



**HAL**  
open science

# Ploidy-dependent changes in the epigenome of symbiotic cells correlate with specific patterns of gene expression

Marianna Nagymihály

## ► To cite this version:

Marianna Nagymihály. Ploidy-dependent changes in the epigenome of symbiotic cells correlate with specific patterns of gene expression. *Vegetal Biology*. Université Paris Saclay (COMUE); University of Szeged (Hongrie), 2017. English. NNT : 2017SACLS399 . tel-01707868

**HAL Id: tel-01707868**

**<https://theses.hal.science/tel-01707868>**

Submitted on 13 Feb 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NNT : 2017SACLS399



Comprendre le monde,  
construire l'avenir



THESE DE DOCTORAT  
DE L'UNIVERSITE PARIS-SACLAY,  
préparée à l'Université Paris-Sud  
ET  
DE L'UNIVERSITE SZEGED

ÉCOLE DOCTORALE N° 567  
Sciences du Végétal : du Gène à l'Ecosystème

Spécialité de doctorat (Biologie)

Par

**Marianna NAGYMIHÁLY**

Ploidy-dependent changes in the epigenome of symbiotic cells  
correlate with specific patterns of gene expression

**Thèse présentée et soutenue à Szeged, le 15/11/2017:**

**Composition du Jury :**

M, VENETIANER, Pál, Professeur émérite, BRC, Szeged, Rapporteur  
M, ROUDIER, François, Directeur de recherche, ENS, Lyon, Rapporteur  
M, FEHÉR, Attila, Directeur de recherche, BRC, Szeged, Président  
M, CHEVALIER, Christian, Directeur de recherche, INRA, Bordeaux-Aquitaine, Examineur  
Mme, CSISZÁR, Jolán, Professeur associé, University Szeged, Examinatrice  
M, RATET, Pascal, Directeur de recherche, IPS2, Gif-sur Yvette, Examineur  
M, MERGAERT, Peter, Directeur de recherche, I2BC, Gif-sur-Yvette, Directeur de thèse  
M, KERESZT, Attila, Directeur de recherche, BRC, Szeged, Co-directeur de thèse



**Titre :** Des changements ploïdie-dépendant dans l'épigénome de cellules symbiotiques sont corrélés avec des profils spécifiques d'expression génique

**Mots clés :** Symbiose rhizobium-legumineuse, peptides NCR, polyploïdie, endoréplication, *Medicago truncatula* Gene Expression Atlas (MtGEA), épigénétiques

**Résumé** Les légumineuses peuvent interagir avec les bactéries du sol de la famille des *Rhizobiaceae*. Cette interaction aboutit à la formation d'un organe spécialisé appelé nodosité. Au sein des cellules symbiotiques des nodosités, les rhizobia sont capables de fixer l'azote atmosphérique et de la convertir en ammoniac, qui est une source d'azote assimilable par les plantes. Chez la Légumineuse *Medicago truncatula*, les cellules symbiotiques produisent une large famille de peptides riches en cystéines appelées (NCRs) spécifiquement exprimés dans les nodosités. Ces NCRs induisent la différenciation des bactéroïdes qui se traduit par un allongement cellulaire couplé à une forte endoréplication du génome (les bactéroïdes deviennent polyploïdes) contribuant ainsi à une augmentation importante de la taille des cellules, ainsi qu'une perméabilité membranaire accrue et une perte de toute capacité reproductrice. Les peptides NCRs ressemblent à des défensines, des peptides antimicrobiens, acteurs clés de l'immunité innée. L'analyse de l'expression de 334 gènes *NCR* dans 267 différentes conditions expérimentales en utilisant la base de données MtGEA (*Medicago truncatula* Gene Expression Atlas) a révélé que l'ensemble des gènes *NCR* testés (sauf quatre) n'est exprimé que dans les nodosités, ils ne sont pas exprimés dans d'autres organes de la plante, ni lors d'une infection par des agents pathogènes. De plus l'expression des *NCRs* n'est induite en réponse à aucune interaction biotique ou abiotique testée ou à des facteurs Nod. Les gènes *NCR* sont activés en vagues successives au cours de l'organogenèse nodulaire et ce profil temporel est en corrélation avec une localisation spatiale spécifique de leurs transcrit de la zone apicale à la partie proximale de nodosités. En outre, nous avons montré que les *NCRs* ne sont pas induites pendant la sénescence des nodules. Ces analyses expérimentales ensemble avec des calculs

d'entropie de Shannon, une métrique pour la spécificité d'expression, montrent que les gènes *NCR* sont parmi les gènes les plus fortement et le plus spécifiquement exprimés chez *M. truncatula*. Ainsi, l'expression des *NCRs* est soumise à une régulation extrêmement stricte et ils sont activés exclusivement pendant l'organogenèse et au cours du développement nodulaire dans les cellules symbiotiques polyploïdes. Cette analyse a suggéré l'implication de la régulation épigénétique des gènes *NCR*. La formation des cellules symbiotiques s'exerce par une endoreplication et est associée à une reprogrammation transcriptionnelle. En utilisant le tri par cytométrie en flux des noyaux, en fonction de leur contenu en ADN, nous avons montré que les vagues transcriptionnelles sont en corrélation avec les niveaux croissants de ploïdie et résultent des modifications épigénétiques durant les cycles d'endoréplication. Nous avons étudié la méthylation de l'ADN génomique et l'accessibilité à la chromatine, ainsi que la présence des marqueurs répresseurs (H3K27me3) ou activateurs transcriptionnels (H3K9ac) sur des gènes spécifiques des nodosités. La méthylation différentielle de l'ADN n'a été trouvée que dans un petit sous-ensemble de gènes symbiotiques spécifiques aux nodosités. Néanmoins, plus que la moitié des gènes *NCR* était différentiellement méthyles. D'autre part, l'expression des gènes était corrélée avec la décondensation de la chromatine (ouverture), un enrichissement du marqueur H3K9ac et une diminution du marqueur H3K27me3. Nos résultats suggèrent que l'endoréplication, pendant la différenciation cellulaire dans les nodosités, fasse partie des mécanismes qui lèvent l'inactivation transcriptionnelle des gènes spécifiques des nodosités, ceci résultant de modifications des codes épigénétiques au niveau de la chromatine.

**Title :** Ploidy-dependent changes in the epigenome of symbiotic cells correlate with specific patterns of gene expression

**Keywords :** rhizobia-legume symbiosis, Nodule-Specific Cysteine-Rich peptides, polyploidy, endoreduplication, *Medicago truncatula* Gene Expression Atlas, epigenetics

**Abstract:** Legume plants are able to interact with soil bacteria from the *Rhizobiaceae* family. This interaction leads to the development of a specialized organ called root nodule. Inside the symbiotic nodule cells, rhizobia are capable to fix atmospheric nitrogen and convert it to ammonia, which is a usable nitrogen source for the plant. In the legume *Medicago truncatula* the symbiotic cells produce high amounts of Nodule-Specific Cysteine-Rich (NCR) peptides which induce differentiation of the rhizobia into enlarged, polyploid and non-cultivable bacterial cells. NCRs are similar to innate immunity antimicrobial peptides. The *NCR* gene family is extremely large in *Medicago* with about 600 genes. The expression analysis of 334 *NCR* genes in 267 different experimental conditions using the *Medicago truncatula* Gene Expression Atlas (MtGEA) revealed that all the *NCR* genes except five are exclusively expressed in nodules. No *NCR* expression is induced in any other plant organ or in response to biotic, abiotic stress tested or to Nod factors. The *NCR* genes are activated in consecutive waves during nodule organogenesis, which correlated with a specific spatial localization of their transcripts from the apical to the proximal nodule zones. Moreover, we showed that *NCRs* are not induced during nodule senescence. According to their Shannon entropy, a metric for tissue specificity, *NCR* genes are among the most specifically and highest expressed genes in *M. truncatula*.

Thus, *NCR* gene expression is subject to an extreme tight regulation since they are only activated during nodule organogenesis in the polyploid symbiotic cells. This analysis suggested the involvement of epigenetic regulation of the *NCR* genes. The formation of the symbiotic cells is driven by endoreduplication and is associated with transcriptional reprogramming. Using sorted nodule nuclei according to their DNA content, we demonstrated that the transcriptional waves correlate with growing ploidy levels and investigated how the epigenome changes during endoreduplication cycles. We studied genome-wide DNA methylation and chromatin accessibility as well as the presence of repressive H3K27me3 and activating H3K9ac histone tail modifications on selected genes. Differential DNA methylation was found only in a small subset of symbiotic nodule-specific genes, including over half of the *NCR* genes, while in most genes DNA methylation was unaffected by the ploidy levels and was independent of the genes' active or repressed state. On the other hand, expression of these genes correlated with ploidy-dependent opening of the chromatin and in a subset of tested genes with reduced H3K27me3 levels combined with enhanced H3K9ac levels. Our results suggest that endoreduplication-dependent epigenetic changes contribute to transcriptional reprogramming in differentiation of symbiotic cells.

## **ACKNOWLEDGEMENTS**

I would like to thank Dr. Éva Kondorosi for accepting me into her lab in Szeged and providing me a productive environment to conduct my PhD research while showing great encouragement, enthusiasm, patience and support during my time as her student. I am very grateful to Dr. Peter Mergaert for mentoring me through my time in the laboratory in Gif-sur-Yvette and providing me helpful discussion about my project. I am also very thankful to Dr. Attila Kereszt to supervising me and always providing support and insight during my thesis work. I would like to thank to Dr. Moussa Benhamed, Dr. Federico Ariel, Dr. Christian Chevalier to accepting to participate as members of my thesis committee and providing me very helpful advices. I am very thankful to my opponents Dr. Pál Venetianer and Dr. Francois Roudier for critical reviewing of my thesis. I would also like to thank to Dr. Mickael Bourge for helping me with the flow cytometry experiments. I also need to thank to all of my colleagues in Szeged and Gif-sur-Yvette for providing me feedback and always being supportive. Additional thanks to Campus France for the financial support, which enabled my 15-month stay and research in Paris. I must thank the Biology department at the University of Szeged in Hungary and at Paris-Sud in France for providing me the opportunity to complete my thesis research. Finally, but definitely not least, I would like to thank my family especially my Mother and Father for encouraging me to pursue my goals and dreams, they are true inspiration in my life and has taught me how to endure through hard times and to never give up.

## **ABBREVIATIONS**

5-mC: 5'-methyl cytosine

ATAC: assay for transposase accessible chromatin

BacA: bacterial development factor A

BAP: 6-benzylaminopurine, synthetic cytokinin

BS-seq: bisulphite sequencing

C: haploid DNA content

Ccs52: Cell cycle switch gene

ChIP: chromatin immunoprecipitation

CLF: CURLY LEAF H3K27me3 methyltransferase gene,

CS: cell suspension

DAPI: 4',6-diamidine-2'-phenylindole dihydrochloride DNA staining dye

DME: DNA glycosylase/demethylase

DmC: differentially methylated cytosine

DmR: differentially methylated regions

DmRG: differentially methylated region associated gene

dpi: days post inoculation

ENOD12: early nodulin 12

ENR: endoreduplication

EST: expressed sequence tag

FIE: FERTILIZATION INDEPENDENT ENDOSPERM gene, component of PRC2

FLC: flowering locus C

G: gap phase of cell cycle

GFP: green fluorescent protein

GO: gene ontology

GOGAT: glutamate synthase

GRP: glycine-rich peptide

GS: glutamine synthase

GUS:  $\beta$ -glucuronidase reporter gene

H3K9ac: acetylation of histone H3 lysine 9

H3K27me3: trimethylation of histone H3 lysine 27

IRLC: Inverted Repeat Lacking Clade of legumes

IP: immunoprecipitation

IT: infection thread

kb: kilobase

Lb: leghemoglobin

LCM: laser-capture microdissection

LCO: lipo-chitooligosaccharides

LND: low nucleosome density region

M: cell division, mitosis

MBD2: methyl-CpG-binding domain protein 2

mC: methylated cytosines

M-CDK: mitotic cyclin-dependent kinase

MeDIP: methylated DNA immunoprecipitation

MJ: methyl jasmonate

MtGEA: *Medicago truncatula* Gene Expression Atlas

N: nitrogen

NAA: 1-naphthaleneacetic acid, synthetic phytohormone auxin

NCR: Nodule-specific Cysteine-Rich peptide

NF: Nod factor

NFR: Nod factor receptor

NH<sub>3</sub>: ammonia

NH<sub>4</sub><sup>+</sup>: ammonium

NO<sub>3</sub><sup>-</sup>: nitrate

NPA: N-1-naphthylphthalamic acid, auxin transport inhibitor

PcG: Polycomb Group protein

PRC1, 2: Polycomb Repressive Complex 1, 2

PTM: post-translational histone modification

RdDM: RNA-dependent DNA methylation

ROS1: REPRESSOR OF SILENCING 1 demethylase

RRBS: Reduced Representation Bisulphite Sequencing

RT-qPCR: reverse transcription quantitative real-time PCR

S: DNA synthesis phase of cell cycle

SNARP: small nodulin acidic RNA-binding protein

SNF: symbiotic nitrogen fixation

SWN: SWINGER H3K27me3 methyltransferase gene

T3SS: type III secretion system

TE: transposable element

TIBA: 2,3,5-triiodobenzoic acid, auxin transport inhibitor

TF: transcription factor

TSS: translation start site

VAG1: vagrant infection thread 1 gene

wpi: weeks post inoculation

YE: yeast elicitor

ZI: nodule zone I, meristem

ZII<sub>d</sub>: distal nodule zone II, invasion zone

ZII<sub>p</sub>: proximal nodule zone II, invasion zone

IZ: nodule zone II-III, interzone

ZIII: nodule zone III, nitrogen-fixing zone

ZIV: nodule zone IV, senescence zone



# TABLE OF CONTENT

<b>ACKNOWLEDGEMENTS</b> .....	<b>4</b>
<b>ABBREVIATIONS</b> .....	<b>5</b>
<b>RÉSUMÉ (traduit en utilisant Google Translate)</b> .....	<b>13</b>
<b>1. INTRODUCTION</b> .....	<b>23</b>
1.1. The <i>Rhizobium</i> -legume symbiosis .....	23
1.1.1. The agronomic and ecological importance of legume plants.....	23
1.1.2. Forms of mutual dependence: Nitrogen-fixing symbiotic associations .....	25
1.1.3. Rhizobial infection and nodule development.....	28
1.1.4. Polyploidisation of symbiotic cells is a prerequisite for nodule organogenesis ....	31
<b>1.2. The Nodule-specific Cysteine-Rich (NCR) peptide family</b> .....	<b>34</b>
1.2.1. <i>NCR</i> genes are specific for the IRLC and Dalbergioid clade legumes .....	34
1.2.2. Bacteroid morphotypes: NCRs direct irreversible bacteroid differentiation .....	36
1.2.3. Possible mode of action and intracellular targets of the NCR peptides .....	38
1.2.4. <i>NCRs</i> are specifically induced during nodulation: Early and late genes.....	40
1.2.5. Transcriptional regulation of <i>NCRs</i> : Epigenetic mechanisms might be involved .	42
<b>1.3. Epigenetic control of gene expression</b> .....	<b>43</b>
1.3.1. DNA methylation dynamics.....	44
1.3.2. Chromatin modifications: The histone code .....	47
1.3.2.1. H3K9ac marks active genes .....	48
1.3.2.2. H3K27me3 marks developmentally regulated genes .....	49
1.3.3. DNA accessibility and chromatin remodelling .....	50
<b>1.4. Plant processes under epigenetic control</b> .....	<b>52</b>
1.4.1. Epigenetic changes during development.....	52
1.4.2. Epigenetic changes in response to biotic and abiotic stress .....	55
<b>2. OBJECTIVES OF THE WORK</b> .....	<b>58</b>

<b>3. MATERIALS AND METHODS .....</b>	<b>62</b>
<b>3.1. NCR expression analysis and entropy calculation .....</b>	<b>62</b>
3.1.1. Analysis of the <i>Medicago truncatula</i> Gene Expression Atlas (MtGEA) data .....	62
3.1.2. Transcriptome analysis of hand-dissected nodule zones.....	62
3.1.3. Laser-capture microdissection coupled to RNA-Seq (LCM RNA-seq).....	63
3.1.4. Shannon entropy calculations.....	63
<b>3.2. Transgenic plants and <math>\beta</math>-glucuronidase reporter (GUS) analysis .....</b>	<b>63</b>
3.2.1. Promoter-GUS constructs .....	63
3.2.2. Stable transgenic GUS plant lines .....	64
3.2.3. GUS staining .....	64
3.2.4. Pathogen assays.....	64
<b>3.3. Antibodies and immunolocalization.....</b>	<b>65</b>
<b>3.4. Preparation of plant material for flow cytometry .....</b>	<b>65</b>
3.4.1. Seed sterilisation and germination .....	65
3.4.2. Plant inoculation with <i>Rhizobium</i> and nodulation.....	65
3.4.3. Flow cytometry sorting of different ploidy level nodule nuclei .....	66
<b>3.5. Nuclear RNA isolation, reverse transcription and quantitative PCR analysis .....</b>	<b>66</b>
3.5.1. Nuclear RNA extraction.....	66
3.5.2. DNase treatment.....	67
3.5.3. Reverse transcription.....	67
3.5.4. Quantitative PCR (qPCR) .....	67
3.5.5. Statistical analysis .....	68
<b>3.6. Genomic DNA extraction and Methylated DNA Immunoprecipitation (MeDIP) .....</b>	<b>69</b>
3.6.1. Genomic DNA extraction from sorted nuclei .....	69
3.6.2. Shearing the genomic DNA .....	69
3.6.3. MeDIP .....	69
3.6.4. Methylated DNA clean-up .....	70

<b>3.7. Chromatin isolation and chromatin immunoprecipitation (ChIP)</b> .....	<b>70</b>
3.7.1. Isolation and shearing of the chromatin .....	70
3.7.2. ChIP.....	71
<b>3.8. Quantitative real-time PCR analysis of MeDIP and ChIP DNA</b> .....	<b>71</b>
3.8.1. qPCR .....	71
3.8.2. Data analysis .....	72
3.8.3. Statistical analysis .....	72
<b>3.9. Reduced representation bisulphite sequencing (RRBS)</b> .....	<b>72</b>
3.9.1. Methyl-MidiSeq library construction.....	73
3.9.2. Methyl-MidiSeq sequence alignments and methylated cytosine (mC) calling.....	73
3.9.3. Detection of differentially methylated regions (DmRs).....	74
<b>3.10.Data of RNA-Seq coupled to laser microdissection of different nodule zone</b> .....	<b>74</b>
<b>3.11.Assay for transposase-accessible chromatin using sequencing (ATAC-Seq)</b> .....	<b>75</b>
3.11.1. Transposition and library preparation.....	75
3.11.2. Library amplification .....	76
3.11.3. Sequencing and data analysis .....	76
<b>4. RESULTS</b> .....	<b>78</b>
<b>4.1. Extreme specificity of <i>NCR</i> gene expression in <i>Medicago truncatula</i></b> .....	<b>78</b>
4.1.1. Global analysis of <i>NCR</i> gene expression .....	78
4.1.2. Spatio-temporal expression of <i>NCR</i> genes in nodules .....	81
4.1.3. <i>NCR</i> genes are not directly involved in nodule senescence .....	86
4.1.4. Promoter-GUS analysis and immunolocalization of selected <i>NCRs</i> in nodules....	88
4.1.5. <i>NCR</i> genes are not expressed after biotic and abiotic stress .....	90
4.1.6. Promoter-GUS analysis of <i>NCR</i> expression during pathogenic interactions.....	93
4.1.7. <i>NCR</i> genes have very high tissue specificity as measured by Shannon entropy....	95
4.1.8. Conclusion.....	97

<b>4.2. Ploidy-dependent changes in the epigenome of symbiotic cells correlate with specific patterns of gene expression.....</b>	<b>99</b>
4.2.1. Differential expression of nodule-specific genes depends on the ploidy levels of the symbiotic cells. ....	100
4.2.2. Only a subset of genes display ploidy-dependent differential DNA methylation	101
4.2.3. Chromatin accessibility varies with ploidy and correlates with gene expression	107
4.2.4. Ploidy-dependent variation in H3K27me3 and H3K9ac chromatin marks correlates with the expression pattern of <i>NCR</i> genes .....	109
4.2.5. Conclusions .....	111
<b>5. DISCUSSION .....</b>	<b>112</b>
5.1. <i>NCR</i> genes are expressed during symbiosis in the rhizobia infected nodule cells ....	112
5.2. <i>NCRs</i> are not involved in innate immunity .....	113
5.3. <i>NCRs</i> are activated sequentially in different spatio-temporal waves .....	113
5.4. <i>NCR</i> genes are differentially expressed according to the ploidy level of the differentiating symbiotic cells.....	114
5.5. <i>NCR</i> genes are among the most specifically expressed genes in <i>M. truncatula</i> .....	115
5.6. Differential DNA methylation is rare but predominant in the <i>NCR</i> gene family .....	116
5.7. Chromatin accessibility changes with the ploidy levels .....	117
5.8. Histone modifications: The level of H3K27me3 is crucial for gene expression and is epistatic over H3K9ac.....	118
5.9. Epigenome reconfiguration in endoreduplication-driven cell differentiation .....	119
5.10. Concluding remarks .....	119
<b>REFERENCES .....</b>	<b>121</b>
<b>APPENDIX .....</b>	<b>143</b>
Supplementary information 1. ....	143
Supplementary information 2. ....	143

Supplementary information 3. ....	143
Table 1. Primers used for the amplification and cloning of the promoters. ....	144
Table 2. Primers used for RT-qPCR experiment. ....	145
Table 3. Primers used for MeDIP-qPCR and ChIP-qPCR experiment. ....	146
Table 4. ATAC-seq oligos used for PCR. ....	148
Table 5. GO term enrichment of <i>M. truncatula</i> genes with differential DNA methylation	149

## RÉSUMÉ

Les plantes légumineuses peuvent interagir avec les bactéries du sol de la famille des rhizobiacées. Cette interaction conduit au développement d'un organe spécialisé appelé nodule racinaire. À l'intérieur des cellules végétales symbiotiques, les rhizobiums sont capables de fixer l'azote atmosphérique et de le convertir en ammoniac, qui est une source d'azote utilisable pour la plante. Les cellules symbiotiques proviennent de la division des cellules progénitrices dans le méristème nodulaire. Pendant la différenciation cellulaire symbiotique, les cellules post-méristématiques sortent du cycle de division cellulaire et entrent dans un cycle d'endoréduplication, un cycle cellulaire modifié avec réplication répétée du génome sans mitose et cytokinèse, conduisant à des cellules polyploïdes avec une teneur en ADN accrue (ploïdie de 2C à 64C) et l'agrandissement des cellules jusqu'à 80 fois par rapport à la cellule méristématique (Vinardell et al. 2003). Les cellules symbiotiques de différenciation sont progressivement infectées et remplies de symbiomes formant des rhizobiums, des structures ressemblant à des organites fixant l'azote. Les nodules de *Medicago truncatula* sont de type indéterminé, contenant un méristème apical persistant (ZI), une zone d'infection (ZIIId distale et ZIIp proximale), une interzone (IZ) et une zone de fixation d'azote (ZIII) et dans des nodules plus âgés une zone de sénescence (ZIV). Les cellules matures de la zone III sont entièrement remplies de symbiomes et leur physiologie est adaptée à la symbiose, alimentant les microsymbiotes et assimilant l'azote fixé.

La génération de cellules symbiotiques différenciées dans la légumineuse *Medicago truncatula* implique une reprogrammation massive du transcriptome nodulaire avec l'activation de gènes spécifiques des nodules dans des ondes spatio-temporelles successives (Maunoury et al. 2010). Beaucoup de gènes induits par les nodules sont également exprimés dans d'autres tissus végétaux, mais une grande partie du transcriptome nodulaire est composée de gènes exclusivement exprimés dans les nodules et maintenus silencieux dans tous les autres tissus végétaux. Parmi les gènes activés, on trouve la famille NCR, produisant des centaines de peptides NCR différents qui guident la différenciation terminale des endosymbiotes, les convertissant en grands bactéroïdes polyploïdes non-cultivables fixateurs d'azote. Les gènes NCR sont soumis à un contrôle transcriptionnel strict car ils sont exclusivement exprimés dans les cellules nodulaires symbiotiques et leur activation nécessite la présence intracellulaire de rhizobiums (Van de Velde et al. 2010, Maunoury et al. 2010, Mergaert et al. 2003). L'expression des gènes NCR dans les cellules symbiotiques suggère une connexion directe entre l'endoréduplication et l'expression des gènes NCR. Comment endoreduplication ou l'état

polyploïde du génome contrôle l'expression des gènes est entièrement inconnue. La spécificité stricte des nodules et le lien possible entre l'expression génique et les niveaux de ploïdie des cellules symbiotiques suggèrent que les modifications épigénétiques le long des cycles d'endoréduplication pourraient jouer un rôle crucial dans la régulation transcriptionnelle des *NCR* et d'autres gènes spécifiques des nodules. La reprogrammation récemment rapportée de la méthylation de l'ADN dans le développement du nodule de *M. truncatula*, régulée par le gène DEMETER de l'ADN déméthylase (MtDME), est en accord avec cette hypothèse (Satgé et al. 2016).

Les changements dans la méthylation de la 5-cytosine de l'ADN et dans les modifications de la queue des histones, seuls ou en combinaison, sont des déterminants importants de la structure de la chromatine et de l'expression des gènes. Le contexte de la séquence (CG, CHG, CHH, où H = A, C ou T) et la localisation de la méthylation de l'ADN dans le promoteur, le corps du gène ou la région régulatrice 3' peuvent avoir des effets différents sur la transcription génique. En général, il existe une forte corrélation entre la méthylation de l'ADN dans la région du promoteur et la diminution de l'expression génique (Zhang et al. 2006, Zilberman et al. 2007, Cokus et al. 2008, Garg et al. 2015). Parmi les modifications des queues d'histones, la triméthylation de l'histone H3 lysine 27 (H3K27me3) (Feng & Jacobsen 2011) entraîne également un faible niveau d'expression et une haute spécificité tissulaire (Zhang et al. 2007) alors que l'histone H3 lysine 9 acétylation (H3K9ac) activation (Kurdistani et al. 2004, Schubeler et al. 2004, Roh et al. 2005, Zhou et al. 2010).

Ainsi, les cellules nodulaires symbiotiques de *Medicago truncatula* constituent un excellent système modèle pour étudier la reprogrammation transcriptionnelle au niveau spécifique du stade de développement car la différenciation cellulaire symbiotique implique une endoréduplication et une reprogrammation massive de leur transcriptome. De plus, les nodules indéterminés de *M. truncatula* contiennent des cellules à tous les stades de différenciation symbiotique, à partir des cellules méristématiques indifférenciées (2C / 4C), au stade post-mitotique lorsqu'elles ont arrêté la division cellulaire et ont une ploïdie 4C et l'intermédiaire, stades de différenciation avec des niveaux de ploïdie de 8C-16C, jusqu'aux cellules matures fixatrices d'azote avec des niveaux de ploïdie de 32C-64C. Cette propriété de ces nodules facilite l'échantillonnage de toutes les étapes possibles de différenciation des cellules symbiotiques.

De la spécificité élevée des gènes spécifiques des cellules nodulaires symbiotiques et du lien apparent de leur activation transcriptionnelle lors de l'endoréduplication dans les cellules symbiotiques, nous avons supposé que des mécanismes épigénétiques impliquant des

modifications spécifiques des histones et / ou des méthylations d'ADN pourraient être impliqués dans la détermination du gène programme d'expression dans les cellules symbiotiques. Les structures de la chromatine répressive peuvent garder les gènes silencieux pendant toutes les étapes du développement de la plante. L'inversion de ces marques dans les cellules symbiotiques de différenciation serait nécessaire pour les rendre compétentes pour la transcription par des facteurs de transcription spécifiques. De plus, nous proposons que l'endoréduplication des cellules nodulaires symbiotiques sert de mécanisme pour contrôler l'expression des gènes. L'endoréduplication pourrait mener activement à l'activation des "effaceurs" des marques répressives de la chromatine. Alternativement, les marques répressives, qui sont normalement fidèlement transmises au cours du cycle cellulaire, pourraient être perdues passivement par dilution au cours des endocycles consécutifs. Ces hypothèses constituent la base de mon travail de thèse.

## **OBJECTIFS**

### **A. Caractériser en détail le comportement d'expression d'une famille de gènes NCR (Nodule-Specific Cysteine-Rich).**

1. La base de données du transcriptome fournie par l'Atlas d'expression génique *Medicago truncatula* (MtGEA) (Benedito et al. 2008; He et al. 2009) est générée avec le génome entier Affymetrix *Medicago* Gene Chip et compile les données de la puce de *M. truncatula*. gènes (50 900 ensembles de sondes) sur un large ensemble d'expériences (267 expériences différentes) comprenant différents organes végétaux, biotiques, conditions de stress abiotique et différents types de tissus. La base de données MtGEA est actuellement la ressource la plus riche pour analyser le profil d'expression et la spécificité des gènes *NCR* de *M. truncatula* dans diverses conditions.

2. Les nodules de *M. truncatula* sont de type indéterminé; par conséquent, les nodules matures sont organisés en zones histologiques bien définies. Cette structure nodulaire suggère que les profils d'expression *NCR* temporels pourraient être corrélés avec un profil d'expression spatial dans les tissus nodulaires. Pour tester cette possibilité, nous avons analysé des données de transcriptome à partir de nodules de 4 semaines (28 jour après inoculation, dpi) qui ont été sectionnés à la main dans cinq parties différentes (ZI, ZIIId, ZIIp, IZ et ZIII).



3. Afin de confirmer les données d'expression de MtGEA, des lignées de *M. truncatula* R108 transgéniques stables ont été générées portant des constructions de fusion promoteur- $\beta$ -glucuronidase reporter (GUS) pour 3 gènes *NCR* différents, représentant différentes classes temporelles de *NCR*, ainsi qu'un anticorps spécifique pour un peptide *NCR*.

4. En outre, le recueil MtGEA contient des ensembles de données provenant de différentes interactions pathogènes, mais il s'agit exclusivement de pathogènes racinaires. Pour confirmer la spécificité des profils d'expression *NCR* dans diverses conditions, en particulier lors d'infections foliaires ou pathogènes englobant également d'autres interactions trophiques (bio / hémibio / nécrotrophe), des lignées de *M. truncatula* R108 transgéniques stables portant des constructions de fusion rapporteur *NCR*-GUS ont été utilisées.

**B. L'implication de la régulation épigénétique dans la reprogrammation transcriptionnelle au cours de la différenciation cellulaire symbiotique a été étudiée à la résolution du stade de différenciation des cellules symbiotiques en purifiant les noyaux nodaux de niveau de ploïdie défini avec un cytomètre en flux couplé à un trieur cellulaire. Ces expériences démontrent si les changements d'expression génique sont liés à la ploïdie croissante au cours de la différenciation des cellules nodulaires symbiotiques.**

5. L'expression *in situ* de divers gènes spécifiques de *NCR* et de nodules suggère que leur expression pourrait être couplée à des niveaux de ploïdie spécifiques des cellules nodulaires. Pour confirmer un lien possible entre l'expression de gènes spécifiques de nodules dans les ondes transcriptionnelles successives et les niveaux de ploïdie des cellules symbiotiques différenciées, le niveau d'expression des gènes sélectionnés différenciellement exprimés a été mesuré dans les noyaux 4C, 8C, 16C et 32C cytomètre.

6. Comme la majorité des gènes spécifiques des nodules sont exprimés dans les cellules symbiotiques matures (32C) et réprimés dans les cellules non infectées (4C), nous avons analysé les différences de génome de l'ADN (5-mC) dans les noyaux isolés 4C et 32C et les profils de méthylation de l'ADN ont été appariés aux données d'ARN *in situ* publiées sur les zones de nodules obtenues par microdissection par capture laser (Roux et al. 2014). De plus, la méthylation différentielle des gènes spécifiques des nodules et des nodules exprimés dans les noyaux 4C, 8C, 16C et 32C a été analysée.

7. Pour savoir comment le compactage de la chromatine et l'accessibilité changent au cours de la différenciation des nodules et comment cela influence l'expression des gènes, nous avons étudié l'accessibilité de la chromatine dans les noyaux 4C, 8C, 16C et 32C. Les profils des modifications de la chromatine ont été appariés aux données publiées in situ sur l'ARN-seq des zones de nodules obtenues par microdissection par capture laser (Roux et al. 2014).

8. Pour analyser la corrélation entre les modifications de la chromatine et l'expression des gènes, nous avons étudié comment les marques de chromatine antagonistes H3K27me3 et H3K9ac sont en corrélation avec l'activation et la répression de gènes sélectionnés, différemment exprimés dans différents noyaux de ploïdie.

## RÉSULTATS

1. À partir de l'extraction de données du transcriptome et de la confirmation expérimentale de l'expression de *NCR*, nous pouvons conclure que, à quelques exceptions près, les centaines de gènes *NCR* codant des peptides de type défensine ne sont activés que pendant la formation des nodules. Le profil d'expression de 334 gènes *NCR* a été analysé sur 267 expériences différentes incluant 9 organes végétaux, des conditions de croissance biotiques et abiotiques et différents stades de développement en utilisant le MtGEA et d'autres données de transcriptome publiées. Les gènes *NCR* sont exprimés dans les nodules, mais dans aucun autre organe de la plante ou en réponse aux phytohormones ou à la sécheresse et au stress salin. Pendant la symbiose, aucun des gènes *NCR* n'est induit par les facteurs Nod et ils ne sont pas activés dans les nodules avant que les rhizobiums soient libérés dans les cellules hôtes et que les cellules symbiotiques ne se forment pas. Nous avons également montré que les gènes *NCR* ne sont pas impliqués dans la dégradation des cellules nodulaires symbiotiques et des bactéroïdes durant la sénescence nodulaire puisque leur expression a été arrêtée lorsque la sénescence est initiée; par conséquent, les peptides *NCR* n'ont aucun rôle dans la mort et la désintégration des rhizobiums dans les cellules végétales sénescences. Les *NCR* ressemblent à des peptides antimicrobiens tels que les défensines de l'immunité innée et de nombreux peptides *NCR*. In vitro, les peptides cationiques particulièrement à haute charge positive ont une forte activité antimicrobienne contre une diversité de bactéries Gram négatif et Gram positif incluant les pathogènes humains et végétaux ainsi que les champignons (Van de Velde et al. 2010, Tiricz et al. 2013, Ördögh et al. 2014). Malgré cela, les gènes *NCR* ne sont pas induits lors d'une attaque pathogène par des bactéries, des champignons, des oomycètes ou des nématodes. Ils ne

sont exprimés ni dans les organes «sensibles à l'infection» comme les feuilles, les graines et les fleurs qui expriment souvent des niveaux élevés de peptides antimicrobiens à immunité innée (Sels et al. 2008). Par conséquent, il semble que les peptides NCR n'ont aucune fonction dans l'immunité innée. Les gènes *NCR* ne sont pas activés par les facteurs Nod ou durant les tout premiers stades de l'organogenèse des nodules avant la formation des cellules infectées (ce travail, Nallu et al. 2013, Maunoury et al. 2010). En outre, quantifier la spécificité de l'expression avec le facteur d'entropie de Shannon révèle que les gènes *NCR*, et plus généralement, les gènes spécifiques des nodules sont parmi les gènes les plus spécifiquement exprimés chez *M. truncatula*. De plus, lorsqu'il est activé dans les nodules, leur niveau d'expression est parmi les plus élevés de tous les gènes. Ensemble, ces données montrent que l'expression du gène *NCR* est soumise à une régulation extrêmement stricte et n'est activée que pendant l'organogenèse des nodules dans les cellules symbiotiques.

2. Nous avons montré que les gènes *NCR* sont activés au cours du développement des nodules dans au moins trois ondes temporelles correspondant à des modèles d'expression spatiale spécifiques. Les gènes activés précocement dans le développement des nodules sont exprimés dans les parties les plus distales des nodules (proches de l'apex, ZII) tandis que les gènes activés tardivement au cours du développement sont exprimés dans les tissus des nodules proximaux (IZ-ZIII). De plus, certaines grappes de gènes, une fois activées, conservent leur activité au cours du vieillissement des cellules nodulaires, tandis que d'autres sont caractérisées par une absence d'expression dans les cellules nodulaires plus âgées. Notre analyse spatiale de l'expression de *NCR* est en bon accord avec une étude récemment publiée qui utilisait la LCM des zones nodulaires couplées au séquençage de l'ARN (Roux et al. 2014).

3. En accord avec sa régulation temporelle au cours de la nodulation, l'expression de *NCR121* a été induite chez de jeunes primordia nodulaires dès 5 dpi et restée exprimée dans toute la zone d'infection II et la zone de fixation III des nodules matures. *NCR084* expression a été détectée à partir de 11 dpi sur et a été principalement confinée à la zone d'infection proximale, l'interzone II-III et à la partie distale de la zone de fixation III. L'expression de *NCR001* était détectable à partir de 11 dpi dans la zone de fixation en développement III et son expression s'étendait les jours suivants lorsque la zone de fixation se développait. Les 3 gènes sont seulement exprimés dans les cellules nodulaires symbiotiques. L'expression de GUS dans les trois lignées transgéniques n'a pas été détectée dans les extrémités des racines ou d'autres parties de la racine. Dans les nodules plus âgés, à 30 dpi, présentant une zone de sénescence, l'expression de *NCR* n'a jamais été détectée dans les tissus sénescents et leur expression dans la

zone proximale adjacente au tissu sénescence n'a pas été confirmée, confirmant que les gènes *NCR* ne sont pas impliqués dans la sénescence.

Le profil d'expression particulier de *NCR122* avec sa spécificité tissulaire relâchée et son expression apparente dans les cellules nodulaires non infectées nous a incités à analyser spécifiquement la localisation du peptide *NCR122* dans les nodules avec l'anticorps anti-*NCR122*. L'immunolocalisation du peptide révèle en effet une présence spécifique de *NCR122* dans les cellules non infectées de la zone centrale d'un nodule mature ainsi que dans les cellules corticales non infectées du nodule.

4. Les résultats des essais pathologiques sont en accord avec une étude récente montrant que l'expression de *NCR* n'a pas été détectée lors de l'interaction entre *M. truncatula* et le pathogène biotrophique du sol *Phytophthora medicaginis*, le pathogène hémibiotrophique des feuilles *Colletotrichum trifolii* (Tesfaye et al. 2013) et le champignon nécrotrophique gris causant la moisissure *Botrytis cinerea*. Nos données sont en accord avec l'ensemble de données MtGEA et élargissent la conclusion que les *NCR* ne sont pas impliqués dans la réponse aux agents pathogènes, quel que soit le système trophique (bio / hémibio / nécrotrophe). De plus, en tant qu'herbivorie et plus généralement, l'effet de la blessure mécanique, qui peut induire des défenses de la plante, a également été testé, mais encore une fois, aucune expression de *NCR* n'a pu être détectée dans le feuillet blessé.

5. Nous avons montré dans cette étude que l'organogenèse des nodules s'accompagne de changements majeurs dans le programme d'expression génique (Maunoury et al. 2010, Roux et al. 2014, ce travail). La localisation des transcrits *in situ*, les fusions de gènes promoteurs-reporters ont révélé que les gènes précoces sont activés dans les jeunes cellules symbiotiques différenciées alors que les gènes tardifs dans les cellules fixatrices d'azote matures (Mergaert et al. 2010; Van de Velde et al. 2010, Farkas et al. 2014, Roux et al. 2014, ce travail) et ce modèle suggère que leur expression pourrait être couplée à des niveaux de ploïdie spécifiques. En utilisant des noyaux de nodules triés en fonction de leur teneur en ADN, le profil d'expression de gènes spécifiques de nodules exprimés et de nodules spécifiques était corrélé avec leur profil d'expression spatiale dans les nodules (Roux et al. 2014). Nous avons trouvé que l'activation des premiers gènes se produit dans les cellules post-méristématiques 4C tandis que les gènes de la deuxième vague sont activés dans les parties distales et proximales de ZII avec 8C et 16C tandis que les gènes tardifs de la troisième onde transcriptionnelle sont induits stades plus avancés du développement des nodules dans les IZ et ZIII dans les cellules de niveau élevé de

ploïdie de 32C et 64C. Nous avons trouvé une très bonne corrélation entre les profils d'expression *NCR* spatiaux connus et les profils d'expression dans les noyaux de différents niveaux de ploïdie.

6. En réponse aux molécules de signal de *Rhizobium*, les facteurs Nod, la prolifération des cellules corticales des racines conduit à l'établissement du méristème nodulaire. La méthylation de l'ADN des cellules méristématiques représente le statut initial et probablement le plus méthylé de l'ADN. La dynamique de la méthylation de l'ADN à la résolution d'une seule base a été évaluée dans des cellules non infectées (4C) et symbiotiques (32C). La plus forte fraction de mCs a été observée dans un contexte de séquence CG (75%) suivie par CHG (20%) et CHH (5%) dans les deux types de cellules, ce qui était cohérent avec les études précédentes. La méthylation différentielle de l'ADN n'a été trouvée que dans un petit sous-ensemble de gènes spécifiques des nodules symbiotiques, dont plus de la moitié des gènes *NCR*, alors que la méthylation de l'ADN n'était pas affectée par les niveaux de ploïdie. Cette découverte indique que l'état de méthylation de l'ADN des cellules 4C a été copié au cours des cycles répétés d'endoréduplication. Contrairement à la plupart des gènes codant pour les protéines, les *NCR* se comportaient différemment car sur 375 gènes, 164 devenaient hypométhylés, ce qui correspond à la régulation à la hausse de la déméthylase de l'ADN MtDME spécifique des nodules dans les cellules interzone ainsi qu'à la régulation négative des gènes *NCR* dans le MtDME. Lignes d'interférence ARN (Satgé et al. 2016). Les modèles de méthylation différentielle que nous avons observés pourraient être en grande partie la conséquence de l'action de MtDME. Pourquoi l'activation de certains gènes *NCR* est associée à la déméthylation et pas les autres, nécessite une enquête plus approfondie. Une possibilité intéressante de l'importance de la méthylation différentielle de l'ADN pour l'expression d'un sous-ensemble de gènes *NCR* pourrait être liée à la présence d'éléments transposables (TE) (Satgé et al. 2016). Les génomes végétaux sont riches en TE, qui sont habituellement maintenus silencieux transcriptionnellement par la méthylation de l'ADN. Cette extinction de TE peut également affecter les gènes voisins et l'activation du gène *NCR* peut donc nécessiter une déméthylation médiée par MtDME.

7. Nous avons trouvé que 44% des gènes *NCR* étaient hypométhylés dans les cellules symbiotiques 32C par rapport à 4C et cela a également été validé avec MeDIP, suggérant que la diminution de la méthylation de l'ADN, en particulier dans leur région 1-kb amont, contribue à leur expression. Mais environ 53% des gènes *NCR* ne diffèrent pas dans la méthylation de l'ADN à 4C et 32C. Cela signifie que ces gènes pourraient être actifs dans les cellules 8C et

16C et afficher une hypométhylation dans ces noyaux. Nous avons testé cette hypothèse sur des gènes sélectionnés le long de leurs régions amont et aval à 1 kb du site de départ de traduction (TSS), mais l'étendue de la méthylation de l'ADN n'a pas été affectée par les niveaux de ploïdie confirmant les résultats du Séquençage de bisulfite à représentation réduite (RRBS). Cela pourrait signifier que la méthylation de l'ADN ne joue pas un rôle central dans leur régulation et que les gènes dont la méthylation de l'ADN n'est pas modifiée pourraient nécessiter des modifications de la chromatine.

8. La structure de la chromatine a été impliquée dans la régulation développementale et tissulaire d'un certain nombre de gènes de plantes et d'animaux. L'ouverture et la fermeture dynamiques de la chromatine à différents niveaux de ploïdie des cellules nodulaires étaient bien corrélées avec l'état d'expression actif / réprimé des gènes. Nous avons également constaté que la chromatine ouverte seule n'était pas suffisante pour l'activation du gène. Bien que l'accessibilité de la chromatine ait montré de petites variations aux niveaux de 4 ploïdies, l'état de chromatine moins condensée a été trouvé dans les cellules symbiotiques matures avec 32C. Ainsi, des profils d'expression spécifiques au tissu et au stade de développement dans le nodule s'accompagnent d'altérations dynamiques de la structure de la chromatine. La réorganisation de la chromatine était particulièrement importante à proximité ou sur le site de démarrage de la traduction. Les gènes précoces présentaient une accessibilité maximale dans les cellules 4C et 8C, tandis que les gènes tardifs dans les cellules 16C et 32C. Nous avons également constaté que la modification de l'accessibilité de la chromatine n'était pas étroitement liée à la méthylation de l'ADN, car elle se produisait également lorsque l'état de méthylation de l'ADN n'était pas affecté par la ploïdie. De plus, le degré d'ouverture de la chromatine n'a pas toujours montré une corrélation directe avec le niveau de méthylation de l'ADN.

9. En outre, nous avons analysé la répression H3K27me3 et le profil d'activation H3K9ac des gènes spécifiques de la symbiose sélectionnés à partir de différentes classes d'expression spatiales et temporelles dans les noyaux de ploïdie en croissance. Les marques H3K27me3 étaient prédominantes dans les cellules 4C et sont probablement essentielles pour l'état réprimé des *NCR* et d'autres gènes symbiotiques dans ces cellules. En cas de gènes précoces, les marques H3K27me3 ont été radicalement réduites dans les cellules 8C et 16C et augmentées à nouveau dans les cellules 32C corrélées avec leur répression, tandis que dans le cas des gènes tardifs, les marques H3K27me3 réduites combinées à la présence des marques actives H3K9ac et accessibles. la structure de la chromatine a contribué à l'activation du gène. Les *NCR* dans les cellules symbiotiques sont habituellement exprimées à des niveaux très élevés

qui, en plus de H3K27me3 réduit, ont également les marques actives H3K9ac dans le corps du gène et dans le promoteur. Le rôle principal de la marque H3K27me3 dans la régulation de ces gènes est conforme aux observations précédentes selon lesquelles le dépôt et la régulation dynamique de la marque H3K27me3 sont importants pour contrôler l'expression génique spécifique des tissus et la différenciation des plantes (Lafos et al. 2011).

En identifiant la méthylation de l'ADN spécifique à la ploïdie, l'H3K27me3 répressive et l'activation des marques d'histones H3K9ac et les profils d'accessibilité de la chromatine des gènes différentiellement exprimés, nous obtenons une première vue d'un contrôle épigénétique multicouche de la différenciation cellulaire symbiotique. Nos résultats constituent un pas en avant dans la compréhension de la régulation extrêmement stricte des gènes *NCR*. De plus, les changements dynamiques observés dans l'épigénome suggèrent que l'accessibilité de la chromatine et les modifications de la queue d'histone régulent l'état transcriptionnellement actif ou inerte des gènes.

# 1. INTRODUCTION

## 1.1. The *Rhizobium*-legume symbiosis

### 1.1.1. The agronomic and ecological importance of legume plants

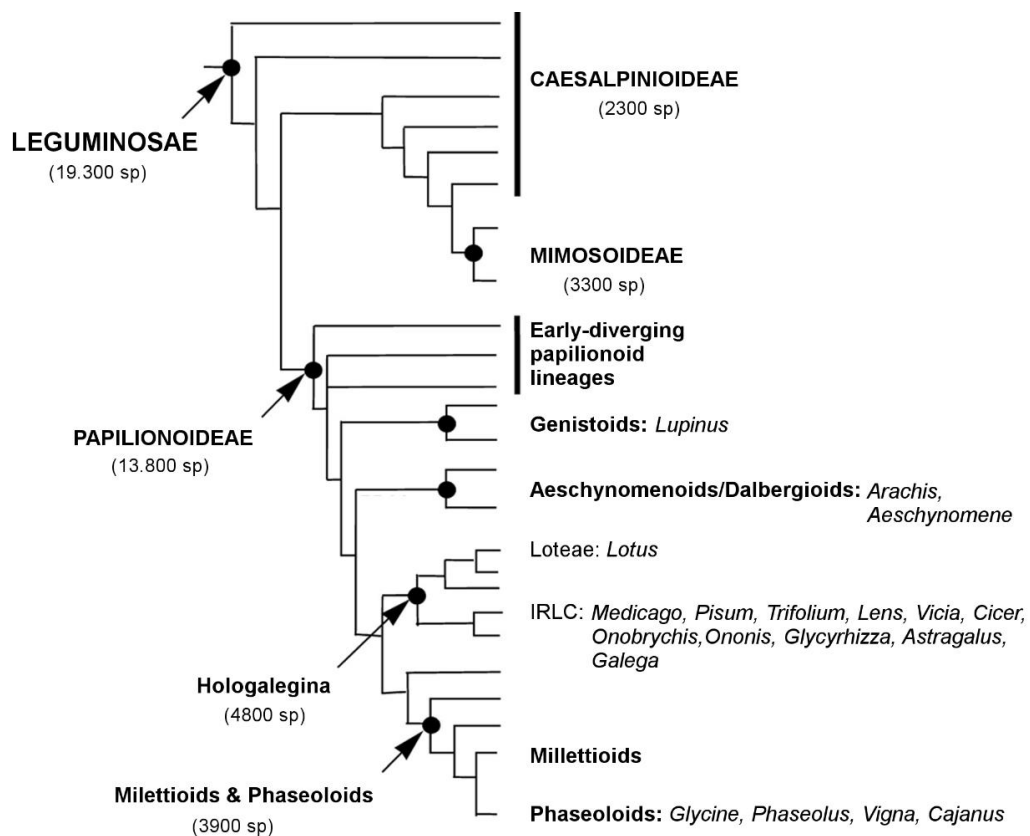
*Leguminosae* (or *Fabaceae*) is the third largest family in the plant kingdom (Delprete & Mabberley 1998) with 946 genera and over 24,500 species worldwide (The Plant List 2013). Legumes are second to grasses (*Poaceae*) in agricultural importance. Moreover, they have a tremendous ecological impact through their major input of combined (fixed) nitrogen (N) in the biosphere, which results from symbiotic nitrogen fixation (SNF). N, together with carbon (C), oxygen (O) and hydrogen (H) is a major elemental component, essential for proper cellular functioning and its adequate supply is crucial for plant growth and development (Zheng 2009). Plants can obtain unlimited C by photosynthesis, O and H from water or air while sources of N are limited. Physiological and biochemical studies showed that N deficiency has negative impact on photosynthetic output of plants, which could be recovered by added N (Coruzzi & Zhou 2001; Coruzzi & Bush 2001). Similarly, increasing C supply promoted N uptake and assimilation. Plants take up N in the form of nitrate ( $\text{NO}_3^-$ ) and ammonium ( $\text{NH}_4^+$ ) from soil and/or added fertilizers. Since the weathering of rocks releases these ions slowly, it has a negligible effect on the availability of combined N in soils. In agriculture, the N demand is compensated by the use of nitrogen fertilizers, which increases greenhouse effect, nitrate pollution of waters and represent significant expense for growth of crop plants and thus food production. To overcome the problem of N starvation, legumes can access  $\text{N}_2$  through a beneficial symbiosis with  $\alpha$ - and  $\beta$ -proteobacteria collectively referred as rhizobium or rhizobia (Chen et al. 2003; MacLean et al. 2007) which can convert  $\text{N}_2$  into ammonia. Therefore, rhizobia-legume symbiosis is a major source of fixed N for plant growth. Worldwide, SNF contributes approximately 40 million tonnes of N each year (Herridge et al. 2008). Improved understanding of the rhizobia-legume symbiosis has implications for sustainable agriculture and the ecosystem. Since the identification of rhizobia as the source of fixed N (Hellriegel & Wilfarth 1888) in legume root nodules, scientists have wondered whether plants outside the *Fabaceae* could be manipulated to associate with rhizobia for improvement and future extension of the eco-friendly biological nitrogen fixation. As the human population grows and the need for agricultural production increases, the engineering of non-fixing food crops such as



cereals to associate with suitable microbes to access atmospheric nitrogen is becoming more and more important.

Legume plants, besides their contribution to SNF, are for humans and animals a significant source of protein, dietary fibre, carbohydrates and minerals. Grain legumes are cultivated for their seeds, which are important for human and animal consumption. The major commercial species are soybean (*Glycine max*), pea (*Pisum sativum*), peanut (*Arachis hypogaea*), chickpea (*Cicer arietinum*), faba bean (*Vicia faba*) and lentil (*Lens culinaris*). Moreover, legumes such as lucerne (*Medicago sativa*) are used as livestock forage, silage, and as soil-enhancing green manure.

The legume family is extremely diverse and found in most terrestrial environments. The *Leguminosae* are divided in three subfamilies: *Caesalpinioideae*, *Mimosoideae*, and *Papilionoideae*. The *Papilionoideae* is the largest of the three subfamilies containing economically important crop species and the major model legumes like *Medicago truncatula* and *Lotus japonicus* (**Fig. 1**) (Doyle & Luckow 2003).



**Figure 1. Schematic phylogenetic tree of the legume family** (modified from Gepts et al. 2005). The three subfamilies (*Caesalpinioideae*, *Mimosoideae*, and *Papilionoideae*) and the major subclades are shown in bold and their positions are indicated by black circles.

The four important clades within *Papilionoideae* are the genistoid clade with the genus *Lupinus*, the aeschynomenoid/dalbergioid clade represented by peanut (*Arachis hypogaea*) and *Aeschynomene* species and two other clades that share a common ancestor: the phaseoloid/millettioid clade and the Hologalegina clade. The phaseoloid/millettioid clade is composed of tropical and subtropical legumes such as common bean (*Phaseolus vulgaris*), cowpea (*Vigna unguiculata*), pigeon pea (*Cajanus cajan*), and soybean (*Glycine max*). The Hologalegina clade includes most of the temperate herbaceous species including clover (*Trifolium*), faba bean (*Vicia faba*), pea (*Pisum sativum*), lentil (*Lens culinaris*) and the most important model legumes like *Lotus japonicus*, *Medicago truncatula* and alfalfa (*Medicago sativa*).

### **1.1.2. Forms of mutual dependence: Nitrogen-fixing symbiotic associations**

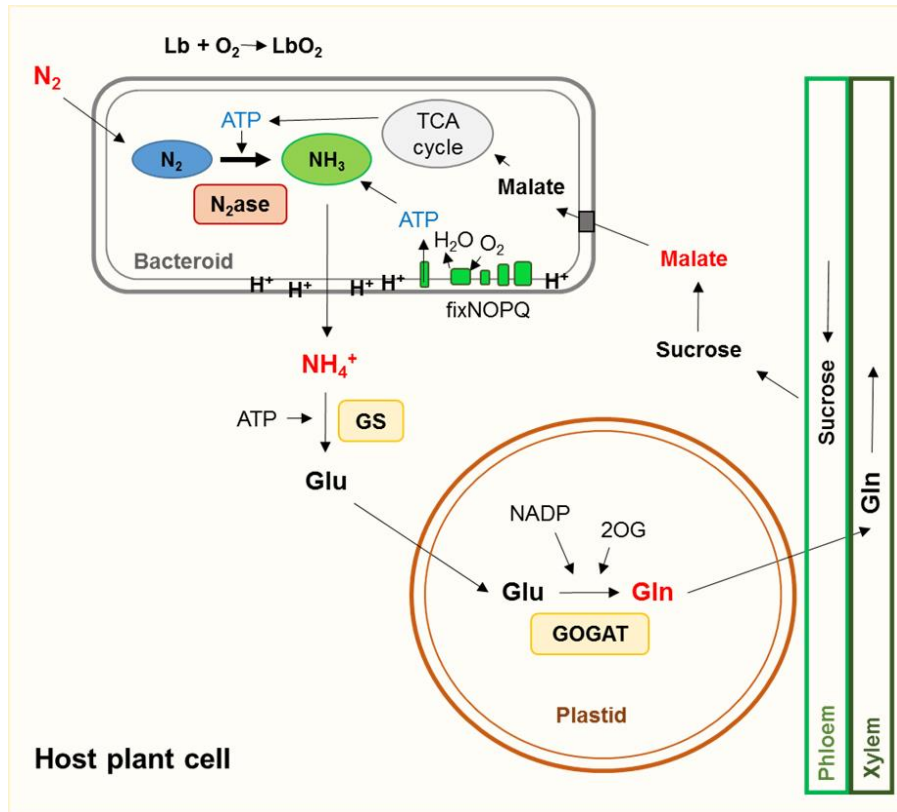
The driving force of symbiosis is the N limitation in the soil for plant growth and later the nutrient exchange between the host plant and the microsymbiont. The plant in return for the fixed N provides the symbiont carbon compounds as energy source to sustain metabolism including N fixation (Mus et al. 2016; Lodwig et al. 2003). Both organisms change considerably their metabolism during symbiosis (Lodwig et al. 2003): the plant provides amino acids to the symbiont enabling the shutdown of its ammonium assimilation and therefore the release of N, while the nitrogen-fixing bacterium (bacteroid) acts like plant organelle and provides the host with ammonium. Reduction of N<sub>2</sub> to ammonia (NH<sub>3</sub>) using the nitrogenase enzyme complex can be performed only by some prokaryotes. Breaking the stable triple bond of N<sub>2</sub> is a highly energy demanding, oxygen-sensitive process (Mylona et al. 1995). There are free-living nitrogen-fixing bacteria (e.g. *Azotobacter*, purple- and green sulphur bacteria) and SNF organisms such as *Rhizobium*. Leguminous plants can form symbiosis with  $\alpha$ -proteobacteria including the genera *Rhizobium*, *Ensifer* (*Sinorhizobium*) *Mesorhizobium*, *Bradyrhizobium*, and *Azorhizobium*, as well as members of the  $\beta$ -proteobacteria genera *Burkholderia*, *Cupriavidus* and *Herbaspirillum* (Velázquez et al. 2010). Under N limiting conditions, rhizobia infect the legume roots, induce the formation of root nodules, which later host and feed the bacterial symbionts.

The role of the nodule is to provide microaerobic conditions for nitrogen fixation. The plant accomplishes this by the production of a heme-containing protein called leghemoglobin (Lb), which gives the pink color of the functioning nodule. Lb is an O<sub>2</sub>-scavenger that binds oxygen and transfers it to the bacterial electron transport chain. In addition, a bacteroid terminal

oxidase with a high affinity for O<sub>2</sub> (i.e., fixNOPQ) also contributes to low free-oxygen concentration in the host cell for proper nitrogenase functioning (Delgado et al. 1998). The NH<sub>3</sub> produced by nitrogenase can diffuse through the bacteroid membranes and enter the plant cytoplasm. During diffusion through the symbiosome membranes (acidic intramembrane/peribacteroid space) NH<sub>3</sub> becomes protonated and assimilated by the plant into amino acids, via the action of the glutamine synthetase-glutamate synthase (GS-GOGAT) pathway (Gene et al. 1981; Cordoba et al. 2003). Nodules export the assimilated ammonia in the form of glutamine, which is metabolised in leaves (Mus et al. 2016). Some legume nodules produce and export ureides (allantoin and allantoic acid) rather than glutamate (Tajima et al. 2004). The plant provides carbon to the bacteroids as dicarboxylic acids, such as malate, which is catabolised through the tricarboxylic acid (TCA) cycle to produce ATP for nitrogen fixation (Finan et al. 1988) (**Fig. 2**). Moreover, the high-energy demand of the nitrogen fixation is also supported by ATP synthesis during oxidative phosphorylation. The nitrogenase enzyme complex contains two metalloproteins, the iron (Fe) protein and the molybdenum-iron (MoFe) protein. While these proteins are encoded by the bacterium, in some legumes the homocitrate component of the iron-molybdenum cofactor derives from the host cells. Interestingly, the homocitrate synthase gene is absent from most rhizobial species, indicating the indispensable relationship and molecular complementation between rhizobia and the host plant (Hakoyama et al. 2009). Some rhizobia however, particularly those that can fix N also in free-living state can produce homocitrate and therefore do not rely on host-supply (Nouwen et al. 2017).

*Rhizobium* strains can nodulate only particular legumes and not all legumes are able to form nitrogen-fixing symbiosis with a given rhizobium strain. The specificity of symbiotic interactions depends on the repertoire of bacterial and plant signalling molecules. For example, *Mesorhizobium muleiense* CCBAU 83963T only forms an association with *Cicer arietinum* (Zhang et al. 2012) while other strains, such as *Rhizobium fredii* NGR234 or *Rhizobium fredii* USDA257 (Pueppke & Broughton 1999) have a broad host range and can nodulate plants from the three distinct subfamilies of *Leguminosae* (136 and 66 species of legumes, respectively). Similarly to rhizobia, some legumes, like *Astragalus sinicus* (Wang et al. 2014) and *Medicago truncatula* (Barker et al. 1990) are specific for certain *Rhizobium* species or strains. Others, such as *Phaseolus vulgaris* (common bean) (Michiels et al. 1998), *Sophora flavescens* (Jiao et al. 2015) and *Vigna unguiculata* (Guimaraes et al. 2012) has been described as promiscuous hosts. In the majority of legumes, host specificity and nodule initiation depend on host flavonoids and *Rhizobium* Nod factor production (Perret et al. 2000). Moreover, other molecular signals like extracellular polysaccharides needed for nodule invasion as well as

proteins secreted directly into the host cytoplasm by type III secretion systems (T3SS) are important for the nodulation process. However, even within compatible interactions, both the genotype of the host plant and the bacterial symbiont will influence the efficiency of the nodule functioning and the nitrogen-fixation activity (Sugawara et al. 2013; Kazmierczak et al. 2017).



**Figure 2. Schematic representation of metabolite exchange between the bacterial symbiont and the host plant cell during nitrogen-fixing symbiosis.** In the nodules, bacteria differentiate into bacteroids and reduce diffused nitrogen gas ( $N_2$ ) into ammonia ( $NH_3$ ) by the nitrogenase enzyme complex ( $N_2ase$ ). The microaerobic condition required for nitrogenase functioning is provided by the nodule structure and by leghemoglobin (Lb), which binds and transports oxygen ( $O_2$ ).  $NH_3$  produced by nitrogenase becomes protonated ( $NH_4^+$ ) when it diffuses through the bacteroid membranes. In the plant cytoplasm,  $NH_4^+$  is converted to glutamine (Glu) by glutamine synthetase (GS) and into glutamate (Gln) by glutamate synthase (GOGAT) using 2-oxoglutarate (2OG) in the plastid. The produced Gln is transported to the leaf. The plant provides malate to the bacteroids, which in the tricarboxylic acid (TCA) cycle produce ATP for nitrogen fixation. Nitrogen-fixing bacteroids use a high-affinity symbiosis specific terminal oxidase encoded by the *fixNOPQ* operon that transfer electrons to  $O_2$ . Oxidative phosphorylation leads to ATP synthesis in the bacteroid membrane and provides energy for nitrogenase functioning.

SNF is best studied within the legume family, but some other plant groups can also form nitrogen-fixing symbioses. The only known non-legume plant genus forming nitrogen-fixing nodules upon infection with rhizobia is *Parasponia* (*Ulmaceae*) (Trinick 1979; Behm et al., 2014). *Frankia*, filamentous bacteria that live in symbiosis with actinorhizal plants, similarly

to legumes able to induce formation of root nodules (Callaham et al. 1978). In some species of *Gunnera*, a symbiotic cyanobacterium (*Nostoc punctiforme*) resides in stem glands and fix nitrogen. This association is usually not considered as a true form of bacterial endosymbiosis since no novel organ is formed (Gualtieri & Bisseling 2000). The common feature of all three symbiotic systems is the intracellular nitrogen fixation by the prokaryote, and the host-derived membrane, which separates the microsymbiont from the plant cytoplasm.

Thus, nitrogen-fixing symbiosis has evolved in several lineages during evolution (Kistner & Parniske 2002) and nodulation has arisen several times independently (Doyle & Luckow 2003).

### **1.1.3. Rhizobial infection and nodule development**

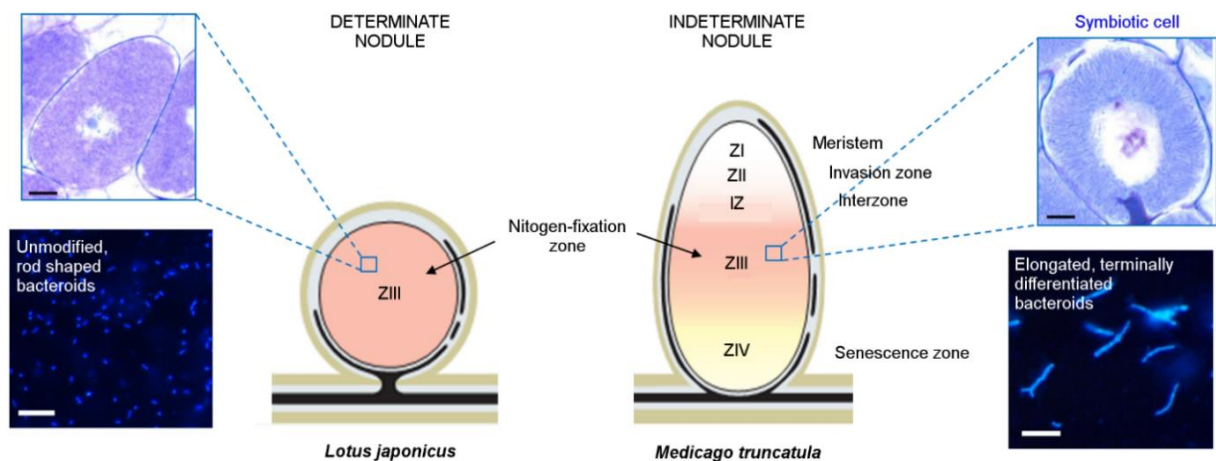
Even if the symbiotic interaction is beneficial for both partners, it follows rules that are strictly dictated by the plant. Because billions of microorganisms are present in the soil and particularly around the root system, legumes and rhizobia must recognize each other. The signal exchange is important for the plant to distinguish between compatible symbionts, non-symbionts and pathogens and for rhizobia as well to distinguish their hosts from other legumes. The plant initiates symbiosis by releasing flavonoids and isoflavonoids that are chemoattractants for the microsymbiont, but more importantly, they bind the transcriptional regulator NodD and activate expression of rhizobial nodulation genes in their symbiotic partner (Kondorosi et al. 1986). There have been common nodulation genes (*nodABC* and *D*) identified that are conserved in all rhizobia, both functionally and at the DNA level (Kondorosi et al. 1985, Mulligan & Long 1985) while other sets of genes (*hnsABC* and *D*) determine the host-specificity of nodulation (Horvath et al. 1986). The activation of *nod* genes leads to the synthesis of Nod factors (NFs) by rhizobia that are lipo-chitooligosaccharides (LCOs). The strain-specific chemical structures (length of the oligosaccharide backbone, modifications, size and saturation of the fatty acid chain) of the NFs are the primary determinants of host specificity (Oldroyd & Downie 2008; Denarie et al. 1996; Schultze & Kondorosi 1998). Additional bacterial components such as exopolysaccharides, type III and type IV secretion systems are also required for an effective infection (Perret et al., 2000; Saeki, 2011). The specific structure of LCOs is important for recognition by NF receptors (NFR) which are LysM domain receptor kinases (Radutoiu et al. 2007). Moreover, NFs also play important role to suppress plant immunity and thus permit rhizobial invasion and persistence in the host cell (Liang et al. 2013, Gourion et al. 2015). Upon recognition of NFs by NFRs in the compatible legume host initiates

a complex signalling pathway leading to nodule formation. The process begins with a series of signaling events triggered in the root cells, including nuclear calcium-oscillations (calcium spiking), followed by curling of the root hairs around the invading rhizobia (Esseling et al. 2003). Concomitantly, cell divisions are activated in the root cortex underlying the activated root hairs, leading to the formation of a nodule primordium (Downie & Walker 1999). Rhizobia enter the plant cells through infection threads (ITs), which are initiated in root hairs and grow onward into the root cortex transporting the bacteria to the nodule primordium. When the IT reaches the nodule primordium, rhizobia enter into the plant cells (Oldroyd & Downie 2008) and form organelle-like structures called symbiosomes. The symbiosome contains single or multiple bacterial cells, depending on the host plant, surrounded by a plant-derived peribacteroid membrane and the peribacteroid space. The rhizobia in the symbiosomes differentiate into nitrogen-fixing bacteroids, which are physiologically distinct from the free-living bacteria form. In some legumes, the formation of bacteroids is accompanied with a plant-guided morphological differentiation process, such as genome amplification and membrane alteration (see Section 1.2.2). In parallel with the formation of the bacteroids, also the host cells differentiate into a distinct cell type, called the symbiotic nodule cell, highly adapted to the symbiotic process, housing large quantities of rhizobia and assimilating  $\text{NH}_4^+$ .

There are two morphological nodule types: the spherical, determinate and the elongated indeterminate nodule types. The type of the nodule that will be formed is dependent on the host plant (Kondorosi et al. 2013). The main difference between them is the origin and the activity of their meristem (**Fig. 3**). Determinate nodules lose their meristem at an early stage of development. The differentiation of the infected cells is synchronous, therefore there is no developmental gradient and the nodule shape is spherical. The mature nodules contain a homogenous central tissue composed of infected cells fully packed with nitrogen-fixing bacteroids and also some uninfected cells (Szczyglowski et al. 1998). The symbiotic cells of *Lotus japonicus* are filled with rod shaped bacteroids, which are able to multiply within the symbiosome and when released back into the soil after nodule senescence, they can revert to the free-living lifestyle. Legumes that form determinate nodules are mainly tropical and subtropical species, such as soybean (*Glycine max*), bean (*Phaseolus vulgaris*), and *Lotus japonicus* (**Fig. 3, left**). In contrast to determinate nodules, indeterminate nodules contain a persistent meristem at the nodule apex that originates from inner cortical root cells. The meristem produces continuously new cells that enter the nodule differentiation program. Therefore, the form of the nodule is elongated or cylindrical. Indeterminate nodules are

composed of different zones: the meristem (ZI), the invasion zone (ZII), the interzone (IZ), the nitrogen-fixing zone (ZIII), and in older nodules the senescence zone (ZIV) (**Fig. 3, right**).

Post-meristematic cells leaving ZI enter endoreduplication cycles in ZII (Vinardell et al. 2003) and become infected and filled with rhizobia. In ZIII the differentiation of both the plant and the bacterial cells is terminated, the symbiotic cells have reached their final size and are fully packed with differentiated bacteroids. ZIII is interspersed with uninfected cells (Xiao et al. 2014). In *M. truncatula* bacteroids inside the symbiotic cells are elongated and/or branched and they are terminally differentiated because they have lost their reproductive capacity.



**Figure 3. Differences between determinate and indeterminate nodules of two model legumes** (modified from Kondorosi et al. 2013). (**Left**) In determinate nodules of *Lotus japonicus*, meristematic activity is lost in mature nodules and they become spherical. The central zone III region contains a mixture of uninfected cells and nitrogen-fixing symbiotic cells. Enlargement of the symbiotic cells (blue brackets) show the presence of unmodified, rod shaped bacteroids in the host cytoplasm. (**Right**) Indeterminate nodules of *Medicago truncatula* have an apical meristem (ZI), below a differentiation zone (ZII) where symbiotic cells are formed; zone II-III is an interzone (IZ) and finally a nitrogen-fixation zone (ZIII) with mature symbiotic cells and, in older nodules, a senescent zone (ZIV). The zone III is interspersed with uninfected, specialized cells.

Temperate legumes including alfalfa (*Medicago sativa*), clover (*Trifolium repens*), pea (*Pisum sativum*) and *M. truncatula* form indeterminate nodules. However, it should be noted that the nodule type not necessarily reflects the bacteroid morphotype, since morphological differentiation of the endosymbiont is also seen in determinate nodules, or inversely, reversibly differentiated bacteroids can be found in the indeterminate nodules of some legume species, indicating that bacteroid fate is controlled by the plant (Czernic et al. 2015; Kondorosi et al. 2013) (see Section 1.2.2).

Besides these two main nodule types, lupinoid nodules represent a unique subclass of indeterminate nodules (Lotocka et al. 2000; Gonzalez-Sama et al. 2006). Lupinoid nodules are

formed by legumes of the genus *Lupinus* interacting with *Bradyrhizobium* sp. They are considered indeterminate because of their meristem, which can cause the nodule to grow until it surrounds the root. This nodule type does not show zonation and is spherical like determinate nodules. The central tissue only contains infected symbiotic cells and no uninfected cells (Lotocka et al. 2000). Rhizobium infection occurs in the outer cortex, and the meristematic region contains bacteroids. There are no ITs, rhizobia enter at the junction of the root hair base and the symbiotic cells originate from a single infected cortical cell (Gonzalez-Sama et al. 2006).

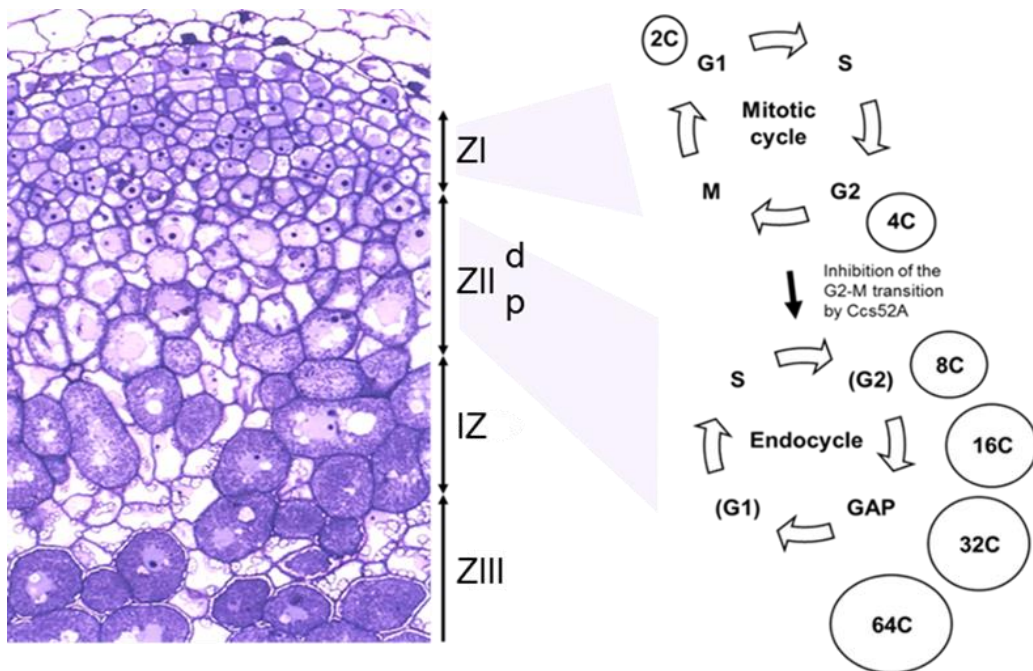
#### **1.1.4. Polyploidisation of symbiotic cells is a prerequisite for nodule organogenesis**

During nodule formation, both the symbiotic plant cell and in certain legumes also the endosymbiont (see Section 1.2.2) undergo drastic cellular modifications, involving cell enlargement and polyploidisation of the genome through endoreduplication (ENR) cycles. Polyploid eukaryotic cells contain multiples of the typical diploid genome and are found in a wide variety of cell types and organs in insects, animals, and plants. Functional specialization of plant cells is generally coupled to cell enlargement and ENR, like in the case of elongation of hypocotyl cells, branching of trichomes, nutrient storage in the endosperm of seeds and pericarp of fruits, nematode-infected giant cells in galls and in symbiotic nodule cells hosting endosymbiotic bacteria (Wildermuth 2010; Sabelli & Larkins 2009; Chevalier et al. 2011). Forming of large polyploid symbiotic cells is essential in both determinate and indeterminate nodule organogenesis of all legumes that have been tested (Cebolla et al. 1999; Vinardell et al. 2003; Gonzalez-Sama et al. 2006).

The diploid *M. truncatula* exhibits various degrees of endoploidy in different organs with the highest level in root nodules (Cebolla et al. 1999). The tissue-specific pattern of endopolyploidy in different organs indicates that ENR constitutes an integral part of the developmental programs. During symbiosis, the host plant has the ability to reset and alter the fate of the differentiated root cortical cells to form a new root organ. Nodule organogenesis is induced by rhizobia-derived NFs, which mitotically reactivate cortical cells to proliferate and form the nodule primordium. The nodule grows by the activity of the meristem that produces continuously new cells, which enter the nodule differentiation programme by cell division arrest, followed by several rounds of ENR cycles. ENR arises from the transition of a mitotic cell cycle into endocycles, where the genome doubles in each cycle resulting from the diploid



2C to 4C, 8C, 16C, 32C, 64C genomes (C is the haploid DNA content). In the classical mitotic cell cycle, the DNA duplicates during the S (synthesis) phase and subsequently divides equally in the mitosis (M) phase and thus each of the two daughter cells has a 2C DNA content. In contrast, there is no M phase during the ENR cycle and repeated S- and gap (G) phases result in consecutive doublings of the genome (**Fig. 4**).



**Figure 4.** In *Medicago truncatula*, the symbiotic cell undergoes several rounds of endoreduplication during differentiation. *M. truncatula* nodule zones (ZI-ZIII) and the genome copy number (C) of nodule cells: Cells in meristem (ZI) are in a mitotic cell cycle, they have 4C DNA content after the S phase and 2C after division. In contrast, cells in ZII do not divide and by endocycles, they become polyploid. ZII<sub>d</sub>: distal, younger 4C/8C symbiotic cells, ZII<sub>p</sub>: proximal, older symbiotic cells 8C/16C. The switch from mitotic to endocycle is controlled by the Ccs52A protein that provokes degradation of mitotic cyclins and thereby inhibits M phase and cell division (Cebolla et al. 1999; Vinardell et al. 2003). Mature symbiotic cells have a 32C or 64C DNA content. ZIII also contains uninfected, low ploidy (4C/8C) level cells.

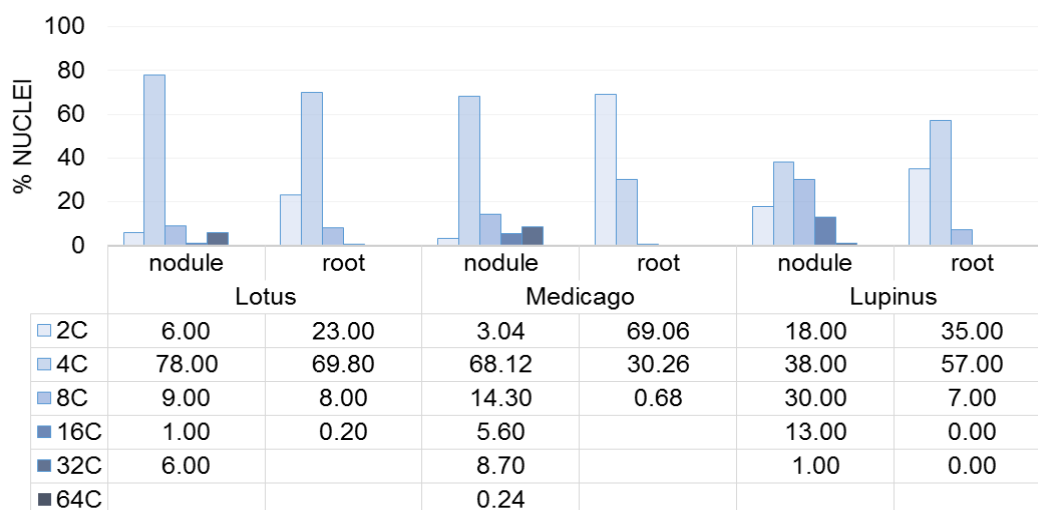
Two cell cycle switch genes - *Ccs52A* and *Ccs52B* - have been identified in *M. truncatula* and *A. thaliana* encoding for the anaphase-promoting complex (APC) activator (Tarayre et al. 2004; Cebolla et al. 1999; Vinardell et al. 2003). *Ccs52A* is a major regulator of symbiotic cell differentiation in the indeterminate nodule of *M. truncatula*, whereas *Ccs52B* is expressed in shoot apices but not in the symbiotic organ (Vinardell et al. 2003). *Ccs52A* promotes the exit of meristematic cells from the cell division cycle by inactivating mitotic cyclin dependent kinase (M-CDK) through ubiquitin-dependent degradation of cyclin B, which leads to G2-M-phase inhibition and switch to ENR cycle (**Fig. 4**). *Ccs52A* antisense nodules,

in which ENR is blocked, were arrested in the nodule primordium phase and underwent early senescence (Vinardell et al. 2003). The *Ccs52A* is highly conserved in the plant kingdom and its expression in determinate and indeterminate nodules indicate important role in cell enlargement and polyploidisation in different nodule types (Cebolla et al. 1999; Gonzalez-Sama et al. 2006). Phylogenetic analysis of the *Ccs52A* proteins suggests that the *Lupinus* *Ccs52A* and *Lotus* *Ccs52A* proteins are orthologues of the *Medicago* *Ccs52A* protein and that they are involved in control of ENR in these legumes as well (Gonzalez-Sama et al. 2006). In *M. truncatula*, *Ccs52A* expression is detected during nodule initiation, in the central region of the nodule primordium before cell differentiation starts, and is maintained in the ZI and ZII of mature nodules (Cebolla et al. 1999; Vinardell et al. 2003). In contrast, in *Lupinus albus*, *Ccs52A* expression is detected in proliferating root cortical cells; and maximum expression is reached in the nodule primordia. In the fully differentiated nitrogen-fixing symbiotic cells of either nodule types, no expression of *Ccs52A* was detected. In *Lotus*, its expression was maximal in young nodules, but transcripts were present in mature nitrogen-fixing nodules as well (Gonzalez-Sama et al. 2006).

Recently, a *vagrant infection thread 1 (VAG1)* gene was identified in *Lotus* that plays an important role in the onset of ENR in cortical cells during early nodule development. *VAG1* encodes for a component of plant DNA topoisomerase VI complex that is involved in the control of ENR (Suzaki et al. 2014). The *vag1* mutant plant could form nodules, but the proportion of endoreduplicated cells (>4C) and the number of rhizobium-colonized infected cells were reduced compared to wild type. The loss of ENR in *vag1* mutant caused misguided elongation of IT and failure in ploidy-dependent cell growth of rhizobial-infected cells (Suzaki et al. 2014). Thus, ENR of cortical cells have important role in guiding ITs and therefore rhizobia to host cells.

By measuring the DNA content of root and nodule tissues in *L. japonicus*, *M. truncatula* and *L. albus*, the highest endoploidy was detected in nodules, although there were differences among the three legumes (**Fig. 5**). Root tissue and nodule primordia contained cells with 2C/4C DNA content corresponding to G1 and G2 cell cycle phases, and a low proportion of 8C cells. Young nodules contained cells with increased 16C ploidy, and in mature nodules 32C or even 64C cells have been detected (Cebolla et al. 1999; Vinardell et al. 2003; Gonzalez-Sama et al. 2006). *Medicago* and *Lotus* have a higher percent of 32C nuclei than lupin nodules, but lupin contained a high proportion of intermediate ploidy level 8C/16C cells. These differences likely indicate different nodule developmental programs. In *Medicago* and *Lotus*, the endoreduplicated and enlarged cells are invaded by rhizobia, while diploid cells that do not

enter the endocycles remain small and uninfected. In contrast, during lupin nodule development, dividing cortical cells/primordium cells (4C) are infected then ENR cycles and symbiotic cell differentiation takes place. Altogether, high endoploidy and formation of huge symbiotic cells are common features of nitrogen-fixing root nodules.



**Figure 5. Ploidy levels in *Lotus japonicus*, *Medicago truncatula*, and *Lupinus albus* plants.** Fraction (%) of nuclei with different DNA content (from 2C up to 64C) in nodule and root measured by flow cytometry (Cebolla et al. 1999; Vinardell et al. 2003; Gonzalez-Sama et al. 2006).

In *M. truncatula*, polyploid symbiotic nodule cells change dramatically their gene expression programs involving the activation of a large number of symbiosis-specific genes and the repression of others (Maunoury et al. 2010). This suggests that ENR of the symbiotic nodule cells may serve as a mechanism in controlling gene expression, since the polyploid genome content of symbiotic cells appears to be a prerequisite for proper nodule organogenesis and for the expression of most symbiotic plant genes (Maunoury et al. 2010). Duplication of the genome without mitosis is believed to result in multiplication of gene copy number and decondensation of the chromatin leading to higher transcriptional activity of the cell, although this has never been convincingly demonstrated or disproven.

## 1.2. The Nodule-specific Cysteine-Rich (NCR) peptide family

### 1.2.1. NCR genes are specific for the IRLC and Dalbergioid clade legumes

Parallel to the host cell maturation, the endosymbiotic partner also undergoes differentiation to adapt to the symbiotic lifestyle. In some legume clades, the differentiation process of the nitrogen-fixing bacteroids is accompanied by morphological and cytological

metamorphosis, including cell enlargement, different cell shape, loss of cell division capacity, genome amplification by ENR cycles, and alteration of the membrane. Such changes of the bacteroids are irreversible, *i.e.* the bacteroids are unable to revert to a free-living lifestyle (Mergaert et al. 2006). In *M. truncatula*, a large family of peptides named NCRs (Nodule-specific Cysteine-Rich peptides) control the bacteroid differentiation (Van de Velde et al. 2010). The NCR peptides are produced by the infected symbiotic cells and are transported to the bacteroid-containing symbiosomes. More than 600 *NCR* genes were identified in *Medicago* encoding for small (60-70 amino acids) secreted symbiotic peptides (Mergaert et al. 2003; Alunni et al. 2007). Despite the fact that the mature NCR peptides are highly divergent, they share several common features: (i) *NCRs* are usually composed of two exons (400-700 bp) separated by ~70-100 bp intron sequence; (ii) the first exon codes for a conserved signal peptide (that directs the NCRs to the secretory pathway) and the second one for the mature cysteine-rich peptides (Mergaert et al. 2003); (iii) the processed mature NCR peptides contain 4 or 6 conserved cysteines in conserved positions with well-defined spacing. *NCR* genes are scattered over the 8 *Medicago* chromosomes but they are found within genomic clusters from 7 up to 364 kb (Alunni et al. 2007). Genes of the same cluster are the most homologous to each other and they likely arose from gene duplication events. It was noticed that *NCR* genes are often in the vicinity of transposable elements suggesting that their multiplication in the genome during evolution arose by co-transfer with activated transposons (Satzg  et al. 2016).

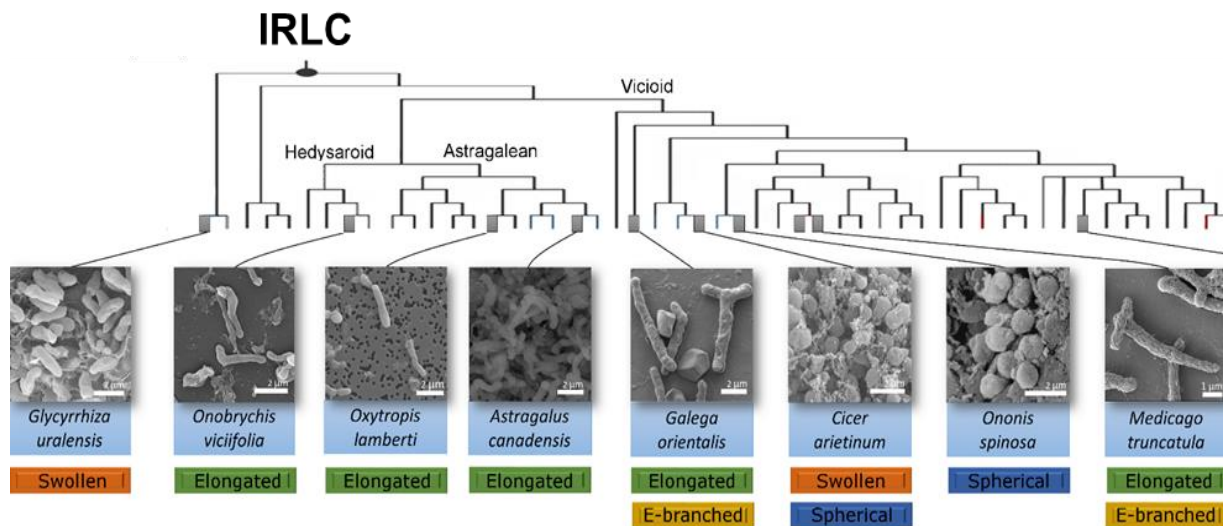
Members of the *NCR* gene family have been found in legumes of the Inverted Repeat-Lacking Clade (IRLC) (Wojciechowski et al. 2004) like *Medicago* spp. (Fedorova et al. 2002; Graham et al. 2004; Gyorgyey et al. 2000; Mergaert et al. 2003), *Pisum sativum* (Kardailsky et al. 1993; Kato et al. 2002; Scheres et al. 1990), *Trifolium repens* (Crockard et al. 2002), *Astragalus sinicus* (Chou et al. 2006), *Galega orientalis* (Kaijalainen et al. 2002), *Vicia faba* (Fruhling et al. 2000) *Cicer arietinum* (Kant et al. 2016), *Astragalus canadensis*, *Onobrychis viciifolia*, *Ononis spinosa*, *Oxytropis lamberti*, and *Glycyrrhiza uralensis* (Montiel et al. 2017). Recently, *NCR*-like genes were also identified in legumes such as *Aeschynomene afraspera*, *Aeschynomene evenia*, *Aeschynomene indica* and *Arachis hypogea* belonging to the more ancient Dalbergioid lineage (Czernic et al. 2015). However, it should be noted that the Dalbergioid *NCR*-like peptides form a distinct family, unrelated to the *NCR* family of the IRLC indicating an independent convergent evolutionary origin. *Aeschynomene* *NCR*-like peptides play a similar role in bacteroid differentiation to those described in IRLC species (Czernic et al. 2015). In legumes with reversible fate of bacteroids like *L. japonicus*, *G. max*, *P. vulgaris* and in any other plants, no homologues of *NCRs* have been identified.

NCRs are targeted to the bacteroids and ~150 of them were found in the bacteroids (Van de Velde et al. 2010; Durgo et al. 2015). Some of these peptides provoke symptoms of terminal differentiation by inhibiting cell division, inducing cell elongation and polyploidization of *Rhizobium* (Van de Velde et al. 2010) *e.g.* NCR247 can specifically block bacterial cell division and cause endoreduplication (Penterman et al. 2014). Some peptides accumulate in the cytosol of mature bacteroids, indicating that they may affect bacteroid metabolism as well (Van de Velde et al. 2010; Kereszt et al. 2011; Durgo et al. 2015; Farkas et al. 2014). The structure and gene organization of the NCR peptides are similar to defensin-type of antimicrobial peptides that are part of innate immunity in animals and plants (Maroti et al. 2011). Certain NCR peptides indeed exhibit *in vitro* antimicrobial activity and when applied at high concentration they can kill *Rhizobium* or other bacteria (Van de Velde et al. 2010; Tiricz et al. 2013). The *Sinorhizobium meliloti* *bacA* gene (bacteroid development factor A) was identified to be essential to counteract the antimicrobial activity of the NCR peptides during symbiosis in *M. truncatula*. In the absence of this protein, the bacteroids are immediately killed by the NCR peptides in the nodule cells (Haag et al. 2011). Although, the BacA protein belongs to the ABC transporter family, how it contributes to resistance against NCR peptides is unknown. In symbiosis with *Aeschynomene* legumes, the *Bradyrhizobium* BclA protein, a BacA-like peptide transporter is required for bacteroid differentiation likely in a similar way as the BacA protein of *S. meliloti* in symbiosis with *Medicago* (Guefrachi et al. 2015).

### **1.2.2. Bacteroid morphotypes: NCRs direct irreversible bacteroid differentiation**

The morphology and physiology of bacteroids display striking differences according to the host plant. Bacteroids in the IRLC legumes differ from free-living bacteria because they are adapted for an intracellular, in planta life-style with a specialized metabolic activity and capability for nitrogen fixation. Besides this physiological differentiation, which is common to bacteroids in all legumes, bacteroids can also display a striking morphological differentiation. Different bacteroid morphotypes were described in legumes: the U-morphotype bacteroids (unmodified) are rod-shaped, similar to free-living rhizobia, the S-morphotype bacteroids are swollen/spherical compared to free-living rhizobia and the elongated E-morphotype bacteroids are sometimes elongated Y-shaped, branched cells (Mergaert et al. 2006; Montiel et al. 2016; Czernic et al. 2015; Guefrachi et al. 2015; Oono et al. 2010; Montiel et al. 2017). U-morphotype bacteroids are able to revert to free-living lifestyle, but the differentiation of the other

morphotypes is irreversible and terminal. U-type bacteroids are typical in *Phaseolus*, *Glycine*, *Vigna*, *Lotus* and *Sesbania* species; whereas terminally differentiated bacteroids develop in legumes like *Medicago*, *Pisum*, *Vicia*, *Trifolium*, *Cicer*, *Galega*, *Ononis*, *Onobrychis*, *Astragalus*, *Glycyrrhiza* and *Aeschynomene* species (**Fig. 3, Fig. 6**).



**Figure 6. Bacteroid morphotypes** (modified from Montiel et al. 2017). Phylogenetic tree and morphotype of bacteroids in legumes from the IRLC. Phylogenetic tree modified from Wojciechowski et al. 2004 to highlight the studied species. Scanning electron microscopy images of S- and E-morphotype bacteroids isolated from nodules of the above indicated species.

Besides morphological changes, terminally differentiated bacteroids undergo drastic cytological modifications during differentiation, such as genome amplification by ENR (this is a remarkable similitude in the cytological organization of the bacteroids and their eukaryotic host cells described above), and alteration of their membranes which may contribute to their uncultivable nature. The DNA content can be from 4 to 24 fold higher in the differentiated bacteroids, depending on the host plant and the bacterial species (Mergaert et al. 2006; Montiel et al. 2016; Czernic et al. 2015; Guefrachi et al. 2015). *In vitro* increased membrane permeability of bacteroid membranes is obvious by the slow penetration of propidium iodide in the cytosol which is normally excluded from living cells (Mergaert et al. 2006). Although, morphology and the degree of cell enlargement can vary among terminally differentiated bacteroids, the genome amplification is a common feature.

The diverse morphology of bacteroids in determinate and indeterminate nodules indicates that the nodule type not necessarily reflects the bacteroid morphotype (**Fig. 6**). The distinct behaviour of the same *Rhizobium* strain in different legume hosts (Sen & Weaver 1980; Bonaldi et al. 2011; Mergaert et al. 2006) led to the conclusion that plant factors influence the bacteroid differentiation and determine the morphotype. Until now, different studies

demonstrated in the IRLC and Dalbergioid legumes forming E- and S-type bacteroids, that NCR peptides guide irreversible terminal differentiation of the endosymbionts (Van de Velde et al. 2010; Montiel et al. 2016; Czernic et al. 2015).

It is currently not known why some legumes induce morphological modifications of their microsymbiont. It is not a prerequisite for a functional symbiosis since many legumes do not produce NCR peptides and do not impose this particular differentiation. Phylogenetic analysis (Oono et al. 2010) shows that the unmodified U-morphotype is the ancestral bacteroid form and the E- and S-morphotype appeared several times independently during evolution, suggesting that the process provides a selective advantage. In agreement with this, some studies indicate superior nitrogen-fixation performance of terminally differentiated bacteroids compared to the reversible, U-morphotype bacteroids of the same strain (Sen & Weaver 1980; Oono et al. 2010). However, a shortcoming in these studies is that the nitrogen-fixation efficiency was compared in different host plants and therefore it is unclear what are the effects of bacteroid type and other host factors. At present, no firm experimental proof of a selective advantage for bacteroid differentiation has been delivered, even if this seems to be very plausible.

### **1.2.3. Possible mode of action and intracellular targets of the NCR peptides**

The NCR peptide family is very large (>600 genes) suggesting an extensive functional redundancy among them but their amino acid composition and their isoelectric point (ranging from 3,2 to 11,25, anionic to cationic, respectively) is very diverse, indicating that they could have multiple activities during different stages of bacteroid differentiation and functioning. This is supported by the spatio-temporal expression of the *NCR* genes, which are active in different nodule zones, overlapping with different stages of bacteroid development or functioning (see Section 1.2.4). The *ex planta* effects of certain synthetic cationic NCRs on cell elongation have been demonstrated on the free-living *S. meliloti* cultures. At high concentrations, NCRs disrupt bacterial cell envelope acting both on the outer and inner membranes of the bacteria, via forming pores and destroying the membrane potential (Nagy et al. 2015; Mikuláss et al. 2016), which leads to stress responses, including the slowdown of the cell metabolism and cell division arrest (Tiricz et al. 2013). The high concentrations of and/or longer exposure to cationic NCRs lead to cell death of a wide variety of Gram-negative and Gram-positive bacterial strains (Tiricz et al. 2013; Farkas et al. 2017) as well as unicellular and filamentous fungi (Ördögh et al. 2014). At low concentration the fluorescein isothiocyanate (FITC) labelled NCR247 peptide enter the

bacterial cytosol without membrane damage (Farkas et al. 2014). *In planta*, different cocktails of NCRs are produced and the neutral and anionic peptides could attenuate the *in vitro* killing effect of cationic NCRs and result loss of cell division capacity (Montiel et al. 2017). In *L. japonicus* nodules, where *NCR* genes are absent, expression of NCR035 induced bacteroid elongation, reminiscent of terminally differentiated bacteroids of *Medicago* (Van de Velde et al. 2010).

Although, the first and most general mode of the NCR action is the disruption of the microbial membranes, a growing body of evidence suggests that the peptides are taken up by the bacterial cells and they can have intracellular targets. NCR247 is the best characterized NCR peptide with strong *in vitro* antimicrobial activity (Tiricz et al. 2013). NCR247 has been shown to interact with a remarkably wide range of bacterial proteins like the GroEL chaperon, the cell division protein FtsZ, the ribosome, nitrogenase and energy metabolism proteins (Farkas et al. 2014). Moreover, NCR peptides affect the bacterial cell cycle, treatment of *S. meliloti* with NCR247 down-regulates the *S. meliloti gcrA* and *ctrA* cell cycle regulator genes (Penterman et al. 2014). CtrA may be a central target of the NCR activity because this protein is a principal regulator of the cell cycle in  $\alpha$ -proteobacteria, promoting cell division and inhibiting DNA replication. Indeed, it has been shown that in bacteroids *CtrA* expression is downregulated to undetectable levels and its repression in cultured cells results in polyploid, enlarged cells resembling bacteroid cells (Pini et al. 2013; Pini et al. 2015). In addition, treatment of NCR247 activated the expression of *rpoH1* and several RpoH1-regulated genes involved in cytoplasmic stress response (Penterman et al. 2014) suggesting mechanisms through which *S. meliloti* adapts to NCR peptide-mediated stress *in vitro* and *in planta*.

Genetic evidence have been provided that not all NCRs are redundant but that on the contrary, some peptides are essential and cannot be substituted by any of the other members of the family. NCR211 (Kim et al. 2015) and NCR169 (Horvath et al. 2015) have been shown to be essential for bacterial persistence *in planta*. In *dnf4* mutant plants lacking NCR211 and in *dnf7* mutants lacking NCR169 bacteroid differentiation is incomplete, there is no nitrogen fixation and early nodule senescence is induced (Horvath et al. 2015; Kim et al. 2015). Thus, NCR peptides besides provoking stress responses, also target intracellular regulatory pathways to drive *S. meliloti* genome amplification and differentiation during symbiosis. The high number and diversity of the NCR peptides might ensure that the symbiont cannot escape the differentiation process imposed by the plant.

In addition, two recent studies demonstrated rhizobia-specific effects of individual NCRs in *M. truncatula* (Yang et al. 2017; Wang et al. 2017). *NFS1* and *NFS2* encoding NCRs



control discrimination against incompatible microsymbionts by provoking bacterial cell death and early nodule senescence (Yang et al. 2017, Wang et al. 2017). The negative effect of the two NFS loci on symbiotic performance is proved to be strain-specific. The bacterial components that are responsible for the differential responses to the specific NCR peptides remain to be determined.

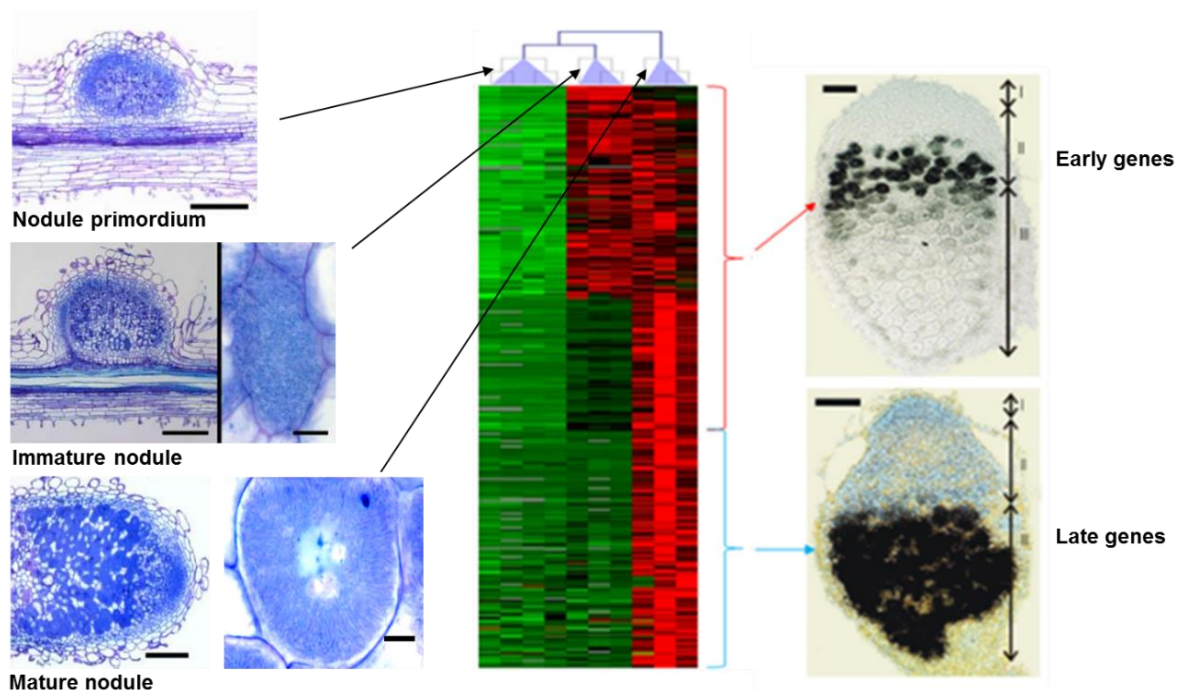
The opposing roles of the different NCRs may indicate an evolutionary mechanism for fine-tuning the symbiotic performance and could explain the high number and diversity of the NCR peptides in *M. truncatula*.

#### **1.2.4. NCRs are specifically induced during nodulation: Early and late genes**

Besides the size of the family, another surprising feature of the *NCR* family is that all *NCR* genes are specifically and only expressed in the nodules. The expression of *NCR* genes has been studied - mostly in *M. truncatula* but also in some other IRLC and Dalbergioid legumes - at the level of individual genes by RT-PCR, *in situ* hybridization, immunolocalization and promoter-reporter gene fusions (Mergaert et al. 2003; Alunni et al. 2007; Van de Velde et al. 2010; Nallu et al. 2013) or at the family level by EST-analysis, microarrays, dedicated microarrays or whole-genome microarrays (Mergaert et al. 2003; Graham et al. 2004; Maunoury et al. 2010; Moreau et al. 2011; Nallu et al. 2013; Tesfaye et al. 2013). For example, analysis of 120,000 expressed sequence tags (ESTs) in 26 different cDNA libraries which together covered most plant organs, different growth conditions including biotic and abiotic stresses, demonstrated that ESTs of the 311 analysed *NCR*s were only found in nodule libraries (Mergaert et al. 2003). In addition, this analysis revealed that all *NCR*s together constitute almost 5% of the total mRNA population in nodules. A more recent work assessing the expression of more than 500 *NCR* genes with dedicated Affymetrix chips confirmed the nodule specificity of the *NCR* gene family (Teskaye et al. 2013; Nallu et al. 2013). *In situ* detection of *NCR* expression by immunolocalization, *in situ* hybridization, or promoter-GUS/GFP/mCherry analysis has demonstrated for all the tested genes that they are expressed in the symbiotic nodule cells but different subsets of *NCR* genes are activated at different stages of differentiation of these host cells (Mergaert et al. 2003; Van de Velde et al. 2010; Nallu et al. 2013). Transcriptome analysis with microarrays or Affymetrix chips extended this pattern to the whole *NCR* gene family (see below, Chapter 1 of the results). The *Medicago truncatula* Gene Expression Atlas (MtGEA) (Benedito et al. 2008; He et al. 2009) is generated with the whole genome Affymetrix *Medicago* Gene Chip and compiles microarray data for the majority

of *M. truncatula* genes (50,900 probe sets) over a large set of experiments (254 different experimental conditions).

During nodule organogenesis, two major waves of transcriptional reprogramming have been described resulting in repression and massive induction of hundreds of genes (**Fig. 7**). The first transcriptional wave occurs along with the establishment of an incipient nodule containing differentiating plant cells, housing the dividing rhizobia, while a second wave takes place when rhizobia differentiate into nitrogen-fixing bacteroids (Maunoury et al. 2010). These two transcriptional waves were defined with the use of custom-made microarrays and the analysis of wild-type nodules at different age (primordia, immature, mature nodule) as well as a collection of bacterial and plant symbiotic mutants forming non-functional nodules.



**Figure 7. Transcriptome switch during nodule organogenesis.** (*Left panel*) Different stages of nodule development from early to late (nodule primordium, immature nodule and mature nodule). (*Middle panel*) A dramatic transcriptome switch revealed by transcriptome analysis and seen on the heat map accompanies the symbiotic cell differentiation (Maunoury et al. 2010). The expression profile in columns correspond to the developmental stage of the nodule. The transcriptome is activated in two consecutive waves. The cluster of genes activated during the first wave (early genes) in differentiating symbiotic cells in zone II are indicated with red bracket, while the second wave genes (late genes) are activated in mature symbiotic cells of interzone and zone III and are indicated with blue bracket. (*Right panel*) *M. truncatula* nodule section shows the tissue organisation of the nodule with at the top meristem (ZI), below the differentiation zone (ZII) where symbiotic cells are formed, the transition zone II-III (IZ) and zone III (ZIII) with mature symbiotic cells. Only the upper part of zone III is seen. Examples of *in situ* transcript detection of nodule-specific genes, which are expressed in young symbiotic cells (early genes) and in mature cells (late genes).

A more recent study, employing laser-capture microdissection of wild-type nodules coupled to RNA-seq analysis, was able to define more precisely the transcriptional changes that take place in differentiating nodule cells and could show the activation of nodule-specific genes in multiple successive spatio-temporal waves (Roux et al. 2014; Maunoury et al. 2010).

Among the activated genes, the *NCR* family with hundreds of members takes the most prominent place, which are all exclusively expressed in the symbiotic nodule cells. Their activation requires intracellular presence of rhizobia (Mergaert et al. 2003; Maunoury et al. 2010; Van de Velde et al. 2010). Along the maturation of symbiotic cells, different subsets, tens or hundreds of *NCR* genes are expressed during the early, the middle and later stages. Thus, the *NCR* genes are under an extremely tight transcriptional control.

### **1.2.5. Transcriptional regulation of *NCRs*: Epigenetic mechanisms might be involved**

How the nodule specific spatio-temporal regulation of the *NCRs* in the symbiotic nodule cells is achieved is currently not known. *NCR* expression correlates with bacterial infection of the host cells, which suggests the involvement of components of the bacterial envelope in their activation. Several transcription factors (TFs) have been reported to be specifically upregulated in nodules (El Yahyaoui 2004). For example, the ethylene response factor group (ERF) transcription factor that is required for the formation of functional nitrogen-fixing nodules may control, directly or indirectly, the expression of the *NCR* genes. In *erf1* mutant nodules many *NCRs* are downregulated and bacteroid differentiation is impaired (Vernie et al. 2008). Another TF is the IPD3/CYCLOPS (Singh et al. 2014; Limpens & Bisseling 2014) which is co-expressed with the *NCRs* (Messinese et al. 2007). In the *ipd3* mutant nodules, symbiotic cells are not formed and the symbiotic cell-specific genes, including the *NCRs*, are not activated (Maunoury et al. 2010). Potential *cis*-elements in the promoters of *NCR* genes were identified. The five different conserved motifs of 41 to 50 bp are specifically enriched in the 1000 bp promoter regions of *NCRs* (Nallu et al. 2013). Some of these motifs show resemblance to previously described motifs conferring nodule-specific gene expression. However, these data are fragmentary and the transcription factors that bind these elements are unknown. Interestingly, some of these motifs comprise Auxin Response Factor binding sites that may suggest a role for auxin in *NCR* regulation (Nallu et al. 2013).

The very tight regulation of the *NCR* genes indicates that besides *cis*- and *trans*-acting factors, regulation at the level of chromatin might also be involved in their activation. Genes

with high tissue-specific expression can be actively silenced during most of the plant growth. Gene silencing can be achieved by different epigenetic mechanisms such as small RNAs, DNA methylation but also by chromatin compaction mediated by trimethylation of histone H3 lysine 27 (H3K27me3) and Polycomb group (PcG) protein complexes. The latter mechanism seems to be particularly important for tissue-specific genes like the *NCRs*. Hundreds of genes have been shown to gain or lose H3K27me3 upon leaf differentiation in *Arabidopsis* (Lafos et al. 2011). H3K27me3 is correlated with gene repression, and its removal has been shown to be required for the proper expression of developmentally important genes in *Arabidopsis* (Zhang et al. 2007). Moreover, ENR seems to be a prerequisite for the activation of the *NCR* genes (Maunoury et al. 2010). It is unknown how successive ENR cycles and the polyploid state of the genome affect transcriptional reprogramming. The strict nodule specificity and the possible linkage between gene expression and ploidy levels of symbiotic cells suggested that epigenetic modifications along the ENR cycles might play a critical role in transcriptional regulation of *NCRs* and other nodule-specific genes.

The nodule-specific expression of retrotransposons, which are usually epigenetically silenced (Guefrachi et al. 2014; Satgé et al. 2016; see Section 4.1.7), the DNA demethylase gene DEMETER (Satgé et al. 2016) and the Dicer 1-like ribonuclease III gene, which encode both enzymes that may have a role in epigenetic regulation, as well as the identification of small RNAs potentially targeting *NCR* genes (Lelandais-Briere et al. 2009) are all in agreement with such an epigenetic control of the symbiotic cell-specific genes. Interestingly, the fungal plant pathogen *Leptosphaeria maculans* produces an arsenal of small secreted proteins, often cysteine-rich peptides acting as effector proteins. The genes encoding these effectors are up-regulated upon plant infection, requiring the removal of chromatin-mediated repression (Soyer et al. 2014). If the hypothesis related to the regulation of *NCR* and other nodule-specific genes is true, then the epigenetic chromatin marks for silencing must be reversed during the differentiation of the symbiotic cells in which ENR could be involved.

### **1.3. Epigenetic control of gene expression**

Epigenetics define mitotically or meiotically heritable modifications which change gene expression and genome structure without altering the primary DNA sequence (Russo et al. 1996). The epigenome comprises alternative chromatin states that can affect gene activity by altering the accessibility of regulatory DNA elements, including promoter and more distal enhancer elements at which both positive- and negative-acting TFs can bind and control

transcription during cell differentiation and development, as well as coordinating transcriptional responses to environmental change (Vergara & Gutierrez 2017; Ong & Corces 2011; Roudier et al. 2009). Packaging DNA into chromatin represents a highly dynamic structure that relies on covalent modification of the DNA by methylation, posttranslational modifications (PTMs) of histones, deposition of histone variants, nucleosome remodelling and small RNAs (Feng & Jacobsen 2011). Upon perceiving a relevant cue, enzyme complexes (that establish, maintain, read, and erase specific epigenetic modifications) can alter the existing chromatin state, thereby defining where and when the transcriptional machinery can access the DNA. Together the different epigenetic modifications, which can be dynamically modulated during development and in response to external signals, serve as recruitment platforms for gene expression regulators or affect chromatin compaction and accessibility for TFs and RNA polymerases.

In recent years, advances have been made to understand the highly dynamic nature of the chromatin, which affects not only transcription but also processes such as genome replication, transposition, DNA repair, recombination and chromosome segregation (Vergara & Gutierrez 2017). Moreover, epigenetic mechanisms have the potential to memorize cell identity or establish new cell fate by reprogramming the epigenome. The potential existence of an epigenetic memory requires mechanisms that actively remove persistent chromatin marks by enzymes or passively through replication-coupled dilution of a repressive modification during cell fate reprogramming (Birnbaum & Roudier 2017). In *Medicago*, root nodule formation is initiated by mitotic reactivation of the root cortical cells in response to rhizobial NF, which implies that first the root cell fate needs to be erased. The cortical cell dedifferentiation and proliferation will lead to the establishment of the nodule meristem. In ZI the post-meristematic cells enter ENR cycle and become infected by rhizobia, which represents the second reprogramming phase towards nitrogen-fixing symbiotic cell differentiation. DNA replication during the consecutive ENR cycles could provide a mechanism to modify the chromatin state and reset or reprogram cell fate.

### **1.3.1. DNA methylation dynamics**

DNA methylation in eukaryotes is a process by which mostly cytosine bases in the genome are modified with a methyl group to form methylcytosine. Adenine methylation, which is frequent in bacteria, is also observed in plants and animals but has received little attention until now (Low et al. 2001; Wu et al. 2016; Greer et al. 2015). Cytosine DNA methylation is involved in various biological processes in plants and is generally associated with

heterochromatin formation and transcriptional gene silencing. It is an important epigenetic mark for transcriptional repression of transposable elements (TEs) and repetitive sequences, stress response and genomic imprinting (Bird 2002; Goll & Bestor 2005; Downen et al. 2012). DNA methylation occurs primarily at CG dinucleotides but in plants cytosine methylation in CHG, and CHH (H = A, C, or T) context have been found also (Henderson & Jacobsen 2007). DNA methyltransferases catalyse the *de novo* formation of 5-cytosine methylation (5-mC) or maintain pre-existing cytosine methylation patterns. Three families of DNA methyltransferases are found in plants. DNA methylation *de novo* is catalysed by DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) and maintained by three different pathways: CG methylation by DNA METHYLTRANSFERASE 1 (MET1), CHG methylation by CHROMOMETHYLASE (CMT3) and CHH methylation by CHROMOMETHYLASE 2 (CMT2) (Chan et al. 2005; Law & Jacobsen 2010; Garg et al. 2014). DRM2 can be targeted to a sequence by small interfering RNAs (siRNAs) termed as RNA-directed DNA methylation (RdDM) (Wassenegger et al. 1994). Although in most cases DNA methylation is a stable epigenetic mark, reduced level of methylation is observed during development in both plants and animals. Methylation can be lost either passively by the failure of maintenance methylation during replication or DNA repair or actively by demethylation enzymes. In *Medicago*, active demethylation is catalysed by DEMETER (DME) and REPRESSOR OF SILENCING 1 (ROS1) probably in combination with the base excision repair pathway (Ikeda & Kinoshita 2009; Zhu 2009; Choi et al. 2002; Gong et al. 2002; Satgé et al. 2016).

DNA methylation is an important epigenetic mechanism in regulating gene transcription, directing deposition of chromatin modifications or chromatin remodelling (Zilberman et al. 2007). However, even if DNA methylation is conserved in eukaryotes, including plants, animals and fungi (although it is lost in some groups including in *Saccharomyces*, *Drosophila* and *Caenorhabditis*), the methylation pattern and level vary drastically among different organisms (Goll & Bestor 2005; Feng et al. 2010). In *Arabidopsis thaliana*, the genome-wide DNA methylation level is 24%, 6.7% and 1.7% for CG, CHG and CHH contexts, respectively (Cokus et al. 2008). The largest fraction of methylated DNA sequences in plants are predominantly represented by TEs and repetitive sequences which are heavily and evenly methylated along their entire length, whereas non-TE genes can be methylated at their promoter, in their transcribed coding region, or at the 3' end (Zhang et al. 2006; Zilberman et al. 2007; Cokus et al. 2008). CG methylation at the promoter and gene body has different effects on the gene expression: promoter methylation usually represses gene expression (Bell & Felsenfeld 2000; Suzuki & Bird 2008) while methylation in gene bodies

does not seem to cause silencing and body methylated genes even tend to be constitutively active (Zhang et al. 2006). Currently, the significance of gene body methylation is not clear, but its enrichment within exons suggests a potential role in pre-mRNA splicing (Lev Maor et al. 2015) or prevention of aberrant transcription from cryptic promoters (Zilberman 2017). In *Arabidopsis*, promoter methylation is more frequent than gene body methylation and genes methylated within their coding sequence are less likely to have tissue-specific expression (Zilberman et al. 2007; Zhang et al. 2006).

Importantly, the methylation level of genetic loci is not a fixed condition but it can change in the course of development or in response to environmental stimuli. It is now well established that the position and the hyper- (increase in methylation) or hypomethylated (undermethylated) status of differentially methylated regions (DmRs) can regulate the expression of proximal genes (Song et al. 2013; Garg et al. 2015). DmRs are generally found at regulatory elements such as enhancers and promoters and display lineage- or cell-type specific methylation patterns. DmRs are more likely to be found near (2 kb upstream and downstream) the genes (42–45%), but less frequently within the gene body (16–18%) (Garg et al. 2015). In addition, genes proximal to hyper-DmRs exhibit lower transcript levels (downregulation), while genes proximal to hypo-DmRs display similar or higher transcript levels (upregulation) compared to all genes (Garg et al. 2015). Since DNA methylation inhibits transcription and spreading of transposons to protect the genome (Kato et al. 2003; Hsieh et al. 2009) this also affects expression of neighbouring genes (Song et al. 2013; Satgé et al. 2016; Garg et al. 2015). Demethylation by DME glycosylase is required to activate gene expression during different plant development processes by establishing a hypomethylated state (Gehring et al. 2006; Satgé et al. 2016), whereas ROS1 demethylation prevents transcriptional gene silencing by maintaining a locus free of methylation (Gong et al. 2002; Zhu et al. 2007). The effect of DNA methylation on gene expression may be mediated either directly or indirectly via transcriptional regulatory proteins, which recognize 5-mCs in the promoter regions or via mechanism involving the recruitment of methyl CG-binding proteins to remodel the chromatin (Nan et al. 1998; Ng & Bird 1999; Watt & Molloy 1988; Medvedeva et al. 2014). The persistence and reversibility of cytosine methylation makes it an ideal mechanism of controlling response to biotic or abiotic stress (Yu et al. 2013; Downen et al. 2012; Bilichak et al. 2012) or plant developmental processes including gametogenesis, embryo, seed and root nodule development (Saze et al. 2003; Gehring et al. 2009; Zemach et al. 2010; Satgé et al. 2016; Hsieh et al. 2009). As a general rule, the extent of genomic methylation is in inverse relationship to developmental potential since pluripotent genomes are highly methylated, whereas

differentiated cells from a variety of lineages display reduced levels of global DNA methylation in animal and plant cells (Gehring et al. 2009; Popp et al. 2010). These data indicate that there is a progressive loss of methylation during differentiation.

To date, much of our understanding about DNA methylation mechanism in plants relies on the research on *A. thaliana*, while methylation analysis in *Medicago* have been published only recently (Satgé et al. 2016). Increasing evidence suggests that investigating DNA methylation alone is not sufficient to get a global picture about the epigenetic landscape of the genome since interplay between DNA methylation and histone modifications is well documented.

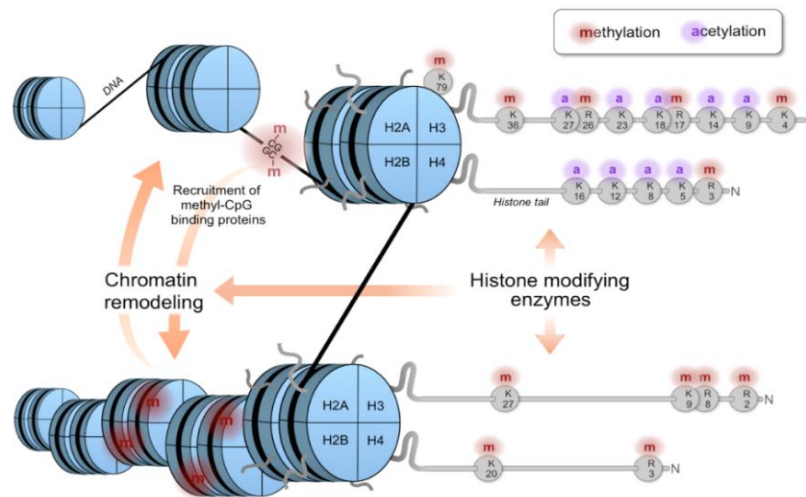
### **1.3.2. Chromatin modifications: The histone code**

In the nuclei of all eukaryotic cells, genomic DNA is folded into nucleosomes composed of histone (H3, H4, H2A, H2B) and non-histone proteins, that are the fundamental building blocks of chromatin. Nucleosomes likely evolved to protect and compact large eukaryotic genomes (Malik & Henikoff 2003). The distinct levels of chromatin organization are dependent on the dynamic structure of nucleosomes, which can restrict the access of DNA-binding transcription factors and RNA polymerases (Li et al. 2007). PTMs of histone tails are important determinants of the higher order chromatin structure (Wolffe & Hayes 1999). PTMs alone or in combination form a “histone code” that is read by other proteins and regulate changes between transcriptionally repressive/structurally condensed- and transcriptionally active/structurally accessible states of the genes and many other cellular processes such as repair, replication and recombination (Strahl & Allis 2000). The histone code considerably extends the information potential of the DNA. Histones can be covalently modified on different amino acid residues, like arginine (R), serine, threonine, tyrosine but mostly on lysine (K) (**Fig. 8**). So far, >60 PTMs of the four core histones have been identified (Kouzarides 2007). Among various kinds of histone PTMs such as acetylation (Ac), methylation (Me), phosphorylation, sumoylation, biotinylation, citrullination, acetylation and methylation of histone H3 and H4 are the best characterized in plants (Grunstein 1997; Kurdistani et al. 2004; Zhou et al. 2010; Zhang et al. 2007). Since the histone proteins and their covalent modifications can alter higher order structure of chromatin, they can influence indirectly the transcriptional “ON/OFF” states of the genes.

Mechanisms for “writing”/depositing, “reading”/translating and “erasing”/removing the histone code is directed by various kind of histone-modifying enzymes encoded by large gene



families in plants and which are highly specific for particular amino acid positions (Berr et al. 2011). The effect of the different histone modifications on transcription in higher eukaryotes are well documented but several modifications appear to have contradictory roles. Therefore we have chosen in our experimental work (see Section 4.2.4) to focus on two widely studied and highly conserved modifications, which play a role in transcriptional activation or in repression, namely the H3 lysine 9 acetylation (H3K9ac) and H3 lysine 27 tri-methylation (H3K27me<sub>3</sub>), respectively (Feng & Jacobsen 2011; Kurdistani et al. 2004; Schubeler et al. 2004; Roh et al. 2005; Zhou et al. 2010; Zhang et al. 2007). However, as the histone code hypothesis states, histone marks can cooperate and in combination or sequentially they will specify the downstream biological function.



**Figure 8. The core histone proteins of nucleosomes and their amino tail modifications** (modified from Keating 2015). Histones are designated as histone H2A, histone H2B, histone H3 and histone H4. The DNA (black) is wrapped around the histone octamer (each histone is present in two copies) building the nucleosome. The N-terminal tails of histones H2A, H2B, H3 and H4 are potential target sites for histone-modifying enzymes. Histone H3 lysine 9 (H3K9) and histone H3 lysine 27 (H3K27) residues are among the best-characterized modifications. K: lysine; R: arginine, a: acetylation of histones; m: methylation of histones or DNA; N: amino terminus. (**Top**) Open/accessible chromatin conformation. (**Bottom**) Closed/inaccessible chromatin.

### 1.3.2.1. H3K9ac marks active genes

The acetylation of lysine residues neutralizes the positive charge of the histone tails and decreases their affinity for negatively charged DNA, thereby promoting the binding of TFs to regulatory DNA sequences. The reduction of DNA-nucleosomal interaction leads to chromatin decondensation and transcription, while deacetylation reverses this effect (Wolffe & Hayes 1999). The level of histone acetylation is regulated by the activity of both histone acetyl transferases (HATs) and histone deacetylases (HDACs), which acetylate and deacetylate lysine

residues of histone tails, respectively. Acetylation of histone H3K9 generally marks active genes and participates in numerous developmental and biological processes in both plants and animals (He et al. 2003; Benhamed et al. 2006; Grunstein 1997; Qiao et al. 2014; Zhou et al. 2010). Histone acetylation is shown to play important roles in the regulation of plant cell cycle, flowering time, response to environmental conditions and hormone signals. H3K9ac is enriched downstream of transcription start site of target genes in *Arabidopsis*, but it can span the gene body and promoter as well. *Arabidopsis* genes containing only H3K9ac modification are highly expressed, while genes with H3K9ac and repressive DNA methylation or H3K27me3 exhibited lower expression level. Furthermore, *Arabidopsis hda19* (histone deacetylase 19) mutant seedlings showed significant increase of the H3K9ac modification level which also resulted in an increased expression level of the target genes (Zhou et al. 2010).

#### **1.3.2.2. H3K27me3 marks developmentally regulated genes**

While stable silencing of transposons is mediated by DNA methylation, silencing of genes during cell differentiation is mediated by Polycomb-group proteins (PcG) that are composed of two major enzyme complexes: the Polycomb Repressive Complex 1 (PRC1) and PRC2. PcG represses gene expression at the chromatin level in eukaryotes. PRC1 catalyses monoubiquitination on histone H2A (Wei et al. 2006), and PRC2 deposits trimethylation marks on histone H3K27 (H3K27me3) (Schwartz et al. 2006; Lee et al. 2006). H3K27 has a key role in gene repression and is critical for normal development of animals and plants (Hennig & Derkacheva 2009; Lee et al. 2006). The silencing of TF- and other regulatory genes during development by H3K27me3 is a conserved mechanism. In mammals, histone H3K27 methylation is catalysed by the histone lysine methyltransferase (HKMT) Enhancer of Zeste 2 (EZH2) (Kuzmichev et al. 2002) which is part of the PRC2 and involved in the repression of many genes involved in development and cell differentiation (Boyer et al. 2006; Bracken et al. 2006). Loss of plant PcG proteins leads to loss of organ identity. PRC1 and PRC2 collectively establish and maintain H3K27me3, respectively while it is removed by KDM6a and KDM6b histone demethylases (Agger et al. 2008). In *Arabidopsis* EZH2 homologs like CURLY LEAF (CLF), MEDEA (MEA) and SWINGER (SWN) are H3K27me3 methyltransferases (Hennig & Derkacheva 2009; Zheng & Chen 2011). However, phylogenetic analysis did not reveal any homolog of KDM6. The RELATIVE OF EARLY FLOWERING 6 (REF6), has been shown to specifically demethylate H3K27me3 in *Arabidopsis* (Lu et al. 2011). The classic recruitment model of PcG complexes in which PRC2-mediated H3K27 trimethylation recruits PRC1 was

recently challenged by data showing that PRC1 activity can also recruit PRC2 (Blackledge et al. 2015; Zhou et al. 2017).

Genome-wide analysis in *Arabidopsis* demonstrated that most H3K27me<sub>3</sub>-marked genes exhibited low expression level and high tissue-specificity with expression in one or a few specific tissues (Zhang et al. 2007; Turck et al. 2007; Roudier et al. 2011). In *Drosophila* and mammals H3K27me<sub>3</sub> forms extensive domains (up to hundreds of kbs) maintaining multiple genes in transcriptionally suppressed state, while in *Arabidopsis* H3K27me<sub>3</sub> regions were significantly shorter, rarely extended to adjacent genes and mostly marked single transcription units (Zhang et al. 2007; Turck et al. 2007; Roudier et al. 2011). In addition, maps of H3K27me<sub>3</sub> in *Arabidopsis* revealed different H3K27me<sub>3</sub> enrichment profiles with distinct regulatory outcomes. A broad domain of H3K27me<sub>3</sub> across the body of genes corresponded to transcriptional inhibition while an enrichment profile with a peak in the promoter of the target genes is associated with active transcription (Young et al. 2011; Zhang et al. 2007). Genes with low nucleosome density (LND) regions in their promoters were among the most highly expressed genes in *Arabidopsis* which showed very low tissue specificity and coded in majority catalytic enzymes involved in a variety of physiological processes, while H3K27me<sub>3</sub> regions were not preferentially associated with LND regions (Zhang et al. 2007). Moreover, it was also shown that H3K27me<sub>3</sub> is a major silencing system that is independent of DNA methylation or RNA interference. Although the silencing function of H3K27me<sub>3</sub> is conserved, plants and animals use distinct mechanisms to regulate H3K27me<sub>3</sub> dynamics. Removal of H3K27me<sub>3</sub> can be achieved by the enzymatic activity of REF6 demethylase or passively by replication-coupled dilution (Cui et al. 2016; Lu et al. 2011; Morao et al. 2016).

The acetylated and methylated lysine residues in histones can be recognized and bound by transcriptional co-regulators and chromatin-remodelling factors and change the expression pattern of the underlying genes. Collectively, although certain mechanisms that deposit and maintain these modifications might differ between plants and animals, the distributions and functions of H3K9ac and H3K27me<sub>3</sub> seem to be well conserved.

### **1.3.3. DNA accessibility and chromatin remodelling**

The accessibility of DNA to regulatory proteins is a major property of the chromatin that restricts or enables transcription. The deposition or the removal of specific histone marks or DNA methylation is one way in which the chromatin state can affect the transcriptional program of the cell but the position of the nucleosomes can also be altered, globally or at the

level of individual genes. It has been proposed that a major function of the multiple combinations of histone modifications is to fine-tune accessibility of the DNA (Henikoff & Shilatifard 2011).

In 1928 Emil Heitz made the distinction between heterochromatin and euchromatin based on their differential compaction in interphase nuclei (Heitz 1928). Heterochromatin represented the densely stained, compacted areas, while euchromatin the sparsely stained, less compact areas. Histological analysis of the chromatin in progenitor and differentiated cells indicated that undifferentiated cells typically have open chromatin conformation and are mostly devoid of heterochromatin (Spangrude et al. 1988; Park et al. 2004). However, it was suggested that this view might be an oversimplification, since various high-molecular weight probes and proteins can enter heterochromatin (Verschure et al. 2003; Görisch et al. 2005) and also transcription of repetitive sequences by specific RNA polymerases requires H3K9-methylation and heterochromatin assembly (Volpe et al. 2002). In *Arabidopsis* Shu et al. showed that genes without the repressive H3K27me<sub>3</sub> are often inaccessible meaning that reduced accessibility is not necessarily the result of “visible” compaction but a local property of the chromatin (Shu et al. 2012). Inactive genes have an intermediate accessibility, which can be greatly reduced by repressive PcG proteins and/or DNA methylation. Conversely, transcription is often associated with a local increase in accessibility, which is restricted to promoters and 5'-sequences. It is possible that accessibility is locally controlled by neighbouring nucleosomes (Misteli et al. 2007). Chromatin-remodelling proteins that can modify histone octamer-DNA contacts can alter the position of nucleosomes. Remodelers use the energy of ATP hydrolysis to move, insert, eject, exchange or restructure nucleosomes. Chromatin remodeling ATPases have been divided into four major subfamilies SWI/SNF, CHD, INO80/SWR1 and ISWI (Knizewski et al. 2008; Clapier & Cairns 2009; Hargreaves & Crabtree 2011; Narlikar et al. 2013). SWI/SNF chromatin remodelers disassemble or slide nucleosomes and interact with ‘activating’ chromatin regulators such as the histone acetyl transferase GCN5 (Cosma et al. 1999). However, recent studies in mammals and plants have shown that SWI/SNF complexes can also directly repress gene expression (Ho et al. 2009; Ho et al. 2011; Hargreaves & Crabtree 2011; Zhao et al. 2015) by facilitating the recruitment of sequence-specific binding proteins or chromatin regulators with repressive function (Ho et al. 2009) or stabilizing nucleosomes at the transcription start site that interfere with transcription (Han et al. 2012). The CHD3-family chromatin remodeller PICKLE (PKL) has opposite activities in *Arabidopsis* (Shen et al. 2015). Many studies have implicated PKL in activation of transcription by counteracting Polycomb repression (Fukaki et al. 2006; Furuta et al. 2011) but others demonstrated a role for promoting Polycomb-repression

(Aichinger et al. 2009; Aichinger et al. 2011; Jing et al. 2013; Zhang et al. 2014) although evidence for association of PKL with Polycomb-repressed loci is not demonstrated yet. Moreover, recently it was also revealed that mutation of PKL results in defects in DNA methylation at a subset of RdDM loci (Yang et al. 2017). Many studies have characterized the involvement of SWI/SNF and CHD3 subgroup ATPases in growth, differentiation and in developmental phase transitions (Kwon et al. 2005; Aichinger et al. 2009; Aichinger et al. 2011; Wagner & Meyerowitz 2002; Wagner, 2003). *Arabidopsis* SWR1 directs the exchange of the canonical histone variant H2A for H2A.Z at the FLOWERING LOCUS C (FLC) which is affecting the stability of the nucleosome and plays a role in repression of flowering (Noh & Amasino 2003; Deal et al. 2007; Deal & Henikoff 2010). On the other hand, INO80 catalyses the opposite reaction by replacing H2A.Z with the canonical H2A (Mizuguchi et al. 2004; Luk et al. 2010). Chromatin remodelers in the ISWI families have been linked to DNA methylation in plants. DECREASE IN DNA METHYLATION 1 (DDM1) can shift nucleosomes *in vitro* (Brzeski & Jerzmanowski 2003) and its mutation causes loss of DNA methylation from some TEs and repeats (Jeddeloh et al. 1999; Zemach et al. 2013) but not from genes (Lippman et al. 2004). DDM1 enables DNMTs to access H1-containing heterochromatin for stable silencing of TEs (Zemach et al. 2013). These findings link nucleosome positioning with the initiation of RdDM. The dual functional role of chromatin remodeling factors in regulating plant development and silencing repeats by RdDM suggests that both processes are more closely connected.

One of the unique feature of plants is that they maintain pools of pluripotent stem cells that can generate new organs from meristems throughout their lives. For this purpose, pluripotency needs to be stably maintained while it also needs to be flexible in response to differentiation signals. Chromatin remodelers play important roles in controlling switches of transcriptional programs and provide an excellent regulatory platform for the plant lifestyle. How chromatin organisation of the nodule meristematic cells is reprogrammed during symbiotic cell differentiation in *Medicago* is entirely unknown.

## **1.4.Plant processes under epigenetic control**

### **1.4.1. Epigenetic changes during development**

One way to actively reprogram the epigenome is to remove methylated cytosines either passively or actively. In passive demethylation the methylated cytosines are replaced by

unmethylated cytosines during DNA replication, while in active demethylation the 5-mC is removed by DNA glycosylases (DME) (Choi et al. 2002; Gong et al. 2002). Several studies revealed the dynamic nature of DNA methylation in different aspects of plant development. Active demethylation of promoters by DME has been shown to be required for the imprinted expression of maternal genes such as MEDEA (MEA), FERTILIZATION INDEPENDENT SEED 2 (FIS2) and FLOWERING WAGENINGEN (FWA) during female gametogenesis in *Arabidopsis* (Zhang et al. 2010; Choi et al. 2002; Kinoshita 2004). Recently it was also demonstrated that DNA demethylation is important in controlling indeterminate nodule organogenesis in *M. truncatula* (Satsgé et al. 2016). MtDME regulates the expression of several genes involved in plant and bacterial cell differentiation. MtDME is mostly expressed in the nodule differentiation zone. The decrease in DME expression resulted in hypermethylation and downregulation of genes associated with nodule differentiation. Moreover, demethylation by DME also reactivated transposon expression (Satsgé et al. 2016). Besides active demethylation by MtDME, passive demethylation is also likely to occur in the nodule differentiation zone where methylation could be passively lost by dilution during the consecutive endocycles.

Several studies demonstrated that DNA methylation differs between plant organs and tissues (Widman et al. 2014; Song et al. 2013), but the extent of variation between different cell types is resolved only recently in *Arabidopsis* (Kawakatsu et al. 2016). To analyse DNA methylome between six root meristematic cell types fluorescence-activated cell sorting (FACS) have been used. The experiment revealed that columella cells are the most highly methylated in *Arabidopsis*. This is similar to our experimental approach where the cell-type specific isolation of nodule nuclei by flow cytometry was based on the different DNA content (ploidy level) of the cells. Kawakatsu et al. (2016) showed that the absence of the chromatin remodelling protein DDM1 and loss of heterochromatin formation allow access of RdDM factors to the genome, and the production an excess of 24-nt small RNAs lead to CHH hypermethylation of TEs in columella cells. Moreover, they showed that genes harbouring TEs exhibited low expression level, suggesting that increase in mCHH within protein-coding genes might be due to the hypermethylation of nearby TEs (Kawakatsu et al. 2016). Thus, DNA methylation has essential roles in the silencing of TEs and imprinted expression of certain loci, but loss of DNA methylation is not sufficient *per se* to induce gene expression (Bestor et al. 2015). Presumably, the expression of these loci requires additional factors, some of which may be related to cell lineage and differentiation.

In addition to DNA methylation, histone modifications can also be reversed in plants. PRC2 is a key regulator of major phase transitions in plants such as embryo-seedling transition

(Bouyer et al. 2011) and vegetative growth to flowering (Berry & Dean 2015). It has been demonstrated that PRC2 is essential to the reprogramming and to the switch from embryonic to seedling phase in *Arabidopsis* (Bouyer et al. 2011). Homozygous loss of the FERTILIZATION INDEPENDENT ENDOSPERM (FIE) gene - component of the PRC2 complex - in seeds resulted in enhanced dormancy and germination defects. Moreover, several genes controlling seed maturation and dormancy were marked by H3K27me3 and were upregulated upon loss of PRC2 (Bouyer et al. 2011). Another example for Polycomb-group protein based control is the silencing of FLOWERING LOCUS C (FLC) in *Arabidopsis* that regulates flowering (Bastow et al. 2004). The coordinated switch of activating histone modifications from H3K4me3/H3K36me3 to repressive H3K27me3 appeared to be important in maintaining the "ON" or "OFF" expression states of FLC. In winter (in annual accessions of *Arabidopsis*) FLC represses flowering. During vernalisation (prolonged cold treatment) FLC expression is gradually repressed epigenetically by H3K27me3, and silencing of FLC persists until temperatures becomes warmer, therefore *Arabidopsis* can flower in response to the photoperiod. The "ON" expression state of FLC during winter is characterized by the presence of H3K4me3, H3K36me3, histone acetylation, and active transcription by RNA polymerase II. The PRC2 complex located at FLC is important for the vernalisation response, which comprises FIE which specifically recognize H3K27me3 while CLF and SWN catalyses H3K27me3 formation (Cao et al. 2002). The RNA-binding ability of Polycomb complexes led to the hypothesis that long non-coding RNAs may act as 'recruiters' of PRC2 (Tsai et al. 2010). Indeed, the long non-coding RNA COOLAIR (Heo & Sung 2011) appears to be required during cold exposure to ensure removal of activating chromatin marks and to mediate FLC transcriptional down-regulation. This raises the possibility of the existence of an H3K36me3 demethylase (currently hypothetical), whose targeting to FLC depends on antisense COOLAIR transcripts (Berry & Dean 2015). The coordinated loss of H3K4me3/H3K36me3 and gain of H3K27me3 regulate FLC expression states through *Arabidopsis* life cycle. During cold exposure, the switch from H3K36me3-rich to H3K27me3-rich chromatin mediates FLC repression by the PRC2 complex. FLC expression increases throughout embryogenesis, and reaches a maximum when the seed has fully formed (Sheldon et al. 2008; Choi et al. 2009). H3K27me3 marks on FLC are removed by an unknown mechanism and in the seeds, FLC becomes re-activated. Thus, flowering is inhibited by FLC until the next-generation plants encounter cold weather.

### 1.4.2. Epigenetic changes in response to biotic and abiotic stress

Plants are frequently exposed to abiotic stressors, including water deficit, high-salinity, temperature fluctuation and nutrition deficiency, and biotic threats, such as insect herbivores or microbial pathogens and therefore have evolved sophisticated adaptation and defense mechanisms. Epigenetic modifications may provide a mechanistic basis for a stress memory, which results in rapid, reversible, or trans-generational heritable changes in gene expression. Research over the last decade revealed that plants have stress memory that supports adaptation to reoccurring stress and that transcriptional regulation of abiotic stress-responsive genes and defense-related genes often involves DNA methylation, histone modifications and ATP-dependent chromatin remodeling (Ding & Wang 2015; Downen et al. 2012; Yu et al. 2013; Le et al. 2014; Kim et al. 2015; Bilichak et al. 2012; Sani et al. 2013; Kim et al. 2012; Lämke & Bäurle 2017). Most responses to biotic or abiotic stress exposure that involve chromatin features are transient and return quickly to baseline levels after normal conditions have been restored. Global changes in DNA methylation in response to abiotic stress have been reported in several plant species (Bilichak et al. 2012; Choi & Sano 2007; Song et al. 2012; Steward et al. 2002). Roots of maize seedlings subjected to cold stress reacted by genome-wide demethylation (Steward et al. 2002). Oxidative stress induced active demethylation and transcriptional activation of *NtGPD*L (glycerophosphodiesterase-like protein) in *Nicotiana tabacum* (Choi & Sano 2007) leading to higher stress tolerance, while salinity stress triggered hypomethylation and activation of salt stress-responsive TFs in soybean (Song et al. 2012). Furthermore, it was shown in *A. thaliana* that progeny of plants exposed to salt stress exhibited changes in DNA methylation pattern, histone modifications and gene expression, indicating that such trans-generational stress adaptation depends on numerous epigenetic marks (Bilichak et al. 2012). Chromatin regulators were identified to be involved in the control of stress-responsive genes. For example in response to drought, RNA polymerase II was recruited on the drought-inducible genes *RD20*, *RD29A* and *AtGOLS2* in *Arabidopsis* and rapidly disappeared after rehydration (Kim et al. 2012). Histone modifications like H3K9ac were enriched by drought and rapidly removed from *RD29A*, *RD20*, and *AtGOLS2* by rehydration. In contrast, histone H3K4me3 was gradually decreased by dehydration but was maintained at low levels after rehydration, suggesting that H3K4me3 functions as an epigenetic mark of stress memory (Kim et al. 2012; Ding et al. 2012). These results show that the transcriptional activity and chromatin status are rapidly changed from an active to inactive state during the stress recovery process. Moreover, the transcriptional responsiveness of drought stress-upregulated genes was found to be



dependent on the intensity of the stress (Kim et al. 2008; Kim et al. 2012). Under strong drought conditions, the histone modifications H3K4me3 and H3K9ac on drought stress-upregulated genes were more highly enriched than under moderate drought conditions and nucleosome density on the *RD20* and *RAP2.4* genes was gradually decreased in response to increase in stress (Kim et al., 2012). Hyperosmotic priming (“immunisation”) of *Arabidopsis* seedlings, exposed to mild salt treatment followed by an extensive period of growth in control conditions altered the H3K27me3 epigenomic landscape (Sani et al. 2013). In salt-primed plants, the *HKT1* gene, encoding Na-transporter was more strongly induced than in the non-primed (naïve) plants during a second salt treatment, indicating a somatic long-term memory of adult plants (Sani et al. 2013).

Innate immunity pathways such as the salicylic acid (SA), jasmonic acid (JA), and ethylene (ET) signaling play pivotal roles in plants in defending against biotrophic and necrotrophic pathogens (Glazebrook 2005). Profiling DNA methylomes of plants exposed to the bacterial pathogen *Pseudomonas syringae* pv. *tomato* DC3000 (*Pst*) or to the fungal pathogen *Fusarium oxysporum* revealed numerous stress-induced DMRs, many of which were associated with differentially expressed genes (Downen et al. 2012; Le et al. 2014; Yu et al. 2013). DNA demethylation has been shown to play important role in plant innate immune response by the activation of SA-dependent signaling. In response to SA, TE-associated DMRs were accompanied by upregulation of 21-nt siRNAs, and often coupled to transcriptional changes of the TEs and/or neighbouring genes (Yu et al. 2013). Thus, the primary function of DNA demethylases in plants is to regulate the expression of stress response genes by targeting TEs or repetitive sequences (Downen et al. 2012; Yu et al. 2013; Le et al. 2014). *Arabidopsis* histone deacetylase 19 (HDA19) has been also shown to be involved in the ET/JA signaling during pathogen attack. Expression of *HDA19* is induced by challenge with the necrotrophic pathogen *Alternaria brassicicola*, and by treatment with the plant hormone ET and JA. HDA19 overexpression resulted in plant lines with increased resistance against the pathogen (Zhou et al. 2005). Under unchallenged conditions, HDA19 forms a repressive chromatin environment (low histone acetylation level) that ensures a low basal expression of pathogenesis related (*PR*) genes, which might be critical for preventing overstimulation of defense responses (Choi et al. 2012) while also establishing a chromatin status for rapid *PR* gene induction when the plant is challenged by a pathogen. In contrast, the removal of the repressive histone H3K27me3 by the Jumonji C Domain protein JMJ705 in rice also plays important roles in defense-related gene expression (Li et al. 2013). During pathogen infection, JMJ705 is involved in the JA-induced removal of H3K27me3 and biotic stress-responsive gene activation (Li et al. 2013). Plant

defense can also be affected by chromatin-remodeling factors that control specific pathways within biotic stress signaling networks. The SWR-1-like chromatin remodelling complex that replaces the histone H2A with the histone variant H2A.Z is required for the maintenance of the repressed SA-mediated systemic acquired resistance (SAR) response (March-Díaz et al. 2008) which is associated with the hypersensitive response, characterized by apoptotic-like cell death (Hammond-Kosack 1996). Furthermore, the *Arabidopsis* SWI/SNF class chromatin remodeling ATPase SPLAYED (SYD) is required for resistance against the necrotrophic pathogen *Botrytis cinerea* but not the biotrophic pathogen *Pseudomonas syringae*, indicating that chromatin remodeling is required for selective pathogen resistance (Walley et al. 2008).

Thus, during abiotic stress DNA demethylation and enrichment of H3K9ac and removal of H3K4me3 positively affect gene activation of stress responsive genes. It has been shown that stress response is memorized via histone modification on several stress-upregulated genes (Ding et al., 2012). At trainable genes, H3K4me3 and RNA polymerase II persist as memory marks. During biotic stress, active histone marks enable a basal expression level of defense-related genes and their effective induction when the plant is challenged by a pathogen, while repressive histone modifications can prevent unnecessary activation of defense genes under normal growth conditions. Moreover, histone/chromatin changes that occur in response to biotic or abiotic stresses can be transmitted to the next generation. Emerging evidence suggests that long non-coding RNAs and chromatin remodeling complexes are shaping the dynamic, three dimensional genome topology through chromatin loops to regulate gene expression (Ariel et al. 2014; Jegu et al. 2014) but currently it remains an unexplored area during biotic and abiotic stress response.

## 2. OBJECTIVES OF THE WORK

The generation of differentiated symbiotic cells in the legume *M. truncatula* involves a massive reprogramming of the nodule transcriptome with the activation of nodule-specific genes in successive spatio-temporal waves (Maunoury et al. 2010). Many of the nodule-induced genes are also expressed in other plant tissues, but a large proportion of the nodule transcriptome is composed of genes that are exclusively expressed in the nodules and are maintained silent in all other plant tissues. Among the activated genes are the *NCR* family, producing hundreds of different *NCR* peptides that guide terminal differentiation of the endosymbionts converting them to large polyploid non-cultivable nitrogen-fixing bacteroids. The *NCR* genes are under tight transcriptional control since they are exclusively expressed in the symbiotic nodule cells and their activation requires the intracellular presence of rhizobia (Van de Velde et al. 2010; Maunoury et al. 2010; Mergaert et al. 2003). The expression of the *NCR* genes in the symbiotic cells suggests a direct connection between the ENR and the expression of the *NCR* genes. How ENR or the polyploid state of the genome controls gene expression is entirely unknown. The strict nodule specificity and the possible linkage between gene expression and ploidy levels of symbiotic cells suggested that epigenetic modifications along the ENR cycles might play a critical role in transcriptional regulation of *NCRs* and other nodule-specific genes. The recently reported reprogramming of DNA methylation in *M. truncatula* nodule development, regulated by the DNA demethylase gene *DEMETER* (*MtDME*), is in agreement with this hypothesis (Satgé et al. 2016). Changes in 5-cytosin methylation of the DNA and in histone tail modifications, alone or in combination, are important determinants of the chromatin structure and gene expression. The sequence context (CG, CHG, CHH, where H = A, C, or T) and the location of the DNA methylation in the promoter, gene body or 3' regulatory region can have different effects on gene transcription. In general, there is a strong correlation between DNA methylation in the promoter region and decrease in gene expression (Zhang et al. 2006; Zilberman et al. 2007; Cokus et al. 2008; Garg et al. 2015). Among the histone tail modifications, H3K27me3 (Feng & Jacobsen 2011) also results in low expression level and high tissue-specificity (Zhang et al. 2007) whereas H3K9ac leads to gene activation (Kurdistani et al. 2004; Schubeler et al. 2004; Roh et al. 2005; Zhou et al. 2010).

A picture is emerging that the epigenome is specific for a given cell type or phase of the cell cycle and it is important to determine how the epigenome between distinct cell types changes during cellular differentiation to fully evaluate the impact of the different epigenetic

modifications in the regulation of gene activity. Thus, the symbiotic nodule cells of *M. truncatula* constitute an excellent model system to study transcriptional reprogramming at developmental-stage specific level because of the symbiotic cell differentiation involves ENR and a massive reprogramming of their transcriptome. Moreover, the indeterminate *M. truncatula* nodules contain cells at all the stages of symbiotic cell differentiation, from the undifferentiated meristematic cells (2C/4C), over the post-mitotic stage when they have stopped cell division and have a 4C ploidy and the intermediate differentiation stages with ploidy levels of 8C-16C, until the mature nitrogen-fixing cells with ploidy levels of 32C-64C. This property of these nodules facilitates sampling of all possible differentiation stages of the symbiotic cells.

From the high specificity of the symbiotic nodule cell-specific genes and the apparent link of their transcriptional activation during ENR in the symbiotic cells, we hypothesise that epigenetic mechanisms involving specific histone modifications and/or DNA methylations may be implicated in the determination of the gene expression program in the symbiotic cells. Repressive chromatin structures could keep the genes silent during all stages of plant development. Reversal of these marks in the differentiating symbiotic cells would be required to make them competent for transcription by specific TFs. The combination of repressive chromatin marks, their reversion and the TFs could account for the extremely robust symbiotic nodule cell specific expression. Moreover, we propose that ENR of the symbiotic nodule cells serves as a mechanism in controlling gene expression. ENR could actively lead to the activation of “erasers” of the repressive chromatin marks. Alternatively, repressive marks, which are normally faithfully transmitted during the cell cycle, could be passively lost by dilution during the consecutive endocycles. These hypotheses form the basis for my thesis work.

**In the first part of the thesis, the transcriptome database provided by the *Medicago truncatula* Gene Expression Atlas (MtGEA) was used to characterize in detail the expression behaviour of a Nodule-Specific Cysteine-Rich (NCR) gene family.**

1. The MtGEA transcriptome database (Benedito et al. 2008; He et al. 2009) is generated with the whole genome Affymetrix *Medicago* Gene Chip and compiles microarray data for the majority of *M. truncatula* genes (50,900 probe sets) over a large set of experiments (267 different experiments) including different plant organs, biotic-, abiotic stress conditions and different tissue types. The MtGEA database is currently the richest resource for analysing expression pattern and specificity of the *M. truncatula* NCR genes in various conditions.

2. Nodules of *M. truncatula* are indeterminate type therefore, mature nodules are organized in well-defined histological zones. This nodule structure suggests that the temporal *NCR* expression profiles could correlate with a spatial expression pattern in the nodule tissues. To test this possibility, we analyzed transcriptome data from 4-weeks-old nodules that were hand-sectioned in five different parts (ZI, ZII<sub>d</sub>, ZII<sub>p</sub>, IZ and ZIII).
3. In order to confirm the expression data from MtGEA, stable transgenic *M. truncatula* R108 lines were generated carrying promoter- $\beta$ -glucuronidase reporter (GUS) fusion constructs for three different *NCR* genes, representing different temporal classes of *NCRs*, as well as a specific antibody for one *NCR* peptide.
4. The MtGEA compendium contains datasets from different pathogenic interactions but these are exclusively root pathogens. To confirm *NCR* expression pattern specificity in various conditions, most particularly during leaf or stem pathogen infection encompassing also other trophic interactions (bio/hemibio/necrotrophic), stable transgenic *M. truncatula* R108 lines carrying *NCR* promoter-GUS reporter fusion constructs were used.

**In the second part of the thesis, the involvement of epigenetic regulation in the transcriptional reprogramming during symbiotic cell differentiation was investigated at the resolution of the differentiation stage of the symbiotic cells by purifying nodule nuclei of defined ploidy level with a flow cytometer coupled to a cell sorter.**

5. *In situ* expression of various *NCR* and nodule-specific genes suggested that their expression could be coupled to specific ploidy levels of the nodule cells. To confirm a possible linkage between the expression of nodule-specific genes in successive transcriptional waves and the ploidy levels of differentiating symbiotic cells, the expression level of selected, differentially expressed genes was measured in flow-cytometry isolated 4C, 8C, 16C and 32C nuclei, by reverse transcription quantitative PCR (RT-qPCR).
6. As the majority of nodule-specific genes are expressed in the mature symbiotic cells (32C) and repressed in the uninfected (4C) cells, we analysed differences in DNA methylation (5-mC) genome-wide in isolated 4C and 32C nuclei using Reduced Representation Bisulphite Sequencing (RRBS). Patterns of DNA methylation were matched to published *in situ* RNA-seq data of the nodule zones obtained by LCM (Roux et al. 2014). Moreover, differential methylation of nodule-specific- and nodule expressed genes in 4C, 8C, 16C and 32C nuclei were analysed using Methylated DNA Immunoprecipitation (MeDIP).

7. To learn how chromatin compaction and accessibility changes during the course of nodule differentiation and how this influences gene expression, we studied genome-wide the chromatin accessibility in 4C, 8C, 16C and 32C nuclei using Assay for Transposase-Accessible Chromatin using sequencing (ATAC-Seq). Patterns of chromatin accessibility changes were matched to published *in situ* RNA-seq data (Roux et al. 2014).

8. To analyse the correlation between chromatin modifications and gene expression, we investigated how the antagonistic chromatin marks H3K27me3 and H3K9ac correlate with activation and repression of selected, differentially expressed genes in 4C, 8C, 16C and 32C nuclei using Chromatin Immunoprecipitation (ChIP).

The presented results of my thesis work were published in two papers. The data mining study on the *NCR* expression and the experimental validations was published in *BMC Genomics*: Ibtissem Guefrachi, Marianna Nagymihály, Catalina I Pislariu, Willem Van de Velde, Pascal Ratet, Mohamed Mars, Michael K Udvardi, Eva Kondorosi, Peter Mergaert and Benoit Alunni. **Extreme specificity of *NCR* gene expression in *Medicago truncatula*** *BMC Genomics* **2014, 15:712**. I am co-first author on this paper. The epigenetics study was published in the *Proceedings of the National Academy of Sciences*: Marianna Nagymihály, Alaguraj Veluchamy, Zoltán Györgypál, Federico Ariel, Teddy Jégu, Moussa Benhamed, Attila Szűcs, Attila Kereszt, Peter Mergaert, and Éva Kondorosi. **Ploidy-dependent changes in the epigenome of symbiotic cells correlate with specific patterns of gene expression.** *Proc. Natl. Acad. Sci. USA* **2017, 114:4543-4548**.

In addition, I contributed during my thesis period to a study of the symbiotic efficiency of a collection of *Sinorhizobium* strains on different *M. truncatula* accessions and *Medicago sativa* cultivars. This study was motivated in the first place to identify a bacterial strain that has a high performance on *M. truncatula* because the widely used model strain Sm1021 is a notoriously bad symbiont of *M. truncatula*. In this work, we showed that the endoreduplication level of the symbiotic nodule cells is a good parameter that correlates with the efficiency of the symbiosis. This work was recently published in *Molecular Plant-Microbe Interactions* and I am co-first author of this paper (Kazmierczak, Nagymihály et al., 2017). Also in this study, we identified a new strain, *Sinorhizobium meliloti* FSM-MA, which is uniquely highly performant on all tested *Medicago* hosts. This strain might be useful for the scientific community working on *M. truncatula*. Therefore, we have determined the genome sequence of this strain and this work has been submitted recently for publication in *Standards in Genomic Sciences*. I am first author of that paper.

### 3. MATERIALS AND METHODS

#### 3.1. *NCR* expression analysis and entropy calculation

##### 3.1.1. Analysis of the *Medicago truncatula* Gene Expression Atlas (MtGEA) data

The MtGEA transcriptome compendium was downloaded from the website of the Samuel Roberts Noble foundation (<http://mtgea.noble.org/v3/>). Data were also obtained from the NCBI Gene Expression Omnibus (accession n° GSE53406 and GSE43354, respectively) (Jayaraman et al. 2014; Limpens et al. 2013). All the data were imported in Excel (**Appendix, Supplementary information 1**; [https://static-content.springer.com/esm/art%3A10.1186%2F1471-2164-15-712/MediaObjects/12864\\_2014\\_6439\\_MOESM1\\_ESM.xlsx](https://static-content.springer.com/esm/art%3A10.1186%2F1471-2164-15-712/MediaObjects/12864_2014_6439_MOESM1_ESM.xlsx)) for extracting the expression profiles of the 334 *NCR*s for further analysis. The *NCR* probe-sets on the Affymetrix *Medicago* GeneChip were obtained by BLASTn searches on the MtGEA website (**Appendix, Supplementary information 1**). Each individual *NCR* nucleotide sequence resulted in the identification of multiple probe-sets due to the homology between *NCR* gene sequences. In total 334 different probe-sets were retrieved. This collection represents nearly all *NCR* probe-sets present on the Affymetrix *Medicago* GeneChip and the remaining genes identified in Young et al. (2011) and Zhou et al. (2013) are missing from these arrays because they were not yet annotated at the time of array design.

Cluster analysis of the complete MtGEA dataset was performed using the MeV software package. Briefly, the Excel datasheet extracted from MtGEA was analysed using the Euclidean distance application with average linkage settings. Heat maps were generated with MeV and histograms and graphs with Excel.

##### 3.1.2. Transcriptome analysis of hand-dissected nodule zones

Using the leghemoglobin color gradient along the nodule as guideline, five regions (A to E) corresponding to meristem-, invasion zone-, interzone-, nitrogen fixation- and senescence zone-enriched tissue were hand-dissected from 28 dpi nodules, as previously described (Zhou et al. 2011). It should be noted that each hand-dissected sample are enriched for the indicated zone but can contain cell layers from the adjacent zones as well.

Total RNA extraction and purification were conducted as described (Benedito et al. 2008). For hybridization onto the Affymetrix *Medicago* Genechip Array probes were synthesized and labelled from 500 ng RNA using the Gene Chip 3' IVT express kit following manufacturer's guidelines (Affymetrix). Global normalization of expression was carried out using the Robust Multiarray Average Express software (Irizarry et al. 2003).

### **3.1.3. Laser-capture microdissection coupled to RNA-Seq (LCM RNA-seq)**

The expression profiles for the *NCR* gene-set described in this study were extracted from the data obtained by LCM RNA-Seq from the website <https://iant.toulouse.inra.fr/symbimics> (Roux et al. 2014).

### **3.1.4. Shannon entropy calculations**

Calculations were performed on the MtGEA dataset in Excel. For the normalization of expression levels in  $N$  tissues, the relative expression  $P_{t/g}$  of a gene  $g$  in a tissue  $t$  was calculated as  $P_{t/g} = W_{t/g} / \sum_{1 \leq t \leq N} W_{t/g}$  where  $W_{t/g}$  is the expression level of the gene  $g$  in the tissue  $t$ . The Shannon entropy  $E_g$  of gene  $g$  is calculated as  $E_g = -\sum_{1 \leq t \leq N} P_{t/g} \log_2(P_{t/g})$ .  $E_g$  ranges from zero for genes expressed in a single tissue to  $\log_2(N)$  for genes expressed uniformly in all tissues considered. Heat maps of entropy values were generated by the MeV software package.

## **3.2. Transgenic plants and $\beta$ -glucuronidase reporter (GUS) analysis**

### **3.2.1. Promoter-GUS constructs**

The promoters of *NCR001*, *NCR084* and *NCR121* (respectively 2.5 kb, 1.5 kb and 1 kb fragments upstream of the ATG) were obtained by an Amplified Fragment-Length Polymorphism (AFLP) based PCR protocol as described (Ratet et al. 2010) and recombined in the Gateway vector pDONRP4-P1R according to the manufacturer's instructions (Invitrogen). Primers used for the amplification and cloning of the promoters are listed in **Appendix, Table 1**. Entry clones for the GUS ORF and the 35S terminator were obtained in the Gateway vectors pDONR221 and pDONRP2R-P3, respectively (Van de Velde et al. 2010). Entry clones were recombined in the binary vector pKm43GW (Karimi et al. 2005).



### 3.2.2. Stable transgenic GUS plant lines

Leaf explants from the *M. truncatula* line R108 were transformed using *Agrobacterium tumefaciens* according to the method described in Cosson et al. (2006). For GUS analysis, three independent T2 transgenic lines were each time analysed to exclude positional effects of the transgene insertion. No pattern variations were observed between independent lines. Untransformed plants and the constitutive GUS line pG3.3 (35S promoter fused to GUS) (Mondy et al. 2014) were used as negative and positive controls, respectively. For nodulation kinetics, R108 plants were cultivated on BNM agar plates and inoculated with OD<sub>600</sub>=0.1 suspensions of *S. meliloti* strain FSM-MA (Nagymihaly et al. 2017, submitted) which is an excellent symbiont of *M. truncatula* R108 forming numerous large, nitrogen fixing nodules (Kazmierczak et al. 2017).

### 3.2.3. GUS staining

Samples were collected at indicated time points and embedded in 6% agarose. Tissue sections of 70 µm were prepared with a Leica VT1200S vibratome. GUS staining was done as described (Vanstraelen et al. 2009) and was allowed to proceed for 1h. Overnight staining did not alter the expression patterns.

### 3.2.4. Pathogen assays

For all pathogen assays, plants were cultivated on perlite/sand (3/1 vol/vol) substrate and watered with a commercial nutrient solution. Six weeks old plants were transferred to a growth chamber with saturating humidity the day before the inoculations and stayed in these conditions all along the assay. *Dickeya dadantii* 3937, *Pseudomonas syringae* pv. *tomato* DC3000 and its *hrcC* derivative strain were cultivated at 30°C in LB medium. Inocula of OD<sub>600</sub>=0.1 were resuspended in 10 mM MgCl<sub>2</sub> and were syringe infiltrated in the terminal leaflet of 5-8 leaves per plant. Sterile 10 mM MgCl<sub>2</sub> solution was infiltrated as mock control. *Botrytis cinerea* strain B05.10 was cultivated on PDA medium (Amselem et al. 2011) at 20°C. Spores were collected in ½ potato dextrose broth with 0.01% Tween 20 and inocula were normalized to 10<sup>6</sup> spores/mL using a Malassez cell. Five microliter drops of mock/inoculum were put on 5 to 8 terminal leaflets per plant. Symptoms were scored at 1, 2 or 7 dpi and leaflets were collected for GUS staining. For wounding experiments, the terminal leaflet of 5-8 leaves per plant were pinched with forceps and collected 24 hours post wounding. Staining for all

infections or treatments was allowed for 24 hours in the GUS staining solution at 37°C. The leaflets were transferred to bleach to remove chlorophyll before photographing.

### **3.3. Antibodies and immunolocalization**

Part of the *NCR122* gene coding for the mature peptide was amplified from cDNA and cloned into the expression vector pBADgIII/A (Invitrogen). Recombinant proteins were purified according to the manufacturer's instructions and used for immunization of rabbits by a commercial service (Agro-bio). Immunolocalizations were done exactly as described before (Van de Velde et al. 2010). For the SYTO13 nucleic acid staining, nodules sections were incubated for 5 minutes with 1  $\mu$ M SYTO13 in water. Immuno- or SYTO13-stained sections were mounted in deionised water for confocal imaging. Fluorescence images were acquired at 1024x1024 pixels resolution with the confocal laser scanning microscope TCS SP2 from Leica, using 10X water-immersion and 63X oil-immersion objectives and Leica software. Images were processed with Adobe Photoshop for adjustment of contrast and brightness.

### **3.4. Preparation of plant material for flow cytometry**

#### **3.4.1. Seed sterilisation and germination**

*M. truncatula* wild-type A17 seed were scarified with sulphuric acid (95-98%) covering all seeds, for 8 min. After removing the sulphuric acid with a glass pipette, seeds were washed five times with large volumes of ice cold (to avoid heat stress and burning of the seeds) sterile water. Then seeds were immersed in 0,1% HgCl<sub>2</sub> solution for 30 min. Seeds were washed extensively with sterile water at least five times and seeds were germinated on 0,7% water-agar (agar with tap water, sterilized) plates up-side-down (agar layer on top) wrapped in aluminium foil and incubated at 4°C for two days and one day at room temperature until rootlets were 1 cm (Boisson-Dernier et al. 2001).

#### **3.4.2. Plant inoculation with *Rhizobium* and nodulation**

*Sinorhizobium medicae* WSM419 strain was grown in YEB medium (0.5% beef extract; 0.1% yeast extract; 0.5% Peptone; 0.5% sucrose; 0.04% MgSO<sub>4</sub>.7H<sub>2</sub>O; pH 7.5) supplemented with chloramphenicol (25  $\mu$ g/mL) at 30°C. Cultures were centrifuged and the bacterial pellets re-suspended in water at OD<sub>600</sub>=0.05 were used for plant inoculation. Seedlings were transferred to Perlite/Sand (3:1 v/v) for a week (24°C, photoperiod 16h Light – 8h Dark,

humidity 60%) and inoculated with 50 mL bacterial culture per pot, each pot containing 5 plantlets. Plants were watered with a commercial N-free fertilizer (PlantProd solution (N-P-K: 0-15-40) at 1 g/L; Fertil, Ref. 211.00). Nodules were harvested at 28 dpi in dry ice and kept at -80°C for subsequent nuclei sorting. For ChIP experiment nodules were collected on ice in 1% formaldehyde and crosslinking was performed under vacuum for 15 min. Formaldehyde was quenched with 125 mM glycine (5 min under vacuum) and nodules were washed twice with water, dried and stored at -80°C until further use.

### **3.4.3. Flow cytometry sorting of different ploidy level nodule nuclei**

Nodules were chopped with a razor blade in buffer (45 mM MgCl<sub>2</sub>, 30 mM trisodium citrate, 20 mM 3-(N-morpholino) propanesulfonic acid (MOPS), 0.1% triton X-100, pH 7.2-7.4., stored at 4°C) to break cells and release nuclei. For nuclear RNA isolation the sorting buffer was supplemented with 1 unit (U) ribonuclease inhibitor (RNasin, Promega) and for chromatin isolation with protease inhibitor cocktail (10 µL/mL) (P9599, Sigma). Suspensions were then filtered through 30 µm Celltrics filters (Partec, Germany). Nuclei were stained by 4',6-diamidino-2-phenylindole (DAPI) (5 µg/mL) and analysed using a Beckman MoFlow Astrios (Beckman Coulter, USA) flow cytometer with a 488-nm laser for scattering and 355-nm for DAPI excitation. Nuclei were gated in a DNA/side scatter plot (DAPI vs SSC) according to their DNA content and granularity and sorted from 2C up to 64C in 6 channels. Different ploidy level nuclei were collected at 4°C in 1,5 mL sterile Eppendorf tubes and transferred to -80°C until further RNA, DNA or chromatin isolation. 1 g nodules (5 mL) yielded 1,8x10<sup>5</sup> in 2C, 7x10<sup>5</sup> nuclei in 4C, 1,7x10<sup>5</sup> in 8C, 4x10<sup>4</sup> in 16C 1,3x10<sup>5</sup> in 32C and 1x10<sup>4</sup> in 64C. Yields of 2C and 64C nuclei were too low for further experimentation.

## **3.5. Nuclear RNA isolation, reverse transcription and quantitative PCR analysis**

### **3.5.1. Nuclear RNA extraction**

Nuclear RNA was extracted from 4C, 8C, 16C, 32C ploidy level nuclei (7x10<sup>5</sup>-5x10<sup>4</sup> nuclei per sample in sorting buffer) with Tri-Reagent (Sigma). 750 µL Tri-Reagent was added to 250 µL nuclei, homogenised, incubated 5 min at room temperature. Afterwards 200 µL chloroform was added, mixed, incubated 2 min at room temperature and centrifuged at 12.000 g for 15 min at 4°C. The aqueous phase with the RNA was removed and transferred to a new

tube. RNA precipitation was carried out by adding 20 µg glycogen as carrier (ThermoFisher) and 500 µL 100% isopropanol. Samples were incubated 10 min at room temperature and centrifuged at 12.000 g 10 min at 4°C. The supernatant was removed, the RNA pellet was washed with 1 mL 75% ethanol and centrifuged at 7500 g for 5 min at 4°C. After removing the supernatant, the RNA pellet was air dried and dissolved in 50 µL sterile water.

### **3.5.2. DNase treatment**

DNase treatment was carried out by TURBO DNase (ThermoFisher) to remove contaminating genomic DNA from the RNA sample prior to cDNA transcription. Reaction mixtures were composed as follows: 10x DNase Buffer 6 µL; RNase inhibitor (2U) 1 µL; TURBO DNase (2U) 1 µL; Nuclear RNA 50 µL; Nuclease-free water 2 µL in a total volume of 60 µL. Samples were incubated 30 min at 37°C.

To remove the DNase from the RNA sample, 290 µL sterile water and 350 µL phenol/chloroform/isoamyl alcohol (25:24:1 v/v) were added, samples were vortexed and centrifuged at 16.000 g for 15 min at 4°C. The aqueous phase (350 µL) was removed and mixed with 35 µL NaOAc 3 M (1:10 v/v), 20 µg glycogen and 875 µL (1:2.5 v/v) ice cold ethanol (ThermoFisher), kept 2h at -20°C and centrifuged at 16.000 g for 30 min at 4°C. After removal of the supernatant, the RNA pellet was washed with 500 µL 70% ice cold ethanol, centrifuged at 16.000 g for 5 min. The RNA pellet was air dried at room temperature and resuspended in 20 µL nuclease-free water.

### **3.5.3. Reverse transcription**

cDNA was synthesized using 500 ng nuclear RNA (measured by NanoDrop2000 Spectrophotometer, ThermoFisher) using SuperScript® VILO Master Mix (ThermoFisher). The reaction mixture was composed as follows: Nuclear RNA 20 µL; Nuclease-free water 18 µL; SuperScript Vilo Master Mix 2 µL in a total volume of 40 µL. The samples were incubated at 25°C for 10 min, At 42°C for 120 min and at 85°C for 5 min. The cDNA was diluted two times and used for quantitative PCR.

### **3.5.4. Quantitative PCR (qPCR)**

The synthesized cDNA was analysed by qPCR using PowerUp SYBR Green Master Mix (ThermoFisher). Relative quantification (Pfaffl 2001) was used to measure the expression level of a target gene compared to the constitutively expressed reference genes *Mtc27*

(Györgyey et al. 1991) and *40S RNA* (Van de Velde et al. 2006). The amplification efficiency of the primers was between 1.9-2.0. All primer sequences are provided in **Appendix, Table 2**.

qPCR mixes were composed as follows: PowerUP Sybr Green Master Mix (2x) 3,5 µL; Forward primer (10 µM) 0,5 µL; Reverse primer (10 µM) 0,5 µL; Nuclease-free water 1,5 µL; cDNA 1 µL in 7 µL total volume.

The StepOne Plus System (ThermoFisher) was used for the real-time PCR experiment. Two technical replicates of each reaction and a no template control (NTC) reaction were set up. The NTC reaction containing all reaction components (PowerUP SYBR Green Master Mix, primers, water) except the cDNA sample was used to identify eventual PCR contamination. Standard cycling conditions were used as follows: UDG activation\* 50°C 2 min hold; Polymerase activation 95°C 2 min hold and 40 cycles; Denaturation 95°C 15 sec; Annealing /extension 60°C 1 min. To check for nonspecific amplification dissociation/melting curves were also included with conditions as follows: Step 1 95°C 15 sec 1,6°C/s ramp increment rate; Step 2 60°C 1 min 1,6°C/s ramp increment rate; Step 3/dissociation 95°C 15 sec 0,3°C/s ramp increment rate; Step 4/UDG inactivation 95°C 10 min.

\*The reaction mixture contains uracil-DNA glycolase (UDG) which prevents carry-over contamination of PCRs if the contaminants contain uracils in place of thymines.

For relative quantitation, the  $\Delta C_t$  model was used (Perkin Elmer, Applied Biosystems):

$$\text{Relative expression} = 2^{-[\Delta C_t (\text{target}) - \Delta C_t (\text{reference})]}$$

$C_t$  is threshold cycle/crossing point obtained from the exponential phase of the qPCR, target is the gene of interest and reference is the constitutively expressed control gene.

### **3.5.5. Statistical analysis**

To determine significant expression differences of the target genes between each ploidy levels, one-way ANOVA with Student-Newman-Keuls *post hoc* (multiple comparisons) test of three independent biological replicates with two technical replications were performed (GraphPad Prism 6).

## **3.6. Genomic DNA extraction and Methylated DNA Immunoprecipitation (MeDIP)**

### **3.6.1. Genomic DNA extraction from sorted nuclei**

Genomic DNA was extracted from 4C, 8C, 16C, 32C nuclei ( $1,8 \times 10^6$ - $1,4 \times 10^5$  nuclei per sample,  $3 \times 10^5$  nuclei per mL) using the DNeasy® Blood and Tissue kit (Qiagen): 1 volume nuclei in sorting buffer and 1 volume lysis buffer (Buffer AL) containing Proteinase K (20 mg/mL) was incubated at 56°C for 10 min. After incubation, 1 volume ethanol was added to the lysed nuclei, samples were mixed by vortexing, transferred to the spin-column (1 column/4 mL nuclei with Buffer AL) and centrifuged at 8000 g for 1 min. The flow through was discarded and the spin column washed with 500  $\mu$ L AW1 Buffer, centrifuged at 8000 g for 1 min, washed again with 500  $\mu$ L AW2 Buffer, centrifuged at 20.000 g for 3 min. DNA was eluted in a clean Eppendorf tube with 50  $\mu$ L elution buffer (AE) for 2 min and centrifuged at 20.000 g for 1 min. This step was repeated and the two eluates were combined to obtain a final elution volume of 100  $\mu$ L. The DNA concentration was measured with a Qubit 2.0 Fluorometer (ThermoFisher) and the DNA was stored at -20°C in elution buffer for further use.

### **3.6.2. Shearing the genomic DNA**

Three hundred ng DNA in 100  $\mu$ L final volume were sheared to ~300 bp by sonication using a Covaris S2 ultrasonicator (Covaris, Inc.) as follows: duty cycle 10%, intensity 4, burst 200 for 40 sec and 2 cycles at 4°C.

### **3.6.3. MeDIP**

The methylated DNA was immunoprecipitated (IP) using the Methyl-CpG-binding domain protein 2 (MBD2) (NEBNext Microbiome DNA Enrichment Kit, New England Biolabs). 80  $\mu$ L MBD2-Fc protein with Protein A magnetic beads were used per 300 ng sheared DNA sample. To prepare the bead-protein mixture, 8  $\mu$ L MBD2-Fc and 80  $\mu$ L of Protein A Magnetic beads were placed on a rotating mixer for 10 min at room temperature, beads were washed on a magnetic rack with 1 mL cold 1x Bind/Wash Buffer and incubated on a rotating mixer for 3 min at room temperature. The washing step was repeated. Finally, the MBD2-Fc-Protein A bead mixture was resuspended in 80  $\mu$ L 1x Bind/Wash Buffer. To capture the methylated DNA 80  $\mu$ L sheared DNA were mixed with 20  $\mu$ L 5x Bind/Wash Buffer and

incubated on a rotating mixer for 15 min at room temperature with the 80  $\mu$ L MBD2-Fc-Protein A magnetic beads. Moreover, 1  $\mu$ L methylated DNA (met+) and unmethylated DNA (met-) from the MethylMiner Methylated DNA Enrichment Kit (Invitrogen) were also included in the IP reaction as internal controls. An aliquot of 8  $\mu$ L (10% of input) untreated sonicated DNA was processed in parallel and used as the total input DNA control. The bead-DNA mixture was washed with 1 mL cold 1x Bind/Wash Buffer and the captured methylated DNA was eluted using 150  $\mu$ L 1xTE and 40  $\mu$ g Proteinase K (ThermoFisher), vortexed and incubated at 65°C for 20 min, placed on a magnetic rack and the DNA was transferred into a new Eppendorf tube.

#### **3.6.4. Methylated DNA clean-up**

Purification of the enriched DNA was performed using Agencourt AMPure XP (Beckman Coulter). The IP DNA was cleaned-up and concentrated using Agencourt AMPure XP magnetic beads. 160  $\mu$ L IP DNA (obtained as in Section 3.6.3) was mixed with 289  $\mu$ L AMPure XP beads (1/1.8 v/v) and incubated for 5 min at room temperature on a rotating mixer. The 8  $\mu$ L unenriched control was diluted with 92  $\mu$ L nuclease-free water and mixed with 180  $\mu$ L AMPure beads and purified parallel to the IP sample. After incubation, the beads were transferred to a magnetic rack and washed twice with 400  $\mu$ L of freshly prepared 80% ethanol. The ethanol was removed and the magnetic beads with the purified DNA were air dried at room temperature. DNA was eluted from the beads with 50  $\mu$ L 1x TE at room temperature for 5 min and transferred to a new Eppendorf tube. The IP DNA was stored on -20°C and diluted 10 times before qPCR quantitation.

### **3.7. Chromatin isolation and chromatin immunoprecipitation (ChIP)**

#### **3.7.1. Isolation and shearing of the chromatin**

Cross-linked chromatin was isolated from sorted 4C, 8C, 16C, 32C nuclei ( $3 \times 10^6$ - $4 \times 10^5$  nuclei per sample) that were prepared from formaldehyde-fixed nodules (see Section 3.4.2). Prior to IP, sorted nuclei were centrifuged at 13.000 g 30 min and washed with 500  $\mu$ L ice cold phosphate buffered saline (PBS; 8 g/L NaCl; 0,2 g/L KCl; 1,44 g/L Na<sub>2</sub>HPO<sub>4</sub>; 0,24 g/L KH<sub>2</sub>PO<sub>4</sub>; pH 7,4; autoclaved), centrifuged at 13.000 g 15 min and resuspended in 110  $\mu$ L freshly prepared Nuclei Lysis Buffer (50 mM Tris-HCl, pH 8; 1% SDS; 10 mM EDTA; 100  $\mu$ L Protease inhibitor cocktail (10  $\mu$ L/mL) for plant cell extracts (P9599, Sigma)). Samples

were sonicated to obtain DNA fragments of 300-500 bp using a Covaris S2 ultrasonicator as follows: duty cycle 5%, intensity 4, burst 200 for 60 sec, rest 30 sec for 9 cycles at 4°C.

### **3.7.2. ChIP**

The 80 µL sonicated chromatin was diluted with 1300 µL ChIP Dilution Buffer (1,1% triton X-100; 1,2 mM EDTA; 16,7 mM Tris-HCl, pH 8; 167 mM NaCl) and immunoprecipitated with 5 µg anti-H3K27me3 or anti-H3K9ac antibodies (Millipore) overnight at 4°C with gentle shaking. Next, samples were incubated for 1 hr at 4°C with 40 µL of Protein AG UltraLink Resin (ThermoFisher). Protein AG Magnetic beads were washed in 1 mL of the consecutive wash buffers on a rotating device for 5 min at room temperature followed by separation from buffer on a magnetic rack. with 1 mL cold Wash Buffer 1. Beads were washed twice with 1 mL Wash Buffer 1 (150 mM NaCl; 0,1% SDS; 1% triton X-100; 2 mM EDTA; 20 mM Tris-HCl; pH 8; stored at 4°C), twice with 1 mL Wash Buffer 2 (500 mM NaCl; 0,1% SDS; 1% triton X-100; 2 mM EDTA; 20 mM Tris-HCl; pH 8; stored at 4°C), once with 1 mL Wash Buffer 3 (0,25 M LiCl; 1% sodium-deoxycholate; 1% triton X-100; 1 mM EDTA; 10 mM Tris-HCl; pH 8; stored at 4°C), once 1 mL Wash Buffer 4 (10 mM Tris-HCl; pH 8; 1 mM EDTA) and the IP chromatin was finally eluted twice with freshly prepared 250 µL Elution Buffer (1%SDS; 0,1 M NaHCO<sub>3</sub>) at room temperature for 15 min. 500 µL IP chromatin was reverse cross-linked with 200 mM NaCl at 65°C overnight and 1 h at 45°C with 8 mM Tris-HCl (pH 6.5), 10 mM EDTA and 20 mg Proteinase K (ThermoFisher). IP DNA was recovered using Agencourt AMPure XP beads (Beckman Coulter) (see Section 3.6.4) and analysed by qPCR. IgG antibody (Millipore) was used as a control to measure the background IP signal. An aliquot of 8 µL (10% of input) untreated sonicated chromatin was processed in parallel and used as the total input DNA control.

## **3.8. Quantitative real-time PCR analysis of MeDIP and ChIP DNA**

### **3.8.1. qPCR**

Input and IP DNA from the MeDIP and ChIP assays on the different ploidy nuclei were analysed by quantitative real-time PCR (qPCR). Quantitative PCR was performed using PowerUp SYBR Green Master Mix and a StepOne Plus Real-Time PCR System (Applied Biosystems) according to previously described protocol in section 3.5.4. Primers were designed to amplify 150-200 bp DNA fragments covering 1-kb upstream region (from ATG) of the gene



of interest, the gene body and the 0.4-0.6 kb upstream region. Primers are listed in **Appendix, Table 3**.

### **3.8.2. Data analysis**

The DNA recovery rate of MeDIP and ChIP experiment of the particular genomic region was calculated from Ct values of qPCR data and reported as a percentage of starting input DNA taking into account the dilution of the input (1:10 v/v).

$$\% \text{ (DNA-IP/ Total input)} = 2^{[(Ct \text{ (10\% input)} - 3.32) - Ct \text{ (DNA-IP)}]} \times 100\%$$

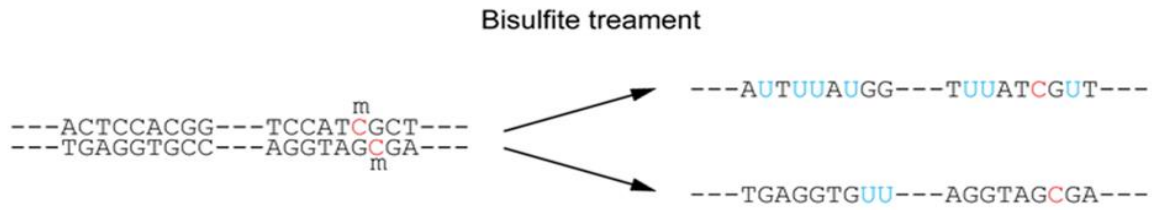
Here 2 is the amplification efficiency; Ct (DNA-IP) and Ct (10% input) are threshold values obtained from the exponential phase of qPCR for the MeDIP or ChIP DNA sample and input sample respectively; the compensatory factor (3.32) is used to take into account the dilution of the input (1:10 v/v). The recovery is the % (DNA-IP/Total input).

### **3.8.3. Statistical analysis**

To determine significance of differences in methylation level and chromatin modification of the target genes between each ploidy levels, two-way ANOVA with Tukey's multiple comparisons test of two independent biological replicates with two technical replicates were performed with P-value of <0.01 and <0.05 in case of the MeDIP and ChIP experiments, respectively (GraphPad Prism 6).

## **3.9.Reduced representation bisulphite sequencing (RRBS)**

To analyse 5-mC methylation of the DNA at single base resolution, the gold-standard method is bisulphite sequencing (BS-seq) (Frommer et al. 1992). Due to high costs of whole-genome BS-seq, Meissner et al. 2005 described a method called reduced representation bisulphite sequencing (RRBS) to reduce the amount of nucleotides needed to be sequenced. This technique combines enrichment for the areas of the genome that have high CpG content and bisulphite sequencing. The DNA fragments that comprise the reduced genome still include the majority of the regulatory elements, gene bodies, and repeated DNA sequences. Bisulfite treatment of purified DNA leaves methylated cytosines (mCs) unconverted while mediates the deamination of non-methylated cytosine (C) into uracil (U) which in subsequent PCR amplification is converted into thymine (T) (**Fig. 9**).



**Figure 9.** DNA fragments converted during bisulphite treatment. Nucleotides in blue are unmethylated cytosines (C) converted to uracils (U) by bisulfite, while red nucleotides are 5-methylcytosines (5-mCs) resistant to conversion.

Bisulfite reads are transformed into C-to-T on the forward- and to G-to-A on the reverse strand. Furthermore, in plants, the surrounding sequences are also taken into consideration to discriminate between Cs in CpG, CHG and CHH context.

### 3.9.1. Methyl-MidiSeq library construction

Sequencing library construction was performed by Zymo Research using 4C and 32C genomic DNA isolated as described previously in section 3.6.1. Three hundred ng of genomic DNA was digested with 40U of *Bfa*I, 40U of *Mse*I, and 80U of *Msp*I and the fragments produced were ligated to pre-annealed adapters containing 5'-methyl-cytosine (5mC) instead of cytosine (C). Adapter-ligated fragments were filled in and 3'-terminal-A extended, then purified using the DNA Clean & Concentrator kit (Zymo Research). Bisulfite treatment of the fragments was done using the EZ DNA Methylation Lightning kit (Zymo Research). PCR was performed and the size and concentration of the fragments were confirmed on the Agilent 2200 TapeStation, then sequenced on the Illumina HiSeq genome analyser.

### 3.9.2. Methyl-MidiSeq sequence alignments and methylated cytosine (mC) calling

The bisulfite-treated 4C and 32C DNA libraries were sequenced on Illumina HiSeq 2000 as paired-end (PE) 50 bp reads (45-50 million reads with median Phred score of > 30). Bismark software package was used to map and determine the methylation state of the cytosines (Cs) in each read (Krueger & Andrews 2011). Sequence reads were first transformed into fully bisulfite-converted forward (C -to- T) and reverse reads (G -to- A) before aligning them by Bowtie2 (Langmead & Salzberg 2012) to similarly converted versions of the JCVI *Medicago* Mt4.0v1 indexed reference genome. The number of mismatches in the induced alignment was then counted between the unconverted read and reference, ignoring cases in which a T in the

unconverted read is matched to a C in the unconverted reference. For a given read, the best alignment was kept. If there was more than one best alignment, the read was discarded as non-unique. To generate the reference methylome, Fisher's exact test was performed for each C site, which has a minimum coverage of five times. Mapping with seed length of 28 bp and mismatch of 1 bp resulted in bisulfite conversion rate of 83% each.

### **3.9.3. Detection of differentially methylated regions (DmRs)**

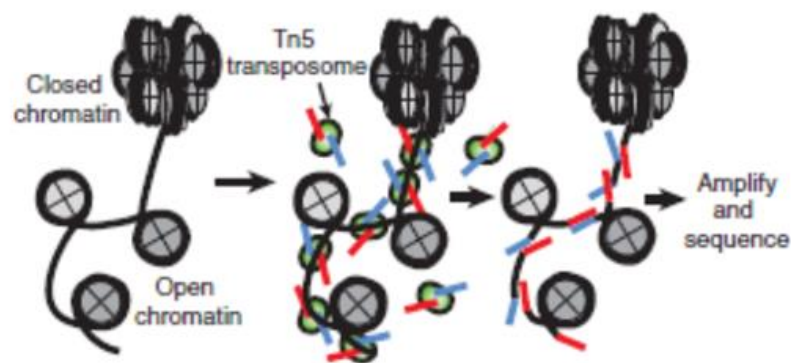
A hidden Markov model (HMM) based approach was used for *de novo* detection of differentially methylated regions (DmRs) between 4C and 32C samples (Saito et al. 2014). HMM-based detection of DmRs integrates identification of differentially methylated Cs (DmCs) between 4C and 32C and grouping them at neighbouring positions into contiguous DmRs with a framework called "comparative methylomics" (ComMet). ComMet uses automated adjustment of DmC chaining criteria and in addition, ComMet does not require biological replicates for DmR detection. The framework uses two modules where bisulphite-converted reads are aligned to a reference genome and the mC level is estimated as a ratio of C-C matches (mCs) by the count of all reads mapped at the same position and a module where neighbouring, differentially methylated Cs between 4C and 32C samples are grouped into a DmR according to the direction of the state transition: hypermethylated (UP), hypomethylated (DOWN) or no change (NoCh). ComMet detects DmRs based on log-likelihood ratio scores ( $\log[p(M)/p(N)]$ , where p is probability, M is change in methylation, N is no change in methylation. The log-likelihood ratio score represents a probability of the region with the corresponding state transition and reflects the difference in methylation between 4C and 32C. The higher the score the higher the methylation change between the two samples. The criteria for detecting DmRs are as follows: number of training iteration: 500, threshold of 0.01 for log-likelihood ratio scores.

### **3.10. Data of RNA-Seq coupled to laser microdissection of different nodule zone**

RNA-seq reads generated previously by Roux et al. 2014 were downloaded (<https://iant.toulouse.inra.fr/symbimics>) and mapped to the reference *Medicago truncatula* A17 (JCVI Mt4.0) genome with the CLC Genomics Workbench. To identify up- and downregulated genes, RPKM values (reads per kilobase of gene per million mapped reads) were calculated and compared between ZI (4C), ZIIId (8C), ZIIp (16C) and ZIII (32C).

### 3.11. Assay for transposase-accessible chromatin using sequencing (ATAC-Seq)

The compaction of nucleosomes into chromatin separates inactive genomic regions and maintains active regulatory elements accessible to the transcription machinery (Henikoff & Shilatifard 2011). As we have seen, on top of this physical compaction, a dynamic epigenetic code operates that includes DNA methylation, histone variants, and modification as well as TFs, chromatin remodelers, and non-coding RNAs that influence gene expression. Currently, the assay of transposase accessible chromatin (ATAC-seq) (Buenrostro et al. 2013) is used for genome-wide analysis of chromatin accessibility at regulatory elements. ATAC-seq uses hyperactive Tn5 transposase (Goryshin & Reznikoff 1998) to simultaneously cut and ligate adapters for high-throughput sequencing at regions of increased accessibility. The prokaryotic Tn5 transposase is loaded *in vitro* with Illumina adapters for high-throughput DNA sequencing and is able to integrate into *in vivo* open chromatin (marking active regulatory elements). The transposase thereby simultaneously fragments and tags the target genome with sequencing adapters (Fig. 10). The amplifiable DNA fragments are preferentially generated at locations of accessible chromatin, therefore acting as a probe for measuring chromatin accessibility genome-wide. With this approach, we are able to assess how chromatin accessibility changes during symbiotic cell differentiation genome-wide and at the level of individual genes.



**Figure 10. Schematic representation of the ATAC-seq method** (modified from Buenrostro et al. 2013). Tn5 Transposase (green), loaded with sequencing adaptors (red and blue), inserts between nucleosomes (grey) into open chromatin and generates sequencing-library fragments that can be PCR-amplified.

#### 3.11.1. Transposition and library preparation

The transposition was carried out according to a published protocol (Buenrostro et al. 2013). Unfixed, 4C, 8C, 16C and 32C ploidy level sorted nuclei ( $2.5 \times 10^6$ - $3.5 \times 10^5$ ) were

centrifuged at 2600 g for 20 min at 4°C and supernatant was removed. The pellet was resuspended in the transposase reaction mix (Nextera, Illumina): 25 µL 2x TD buffer, 2.5 µL Tn5 transposase and 22.5 µL nuclease-free water. The transposition reaction was carried out for 30 min at 37°C. Directly following transposition the sample was purified using a Qiagen MinElute kit and eluted in 11 µL Elution Buffer.

### **3.11.2. Library amplification**

Following purification, library fragments were amplified using 1x NEBnext PCR master mix and 1.25 µM of custom Nextera PCR primers 1 and 2 (**Appendix, Table 4**) using the following PCR conditions: 72°C for 5 min; 98°C for 30 sec; and 5 cycles at 98°C for 10 sec, 63°C for 30 sec and 72°C for 1 min. In order to stop PCR amplification before saturation, qPCR was used to determine the number of cycles needed for amplification. The full libraries were amplified for five cycles, an aliquot of the PCR reaction was taken and ran for 20 cycles to determine the additional number of cycles needed for the remaining 45 µL reaction. The libraries were purified using a Qiagen PCR cleanup kit yielding a final library concentration of ~30 nM in 20 µL. Libraries were amplified for a total of 10–12 cycles. During the ATAC-seq protocol, the size-selection step was avoided to maximize the library complexity.

### **3.11.3. Sequencing and data analysis**

The libraries were sequenced on the Illumina HiSeq as PE 50 base reads (100-145 million reads with median Phred score of > 30) following Illumina's instructions. Reads were mapped against JCVI Mt4.0v1 reference genome with CLC Genomics Workbench. Genes from different clusters were ordered into three quartiles according to the number of reads in the 1 kb upstream genic region. The first quartile (0-33th percentile) genes represent the low accessible genes while the third quartile (66-100th percentile) genes the highly accessible ones. Decreasing/increasing accessibility of the selected five genes indicates decreasing/increasing sequencing tag density, respectively in the 1 kb upstream region of a given gene compared between different ploidy levels. Analysis and visualization of sequencing tracks were performed in Genomic viewer IGV. To evaluate the association between expression and accessibility, Pearson correlation coefficients were determined between the RPKM values and ATAC-seq tag counts for individual genes at 4C to 32C ploidy levels. Random generated data sets with mean and variance identical to the experimental data set were used as a baseline control. Using Wilcoxon-test the correlations derived from experimental data were found to be

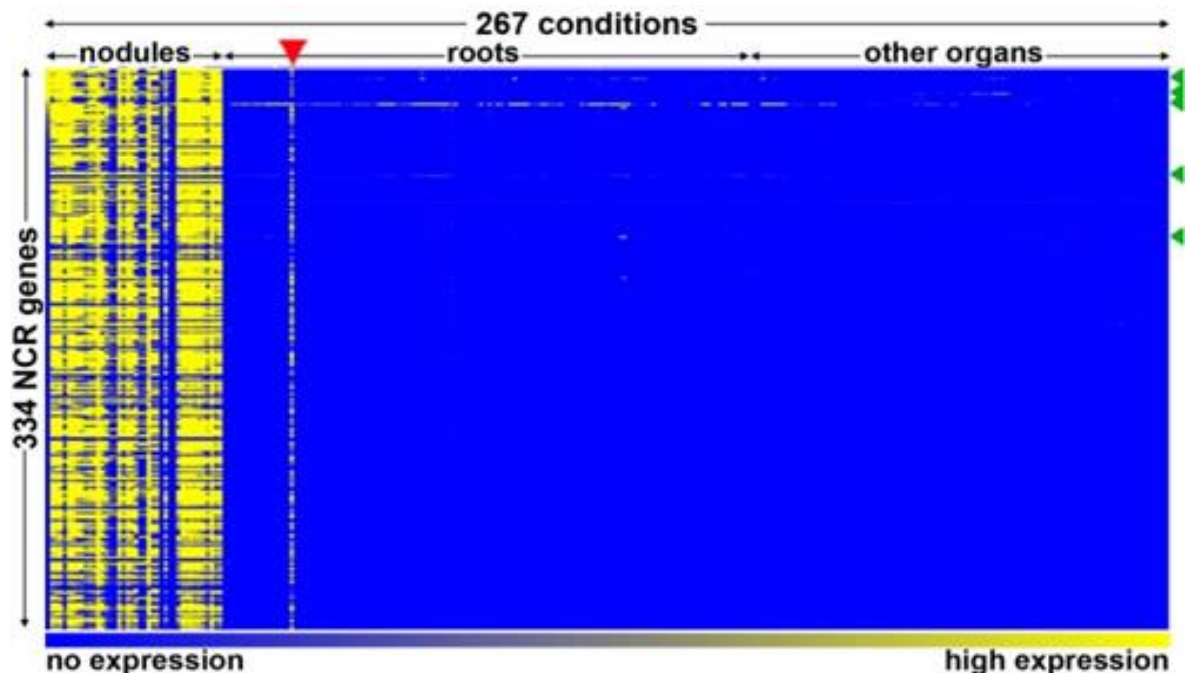
significantly higher than correlations generated from comparable randomized samples ( $P < 0.0001$ ), indicating positive correlation of gene expression and chromatin accessibility in the case of late genes.

## 4. RESULTS

### 4.1. Extreme specificity of *NCR* gene expression in *Medicago truncatula*

#### 4.1.1. Global analysis of *NCR* gene expression

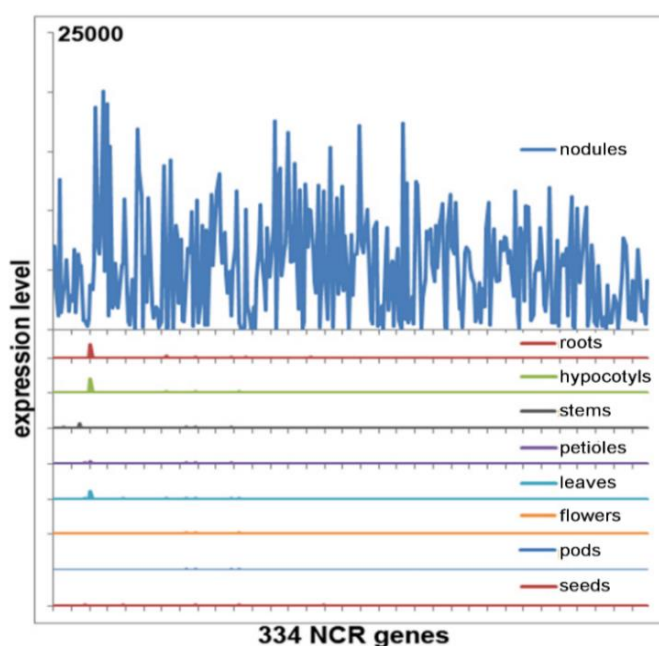
The Affymetrix *Medicago* GeneChip microarray probe-sets and the expression data of the *NCRs* (Mergaert et al. 2003; Alunni et al. 2007) were downloaded from the *Medicago truncatula* Gene Expression Atlas MtGEA (version 3) (He et al. 2009; Benedito et al. 2008). The expression pattern of 334 *NCRs* in 267 different experimental conditions (254 from MtGEAv3, 5 unpublished conditions, 3 from Jayaraman et al. 2014 and 5 from Limpens et al. 2013) are summarised in the heat map of **Fig. 11**. The transcriptome compendium is mostly derived from the *M. truncatula* line A17 but also contains data from the R108 and F83005.5 plant lines, although all experiments discussed here were obtained with the A17 line. The compendium covers the plant's major organs (root, nodule, leaf, petioles, stem, bud, flower, seed, pod), various kinds of abiotic- and biotic stresses and specific cell- and tissue types.



**Figure 11. Heat map of *NCR* gene expression in the MtGEA.** The heat map shows the expression of 334 *NCR* genes (rows) in 267 experimental conditions (columns). Experiments are ordered as indicated above the columns. The color scale bar indicates the expression from background level (blue) to maximum level (yellow). The red arrowhead indicates the mycorrhizal sample that is contaminated with nodules. The green arrowheads (on the right) locate the *NCR* genes with relaxed specificity; from top to down: *NCR247*, *NCR235*, *NCR122*, *NCR218* and *NCR077*. The dataset of the heat map is provided in **Appendix, Supplementary information 1**.

The experiments are organized in three major groups: nodule samples, root samples and samples of different other plant organs (**Fig. 11, and Appendix, Supplementary information 1**, [https://static-content.springer.com/esm/art%3A10.1186%2F1471-2164-15-712/MediaObjects/12864\\_2014\\_6439\\_MOESM1\\_ESM.xlsx](https://static-content.springer.com/esm/art%3A10.1186%2F1471-2164-15-712/MediaObjects/12864_2014_6439_MOESM1_ESM.xlsx)). The heat map reveals instantly an obvious global expression pattern; the nearly complete 334 *NCR* gene-set is only expressed in nodules except for one experiment marked with the red arrowhead in **Fig. 11**, which corresponds to a sample annotated as a mycorrhizal root sample. The *NCR* expression in mycorrhizal roots was inconsistent with previous results (Mergaert et al. 2003) and also with other mycorrhizal samples in the MtGEA in which the *NCR* genes are not expressed raising the possibility that the mycorrhizal roots might have nodules. Moreover, nodulation-specific genes were found to be active in this sample as well.

Previously, *NCR* expression beside symbiotic conditions was tested by EST analysis (Mergaert et al. 2003) and with microarrays (Tesfaye et al. 2013) indicating the absence of expression. In the MtGEA database besides nodules, 8 other plant organs were included for testing *NCR* expression (**Fig. 12**) and in none of these plant organs were *NCR* genes expressed except for the five genes with relaxed expression.

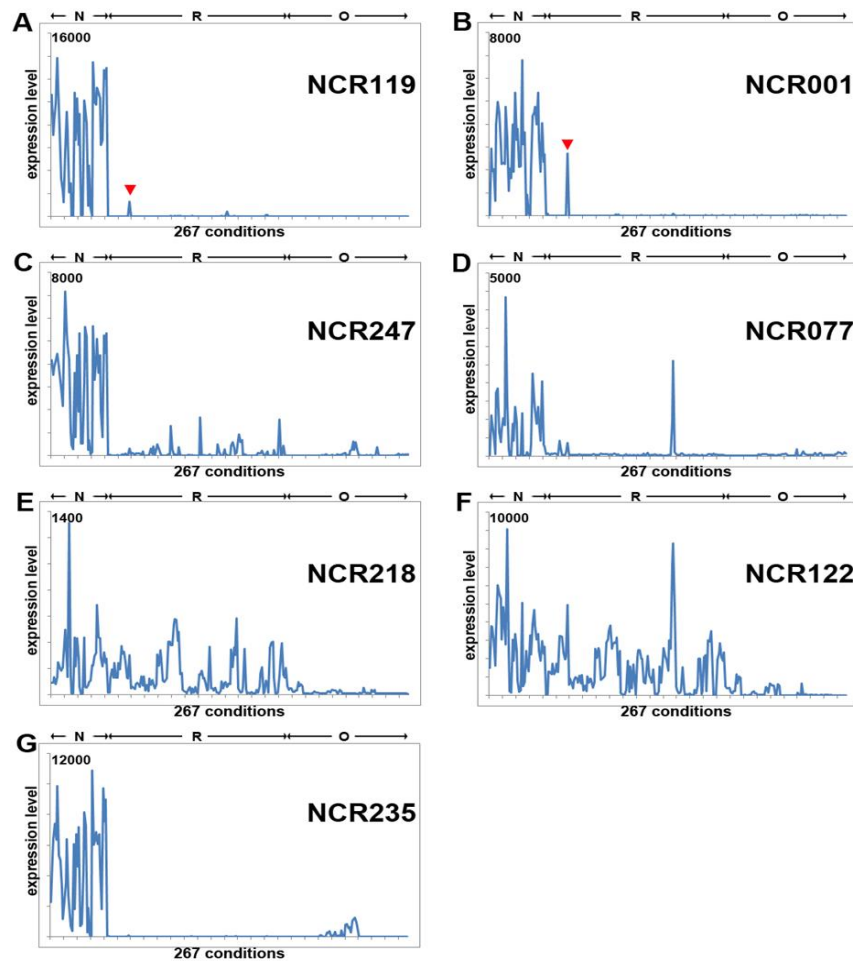


**Figure 12. *NCR* expression in plant tissues.** The expression pattern of 334 *NCR*s is shown for nodules, roots, hypocotyls, stems, petioles, leaves, flowers, seed pods and seeds. The graphs for all organs are at the same scale (maximum hybridization signal value is 25,000). The small peaks that can be seen in some of the organ samples correspond to the *NCR* genes with relaxed expression (Fig. 11).

Analysing the expression profiles of individual *NCR* genes revealed very high expression level in the nodule and only background signals in other experiments (examples are in **Fig. 13A**,



*B*). Such expression profile is typical for the large majority of the *NCR* genes (**Appendix, Supplementary information 1**), however, five exceptions were discovered with atypical, relaxed nodule-specific expression patterns (**Fig. 11, green arrowheads, Fig. 13C-G**). *NCR247* and *NCR077* are mainly active in nodules but are also weakly expressed in other conditions. The *NCR247* is expressed in different root-, stem- and aerial samples although at much lower levels than in nodules (**Fig. 13C**). It is not evident from the available information what signals activate its expression.



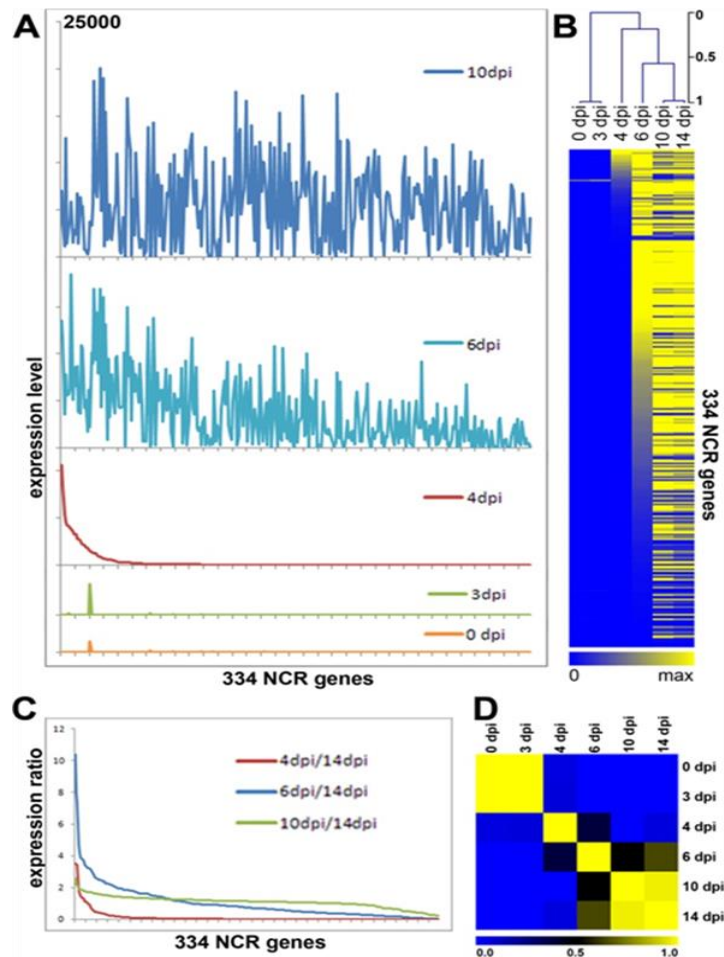
**Figure 13. Expression profile examples of seven *NCR* genes.** (A-G) The expression pattern of seven *NCR* genes in 267 experiments (x-axis) in nodules (N), in roots (R) and in other organs (O). The experiments are ordered as in Figure 11. The y-axis is the expression level according to the fluorescence hybridization signal and is scaled according to the maximum hybridization signal as indicated on the top of the axis. The red arrowheads in panels A and B indicate the mycorrhizal sample that is contaminated with nodules.

*NCR077* has a higher than usual background level, possibly because of a less specific probe-set, but the gene is also expressed in some mycorrhizal samples (**Fig. 13D**) including a laser-capture microdissection (LCM) sample of arbuscule-containing cells (Gaude et al. 2012). On the other hand, *NCR218* and *NCR122* have a completely relaxed specificity and they are

expressed to similar levels in nodules and in other conditions, mostly roots (**Fig. 13E, F**). *NCR122* is also particular in that sense, that it is expressed in the uninfected but not in the infected nodule cells. Expression of *NCR218* has not been investigated on cellular level yet. *NCR235* expression is similarly nodule specific as most other *NCR* genes except for a weak expression in stems and shoots, which is about 10- to 100-fold lower than in nodules (**Fig. 13G**). Thus except for these five genes, the complete set of tested *NCR* genes is only expressed in nodules and in none of the other conditions that are present in MtGEA.

#### 4.1.2. Spatio-temporal expression of *NCR* genes in nodules

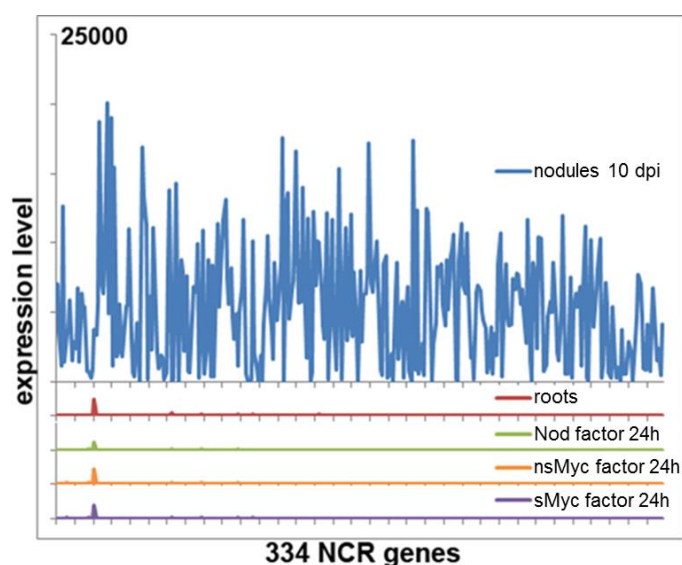
The MtGEA compendium contains 42 different nodule samples including samples of wild type nodules harvested at different days post inoculation (dpi) with *S. meliloti* and thus at different stages of nodule development (**Fig. 14**).



**Figure 14. Successive activation of *NCR* genes during nodule development.** (A) Expression level of the 334 *NCR* genes (x-axis) at 0, 3, 4, 6 and 10 dpi. The y-axis scale is the same for all graphs and is 25,000 at maximum. (B) Heat map of the same expression data as in (A) and 14 dpi. (C) The ratios of *NCR* expression levels at 4 dpi, 6 dpi or 10 dpi compared to 14 dpi. For each time point, the genes were

ordered from high ratio to low ratio. **(D)** Heat map of the Pearson correlation of the *NCR* expression profiles at 0, 3, 4, 6, 10 and 14 dpi.

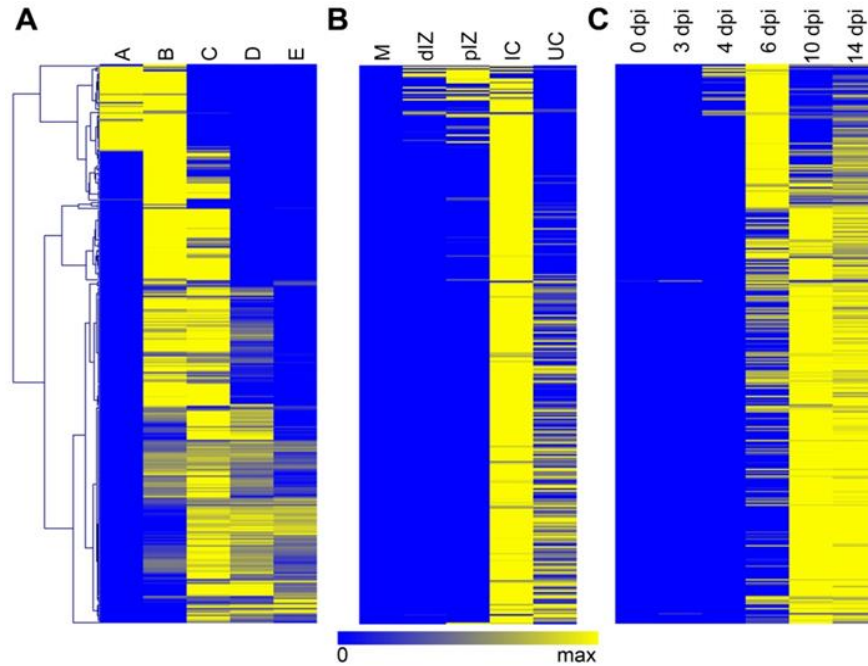
In samples of uninoculated roots and nodule primordia of 3 dpi, none of the *NCR* genes are activated (**Fig. 14A, B**). The nodule transcriptomes at 0 dpi and 3 dpi have high correlation coefficient (**Fig. 14D**). This suggests that *NCR*s have no function in the early stages of the interaction. In agreement with this, treatments with Nod- or Myc factors also do not activate *NCR* gene expression (**Fig. 15**). Myc factors are similar to the lipochitoooligosaccharide Nod factors and are produced by mycorrhizal fungi symbionts, inducing mycorrhiza formation (Maillet et al. 2011).



**Figure 15. *NCR* gene expression in response to Nod factors and Myc factors.** The expression pattern of 334 *NCR* probe-sets (x-axis) is shown from top to bottom as follows: for nodules 10 dpi, untreated control roots, 24h 10-8M Nod factor treatment, 24h 10-7M non-sulfated Myc factor treatment (nsMyc) and 24h 10-8M sulfated Myc factor treatment (sMyc) (Czaja et al. 2012). The y-axis represents the strength of the hybridization signal and the graphs for all treatments are at the same scale (maximum 25,000).

At 4 dpi of rhizobia induced nodulation a subset of ~70 genes is already activated, some of them already at their maximal expression level while others are only partially induced (**Fig. 14**). At 6 dpi, most genes are activated but many of them not yet at their maximal level (**Fig. 14A-C**). At nodulation time points of 10 dpi and 14 dpi the *NCR* transcriptome is fully activated. They have a correlation coefficient close to 1 (**Fig. 14D**). Together this pattern reveals that *NCR* genes are activated in different waves and indicates a link between *NCR* activation and the progression of bacterial infection in the incipient nodules.

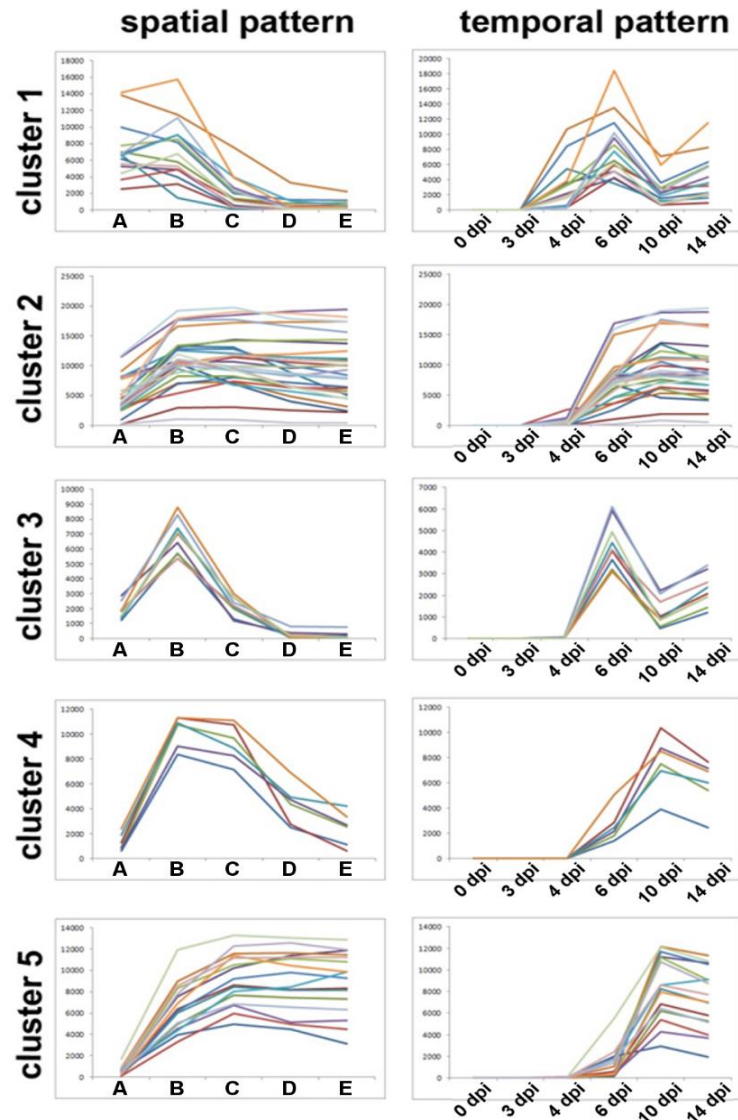
Nodules of *M. truncatula* are of the indeterminate type and therefore, mature nodules are organized in well-defined histological zones (**Fig. 3**). This nodule structure suggests that the temporal *NCR* expression profiles could correlate with a spatial expression pattern in the nodule tissues (**Fig. 16**). To test this possibility, we performed transcriptome analysis of 4-weeks-old nodules that were hand-sectioned in five different parts (see Section 3.1.2).



**Figure 16. Spatio-temporal expression of the *NCR* gene family in nodules.** (A) Heat map and hierarchical clustering of the 334 *NCR*s according to their spatial expression in nodules obtained by hand-dissection of nodules in this study. A: meristem, zone I, B: infection and differentiation zone II, C: interzone II-III, D: nitrogen-fixing zone III, E: senescence zone IV. Clustering was performed using Pearson correlation. (B) Comparison of data (A) to the corresponding profiles obtained by LCM coupled to Affymetrix microarray analysis described by Limpens et al 2013. M corresponds to meristem, dIZ to the distal zone II or infection zone, pIZ to the proximal zone II, IC to infected cell and UC to uninfected cell. (C) The corresponding *NCR* expression profiles in function of nodule age (dpi) (same data as on Fig. 14). For all data sets, the expression patterns in the different dissected nodule zones are expressed in percentage from their total. The *NCR* genes are ordered identically in all three panels according to a hierarchical clustering of the dataset from panel (A).

These five samples correspond to the tissues from the most apical part of the nodule to the most proximal part containing the oldest symbiotic cells. Sample A is the meristem and the underlying few cell layers of post-meristematic cells, which start the infection and differentiation process. Sample B corresponds mainly to the infection and differentiation zone II. Sample C corresponds to the interzone II-III. Sample D is the nitrogen-fixation zone III, easily characterized by its pink color due to the accumulation of leghemoglobin and finally sample E is the senescence zone IV that is recognized by its green color resulting from the

accumulation of biliverdin, a product of the catabolism of leghemoglobin-derived heme (**Fig. 16A**). It should be noted that each of these hand-dissected samples is enriched for the indicated zone but can contain cell layers from the adjacent zones as well. Cluster analysis of the *NCR* abundance profiles in these five samples distinguishes groups of *NCR* genes that have preferential expression in defined nodule zones and that are sequentially activated from nodule apex to proximal tissues (**Fig. 16A and Fig. 17**).



**Figure 17. Representative clusters of spatial and temporal *NCR* expression profiles.** Five representative spatial (*Left*) and temporal expression patterns (*Right*) are shown for clusters 1 to 5.

Genes in cluster 1 display the highest expression in the sample A, while in sample B are still high and then decline rapidly. Cluster 2 genes are already activated in sample A but their expression increases in sample B and remains high in the whole nodule. Cluster 3 genes have a very sharp peak in sample B due to a specific high expression here that is also visible at the 6

dpi sample, while in sample A and C their expression is low and in the samples D and E close to zero. The genes of cluster 4 have a similar profile but are expressed in a broader zone with high levels of expression in samples B and C. Finally, cluster 5 genes are absent in sample A, partially activated in sample B and then fully activated in sample C, D and E.

The transcriptome of specific nodule tissues and cells was also obtained by laser-capture microdissection (LCM) in a French lab (Limpens et al. 2013) and their data are publicly available in the MtGEA for further analysis and comparison. The transcriptome was analyzed in nodule meristems (M), the distal (dIZ) and the proximal infection zone (pIZ), the infected cells of the fixation zone (IC) and the uninfected cells of the fixation zone (UC) (**Fig. 16B**). Although, the type of samples is not entirely overlapping, a good correspondence can be observed between the LCM dataset and the hand-dissected dataset (**Fig. 16A, B**). The LCM dataset shows that *NCR* genes are not expressed in the meristem which is free of rhizobia (**Fig. 16B**) and that their activation is coinciding with rhizobia infection. Moreover, it also indicates that activation of the early *NCR* genes in the hand-dissected sample A is coming from the presence of cells from the infection zone. *NCR* genes were reported to be expressed in infected cells (Kondorosi et al. 2013). Unexpectedly, the LCM dataset revealed a relatively high expression of several *NCRs* in the uninfected cells including *NCR001*, *NCR084*, *NCR035* and *NCR247* (**Fig. 16B, sample UC**). This sample, however, could contain some infected cells because *in situ* hybridization, promoter-GUS- or mCherry fusion of these genes or immunolocalization of peptides demonstrated their specific expression in the infected symbiotic nodule cells but not in uninfected cells (Van de Velde et al. 2010; Farkas et al. 2014; Mergaert et al. 2003). Possibly, the signal in the sample of the uninfected cells is the result of contamination with infected cells or due to background hybridization coming from the RNA amplification procedure used with the very low amount of RNA obtained from the LCM samples (Limpens et al. 2013). Nevertheless, when comparing the expression level of all *NCRs* in the uninfected and infected cells, only two genes had a significantly higher expression in the uninfected cells. Those two genes are *NCR218* (9,6 fold higher; t-test  $P=0,002$ ) and *NCR122* (3,2 fold higher; t-test  $P=0,018$ ) which are the only 2 *NCRs* that are consistently expressed to high levels in roots (**Fig. 13E, F**).

When matching the spatial- and temporal *NCR* expression patterns, good correspondence can be observed: genes expressed in the apex are fully activated in the early stages of nodule development at 4 or 6 dpi, while genes expressed in the proximal tissues are activated in later stages of nodule development at 10 and 14 dpi (**Fig. 16C**). The correlation between the spatial- and temporal pattern of *NCR* expression is obvious when considering the

*NCR* gene clusters representing the major expression profiles (**Fig. 17**). For example, genes of clusters 1 and 2 (**Fig. 17**) are expressed in the most apical part of the nodule and this correlates well with their early activation at 4 or 6 dpi in the temporal pattern. On the other hand, genes in cluster 5 have maximum expression in sample C and this corresponds with their expression later in nodule development at 10 dpi (**Fig. 17**).

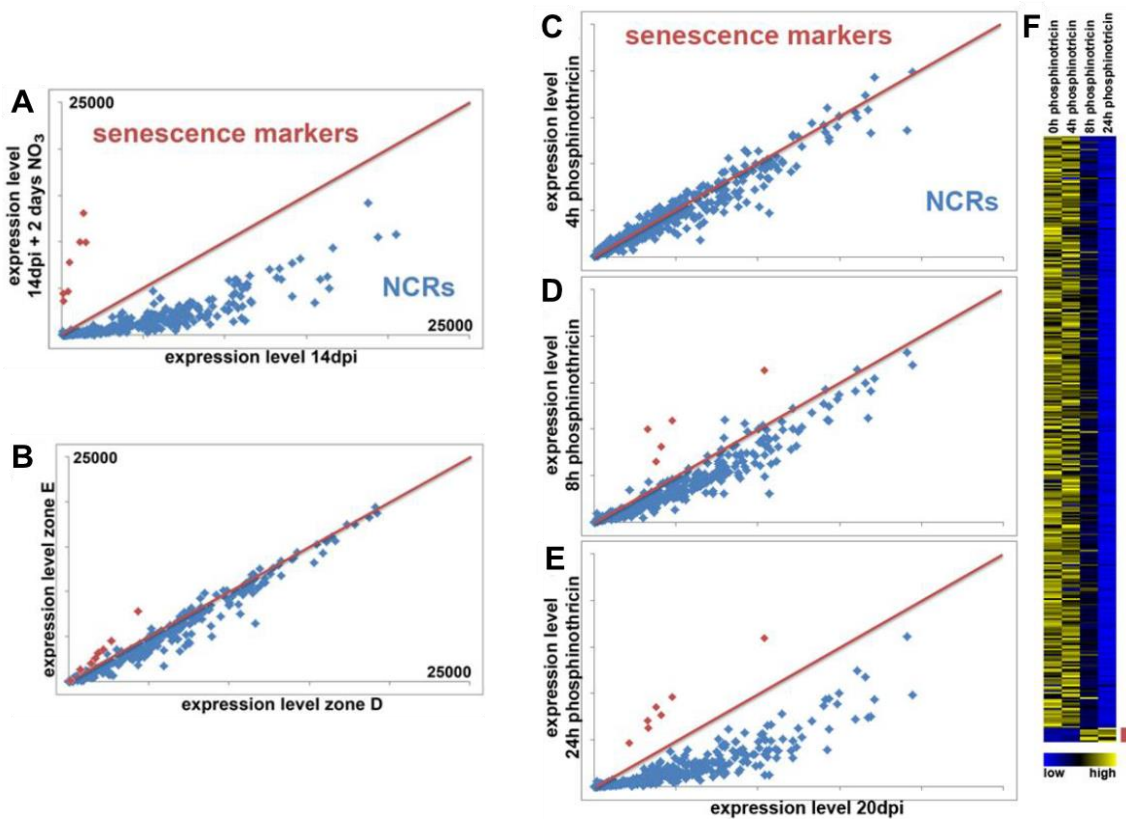
Together these spatio-temporal patterns reveal that the *NCR* genes are activated in different waves, in agreement with our previous results that identified two key points in nodule development associated with major transcriptional activation, one at the formation of symbiotic cells and another one when bacteroids differentiate (Maunoury et al. 2010). Nevertheless, the present analysis is refining this description and shows that *NCR* genes are activated in at least three waves and they can be distinguished by the maintenance or the decline of their expression in the older nodule cells. These analyses of *NCR* expression during nodule development are also in very good agreement with data from a recently published study (Roux et al. 2014) that used LCM of nodule zones coupled to RNA-Sequencing (see Section 3.1.1; **Fig. 28D**) and my RT-qPCR analysis of selected *NCR* genes in purified nodule nuclei with different ploidy levels which represent distinct stages of symbiotic nodule cell differentiation (**Fig. 28F**).

#### **4.1.3. *NCR* genes are not directly involved in nodule senescence**

Senescence is a natural process activated in aging nodules at 30-35 dpi involving a complete digestion of the bacteroids and later the symbiotic host cells (Van de Velde et al. 2006). Because NCRs resemble to the defensin-type of antimicrobial peptides (Mergaert et al. 2003) and because they have *in vitro* and *in vivo* antimicrobial activity (Tiricz et al. 2013; Van de Velde et al. 2010; Haag et al. 2011; Farkas et al. 2014) it was feasible to predict that part of the NCR family could be involved in the killing of the rhizobia during nodule senescence. This process is associated with the transcriptional activation of genes that are involved in the digestion of macromolecules and the remobilization of the liberated nutrients (Van de Velde et al. 2006). Nodule senescence can be induced artificially by nitrate supply to roots, which will prefer this new nitrogen source over the energy costly nitrogen fixation or alternatively, by keeping plants in the dark or applying a herbicide, cutting off the nutrient supply to nodules.

Another opportunity in the MtGEA database to test the effect of senescence on *NCR* expression is provided by samples corresponding to a treatment of *M. truncatula* with nitrate and the herbicide phosphinothricin (*a.k.a.* BASTA) which also induces nodule senescence (Seabra et al. 2012) (**Fig. 18A, C-E**). To test whether *NCR* genes are induced by senescence,

nodules at 14 dpi were compared with nodules of 14 dpi followed by 2 days of nitrate supply (**Fig. 18A**). Beside the *NCR* genes, 8 senescence marker genes were included in the analysis (**Appendix, Supplementary information 1**) (Van de Velde et al. 2006). The downregulation of *NCR* gene expression is very rapidly detectable, within 4 hours of nitrate application (Cabeza et al. 2014) or 8 hours of phosphinothricin treatment (**Fig. 18C-E**). Both treatments induced senescence as indicated by the strong induction of the senescence marker genes. This conclusion is also confirmed by comparing the hand-dissected nodule samples D and E, which are enriched for the nitrogen fixation zone III and the senescence zone IV respectively. None of the *NCRs* are significantly (t-test,  $P < 0.05$ ) higher expressed in sample E (**Fig. 18B**).



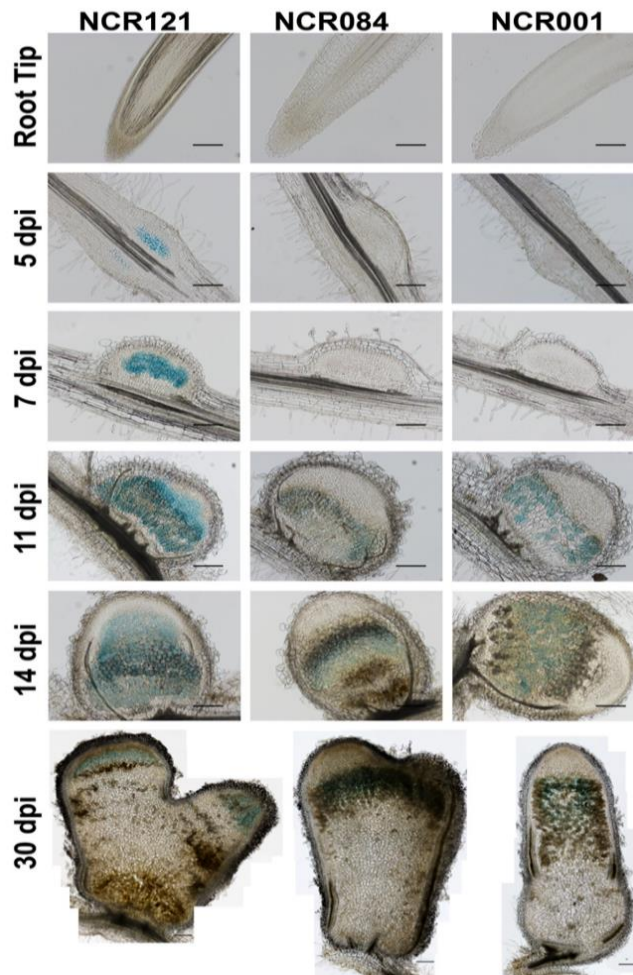
**Figure 18. *NCR* expression during nodule senescence.** (A) Scatter plot of the expression of 334 *NCRs* (blue) and 8 senescence marker genes (red) in 14 dpi nodules (x-axis) and 14 dpi nodules treated with nitrate for 2 days (y-axis). (B) Scatter plot of the expression of 334 *NCRs* (blue) and 8 senescence marker genes (red) in hand-sectioned nodule zone D (x-axis) and in hand-sectioned nodule zone E (y-axis). (C-E) Scatter plot of gene expression levels in 20 dpi nodules compared to 20 dpi nodules 4h (C), 8h (D) and 24h (E) post treatment with the herbicide phosphinothricin. The scale of all axes is the same and is 25,000 at maximum. The red line in the three graphs indicates a ratio of 1 between the two conditions. The probe-sets for the senescence marker genes, encoding cysteine proteinases, a chitinase, a nuclease a nucleoside transporter and a metal-nicotinamide transporter, are provided in **Appendix, Supplementary information 1**. (F) Heat map of *NCR* and senescence markers expression before and 4, 8 and 24 h post phosphinothricin treatment. Expression levels vary from low (blue) to high (yellow). The red bar next to the heat map indicates the location of the senescence marker genes.



All the *NCR* genes are reduced in expression by the senescence-inducing treatments, suggesting that none of the *NCR* genes has a direct role in senescence. The down-regulation of the *NCRs* most likely reflects the shut-down of the symbiotic process.

#### 4.1.4. Promoter-GUS analysis and immunolocalization of selected *NCRs* in nodules

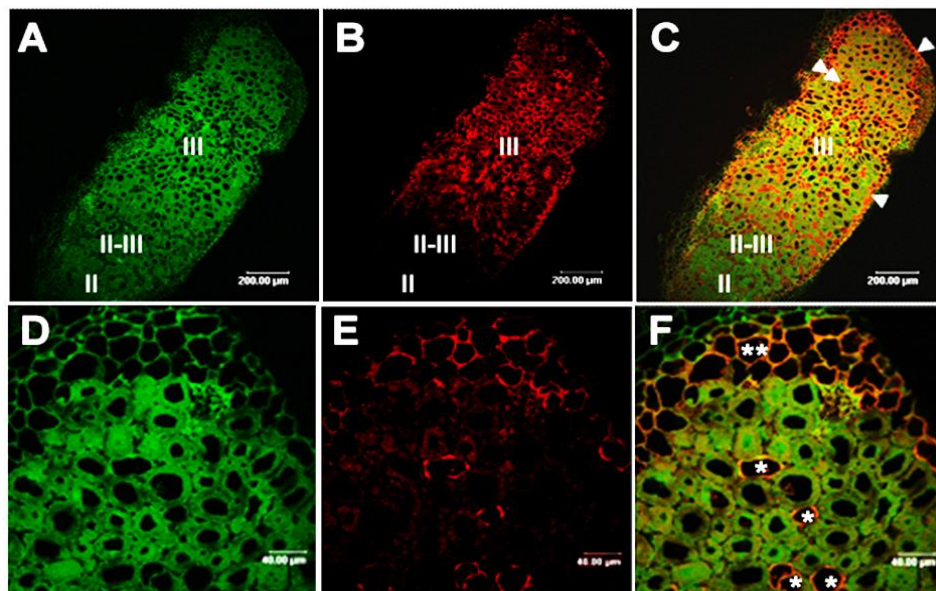
In order to confirm the expression data from MtGEA, stable transgenic *M. truncatula* R108 lines generated carrying promoter-GUS fusion constructs for three different *NCR* genes of different temporal expression classes (**Fig. 19**). *NCR001* is not activated before the late stages of the nodule formation (**Fig. 17, cluster 5**), while *NCR084* is slightly induced in early time points (4 dpi) and fully activated at the mature stage of the nodule (**Fig 17. cluster 2**) and finally *NCR121* is an early gene, which is already fully activated at 4 dpi (**Fig 17, cluster 1**).



**Figure 19. Promoter-GUS analysis of *NCR* genes in nodules.** Sections (70 µm) of root tips and nodules of different ages of transgenic plants carrying promoter-GUS fusions for the *NCR121*, *NCR084* and *NCR001* genes were stained for GUS activity observed as blue color. Scale bars are 50 µm.

The transgenic plants were inoculated with *S. meliloti* strains Sm1021 or with FSM-MA. GUS expression in the three different transgenic lines was not detected in root tips or other root parts (**Fig. 19**). In agreement with its temporal regulation during nodulation, *NCR121* expression was induced in young nodule primordia as early as 5 dpi and remained expressed throughout the experiment in the entire infection zone II and the fixation zone III of mature nodules (**Fig. 19**). *NCR084* expression was detected from 11 dpi on and was mainly confined to the proximal infection zone, the interzone II-III and to the distal part of the fixation zone III (**Fig. 19**). *NCR001* expression was detectable from 11 dpi in the developing fixation zone III and its expression extends in the following days as the fixation zone is growing (**Fig. 19**). All three genes are only expressed in the symbiotic nodule cells. In older nodules, at 30 dpi, displaying a senescence zone, *NCR* expression was never detected in the senescing tissues, nor was their expression enhanced in the proximal fixation zone adjacent to the senescent tissue (**Fig. 19**), confirming that *NCR* genes are not involved in the senescence process. Overall, the temporal and spatial promoter-GUS expression patterns are in very good agreement with the expression profiles deduced from MtGEA.

The particular expression pattern of *NCR122* with its relaxed tissue specificity (**Fig. 13F**) and its apparent expression in the uninfected nodule cells prompted us to analyze the localization of the NCR122 peptide in nodules with anti-NCR122 antibody (**Fig. 20**).



**Figure 20. Immunolocalization of NCR122 in nodule sections.** Sections of nodules were stained with the DNA label SYTO13 (green) to reveal the bacteria (**A**, **D**) and were immunolabelled with the anti-NCR122 antibody (red) (**B**, **E**). (**C**, **F**) Overlays of (**A**, **B**) and (**D**, **E**) respectively. (**D-F**) Enlargement of the top part of the nodule shown in (**A-C**). Arrowheads or double asterisks mark NCR122-labeled cells in the nodule cortex and the double arrowhead or asterisks mark uninfected cells of the nitrogen

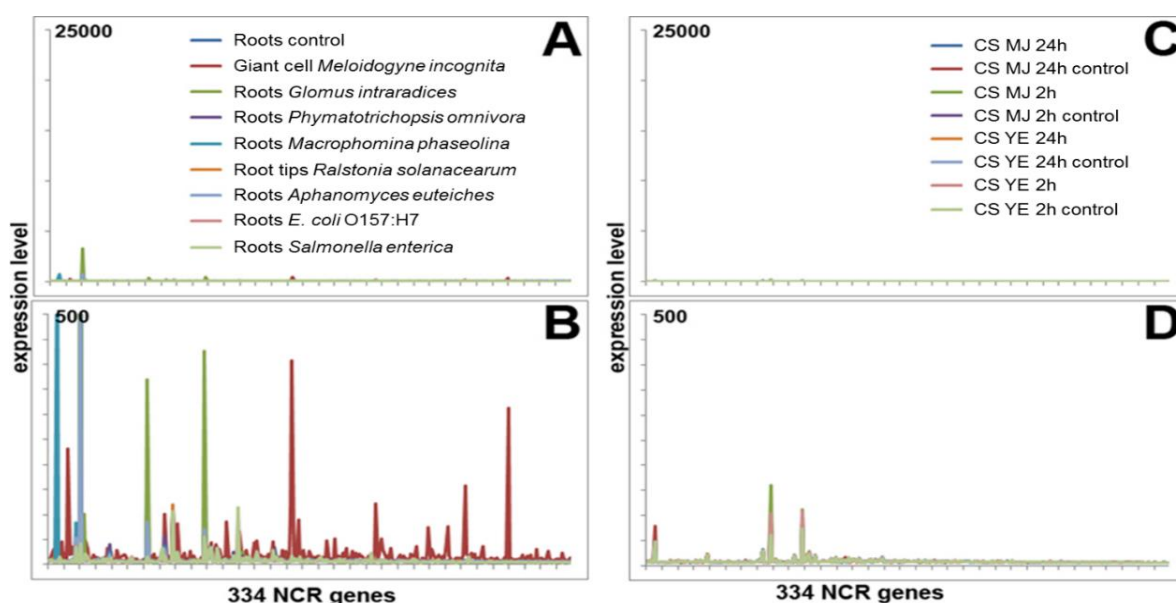
fixation zone III labeled by NCR122. II is the infection zone; II-III is the interzone. Scale bars are 200  $\mu\text{m}$  in panels A-C and 40  $\mu\text{m}$  in panels D-F.

Immunolocalization of the peptide revealed indeed a specific presence of NCR122 in the uninfected cells of the central zone of a mature nodule as well as in the uninfected cortical cells of the nodule (**Fig. 20**). Together with the transcriptome data, this indicates that NCR122 and most likely NCR218 are the only NCR peptides that are specific to uninfected root and nodule cells.

#### 4.1.5. NCR genes are not expressed after biotic and abiotic stress

NCRs resemble the defensin-type of antimicrobial peptides and defensins are often strongly expressed in “infection-sensitive” organs like flowers or seeds. The complete lack of expression of NCRs in these organs (**Fig. 12**) suggests that they probably do not have a defensive function in these organs. Nevertheless, many non-NCR defensin-like genes were found to be expressed in seeds, potentially involved in their protection (Tesfaye et al. 2013).

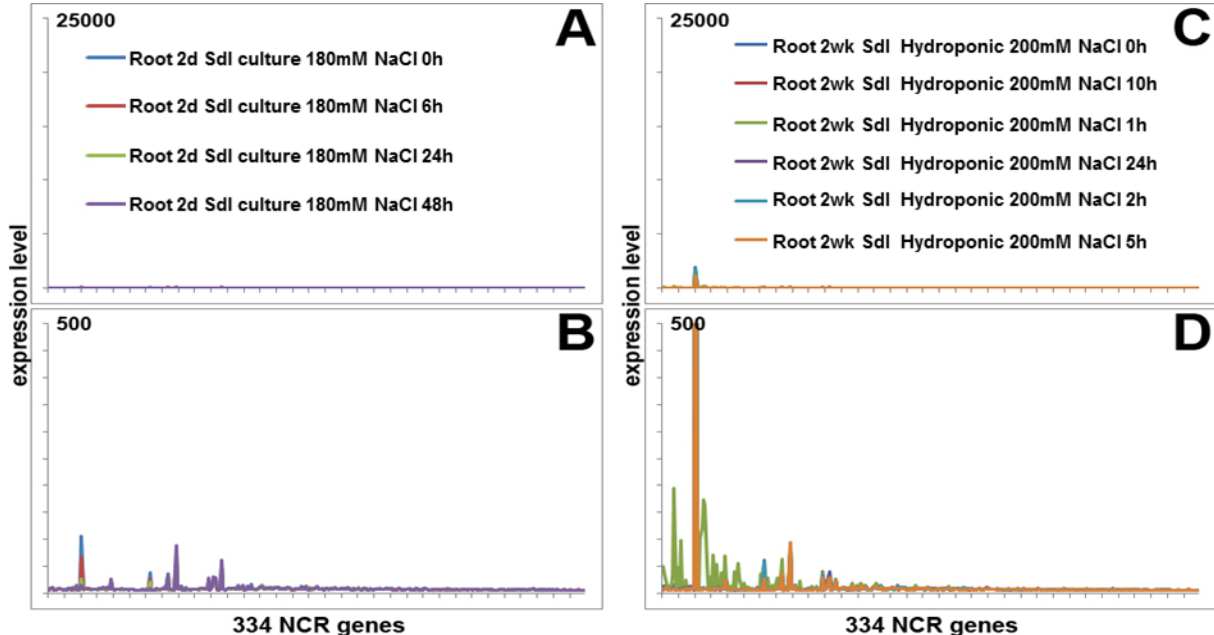
Defensins are also induced during pathogen infection (Maroti et al. 2011; Sels et al. 2008; Coninck et al. 2010; De Coninck et al. 2013) therefore expression of the NCR family was tested using data from roots infected with different pathogens. (**Fig. 21**).



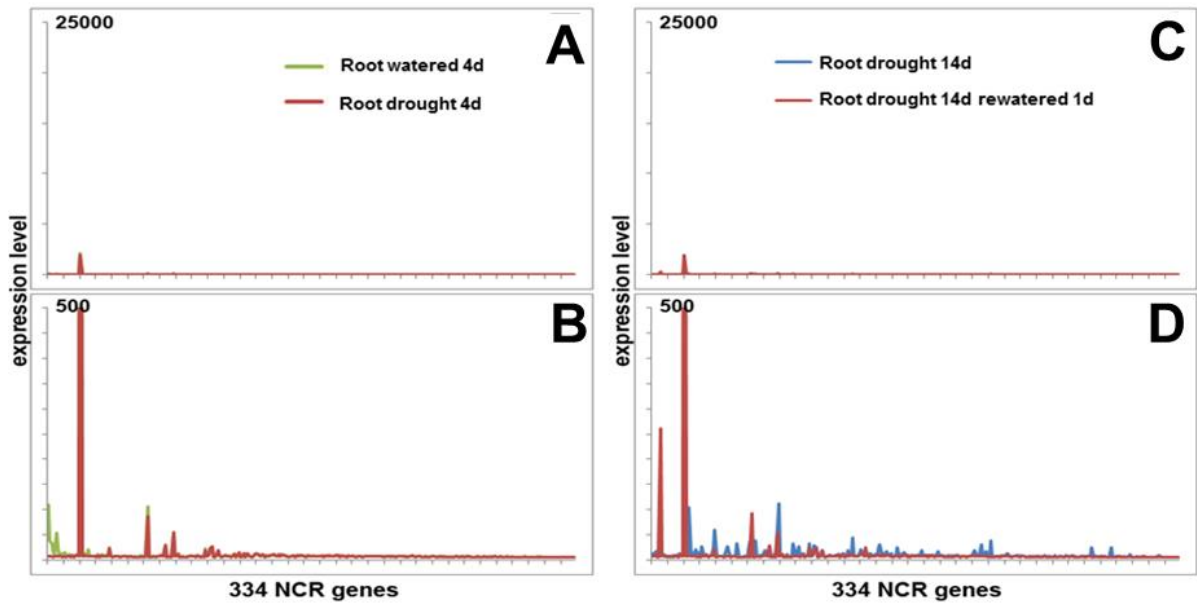
**Figure 21. Expression of NCR genes during microbial infections and elicitor treatment.** (A, B) NCR expression in *M. truncatula* control roots and roots infected with *Phymatotrichopsis omnivora* (Uppalapati et al. 2009), *Macrophomina phaseolina* (Mah et al. 2012), *Glomus intraradices* (Hogekamp et al. 2011), *Aphanomyces euteiches* (Rey et al. 2013), *Ralstonia solanacearum*, *Meloidogyne incognita* (Damiani et al. 2012), *Escherichia coli* O157:H7 and *Salmonella enterica* (Jayaraman et al. 2014). (C, D) NCR expression in a *M. truncatula* cell suspension (CS) derived from root cells after treatments for

2h or 24h with 50 µg glucose equivalents ml<sup>-1</sup> yeast elicitor (YE) or 500 µM methyl jasmonate (MJ) (Naoumkina et al. 2007). (A) and (B) on the one hand and (C) and (D) on the other hand represent the same data but the scale of the graphs is at relative expression level 25,000 maximum in (A) and (C) and at maximum level 500 in (B) and (D).

Pathogenic fungi *Phymatotrichopsis omnivore* (Uppalapati et al. 2009) and *Macrophomina phaseolina* (Mah et al. 2012), the symbiotic mycorrhizal fungus *Glomus intraradices* (Hogekamp et al. 2011), the oomycete *Aphanomyces euteiches* (Rey et al. 2013), the bacterial pathogen *Ralstonia solanacearum* and the nematode *Meloidogyne incognita* (Damiani et al. 2012) and also the human enteric bacterial pathogens *Escherichia coli* O157:H7 and *Salmonella enterica*, which are frequent sources of legume food contamination were also included because they are capable of surface and internal colonization of *M. truncatula* roots (Jayaraman et al. 2014) (**Fig. 21A, B**). In addition, transcriptomes in *M. truncatula* cell suspensions (CS) derived from root cells treated with the defense response-inducing yeast elicitor (YE) or methyl jasmonate signals (MJ) (Naoumkina et al. 2007) were also included (**Fig. 21C, D**) as well as, salt and drought stresses that induce defensin genes in plants (De Coninck et al. 2013). Thus, transcriptomes in NaCl-treated (**Fig. 22**) (Li et al. 2009) or drought stressed roots (**Fig. 23**) (Zhang et al. 2014) were also analysed.



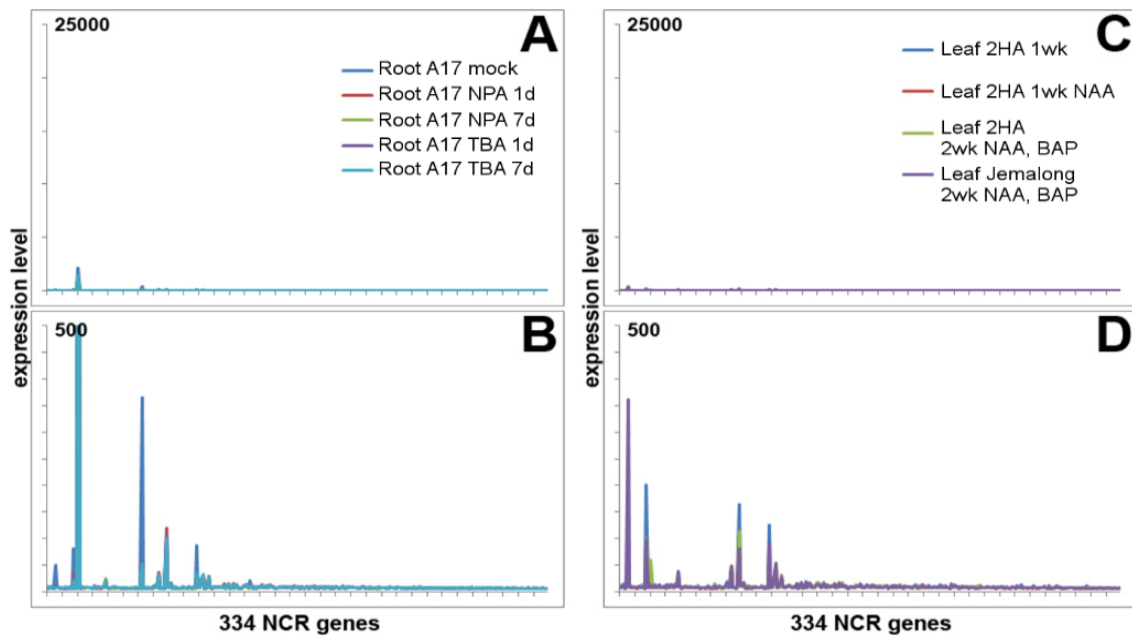
**Figure 22. Expression of NCR genes during salt stress.** (A-D) NCR expression in roots exposed to NaCl salt stress (Li et al. 2009). (A) and (B), (C) and (D) represent pairwise the same data but the scale of the graphs is at relative expression level 25,000 maximum in (A), (C) and at level 500 in (B), (D).



**Figure 23. Expression of *NCR* genes during drought stress.** (A-D) *NCR* expression in roots exposed to drought stress (Zhang et al. 2014). (A) and (B), (C) and (D) represent pairwise the same data but the scale of the graphs is at relative expression level 25,000 maximum in (A), (C) and at level 500 in (B), (D).

Not considering the five *NCR* genes with relaxed specificity, none of the *NCR*s showed expression in any of the data sets (with the possible exception of giant cells formed by the nematode *M. incognita* (**Fig. 21B**)). Several *NCR*s exhibited a hybridization signal in giant cells although the level was about 1 to 2 orders of magnitude lower than the signal in nodules for the same *NCR* gene. However, it should be noted that the giant cells were isolated by LCM and that the array hybridization was performed with an amplified cDNA sample (Damiani et al. 2012), which could be a source of background hybridization. In any case, besides the possible exception of the giant cells, the data indicate that the *NCR* genes seem not to be used by the plant to control infections other than the *Rhizobium* bacteria in nodules.

In the MtGEA data for plant treatments with the auxin 1-naphthaleneacetic acid (NAA), the cytokinin 6-benzylaminopurine (BAP) and the auxin transport inhibitors 2,3,5-triiodobenzoic acid (TIBA) or N-1-naphthylphthalamic acid (NPA) (**Fig. 24**) are also available for testing *NCR* expression (Imin et al. 2008; Rightmyer & Long 2011). Interestingly, treatment of roots with the auxin transport inhibitors TIBA or NPA and also cytokinin leads to the formation of nodule-like structures (Rightmyer & Long 2011). However, in none of these conditions, *NCR* genes were expressed except for the five genes with relaxed expression (**Fig. 24**).

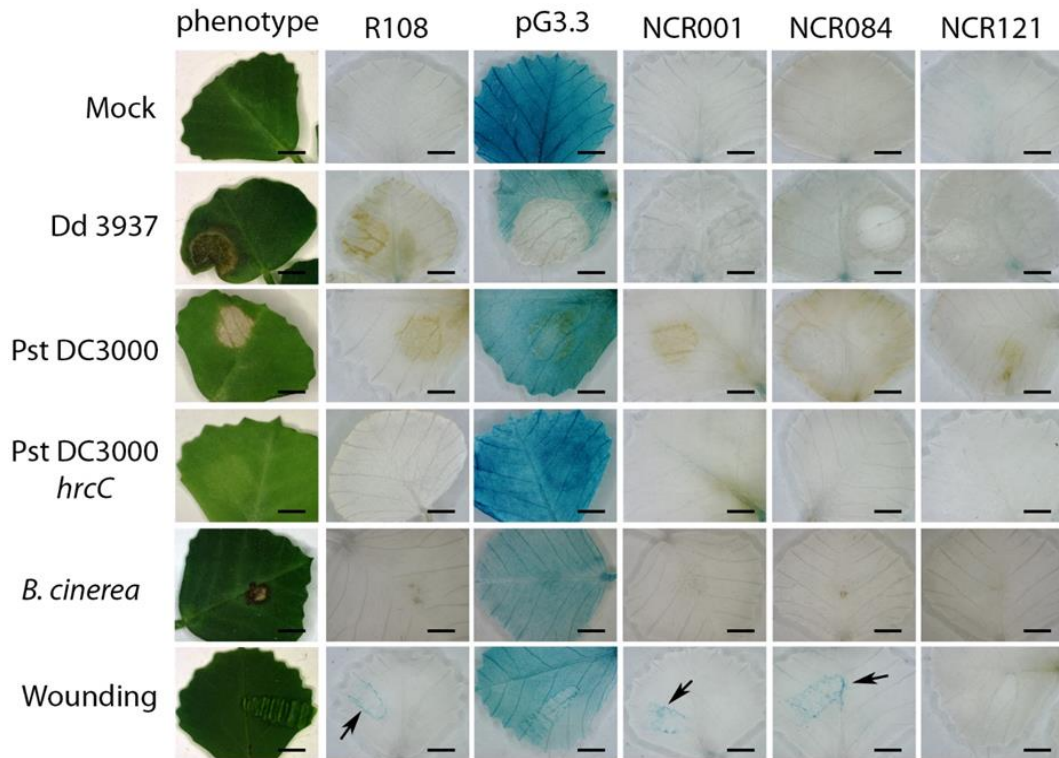


**Figure 24. NCR expression in response to phytohormones.** (A, B) NCR expression after treatments for 1 (1d) or 7 days (7d) of *Medicago* line A17 roots with the auxin transport inhibitors naphthylphthalamic acid (NPA; 200  $\mu$ M) and 2,3,5-triindobenzoid acid (TIBA; 200  $\mu$ M) (Rightmyer & Long 2011). (C, D) NCR expression after treatments of *Medicago* line 2HA or Jemalong leaves for 1 (1wk) or 2 (2wk) weeks with the auxin 1- naphthaleneacetic acid (NAA; 10  $\mu$ M) or the cytokinin 6-benzylaminopurine (BAP; 4  $\mu$ M) (Imin et al. 2008). (A) and (B) on the one hand and (C) and (D) on the other hand represent the same data but the scale of the graphs is at relative expression level 25,000 maximum in (A) and (C) and at maximum level 500 in (B) and (D).

#### 4.1.6. Promoter-GUS analysis of NCR expression during pathogenic interactions

The MtGEA compendium contains datasets from different pathogenic interactions but these are exclusively root pathogens (*R. solanacearum*, *A. euteiches*, *M. phaseolina* and *P. omnivore*). In these datasets, transcription of NCR genes is not detected. To confirm the absence of NCR expression during leaf or stem pathogen infection that encompass also other trophic interactions and infection strategies, the previously described three NCR promoter-GUS reporter lines were used (see Section 4.1.4). Leaflet inoculation of the necrotrophic soft rotting bacterium *Dickeya dadantii* 3937 induced maceration symptoms from 1 dpi on, but failed to induce NCR expression (Fig. 25). Similarly, infiltration of the virulent strain *Pseudomonas syringae* pv. *tomato* DC3000 (*Pst*) induced necrosis in the infiltrated zone within 2 dpi, whereas the *hrcC* mutant strain that is unable to form a functional T3SS did not induce any visible reaction. Although, *Pst* DC3000 is not described as a natural pathogen of *M. truncatula*, our data suggest that at least some bacterial effector proteins can be specifically recognized by the

plant resulting in a necrosis similar to the hypersensitive response of the non-host *Nicotiana benthamiana* plants (Hann & Rathjen 2007). This observation is also in agreement with the absence of detectable reaction in plants inoculated by a T3SS mutant strain. *NCR* expression was not detected in the infiltrated leaflets in any condition (**Fig. 25**). Similarly, inoculation of the same *M. truncatula* lines with the fungus *Botrytis cinerea* causing necrotrophic grey mold yielded symptoms at 7 dpi without any detectable *NCR* expression (**Fig. 25**).



**Figure 25. Promoter-GUS analysis of *NCR* genes in infected and wounded leaves.** Wild type *M. truncatula* R108 (R108) and transgenic R108 carrying *NCR* promoter-GUS fusions (*NCR001*, *NCR084*, *NCR121*) or a constitutive GUS under the control of the 35S promoter (pG3.3) were mock infected, infected with *Dickeya dadantii* 3937 (*Dd* 3937), *Pseudomonas syringae* pv. *tomato* DC3000 (*Pst* DC3000), the *Pseudomonas syringae* pv. *tomato* DC3000 *hrcC* T3SS mutant (*Pst* DC3000 *hrcC*) or *Botrytis cinerea* (*B. cinerea*), or wounded and then stained for GUS activity, which is observed as blue color. Note that in certain leaves, the leaf base and the veins are faintly stained blue. In addition, the wounding print is slightly stained (arrows). This staining is also observed in untransformed R108 and thus represents background signal. The left panels show the leaf phenotype after treatment and before staining. Scale bars are 2 mm.

The results of the pathoassays are also in line with a recent study showing that *NCR* expression was not detected during the interaction between *M. truncatula* and the hemibiotrophic leaf pathogen *Colletotrichum trifolii* (Tesfaye et al. 2013). Similarly, they reported that the biotrophic soil pathogen *Phytophthora medicaginis* failed to induce *NCR* expression in host roots (Tesfaye et al. 2013). Altogether, our data and the study of Tesfaye et

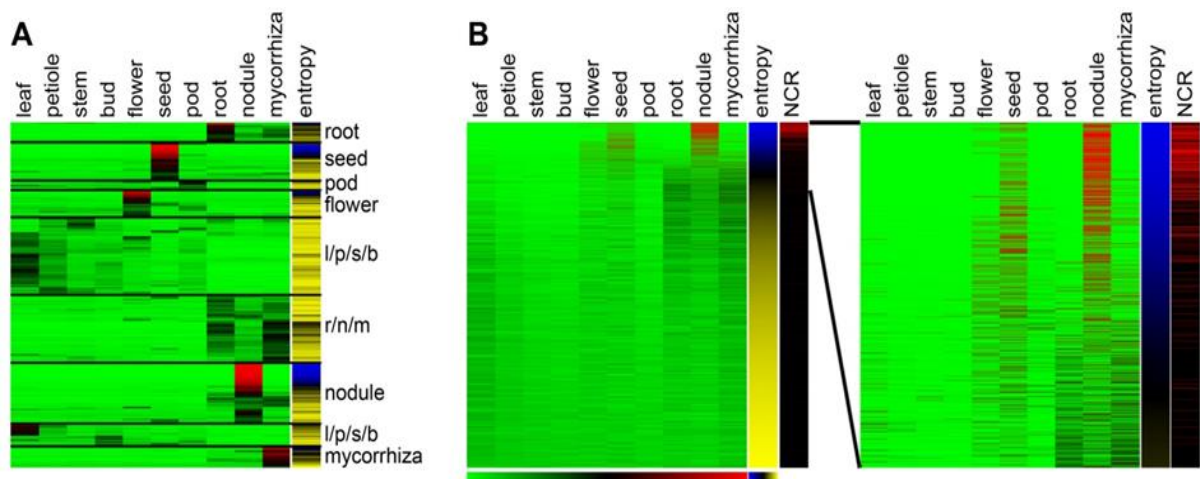
al. 2013 are in agreement with the MtGEA dataset and broaden the conclusion that *NCRs* are not involved in pathogen response, whatever the trophic- (bio/hemibio/necrotrophic), the host/non-host status and the output of the interaction (disease or resistance). Finally, as herbivory and more generally, wounding may induce plant defenses around the wounded zone, the effect of mechanical wounding of leaflets was also tested but again no *NCR* expression could be detected (**Fig. 25**).

#### **4.1.7. *NCR* genes have very high tissue specificity as measured by Shannon entropy**

The above analyses revealed high expression specificity for the *NCR* gene family: the genes are only expressed in nodules and not in any other organ or under any physiological condition. To express this specificity quantitatively and to compare it to other types of specifically expressed genes, their Shannon entropy was calculated. This parameter is a metric for characterizing the uniformity of the expression pattern of a gene over the tested conditions (Schug et al. 2005; Zhang et al. 2006). Low entropy values indicate high tissue specificity, while high entropy levels on the contrary mean ubiquitous expression. Ten different tissues were taken into consideration: leaf, petiole, stem, bud, flower, seed, pod, root, nodule and mycorrhiza. In case of nodules, seed and root the mean value was used of different developmental stages or experiments (for nodules: 4, 10, 14 and 28 dpi stages; for seeds: 10, 12, 16, 20, 24 and 36 dap; for root was used the 0 dpi control for nodulation and an independent experiment); thus in total 19 experiments were used. The Shannon entropy was calculated as described (see 3.1.4) for each of the 50,900 probe sets using these 10 tissue datasets (**Appendix, Supplementary information 2**, [https://static-content.springer.com/esm/art%3A10.1186%2F1471-2164-15-712/MediaObjects/12864\\_2014\\_6439\\_MOESM3\\_ESM.xlsx](https://static-content.springer.com/esm/art%3A10.1186%2F1471-2164-15-712/MediaObjects/12864_2014_6439_MOESM3_ESM.xlsx)). The 9000 probe-sets with the lowest entropy (and therefore the higher tissue specificity) were selected for further analysis and hierarchical clustering of these genes was performed (**Fig. 26A**). Clusters of tissue-specific genes could be distinguished for root, seed, pod, flower, aerial tissues, root tissues, nodule and mycorrhiza. The seed, flower, mycorrhiza and especially the nodule clusters were enriched in genes with low entropy value (**Fig. 26A**). The strong over representation of nodule-specific genes among the low entropy genes became more obvious when the dataset was ordered according to increasing entropy levels (**Fig. 26B**). This analysis showed that in *M. truncatula* the genes with the lowest entropy and thus the highest tissue specificity are mostly nodule-



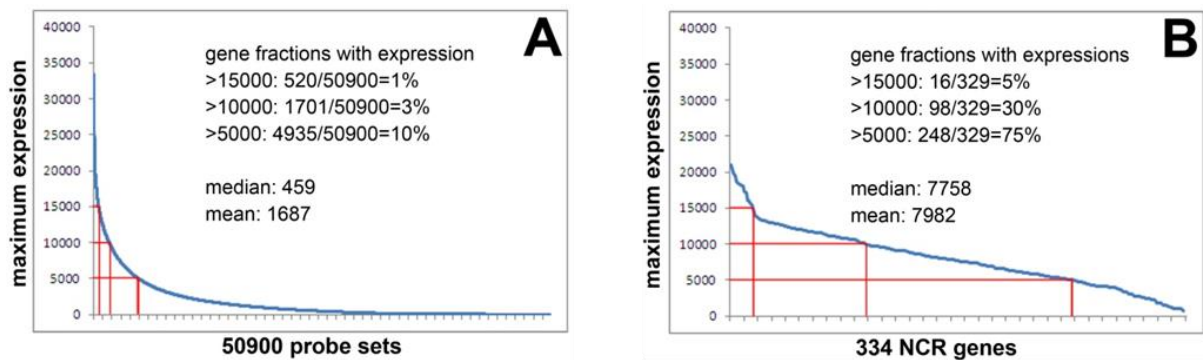
specific genes and among them the *NCRs* are the most represented ones although many other known nodule-specific genes also exhibited very low entropy (**Appendix, Supplementary information 2**). These include for example leghemoglobin genes, the glycine-rich peptide (GRPs) genes (Kevei et al. 2002; Alunni et al. 2007), the small nodulin acidic RNA-binding protein (SNARP) gene family (Laporte et al. 2010), genes encoding a small family of secretory calmodulin-like proteins (Mergaert et al. 2003; Liu et al. 2006), the *DNF2* (defective in nitrogen fixation 2) gene involved in suppression of defence responses in the symbiotic cells (Bourcy et al. 2013) and others. Most interestingly, also putative retrotransposons (probe-sets Mtr.9294.1.S1\_at and Mtr.636.1.S1\_at) and a Dicer 1-like ribonuclease III gene (probe-set Mtr.41531.1.S1\_at) are among the nodule specific low entropy genes (**Appendix, Supplementary information 2**).



**Figure 26. Shannon entropy.** (A) Hierarchical clustering of the 9000 probe-sets with the highest tissue specificity (lowest entropy values  $E_g$ ). The expression heat map is in green-black-red color scheme. The entropy heat map is in blue-black-yellow scheme. (B) The data set of panel (A) was ordered according to increasing entropy of the genes. The left panel shows the relative expression level of the 9000 genes in the 10 tissues (green-black-red heat map) and the entropy values  $E_g$  (blue-black-yellow heat map). The location of the *NCR* genes is indicated with the black-red heat map (*NCR*): red means *NCR*, black means other gene type. The right panel is an enlargement for the first 1800 genes (entropy values  $E_g$  from 0.07 to 1.64). The scale bar for the expression level heat map is green: 0, black: 0.5 and red: 1 and for the entropy heat map blue: 0, black 1.44 and yellow: 2.88 (maximum entropy in the complete data set is 3.32 ( $\log_2(N)$ )).

Besides the high tissue-specificity, another feature in which the *NCRs* stands out from the average *M. truncatula* genes is the strength of expression. The microarray hybridization signal was used to estimate gene expression levels. The maximal signal for each of the 50,900 probe sets in the 267 experiments was searched. These hybridization signals vary from 33,500 for the strongest expressed gene to 9 (background) for the weakest. One percent of the probe

sets have an expression level higher than 15,000, 3% higher than 10,000 and 10% higher than 5000. The mean (average) signal is 1687 and the median 459 (**Fig. 27A**).



**Figure 27. *NCR* genes are among the most actively expressed genes in *M. truncatula*.** (A) Maximum expression level of the 50,900 *M. truncatula* probe-sets. (B) Maximum expression level of 334 *NCR* genes.

The same analysis on the subset of probes corresponding to the *NCR* genes gives a completely different picture: 5% of *NCR* genes have signals above 15,000, 30% above 10,000 and 75% above 5000 with a mean signal of 7982 and a median of 7758 (**Fig. 27B**). Thus, the *NCR* genes are among both the most specifically expressed and the most strongly expressed genes in the genome of *M. truncatula*.

#### 4.1.8. Conclusion

From the transcriptome data mining and experimental confirmation described here, we can conclude that apart from very few exceptions, the hundreds of *NCR* genes encoding defensin-like peptides are only activated during nodule formation. They are not expressed in other plant organs, during pathogen attack or abiotic stress. In nodules, they are not yet activated during the very early stages (initial cell proliferation period or in the meristem) before symbiotic nodule cells are formed and rhizobia are released in symbiosomes within the host cells. *NCR* genes are also not involved in symbiosome and bacteroid degradation during nodule senescence since their gene expression shuts down when senescence is initiated. However, the expression pattern of *NCRs* in successive waves during nodule formation suggest that the bacteroids are the only targets of the peptides and that subsets of the peptides might be involved in bacteroid differentiation and other subsets in bacteroid functioning. The *NCR* genes are among the most specifically expressed genes in *M. truncatula*. Moreover, when activated in nodules, their expression level is among the highest of all genes. Together, these data show that the *NCR* gene expression is subject to an extremely tight regulation and is only activated during nodule

organogenesis in the symbiotic cells. However, very little is known about how the very specific regulation of *NCRs* is achieved. Since their expression correlates with bacterial infection of the symbiotic cells, the perception of bacterial signals such as components of the bacterial envelope could be involved. It will be interesting to analyse nodule-specific transcription factors such as EFD (Vernie et al. 2008), IPD3 (Messinese et al. 2007) and others as well as *cis*-regulatory elements in the promoters of the *NCR* genes for their involvement in this particular regulation of *NCR* expression.

Moreover, the very tight regulation of the *NCR* genes might indicate that besides control by specific transcription factors and *cis*-regulatory elements, regulation at the level of chromatin might also be involved in the activation of the *NCR* genes. Genes with high tissue-specific expression are often actively silenced during most of the plant growth by epigenetic mechanisms. Since in *M. truncatula* the nodule-specific genes display the highest level of expression specificity, it might be worthwhile to investigate if epigenetic control is important in the regulation of the symbiotic cell-specific genes. The nodule-specific expression of putative retrotransposons (this analysis and Satgé et al. 2016), which are usually epigenetically silenced, and the Dicer 1-like ribonuclease III gene, which may have a role in epigenetic regulation, as well as the identification of small RNAs potentially targeting *NCR* genes (Lelandais-Briere et al. 2009) are all in agreement with such an epigenetic control of the symbiotic cell-specific genes. In the next part of my thesis work, I have analysed this hypothesis.

## **4.2. Ploidy-dependent changes in the epigenome of symbiotic cells correlate with specific patterns of gene expression**

In the previous chapter, I have described that the formation of symbiotic nodule cells in *M. truncatula*, which is driven by successive endoreduplication cycles, requires a drastic transcriptional reprogramming in different temporal waves including the activation of hundreds of cysteine-rich *NCR* genes expressed only in nodules. Based on those results, we have postulated that epigenetic mechanisms are involved in this transcriptional reprogramming as well as a role for the endoreduplication cycles in mediating epigenetic changes. In this chapter I describe the experiments that we have done to test these hypotheses. We show that the transcriptional waves indeed correlate with growing ploidy levels and we have investigated how DNA methylation, histone modifications and chromatin compaction change during endoreduplication cycles. In most *Medicago* genes DNA methylation was unaffected by the ploidy levels and was independent of the genes' active or repressed state. Differential DNA methylation was only found in a small subset of nodule-specific genes but it was detected in more than half of the *NCR* genes. On the other hand, expression of nodule-specific genes correlated with ploidy-dependent opening of the chromatin as well as, in a subset of tested genes, with reduced H3K27me3 levels combined with enhanced H3K9ac levels. Our results suggest that endoreduplication-dependent epigenetic changes contribute to transcriptional reprogramming in the differentiation of symbiotic cells.

The data of this chapter were published in PNAS in 2017. Marianna Nagymihály, Alaguraj Veluchamy, Zoltán Györgypál, Federico Ariel, Teddy Jégu, Moussa Benhamed, Attila Szűcs, Attila Kereszt, Peter Mergaert, and Éva Kondorosi. Ploidy-dependent changes in the epigenome of symbiotic cells correlate with specific patterns of gene expression. Proc. Natl. Acad. Sci. USA 2017, 114:4543-4548.

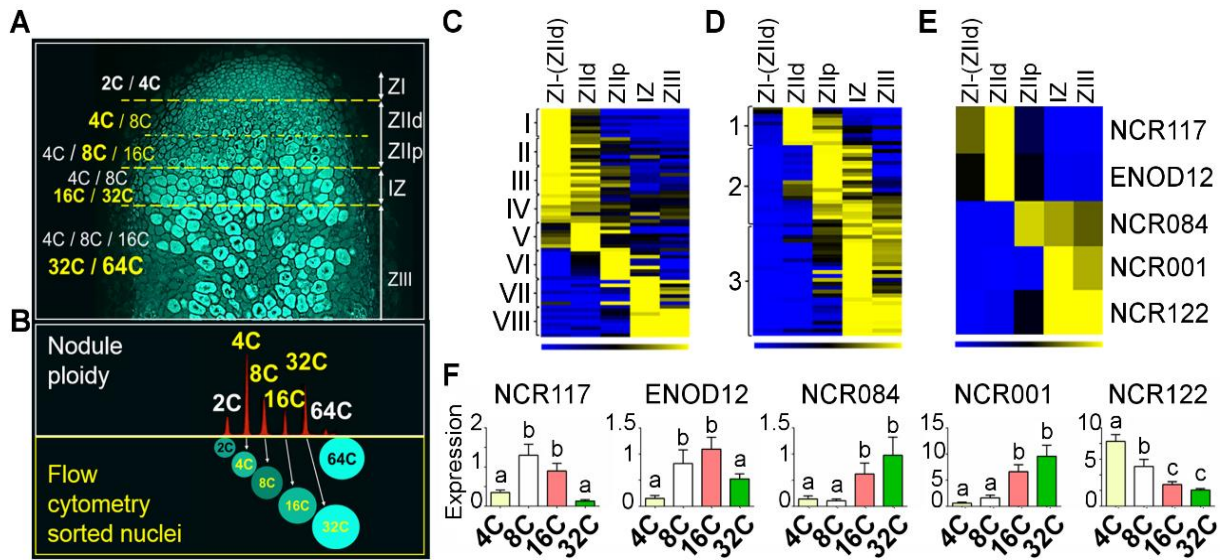
#### 4.2.1. Differential expression of nodule-specific genes depends on the ploidy levels of the symbiotic cells.

Mature *M. truncatula* nodules contain nuclei of 2C, 4C, 8C, 16C, 32C and 64C (**Fig. 28A, B**). The high ploidy nuclei result from the ENR-driven differentiation of post-meristematic cells into mature symbiotic cells (**Fig. 28A**) (Vinardell et al. 2003). To confirm a possible linkage between the above described expression of nodule-specific genes in successive waves and the ploidy levels of differentiating symbiotic cells, the expression level of selected, differentially expressed genes was measured in 4C, 8C, 16C and 32C flow-cytometry isolated nuclei (**Fig. 28B**) with quantitative reverse transcription PCR (RT-qPCR). In this and our further analyses the 2C and 64C nuclei were not included since they could not be isolated in sufficient amounts. Since nuclear and total cellular mRNA pools are comparable (Deal & Henikoff 2010) gene expression levels can be faithfully determined using nuclear RNA. Published *in situ* LCM RNA-seq data of nodule zones (Roux et al. 2014) (**Fig. 28C, D**) were correlated with the expression pattern of the selected genes in the nuclei (**Fig. 28E**). As examples of early genes *NCR117* and *ENOD12* were chosen. *NCR117* is one of the earliest *NCR* genes which is active in the distal part of zone II (ZIIId) of a mature nodule (**Fig. 28E**, Roux et al. 2014) while *ENOD12* is expressed in the root hairs and root epidermal cells as well as in ZIIId cells (Pichon et al. 1992). As later genes, *NCR084*, expressing in ZIIp and IZ and *NCR001*, which is activated later and is expressed in the IZ and ZIII cells (**Fig. 19**) was selected. *NCR122* was also tested because unlike all other *NCRs*, this gene is expressed in the root and in the uninfected nodule cells but not in the infected ones (**Fig. 20**).

Both the early genes, *NCR117* and *ENOD12* were the most strongly expressed in the 8C and 16C nuclei while the late genes *NCR084* and *NCR001* in the 16C and 32C cells coinciding with their later function in symbiotic cell development (**Fig. 28F**). *NCR122* expression was maximal in the 4C cells and then decreased with the increasing ploidy levels, in agreement with the expression of this gene in the low-ploidy cortical cells and uninfected cells of the IZ and ZIII (**Fig. 28F**). Expression of *NCR001* and *NCR122* was significantly higher than that of the other tested genes.

Overall, expression of these nodule-specific genes according to the ploidy level correlated well with their spatial expression pattern in nodules identified previously with *in situ* hybridization, promoter-GUS gene fusions and LCM RNA-seq (**Fig. 19 and Fig. 20**) (Pichon et al. 1992). Therefore, the nuclei purified by flow cytometry according to the ploidy level are

suitable to analyse the dynamics of the epigenetic state associated with the activation and inactivation of *NCR* gene expression.



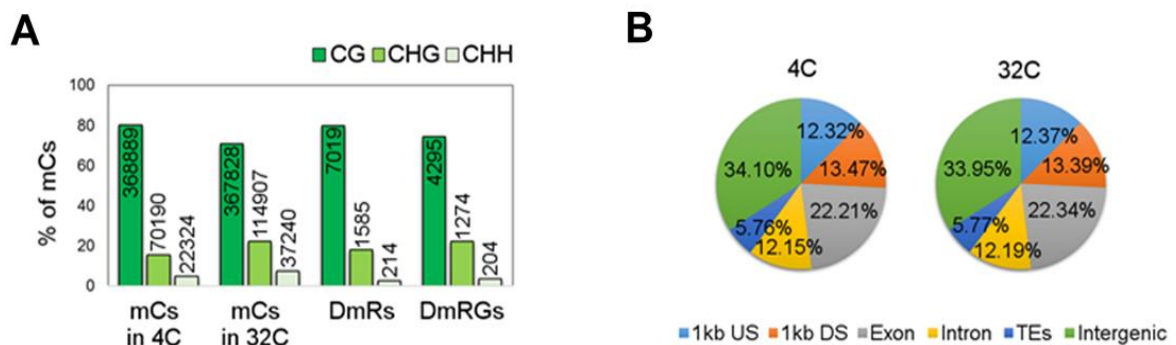
**Figure 28. Differential expression of *M. truncatula* genes during nodule development and at different ploidy levels of the symbiotic cells.** (A) *M. truncatula* nodule zones (ZI-ZIII) and the genome copy number (C) of nodule cells in a longitudinal nodule section visualized by autofluorescence. ZI: meristem; ZII: invasion zone (ZII<sub>d</sub>: distal younger 4C/8C symbiotic cells, ZII<sub>p</sub>: proximal older growing symbiotic cells 8C/16C); IZ: interzone large symbiotic cells (16C/32C) and ZIII: nitrogen-fixing zone huge symbiotic cells with large vacuoles (32C/64C) as well as uninfected cells (4C/8C) in the central region and nodule cortex. Predominant genome copy number of symbiotic cells is in yellow. (B) Nodule ploidy profile (Top) and flow cytometry sorting of different ploidy level nuclei (Bottom). Yellow indicates the ploidy of nuclei used for further experiments. (C-E) Hierarchical clustering analysis of *M. truncatula* gene sets (C, clusters I-VIII), *NCR* genes (D, groups 1-3) and selected genes (E) from RNA-seq experiment of nodule zones obtained by laser-capture microdissection (Data from Roux et al., 2014). Blue-(black)-yellow colors correspond from low to high expression levels. (F) Expression levels determined by RT-qPCR of *NCR117*, *ENOD12*, *NCR084*, *NCR001* and *NCR122* genes in 4C, 8C, 16C and 32C nodule nuclei. The expression levels are relative to the constitutive genes *Mtc27* and *40S*. Values are averages  $\pm$  SEM from three independent experiments. Letters a, b, c indicates significant differences (Student-Newman-Kuels test,  $P < 0.05$ ).

#### 4.2.2. Only a subset of genes display ploidy-dependent differential DNA methylation

As the majority of nodule-specific genes are expressed in the mature symbiotic cells (32C) and repressed in the uninfected (4C) cells, we analysed differences in DNA methylation (5-mC) genome-wide in isolated 4C and 32C nuclei with RRBS (Meissner et al. 2005). In total 461,403 methylated cytosines (mCs) were found in the 4C cells; 80% in CG, 15% in CHG and 5% in CHH sequence context whereas in 32C cells from the 519,975 mCs 71% were in CG,

22% in CHG and 7% in CHH sequence context (**Fig. 29A**). The number of detected mCs is, however, an underestimation due to the RRBS approach, the sequencing coverage and the stringent parameters used for mC calling (see Materials and Methods). The 71 to 80% mCs in the CG context was higher than reported in other plants in which typically 40 to 55% mCs are in the CG context (Garg et al. 2015; Shangguan et al. 2013; Song et al. 2013). The reason for this difference is potentially related to the CG enrichment during the library preparation. Moreover, the *M. truncatula* genome has lower GC content (27%) compared to other plant species (32-46%) (Lister et al. 2008). The overall distribution of mCGs on the 1 kb upstream or downstream regions, exons, introns, intergenic regions or transposons was similar in 4C and 32C nuclei (**Fig. 29B**). A considerable number of mCs were observed in the intergenic (34%) and gene body regions (22%) as well.

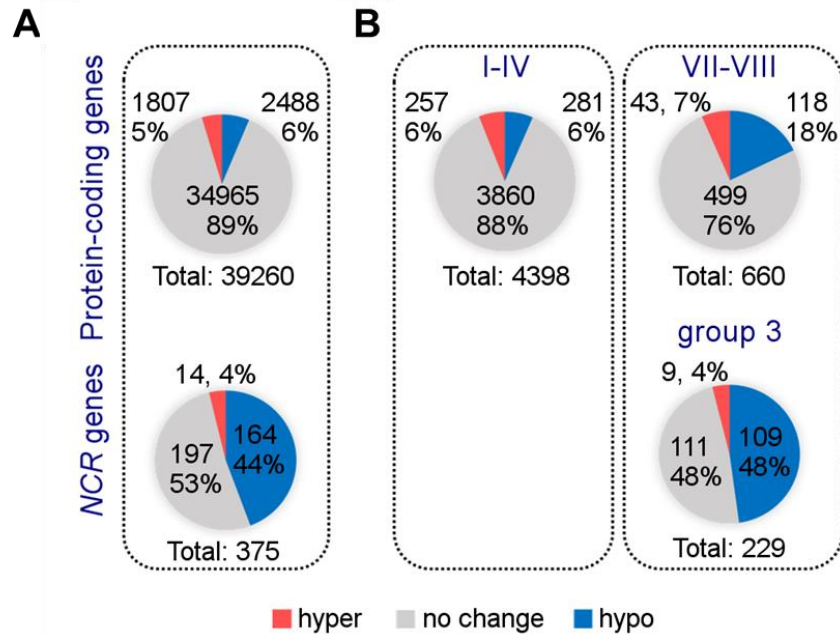
To investigate the differential methylation in 4C and 32C nuclei, mCs at neighbouring positions were grouped and differentially methylated regions (DmRs) and within them DmR-associated genes (DmRGs) were identified. Most DmRs (7019, 79%) and DmRGs (4295, 74%) were found in CG sequences, suggesting the relevance of mCGs in gene regulation (**Fig. 29A**).



**Figure 29. Summary of DNA methylation in 32C versus 4C nodule cells.** (A) Percentage and number of methylated cytosines (mCs) in 4C and 32C, differentially methylated regions (DmRs) and genes (DmRGs) in 32C compared to 4C in CG, CHG and CHH context. (B) Distribution of mCGs on the 1 kb upstream (US) and downstream (DS) regions, exons, introns, intergenic regions or transposons (TEs).

Out of the detected 39,260 protein-coding genes in the *M. truncatula* genome, 11% (4295) showed differential methylation; 6% (2488) being hypo- and 5% (1807) hypermethylated in the 32C cells (**Fig. 30A**). Gene Ontology (GO) term distribution analysis of hypo-DmRGs revealed a strong over-representation of the “nodule morphogenesis” (GO:0009878;  $P=1.4E^{-31}$ ;  $FDR=5.5E^{-29}$ ) and “metal ion binding” (GO:0046872;  $P=4.50E^{-07}$ ;  $FDR=8.60E^{-05}$ ) GO categories (**Appendix, Table 5**). In these two categories of hypo-DmRGs, the *NCRs* and other nodule-specific genes accounted for 96.6% and 60.0% of genes,

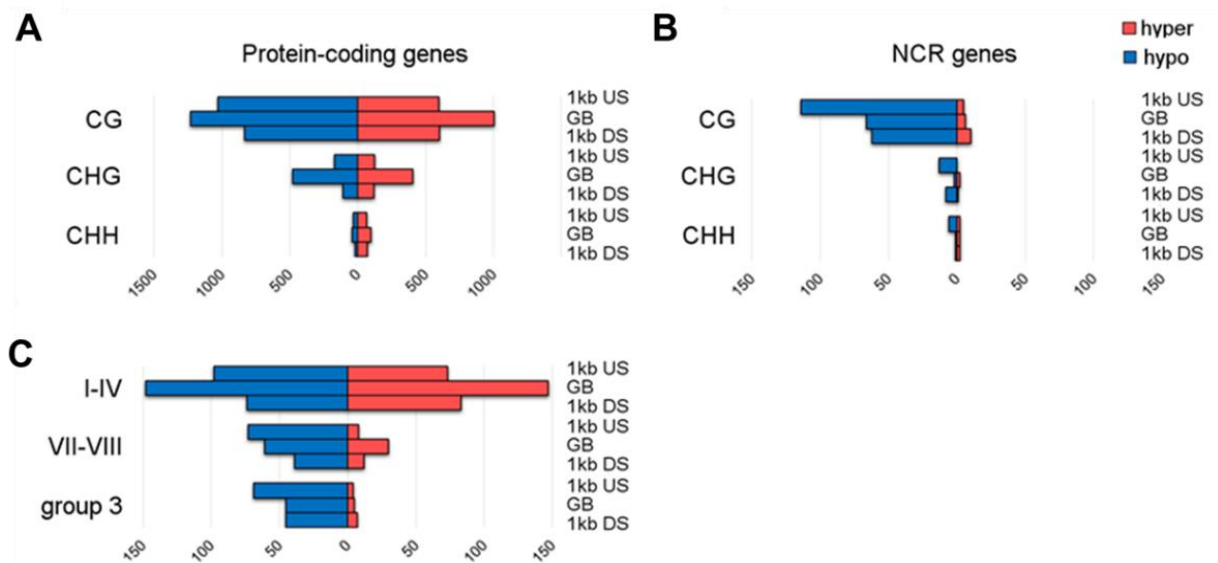
respectively. Out of the 375 detected *NCRs*, 44% (164) were hypomethylated, and only 4% (14) hypermethylated at 32C ploidy level (**Fig. 30A**). Differential methylation of protein-coding genes was observed on the entire coding region including the 1 kb upstream and downstream sequences with predominance in the gene body and primarily in the CG context (**Fig. 31A**) while in case of the *NCRs* differential methylation was predominant in their 1 kb upstream regions and in the CG context (**Fig. 31B**).



**Figure 30. Differential methylation of nodule-specific and nodule expressed genes in 32C versus 4C nodule cells.** Percentage and number of protein-coding genes and *NCR* genes showing hyper- or hypomethylation (A) genome-wide and (B) according to the expression profile: clusters I-IV genes preferentially expressed in 4C (Left), and clusters VII-VIII and the group 3 *NCR* genes, preferentially expressed in 32C (Right). Hypermethylation is shown in red, no change in grey and hypomethylation in blue.

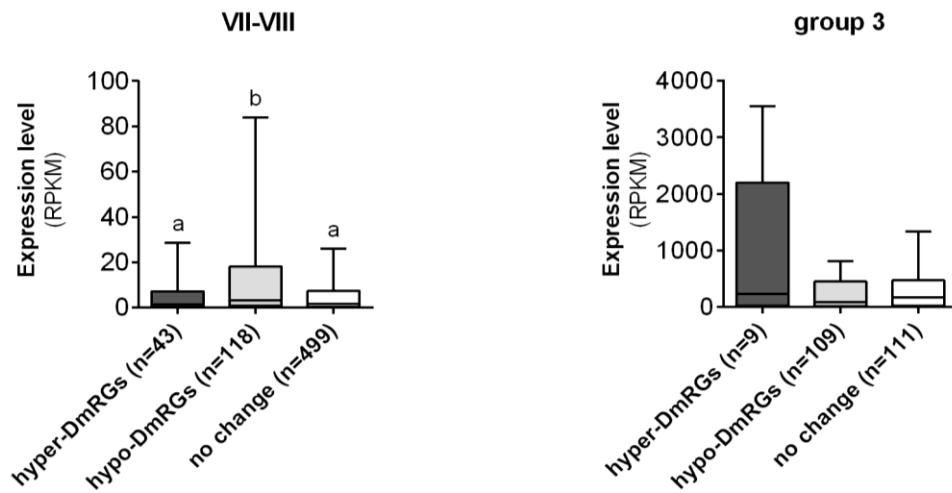
To investigate the relationship between changes in CG methylation and differential gene expression, clusters I-IV protein-coding genes (**Fig. 28C**) expressed in 4C and downregulated in 32C, clusters VII-VIII genes (**Fig. 28C**) upregulated in 32C and not expressed in 4C and the 32C expressed *NCRs* (group 3, **Fig. 28D**) were analysed (**Fig. 30B**). In clusters I-IV, 6% (281) of the genes were hypomethylated and unexpectedly only 6% (257) were hypermethylated in 32C suggesting that DNA methylation plays a minor role in their repression. Moreover, these methylation changes were mostly visible in the gene body (**Fig. 31C**). In the case of the 32C upregulated genes, the fraction of hypo-DmRGs was significantly higher ( $P < 0.0001$ , Fisher's exact test): 18% (118) in clusters VII-VIII and 48% (109) in group 3 *NCRs* (**Fig. 30B**). These changes occurred mostly in the 1 kb promoter region (**Fig. 31C**). These results suggest that decrease in DNA methylation might contribute to their activation or upregulation in 32C.





**Figure 31. DNA methylation change of protein-coding- and *NCR* genes in different sequence context and genic regions.** (A) Genome-wide DNA methylation changes in the protein-coding genes and (B) *NCR* genes in CG/CHG/CHH context in the US, gene body (GB) and DS genic regions. (C) DNA methylation changes of protein-coding genes and *NCRs* from different expression clusters in CG sequences in the 1 kb flanking and in the gene body region. Hypomethylation is shown in blue, and hypermethylation in red.

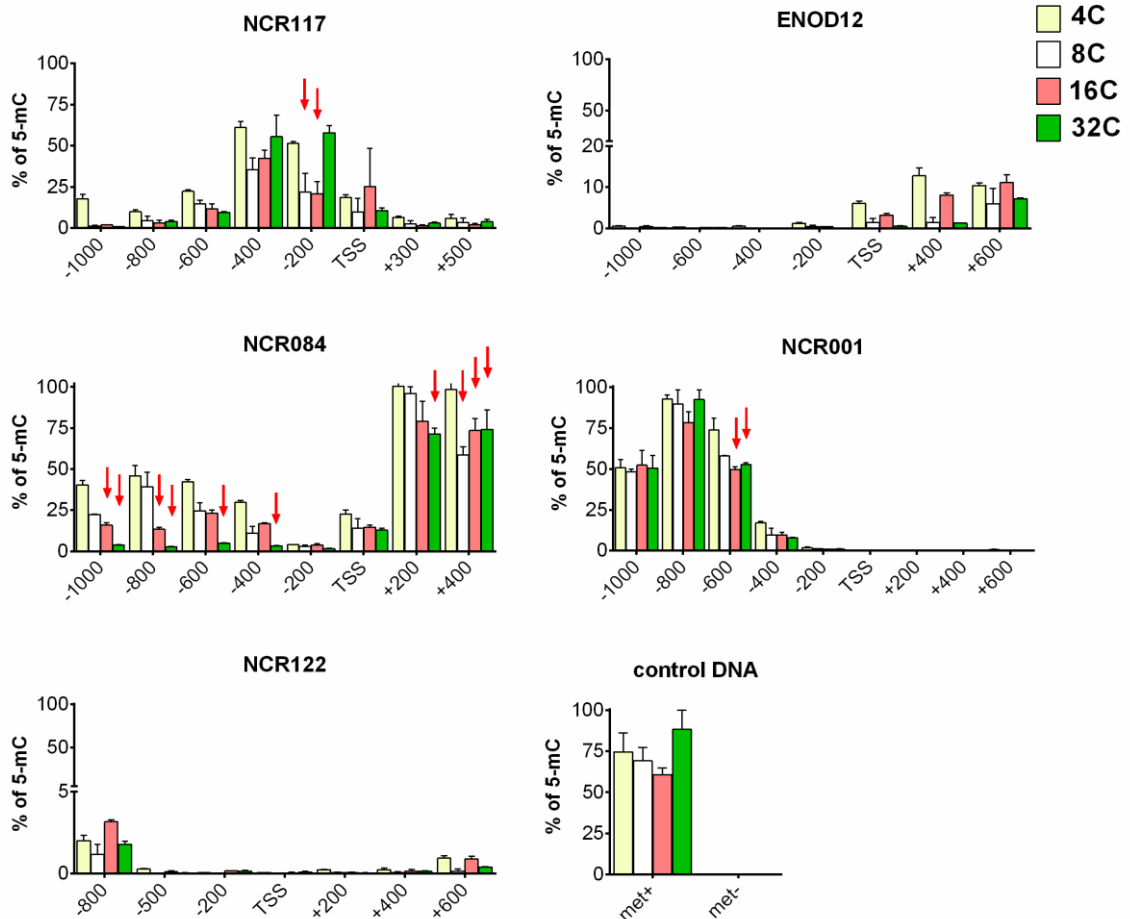
To investigate if differential methylation has quantitative effects on gene expression, transcript abundance of hyper- and hypo-DmRGs were compared using the published *in situ* RNA-seq data of the nodule zones (Fig. 32, Appendix, Supplementary information 3, <http://www.pnas.org/content/suppl/2017/04/12/1704211114.DCSupplemental/pnas.1704211114.sd01.xlsx>). We found that hypo- and hypermethylation or no change in DNA methylation affected differently the expression levels of the protein coding- and *NCR* genes (Fig 32). The lower expression levels of protein-coding genes in 32C cells correlated with their hypermethylation or no change in their methylation status while the transcript abundance of hypomethylated protein-coding genes was significantly higher. This is in agreement with the prevailing view that demethylation is a positive regulator of gene expression. However, such a correlation was not observed for the group 3 *NCR* genes and unexpectedly, expression levels of the few hypermethylated *NCRs* were particularly high. Moreover, hypomethylated *NCR* genes exhibited similar expression level than the unmethylated ones. Nevertheless, the expression of the group 3 *NCR* genes, independent of their methylation status, was much above the average gene expression of the cluster VII-VIII genes. Taken together, these data suggests that demethylation might be involved in the activation of a large subset of *NCR* genes but is not correlated with their extremely high expression level.



**Figure 32. Differential expression of differentially methylated nodules expressed and nodules-specific genes in 4C and 32C nodules cells.** Expression level (RPKM) of protein-coding genes of cluster VII-VIII and group 3 *NCRs* (Roux et al. 2014) showing hyper- or hypomethylation either in their 1 kb upstream, gene body or in their 1 kb downstream region. Box plots show the median, upper and lower quartiles, while whiskers show the 10-90th percentile expression (RPKM) values of protein-coding genes and *NCRs*. n indicates gene number. Letters indicate significant differences (Mann-Whitney test,  $P < 0.05$ ).

About 53% of the *NCR* genes did not differ in DNA methylation at 4C and 32C (**Fig. 30A**). These genes might be active in the 8C and 16C cells and display hypomethylation in these nuclei or alternatively, it could mean that DNA methylation does not play a pivotal role in their regulation. In our selected gene set, *NCR117*, *ENOD12*, *NCR001* and *NCR122* belonged to this category while *NCR084* was in the hypo-DmRG category, which allowed to test these possibilities. By carrying out methylated DNA immunoprecipitation coupled to quantitative PCR (MeDIP-qPCR) in 4C, 8C, 16C and 32C nuclei and measuring DNA methylation of the five selected genes along their 1 kb upstream and the 500-600 bp downstream regions from the translational start site (TSS), strikingly different patterns and extent of DNA methylation were found (**Fig. 33**). Despite its high expression, the highest level of DNA methylation was observed in the *NCR001* gene, between -1000 and -600 bp that peaked at -800 bp and then the DNA methylation decreased at -400 bp and even further at -200 bp and was not detectable in the coding sequence. This profile was unaffected by the ploidy levels between -1000 bp and -800 bp and compared to 4C only slight decrease in methylation was observed at -600 bp in the higher ploidy nuclei. *NCR117* had a lower level of methylation between -600 bp and -200 bp with the highest level at -400 bp at all ploidy levels. Significant decrease was observed in 8C and 16C at -200 bp. *ENOD12* was methylated at low levels in the gene body but not in the 1 kb upstream region. In the case of *NCR122*, neither the upstream region nor gene body was

methylated. Also the strongly methylated gene body of *NCR084* was unaffected by the ploidy levels, unlike its 1000 bp upstream region where DNA methylation decreased significantly by the growing ploidy levels in line with the bisulphite sequencing analysis and its expression in 16C and 32C cells.

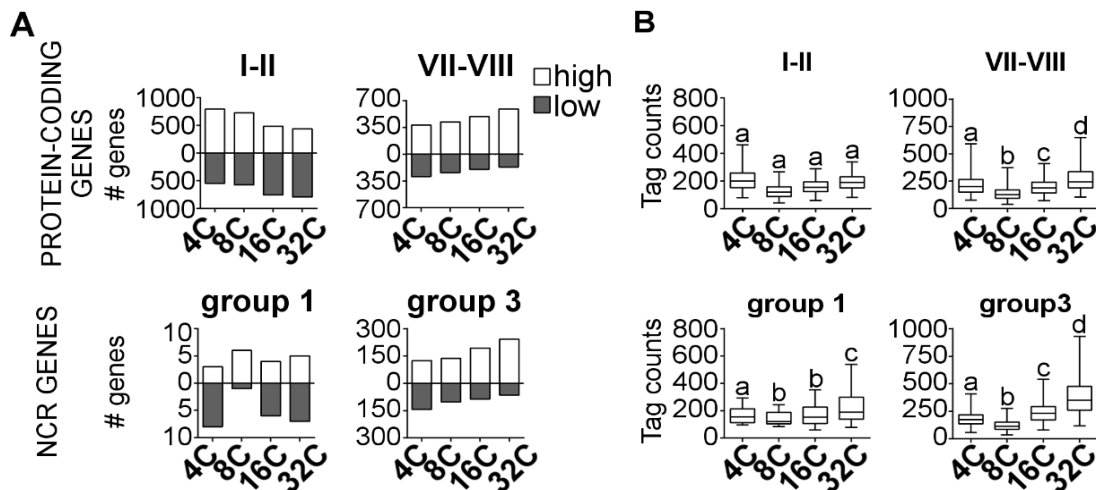


**Figure 33. Differential methylation of the studied genes in 4C, 8C, 16C and 32C nuclei.** DNA methylation levels of *NCR117*, *ENOD12*, *NCR084*, *NCR001*, *NCR122* genes and control methylated (met+) and unmethylated (met-) DNA at different ploidy levels of nodule cells with MeDIP-qPCR. 5-mC levels are shown 1 kb upstream and downstream of the translational start site (TSS) as percentages of input. Values are averages  $\pm$  SEM from two independent experiments. Arrows indicate significant difference compared to 4C (two-way ANOVA with Tukey's multiple comparison test,  $P < 0.01$ ).

Our findings show that the clusters VII-VIII and the *NCR* genes are overrepresented in hypomethylated regions, suggesting that decrease in DNA methylation, particularly in the 1-kb upstream region, contributes to their expression while other genes with unaltered DNA methylation from these groups might require chromatin modifications.

### 4.2.3. Chromatin accessibility varies with ploidy and correlates with gene expression

To learn how chromatin compaction and accessibility changes during the course of nodule differentiation and how this influences gene expression, we studied genome-wide the chromatin accessibility in 4C, 8C, 16C and 32C nuclei using the „Assay for Transposase-Accessible Chromatin sequencing” (ATAC-seq) (Buenrostro et al. 2013) (**Fig. 34**). At whole-genome scale the chromatin accessibility showed considerable differences with growing ploidy levels. First, the early- and late protein coding- and *NCR* genes were classified according to the accessibility of their chromatin (**Fig. 34A**). We found that the number of genes with high- or low compactness in their 1 kb upstream region correlated well with the up- or downregulation of genes in the respective gene clusters.

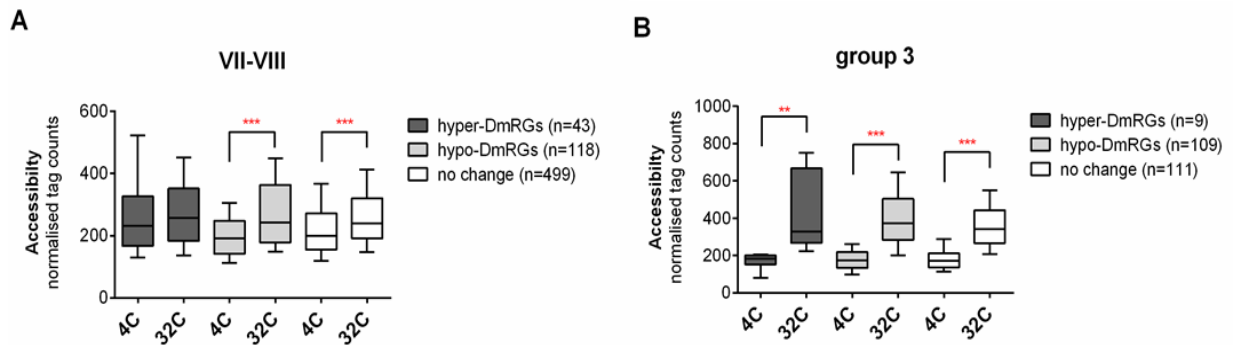


**Figure 34. Differential chromatin accessibility of nodule-specific and nodule expressed genes in 4C, 8C, 16C and 32C cells.** (A) Number of early (I-II, group 1) and late (VII-VIII, group 3) protein coding- and *NCR* genes showing high- (white) or low (dark grey) chromatin accessibility at 4C, 8C, 16C and 32C nuclei. # indicates number of genes. (B) ATAC-seq sequencing tag counts in the 1 kb upstream region of protein coding- and *NCR* genes from early (I-II, group 1) and late (VII-VIII, group 3) gene clusters at 4C, 8C, 16C and 32C nuclei. Box plots show the median, upper and lower quartiles, while whiskers show the 2.5-97.5% percentiles of the accessibility values, respectively. Letters indicates significant differences (Wilcoxon’s rank sum test,  $P < 0.05$ ).

The number of cluster I and II genes with high chromatin accessibility was the largest in 4C and 8C and was lower in 16C and 32C, whereas inversely it increased in clusters VII and VIII with the growing ploidy levels. The number of early group 1 *NCR* genes with high chromatin accessibility was the highest in the 8C nuclei while that of the late group 3 *NCR* genes was highly accessible in 16C and even more in 32C. Distribution of the ATAC

sequencing tag counts displayed only small variations at the 4 ploidy levels in the case of cluster I-II genes while the cluster VII-VIII genes have significantly increased numbers of reads in the 1 kb upstream region in 32C nuclei (**Fig. 34B**). For *NCRs* in group 3, the read counts increased in 16C and were even more pronounced in the 32C nuclei (**Fig. 34B**). To evaluate the correlation between expression and accessibility, Pearson correlation coefficients (PCC) were determined between the RPKM values and ATAC-seq tag counts for protein-coding genes and *NCR* genes belonging to different expression groups from 4C to 32C ploidy levels. No correlation was found (PCC=0,28, SEM=0,01, SD=0,51) in case of early expressed protein-coding genes (clusters I-II) while in case of the late expressed genes (cluster VII-VIII, PCC=0,52, SEM= 0,02, SD= 0,44) and in group 3 *NCRs* (PCC=0,84, SEM=0,01, SD=0,20) the correlation was evident.

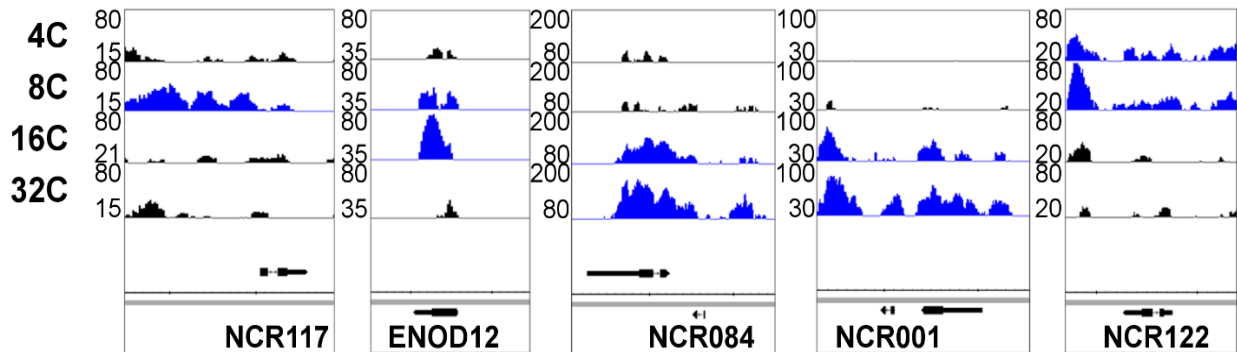
Furthermore, the correlation between the methylation status of genes and chromatin accessibility was assessed. Late genes of clusters VII-VIII and *NCR* genes of group 3 showed an increase in chromatin accessibility in 32C compared to 4C, independently of their hyper-, hypo- or unchanged methylation state (**Fig. 35**), indicating that chromatin opening can happen also in the absence of demethylation of the genes.



**Figure 35. Chromatin accessibility of differentially methylated nodule-specific and nodule expressed genes in 32C cells.** (A) Chromatin accessibility of 32C hyper- and hypomethylated VII-VIII cluster protein-coding genes and (B) group three *NCRs* at 32C. Box plots show the median, upper and lower quartiles, while whiskers show the 10-90th percentiles of the accessibility values. Asterisk indicates significant differences (Wilcoxon's rank sum test,  $P < 0.0039$ ). n indicates gene number

The local chromatin state of the five selected genes is shown in **Fig. 36**. Based on the ATAC-seq sequencing tag frequency, the chromatin shows relative low accessibility in the nuclei in which the corresponding gene is not expressed, while the chromatin strongly opens in nuclei in which the gene is transcriptionally active. In the *NCR117* region, the chromatin was relatively the most open in 8C in line with its early expression. *ENOD12* is present in a rather

open chromatin region where the accessibility was the highest in the 8C and 16C cells. For *NCR084* and *NCR001* the chromatin became highly open in 16C and 32C while in the case of *NCR122*, it was the most accessible in 4C and 8C. The genome viewer at the five selected loci revealed that changes in the local chromatin state are not limited to the promoter regions but can extend over the neighbouring regions.



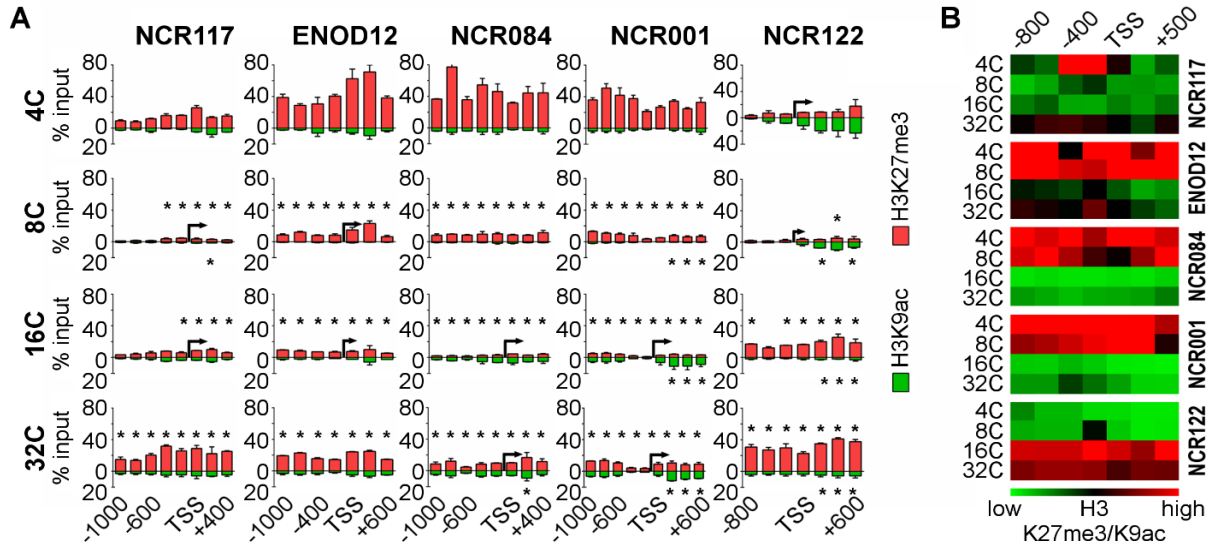
**Figure 36.** ATAC-Seq sequencing tag frequency profile along the *NCR117*, *ENOD12*, *NCR084*, *NCR001* and *NCR122* genes at different ploidy levels. Black indicates repressed state whereas blue the active state of the genes.

Thus taken together, the differential accessibility of chromatin regions at various ploidy levels was in line with the activity of the genes as well as with the expression of several neighbouring genes in the same chromatin microenvironment.

#### 4.2.4. Ploidy-dependent variation in H3K27me3 and H3K9ac chromatin marks correlates with the expression pattern of *NCR* genes

To analyse the correlation between chromatin modifications and gene expression, we investigated how the antagonistic chromatin marks H3K27me3 and H3K9ac correlate with activation of weakly (*NCR117*, *ENOD12*, *NCR084*) and strongly (*NCR001*, *NCR122*) expressed genes in different ploidy nuclei, using chromatin immunoprecipitation coupled to quantitative PCR (ChIP-qPCR) (**Fig. 37**). On the *NCR117* locus the level of H3K27me3 was relatively high in the 4C and 32C nuclei where this gene is repressed while it was very low especially close to the TSS in 8C and low in 16C where the gene is active ( $P < 0.05$ ). H3K9ac was observed in the 4C, 16C and 32C nuclei. In the case of *ENOD12*, *NCR084* and *NCR001* the H3K27me3 level was high in the 4C cells, then decreased throughout the entire region in the 8C and even further in the 16C cells before increasing again in the 32C cells. Similarly to *NCR117*, at all the three loci H3K9ac marks were observed in 4C, 16C and 32C nuclei while 8C was devoid of H3K9ac. The extent of H3K9ac in the gene body of *NCR084* and *NCR001* was particularly important in

16C and 32C. In the case of *NCR122*, low level of H3K27me3 together with high H3K9ac in the 4C and 8C nuclei coincided with gene activity while decreasing H3K9ac and increasing H3K27me3 with gradual repression of *NCR122* in 16C and 32C. As both H3K27me3 and H3K9ac contribute to gene regulation, we calculated the ratio of H3K27me3 and H3K9ac for these five genes upstream and downstream from the TSS (**Fig. 37B**).



**Figure 37. H3K27me3 and H3K9ac modifications of the *NCR117*, *ENOD12*, *NCR084*, *NCR001* and *NCR122* genes at different ploidy levels (4C-32C) detected with ChIP-qPCR. (A)** H3K27me3 (red) and H3K9ac (green) levels shown as percentages of input chromatin. X-axis represents the distance from TSS. Each column corresponds to a given primer pair used for this analysis. Black arrows at the TSS indicate those ploidy levels where the gene is active. Percent (%) of input values are averages  $\pm$  SEM of three independent experiments. **(B)** High (repression) to low (activation) ratio of H3K27me3 and H3K9ac levels for the five selected loci (at the indicated gene positions relative to the TSS) is indicated with red to green. Asterisks indicate significant difference compared to 4C (two-way ANOVA with Tukey's multiple comparison test,  $P < 0.05$ ).

In *NCR117*, the only presence of H3K27me3 in the -400 – TSS region in 4C cells could be responsible for gene silencing. In *NCR084* and *NCR001*, the high H3K27me3/H3K9ac ratio in 4C and 8C coincides with their repression while the lower values with their activation. The same tendency though less pronounced was valid also for *ENOD12*. In contrast, for *NCR122* the H3K27me3/H3K9ac ratio was low in 4C and 8C in line with high gene expression and an increased H3K27me3/H3K9ac ratio in 16C and 32C correlated with decreased expression. In general, H3K27me3 modification showed dynamic changes in both the promoter and the gene body as a function of ploidy levels while changes in H3K9ac were most important in the gene body. The repressed state of the genes coincided with high level of H3K27me3 and chromatin compaction while gene activation with a reduced level of H3K27me3 and an increased

H3K9ac/H3K27me3 ratio. The increased H3K9ac levels correlated with higher gene expression and open chromatin conformation. These data suggest that activities of the *NCRs* are dynamically regulated by their chromatin modifications.

#### **4.2.5. Conclusions**

These analyses show that the temporal/spatial expression of *NCRs* in nodules correlate with growing ploidy levels and changes in the epigenome during the ENR cycles, contributing likely to the regulation of *NCR* expression. Differential DNA methylation was found in only a small subset of symbiotic nodule-specific genes but it was detected in more than half of the *NCR* genes. This suggests that, surprisingly, the *NCR* genes are regulated differently from other nodule-specific genes which is in line with a report showing that the nodule specific MtDME DNA demethylase is key for the expression of many *NCR* genes (Satgé et al. 2016). Expression of nodule-specific genes correlated with ploidy-dependent opening of the chromatin as well as, in at least a subset of tested genes, with reduced H3K27me3 levels combined with enhanced H3K9ac levels. Our results suggest that ENR-dependent epigenetic changes contribute to transcriptional reprogramming and differentiation of symbiotic cells.



## 5. DISCUSSION

In this study, we showed that *NCR* genes are among the most specifically expressed genes in *M. truncatula*. Moreover, endoreduplication and the intracellular presence of rhizobia is required for *NCR* gene activation. The particular expression pattern of the *NCR* gene family suggested the involvement of epigenetic mechanisms in their transcriptional control and therefore, we investigated at whole genome level how DNA methylation, specific histone modifications and opening or closing of the chromatin correlate with gene expression. By identifying ploidy-specific DNA methylation, repressive H3K27me3 and activating H3K9ac histone marks and chromatin accessibility profiles of differentially expressed genes, we get a first view of a multi-layered epigenetic control of symbiotic cell differentiation. Our findings are a step forward in the understanding of the extremely tight regulation of *NCR* genes.

### 5.1. *NCR* genes are expressed during symbiosis in the rhizobia infected nodule cells

The *NCR* genes represent one of the largest and most diverse multigenic families in plants, which has been discovered in nodules of *M. truncatula* (Mergaert et al. 2003). *NCR* expression has been analysed so far with various methods (RT-PCR, *in situ* hybridization, immunolocalization, promoter-marker gene fusions, EST-analysis, macro- and microarrays, RNA-seq) and these studies detected *NCR* expression exclusively in nodules (Van de Velde et al. 2010; Farkas et al. 2014; Mergaert et al. 2003; Alunni et al. 2007; Nallu et al. 2013; Maunoury et al. 2010; Tesfaye et al. 2013; Sels et al. 2008; Roux et al. 2014). In this study, the expression pattern of 334 *NCR* genes was analysed over 267 different experiments including 9 plant organs, biotic and abiotic growth conditions and different developmental stages using the MtGEA and additional published transcriptome data. We showed that the hundreds of *NCR* genes encoding defensin-like peptides are activated only in nodules, except for five *NCR*s, which showed relaxed specificity. Out of these atypically expressed genes, *NCR218* and *NCR122* showed high expression level in uninfected nodule and root cells. Immunolocalization of the *NCR122* peptide revealed its presence in the uninfected nodule cells. Thus, *NCR122* and most likely also *NCR218* are the only *NCR* peptides that are specific to uninfected root and nodule cells. *NCR* genes are expressed in nodules but in no other plant organs or in response to phytohormones or drought and salt stress. During symbiosis, none of the *NCR* genes are induced by Nod factors and they are not activated in nodules before rhizobia are released in the host

cells and symbiotic cells are not formed. We also showed that *NCR* genes are also not involved in degradation of symbiotic host cells and bacteroids during nodule senescence since their expression shuts down when senescence is initiated therefore *NCR* peptides are not used by the plant to kill rhizobia.

## **5.2.NCRs are not involved in innate immunity**

NCRs resemble antimicrobial peptides such as defensins of the innate immunity and many NCR peptides, in particular the most cationic ones, have a strong *in vitro* antimicrobial activity against a diversity of Gram-negative and Gram-positive bacteria including human and plant pathogens as well as fungi (Van de Velde et al. 2010; Tiricz et al. 2013; Ördögh et al. 2014; Farkas et al. 2017; Balogh et al. 2014). This *ex-planta* killing effect is associated with pore formation and membrane permeabilization (Nagy et al. 2015; Mikuláss et al. 2016). Despite this, the *NCR* genes are not expressed during pathogen attack included interactions with bacteria, fungi, oomycetes and nematodes. They are also not expressed in “infection sensitive” organs like leaves, seeds and flowers which often express high levels of innate immunity antimicrobial peptides (Sels et al. 2008). Therefore, it seems that the NCR peptides have no function in innate immunity.

## **5.3.NCRs are activated sequentially in different spatio-temporal waves**

The *NCR* genes are activated in temporal waves during nodule organogenesis, which correlated well with bacterial infection of the nodule cells and with a specific spatial localization of their transcripts from the apical to the proximal nodule zones. *In situ* detection of *NCR* expression has demonstrated for all the tested genes that they are specifically expressed in the symbiotic nodule cells but different subsets of *NCR* genes are activated at different developmental stages (Van de Velde et al. 2010; Farkas et al. 2014; Mergaert et al. 2003, this work). The *NCR* genes are not activated by Nod factors during the very early stages of the nodule organogenesis when infected cells are not yet formed and in the nodule meristem (this work, Nallu et al. 2013; Maunoury et al. 2010). During the development of wild-type nodules, they are activated in consecutive waves and their first appearance coincides with the formation of infected symbiotic cells (Maunoury et al. 2010). We showed that *NCR* genes are activated during nodule development in at least three temporal waves corresponding to specific spatial expression patterns. Genes activated early in nodule development are expressed in the more distal nodule parts (close to the apex) while genes activated late during development are

expressed in the proximal nodule tissues. In addition, certain clusters of genes, once activated, maintain their activity when the tissues grow older while other clusters are characterized by a decline of their expression in the older nodule cells. Our spatial analysis of *NCR* expression is in strong agreement with a recently published study that used LCM of nodule zones coupled to RNA-Sequencing (Roux et al. 2014).

Transcriptome analysis of non-functional nodules formed by bacterial or plant symbiotic mutants arrested at different stages of nodule development, is also in agreement with the symbiotic nodule cell specific expression of all *NCR* genes and their transcriptional activation in the polyploid symbiotic cells (Maunoury et al. 2010). For example, in nodules of the *M. truncatula* TE7 mutant which is affected in the *IPD3* gene (Horvath et al. 2011; Ovchinnikova et al. 2011) and in nodules infected by the *S. meliloti* *exoY* mutant, no infected symbiotic cells are formed and nodule ploidy is lower, consequently these nodules do not express any of the *NCR* genes (Maunoury et al. 2010). On the contrary, in nodules infected by the *S. meliloti* *bacA* mutant which contain symbiotic cells with undifferentiated bacteroids, a subset of *NCR* genes is activated. In other mutants, forming normal symbiotic cells with differentiated bacteroids, *NCR* genes are activated to a similar extent as in the wild type (Nallu et al. 2013; Maunoury et al. 2010). Together, the expression pattern of the *NCR* genes suggests that the endosymbiotic rhizobia in the host cells are the only targets of the peptides. However, the distinct spatio-temporal profiles and their high sequence variety clearly indicate that *NCR* peptides have many different roles. Subsets of *NCR* genes that are expressed during the early stages of symbiotic cell formation might be involved in cell division arrest, elongation and polyploidization of the bacteroids, while others that are active in later stages of symbiotic cell formation or even after the completion of the symbiotic cell differentiation might have other functions in the bacteroids like influencing their metabolism (Kereszt et al. 2011).

#### **5.4. *NCR* genes are differentially expressed according to the ploidy level of the differentiating symbiotic cells**

Among other things, we showed in this study that nodule organogenesis is accompanied with major changes in the gene expression program (Maunoury et al. 2010; Roux et al. 2014, this work). *In situ* transcript localization, promoter-reporter gene fusions revealed that early genes are activated in young differentiating symbiotic cells while late genes in mature nitrogen-fixing cells (Mergaert et al. 2003; Maunoury et al. 2010; Van de Velde et al. 2010; Farkas et al. 2014; Roux et al. 2014, this work) and this pattern suggested that their expression could be

coupled to specific ploidy levels. The expression pattern of selected nodule-expressed and nodule-specific genes in sorted 4C, 8C, 16C and 32C nuclei was correlated with their spatial expression pattern in nodules (Roux et al. 2014). We found that activation of the early clusters I-IV genes occurs in the 4C post-meristematic cells whereas cluster V-VI and group 1 *NCR* genes are activated in the distal (ZII<sub>d</sub>) and proximal parts (ZII<sub>p</sub>) of ZII with 8C and 16C while the genes in clusters VII-VIII and group 3 *NCRs* are induced during later stages of nodule development in the IZ and ZIII in high ploidy level cells of 32C and 64C. We found very good correlation between the known spatial *NCR* expression patterns and the expression profiles in the nuclei of different ploidy levels.

### **5.5. *NCR* genes are among the most specifically expressed genes in *M. truncatula***

Quantifying the specificity of expression with the Shannon entropy factor reveals that the *NCR* genes, and more generally, nodule-specific genes are among the most specifically expressed genes in *M. truncatula*. This suggests that nodulation in *Medicago* is in large part depending on genes solely dedicated to this symbiotic process. These genes may be the results of gene duplications followed by neo-functionalization (for example the DNF2 protein (Bourcy et al. 2013) which has non-symbiotic homologues) or they may be unique for the symbiosis (possibly the *NCRs*, *GRPs*, *SNARPs* and others). In addition to that, the expression of the *NCR* genes in nodules reaches very high levels. Even if certain *NCR* genes are expressed at a low level, the majority of them are among the highest expressed genes in the whole genome of *Medicago*. This is in agreement with a previous estimation, based on EST counts, that all *NCR* mRNAs together constitute almost 5% of the total mRNA population in nodules (Mergaert et al. 2003). Genes with high tissue-specific expression (expression in a single or limited number of tissues as opposed to ubiquitous expression in all tissues) are silenced during most of the plant growth. Gene silencing can be achieved by different mechanisms such as small RNAs, DNA methylation but also by chromatin compacting mediated by H3K27me3 and PcG protein complexes.

## **5.6. Differential DNA methylation is rare but predominant in the *NCR* gene family**

Nodules derive from cortical cells in the root elongation-differentiation zone. Induced by the *Rhizobium* signal molecules, the Nod factors, cortical cells dedifferentiate and enter the mitotic cycle. Proliferation of these cells leads to the establishment of the meristem. DNA methylation in the meristematic cells represents the initial and likely, the most methylated status of DNA. Cells leaving the meristem and infected with rhizobia undergo consecutive waves of transcriptional reprogramming and this differential gene expression is coupled to ploidy level increases. The dynamics of DNA methylation at single-base resolution was assessed in uninfected (4C) and symbiotic cells (32C). The highest fraction of mCs was observed in CG context followed by CHG and CHH in both cell types, which was consistent with previous studies and indicates the importance of CG methylation. The effect of DNA methylation on repressing gene expression may be mediated either directly or indirectly via some transcriptional regulatory proteins, which can recognize mCs in the promoter regions and via recruitment of methyl CG-binding proteins to remodel chromatin and regulate gene expression. Another mechanism of gene silencing via inhibition of transcription activator binding due to promoter DNA methylation has also been reported (Watt & Molloy 1988; Medvedeva et al. 2014; Nan et al. 1998). However, evidence is emerging that some TFs can bind to methylated DNA and activate gene expression (Zhu et al. 2016). High level of DNA methylation of promoters usually provokes gene silencing, while demethylation was expected to play a major role in gene activation. In addition, we hypothesised that endoreduplication of the symbiotic nodule cells could lead to passive loss of DNA methylation during the consecutive endocycles, which is in line with the downregulation of *MtMET* methyltransferase gene in the symbiotic cells. However, only a small subset of early- and late genes exhibited methylation change between the 4C and 32C cells, which indicates that the DNA methylation state of the 4C cells has been copied for many genes during the repeated endoreduplication cycles. Unlike most protein-coding genes, *NCRs* behaved differently as out of 375 genes 164 became hypomethylated. This finding is in line with the recently reported upregulation of the nodule-specific *MtDME* DNA demethylase in the interzone cells as well as with downregulation of *NCR* genes in the *MtDME* RNA interference lines (Satgé et al. 2016). The differential methylation patterns that we observed could be in large part the consequence of the action of *MtDME*. Why the activation of certain *NCR* genes is associated with demethylation and not the others requires further investigation.

An attractive possibility for the importance of differential DNA methylation for the expression of a subset of *NCR* genes might be related to the presence of TEs (Satzé et al. 2016). Plant genomes are rich in TEs, which are usually maintained transcriptionally silent. This silencing of TEs may also affect neighbouring genes and *NCR* gene activation may thus require demethylation mediated by MtDME. Similarly, in *Arabidopsis* active demethylation of TEs correlated with up-regulation of neighbouring stress responsive genes (Le et al. 2014). Many transposon genes in *Medicago* are activated concomitantly with the *NCRs* during symbiotic cell differentiation (Satzé et al. 2016). Moreover, TEs can provide cryptic promoters or *cis*-acting regulatory sites behaving as enhancers, which can influence expression patterns or expression level of proximal genes.

### **5.7. Chromatin accessibility changes with the ploidy levels**

Chromatin structure has been implicated in developmental and tissue-specific regulation of a number of genes in both plants and animals. Dynamic opening and closing of the chromatin at different ploidy levels of the nodule cells correlated well with the active/repressed expression state of the genes. We also found that open chromatin alone was not sufficient for gene activation. Although, the chromatin accessibility showed small variations at the 4 ploidy levels the less condensed chromatin state was found in mature symbiotic cells with 32C. Thus, tissue-specific and developmental-stage specific expression patterns in the nodule are accompanied by dynamic alterations in chromatin structure. Chromatin reorganisation was particularly important close to or at the TSS. Early genes exhibited highest accessibility in 4C and 8C cells while late genes in 16C and 32C. Various studies have demonstrated that while increased chromatin accessibility generally precedes transcription activation, structural changes alone are often insufficient for gene activation (Armstrong & Emerson 1998; Pazin et al. 1994). Differential nodule-specific expression patterns are accompanied by discrete changes in local chromatin structure of the analysed genes. Growing evidence suggests that DNA methylation can influence other epigenetic modifications and can have an effect directly on the chromatin structure (Hussein et al. 2014; Lee et al. 2014). We also found that change in chromatin accessibility was not tightly linked to DNA methylation, as it occurred also when the DNA methylation status was not affected by the ploidy. Moreover, the degree of chromatin opening did not always show a direct correlation with the level of DNA methylation. TEs may change chromatin modifications in regions near genes, which in turn can influence gene expression

levels. Thus, increase in transposable element expression during differentiation further supported that reprogramming led to a transient opening of the chromatin.

### **5.8.Histone modifications: The level of H3K27me3 is crucial for gene expression and is epistatic over H3K9ac**

We analysed the repressive H3K27me3 and the activating H3K9ac profile of selected symbiosis-specific genes from different spatial and temporal expression classes in growing ploidy nuclei. The H3K27me3 marks were predominant in the 4C cells and are likely essential for the repressed state of *NCRs* and other symbiotic genes. The H3K27me3 marks were dramatically reduced in the 8C and 16C cells and increased again in the 32C cells. Such a reduction of the H3K27me3 marks could be sufficient for gene activation of the *NCR117* and *ENOD12* genes, which are already activated in 8C cells. For the *NCR084* and *NCR001* genes, which are activated from 16C on, reduced H3K27me3 marks together with the presence of the active H3K9ac marks and accessible chromatin structure contributed to gene activation. *NCRs* in 32C cells are usually expressed at very high levels, which in addition to reduced H3K27me3 have also the active H3K9ac marks in the gene body and the promoter and an open chromatin structure. On the other hand, in the case of early genes, the increase of H3K27me3 marks correlated with their repression in the 32C cells. The primary role of the H3K27me3 mark in regulation of these genes is in line with previous observations that deposition and dynamic regulation of the H3K27me3 mark are important in controlling tissue-specific gene expression and plant cell differentiation (Lafos et al. 2011). The widespread loss of H3K27me3 was accompanied with expression of the H3K27me3 demethylases KDM6a and low expression of the PRC2 complex members that catalyse trimethylation of H3K27 and a general opening of the chromatin (Hussein et al. 2014). Moreover, the global reduction of H3K27me3 lead to a loss of heterochromatin on TEs - since their silencing is linked to heterochromatin formation – and their expression was high during reprogramming. The two antagonistic histone modifications described here are consistent with previous studies in yeast, mouse, human and plants, and support a model in which H3K9ac enhances gene expression, while H3K27me3 represses it. Moreover, H3K27me3 and DNA methylation can suppress the positive effect of H3K9ac on gene expression level and that the combination of the two repressive modifications is probably even more effective (Zhou et al. 2010; Zilberman et al. 2007; Zhang et al. 2007).

## **5.9. Epigenome reconfiguration in endoreduplication-driven cell differentiation**

Our results are consistent with studies of somatic cell reprogramming and cellular differentiation in plants and animals, showing that gene expression is controlled by multiple regulatory processes from nucleosome positioning, through histone modification and DNA methylation (Hussein et al. 2014; Lee et al. 2014; Huang et al. 2015). We propose that the dynamic changes in the chromatin environment that we described here determine the cell-type specific transcriptional program by turning on or off the temporally and spatially expressed nodule-specific genes and therefore contribute to symbiotic cell differentiation. Importantly, this remodelling is associated with the ENR process that drives this differentiation. ENR is widespread in plants and can be found in many cell types, especially in those undergoing differentiation and expansion (Kondorosi et al. 2000; Larkins et al. 2001; Bramsiepe et al. 2010). Well studied examples are the endosperm in seeds, the root hairs and trichomes or the root cortical cells, which are all formed through ENR (Vanstraelen et al. 2009; Sugimoto-Shirasu et al. 2005; Bourdon et al. 2012; Wildermuth 2010). Bramsiepe et al. (2010) demonstrated that by compromising ENR, trichome fate could change to epidermal pavement cell fate and conversely they could restore the trichome fate by promoting ENR. In addition, the establishment of many plant-biotrophic interactions, and notably the formation of the plant cells that are at the interface with the microbial partner, require ENR. These interactions include the nodules discussed here but also mycorrhiza, or pathogenic interactions with powdery mildews and nematodes (Wildermuth 2010). Similar epigenetic processes as the ones we described here during the differentiation of nodule symbiotic cells could also be of importance in these other ENR-mediated differentiation programs. On the other hand, the formation of symbiotic nodule cells in *M. truncatula* involves the activation of a very specific transcriptional program, involving genes only expressed in these cells (Roux et al. 2014; Maunoury et al. 2010, this work). It is unclear if these other cellular differentiation processes involve a similarly profound transcriptional reprogramming and if they therefore require the same strict regulatory controls as the symbiotic nodule cells.

## **5.10. Concluding remarks**

Symbiotic nodule cells represent an attractive model for studying ploidy-dependent epigenetic changes. During symbiotic cell differentiation in *M. truncatula*, patterns of



developmental stage-specific gene expression are established. It is generally assumed that tissue-specificity is achieved through the action of selective, cell-type restricted TFs. However, additional levels of control *in vivo* strongly relies on the chromatin structure that restricts access to regulatory proteins. We showed that *NCR* genes are tightly regulated in a spatio-temporal manner during nodulation by epigenetic modifications, which play a crucial role in the modulation of the chromatin structure.

In the future by correlating ploidy-specific transcriptomes with the map of open chromatin assessed by ATAC-seq will enable to identify potential *cis*-regulatory elements that operate in differentiation-stage specific manner in the nodule cells. By identifying 4C, 8C, 16C and 32C specific ATAC-seq peaks overlapping with known TF binding sites or using *de novo* TF binding motif analysis transcriptional regulatory elements such as enhancers involved in the very specific regulation of symbiotic cell-specific genes could be discovered.

## REFERENCES

- Agger, K. et al., 2008. The emerging functions of histone demethylases. *Current Opinion in Genetics and Development*, 18(2), pp.159–168.
- Aichinger, E. et al., 2009. CHD3 proteins and polycomb group proteins antagonistically determine cell identity in *Arabidopsis*. *PLoS Genetics*, 5(8): e1000605
- Aichinger, E. et al., 2011. The CHD3 chromatin remodeler PICKLE and polycomb group proteins antagonistically regulate meristem activity in the *Arabidopsis* root. *The Plant cell*, 23(3), pp.1047–60.
- Alunni, B. et al., 2007. Genomic organization and evolutionary insights on GRP and NCR genes, two large nodule-specific gene families in *Medicago truncatula*. *Molecular Plant-Microbe Interactions*, 20(9), pp.1138–1148.
- Amselem, J. et al., 2011. Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genetics*, 7(8): e1002230.
- Ariel, F. et al., 2014. Noncoding transcription by alternative rna polymerases dynamically regulates an auxin-driven chromatin loop. *Molecular Cell*, 55(3), pp.383–396.
- Armstrong, J.A. & Emerson, B.M., 1998. Transcription of chromatin: These are complex times. *Current Opinion in Genetics and Development*, 8(2), pp.165–172.
- Balogh, E.P. et al., 2014. Anti-Chlamydial effect of plant peptides. *Acta Microbiologica Et Immunologica Hungarica*, 61(2), pp.229–239.
- Barker, D.G. et al., 1990. *Medicago truncatula*, a model plant for studying the molecular genetics of the *Rhizobium*-legume symbiosis. *Plant Molecular Biology Reporter*, 8(1), pp.40–49.
- Bastow, R. et al., 2004. Vernalization requires epigenetic silencing of FLC by histone methylation. *Nature*, 427(6970), pp.164–167.
- Becker, A. et al., 2004. Global changes in gene expression in *Sinorhizobium meliloti* 1021 under microoxic and symbiotic conditions. *Mol Plant Microbe Interact*, 17(3), pp.292–303.
- Bell, A.C. & Felsenfeld, G., 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, 405 (6785), pp.482–485.
- Benedito, V.A. et al., 2008. A gene expression atlas of the model legume *Medicago truncatula*. *Plant Journal*, 55(3), pp.504–513.
- Benhamed, M. et al., 2006. *Arabidopsis* GCN5, HD1, and TAF1/HAF2 interact to regulate histone acetylation required for light-responsive gene expression. *The Plant cell*, 18(11), pp.2893–2903.
- Berr, A., Shafiq, S. & Shen, W.-H., 2011. Histone modifications in transcriptional activation during plant development. *Biochimica et biophysica acta*, 1809(10), pp.567–76.

- Berry, S. & Dean, C., 2015. Environmental perception and epigenetic memory: Mechanistic insight through FLC. *Plant Journal*, 83(1), pp.133–148.
- Bestor, T.H., Edwards, J.R. & Boulard, M., 2015. Notes on the role of dynamic DNA methylation in mammalian development. *Proceedings of the National Academy of Sciences*, 112(22), pp.6796–6799.
- Bilichak, A. et al., 2012. The progeny of *Arabidopsis thaliana* plants exposed to salt exhibit changes in DNA methylation, histone modifications and gene expression. *PLoS ONE*, 7(1): e30515.
- Bird, A., 2002. DNA methylation patterns and epigenetic memory. *Genes and Development*, 16(1), pp.6–21.
- Birnbaum, K.D. & Roudier, F., 2017. Epigenetic memory and cell fate reprogramming in plants. *Regeneration*, 4(1), pp.15–20.
- Blackledge, N.P., Rose, N.R. & Klose, R.J., 2015. Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nature reviews. Molecular cell biology*, 16(11), pp.643–649.
- Boisson-Dernier, A. et al., 2001. *Agrobacterium rhizogenes*-transformed roots of *Medicago truncatula* for the study of nitrogen-fixing and endomycorrhizal symbiotic associations. *Mol Plant Microbe Interact*, 14(6), pp.695–700.
- Bonaldi, K. et al., 2011. Nodulation of *Aeschynomene afraspera* and *A. indica* by photosynthetic *Bradyrhizobium* Sp. strain ORS285: the nod-dependent versus the nod-independent symbiotic interaction. *Mol Plant Microbe Interact*, 24(11), pp.1359–1371.
- Bourcy, M. et al., 2013. *Medicago truncatula* DNF2 is a PI-PLC-XD-containing protein required for bacteroid persistence and prevention of nodule early senescence and defense-like reactions. *New Phytologist*, 197(4), pp.1250–1261.
- Bourdon, M. et al., 2012. Evidence for karyoplasmic homeostasis during endoreduplication and a ploidy-dependent increase in gene transcription during tomato fruit growth. *Development*, 139(20), pp.3817–3826.
- Bouyer, D. et al., 2011. Polycomb repressive complex 2 controls the embryo-to-seedling phase transition. *PLoS Genetics*, 7(3): e1002014.
- Boyer, L.A. et al., 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*, 441(7091), pp.349–53.
- Bracken, A.P. et al., 2006. Genome-wide mapping of polycomb target genes unravels their roles in cell fate transitions. *Genes and Development*, 20(9), pp.1123–1136.
- Bramsiepe, J. et al., 2010. Endoreplication controls cell fate maintenance. *PLoS Genet*, 6(6): e1000996.
- Brzeski, J. & Jerzmanowski, A., 2003. Deficient in DNA methylation 1 (DDM1) defines a novel family of chromatin-remodeling factors. *Journal of Biological Chemistry*, 278(2), pp.823–828.

- Buenrostro, J.D. et al., 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*, 10(12), pp.1213–1218.
- Cabeza, R. et al., 2014. An RNA sequencing transcriptome analysis reveals novel insights into molecular aspects of the nitrate impact on the nodule activity of *Medicago truncatula*. *Plant physiology*, 164(1), pp.400–11.
- Callaham, D., Deltredici, P. & Torrey, J.G., 1978. Isolation and Cultivation In vitro of Actinomycete Causing Root Nodulation in Comptonia. *Science*, 199(4331), pp.899–902.
- Cao, R. et al., 2002. Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. *Science*, 298(5595), pp.1039–1043.
- Cebolla, A. et al., 1999. The mitotic inhibitor ccs52 is required for endoreduplication and ploidy-dependent cell enlargement in plants. *EMBO J*, 18(16), pp.4476–4484.
- Chan, S.W.L., Henderson, I.R. & Jacobsen, S.E., 2005. Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nature reviews Genetics*, 6(5), pp.351–360.
- Chen, W.M. et al., 2003. Legume symbiotic nitrogen fixation by beta-proteobacteria is widespread in nature. *J Bacteriol*, 185(24), pp.7266–7272.
- Chevalier, C. et al., 2011. Elucidating the functional role of endoreduplication in tomato fruit development. *Annals of Botany*, 107(7), pp.1159–1169.
- Choi, C.S. & Sano, H., 2007. Abiotic- stress induces demethylation and transcriptional activation of a gene encoding a glycerophosphodiesterase- like protein in tobacco plants. *Molecular Genetics and Genomics*, 277, pp.589–600.
- Choi, J. et al., 2009. Resetting and regulation of FLOWERING LOCUS C expression during *Arabidopsis* reproductive development. *Plant Journal*, 57(5), pp.918–931.
- Choi, S.M. et al., 2012. HDA19 is required for the repression of salicylic acid biosynthesis and salicylic acid-mediated defense responses in *Arabidopsis*. *Plant Journal*, 71(1), pp.135–146.
- Choi, Y. et al., 2002. DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. *Cell*, 110(1), pp.33–42.
- Chou, M.X. et al., 2006. Thirteen nodule-specific or nodule-enhanced genes encoding products homologous to cysteine cluster proteins or plant lipid transfer proteins are identified in *Astragalus sinicus* L. by suppressive subtractive hybridization. *J Exp Bot*, 57(11), pp.2673–2685.
- Clapier, C.R. & Cairns, B.R., 2009. The biology of chromatin remodeling complexes. *Ann Rev Biochem*, 78, pp.273–304.
- Cokus, S.J. et al., 2008. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature*, 452(7184), pp.215–219.

- De Coninck, B. et al., 2013. Mining the genome of *Arabidopsis thaliana* as a basis for the identification of novel bioactive peptides involved in oxidative stress tolerance. *Journal of Experimental Botany*, 64(17), pp.5297–5307.
- Coninck, B.M. a De et al., 2010. *Arabidopsis thaliana* plant defensin AtPDF1.1 is involved in the plant response to biotic stress. *New Phytologist*, pp.1075–1088.
- Cordoba, E. et al., 2003. Antisense inhibition of NADH glutamate synthase impairs carbon/nitrogen assimilation in nodules of alfalfa (*Medicago sativa* L.). *Plant Journal*, 33(6), pp.1037–1049.
- Coruzzi, G. & Bush, D.R., 2001. Nitrogen and Carbon Nutrient and Metabolite Signaling in Plants. *Plant Physiology*, 125(January), pp.61–64.
- Coruzzi, G.M. & Zhou, L., 2001. Carbon and nitrogen sensing and signaling in plants: Emerging “matrix effects.” *Current Opinion in Plant Biology*, 4(3), pp.247–253.
- Cosma, M.P., Tanaka, T. & Nasmyth, K., 1999. Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter. *Cell*, 97(3), pp.299–311.
- Cosson, V. et al., 2006. *Medicago truncatula* transformation using leaf explants. *Methods in Molecular Biology*, 343, pp.115–127.
- Crockard, A. et al., 2002. A white clover nodulin gene, dd23b, encoding a cysteine cluster protein, is expressed in roots during the very early stages of interaction with *Rhizobium leguminosarum* biovar trifolii and after treatment with chitolipooligosaccharide Nod factors. *J Plant Res*, 115(6), pp.439–447.
- Cui, X. et al., 2016. REF6 recognizes a specific DNA sequence to demethylate H3K27me3 and regulate organ boundary formation in *Arabidopsis*. *Nature Genetics*, 48(6), pp.694–699.
- Czaja, L.F. et al., 2012. Transcriptional responses toward diffusible signals from symbiotic microbes reveal MtNFP- and MtDMI3-dependent reprogramming of host gene expression by arbuscular mycorrhizal fungal lipochitoooligosaccharides. *Plant Physiology*, 159(4), pp.1671–1685.
- Czernic, P. et al., 2015. Convergent Evolution of Endosymbiont Differentiation in Dalbergioid and Inverted Repeat-Lacking Clade Legumes Mediated by Nodule-Specific Cysteine-Rich Peptides. *Plant Physiology*, 169(2), pp.1254–1265.
- Damiani, I. et al., 2012. Plant genes involved in harbouring symbiotic rhizobia or pathogenic nematodes. *New Phytologist*, 194(2), pp.511–522.
- Deal, R.B. et al., 2007. Repression of flowering in *Arabidopsis* requires activation of FLOWERING LOCUS C expression by the histone variant H2A.Z. *The Plant cell*, 19(1), pp.74–83.
- Deal, R.B. & Henikoff, S., 2010. A Simple Method for Gene Expression and Chromatin Profiling of Individual Cell Types within a Tissue. *Developmental Cell*, 18(6), pp.1030–1040.

- Deal, R.B. & Henikoff, S., 2010. Gene regulation: A chromatin thermostat. *Nature*, 463(7283), pp.887–8.
- Delgado MJ, Bedmar EJ, D.J., 1998. Genes involved in the formation and assembly of rhizobial cytochromes and their role in symbiotic nitrogen fixation. *Adv Microb Physiol*, 40, pp.191–231.
- Delprete, P. & Mabberley, D.J., 1998. The Plant-Book: A Portable Dictionary of the Vascular Plants. *Brittonia*, 50, p.466
- Denarie, J., Debelle, F. & Prome, J.C., 1996. Rhizobium lipo-chitoooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. *Annu Rev Biochem*, 65, pp.503–535.
- Ding, B. & Wang, G.-L., 2015. Chromatin versus pathogens: the function of epigenetics in plant immunity. *Frontiers in Plant Science*, 6, p.675.
- Ding, Y., Fromm, M. & Avramova, Z., 2012. Multiple exposures to drought “train” transcriptional responses in *Arabidopsis*. *Nature Communications*, 3, p.740.
- Doris, W., 2003. Chromatin regulation of plant development. *Current Opinion in Plant Biology*, 6(1), pp.20–28.
- Downen, R.H. et al., 2012. Widespread dynamic DNA methylation in response to biotic stress. *Proc Natl Acad Sci U S A*, 109(32), pp.E2183-91.
- Downie, J.A., 2014. Legume nodulation. *Curr Biol*, 24(5), pp.R184-90.
- Downie, J.A. & Walker, S.A., 1999. Plant responses to nodulation factors. *Current Opinion in Plant Biology*, 2(6), pp.483–489.
- Doyle, J.J. & Luckow, M.A., 2003. The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. *Plant Physiology*, 131(3), pp.900–910.
- Durgo, H. et al., 2015. Identification of nodule-specific cysteine-rich plant peptides in endosymbiotic bacteria. *Proteomics*, 15(13), pp.2291–2295.
- Esseling, J.J., Lhuissier, F.G. & Emons, A.M., 2003. Nod factor-induced root hair curling: continuous polar growth towards the point of nod factor application. *Plant Physiology*, 132(4), pp.1982–1988.
- Fang, G., Yu, H. & Kirschner, M.W., 1998. Direct binding of CDC20 protein family members activates the anaphase-promoting complex in mitosis and G1. *Mol Cell*, 2(2), pp.163–171.
- Farkas, A. et al., 2017. Comparative analysis of the bacterial membrane disruption effect of two natural plant antimicrobial peptides. *Frontiers in Microbiology*, 8(51), pp.1–12.
- Farkas, A. et al., 2014. *Medicago truncatula* symbiotic peptide NCR247 contributes to bacteroid differentiation through multiple mechanisms. *Proc Natl Acad Sci U S A*, 111(14), pp.5183–5188.

- Fedorova, M. et al., 2002. Genome-wide identification of nodule-specific transcripts in the model legume *Medicago truncatula*. *Plant Physiology*, 130(2), pp.519–537.
- Feng, S. et al., 2010. Conservation and divergence of methylation patterning in plants and animals. *Proc Nat Acad Sci U S A*, 107(19), pp.8689–94.
- Feng, S. & Jacobsen, S.E., 2011. Epigenetic modifications in plants: an evolutionary perspective. *Current Opinion in Plant Biology*, 14(2), pp.179–186.
- Finan, T.M., Oresnik, I. & Bottacin, A., 1988. Mutants of *Rhizobium meliloti* defective in succinate metabolism. *Journal of Bacteriology*, 170(8), pp.3396–3403.
- Frommer, M. et al., 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Nat Acad Sci U S A*, 89(5), pp.1827–31.
- Fruhling, M. et al., 2000. A small gene family of broad bean codes for late nodulins containing conserved cysteine clusters. *Plant Science*, 152(1), pp.67–77.
- Fukaki, H., Taniguchi, N. & Tasaka, M., 2006. PICKLE is required for SOLITARY-ROOT/IAA14-mediated repression of ARF7 and ARF19 activity during *Arabidopsis* lateral root initiation. *Plant Journal*, 48(3), pp.380–389.
- Furuta, K. et al., 2011. The CKH2/PKL chromatin remodeling factor negatively regulates cytokinin responses in arabidopsis calli. *Plant and Cell Physiology*, 52(4), pp.618–628.
- Garg, R. et al., 2015. Divergent DNA methylation patterns associated with gene expression in rice cultivars with contrasting drought and salinity stress response. *Scientific Reports*, 5, p.14922.
- Garg, R. et al., 2014. Genomic survey, gene expression analysis and structural modeling suggest diverse roles of DNA methyltransferases in legumes. *PLoS ONE*, 9(2): e88947.
- Gaude, N. et al., 2012. Arbuscule-containing and non-colonized cortical cells of mycorrhizal roots undergo extensive and specific reprogramming during arbuscular mycorrhizal development. *Plant Journal*, 69(3), pp.510–528.
- Gehring, M. et al., 2006. DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell*, 124(3), pp.495–506.
- Gehring, M., Bubb, K.L. & Henikoff, S., 2009. Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science*, 324(5933), pp.1447–1451.
- Gene, R., And, G. & Vance2, C.P., 1981. Root Nodule Enzymes of Ammonia Assimilation in Alfalfa (*Medicago sativa* L.)’ developmental patterns and response to applied nitrogen. *Plant Physiol*, 67, pp.1198–1203.
- Gepts, P. et al., 2005. Legumes as a model plant family. Genomics for food and feed report of the cross-legume advances through genomics conference. *Plant Physiology*, 137(4), pp.1228–1235.

- Glazebrook, J., 2005. Contrasting Mechanisms of Defense Against Biotrophic and Necrotrophic Pathogens. *Annual Review of Phytopathology*, 43(1), pp.205–227.
- Goll, M.G. & Bestor, T.H., 2005. Eukaryotic cytosine methyltransferases. *Annu Rev Biochem*, 74, pp.481–514.
- Gong, Z. et al., 2002. ROS1, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase. *Cell*, 111(6), pp.803–814.
- Gonzalez-Sama, A. et al., 2006. Nuclear DNA endoreduplication and expression of the mitotic inhibitor Ccs52 associated to determinate and lupinoid nodule organogenesis. *Mol Plant Microbe Interact*, 19(2), pp.173–180.
- Goryshin, I.Y. & Reznikoff, W.S., 1998. Tn5 in vitro transposition. *Journal of Biological Chemistry*, 273(13), pp.7367–7374.
- Görisch, S.M., Lichter, P. & Rippe, K., 2005. Mobility of multi-subunit complexes in the nucleus: Accessibility and dynamics of chromatin subcompartments. *Histochemistry and Cell Biology*, 123(3), pp.217–228.
- Graham, M.A. et al., 2004. Computational identification and characterization of novel genes from legumes. *Plant Physiology*, 135(3), pp.1179–1197.
- Greer, E.L. et al., 2015. DNA methylation on N6-adenine in *C. elegans*. *Cell*, 161(4), pp.868–878.
- Grunstein, M., 1997. Histone acetylation in chromatin structure and transcription. *Nature*, 389(6649), pp.349–352.
- Gualtieri, G. & Bisseling, T., 2000. The evolution of nodulation. *Plant Mol Biol*, 42(1), pp.181–194.
- Guefrachi, I. et al., 2015. *Bradyrhizobium* BclA Is a Peptide Transporter Required for Bacterial Differentiation in Symbiosis with *Aeschynomene* Legumes. *Mol Plant Microbe Interact*, 28(11), pp.1155–1166.
- Guefrachi, I. et al., 2014. Extreme specificity of NCR gene expression in *Medicago truncatula*. *Bmc Genomics*, 15(1), p.712.
- Guimaraes, A.A. et al., 2012. Genetic and symbiotic diversity of nitrogen-fixing bacteria isolated from agricultural soils in the western Amazon by using cowpea as the trap plant. *Appl. Environ. Microbiol.* 78, pp.6726–6733.
- Gourion, B. et al., 2015. Rhizobium-legume symbioses: the crucial role of plant immunity. *Trends Plant Sci*, 20(3), pp.186-94.
- Gyorgyey, J. et al., 1991. Alfalfa heat shock genes are differentially expressed during somatic embryogenesis. *Plant Mol Biol*, 16, pp.999-1007.
- Gyorgyey, J. et al., 2000. Analysis of *Medicago truncatula* nodule expressed sequence tags. *Mol Plant Microbe Interact*, 13(1), pp.62–71.



- Haag, A.F. et al., 2011. Protection of *Sinorhizobium* against Host Cysteine-Rich Antimicrobial Peptides Is Critical for Symbiosis. *Plos Biology*, 9(10).
- Hammond-Kosack, K.E., 1996. Resistance Gene-Dependent Plant Defense Responses. *The Plant Cell Online*, 8(10), pp.1773–1791.
- Han, S.-K. et al., 2012. The SWI2/SNF2 chromatin remodeling ATPase BRAHMA represses abscisic acid responses in the absence of the stress stimulus in *Arabidopsis*. *The Plant cell*, 24(12), pp.4892–4906.
- Hann, D.R. & Rathjen, J.P., 2007. Early events in the pathogenicity of *Pseudomonas syringae* on *Nicotiana benthamiana*. *Plant Journal*, 49(4), pp.607–618.
- Hargreaves, D.C. & Crabtree, G.R., 2011. ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell research*, 21(3), pp.396–420.
- He, J. et al., 2009. The *Medicago truncatula* gene expression atlas web server. *BMC Bioinformatics*, 10, pp.441.
- He, Y., Michaels, S.D. & Amasino, R.M., 2003. Regulation of flowering time by histone acetylation in *Arabidopsis*. *Science*, 302(5651), pp.1751–1754.
- Heitz E, 1928. Das heterochromatin der moose. *I Jahrb Wiss Botanik.*, 69, pp.762–818.
- Hellriegel, H. & Wilfarth, H., 1888. Untersuchungen über die Stickstoffnahrung der Gramineen und Leguminosen. *Beilagenheft zur Zeitschrift des Vereins f. d. Rübenzucker-Industrie d. D. R.*, pp 234.
- Henderson, I.R. & Jacobsen, S.E., 2007. Epigenetic inheritance in plants. *Nature*, 447(7143), pp.418–424.
- Henikoff, S. & Shilatifard, A., 2011. Histone modification: Cause or cog? *Trends in Genetics*, 27(10), pp.389–396.
- Hennig, L. & Derkacheva, M., 2009. Diversity of Polycomb group complexes in plants: same rules, different players? *Trends in Genetics*, 25(9), pp.414–423.
- Heo, J.B. & Sung, S., 2011. Vernalization-Mediated Epigenetic Silencing by a Long Intronic Noncoding RNA. *Science*, 331(6013), pp.76–79.
- Herridge, D.F., Peoples, M.B. & Boddey, R.M., 2008. Global inputs of biological nitrogen fixation in agricultural systems. *Plant and Soil*, 311(1–2), pp.1–18.
- Ho, L. et al., 2009. An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proc Nat Acad Sci U S A*, 106(13), pp.5187–91.
- Ho, L. et al., 2011. esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. *Nature cell biology*, 13(8), pp.903–13.

- Hogekamp, C. et al., 2011. Laser Microdissection Unravels Cell-Type-Specific Transcription in Arbuscular Mycorrhizal Roots, Including CAAT-Box Transcription Factor Gene Expression Correlating with Fungal Contact and Spread. *Plant Physiology*, 157(4), pp.2023–2043.
- Horvath, B. et al., 1986. Organization, structure and symbiotic function of *Rhizobium meliloti* nodulation genes determining host specificity for alfalfa. *Cell*, 46, pp.335-343.
- Horvath, B. et al., 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(49), pp.15232–15237.
- Horvath, B. et al., 2011. *Medicago truncatula* IPD3 is a member of the common symbiotic signaling pathway required for rhizobial and mycorrhizal symbioses. *Mol Plant Microbe Interact*, 24(11), pp.1345–1358.
- Horvath, B. et al., 1986. Organization, structure and symbiotic function of *Rhizobium meliloti* nodulation genes determining host specificity for alfalfa. *Cell*, 46(3), pp.335–343.
- Hsieh, T.-F. et al., 2009. Genome-wide demethylation of *Arabidopsis* endosperm. *Science*, 324(5933), pp.1451–4.
- Huang, K. et al., 2015. Dynamically reorganized chromatin is the key for the reprogramming of somatic cells to pluripotent cells. *Sci Rep*, 5, pp.17691.
- Hussein, S.M. et al., 2014. Genome-wide characterization of the routes to pluripotency. *Nature*, 516(7530), pp.198–206.
- Ikeda, Y. & Kinoshita, T., 2009. DNA demethylation: A lesson from the garden. *Chromosoma*, 118(1), pp.37–41.
- Imin, N. et al., 2008. Genome-wide transcriptional analysis of super-embryogenic *Medicago truncatula* explant cultures. *BMC plant biology*, 8, p.110.
- Irizarry, R.A. et al., 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4(2), pp.249–264.
- Jayaraman, D. et al., 2014. Response of *Medicago truncatula* Seedlings to Colonization by *Salmonella enterica* and *Escherichia coli* O157:H7. *Plos One*, 9(2): e87970.
- Jeddeloh, J. a, Stokes, T.L. & Richards, E.J., 1999. Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nature genetics*, 22(1), pp.94–97.
- Jegu, T. et al., 2014. The BAF60 Subunit of the SWI/SNF Chromatin-Remodeling Complex Directly Controls the Formation of a Gene Loop at FLOWERING LOCUS C in *Arabidopsis*. *The Plant Cell*, 26(2), pp.538–551.
- Jiao, Y.S. et al., 2015. Rhizobial Diversity and Nodulation Characteristics of the Extremely Promiscuous Legume *Sophora flavescens*. *Mol Plant Microbe Interact*, 28(12), pp.1338–1352.

- Jing, Y. et al., 2013. *Arabidopsis* Chromatin Remodeling Factor PICKLE Interacts with Transcription Factor HY5 to Regulate Hypocotyl Cell Elongation. *The Plant Cell*, 25(1), pp.242–256.
- Kajjalainen, S., Schroda, M. & Lindstrom, K., 2002. Cloning of nodule-specific cDNAs of *Galega orientalis*. *Physiol Plant*, 114(4), pp.588–593.
- Kant, C., Pradhan, S. & Bhatia, S., 2016. Dissecting the root nodule transcriptome of chickpea (*Cicer arietinum* L.). *PLoS ONE*, 11(6), pp.1–21.
- Kardailsky, I. et al., 1993. The pea late nodulin gene PsNOD6 is homologous to the early nodulin genes PsENOD3/14 and is expressed after the leghaemoglobin genes. *Plant Mol Biol*, 23(5), pp.1029–1037.
- Karimi, M., De Meyer, B. & Hilson, P., 2005. Modular cloning in plant cells. *Trends in Plant Science*, 10(3), pp.103–105.
- Kato, M. et al., 2003. Role of CG and non-CG methylation in immobilization of transposons in *Arabidopsis*. *Current Biology*, 13(5), pp.421–426.
- Kato, T. et al., 2002. Expression of genes encoding late nodulins characterized by a putative signal peptide and conserved cysteine residues is reduced in ineffective pea nodules. *Mol Plant Microbe Interact*, 15(2), pp.129–137.
- Kawakatsu, T. et al., 2016. Unique cell-type-specific patterns of DNA methylation in the root meristem. *Nat Plants*, 2(5), pp.16058.
- Kazmierczak, T. et al., 2017. Specific host-responsive associations between *Medicago truncatula* accessions and *Sinorhizobium* strains. *Mol Plant Microbe Interact*, 30(5), pp.399–409.
- Keating, S.T., 2015. Current epigenetic perspective on diabetes - Who regulates the regulators? *Chromatin Modifications in Health and Disease*, p.<https://fortyfourandtwo.net/tag/transcriptome/>.
- Kereszt, A., Mergaert, P. & Kondorosi, E., 2011. Bacteroid Development in Legume Nodules: Evolution of Mutual Benefit or of Sacrificial Victims? *Mol Plant Microbe Interact*, 24(11), pp.1300–1309.
- Kevei, Z. et al., 2002. Glycine-rich proteins encoded by a nodule-specific gene family are implicated in different stages of symbiotic nodule development in *Medicago* spp. *Mol Plant Microbe Interact*, 15(9), pp.922–931.
- Kim, J.-M. et al., 2015. Chromatin changes in response to drought, salinity, heat, and cold stresses in plants. *Frontiers in Plant Science*, 6, pp.114.
- Kim, J.M. et al., 2008. Alterations of Lysine Modifications on the Histone H3 N- Tail under Drought Stress Conditions in *Arabidopsis thaliana*. *Plant Cell Physiology*, 49, pp.1580–1588.
- Kim, J.M. et al., 2012. Transition of chromatin status during the process of recovery from drought stress in *Arabidopsis thaliana*. *Plant and Cell Physiology*, 53(5), pp.847–856.

- Kim, M. et al., 2015. An antimicrobial peptide essential for bacterial survival in the nitrogen-fixing symbiosis. *Proc Natl Acad Sci U S A*, 112(49), pp.15238–15243.
- King, R.W. et al., 1995. A 20s complex containing CDC27 and CDC16 catalyzes the mitosis-specific conjugation of ubiquitin to cyclin B. *Cell*, 81(2), pp.279–288.
- Kinoshita, T., 2004. One-Way Control of FWA Imprinting in Arabidopsis Endosperm by DNA Methylation. *Science*, 303(5657), pp.521–523.
- Kistner, C. & Parniske, M., 2002. Evolution of signal transduction in intracellular symbiosis. *Trends Plant Sci*, 7(11), pp.511–518.
- Knizewski, L., Ginalski, K. & Jerzmanowski, A., 2008. Snf2 proteins in plants: gene silencing and beyond. *Trends in Plant Science*, 13(10), pp.557–565.
- Kondorosi, A. et al., 1985. Identification and organization of *Rhizobium meliloti* genes relevant to the initiation and development of nodules. *Nitrogen Fixation Research Progress*, pp.73–78.
- Kondorosi, E. & Kondorosi, A. 1986. Nodule induction on plant roots by *Rhizobium*. *Trends in Biochemical Sciences*, 11(7), 296-299.
- Kondorosi, E., Mergaert, P. & Kereszt, A., 2013. A Paradigm for Endosymbiotic Life: Cell Differentiation of *Rhizobium* Bacteria Provoked by Host Plant Factors. *Annual Review of Microbiology*, 67, pp.611–628.
- Kondorosi, E., Roudier, F. & Gendreau, E., 2000. Plant cell-size control: growing by ploidy? *Current Opinion in Plant Biology*, 3(6), pp.488–492.
- Kouzarides, T., 2007. Chromatin Modifications and Their Function. *Cell*, 128(4), pp.693–705.
- Krueger, F. & Andrews, S.R., 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, 27(11), pp.1571–1572.
- Kurdistani, S.K., Tavazoie, S. & Grunstein, M., 2004. Mapping global histone acetylation patterns to gene expression. *Cell*, 117(6), pp.721–733.
- Kuzmichev, A. et al., 2002. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes & development*, 16(22), pp.2893–905.
- Kwon, C.S., Chen, C. & Wagner, D., 2005. WUSCHEL is a primary target for transcriptional regulation by SPLAYED in dynamic control of stem cell fate in *Arabidopsis*. *Genes Dev*, 19(8), pp.992–1003.
- Lafos, M. et al., 2011. Dynamic regulation of H3K27 trimethylation during *Arabidopsis* differentiation. *PLoS Genet*, 7(4), p.e1002040.
- Lämke, J. & Bäurle, I., 2017. Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants. *Genome Biology*, 18(1), pp.124.

- Langmead, B. & Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 9(4), pp.357–359.
- Laporte, P. et al., 2010. A novel RNA-binding peptide regulates the establishment of the *Medicago truncatula*-*Sinorhizobium meliloti* nitrogen-fixing symbiosis. *Plant Journal*, 62(1), pp.24–38.
- Larkins, B.A. et al., 2001. Investigating the hows and whys of DNA endoreduplication. *J Exp Bot*, 52(355), pp.183–192.
- Law, J.A. & Jacobsen, S.E., 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature reviews. Genetics*, 11(3), pp.204–20.
- Le, T.-N. et al., 2014. DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in Arabidopsis. *Genome Biology*, 15(9), pp.458.
- Lee, D.S. et al., 2014. An epigenomic roadmap to induced pluripotency reveals DNA methylation as a reprogramming modulator. *Nat Commun*, 5, p.5619.
- Lee, T.I. et al., 2006. Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells. *Cell*, 125(2), pp.301–313.
- Lelandais-Briere, C. et al., 2009. Genome-wide *Medicago truncatula* small RNA analysis revealed novel microRNAs and isoforms differentially regulated in roots and nodules. *Plant Cell*, 21(9), pp.2780–2796.
- Lev Maor, G., Yearim, A. & Ast, G., 2015. The alternative role of DNA methylation in splicing regulation. *Trends Genet*, 31(5), pp.274–280.
- Li, B., Carey, M. & Workman, J.L., 2007. The Role of Chromatin during Transcription. *Cell*, 128(4), pp.707–719.
- Li, D. et al., 2009. An expression database for roots of the model legume *Medicago truncatula* under salt stress. *BMC genomics*, 10(1), p.517.
- Li, T. et al., 2013. Jumonji C Domain Protein JMJ705-Mediated Removal of Histone H3 Lysine 27 Trimethylation Is Involved in Defense-Related Gene Activation in Rice. *The Plant Cell*, 25(11), pp.4725–4736.
- Liang, Y. et al., 2013. Nonlegumes respond to rhizobial Nod factors by suppressing the innate immune response. *Science*, 341, pp. 1384–1387
- Limpens, E. et al., 2013. cell- and tissue-specific transcriptome analyses of *Medicago truncatula* root nodules. *Plos One*, 8(5), p.e64377.
- Limpens, E. & Bisseling, T., 2014. CYCLOPS: a new vision on rhizobium-induced nodule organogenesis. *Cell Host Microbe*, 15(2), pp.127–129.
- Lippman, Z. et al., 2004. Role of transposable elements in heterochromatin and epigenetic control. *Nature*, 430(6998), pp.471–476.

- Lister, R. et al., 2008. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell*, 133(3), pp.523–536.
- Liu, J. et al., 2006. Recruitment of novel calcium-binding proteins for root nodule symbiosis in *Medicago truncatula*. *Plant physiology*, 141(1), pp.167–177.
- Lodwig, E.M. et al., 2003. Amino-acid cycling drives nitrogen fixation in the legume–*Rhizobium* symbiosis. *Nature*, 422(6933), pp.722–726.
- Lotocka, B. et al., 2000. Formation and abortion of root nodule primordia in *Lupinus luteus* L. *Acta Biologica Cracoviensia Series Botanica*, 42(1), pp.87–102.
- Low, D.A., Weyand, N.J. & Mahan, M.J., 2001. Roles of DNA adenine methylation in regulating bacterial gene expression and virulence. *Infection and Immunity*, 69(12), pp.7197–7204.
- Lu, F. et al., 2011. *Arabidopsis* REF6 is a histone H3 lysine 27 demethylase. *Nature genetics*, 43(7), pp.715–719.
- Luk, E. et al., 2010. Stepwise histone replacement by SWR1 requires dual activation with histone H2A.Z and canonical nucleosome. *Cell*, 143(5), pp.725–736.
- MacLean, A.M., Finan, T.M. & Sadowsky, M.J., 2007. Genomes of the symbiotic nitrogen-fixing bacteria of legumes. *Plant Physiology*, 144(2), pp.615–622.
- Mah, K.M. et al., 2012. Gene expression profiling of *Macrophomina phaseolina* infected *Medicago truncatula* roots reveals a role for auxin in plant tolerance against the charcoal rot pathogen. *Physiological and Molecular Plant Pathology*, 79, pp.21–30.
- Maillet, F. et al., 2011. Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature*, 469(7328), pp.58–63.
- Malik, H.S. & Henikoff, S., 2003. Phylogenomics of the nucleosome. *Nature structural biology*, 10(11), pp.882–91.
- March-Díaz, R. et al., 2008. Histone H2A.Z and homologues of components of the SWR1 complex are required to control immunity in *Arabidopsis*. *Plant Journal*, 53(3), pp.475–487.
- Maroti, G. et al., 2011. Natural roles of antimicrobial peptides in microbes, plants and animals. *Research in Microbiology*, 162(4), pp.363–374.
- Maunoury, N. et al., 2010. Differentiation of symbiotic cells and endosymbionts in *Medicago truncatula* nodulation are coupled to two transcriptome-switches. *Plos One*, 5(3), p.e9519.
- Medvedeva, Y.A. et al., 2014. Effects of cytosine methylation on transcription factor binding sites. *BMC genomics*, 15(1), p.119.
- Meissner, A. et al., 2005. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res*, 33(18), pp.5868–5877.

- Mergaert, P. et al., 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 Physiology*, 132(1), pp.161–173.
- Mergaert, P. et al., 2006. Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*-legume symbiosis. *Proc Nat Acad Sci U S A*, 103(13), pp.5230–5235.
- Messinese, E. et al., 2007. A novel nuclear protein interacts with the symbiotic DMI3 calcium- and calmodulin-dependent protein kinase of *Medicago truncatula*. *Mol Plant Microbe Interact*, 20(8), pp.912–921.
- Michiels, J. et al., 1998. *Phaseolus vulgaris* is a non-selective host for nodulation. *Fems Microbiology Ecology*, 26(3), pp.193–205.
- Mikuláss, K.R. et al., 2016. Antimicrobial nodule-specific cysteine-rich peptides disturb the integrity of bacterial outer and inner membranes and cause loss of membrane potential. *Annals of Clinical Microbiology and Antimicrobials*, 15(1), p.43.
- Misteli, T. et al., 2007. Beyond the Sequence: Cellular Organization of Genome Function. *Cell*, 128(4), pp.787–800.
- Mizuguchi, G. et al., 2004. ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science*, 303(5656), pp.343–8.
- Mondy, S. et al., 2014. GOLLUM [FeFe]-hydrogenase-like proteins are essential for plant development in normoxic conditions and modulate energy metabolism. *Plant Cell and Environment*, 37(1), pp.54–69.
- Montiel, J. et al., 2017. Morphotype of bacteroids in different legumes correlates with the number and type of symbiotic NCR peptides. *Proc Nat Acad Sci*, 114(19), pp.5041-5046.
- Montiel, J. et al., 2016. Terminal Bacteroid Differentiation Is Associated With Variable Morphological Changes in Legume Species Belonging to the Inverted Repeat-Lacking Clade. *Mol Plant Microbe Interact*, 29(3), pp.210–219.
- Morao, A.K., Bouyer, D. & Roudier, F., 2016. Emerging concepts in chromatin-level regulation of plant cell differentiation: timing, counting, sensing and maintaining. *Current Opinion in Plant Biology*, 34, pp.27–34.
- Moreau, S. et al., 2011. Transcription reprogramming during root nodule development in *Medicago truncatula*. *Plos One*, 6(1), p.e16463.
- Mulligan, J. T., and Long, S. R. 1985. Induction of *Rhizobium meliloti* nodC expression by plant exudate requires nodD. *Proc. Natl. Acad. Sci. U.S.A.* 82, 6609–6613.
- Mus, F. et al., 2016. Symbiotic nitrogen fixation and the challenges to its extension to nonlegumes. *Applied and Environmental Microbiology*, 82(13), pp.3698–3710.
- Mylona, P., Pawlowski, K. & Bisseling, T., 1995. Symbiotic Nitrogen Fixation. *Plant Cell*, 7(7), pp.869–885.

- Nagy, K. et al., 2015. Interaction of cysteine-rich cationic antimicrobial peptides with intact bacteria and model membranes. *General physiology and biophysics*, 34(2), pp.135–144.
- Nallu, S. et al., 2013. Regulatory patterns of a large family of defensin-like genes expressed in nodules of *Medicago truncatula*. *Plos One*, 8(4), p.e60355.
- Nan, X. et al., 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, 393(6683), pp.386–389.
- Naoumkina, M. et al., 2007. Different mechanisms for phytoalexin induction by pathogen and wound signals in *Medicago truncatula*. *Proc Nat Acad Sci U S A*, 104(46), pp.17909–15.
- Narlikar, G.J., Sundaramoorthy, R. & Owen-Hughes, T., 2013. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell*, 154(3), pp.490–503.
- Ng, H.H. & Bird, A., 1999. DNA methylation and chromatin modification. *Current opinion in genetics & development*, 9(2), pp.158–63.
- Noh, Y.S. & Amasino, R.M., 2003. PIE1, an ISWI family gene, is required for FLC activation and floral repression in *Arabidopsis*. *Plant Cell*, 15(7), pp.1671–1682.
- Nouwen, N. et al., 2017. The role of rhizobial (NifV) and plant (FEN1) homocitrate synthases in *Aeschynomene*/photosynthetic *Bradyrhizobium* symbiosis. *Scientific Reports*, 7(1), p.448.
- Oldroyd, G.E.D. & Downie, J.M., 2008. Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annu Rev Plant Biol*, 59, pp.519–546.
- Ong, C. & Corces, V., 2011. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature Reviews Genetics*, 12(4), pp.283–93.
- Oono, R. et al., 2010. Multiple evolutionary origins of legume traits leading to extreme rhizobial differentiation. *New Phytologist*, 187(2), pp.508–520.
- Ovchinnikova, E. et al., 2011. IPD3 controls the formation of nitrogen-fixing symbiosomes in pea and *Medicago* Spp. *Mol Plant Microbe Interact*, 24(11), pp.1333–1344.
- Ördögh, L. et al., 2014. Symbiotic Plant Peptides Eliminate *Candida albicans* Both *in Vitro* and in an Epithelial Infection Model and Inhibit the Proliferation of Immortalized Human Cells. *BioMed Research International*, 2014:320796.
- Park, S.-H. et al., 2004. Ultrastructure of Human Embryonic Stem Cells and Spontaneous and Retinoic Acid-Induced Differentiating Cells. *Ultrastructural Pathology*, 28(4), pp.229–238.
- Pazin, M.J., Kamakaka, R.T. & Kadonaga, J.T., 1994. ATP-dependent nucleosome reconfiguration and transcriptional activation from preassembled chromatin templates. *Science*, 266(5193), pp.2007–11.
- Penterman, J. et al., 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(9), pp.3561–3566.



- Perret, X., Staehelin, C. & Broughton, W.J., 2000. Molecular basis of symbiotic promiscuity. *Microbiol Mol Biol Rev*, 64(1), pp.180–201.
- Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*, 29(9), p.e45.
- Pichon, M. et al., 1992. *Rhizobium meliloti* elicits transient expression of the early nodulin gene ENOD12 in the differentiating root epidermis of transgenic alfalfa. *Plant Cell*, 4(10), pp.1199–1211.
- Pini, F. et al., 2015. Cell Cycle Control by the Master Regulator CtrA in *Sinorhizobium meliloti*. *PLoS Genet*, 11(5), p.e1005232.
- Pini, F. et al., 2013. The DivJ, CbrA and PleC system controls DivK phosphorylation and symbiosis in *Sinorhizobium meliloti*. *Mol Microbiol*, 90(1), pp.54–71.
- Popp, C. et al., 2010. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature*, 463(7284), pp.1101–5.
- Pueppke, S.G. & Broughton, W.J., 1999. *Rhizobium* sp. strain NGR234 and *R. fredii* USDA257 share exceptionally broad, nested host ranges. *Mol Plant Microbe Interact*, 12(4), pp.293–318.
- Qiao, Y. et al., 2014. Dual Roles of Histone H3 Lysine 9 Acetylation in Human Embryonic Stem Cell Pluripotency and Neural Differentiation. *The Journal of biological chemistry*. 290(16), pp.9949
- Radutoiu, S. et al., 2007. LysM domains mediate lipochitin-oligosaccharide recognition and Nfr genes extend the symbiotic host range. *EMBO J*, 26(17), pp.3923–35.
- Ratet P, Wen J, Cosson V, Tadege M, M.K., 2010. Tnt1 Induced Mutations in Medicago: Characterisation and Applications. W. Edited by Meksem K, Kahl G. KGaA, ed., Wiley-VCH Verlag GmbH & Co.
- Rey, T. et al., 2013. NFP, a LysM protein controlling Nod factor perception, also intervenes in *Medicago truncatula* resistance to pathogens. *New Phytologist*, 198(3), pp.875–886.
- Rightmyer, A.P. & Long, S.R., 2011. Pseudonodule formation by wild-type and symbiotic mutant *Medicago truncatula* in response to auxin transport inhibitors. *Mol Plant Microbe interact*, 24(11), pp.1372–84.
- Roh, T.Y., Cuddapah, S. & Zhao, K., 2005. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes & Development*, 19(5), pp.542–552.
- Roudier, F. et al., 2011. Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. *The EMBO Journal*, 30(10), pp.1928–1938.
- Roudier, F., Teixeira, F.K. & Colot, V., 2009. Chromatin indexing in *Arabidopsis*: an epigenomic tale of tails and more. *Trends in Genetics*, 25(11), pp.511–517.

- Roux, B. et al., 2014. An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant Journal*, 77(6), pp.817–837.
- Russo, V.E.A., Martienssen, R.A. & Riggs, A.D., 1996. Epigenetic mechanisms of gene regulation. *Cold Spring Harbor Laboratory Press*, p.692.
- Sabelli, P.A. & Larkins, B.A., 2009. The contribution of cell cycle regulation to endosperm development. *Sexual Plant Reproduction*, 22(4), pp.207–219.
- Saito, Y., Tsuji, J. & Mituyama, T., 2014. Bisulfighter: accurate detection of methylated cytosines and differentially methylated regions. *Nucleic Acids Res*, 42(6), p.e45.
- Sani, E. et al., 2013. Hyperosmotic priming of *Arabidopsis* seedlings establishes a long-term somatic memory accompanied by specific changes of the epigenome. *Genome Biology*, 14(6), p.R59.
- Satgé, C. et al., 2016. Reprogramming of DNA methylation is critical for nodule development in *Medicago truncatula*. *Nat Plants*, 2(11), p.16166.
- Saze, H., Mittelsten Scheid, O. & Paszkowski, J., 2003. Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nat Genet*, 34(1), pp.65–9.
- Scheres, B. et al., 1990. Sequential induction of nodulin gene expression in the developing pea nodule. *Plant Cell*, 2(8), pp.687–700.
- Schubeler, D. et al., 2004. The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes & Development*, 18(11), pp.1263–1271.
- Schug, J. et al., 2005. Promoter features related to tissue specificity as measured by Shannon entropy. *Genome biology*, 6(4), p.R33.
- Schultze, M. & Kondorosi, A., 1998. Regulation of symbiotic root nodule development. *Annu Rev Genet*, 32, pp.33–57.
- Schwartz, Y.B. et al., 2006. Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nature genetics*, 38(6), pp.700–705.
- Seabra, A.R. et al., 2012. Inhibition of glutamine synthetase by phosphinothricin leads to transcriptome reprogramming in root nodules of *Medicago truncatula*. *Mol Plant Microbe Interact*, 25(7), pp.976–92.
- Sels, J. et al., 2008. Plant pathogenesis-related (PR) proteins: A focus on PR peptides. *Plant Physiology and Biochemistry*, 46(11), pp.941–950.
- Sen, D. & Weaver, R.W., 1980. Nitrogen-Fixing Activity of Rhizobial Strain-32h1 in Peanut and Cowpea Nodules. *Plant Science Letters*, 18(3), pp.315–318.
- Shangguan, L. et al., 2013. Evaluation of genome sequencing quality in selected plant species using expressed sequence tags. *Plos One*, 8(7), p.e69890.

- Sheldon, C.C. et al., 2008. Resetting of FLOWERING LOCUS C expression after epigenetic repression by vernalization. *Proceedings of the National Academy of Sciences*, 105(6), pp.2214–2219.
- Shen, Y. et al., 2015. Chromodomain, Helicase and DNA-binding CHD1 protein, CHR5, are involved in establishing active chromatin state of seed maturation genes. *Plant Biotechnology Journal*, 13(6), pp.811–820.
- Shu, H. et al., 2012. Distinct modes of DNA accessibility in plant chromatin. *Nat Commun*, 3, p.1281.
- Singh, S. et al., 2014. CYCLOPS, a DNA-binding transcriptional activator, orchestrates symbiotic root nodule development. *Cell Host Microbe*, 15(2), pp.139–152.
- Song, Q.X. et al., 2013. Genome-wide analysis of DNA methylation in soybean. *Mol Plant*, 6(6), pp.1961–1974.
- Song, Y. et al., 2012. The dynamic changes of DNA methylation and histone modifications of salt responsive transcription factor genes in soybean. *PLoS ONE*, 7(7).
- Soyer, J.L. et al., 2014. Epigenetic control of effector gene expression in the plant pathogenic fungus *Leptosphaeria maculans*. *PLoS Genet*, 10(3), p.e1004227.
- Spangrude, G.J., Heimfeld, S. & Weissman, I.L., 1988. Purification and characterization of mouse hematopoietic stem cells. *Science*, 241(4861), pp.58–62.
- Steward, N. et al., 2002. Periodic DNA methylation in maize nucleosomes and demethylation by environmental stress. *Journal of Biological Chemistry*, 277(40), pp.37741–37746.
- Strahl, B.D. & Allis, C.D., 2000. The language of covalent histone modifications. *Nature*, 403(6765), pp.41–45.
- Sugawara, M. et al., 2013. Comparative genomics of the core and accessory genomes of 48 *Sinorhizobium* strains comprising five genospecies. *Genome Biology*, 14, p.R17.
- Sugimoto-Shirasu, K. et al., 2005. RHL1 is an essential component of the plant DNA topoisomerase VI complex and is required for ploidy-dependent cell growth. *Proc Natl Acad Sci U S A*, 102(51), pp.18736–18741.
- Suzaki, T. et al., 2014. Endoreduplication-mediated initiation of symbiotic organ development in *Lotus japonicus*. *Development*, 141(12), pp.2441–2445.
- Suzuki, M.M. & Bird, A., 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nature reviews. Genetics*, 9(6), pp.465–76.
- Szczyglowski, K. et al., 1998. Nodule organogenesis and symbiotic mutants of the model legume *Lotus japonicus*. *Mol Plant-Microbe Interact*, 11(7), pp.684–697.
- Tajima, S., Nomura, M. & Kouchi, H., 2004. Ureide Biosynthesis In Legume Nodules. *Frontiers in Bioscience*, 9, pp.1374–1381.

- Tarayre, S. et al., 2004. Two classes of the Cdh1-type activators of the anaphase-promoting complex in plants: Novel functional domains and distinct regulation. *Plant Cell*, 16(2), pp.422–434.
- Tesfaye, M. et al., 2013. Spatio-Temporal Expression Patterns of Arabidopsis thaliana and *Medicago truncatula* Defensin-Like Genes. *Plos One*, 8(3): e58992.
- The Plant List, 2013. Leguminosae. Available at: <http://www.theplantlist.org/1.1/browse/A/Leguminosae/>.
- Tiricz, H. et al., 2013. Antimicrobial Nodule-Specific Cysteine-Rich Peptides Induce Membrane Depolarization-Associated Changes in the Transcriptome of *Sinorhizobium meliloti*. *Appl Environ Microbiol*, 79(21), pp.6737–6746.
- Trinick, M.J., 1979. Structure of nitrogen-fixing nodules formed by *Rhizobium* on roots of *Parasponia andersonii* Planch. *Can J Microbiol*, 25(5), pp.565–578.
- Tsai, M.-C. et al., 2010. Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. *Science*, 329(5992), pp.689–693.
- Turck, F. et al., 2007. Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. *PLoS Genetics*, 3(6), pp.0855–0866.
- Uppalapati, S.R. et al., 2009. Global gene expression profiling during *Medicago truncatula*-*Phymatotrichopsis omnivora* interaction reveals a role for jasmonic acid, ethylene, and the flavonoid pathway in disease development. *Mol Plant Microbe Interact*, 22(1), pp.7–17.
- Vanstraelen, M. et al., 2009. APC/C-CCS52A complexes control meristem maintenance in the Arabidopsis root. *Proc Natl Acad Sci U S A*, 106(28), pp.11806–11811.
- Velázquez, E. et al., 2010. Bacteria Involved in Nitrogen-Fixing Legume Symbiosis: Current Taxonomic Perspective. *Microbes for legume improvements*. Springer, pp. 1–25.
- Van de Velde, W. et al., 2006. Aging in legume symbiosis. A molecular view on nodule senescence in *Medicago truncatula*. *Plant physiology*, 141(2), pp.711–20.
- Van de Velde, W. et al., 2010. Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science*, 327(5969), pp.1122–1126.
- Vergara, Z. & Gutierrez, C., 2017. Emerging roles of chromatin in the maintenance of genome organization and function in plants. *Genome Biology*, 18(1), p.96.
- Vernie, T. et al., 2008. EFD Is an ERF transcription factor involved in the control of nodule number and differentiation in *Medicago truncatula*. *Plant Cell*, 20(10), pp.2696–2713.
- Verschure, P.J. et al., 2003. Condensed chromatin domains in the mammalian nucleus are accessible to large macromolecules. *EMBO Rep*, 4(9), pp.861–866.
- Vinardell, J.M. et al., 2003. Endoreduplication mediated by the anaphase-promoting complex activator CCS52A is required for symbiotic cell differentiation in *Medicago truncatula* nodules. *Plant Cell*, 15(9), pp.2093–2105.

- Volpe, T.A. et al., 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science*, 297(5588), pp.1833–1837.
- Wagner, D. & Meyerowitz, E.M., 2002. SPLAYED, a novel SWI/SNF ATPase homolog, controls reproductive development in Arabidopsis. *Current Biology*, 12(2), pp.85–94.
- Walley, J.W. et al., 2008. The chromatin remodeler SPLAYED regulates specific stress signaling pathways. *PLoS Pathogens*, 4(12): e1000237.
- Wang, Q. et al., 2017. Host-secreted antimicrobial peptide enforces symbiotic selectivity in *Medicago truncatula*. *Proc Nat Acad Sci U S A*, 114(26), pp.6854–6859.
- Wang, S.M. et al., 2014. Whole-genome sequencing of *Mesorhizobium huakuii* 7653R provides molecular insights into host specificity and symbiosis island dynamics. *BMC Genomics*, 15, pp.440.
- Wassenegger, M. et al., 1994. RNA-directed de novo methylation of genomic sequences in plants. *Cell*, 76(3), pp.567–576.
- Watt, F. & Molloy, P.L., 1988. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes & development*, 2(9), pp.1136–1143.
- Wei, J. et al., 2006. Role of Bmi1 in H2A ubiquitylation and Hox gene silencing. *Journal of Biological Chemistry*, 281(32), pp.22537–22544.
- Widman, N. et al., 2014. Epigenetic differences between shoots and roots in Arabidopsis reveals tissue-specific regulation. *Epigenetics : official journal of the DNA Methylation Society*, 9(2), pp.236–242.
- Wildermuth, M.C., 2010. Modulation of host nuclear ploidy: a common plant biotroph mechanism. *Current Opinion in Plant Biology*, 13(4), pp.449–458.
- Wojciechowski, M.F., Lavin, M. & Sanderson, M.J., 2004. A phylogeny of legumes (Leguminosae) based on analyses of the plastid matK gene resolves many well-supported subclades within the family. *American Journal of Botany*, 91(11), pp.1846–1862.
- Wolffe, A.P. & Hayes, J.J., 1999. Chromatin disruption and modification. *Nucleic Acids Research*, 27(3), pp.711–720.
- Wu, T.P. et al., 2016. DNA methylation on N6-adenine in mammalian embryonic stem cells. *Nature*, 532(7599), pp.1–18.
- Xiao, T.T. et al., 2014. Fate map of *Medicago truncatula* root nodules. *Development*, 141(18), pp.3517–3528.
- El Yahyaoui, F., 2004. Expression Profiling in *Medicago truncatula* Identifies More Than 750 Genes Differentially Expressed during Nodulation, Including Many Potential Regulators of the Symbiotic Program. *Plant Physiology*, 136(2), pp.3159–3176.

- Yang, R. et al., 2017. The developmental regulator PKL is required to maintain correct DNA methylation patterns at RNA-directed DNA methylation loci. *Genome Biology*, 18(1), p.103.
- Yang, S. et al., 2017. Microsymbiont discrimination mediated by a host-secreted peptide in *Medicago truncatula*. *Proc Nat Acad Sci U S A*, 114(26), pp.6848-6853.
- Young, M.D. et al., 2011. ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. *Nucleic Acids Research*, 39(17), pp.7415–7427.
- Young, N.D. et al., 2011. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature*, 480(7378), pp.520–524.
- Yu, A. et al., 2013. Dynamics and biological relevance of DNA demethylation in *Arabidopsis* antibacterial defense. *Proc Nat Acad Sci U S A*, 110(6), pp.2389–94.
- Zemach, A. et al., 2010. Local DNA hypomethylation activates genes in rice endosperm. *Proc Nat Acad Sci U S A*, 107(43), pp.18729–18734.
- Zemach, A. et al., 2013. The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell*, 153(1), pp.193–205.
- Zhang, D. et al., 2014. The Chromatin-Remodeling Factor PICKLE Integrates Brassinosteroid and Gibberellin Signaling during Skotomorphogenic Growth in *Arabidopsis*. *The Plant cell*, 26(6), pp.2472–2485.
- Zhang, J.J. et al., 2012. *Mesorhizobium muleiense* sp nov., nodulating with *Cicer arietinum* L. *Int J Syst Evol Microbiol*, 62, pp.2737–2742.
- Zhang, J.Y. et al., 2014. Global reprogramming of transcription and metabolism in *Medicago truncatula* during progressive drought and after rewatering. *Plant, Cell and Environment*, 37(11), pp.2553–2576.
- Zhang, M. et al., 2010. DNA cytosine methylation in plant development. *Journal of Genetics and Genomics*, 37(1), pp.1–12.
- Zhang, X.Y. et al., 2006. Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell*, 126(6), pp.1189–1201.
- Zhang, X.Y. et al., 2007. Whole-genome analysis of histone H3 lysine 27 trimethylation in *Arabidopsis*. *Plos Biology*, 5(5), pp.1026–1035.
- Zhao, M. et al., 2015. *Arabidopsis* BREVIPEDICELLUS Interacts with the SWI2/SNF2 Chromatin Remodeling ATPase BRAHMA to Regulate KNAT2 and KNAT6 Expression in Control of Inflorescence Architecture. *PLoS Genetics*, 11(3): e1005125.
- Zheng, B. & Chen, X., 2011. Dynamics of histone H3 lysine 27 trimethylation in plant development. *Current Opinion in Plant Biology*, 14(2), pp.123–129.
- Zheng, Z.L., 2009. Carbon and nitrogen nutrient balance signaling in plants. *Plant Signal Behav*, 4(7), pp.584–591.

- Zhou, C. Zhang, L. Duan, J. Miki, B. Wu, K., 2005. HISTONE DEACETYLASE19 Is Involved in Jasmonic Acid and Ethylene Signaling of Pathogen Response in *Arabidopsis*. *The Plant Cell Online*, 17(4), pp.1196–1204.
- Zhou, C. et al., 2011. From model to crop: functional analysis of a STAY-GREEN gene in the model legume *Medicago truncatula* and effective use of the gene for alfalfa improvement. *Plant Physiol*, 157(3), pp.1483–1496.
- Zhou, J. et al., 2010. Genome-wide profiling of histone H3 lysine 9 acetylation and dimethylation in *Arabidopsis* reveals correlation between multiple histone marks and gene expression. *Plant Mol Biol*, 72(6), pp.585–595.
- Zhou, P. et al., 2013. Detecting small plant peptides using SPADA (Small Peptide Alignment Discovery Application). *BMC bioinformatics*, 14(1), p.335.
- Zhou, Y. et al., 2017. H2A monoubiquitination in *Arabidopsis thaliana* is generally independent of LHP1 and PRC2 activity. *Genome biology*, 18(1), p.69.
- Zhu, H., Wang, G. & Qian, J., 2016. Transcription factors as readers and effectors of DNA methylation. *Nature reviews. Genetics*, 17(9), pp.551–65.
- Zhu, J.K., 2009. Active DNA demethylation mediated by DNA glycosylases. *Annual review of genetics*, 43, pp.143–66.
- Zhu, J.K. et al., 2007. The DNA Glycosylase/Lyase ROS1 Functions in Pruning DNA Methylation Patterns in *Arabidopsis*. *Current Biology*, 17(1), pp.54–59.
- Zilberman, D., 2017. An evolutionary case for functional gene body methylation in plants and animals. *Genome Biology*, 18(1), p.87.
- Zilberman, D. et al., 2007. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genetics*, 39(1), pp.61–69.

## **APPENDIX**

### **Supplementary information 1.**

[https://static-content.springer.com/esm/art%3A10.1186%2F1471-2164-15-712/MediaObjects/12864\\_2014\\_6439\\_MOESM1\\_ESM.xlsx](https://static-content.springer.com/esm/art%3A10.1186%2F1471-2164-15-712/MediaObjects/12864_2014_6439_MOESM1_ESM.xlsx)

### **Supplementary information 2.**

[https://static-content.springer.com/esm/art%3A10.1186%2F1471-2164-15-712/MediaObjects/12864\\_2014\\_6439\\_MOESM3\\_ESM.xlsx](https://static-content.springer.com/esm/art%3A10.1186%2F1471-2164-15-712/MediaObjects/12864_2014_6439_MOESM3_ESM.xlsx)

### **Supplementary information 3.**

<http://www.pnas.org/content/suppl/2017/04/12/1704211114.DCSupplemental/pnas.170421114.sd01.xlsx>



**Table 1. Primers used for the amplification and cloning of the promoters.**

attB4FWD_NCR001	GGGGACAACCTTTGTATAGAAAAGTTGGTTGTCCTTATTAGAGCGCC
attB1REV_NCR001	GGGGACTGCTTTTTTGTACAACTTGTATGTTTCATCCTTTGAACG
attB4FWD_NCR084	GGGGACAACCTTTGTATAGAAAAGTTGGCGAGAAAGGAAGGGAAGAA
attB1REV_NCR084	GGGGACTGCTTTTTTGTACAACTTGTATTTTTCTCCCTTTACATG
attB4FWD_NCR121	GGGGACAACCTTTGTATAGAAAAGTTGTCCTTCTATGCATGTTCAA
attB1REV_NCR121	GGGGACTGCTTTTTTGTACAACTTGGTTTTTCCCTCTTTATAGGT

**Table 2. Primers used for RT-qPCR experiment.**

<b>Mt4.0v1-JCVI</b>	<b>Primer name</b>	<b>Primer sequence</b>
Medtr6g071090	Mtc27_F	GGAGGTTGAGGGAAAGTGG
	Mtc27_R	CACCAACAAAGAATTGAAGG
Medtr3g013635	Mt40S_F	CCGCAAGGACTGTTCAAGAT
	Mt40S_R	TATCGGTCCATTCTGGAAGC
Medtr4g088285	NCR117_F	ACGTCAGACGTCAGACAGAGC
	NCR117_R	TCGCGACAATGACAATTACC
Medtr3g415650	ENOD12_F	AAGGAATCACCGACGCATAG
	ENOD12_R	ATGCTCTTGACCTGGCCATA
Medtr3g065710	NCR084_F	AAAGTTTTGCCACGGGTATG
	NCR084_R	TTTCCAATGGCCAATAGAT
Medtr6g463200	NCR001_F	GAGGACTGAAACTAGAATGCTTACC
	NCR001_R	CACCCACAAAAACCATCAAA
Medtr6g021615	NCR122_F	GCTTCTGTCTTAGCAGTGGCA
	NCR122_R	TTGAGGAGCCCAACTCACAT

**Table 3. Primers used for MeDIP-qPCR and ChIP-qPCR experiment.**

<b>Primer name</b>	<b>Primer sequence</b>
NCR117-1F	TGGCTATAAACATTTTACATACCTTGA
NCR117-1R	GGTATCACTGCACCAAACGA
NCR117-2F	TCGTTTGGTGCAGTGATACC
NCR117-2R	TCCCGACCAAATTTTAGCTG
NCR117-3F	CAGCTAAAATTTGGTCGGGA
NCR117-3R	TGCAACCATATATCCCCATACA
NCR117-4F	TGTATGGGGATATATGGTTGCA
NCR117-4R	CTCACGGCGATACATCACAT
NCR117-5F	ATGTGATGTATCGCCGTGAG
NCR117-5R	CCATATTTTGCCTCCCTTGA
NCR117-6F	TCAAGGGAGGCAAAATATGG
NCR117-6R	AGGCGGCTCTGTCTCTGTAA
NCR117-7F	TTACAGAGACAGAGCCGCCT
NCR117-7R	AAATCCACGGAAGGAACAAA
NCR117-8F	TTTGTTCCCTCCGTGGATT
NCR117-8R	AAAATCAACATGATCTTTGTTCAAG
ENOD12-1F	GTATTGGCGCTTTTTGAGGA
ENOD12-1R	CCCTTCAAGCTTCAGGAAAG
ENOD12-2F	CGGTTTCCTCAAATTCATGG
ENOD12-2R	TGGTTGCAAGTACTCGGATG
ENOD12-3F	ATTGCAGGCCTTCCATAAAA
ENOD12-3R	AATTC AATTTTGGCCTTGAAA
ENOD12-4F	ATTCCACGTACACCAATACGTTC
ENOD12-4R	AGTAGTTAGGGGCCTCAATGG
ENOD12-5F	CCTGATTATTCCCTAAATCCCTA
ENOD12-5R	GTGGTGGCCTATAAGCAGGA
ENOD12-6F	AAGGAATCACCGACGCATAG
ENOD12-6R	ATGCTCTTGACCTGGCCATA
ENOD12-8F	AAACCAAAATGATTTAGCAATGG
ENOD12-8R	GGTTTGATTTCAAATTAATTTATTGC
NCR084-1F	CGCCATCTTTATCAATCAACAA
NCR084-1R	TTTTCAGCGGAGACAATTCC
NCR084-2F	GCCTACTCCCTACGTCCATT
NCR084-2R	GGAAAACATGGGTGAGATGG
NCR084-3F	TTTCCCATATGCCCTCTTC
NCR084-3R	GGCATTTGCCAGTTGAGCTA
NCR084-4F	TTTTCAGTGGATTACCACTTCATC
NCR084-4R	TGGCAAACTTTCTGCAACA
NCR084-5F	TGAATTTTGT TTTGTGAAGGATTC
NCR084-5R	TTTCCCAATGGCCAATAGAT
NCR084-6F	CGGTTTCCATTCATCTACTGTG
NCR084-6R	AGATCGAGGAGGGAAGATCC

NCR084-7F	CACTCCTCGGATTTGAAAAA
NCR084-7R	TAAACTTCGCCACGAGTTCC
NCR084-8F	CAACTTGCACCTTGACGAACC
NCR084-8R	AAGCATACATGCCCTTCACC
NCR001-15F	TTGGGGTTTTTCAGAAGAGC
NCR001-15R	GGGTTTTGTTGCTGAGTTCC
NCR001-17F	TGAAGAACATGGAGGTGCAA
NCR001-17R	TTCCTCTCCTCGAGCTATGC
NCR001-19F	GATTTTGGATCGTTGAACGC
NCR001-19R	GAGAAAGGGAACCAAACACAA
NCR001-21F	GCTGGTTTGGTAGGTAAGTCG
NCR001-21R	CCAAAAAATAATTCAACTTTATGATG
NCR001-23F	GTTTAATGTGAAATGCATGAAAGA
NCR001-23R	TGCTCTTAGAAGTAACAAAAAGCAA
NCR001-25F	TTCAAAGGATGAAACATATGGC
NCR001-25R	TGGTGTGAAAAATGCTTACG
NCR001-27F	TCCTTGCACTTCTGACGCTA
NCR001-27R	TGTGATATCCCCTGGTTTCG
NCR001-29F	TGTGTATGCTGCACTCTTTTT
NCR001-29R	CTATGCCAAAACCCAAGGAA
NCR001-31F	TGATGCCTATGTACGTCGTTG
NCR001-31R	TGCAATATCTTTAATGACCAAATGA
NCR122-1F	GGACAACCATGACACATCAAA
NCR122-1R	TTGCCTACTCATATTTGTCTCG
NCR122-2F	CGAGACAAATATGAGTAGGCAA
NCR122-2R	CAAAAACAACATGCCATCCTT
NCR122-4F	TTGGAATGCTTTCGTAAAGA
NCR122-4R	TGTGTAGCAAATAGAAGATTGTTTGA
NCR122-5F	TCAAACAATCTTCTATTTGCTACACA
NCR122-5R	GGAATCTGAGGTCCGGTACA
NCR122-6F	TGTACCGGACCTCAGATTCC
NCR122-6R	ATGCTTCTTTGCCACCTGTC
NCR122-7F	GACAGGTGGCAAAGAAGCAT
NCR122-7R	TTGAGGAGCCCAACTCACAT
NCR122-8F	ATGTGAGTTGGGCTCCTCAA
NCR122-8R	AACCCAGCAAATGCTTGAAC

**Table 4. ATAC-seq oligos used for PCR.**

<b>Ad1</b>	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG
<b>Ad2.1</b>	TAAGGCGACAAGCAGAAGACGGCATAACGAGATTGCCTTAGTCTCGTGGGCTCGGAGATGT
<b>Ad2.2</b>	CGTACTAGCAAGCAGAAGACGGCATAACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT
<b>Ad2.3</b>	AGGCAGAACAAGCAGAAGACGGCATAACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGT
<b>Ad2.4</b>	TCCTGAGCCAAGCAGAAGACGGCATAACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT
<b>Ad2.5</b>	GGACTCCTCAAGCAGAAGACGGCATAACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGT
<b>Ad2.6</b>	TAGGCATGCAAGCAGAAGACGGCATAACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGT
<b>Ad2.7</b>	CTCTCTACCAAGCAGAAGACGGCATAACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGT
<b>Ad2.8</b>	CAGAGAGGCAAGCAGAAGACGGCATAACGAGATCCTCTCTGGTCTCGTGGGCTCGGAGATGT
<b>Ad2.9</b>	GCTACGCTCAAGCAGAAGACGGCATAACGAGATAGCGTAGCGTCTCGTGGGCTCGGAGATGT
<b>Ad2.10</b>	CGAGGCTGCAAGCAGAAGACGGCATAACGAGATCAGCCTCGGTCTCGTGGGCTCGGAGATGT
<b>Ad2.11</b>	AAGAGGCACAAGCAGAAGACGGCATAACGAGATTGCCTCTTGTCTCGTGGGCTCGGAGATGT
<b>Ad2.12</b>	GTAGAGGACAAGCAGAAGACGGCATAACGAGATTCCTCTACGTCTCGTGGGCTCGGAGATGT
<b>Ad2.13</b>	GTCGTGATCAAGCAGAAGACGGCATAACGAGATACACGACGTCTCGTGGGCTCGGAGATGT
<b>Ad2.14</b>	ACCACTGTCAAGCAGAAGACGGCATAACGAGATACAGTGGTGTCTCGTGGGCTCGGAGATGT
<b>Ad2.15</b>	TGGATCTGCAAGCAGAAGACGGCATAACGAGATCAGATCCAGTCTCGTGGGCTCGGAGATGT
<b>Ad2.16</b>	CCGTTTGTCAAGCAGAAGACGGCATAACGAGATACAAACGGGTCTCGTGGGCTCGGAGATGT
<b>Ad2.17</b>	TGCTGGGTCAAGCAGAAGACGGCATAACGAGATACCCAGCAGTCTCGTGGGCTCGGAGATGT
<b>Ad2.18</b>	GAGGGGTTCAAGCAGAAGACGGCATAACGAGATAACCCCTCGTCTCGTGGGCTCGGAGATGT
<b>Ad2.19</b>	AGGTTGGGCAAGCAGAAGACGGCATAACGAGATCCCAACCTGTCTCGTGGGCTCGGAGATGT
<b>Ad2.20</b>	GTGTGGTGCAAGCAGAAGACGGCATAACGAGATCACCACACGTCTCGTGGGCTCGGAGATGT
<b>Ad2.21</b>	TGGGTTTCCAAGCAGAAGACGGCATAACGAGATGAAACCCAGTCTCGTGGGCTCGGAGATGT
<b>Ad2.22</b>	TGGTCACACAAGCAGAAGACGGCATAACGAGATTGTGACCAGTCTCGTGGGCTCGGAGATGT
<b>Ad2.23</b>	TTGACCCTCAAGCAGAAGACGGCATAACGAGATAGGGTCAAGTCTCGTGGGCTCGGAGATGT
<b>Ad2.24</b>	CCACTCCTCAAGCAGAAGACGGCATAACGAGATAGGAGTGGGTCTCGTGGGCTCGGAGATGT

**Table 5. GO term enrichment of *M. truncatula* genes with differential DNA methylation.**

Hypomethylated genes (n=2488)								
GO ID	Term type	Term	Query item	Query total	Back-ground item	Back-ground total	P-value	FDR
GO:0009878	biological process	nodule morphogenesis	90	757	378	15474	1,40E-31	5,50E-29
GO:0046872	molecular function	metal ion binding	145	757	1953	15474	4,50E-07	8,60E-05
GO:0046039	biological process	GTP metabolic process	5	757	16	15474	0,0024	0,21
GO:0008238	molecular function	exopeptidase activity	10	757	74	15474	0,0056	0,8
GO:0051258	biological process	protein polymerization	5	757	22	15474	0,0075	0,59
GO:0009205	biological process	purine ribonucleoside triphosphate metabolic process	10	757	80	15474	0,0091	0,59
GO:0043623	biological process	cellular protein complex assembly	6	757	37	15474	0,014	0,87
GO:0003924	molecular function	GTPase activity	13	757	134	15474	0,02	1
GO:0005529	molecular function	sugar binding	14	757	150	15474	0,021	1
GO:0016840	molecular function	carbon-nitrogen lyase activity	5	757	30	15474	0,022	1
GO:0033177	cellular component	proton-transporting two-sector ATPase complex, proton-transporting domain	5	757	32	15474	0,028	1
GO:0030246	molecular function	carbohydrate binding	14	757	163	15474	0,038	1
GO:0004185	molecular function	serine-type carboxypeptidase activity	7	757	61	15474	0,039	1
GO:0019001	molecular function	guanyl nucleotide binding	18	757	230	15474	0,043	1

Hypermethylated genes (n=1807)								
GO ID	Term type	Term	Query item	Query total	Back-ground item	Back-ground total	P-value	FDR
GO:0016567	biological process	protein ubiquitination	6	534	41	15474	0,0045	1
GO:0070647	biological process	protein modification by small protein conjugation or removal	6	534	42	15474	0,005	1
GO:0004842	molecular function	ubiquitin-protein ligase activity	6	534	43	15474	0,0055	1
GO:0019787	molecular function	small conjugating protein ligase activity	6	534	43	15474	0,0055	1
GO:0016798	molecular function	hydrolase activity, acting on glycosyl bonds	25	534	412	15474	0,0064	1
GO:0019748	biological process	secondary metabolic process	5	534	32	15474	0,0073	1
GO:0000151	cellular component	ubiquitin ligase complex	7	534	67	15474	0,012	1
GO:0043414	biological process	macromolecule methylation	6	534	52	15474	0,013	1
GO:0004553	molecular function	hydrolase activity, hydrolyzing O-glycosyl compounds	23	534	395	15474	0,013	1
GO:0006730	biological process	one-carbon metabolic process	6	534	56	15474	0,017	1
GO:0016881	molecular function	acid-amino acid ligase activity	9	534	118	15474	0,026	1
GO:0043234	cellular component	protein complex	28	534	546	15474	0,029	1
GO:0005975	biological process	carbohydrate metabolic process	32	534	651	15474	0,034	1
GO:0016879	molecular function	ligase activity, forming carbon-nitrogen bonds	10	534	152	15474	0,045	1
GO:0016020	cellular component	membrane	85	534	2049	15474	0,045	1