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

Ph.D THESIS

Marianna NAGYMIHÁLY

Thesis supervisors:

Dr. Peter MERGAERT, I2BC, Gif-sur-Yvette, FRANCE

Dr. Attila KERESZT, BRC, Szeged, HUNGARY

Submitted to the

DOCTORAL SCHOOL IN BIOLOGY,

UNIVERSITY OF SZEGED, HUNGARY

and to the

ECOLE DOCTORAL SCIENCES DU VÉGÉTAL,

UNIVERSITÉ PARIS-SACLAY, FRANCE

2017

I. INTRODUCTION

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 plant cells, rhizobia are capable to fix atmospheric nitrogen and convert it to ammonia, which is a usable nitrogen source for the plant. The symbiotic cells originate from dividing progenitor cells in the nodule meristem. During symbiotic cell differentiation, post-meristematic cells exit the cell division cycle and enter an endoreduplication cycle, a modified cell cycle with repeated replication of the genome without mitosis and cytokinesis, resulting in polyploid cells with increased DNA content (from 2C up to 64C ploidy levels) and cell enlargement up to 80-fold compared to the meristematic cell (Vinardell et al. 2003). The differentiating symbiotic cells are gradually infected and filled with rhizobia forming symbiosomes, nitrogen-fixing organelle-like structures. *Medicago truncatula* nodules are indeterminate type, containing a persistent apical meristem (ZI), an infection zone (distal ZIId and proximal ZIIp), an interzone (IZ) and a nitrogen-fixation zone (ZIII) and in older nodules a senescence zone (ZIV). Mature zone III cells are entirely packed with symbiosomes and their physiology is adapted for symbiosis, feeding the microsymbionts and assimilating the fixed nitrogen.

The generation of differentiated symbiotic cells in the legume *Medicago 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 endoreduplication and the expression of the *NCR* genes. How endoreduplication 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 endoreduplication cycles might play a critical role in

transcriptional regulation of *NCR*s 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 (Satge 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, histone H3 lysine 27 trimethylation (H3K27me3) (Feng & Jacobsen 2011) also results in low expression level and high tissue-specificity (Zhang et al. 2007) whereas histone H3 lysine 9 acetylation (H3K9ac) leads to gene activation (Kurdistani et al. 2004; Schubeler et al. 2004; Roh et al. 2005; Zhou et al. 2010).

Thus, the symbiotic nodule cells of *Medicago truncatula* constitute an excellent model system to study transcriptional reprogramming at developmental-stage specific level because the symbiotic cell differentiation involves endoreduplication 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 endoreduplication in the symbiotic cells, we hypothesised that epigenetic mechanisms involving specific histone modifications and/or DNA methylations might be implicated in the determination of the gene expression program in the symbiotic cells. Repressive chromatin structures can 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 transcription factors. Moreover, we propose that endoreduplication of the symbiotic nodule cells serves as a mechanism in controlling gene expression. Endoreduplication 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 of my thesis work.

II. OBJECTIVES

A. Characterize in great detail the expression behaviour of a Nodule-Specific Cysteine-Rich (NCR) gene family.

1. The transcriptome database provided by 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 (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, ZIId, ZIIp, IZ and ZIII).
3. In order to confirm the expression data from MtGEA, stable transgenic *M. truncatula* R108 lines were generated carrying promoter-β-glucuronidase reporter (GUS) fusion constructs for 3 different *NCR* genes, representing different temporal classes of *NCR*s, as well as a specific antibody for one *NCR* peptide.
4. Moreover, 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.

B. 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. These experiments demonstrate whether gene

expression changes are linked with the increasing ploidy during the differentiation of the symbiotic nodule cells.

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 4C, 8C, 16C and 32C nuclei sorted with flow-cytometer.
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, and patterns of DNA methylation were matched to published *in situ* RNA-seq data of the nodule zones obtained by laser-capture microdissection (Roux et al. 2014). Moreover, differential methylation of nodule-specific and nodule expressed genes in 4C, 8C, 16C and 32C nuclei were analysed.
7. To know 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. Patterns of chromatin changes were matched to published *in situ* RNA-seq data of the nodule zones obtained by laser-capture microdissection (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 different ploidy nuclei.

III. METHODS

A. Global analysis of *NCR* gene expression.

1. Data mining using the MtGEA and Shannon entropy calculations of *Medicago* genes.
2. Transcriptome analysis of hand-dissected nodule zones.
3. Transgenic plants and β-glucuronidase reporter (GUS) analysis of differentially expressed *NCR*s (*NCR121*, *NCR084*, *NCR001*).
4. Pathogen assays for testing biotic stress on *NCR* genes expression and immunolocalization of NCR122 in the nodule.

***B*. Ploidy-specific analysis of epigenetic modifications of nodule-specific genes.**

5. Nuclear RNA isolated from 4C, 8C, 16C, 32C ploidy level nodule nuclei was used for reverse transcription quantitative PCR (RT-qPCR) analysis of selected genes.
6. Genome-wide DNA methylation analysis of uninfected nodule cell (4C) and symbiotic nodule cell (32C) genome using Reduced Representation Bisulphite Sequencing (RRBS).
7. Genomic DNA extraction from different ploidy level nodule nuclei and Methylated DNA Immunoprecipitation (MeDIP) analysis of selected genes.
8. Genome-wide analysis of ploidy dependent changes of chromatin accessibility in different ploidy level nodule nuclei using Assay for Transposase-Accessible Chromatin using sequencing (ATAC-Seq).
9. Chromatin isolation from different ploidy level nodule nuclei and Chromatin Immunoprecipitation (ChIP) analysis of selected genes.

IV. RESULTS

1. From the transcriptome data mining and experimental confirmation of *NCR* expression, we can conclude that apart from very few exceptions, the hundreds of *NCR* genes encoding defensin-like peptides are only activated during nodule formation. 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. *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 the symbiotic nodule cells and bacteroids during nodule senescence since their expression was shut down when senescence is initiated; therefore, *NCR* peptides have no roles in the death and disintegration of rhizobia in the senescing plant cells. NCRs resemble antimicrobial peptides such as defensins of the innate immunity and many NCR peptides. *In vitro*, cationic peptides particularly with high positive charge have a strong 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). Despite this, the *NCR* genes are not induced during pathogen attack neither by bacteria, fungi, oomycetes nor nematodes. They are not expressed

either 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. The *NCR* genes are not activated by Nod factors or during the very early stages of the nodule organogenesis before the infected cells are formed (this work, Nallu et al. 2013; Maunoury et al. 2010). Furthermore, 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*. 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.

2. 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, ZII) while genes activated late during development are expressed in the proximal nodule tissues (IZ-ZIII). In addition, certain clusters of genes, once activated, maintain their activity during ageing of nodule cells while other clusters are characterized by lacking expression in the older nodule cells. Our spatial analysis of *NCR* expression is in good agreement with a recently published study that used LCM of nodule zones coupled to RNA-Sequencing (Roux et al. 2014).

3. In agreement with its temporal regulation during nodulation, *NCR121* expression was induced in young nodule primordia as early as 5 dpi and remained expressed in the entire infection zone II and the fixation zone III of mature nodules. *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. *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. All 3 genes are only expressed in the symbiotic nodule cells. GUS expression in the three transgenic lines was not detected in root tips or other root parts. 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, confirming that *NCR* genes are not involved in the senescence process.

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

4. The results of the pathoassays are in line with a recent study showing that *NCR* expression was not detected during the interaction between *M. truncatula* and the biotrophic soil pathogen *Phytophtora medicaginis,* the hemibiotrophic leaf pathogen *Colletotrichum trifolii,* (Tesfaye et al. 2013) and the necrotrophic grey mold-causing fungus *Botrytis cinerea*. Our data are in agreement with the MtGEA dataset and broaden the conclusion that *NCR*s are not involved in pathogen response, whatever the trophic- (bio/hemibio/necrotrophic). Moreover, as herbivory and more generally the effect of mechanical wounding, which may induce plant defenses were also tested but again no *NCR* expression could be detected in the wounded leaflet.

5. 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. Using sorted nodule nuclei according to their DNA content, the expression pattern of selected nodule-expressed and nodule-specific genes was correlated with their spatial expression pattern in nodules (Roux et al. 2014). We found that activation of the early, first wave genes occurs in the 4C post-meristematic cells whereas early, second wave genes are activated in the distal and proximal parts of ZII with 8C and 16C while late genes of the third transcriptional wave 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.

6. In response to *Rhizobium* signal molecules, the Nod factors, proliferation of root cortical cells leads to the establishment of the nodule meristem. DNA methylation of the meristematic cells represents the initial and likely the most methylated status of DNA. 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 sequence context (75%) followed by CHG (20%) and CHH (5%) in both cell types, which was consistent with previous studies and

indicates the importance of CG methylation. 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. This finding indicates that the DNA methylation state of the 4C cells has been copied during the repeated endoreduplication cycles. Unlike most protein-coding genes, *NCR*s behaved differently as out of 375 genes 164 became hypomethylated which 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 (Satge 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 transposable elements (TEs) (Satge et al. 2016). Plant genomes are rich in TEs, which are usually maintained transcriptionally silent by DNA methylation. This silencing of TEs may also affect neighbouring genes and *NCR* gene activation may thus require demethylation mediated by MtDME.

7. We found that 44% of the *NCR* genes were hypomethylated in the 32C symbiotic cells compared to 4C and this was also validated with MeDIP, suggesting that decrease in DNA methylation, particularly in their 1-kb upstream region, contributes to their expression. But about 53% of the *NCR* genes did not differ in DNA methylation at 4C and 32C. This means that these genes might be active in the 8C and 16C cells and display hypomethylation in these nuclei. We tested this hypothesis on selected genes along their 1 kb upstream and downstream regions from the translational start site (TSS), but the extent of DNA methylation was unaffected by the ploidy levels confirming the RRBS results. It could mean that DNA methylation does not play a pivotal role in their regulation and genes with unaltered DNA methylation might require chromatin modifications.

8. Chromatin structure has been implicated in developmental and tissue-specific regulation of a number of genes of 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 translational start site. Early genes exhibited highest accessibility in 4C and 8C cells, while late genes in 16C and 32C. 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.

9. In addition, 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 probably essential for the repressed state of *NCR*s and other symbiotic genes in these cells. In case of early genes the H3K27me3 marks were dramatically reduced in the 8C and 16C cells and increased again in the 32C cells correlating with their repression, while in case of late genes the reduced H3K27me3 marks together with the presence of the active H3K9ac marks and accessible chromatin structure contributed to gene activation. *NCR*s in the symbiotic 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. 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 differentiation (Lafos et al. 2011).

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. Moreover, the dynamic changes observed in the epigenome suggest that chromatin accessibility together with histone tail modifications regulate the transcriptionally active or inert state of the genes.

PUBLICATIONS LIST

Publications related to the topic of the thesis

Guefrachi I*, Nagymihály M*, Pislariu C*, Ratet P, Mars M, Udvardi M, Kondorosi E, Peter Mergaert P, Alunni B. (2014) Extreme specificity of *NCR* gene expression in *Medicago truncatula.* BMC Genomics 2014, 15:712.
*Equal contribution

Nagymihály M, Veluchamy A, Györgypál Z, Ariel F, Jégu T, Benhamed M, Szűcs A, Kereszt A, Mergaert P, Kondorosi É. (2017) Ploidy-dependent changes in the epigenome of symbiotic cells correlate with specific patterns of gene expression. PNAS, 114(17):4543-4548.

Cumulative IF: 13,409

Other publications

Nagymihály M, Ibijbijen J, Bammou M, Ratet P, Mergaert P, Kondorosi É, Kereszt A. (2017) Whole genome sequence of *Ensifer/Sinorhizobium meliloti* strain FSM-MA, a highly effective nitrogen-fixing microsymbiont of *Medicago truncatula*. *In preparation.*

Kazmierczak T*, Nagymihály M*, Lamouche F, Barrière Q, Guefrachi I, Alunni B, Ouadghiri M, Ibijbijen J, Kondorosi É, Mergaert P, Gruber V. (2017) Specific Host-Responsive Associations Between *Medicago truncatula* Accessions and *Sinorhizobium* Strains. Mol Plant Microbe Interact. 30(5):399-409.
*Equal contribution

Nagymihály M, Attila Szûcs A, Attila Kereszt A. (2015) Next-Generation Sequencing and its new possibilities in medicine. Acta Biol Szeged 59(Suppl.2):323-339.

Hunyadkürti J, Feltóti Z, Horváth B, Nagymihály M, Vörös A, McDowell A, Patrick S, Urbán E, Nagy I. (2011) Complete genome sequence of *Propionibacterium acnes* type IB strain 6609. J Bacteriol. 193(17):4561-2.

Cumulative IF: 8,645

Conference participations:

Oral presentation

Nagymihály M, Szűcs A, Ariel F, Jégu T, Benhamed M, Kereszt A, Mergaert P and Kondorosi É. (2016) Dynamic changes in chromatin structure regulate expression of nodule-specific *NCR* genes in *Medicago truncatula.* Straub days BRC, Szeged, p. 4.

Nagymihály M, Szűcs A, Ariel F, Jégu T, Benhamed M, Kereszt A, Mergaert P and Kondorosi É. (2016) Dynamic changes in chromatin structure regulate expression of nodule-specific *NCR* genes in *Medicago truncatula.* 12th European Nitrogen Fixation Conference, Budapest, p. 119.

Poster presentation

Nagymihály M, Guefrachi I, Pislariu C, Van de Velde W, Ratet P, Mars M, Udvardi M. K, Kondorosi É, Mergaert P, Alunni B. (2014) Extreme specificity of *NCR* gene expression in *Medicago truncatula.* 11emes Rencontres Plantes-Bactéries, Aussois, France, p. 106.

Nagymihály M, Jégu T, Szűcs A, Benhamed M, Kereszt A, Mergaert P and Kondorosi É. (2016) Endoreduplication- and chromatin mediated regulation of gene expression in symbiotic nodule cells of *Medicago truncatula.* 23rd North American Conference on Symbiotic Nitrogen Fixation, Ixtapa, Mexico, p. 63.

Jesus Caballero Award for best poster presentation.

Nagymihály M, Veluchamy A, Györgypál Z, Ariel F, Jégu T, Benhamed M, Szűcs A, Kereszt A, Mergaert P, Kondorosi É. (2017) Ploidy-dependent changes in the epigenome of symbiotic cells correlate with specific patterns of *NCR* gene expression. Frontiers in Beneficial Plant-Microbe Interactions, 3rd Adam Kondorosi Symposium, Gif-sur Yvette, France