**,** 575.
CHEN, J., MOREAU, C., LIU, Y., KAWAGUCHI, M., HOFER, J., ELLIS, N. & CHEN, R. 2012. Conserved genetic determinant of motor organ identity in *Medicago truncatula* and related legumes. *Proc Natl Acad Sci U S A,* 109**,** 11723-8.
CHENG, H. P. & WALKER, G. C. 1998. Succinoglycan is required for initiation and elongation of infection threads during nodulation of alfalfa by *Rhizobium meliloti*. *J Bacteriol,* 180**,** 5183-91.
CHIASSON, D. M., LOUGHLIN, P. C., MAZURKIEWICZ, D., MOHAMMADIDEHCHESHMEH, M., FEDOROVA, E. E., OKAMOTO, M., MCLEAN, E., GLASS, A. D. M., SMITH, S. E., BISSELING, T., TYERMAN, S. D., DAY, D. A. & KAISER, B. N. 2014. Soybean SAT1 (Symbiotic Ammonium Transporter 1) encodes a bHLH transcription factor involved in nodule growth and NH4+ transport. *Proceedings of the National Academy of Sciences,* 111**,** 4814-4819.
CHOW, C. N., LEE, T. Y., HUNG, Y. C., LI, G. Z., TSENG, K. C., LIU, Y. H., KUO, P. L., ZHENG, H. Q. & CHANG, W. C. 2019. PlantPAN3.0: a new and updated resource for reconstructing transcriptional regulatory networks from ChIP-seq experiments in plants. *Nucleic Acids Res,* 47**,** D1155-d1163.
COBA DE LA PEÑA, T., FEDOROVA, E., PUEYO, J. J. & LUCAS, M. M. 2017. The Symbiosome: Legume and Rhizobia Co-evolution toward a Nitrogen-Fixing Organelle? *Front Plant Sci,* 8**,** 2229.
COLLIER, J. 2012. Regulation of chromosomal replication in *Caulobacter crescentus*. *Plasmid,* 67**,** 76-87.
COVINGTON, M. F. & HARMER, S. L. 2007. The circadian clock regulates auxin signaling and responses in Arabidopsis. *PLoS Biol,* 5**,** e222.
COVINGTON, M. F., MALOOF, J. N., STRAUME, M., KAY, S. A. & HARMER, S. L. 2008. Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. *Genome Biol,* 9**,** R130.
CRESPO-RIVAS, J. C., GUEFRACHI, I., MOK, K. C., VILLAECIJA-AGUILAR, J. A., ACOSTA-JURADO, S., PIERRE, O., RUIZ-SAINZ, J. E., TAGA, M. E., MERGAERT, P. & VINARDELL, J. M. 2016. *Sinorhizobium fredii* HH103 bacteroids are not terminally differentiated and show altered O-antigen in nodules of the Inverted Repeat-Lacking Clade legume Glycyrrhiza uralensis. *Environ Microbiol,* 18**,** 2392-404.
CROOK, M. B., LINDSAY, D. P., BIGGS, M. B., BENTLEY, J. S., PRICE, J. C., CLEMENT, S. C., CLEMENT, M. J., LONG, S. R. & GRIFFITTS, J. S. 2012. Rhizobial plasmids that cause impaired symbiotic nitrogen fixation and enhanced host invasion. *Mol Plant Microbe Interact,* 25**,** 1026-33.
CUI, Y., BARAMPURAM, S., STACEY, M. G., HANCOCK, C. N., FINDLEY, S., MATHIEU, M., ZHANG, Z., PARROTT, W. A. & STACEY, G. 2013. Tnt1 retrotransposon mutagenesis: a tool for soybean functional genomics. *Plant physiology,* 161**,** 36-47.

CUSHNIE, T. P. & LAMB, A. J. 2005. Antimicrobial activity of flavonoids. *Int J Antimicrob Agents,* 26**,** 343-56.

CZERNIC, P., GULLY, D., CARTIEAUX, F., et al. 2015. Convergent Evolution of Endosymbiont Differentiation in Dalbergioid and Inverted Repeat-Lacking Clade Legumes Mediated by Nodule-Specific Cysteine-Rich Peptides. *Plant Physiol,* 169**,** 1254-65.

DAI, X., SHEN, Z., WANG, Y. & ZHU, M. 2018. *Sinorhizobium meliloti*, a Slow-Growing Bacterium, Exhibits Growth Rate Dependence of Cell Size under Nutrient Limitation. *mSphere,* 3.

DALLA VIA, V., NARDUZZI, C., AGUILAR, O. M., ZANETTI, M. E. & BLANCO, F. A. 2015. Changes in the Common Bean Transcriptome in Response to Secreted and Surface Signal Molecules of *Rhizobium etli. Plant Physiol,* 169**,** 1356-70.

DAVIDSON, A.L., DASSA, E., ORELLE, C. & CHEN, J., 2008. Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiology and molecular biology reviews*, 72(2), 317-364.

DEIKMAN, J. & HAMMER, P. E. 1995. Induction of Anthocyanin Accumulation by Cytokinins in *Arabidopsis thaliana. Plant Physiol,* 108**,** 47-57.

DOWNIE, J. A. 2010. The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots. *FEMS Microbiol Rev,* 34**,** 150-70.

DOTY, S.L., SHER, A.W., FLECK, N.D., KHORASANI, M., BUMGARNER, R.E., KHAN, Z., KO, A.W., KIM, S.H. & DELUCA, T.H., 2016. Variable nitrogen fixation in wild Populus. *PLoS One*, 11(5), e0155979

DURGO, H., KLEMENT, E., HUNYADI-GULYAS, E., SZUCS, A., KERESZT, A., MEDZIHRADSZKY, K. F. & KONDOROSI, E. 2015. Identification of nodule-specific cysteine-rich plant peptides in endosymbiotic bacteria. *Proteomics,* 15**,** 2291-5.

DVORNYK, V., VINOGRADOVA, O. & NEVO, E. 2003. Origin and evolution of circadian clock genes in prokaryotes. *Proceedings of the National Academy of Sciences,* 100**,** 2495-2500.

EDWARDS, K. D., ANDERSON, P. E., HALL, A., SALATHIA, N. S., LOCKE, J. C., LYNN, J. R., STRAUME, M., SMITH, J. Q. & MILLAR, A. J. 2006. FLOWERING LOCUS C mediates natural variation in the high-temperature response of the Arabidopsis circadian clock. *Plant Cell,* 18**,** 639-50.

ENDO, M., SHIMIZU, H., NOHALES, M. A., ARAKI, T. & KAY, S. A. 2014. Tissue-specific clocks in Arabidopsis show asymmetric coupling. *Nature,* 515**,** 419-422.

ESSELING, J. J., LHUISSIER, F. G. P. & EMONS, A. M. C. 2003. Nod Factor-Induced Root Hair Curling: Continuous Polar Growth towards the Point of Nod Factor Application. *Plant Physiology,* 132**,** 1982-1988.

FARKAS, A., MAROTI, G., DURGO, H., GYORGYPAL, Z., LIMA, R. M., MEDZIHRADSZKY, K. F., KERESZT, A., MERGAERT, P. & KONDOROSI, E. 2014. *Medicago truncatula* symbiotic peptide NCR247 contributes to bacteroid differentiation through multiple mechanisms. *Proc Natl Acad Sci U S A,* 111**,** 5183-8.

FARKAS, A., PAP, B., KONDOROSI, É. & MARÓTI, G. 2018. Antimicrobial Activity of NCR Plant Peptides Strongly Depends on the Test Assays. *Front Microbiol,* 9**,** 2600.

FRANCO-ZORRILLA, J. M., LÓPEZ-VIDRIERO, I., CARRASCO, J. L., GODOY, M., VERA, P. & SOLANO, R. 2014. DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proceedings of the National Academy of Sciences,* 111**,** 2367-2372.

FRANSSEN, H. J., VIJN, I., YANG, W. C. & BISSELING, T. 1992. Developmental aspects of the Rhizobium-legume symbiosis. *In:* SCHILPEROORT, R. A. & DURE, L. (eds.) *10 Years Plant Molecular Biology*. Dordrecht: Springer Netherlands.

FU, X., ADAMS, Z., LIU, R., HEPOWIT, N. L., WU, Y., BOWMANN, C. F., MOSKOVITZ, J. & MAUPIN-FURLOW, J. A. 2017. Methionine Sulfoxide Reductase A (MsrA) and Its Function in Ubiquitin-Like Protein Modification in <em>Archaea</em&gt. *mBio,* 8**,** e01169-17.

GAGE, D. J. 2004. Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *Microbiol Mol Biol Rev,* 68**,** 280-300.

GAVRILOVIC, S., YAN, Z., JURKIEWICZ, A. M., STOUGAARD, J. & MARKMANN, K. 2016. Inoculation insensitive promoters for cell type enriched gene expression in legume roots and nodules. *Plant Methods,* 12**,** 4.

GENDRON, J. M., PRUNEDA-PAZ, J. L., DOHERTY, C. J., GROSS, A. M., KANG, S. E. & KAY, S. A. 2012. Arabidopsis circadian clock protein, TOC1, is a DNA-binding transcription factor. *Proceedings of the National Academy of Sciences,* 109**,** 3167.

GENRE, A. 2008. Confocal microscopy and plant cell biology: A perfect match. *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology,* 142**,** 348-354.

GEPTS, P., BEAVIS, W. D., BRUMMER, E. C., SHOEMAKER, R. C., STALKER, H. T., WEEDEN, N. F. & YOUNG, N. D. 2005. Legumes as a Model Plant Family. Genomics for Food and Feed Report of the Cross-Legume Advances through Genomics Conference. *Plant Physiology,* 137**,** 1228-1235.

GIBON, Y., BLÄSING, O. E., PALACIOS-ROJAS, N., PANKOVIC, D., HENDRIKS, J. H. M., FISAHN, J., HÖHNE, M., GÜNTHER, M. & STITT, M. 2004. Adjustment of diurnal starch turnover to short days: depletion of sugar during the night leads to a temporary inhibition of carbohydrate utilization, accumulation of sugars and post-translational activation of ADP-glucose pyrophosphorylase in the following light period. *The Plant Journal,* 39**,** 847-862.

GIFFORD, M. L., DEAN, A., GUTIERREZ, R. A., CORUZZI, G. M. & BIRNBAUM, K. D. 2008. Cell-specific nitrogen responses mediate developmental plasticity. *Proceedings of the National Academy of Sciences,* 105**,** 803-808.

GLAZEBROOK, J., ICHIGE, A. & WALKER, G. C. 1993. A *Rhizobium meliloti* homolog of the *Escherichia coli* peptide-antibiotic transport protein SbmA is essential for bacteroid development. *Genes Dev,* 7**,** 1485-97.

GLYNN, E. F., CHEN, J. & MUSHEGIAN, A. R. 2006. Detecting periodic patterns in unevenly spaced gene expression time series using Lomb–Scargle periodograms. *Bioinformatics,* 22**,** 310-316.

GOEDHART, J., HINK, M. A., VISSER, A. J., BISSELING, T. & GADELLA, T. W., JR. 2000. In vivo fluorescence correlation microscopy (FCM) reveals accumulation and immobilization of Nod factors in root hair cell walls. *Plant J,* 21**,** 109-19.

GONZALEZ-RIZZO, S., CRESPI, M. & FRUGIER, F. 2006. The *Medicago truncatula* CRE1 Cytokinin Receptor Regulates Lateral Root Development and Early Symbiotic Interaction with *Sinorhizobium meliloti*. *Plant Cell,* 18**,** 2680-93.

GONZALEZ-SAMA, A., DE LA PENA, T. C., KEVEI, Z., MERGAERT, P., LUCAS, M. M., DE FELIPE, M. R., KONDOROSI, E. & PUEYO, J. J. 2006. Nuclear DNA endoreduplication and expression of the mitotic inhibitor Ccs52 associated to determinate and lupinoid nodule organogenesis. *Mol Plant Microbe Interact,* 19**,** 173-80.

GOODSPEED, D., CHEHAB, E. W., MIN-VENDITTI, A., BRAAM, J. & COVINGTON, M. F. 2012. Arabidopsis synchronizes jasmonate-mediated

defense with insect circadian behaviour. *Proceedings of the National Academy of Sciences,* 109**,** 4674-4677.

GOODSPEED, D., LIU, J. D., CHEHAB, E. W., SHENG, Z., FRANCISCO, M., KLIEBENSTEIN, D. J. & BRAAM, J. 2013. Postharvest circadian entrainment enhances crop pest resistance and phytochemical cycling. *Curr Biol,* 23**,** 1235-41.

GRAF, A., SCHLERETH, A., STITT, M. & SMITH, A. M. 2010. Circadian control of carbohydrate availability for growth in Arabidopsis plants at night. *Proceedings of the National Academy of Sciences,* 107**,** 9458-9463.

GREENHAM, K. & MCCLUNG, C. R. 2015. Integrating circadian dynamics with physiological processes in plants. *Nature Reviews Genetics,* 16**,** 598-610.

GRIFFITTS, J. S., CARLYON, R. E., ERICKSON, J. H., MOULTON, J. L., BARNETT, M. J., TOMAN, C. J. & LONG, S. R. 2008. A *Sinorhizobium meliloti* osmosensory two-component system required for cyclic glucan export and symbiosis. *Molecular Microbiology,* 69**,** 479-490.

GUEFRACHI, I., NAGYMIHALY, M., PISLARIU, C. I., VAN DE VELDE, W., RATET, P., MARS, M., UDVARDI, M. K., KONDOROSI, E., MERGAERT, P. & ALUNNI, B. 2014. Extreme specificity of NCR gene expression in *Medicago truncatula. BMC Genomics,* 15**,** 712.

GUEFRACHI, I., PIERRE, O., TIMCHENKO, T., ALUNNI, B., BARRIERE, Q., CZERNIC, P., VILLAECIJA-AGUILAR, J. A., VERLY, C., BOURGE, M., FARDOUX, J., MARS, M., KONDOROSI, E., GIRAUD, E. & MERGAERT, P. 2015. Bradyrhizobium BclA Is a Peptide Transporter Required for Bacterial Differentiation in Symbiosis with Aeschynomene Legumes. *Mol Plant Microbe Interact,* 28**,** 1155-66.

HAAG, A. F., BALOBAN, M., SANI, M., KERSCHER, B., PIERRE, O., FARKAS, A., LONGHI, R., BONCOMPAGNI, E., HEROUART, D., DALL'ANGELO, S., KONDOROSI, E., ZANDA, M., MERGAERT, P. & FERGUSON, G. P. 2011. Protection of Sinorhizobium against host cysteine-rich antimicrobial peptides is critical for symbiosis. *PLoS Biol,* 9**,** e1001169.

HALBERG, F. & CONNER, R. L. Circadian organization and microbiology: Variance spectra and a periodogram on behaviour of *Escherichia coli* growing in fluid culture. Proc Minn Acad Sci, 1961. 227-239.

HARMER, S. L., HOGENESCH, J. B., STRAUME, M., CHANG, H.-S., HAN, B., ZHU, T., WANG, X., KREPS, J. A. & KAY, S. A. 2000. Orchestrated Transcription of Key Pathways in *Arabidopsis* by the Circadian Clock. *Science,* 290**,** 2110-2113.

HE, J., BENEDITO, V. A., WANG, M., MURRAY, J. D., ZHAO, P. X., TANG, Y. & UDVARDI, M. K. 2009. The *Medicago truncatula* gene expression atlas web server. *BMC Bioinformatics,* 10**,** 441.

HE, X. & ZHANG, J. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics,* 169**,** 1157-64.

HECHT, V., KNOWLES, C. L., VANDER SCHOOR, J. K., LIEW, L. C., JONES, S. E., LAMBERT, M. J. & WELLER, J. L. 2007. Pea LATE BLOOMER1 is a GIGANTEA ortholog with roles in photoperiodic flowering, deetiolation, and transcriptional regulation of circadian clock gene homologs. *Plant Physiol,* 144**,** 648-61.

HILLMAN, W. S. 1976. Biological Rhythms and Physiological Timing. *Annual Review of Plant Physiology,* 27**,** 159-179.

HORVATH, B., DOMONKOS, A., KERESZT, A., SZUCS, A., ABRAHAM, E., AYAYDIN, F., BOKA, K., CHEN, Y., CHEN, R., MURRAY, J. D., UDVARDI,

M. K., KONDOROSI, E. & KALO, P. 2015. Loss of the nodule-specific cysteine rich peptide, NCR169, abolishes symbiotic nitrogen fixation in the *Medicago truncatula* dnf7 mutant. *Proc Natl Acad Sci U S A,* 112**,** 15232-7.

HSU, P. Y., DEVISETTY, U. K. & HARMER, S. L. 2013. Accurate timekeeping is controlled by a cycling activator in Arabidopsis. *Elife,* 2**,** e00473.

HSU, P. Y. & HARMER, S. L. 2014. Wheels within wheels: the plant circadian system. *Trends Plant Sci,* 19**,** 240-9.

HUA, J. 2013. Modulation of plant immunity by light, circadian rhythm, and temperature. *Curr Opin Plant Biol,* 16**,** 406-13.

HUANG, W., PEREZ-GARCIA, P., POKHILKO, A., MILLAR, A. J., ANTOSHECHKIN, I., RIECHMANN, J. L. & MAS, P. 2012. Mapping the core of the Arabidopsis circadian clock defines the network structure of the oscillator. *Science,* 336**,** 75-9.

HUBBARD, C. J., BROCK, M. T., VAN DIEPEN, L. T. A., MAIGNIEN, L., EWERS, B. E. & WEINIG, C. 2018. The plant circadian clock influences rhizosphere community structure and function. *The ISME Journal,* 12**,** 400-410.

HWANGBO, K., SON, S. H., LEE, J. S., MIN, S. R., KO, S. M., LIU, J. R., CHOI, D. & JEONG, W. J. 2010. Rapid and simple method for DNA extraction from plant and algal species suitable for PCR amplification using a chelating resin Chelex 100. *Plant Biotechnology Reports,* 4**,** 49-52.

IWASAKI, H., NISHIWAKI, T., KITAYAMA, Y., NAKAJIMA, M. & KONDO, T. 2002. KaiA-stimulated KaiC phosphorylation in circadian timing loops in cyanobacteria. *Proceedings of the National Academy of Sciences,* 99**,** 15788-15793.

JAMET, A., SIGAUD, S., VAN DE SYPE, G., PUPPO, A. & HEROUART, D. 2003. Expression of the bacterial catalase genes during *Sinorhizobium meliloti-Medicago sativa* symbiosis and their crucial role during the infection process. *Mol Plant Microbe Interact,* 16**,** 217-25.

JAYASEKERA, M.M., FOLTIN, S.K., OLSON, E.R. & HOLLER, T.P., 2000. Escherichia coli requires the protease activity of FtsH for growth. *Archives of biochemistry and biophysics*, 380(1), 103-107.

JIN, Y., CHEN, Z., YANG, J., MYSORE, K. S., WEN, J., HUANG, J., YU, N. & WANG, E. 2018. IPD3 and IPD3L Function Redundantly in Rhizobial and Mycorrhizal Symbioses. *Frontiers in Plant Science,* 9.

KAKAR, K., WANDREY, M., CZECHOWSKI, T., GAERTNER, T., SCHEIBLE, W.-R., STITT, M., TORRES-JEREZ, I., XIAO, Y., REDMAN, J. C., WU, H. C., CHEUNG, F., TOWN, C. D. & UDVARDI, M. K. 2008. A community resource for high-throughput quantitative RT-PCR analysis of transcription factor gene expression in Medicago truncatula. *Plant Methods,* 4**,** 18.

KARIMI, M., INZÉ, D. & DEPICKER, A. 2002. GATEWAY vectors for Agrobacterium mediated plant transformation. *Trends in Plant Science,* 7**,** 193-195.

KARPINSKI, S., GABRYS, H., MATEO, A., KARPINSKA, B. & MULLINEAUX, P. M. 2003. Light perception in plant disease defence signaling. *Curr Opin Plant Biol,* 6**,** 390-6.

KATJA KAPP, S. S., MARIUS K. LEMBERG, BERNHARD DOBBERSTEIN 2009. Post-targeting functions of signal peptides. *In:* ZIMMERMANN, R. (ed.) *Protein Transport into the Endoplasmic Reticulum.* Landes Biosciences.

KAWAHARADA, Y., JAMES, E. K., KELLY, S., SANDAL, N. & STOUGAARD, J. 2017. The Ethylene Responsive Factor Required for Nodulation 1 (ERN1) Transcription Factor Is Required for Infection-Thread Formation in Lotus japonicus. *Mol Plant Microbe Interact,* 30**,** 194-204.

KAWAMOTO, N., ITO, H., TOKUDA, I. T. & IWASAKI, H. 2020. Damped circadian oscillation in the absence of KaiA in Synechococcus. *Nat Commun,* 11**,** 2242.

KIM, D., LANGMEAD, B. & SALZBERG, S. L. 2015a. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods,* 12**,** 357-360.

KIM, J. A., KIM, H. S., CHOI, S. H., JANG, J. Y., JEONG, M. J. & LEE, S. I. 2017. The Importance of the Circadian Clock in Regulating Plant Metabolism. *Int J Mol Sci,* 18.

KIM, M., CHEN, Y., XI, J., WATERS, C., CHEN, R. & WANG, D. 2015b. An antimicrobial peptide essential for bacterial survival in the nitrogen-fixing symbiosis. *Proceedings of the National Academy of Sciences,* 112**,** 15238-15243.

KIM, W. Y., FUJIWARA, S., SUH, S. S., KIM, J., KIM, Y., HAN, L., DAVID, K., PUTTERILL, J., NAM, H. G. & SOMERS, D. E. 2007. ZEITLUPE is a circadian photoreceptor stabilized by GIGANTEA in blue light. *Nature,* 449**,** 356-60.

KITAYAMA, Y., IWASAKI, H., NISHIWAKI, T. & KONDO, T. 2003. KaiB functions as an attenuator of KaiC phosphorylation in the cyanobacterial circadian clock system. *Embo j,* 22**,** 2127-34.

KOHLEN, W., NG, J. L. P., DEINUM, E. E. & MATHESIUS, U. 2017. Auxin transport, metabolism, and signaling during nodule initiation: indeterminate and determinate nodules. *Journal of Experimental Botany,* 69**,** 229-244.

KONDO, T., TSINOREMAS, N., GOLDEN, S., JOHNSON, C., KUTSUNA, S. & ISHIURA, M. 1994. Circadian clock mutants of cyanobacteria. *Science,* 266**,** 1233-1236.

KONDOROSI, E., ROUDIER, F. & GENDREAU, E. 2000. Plant cell-size control: growing by ploidy? *Curr Opin Plant Biol,* 3**,** 488-92.

KONG, Y., HAN, L., LIU, X., WANG, H., WEN, L., YU, X., XU, X., KONG, F., FU, C., MYSORE, K. S., WEN, J. & ZHOU, C. 2020. The nodulation and nyctinastic leaf movement is orchestrated by clock gene LHY in *Medicago truncatula. Journal of Integrative Plant Biology,* 62**,** 1880-1895.

KORNELI, C., DANISMAN, S. & STAIGER, D. 2014. Differential Control of Pre-Invasive and Post-Invasive Antibacterial Defense by the Arabidopsis Circadian Clock. *Plant and Cell Physiology,* 55**,** 1613-1622.

LAGUNAS, B., SCHAFER, P. & GIFFORD, M. L. 2015. Housing helpful invaders: the evolutionary and molecular architecture underlying plant root-mutualist microbe interactions. *J Exp Bot,* 66**,** 2177-86.

LAGUNAS B., W. L., HUSSAIN R.M.F., HANDS-PORTMAN I., WOOLLEY-ALLEN K., GIIFORD M.L. 2018. Histological Profiling Over Time to Optimize Root Cell Type-Specific Reporter Lines for Cell Sorting. *In:* RISTOVA D., B. E. (ed.) *Root Development. Methods in Molecular Biology.* New York, NY: Humana Press.

LAI, A. G., DOHERTY, C. J., MUELLER-ROEBER, B., KAY, S. A., SCHIPPERS, J. H. & DIJKWEL, P. P. 2012. CIRCADIAN CLOCK-ASSOCIATED 1 regulates ROS homeostasis and oxidative stress responses. *Proc Natl Acad Sci U S A,* 109**,** 17129-34.

LALOUM, D. & ROBINSON-RECHAVI, M. 2020. Methods detecting rhythmic gene expression are biologically relevant only for strong signal. *PLOS Computational Biology,* 16**,** e1007666.

LAMOUCHE, F., CHAUMERET, A., GUEFRACHI, I., BARRIÈRE, Q., PIERRE, O., GUÉRARD, F., GILARD, F., GIRAUD, E., DESSAUX, Y., GAKIÈRE, B., TIMCHENKO, T., KERESZT, A., MERGAERT, P. & ALUNNI, B. 2019a. From Intracellular Bacteria to Differentiated Bacteroids: Transcriptome and

Metabolome Analysis in Aeschynomene Nodules Using the Bradyrhizobium sp. Strain ORS285 bclA Mutant. *Journal of Bacteriology,* 201**,** e00191-19.

LAMOUCHE, F., GULLY, D., CHAUMERET, A., et al. 2019b. Transcriptomic dissection of *Bradyrhizobium sp.* strain ORS285 in symbiosis with Aeschynomene spp. inducing different bacteroid morphotypes with contrasted symbiotic efficiency. *Environmental Microbiology,* 21**,** 3244-3258.

LAN, G., WOLGEMUTH, C. W. & SUN, S. X. 2007. Z-ring force and cell shape during division in rod-like bacteria. *Proceedings of the National Academy of Sciences,* 104**,** 16110-16115.

LANG, C. & LONG, S. R. 2015. Transcriptomic Analysis of *Sinorhizobium meliloti* and *Medicago truncatula* Symbiosis Using Nitrogen Fixation-Deficient Nodules. *Mol Plant Microbe Interact,* 28**,** 856-68.

LAVIN, M., DOYLE, J. J. & PALMER, J. D. 1990. Evolutionary significance of the loss of the chloroplast-DNA inverted repeat in the leguminosae subfamily papilionoideae. *Evolution,* 44**,** 390-402.

LEE, J. Y., COLINAS, J., WANG, J. Y., MACE, D., OHLER, U. & BENFEY, P. N. 2006. Transcriptional and posttranscriptional regulation of transcription factor expression in Arabidopsis roots. *Proc Natl Acad Sci U S A,* 103**,** 6055-60.

LEVIER, K. & WALKER, G. C. 2001. Genetic analysis of the *Sinorhizobium meliloti* BacA protein: differential effects of mutations on phenotypes. *J Bacteriol,* 183**,** 6444-53.

LI, M., CAO, L., MWIMBA, M., ZHOU, Y., LI, L., ZHOU, M., SCHNABLE, P. S., O'ROURKE, J. A., DONG, X. & WANG, W. 2019. Comprehensive mapping of abiotic stress inputs into the soybean circadian clock. *Proceedings of the National Academy of Sciences,* 116**,** 23840-23849.

LI, M.-W. & LAM, H.-M. 2020. The Modification of Circadian Clock Components in Soybean During Domestication and Improvement. *Frontiers in Genetics,* 11.

LIAO, Y., SMYTH, G. K. & SHI, W. 2013. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics,* 30**,** 923-930.

LIEW, L. C., HECHT, V., LAURIE, R. E., KNOWLES, C. L., VANDER SCHOOR, J. K., MACKNIGHT, R. C. & WELLER, J. L. 2009. DIE NEUTRALIS and LATE BLOOMER 1 contribute to regulation of the pea circadian clock. *Plant Cell,* 21**,** 3198-211.

LINDE, A. M., EKLUND, D. M., KUBOTA, A., PEDERSON, E. R. A., HOLM, K., GYLLENSTRAND, N., NISHIHAMA, R., CRONBERG, N., MURANAKA, T., OYAMA, T., KOHCHI, T. & LAGERCRANTZ, U. 2017. Early evolution of the land plant circadian clock. *New Phytol,* 216**,** 576-590.

LIU, Y., TSINOREMAS, N. F., JOHNSON, C. H., LEBEDEVA, N. V., GOLDEN, S. S., ISHIURA, M. & KONDO, T. 1995. Circadian orchestration of gene expression in cyanobacteria. *Genes Dev,* 9**,** 1469-78.

LIVAK, K. J. & SCHMITTGEN, T. D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods,* 25**,** 402-8.

LONG, S. R. 1996. Rhizobium symbiosis: nod factors in perspective. *Plant Cell,* 8**,** 1885-98.

LOPEZ-GOMEZ, M., SANDAL, N., STOUGAARD, J. and BOLLER, T. 2012. Interplay of flg22-induced defence responses and nodulation in *Lotus japonicus*. *Journal of experimental botany*, 63(1), 393-401.

LOZA-CORREA, M., GOMEZ-VALERO, L. & BUCHRIESER, C. 2010. Circadian clock proteins in prokaryotes: hidden rhythms? *Front Microbiol,* 1**,** 130.

LU, H., MCCLUNG, C. R. & ZHANG, C. 2017. Tick Tock: Circadian Regulation of Plant Innate Immunity. *Annu Rev Phytopathol,* 55**,** 287-311.

LU, S., DONG, L., FANG, C., et al. 2020. Stepwise selection on homeologous PRR genes controlling flowering and maturity during soybean domestication. *Nature Genetics,* 52**,** 428-436.

LU, S. X., KNOWLES, S. M., ANDRONIS, C., ONG, M. S. & TOBIN, E. M. 2009. CIRCADIAN CLOCK ASSOCIATED1 and LATE ELONGATED HYPOCOTYL function synergistically in the circadian clock of Arabidopsis. *Plant Physiol,* 150**,** 834-43.

MA, X., EHRHARDT, D. W. & MARGOLIN, W. 1996. Colocalization of cell division proteins FtsZ and FtsA to cytoskeletal structures in living *Escherichia coli* cells by using green fluorescent protein. *Proc Natl Acad Sci U S A,* 93**,** 12998-3003.

MARCHETTI, M., CATRICE, O., BATUT, J. & MASSON-BOIVIN, C. 2011. *Cupriavidus taiwanensis* bacteroids in *Mimosa pudica* Indeterminate nodules are not terminally differentiated. *Appl Environ Microbiol,* 77**,** 2161-4.

MARGOLIN, W. 2005. FtsZ and the division of prokaryotic cells and organelles. *Nat Rev Mol Cell Biol,* 6**,** 862-71.

MAROTI, G. & KONDOROSI, E. 2014. Nitrogen-fixing Rhizobium-legume symbiosis: are polyploidy and host peptide-governed symbiont differentiation general principles of endosymbiosis? *Front Microbiol,* 5**,** 326.

MARX, H., MINOGUE, C. E., JAYARAMAN, D., RICHARDS, A. L., KWIECIEN, N. W., SIAHPIRANI, A. F., RAJASEKAR, S., MAEDA, J., GARCIA, K., DEL VALLE-ECHEVARRIA, A. R., VOLKENING, J. D., WESTPHALL, M. S., ROY, S., SUSSMAN, M. R., ANE, J. M. & COON, J. J. 2016. A proteomic atlas of the legume *Medicago truncatula* and its nitrogen-fixing endosymbiont *Sinorhizobium meliloti. Nat Biotechnol,* 34**,** 1198-1205.

MCDONALD, K. 2011. *Biologists discover an 'Evening' protein complex that regulates plant growth* [Online]. UCSanDiego. Available: https://ucsdnews.ucsd.edu/archive/newsrel/science/2011_07evening-complex.asp [Accessed 19 Jan 2021 2021].

MELOTTO, M., UNDERWOOD, W., KOCZAN, J., NOMURA, K. & HE, S. Y. 2006. Plant Stomata Function in Innate Immunity against Bacterial Invasion. *Cell,* 126**,** 969-980.

MERGAERT, P. 2018. Role of antimicrobial peptides in controlling symbiotic bacterial populations. *Nat Prod Rep,* 35**,** 336-356.

MERGAERT, P., KERESZT, A. & KONDOROSI, E. 2020. Gene Expression in Nitrogen-Fixing Symbiotic Nodule Cells in *Medicago truncatula* and Other Nodulating Plants. *The Plant Cell,* 32**,** 42.

MERGAERT, P., KIKUCHI, Y., SHIGENOBU, S. & NOWACK, E. C. M. 2017. Metabolic Integration of Bacterial Endosymbionts through Antimicrobial Peptides. *Trends Microbiol,* 25**,** 703-712.

MERGAERT, P., NIKOVICS, K., KELEMEN, Z., MAUNOURY, N., VAUBERT, D., KONDOROSI, A. & KONDOROSI, E. 2003. A novel family in *Medicago truncatula* consisting of more than 300 nodule-specific genes coding for small, secreted polypeptides with conserved cysteine motifs. *Plant Physiol,* 132**,** 161-73.

MERGAERT, P., UCHIUMI, T., ALUNNI, B., EVANNO, G., CHERON, A., CATRICE, O., MAUSSET, A. E., BARLOY-HUBLER, F., GALIBERT, F., KONDOROSI, A. & KONDOROSI, E. 2006. Eukaryotic control on bacterial cell cycle and differentiation in the Rhizobium-legume symbiosis. *Proc Natl Acad Sci U S A,* 103**,** 5230-5.

MICHAEL, T. P., BRETON, G., HAZEN, S. P., PRIEST, H., MOCKLER, T. C., KAY, S. A. & CHORY, J. 2008a. A morning-specific phytohormone gene expression program underlying rhythmic plant growth. *PLoS Biol,* 6**,** e225.

MICHAEL, T. P. & MCCLUNG, C. R. 2003. Enhancer trapping reveals widespread circadian clock transcriptional control in Arabidopsis. *Plant Physiol,* 132**,** 629-39.

MICHAEL, T. P., MOCKLER, T. C., BRETON, G., MCENTEE, C., BYER, A., TROUT, J. D., HAZEN, S. P., SHEN, R., PRIEST, H. D., SULLIVAN, C. M., GIVAN, S. A., YANOVSKY, M., HONG, F., KAY, S. A. & CHORY, J. 2008b. Network Discovery Pipeline Elucidates Conserved Time-of-Day–Specific cis-Regulatory Modules. *PLOS Genetics,* 4**,** e14.

MIERZIAK, J., KOSTYN, K. & KULMA, A. 2014. Flavonoids as important molecules of plant interactions with the environment. *Molecules,* 19**,** 16240-65.

MIN, H., GUO, H. & XIONG, J. 2005. Rhythmic gene expression in a purple photosynthetic bacterium, *Rhodobacter sphaeroides*. *FEBS Letters,* 579**,** 808-812.

MINCHIN, F. R. & PATE, J. S. 1974. Diurnal Functioning of the Legume Root Nodule. *Journal of Experimental Botany,* 25**,** 295-308.

MISHRA, P. & PANIGRAHI, K. C. 2015. GIGANTEA – an emerging story. *Frontiers in Plant Science,* 6.

MITSUI, H., SATO, T., SATO, Y., ITO, N. & MINAMISAWA, K. 2004. *Sinorhizobium meliloti* RpoH1 is required for effective nitrogen-fixing symbiosis with alfalfa. *Mol Genet Genomics,* 271**,** 416-25.

MIZOGUCHI, T., WHEATLEY, K., HANZAWA, Y., WRIGHT, L., MIZOGUCHI, M., SONG, H.R., CARRÉ, I.A. & COUPLAND, G., 2002. *LHY* and *CCA1* are partially redundant genes required to maintain circadian rhythms in *Arabidopsis*. *Developmental cell*, 2(5), 629-641.

MÖGLICH, A., AYERS, R.A. and MOFFAT, K., 2009. Structure and signaling mechanism of Per-ARNT-Sim domains. *Structure*, 17(10), pp.1282-1294.

MOLING, S., PIETRASZEWSKA-BOGIEL, A., POSTMA, M., FEDOROVA, E., HINK, M. A., LIMPENS, E., GADELLA, T. W. & BISSELING, T. 2014. Nod factor receptors form heteromeric complexes and are essential for intracellular infection in medicago nodules. *Plant Cell,* 26**,** 4188-99.

MONTIEL, J., DOWNIE, J. A., FARKAS, A., BIHARI, P., HERCZEG, R., BALINT, B., MERGAERT, P., KERESZT, A. & KONDOROSI, E. 2017. Morphotype of bacteroids in different legumes correlates with the number and type of symbiotic NCR peptides. *Proc Natl Acad Sci U S A,* 114**,** 5041-5046.

MONTIEL, J., SZŰCS, A., BOBOESCU, I. Z., GHERMAN, V. D., KONDOROSI, É. & KERESZT, A. 2016. Terminal Bacteroid Differentiation Is Associated With Variable Morphological Changes in Legume Species Belonging to the Inverted Repeat-Lacking Clade. *Molecular Plant-Microbe Interactions®,* 29**,** 210-219.

MU, X. & LUO, J. 2019. Evolutionary analyses of NIN-like proteins in plants and their roles in nitrate signaling. *Cell Mol Life Sci,* 76**,** 3753-3764.

NAGEL, D. H. & KAY, S. A. 2012. Complexity in the wiring and regulation of plant circadian networks. *Curr Biol,* 22**,** R648-57.

NAGYMIHALY, M., VELUCHAMY, A., GYORGYPAL, Z., ARIEL, F., JEGU, T., BENHAMED, M., SZUCS, A., KERESZT, A., MERGAERT, P. & KONDOROSI, E. 2017. Ploidy-dependent changes in the epigenome of symbiotic cells correlate with specific patterns of gene expression. *Proc Natl Acad Sci U S A,* 114**,** 4543-4548.

NAKAHIRA, Y., KATAYAMA, M., MIYASHITA, H., KUTSUNA, S., IWASAKI, H., OYAMA, T. & KONDO, T. 2004. Global gene repression by KaiC as a master

process of prokaryotic circadian system. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 881-885.
NAKAJIMA, M., IMAI, K., ITO, H., NISHIWAKI, T., MURAYAMA, Y., IWASAKI, H., OYAMA, T. & KONDO, T. 2005. Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation in vitro. *Science,* 308**,** 414-5.
NAKAMICHI, N., KIBA, T., HENRIQUES, R., MIZUNO, T., CHUA, N. H. & SAKAKIBARA, H. 2010. PSEUDO-RESPONSE REGULATORS 9, 7, and 5 are transcriptional repressors in the Arabidopsis circadian clock. *Plant Cell,* 22**,** 594-605.
NALLU, S., SILVERSTEIN, K. A., SAMAC, D. A., BUCCIARELLI, B., VANCE, C. P. & VANDENBOSCH, K. A. 2013. Regulatory patterns of a large family of defensin-like genes expressed in nodules of *Medicago truncatula. PLoS One,* 8**,** e60355.
NALLU, S., SILVERSTEIN, K. A., ZHOU, P., YOUNG, N. D. & VANDENBOSCH, K. A. 2014. Patterns of divergence of a large family of nodule cysteine-rich peptides in accessions of *Medicago truncatula. Plant J,* 78**,** 697-705.
NI, J., DONG, L., JIANG, Z., YANG, X., CHEN, Z., WU, Y. & XU, M. 2018. Comprehensive transcriptome analysis and flavonoid profiling of Ginkgo leaves reveals flavonoid content alterations in day-night cycles. *PLoS One,* 13**,** e0193897.
NUSINOW, D. A., HELFER, A., HAMILTON, E. E., KING, J. J., IMAIZUMI, T., SCHULTZ, T. F., FARRÉ, E. M. & KAY, S. A. 2011. The ELF4–ELF3–LUX complex links the circadian clock to diurnal control of hypocotyl growth. *Nature,* 475**,** 398-402.
OLDROYD, G. E. 2013. Speak, friend, and enter: signaling systems that promote beneficial symbiotic associations in plants. *Nat Rev Microbiol,* 11**,** 252-63.
OONO, R. & DENISON, R. F. 2010. Comparing symbiotic efficiency between swollen versus nonswollen rhizobial bacteroids. *Plant Physiol,* 154**,** 1541-8.
OTTO, M. 2009. Bacterial sensing of antimicrobial peptides. *Contrib Microbiol,* 16**,** 136-149.
PAN, H. & WANG, D. 2017. Nodule cysteine-rich peptides maintain a working balance during nitrogen-fixing symbiosis. *Nat Plants,* 3**,** 17048.
PARK, M.-J., KWON, Y.-J., GIL, K.-E. & PARK, C.-M. 2016. LATE ELONGATED HYPOCOTYL regulates photoperiodic flowering via the circadian clock in Arabidopsis. *BMC Plant Biology,* 16**,** 114.
PATRO, R., DUGGAL, G., LOVE, M. I., IRIZARRY, R. A. & KINGSFORD, C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods,* 14**,** 417-419.
PAULOSE, J. K., WRIGHT, J. M., PATEL, A. G. & CASSONE, V. M. 2016. Human Gut Bacteria Are Sensitive to Melatonin and Express Endogenous Circadian Rhythmicity. *PLOS ONE,* 11**,** e0146643.
PECRIX, Y., STATON, S. E., SALLET, E., et al. 2018. Whole-genome landscape of *Medicago truncatula* symbiotic genes. *Nature Plants,* 4**,** 1017-1025.
PEL, M. J. C. & PIETERSE, C. M. J. 2013. Microbial recognition and evasion of host immunity. *Journal of Experimental Botany,* 64**,** 1237-1248.
PENFIELD, S. & HALL, A. 2009. A Role for Multiple Circadian Clock Genes in the Response to Signals That Break Seed Dormancy in *Arabidopsis. The Plant Cell,* 21**,** 1722-1732.
PENTERMAN, J., ABO, R. P., DE NISCO, N. J., ARNOLD, M. F., LONGHI, R., ZANDA, M. & WALKER, G. C. 2014. Host plant peptides elicit a transcriptional response to control the *Sinorhizobium meliloti* cell cycle during symbiosis. *Proc Natl Acad Sci U S A,* 111**,** 3561-6.

PINI, F., DE NISCO, N. J., FERRI, L., PENTERMAN, J., FIORAVANTI, A., BRILLI, M., MENGONI, A., BAZZICALUPO, M., VIOLLIER, P. H., WALKER, G. C. & BIONDI, E. G. 2015. Cell Cycle Control by the Master Regulator CtrA in *Sinorhizobium meliloti*. *PLoS Genet,* 11**,** e1005232.

PINI, F., FRAGE, B., FERRI, L., DE NISCO, N. J., MOHAPATRA, S. S., TADDEI, L., FIORAVANTI, A., DEWITTE, F., GALARDINI, M., BRILLI, M., VILLERET, V., BAZZICALUPO, M., MENGONI, A., WALKER, G. C., BECKER, A. & BIONDI, E. G. 2013. The DivJ, CbrA and PleC system controls DivK phosphorylation and symbiosis in *Sinorhizobium meliloti*. *Mol Microbiol,* 90**,** 54-71.

POOLE, P., RAMACHANDRAN, V. & TERPOLILLI, J. 2018. Rhizobia: from saprophytes to endosymbionts. *Nature Reviews Microbiology,* 16**,** 291.

PORTILLO, M., LINDSEY, K., CASSON, S., GARCÍA-CASADO, G., SOLANO, R., FENOLL, C. & ESCOBAR, C. 2009. Isolation of RNA from laser-capture-microdissected giant cells at early differentiation stages suitable for differential transcriptome analysis. *Mol Plant Pathol,* 10**,** 523-35.

PRELL, J. & POOLE, P. 2006. Metabolic changes of rhizobia in legume nodules. *Trends Microbiol,* 14**,** 161-8.

PRICE, P. A., TANNER, H. R., DILLON, B. A., SHABAB, M., WALKER, G. C. & GRIFFITTS, J. S. 2015. Rhizobial peptidase HrrP cleaves host-encoded signaling peptides and mediates symbiotic compatibility. *Proc Natl Acad Sci U S A,* 112**,** 15244-9.

R-CORE-TEAM 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

RANOCHA, P., DENANCÉ, N., VANHOLME, R., FREYDIER, A., MARTINEZ, Y., HOFFMANN, L., KÖHLER, L., POUZET, C., RENOU, J. P., SUNDBERG, B., BOERJAN, W. & GOFFNER, D. 2010. Walls are thin 1 (WAT1), an Arabidopsis homolog of *Medicago truncatula* NODULIN21, is a tonoplast-localized protein required for secondary wall formation in fibers. *Plant J,* 63**,** 469-83.

REY, T., NARS, A., BONHOMME, M., BOTTIN, A., HUGUET, S., BALZERGUE, S., JARDINAUD, M.-F., BONO, J.-J., CULLIMORE, J., DUMAS, B., GOUGH, C. & JACQUET, C. 2013. NFP, a LysM protein controlling Nod factor perception, also intervenes in *Medicago truncatula* resistance to pathogens. *New Phytologist,* 198**,** 875-886.

REYES, J. C., MURO-PASTOR, M. I. & FLORENCIO, F. J. 2004. The GATA family of transcription factors in Arabidopsis and rice. *Plant physiology,* 134**,** 1718-1732.

RIBEIRO, C. W., BALDACCI-CRESP, F., PIERRE, O., LAROUSSE, M., BENYAMINA, S., LAMBERT, A., HOPKINS, J., CASTELLA, C., CAZARETH, J., ALLOING, G., BONCOMPAGNI, E., COUTURIER, J., MERGAERT, P., GAMAS, P., ROUHIER, N., MONTRICHARD, F. & FRENDO, P. 2017. Regulation of Differentiation of Nitrogen-Fixing Bacteria by Microsymbiont Targeting of Plant Thioredoxin s1. *Curr Biol,* 27**,** 250-256.

ROBERT-SEILANIANTZ, A., GRANT, M. & JONES, J. D. G. 2011. Hormone Crosstalk in Plant Disease and Defense: More Than Just JASMONATE-SALICYLATE Antagonism. *Annual Review of Phytopathology,* 49**,** 317-343.

RODEN, L. C. & INGLE, R. A. 2009. Lights, rhythms, infection: the role of light and the circadian clock in determining the outcome of plant-pathogen interactions. *Plant Cell,* 21**,** 2546-52.

RON, M., KAJALA, K., PAULUZZI, G., WANG, D., REYNOSO, M. A., ZUMSTEIN, K., GARCHA, J., WINTE, S., MASSON, H., INAGAKI, S.,

FEDERICI, F., SINHA, N., DEAL, R. B., BAILEY-SERRES, J. & BRADY, S. M. 2014. Hairy root transformation using *Agrobacterium rhizogenes* as a tool for exploring cell type-specific gene expression and function using tomato as a model. *Plant Physiol,* 166**,** 455-69.

ROUX, B., RODDE, N., JARDINAUD, M. F., TIMMERS, T., SAUVIAC, L., COTTRET, L., CARRERE, S., SALLET, E., COURCELLE, E., MOREAU, S., DEBELLE, F., CAPELA, D., DE CARVALHO-NIEBEL, F., GOUZY, J., BRUAND, C. & GAMAS, P. 2014. An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant J,* 77**,** 817-37.

ROY, P., ACHOM, M., WILKINSON, H., LAGUNAS, B. & GIFFORD, M. L. 2020. Symbiotic Outcome Modified by the Diversification from 7 to over 700 Nodule-Specific Cysteine-Rich Peptides. *Genes,* 11**,** 348.

ROY, S., LIU, W., NANDETY, R. S., CROOK, A., MYSORE, K. S., PISLARIU, C. I., FRUGOLI, J., DICKSTEIN, R. & UDVARDI, M. K. 2020. Celebrating 20 Years of Genetic Discoveries in Legume Nodulation and Symbiotic Nitrogen Fixation. *The Plant Cell,* 32**,** 15-41.

RUGNONE, M. L., FAIGON SOVERNA, A., SANCHEZ, S. E., SCHLAEN, R. G., HERNANDO, C. E., SEYMOUR, D. K., MANCINI, E., CHERNOMORETZ, A., WEIGEL, D., MAS, P. & YANOVSKY, M. J. 2013. LNK genes integrate light and clock signaling networks at the core of the Arabidopsis oscillator. *Proc Natl Acad Sci U S A,* 110**,** 12120-5.

SANCHEZ, S. E. & KAY, S. A. 2016. The Plant Circadian Clock: From a Simple Timekeeper to a Complex Developmental Manager. *Cold Spring Harb Perspect Biol,* 8.

SARTOR, F., EELDERINK-CHEN, Z., ARONSON, B., BOSMAN, J., HIBBERT, L. E., DODD, A. N., KOVÁCS Á, T. & MERROW, M. 2019. Are There Circadian Clocks in Non-Photosynthetic Bacteria? *Biology (Basel),* 8.

SATGE, C., MOREAU, S., SALLET, E., LEFORT, G., AURIAC, M. C., REMBLIERE, C., COTTRET, L., GALLARDO, K., NOIROT, C., JARDINAUD, M. F. & GAMAS, P. 2016. Reprogramming of DNA methylation is critical for nodule development in *Medicago truncatula*. *Nat Plants,* 2**,** 16166.

SCHIESSL, K., LILLEY, J. L. S., LEE, T., TAMVAKIS, I., KOHLEN, W., BAILEY, P. C., THOMAS, A., LUPTAK, J., RAMAKRISHNAN, K., CARPENTER, M. D., MYSORE, K. S., WEN, J., AHNERT, S., GRIENEISEN, V. A. & OLDROYD, G. E. D. 2019. NODULE INCEPTION Recruits the Lateral Root Developmental Program for Symbiotic Nodule Organogenesis in *Medicago truncatula*. *Current Biology,* 29**,** 3657-3668.e5.

SCHMITTGEN, T., LIVAK, K. 2008. Analyzing real-time PCR data by the comparative CT method. Nat Protoc 3, 1101–1108.

SEVIN-PUJOL, A., SICARD, M., ROSENBERG, C., AURIAC, M.-C., LEPAGE, A., NIEBEL, A., GOUGH, C. & BENSMIHEN, S. 2017. Development of a GAL4-VP16/UAS trans-activation system for tissue specific expression in *Medicago truncatula*. *PLOS ONE,* 12**,** e0188923.

SHIGENOBU, S. & STERN, D. L. 2013. Aphids evolved novel secreted proteins for symbiosis with bacterial endosymbiont. *Proc Biol Sci,* 280**,** 20121952.

SIMSEK, S., OJANEN-REUHS, T., STEPHENS, S. B. & REUHS, B. L. 2007. Strain-ecotype specificity in *Sinorhizobium meliloti-Medicago truncatula* symbiosis is correlated to succinoglycan oligosaccharide structure. *J Bacteriol,* 189**,** 7733-40.

SIMSEK, S., WOOD, K. & REUHS, B. L. 2013. Structural analysis of succinoglycan oligosaccharides from *Sinorhizobium meliloti* strains with different host compatibility phenotypes. *J Bacteriol,* 195**,** 2032-8.

SINGER, A., POSCHMANN, G., MÜHLICH, C., VALADEZ-CANO, C., HÄNSCH, S., HÜREN, V., RENSING, S. A., STÜHLER, K. & NOWACK, E. C. 2017. Massive protein import into the early-evolutionary-stage photosynthetic organelle of the amoeba *Paulinella chromatophora. Current Biology,* 27**,** 2763-2773. e5.

SINGH, M. & MAS, P. 2018. A Functional Connection between the Circadian Clock and Hormonal Timing in Arabidopsis. *Genes (Basel),* 9.

SMIT, P., LIMPENS, E., GEURTS, R., FEDOROVA, E., DOLGIKH, E., GOUGH, C. & BISSELING, T. 2007. Medicago LYK3, an Entry Receptor in Rhizobial Nodulation Factor Signaling. *Plant Physiology,* 145**,** 183-191.

SMITH, T. F. & WATERMAN, M. S. 1981. Identification of common molecular subsequences. *J Mol Biol,* 147**,** 195-7.

SONG, Y. H., ITO, S. & IMAIZUMI, T. 2010. Similarities in the circadian clock and photoperiodism in plants. *Curr Opin Plant Biol,* 13**,** 594-603.

SOYANO, T., SHIMODA, Y., KAWAGUCHI, M. & HAYASHI, M. 2019. A shared gene drives lateral root development and root nodule symbiosis pathways in Lotus. *Science,* 366**,** 1021-1023.

SPRENT, J. I. 2007. Evolving ideas of legume evolution and diversity: a taxonomic perspective on the occurrence of nodulation. *New Phytol,* 174**,** 11-25.

STEVE ROZEN, H. J. S. 2000. Primer3 on the WWW for general users and for biologist programmers. *In:* KRAWETZ S, M. S. (ed.) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Totowa, NJ: Humana Press.

STOUGAARD, J., ABILDSTEN, D. & MARCKER, K. A. 1987. The *Agrobacterium rhizogenes* pRi TL-DNA segment as a gene vector system for transformation of plants. *Molecular and General Genetics MGG,* 207**,** 251-255.

STURTEVANT, R. Circadian patterns in linear growth of *Escherichia Coli.* ANATOMICAL RECORD, 1973a. WILEY-LISS DIV JOHN WILEY & SONS INC, 111 RIVER ST, HOBOKEN, NJ 07030 USA, 453-+.

STURTEVANT, R. P. 1973b. Circadian variability in Klebsiella demonstrated by cosinor analysis. *Int J Chronobiol,* 1**,** 141-6.

SYSKA, C., BROUQUISSE, R., ALLOING, G., PAULY, N., FRENDO, P., BOSSENO, M., DUPONT, L. & BOSCARI, A. 2019. Molecular Weapons Contribute to Intracellular Rhizobia Accommodation Within Legume Host Cell. *Frontiers in Plant Science,* 10.

TADEGE, M., WEN, J., HE, J., TU, H., KWAK, Y., ESCHSTRUTH, A., CAYREL, A., ENDRE, G., ZHAO, P. X. & CHABAUD, M. 2008a. Large-scale insertional mutagenesis using the Tnt1 retrotransposon in the model legume Medicago truncatula. *The Plant Journal,* 54**,** 335-347.

TADEGE, M., WEN, J., HE, J., TU, H., KWAK, Y., ESCHSTRUTH, A., CAYREL, A., ENDRE, G., ZHAO, P. X., CHABAUD, M., RATET, P. & MYSORE, K. S. 2008b. Large-scale insertional mutagenesis using the Tnt1 retrotransposon in the model legume Medicago truncatula. *Plant J,* 54**,** 335-47.

TAKAHASHI, N., HIRATA, Y., AIHARA, K. & MAS, P. 2015. A Hierarchical Multi-oscillator Network Orchestrates the Arabidopsis Circadian System. *Cell,* 163**,** 148-159.

TAKATA, N., SAITO, S., SAITO, C.T. & UEMURA, M., 2010. Phylogenetic footprint of the plant clock system in angiosperms: evolutionary processes of Pseudo-Response Regulators. *BMC evolutionary biology*, 10(1), 1-14.

TAKATA, N., SAITO, S., TANAKA SAITO, C., NANJO, T., SHINOHARA, K. & UEMURA, M., 2009. Molecular phylogeny and expression of poplar circadian clock genes, LHY1 and LHY2. *New Phytologist*, 181(4), 808-819.

TAVORMINA, P., DE CONINCK, B., NIKONOROVA, N., DE SMET, I. & CAMMUE, B. P. 2015. The Plant Peptidome: An Expanding Repertoire of Structural Features and Biological Functions. *Plant Cell,* 27**,** 2095-118.

TAYLOR, J. A., OUIMET, M. C., WARGACHUK, R. & MARCZYNSKI, G. T. 2011. The Caulobacter crescentus chromosome replication origin evolved two classes of weak DnaA binding sites. *Mol Microbiol,* 82**,** 312-26.

TERPOLILLI, J., HILL, Y., TIAN, R., HOWIESON, J., BRÄU, L., GOODWIN, L., HAN, J., LIOLIOS, K., HUNTEMANN, M., PATI, A., WOYKE, T., MAVROMATIS, K., MARKOWITZ, V., IVANOVA, N., KYRPIDES, N. & REEVE, W. 2013. Genome sequence of Ensifer meliloti strain SM1022; a highly effective microsymbiont of the model legume *Medicago truncatula* A17. *Stand Genomic Sci,* 9**,** 315-24.

TERPOLILLI, J. J., O'HARA, G. W., TIWARI, R. P., DILWORTH, M. J. & HOWIESON, J. G. 2008. The model legume *Medicago truncatula* A17 is poorly matched for N2 fixation with the sequenced microsymbiont Sinorhizobium meliloti 1021. *New Phytologist,* 179**,** 62-66.

TIMMERS, A. C., AURIAC, M. C. & TRUCHET, G. 1999. Refined analysis of early symbiotic steps of the Rhizobium-Medicago interaction in relationship with microtubular cytoskeleton rearrangements. *Development,* 126**,** 3617-28.

TIRICZ, H., SZUCS, A., FARKAS, A., PAP, B., LIMA, R. M., MAROTI, G., KONDOROSI, E. & KERESZT, A. 2013. Antimicrobial nodule-specific cysteine-rich peptides induce membrane depolarization-associated changes in the transcriptome of *Sinorhizobium meliloti*. *Appl Environ Microbiol,* 79**,** 6737-46.

TOMITA, J., NAKAJIMA, M., KONDO, T. & IWASAKI, H. 2005. No transcription-translation feedback in circadian rhythm of KaiC phosphorylation. *Science,* 307**,** 251-4.

UCHI, N., FUKUDOME, M., NOZAKI, N., SUZUKI, M., OSUKI, K. I., SHIGENOBU, S. & UCHIUMI, T. 2019. Antimicrobial Activities of Cysteine-rich Peptides Specific to Bacteriocytes of the Pea Aphid *Acyrthosiphon pisum*. *Microbes Environ,* 34**,** 155-160.

UDVARDI, M. & POOLE, P. S. 2013. Transport and metabolism in legume-rhizobia symbioses. *Annu Rev Plant Biol,* 64**,** 781-805.

UEOKA-NAKANISHI, H., YAMASHINO, T., ISHIDA, K., KAMIOKA, M., NAKAMICHI, N. & MIZUNO, T. 2012. Molecular mechanisms of circadian rhythm in *Lotus japonicus* and *Arabidopsis thaliana* are sufficiently compatible to regulate heterologous core clock genes robustly. *Biosci Biotechnol Biochem,* 76**,** 2332-4.

VAN DE VELDE, W., ZEHIROV, G., SZATMARI, A., DEBRECZENY, M., ISHIHARA, H., KEVEI, Z., FARKAS, A., MIKULASS, K., NAGY, A., TIRICZ, H., SATIAT-JEUNEMAITRE, B., ALUNNI, B., BOURGE, M., KUCHO, K., ABE, M., KERESZT, A., MAROTI, G., UCHIUMI, T., KONDOROSI, E. & MERGAERT, P. 2010. Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science,* 327**,** 1122-6.

VAN VELZEN, R., HOLMER, R., BU, F.,et al. 2018. Comparative genomics of the nonlegume *Parasponia* reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proceedings of the National Academy of Sciences,* 115**,** E4700-E4709.

VANCE, C. P. 2001. Symbiotic nitrogen fixation and phosphorus acquisition. Plant nutrition in a world of declining renewable resources. *Plant Physiol,* 127**,** 390-7.
VERNIÉ, T., KIM, J., FRANCES, L., DING, Y., SUN, J., GUAN, D., NIEBEL, A., GIFFORD, M. L., DE CARVALHO-NIEBEL, F. & OLDROYD, G. E. D. 2015. The NIN Transcription Factor Coordinates Diverse Nodulation Programs in Different Tissues of the *Medicago truncatula* Root. *The Plant Cell,* 27**,** 3410-3424.
VICENTE, E. J. & DEAN, D. R. 2017. Keeping the nitrogen-fixation dream alive. *Proceedings of the National Academy of Sciences,* 114**,** 3009-3011.
VOGT, J.H. and SCHIPPERS, J.H. 2015. Setting the PAS, the role of circadian PAS domain proteins during environmental adaptation in plants. *Frontiers in plant science*, 6, p.513.
WALKER, L. 2019. Analyzing the molecular basis of plant root responses to the environment. *PhD thesis*, University of Warwick.
WALKER, L., BODDINGTON, C., JENKINS, D., WANG, Y., GRØNLUND, J. T., HULSMANS, J., KUMAR, S., PATEL, D., MOORE, J. D., CARTER, A., SAMAVEDAM, S., BONOMO, G., HERSH, D. S., CORUZZI, G. M., BURROUGHS, N. J. & GIFFORD, M. L. 2017. Changes in Gene Expression in Space and Time Orchestrate Environmentally Mediated Shaping of Root Architecture. *Plant Cell,* 29**,** 2393-2412.
WALKER, L., LAGUNAS, B. & GIFFORD, M. L. 2020. Determinants of Host Range Specificity in Legume-Rhizobia Symbiosis. *Frontiers in Microbiology,* 11.
WANG, K., BU, T., CHENG, Q., DONG, L., SU, T., CHEN, Z., KONG, F., GONG, Z., LIU, B. & LI, M. 2021. Two homologous LHY pairs negatively control soybean drought tolerance by repressing the abscisic acid responses. *New Phytologist,* n/a.
WANG, Q., LIU, J., LI, H., YANG, S., KORMOCZI, P., KERESZT, A. & ZHU, H. 2018. Nodule-Specific Cysteine-Rich Peptides Negatively Regulate Nitrogen-Fixing Symbiosis in a Strain-Specific Manner in *Medicago truncatula*. *Mol Plant Microbe Interact,* 31**,** 240-248.
WANG, Q., YANG, S., LIU, J., TERECSKEI, K., ÁBRAHÁM, E., GOMBÁR, A., DOMONKOS, Á., SZŰCS, A., KÖRMÖCZI, P., WANG, T., FODOR, L., MAO, L., FEI, Z., KONDOROSI, É., KALÓ, P., KERESZT, A. & ZHU, H. 2017. Host-secreted antimicrobial peptide enforces symbiotic selectivity in Medicago truncatula. *Proc Natl Acad Sci U S A,* 114**,** 6854-6859.
WANG, W., BARNABY, J. Y., TADA, Y., LI, H., TÖR, M., CALDELARI, D., LEE, D.-U., FU, X.-D. & DONG, X. 2011. Timing of plant immune responses by a central circadian regulator. *Nature,* 470**,** 110-114.
WANG, Z. Y. & TOBIN, E. M. 1998. Constitutive expression of the CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) gene disrupts circadian rhythms and suppresses its own expression. *Cell,* 93**,** 1207-17.
WEBB, A. A. R., SEKI, M., SATAKE, A. & CALDANA, C. 2019. Continuous dynamic adjustment of the plant circadian oscillator. *Nature Communications,* 10**,** 550.
XIAO, T. T., SCHILDERINK, S., MOLING, S., DEINUM, E. E., KONDOROSI, E., FRANSSEN, H., KULIKOVA, O., NIEBEL, A. & BISSELING, T. 2014. Fate map of *Medicago truncatula* root nodules. *Development,* 141**,** 3517-3528.
XIE, F., MURRAY, J. D., KIM, J., HECKMANN, A. B., EDWARDS, A., OLDROYD, G. E. D. & DOWNIE, J. A. 2012. Legume pectate lyase required for root infection by rhizobia. *Proceedings of the National Academy of Sciences,* 109**,** 633-638.
XU, G., JIANG, Z., WANG, H. & LIN, R. 2019. The central circadian clock proteins CCA1 and LHY regulate iron homeostasis in Arabidopsis. *Journal of Integrative Plant Biology,* 61**,** 168-181.

XU, Y., MORI, T. & JOHNSON, C. H. 2003. Cyanobacterial circadian clockwork: roles of KaiA, KaiB and the kaiBC promoter in regulating KaiC. *Embo j,* 22**,** 2117-26.

YANG, R. & SU, Z. 2010. Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation. *Bioinformatics,* 26**,** i168-74.

YANG, S., WANG, Q., FEDOROVA, E., LIU, J., QIN, Q., ZHENG, Q., PRICE, P. A., PAN, H., WANG, D., GRIFFITTS, J. S., BISSELING, T. & ZHU, H. 2017. Microsymbiont discrimination mediated by a host-secreted peptide in *Medicago truncatula*. *Proc Natl Acad Sci U S A,* 114**,** 6848-6853.

YOUNG, N. D., DEBELLE, F., OLDROYD, G. E., et al. 2011. The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature,* 480**,** 520-4.

ZAMIOUDIS, C. & PIETERSE, C. M. 2012. Modulation of host immunity by beneficial microbes. *Mol Plant Microbe Interact,* 25**,** 139-50.

ZHANG, C., XIE, Q., ANDERSON, R. G., NG, G., SEITZ, N. C., PETERSON, T., MCCLUNG, C. R., MCDOWELL, J. M., KONG, D., KWAK, J. M. & LU, H. 2013a. Crosstalk between the Circadian Clock and Innate Immunity in Arabidopsis. *PLOS Pathogens,* 9**,** e1003370.

ZHANG, Q., DU, H., LV, D., XIAO, T., PAN, J., HE, H., WANG, G., CAI, R., WENG, Y. & PAN, J. 2018. Efficient Transposition of the Retrotransposon Tnt1 in Cucumber (Cucumis sativus L.). *Horticultural Plant Journal,* 4**,** 111-116.

ZHANG, W., DOU, Z., HE, P., JU, X. T., POWLSON, D., CHADWICK, D., NORSE, D., LU, Y. L., ZHANG, Y., WU, L., CHEN, X. P., CASSMAN, K. G. & ZHANG, F. S. 2013b. New technologies reduce greenhouse gas emissions from nitrogenous fertilizer in China. *Proc Natl Acad Sci U S A,* 110**,** 8375-80.

ZHOU, C., HAN, L., FU, C., CHAI, M., ZHANG, W., LI, G., TANG, Y. & WANG, Z. Y. 2012. Identification and characterization of petiolule-like pulvinus mutants with abolished nyctinastic leaf movement in the model legume *Medicago truncatula*. *New Phytol,* 196**,** 92-100.