

FINAL SCIENTIFIC REPORT

Profit maximalization in symbiosis? Gene for gene interactions in the Medicago-Sinorhizobium partnership

K120122/120300

Project leaders:

Attila Kereszt

Institute of Plant Biology, Biological Research Centre, Szeged

Ágota Domonkos

Institute of Genetics and Biotechnology

Hungarian University of Agriculture and Life Sciences (MATE), Gödöllő

1. Identification of incompatible interactions between the strains of *Sinorhizobium meliloti/medicae* and the ecotypes of the model legume *Medicago truncatula*

Most of the genetic studies to identify genetic determinants required for the development of indeterminate nodules has used *M. truncatula* ecotypes Jemalong and R108 as well as *S. meliloti* strain 1021 and *S. medicae* strain WSM419 although high number of symbiotic mutants is available only in strain 1021. Moreover, none of the two strains are fully effective with the two dominantly used ecotypes (Terpolilli et al., 2008; Kazmierczak et al., 2017) indicating the existence of genetic differences affecting the symbiotic compatibility and efficiency between strains and ecotypes.

To investigate how frequent the incompatibility between the strains and ecotypes of different geographical origin is, our research group and two other research teams (Crook et al., 2012; Liu et al., 2014) screened several ecotypes from *Medicago truncatula* using different *Sinorhizobium meliloti* and *Sinorhizobium medicae* strains as inoculants (Table 1). We identified more than 10 incompatible host-strain interactions in which the host plants showed yellowish leaves and smaller shoots, characteristic phenotypes associated with nitrogen starvation. In particular, we noticed that *S. meliloti* strain Rm41 of Hungarian origin and intensively used in our institute was incompatible with *Medicago truncatula* ecotypes F83005 and Jemalong, while the same strain could nodulate and fix nitrogen normally in association with other *Medicago truncatula* lines (such as A20, DZA315.16, DZA045, SA023859, ...). Although the final outcome of the two incompatible interactions was the same, i.e. nitrogen starved plants, symbiotic development was arrested at different stages in the two plants: In ecotype F83005, strain Rm41 induces root hair curling and cortical cell division but infection thread and – as a consequence – nodule formation is ceased thus only bumps or occasionally empty nodules are formed on the roots. In contrast, nodule formation, bacterial infection and bacteroid development take place in line Jemalong but bacteroids are eliminated from the nodule cells. From these differences we could conclude that the incompatibilities are caused and controlled by different genetic determinants/genes in strain Rm41.

Table 1. symbiosis specificity in different *Medicago truncatula* - *Sinorhizobium* pairs

<i>Sinorhizobium</i> strains	<i>Medicago truncatula</i>					
	Jemalong A17	A20	DZA315.16	DZA220H	F83005	L746
<i>S. meliloti</i> FSM-MA	Fix+	Fix+	Fix+	Fix+	Fix+	Fix+
<i>S. meliloti</i> 1021	Fix+ (-)	Fix+	Fix+	Fix+	Fix+	Fix-
<i>S. meliloti</i> Rm41	Fix-	Fix+	Fix+	Fix+	Fix-	Fix+
<i>S. meliloti</i> GR4	Fix-	Fix+	Fix+	Fix+	Fix+	Fix+
<i>S. meliloti</i> SM11	Fix-	Fix+ (s)	Fix-	Fix-	Fix-	Fix-
<i>S. medicae</i> WSM419	Fix+	Fix+	Fix+	Fix+	Fix+	Fix+
<i>S. medicae</i> ABS7	Fix+	Fix+	Fix+	Fix-	Fix+	Fix+

(s) means early senescence

A novel observation is that the phenotype of a plant mutant may depend on the rhizobium strain used for the phenotypic analysis. We have shown in our recent paper (Kovács et al., 2021) that the nodulation defect of the *nsp2-3* mutant plants is more severe when they are inoculated with strain Sm1021 as compared to strain WSM419. It will be worth to re-test our plant mutants with a collection of wild-type rhizobia and compare the phenotypic outcomes.

2. Establishment of genomic resources to identify the bacterial genes that control the fate of the interactions

Theoretically, there are two different explanations for an ineffective phenotype of a rhizobial strain: it can be caused by 1) either a function that is absent from the bacterium, for example, for the production of a signal molecule/enzyme; 2) or there is an interaction- detrimental biological activity encoded in the genome, for example, for the production of a surface polysaccharides/protein recognized as enemy or an enzyme/protein interfering with the process of nodule/bacteroid development or functioning. These possibilities mean that two different strategies are needed to achieve the restoration of the effective interaction. To identify an interaction-detrimental function, the disruption/impairments of the encoding gene is needed. i.e. random insertion or point mutations in the incompatible strain has to be created. To introduce a gene missing from or inactive in the ineffective strains, genomic libraries from compatible strain(s) have to be constructed.

2.1. Creation of mutant pools of bacteria

There are multiple ways to create bacterial mutants that may lead to different “genetic outcomes”. Transposon insertions result in gene disruption and loss of protein function. Often, the insertion has polar effect meaning that the transcription of downstream genes in the operon is also blocked, thus, the activity of multiple proteins is lost. There are cases, however, when complete loss of protein function is undesirable because it affects multiple biological pathways and/or the viability of the cells. Point mutations, however, can introduce subtle (amino acid) changes that might not be detrimental for the essential activity of the protein, rather, they might affect for example, its interaction(s) with other proteins or substrates. Because of this reasoning, we decided to perform both chemical mutagenesis and random transposon insertion mutagenesis.

2.1.1. Chemical mutagenesis of strain Rm41

To introduce point mutations into the Rm41 genome, we used N-methyl-N'-nitro-N-nitrosoguanidine (MNNG) acting by adding alkyl groups to the O⁶ and O⁴ atoms of guanine and thymine, respectively, as described by Glazebrook et al. (1996). We have not measured the efficiency of the mutagenesis but according to their published results, this mutagenesis protocol resulted in 90 to 95% lethality, and 1 to 2% of the survivors were auxotrophs. As our mutagenesis was performed with 40 populations of 1 mL bacteria at OD₆₀₀ ≈ 0.2 (~2x 10⁷ cells), there might have been ~10⁶ survivors per population, i.e. altogether ~4x10⁷ mutagenized cells. Although these cells might carry

multiple mutations, if we calculate by one mutation per cell and with an approximately 7 Mbp genome size, each nucleotide might be targeted six times on average.

2.1.2. Random transposon insertion mutagenesis of sinorhizobia

For the insertion mutagenesis of the *S. meliloti* and *S. medicae* strains, the *mariner*-based transposon vector, pSAM_R1 (Perry and Yost, 2014) was used. The vector backbone contains an ampicillin resistance gene, an *RK6 γ oriR* origin of replication allowing plasmid maintenance only in cells with a *λ pir* function, an origin of transfer of the RP4 plasmid (*oriT*) as well as the transposase of the *himar1C9* transposon preceded by the promoter and 5' UTR sequence of the *Rhizobium leguminosarum rpoD* gene. The transposon contains the MmeI-adapted mariner inverse repeats (IR_R, IR_L) and the *nptIII* gene of transposon Tn5 (providing kanamycin resistance) followed Rho-independent terminators (*rrnB_T1*, *rrnB_T2*). This vector, which cannot replicate in rhizobia, was conjugated into the Sinorhizobium strains by tri-parental mating and rhizobia with insertion events were selected on minimal medium (where the auxotroph *E. coli* could not grow) supplemented with kanamycin. In two rounds, 105 plus 96, altogether 201 independent mutant pools were generated in the Rm41 background, then 96 mutant pools were established in each of the GR4, Sm11, Sm1021 and ABS7 strains.

Our original testing strategy resulted in not satisfactory results (only two compatible mutants; see later), that is why, to estimate how many independent mutants were collected after the large scale mutagenesis, an Illumina technique-based transposon sequencing approach was applied to selected mutant pools of strain Rm41. To create Illumina sequencing libraries, genomic DNA was isolated from the cells of the pools and digested with the restriction enzyme MmeI, which cleaves the DNA 20/18 nucleotides after the recognition sequence resulting in a two nucleotide 3' overhang. A splinkerette (Devon et al., 1995) Illumina adaptor was ligated to the digested DNA then transposon flanking sequences were amplified using a transposon specific primer, which also carries Illumina-specific sequences for library production, and the splinkerette-specific Illumina primer. The 125-128 bp fragments were isolated from agarose gel, then Illumina adaptors and barcodes were added in a second PCR of few cycles., finally, approximately 250.000 reads were obtained from each library. After trimming the Illumina- and transposon-specific sequences, the 16 nucleotide long reads were mapped to the reference genome and the consensus sequences were generated and counted by the Geneious software. The summary of the sequencing results is shown in Table 2.

As it is shown in Table 2., each pools contains high number of independent transposon insertion mutants, much more than we expected (few hundred clones per pool). If we calculate with the average number of independent mutants, our 201 Rm41 populations may contain more than 1.5 million independently obtained insertion mutants. As this transposon integrates at TA sites into the DNA and there are only 3-400.000 target sites in the genome, each site in genes non-essential for survival could

have been targeted four-five times in average. It is also worth to note that the distribution of insertions among the replicons correlates well with the size of the replicons without bias.

Table 2. Analysis of the results of transposon sequencing from the selected pools

Library	Number of insertions				Total
	chromosome	pSymA	pSymB	Plasmid3	
1st round pool 40	1897	908	1051	127	3983
1st round pool 41	2626	1356	1469	167	5618
2nd round pool 40	5292	2625	3040	289	11246
2nd round pool 41	4484	2284	2519	271	9558
Total	14299	7173	8079	854	30405
Average	3574.75	1793.25	2019.75	213.5	7601.25
Percentage of insertions	47.03%	23.59%	26.57%	2.81%	100.00%
Percentage of genome	51.52%	21.70%	23.32%	3.46%	100.00%
Genome size (bp)	3678504	1549593	1665079	247065	7140241

Based on the results of the sequencing and the nodulation assays, we realized that there was an error in the experimental design: If our estimation based on the number of insertions detected in the four randomly selected pools and the specificity of the transposon insertions (between T and A nucleotides) is correct, then we should have isolated much more mutants of Rm41 that are compatible with ecotype F83005. Indeed, when we checked the transposon insertion sites from the four pools, we found 21 insertions (five of them appeared in two pools) in the gene we identified that is much higher number than the two strains we isolated after inoculating all the pools. It means that we used too many (much higher number than expected) mutants and too few plants in our nodulation assays, thus, in further experiments, we will use only one pool with several hundred plants.

2.2. Construction of genomic libraries for complementation

A missing function can be caused by the absence of or a mutation in a gene or operon that can be complemented by genes/operons from a compatible strain. An ORFeome library carrying all predicted ORFs of the reference strain *S. meliloti* strain 1021 in a Gateway entry vector (Schroeder et al., 2005) was made available for us that could be used for the construction of a library. However, we saw two drawbacks of this library: 1) it contained only individual genes that cannot complement strains where whole operons are missing; 2) strain 1021 is not fully compatible with the *M. truncatula* cultivar Jemalong and is incompatible with a number of other ecotypes. That is why, we decided to create a large-insert size genomic library from strain FSM-MA (see Table 3.) which was shown to form effective symbiosis with the two most commonly used ecotypes, Jemalong and R108 (Kazmierczak et al., 2017) as well as with the most ecotypes we use in our laboratory.

2.2.1. Construction of the strain 1021 ORFeome library

To use the ORFeome library/libraries to complement an incompatibility or a mutation, we had to transfer the ORFs from the entry clones into destination vector(s) that are able to replicate in rhizobia and contain promoter sequences, which ensure the expression of the genes at the proper developmental stage such as during infection or bacteroid development or nitrogen fixation. To create such vectors, first, we constructed three general purpose broad host-range cloning vectors from the relatively small and moderate copy number binary vector pPR97 (Szabados et al., 1995) that contain the *lacZ α* fragment with the pBlueScript multiple cloning site for blue-white selection and provide resistance against kanamycin/neomycin, gentamycin and tetracycline and termed as pPR97lacZ, pPRgmLacZ and pPRtcLacZ, respectively.

In the next step, we selected three rhizobial promoters with different expression pattern (Roux et al., 2014). The nodule specific *nifH* promoter is expressed only in the symbiotic nodule and is activated in the Interzone and reaches its highest activity in the nitrogen-fixing zone (Zone III). The *bacA* gene is expressed both in free-living state and in the nodule but its expression is enhanced in Zone II (infection zone) and is highest in the Interzone where bacteria develop into bacteroids. The *cycHJKL* operon is expressed at lower level than the other two both in cultured and nodule bacteria but constitutively in every zone without any spatial preference. These promoters and the Gateway cassettes were cloned into the vector pPRtcLacZ to obtain the pPRtcLacNifHprGW, the pPRtcLacCycHprGW and the pPRtcLacBacAprGW vectors.

As our earlier results (see later; Wang et al., 2017) showed that the *nifH* promoter is activated during the incompatible interaction between *M. truncatula* cv. Jemalong and strain Rm41, the ORFeome library containing all 6317 annotated ORFs of strain 1021 arranged in 67 pools were transferred from the entry clones into the pPRtcLacNifHprGW vector with the *in vivo* recombination method of House et al. (2004). The 67 pools of the library in the broad host-range vector were then introduced into strain Rm41.

2.2.2. Construction of the large insert size genomic library from strain FSM-MA

In our group, strain FSM-MA was mutagenized with such a modified version of the *himar* transposon we used for the mutagenesis of strain Rm41 that contains a number of recognition sites for restriction enzymes cutting rarely the rhizobial genomes. Genomic DNA from the mutagenized pools were isolated and digested with restriction enzymes SbfI (end compatible with the end generated by PstI), BsrGI (end compatible with the end generated by Acc65I), AclI (end compatible with the end generated by ClaI), XbaI, SpeI, NheI (ends compatible with the end generated by XbaI) and HpaI, DraI and Bst1107I generating blunt end. The resulting DNA fragments were separated by gel electrophoresis in 0.5% agarose gels and the region corresponding to 15-33 kbp were isolated

and ligated into the pZPgmSK2 broad host-range cloning vector constructed in our lab. Approximately 400 pools of 25-30 clones were created that can be mobilized into rhizobia.

To get insight into the quality of the library, 16 clones from the SbfI cloning were isolated and the ends were sequenced with M13 reverse and forward primers. Blast searches against the FSM-MA genome sequences with the obtained reads allowed the positioning of the endpoints on replicons of the genome and the calculation of the insert sizes as shown in Table 3.

As we can see, there are clones from all three replicons of the strain, although pSymA sequences are underrepresented while pSymB fragments are overrepresented in this small sample. It is satisfactory that the average insert size is larger than 10 kbp, thus, we might expect a good representation of full operons.

Table 3. Summary of the sequencing results from randomly chosen SbfI clones

Library clone	Sequence from replicon	Start of sequence with primer		Length of insert (bp)
		M13 reverse	M13 forward	
1	FSMchr	3005784	3019321	13537
2	FSMchr	3284676	3294904	10228
3	FSMpSymB	156807	143195	13612
4	FSMchr	3019321	3005784	13537
5	FSMpSymA	930730	920896	9834
6	FSMchr	1605508	1613739	8231
7	FSMpSymB	1164628	1170604	5976
8	FSMpSymB	782166	789462	7296
9	FSMchr	1432719	1414827	17892
10	FSMchr	67044	83560	16516
11	FSMpSymB	894665	912837	18172
12	FSMpSymB	894671	878748	15923
13	FSMpSymB	878753	864543	14210
14	FSMpSymB	1381174	1366566	14608
15	FSMchr	1465461	1445484	19977
16	FSMchr	1490679	1472843	17836
Average				13587

3. Towards the identification of the genetic determinants responsible for the incompatible interactions

Our aim is to identify the genes, gene variants in both plants and bacteria that are responsible for the incompatibility with the symbiotic partner.

3.1. Incompatibility of the *Sinorhizobium meliloti* strain SM11 with most *Medicago truncatula* lines

3.1.1. Phenotypic characterization of the SM11 induced nodules

The *Sinorhizobium meliloti* strain Sm11 forms incompatible interaction with *M. truncatula* A17 line. A17 plants develop white nodules on the roots 16 days post inoculation (dpi) wherein nitrogen

fixation does not occur (Fix- or ineffective symbiotic interaction). At later time points (20-21 dpi), A17 showed the symptoms of nitrogen deficiency and partially defoliated. We also tested the symbiotic phenotype of some widely used *M. truncatula* genotypes; DZA315, A20 and Jemalong following the inoculation with Sm11. Jemalong line, similar to A17, showed Fix- phenotype, DZA had an intermediate Fix+/Fix- phenotype, while A20 formed Fix+ nodules although these nodules showed early senescence. Jemalong nodules showed typical Fix- phenotypes containing dead Sm11 rhizobia, while the Fix+ A20 nodules were colonized in zones II, II-III and III (the nitrogen fixation zone) (Fig. 1). A20 nodules had the typical zonation of the IRLC nodules although the senescent zone was extended (red rectangle in Fig. 1) compare with the Fix+ nodules at similar age induced by other compatible rhizobium strains (data not shown).

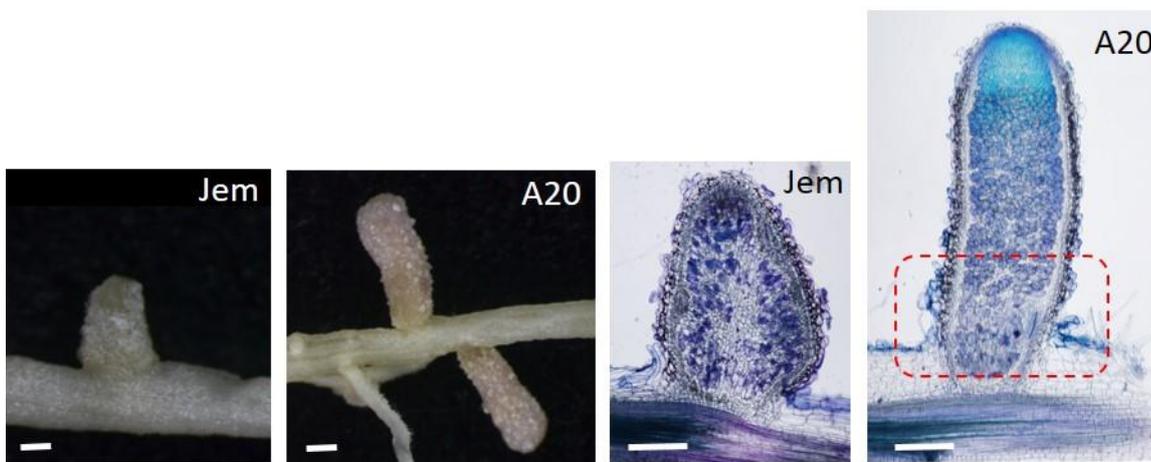


Figure 1. Stereomicroscope images of whole nodules and toluidine blue-stained longitudinal sections of Jemalong (Jem) and A20 nodules 3 wpi with Sm11. Red rectangle shows the extended senescent zone. Bars: 200 μ m

The nodules developed on Jemalong and A20 roots were further analyzed by fluorescent confocal microscopy (Figure 2.). Longitudinal sections of three weeks old nodules were stained with SYTO13 DNA-intercalating fluorescent dye. A20 nodules, which contained more intact DNA (plant and bacteria) show bright green fluorescent staining on the distal part of the nodules wherein the zonation was clearly visible. It was also apparent that the bacteria were long shaped, differentiated and were arranged around the plant cell vacuole in the cells of the infection zone and interzone (Fig 2. A20_a & b). This phenomenon is characteristic for the functional nitrogen fixing nodules of a compatible interaction. The main difference was observed in the proximal part of the nodule, where the cells were full of disintegrated remnants of plant cell nuclei and bacteria (Fig 2. A20_c). This senescent part of the nodule was 3-4-fold larger than in a nodule of a fully compatible interaction. In the smaller Fix- nodules of Jemalong plants some cells contained partially differentiated rhizobia and these cells form a thin layer corresponding to the infection zone of Fix+ nodules (Fig 2 Jem_a). Below this layer, most of cells were full of undifferentiated bacteria (Fig 2. Jem_b & c).

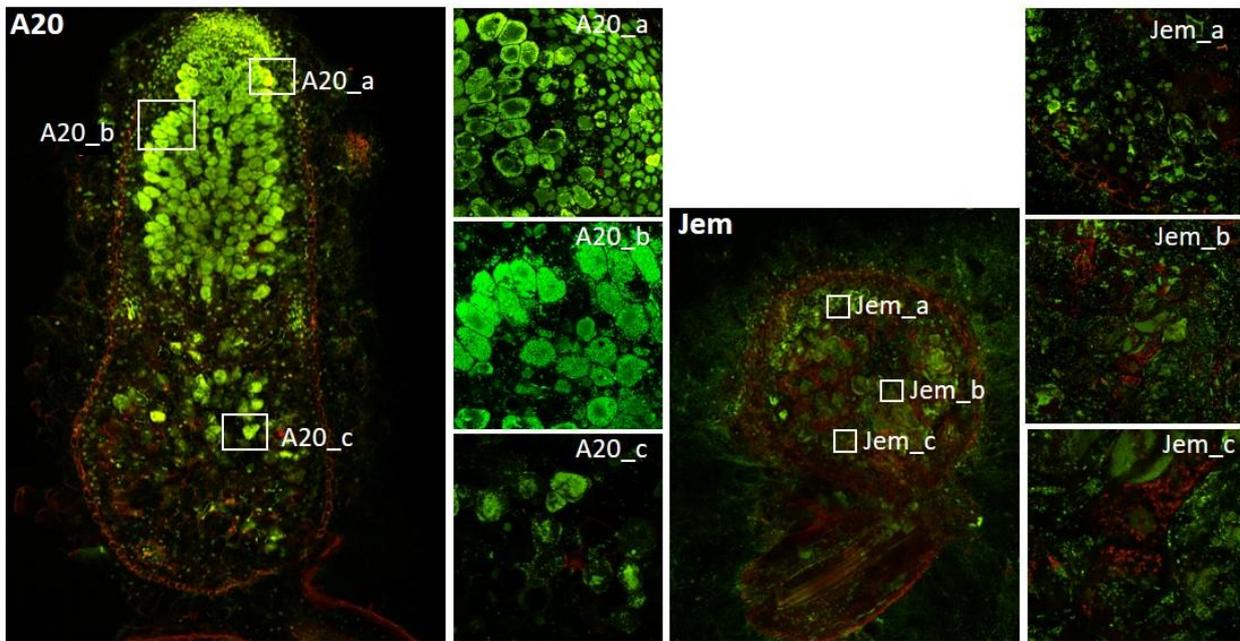


Figure 2. Confocal microscopy of SYTO13-stained sections of A20 and Jemalong nodules 3 weeks post inoculation with strain Sm11. Rectangles on whole nodule images shows the positions, where cells were presented in higher magnification.

One of the aims of the project was the isolation of the gene or genes which is/are responsible for the incompatibility of A17/Jemalong line with Sm11. We first performed genetic mapping on recombinant inbred lines (RILs) of the cross of *M. truncatula* parental lines Jemalong and DZA 315. We tested the progeny of 47 individuals from a Jemalong and DZA 315 (JD RIL) population for the symbiotic phenotype after inoculation with Sm11. Characterization of the symbiotic phenotype of several plants was not clear and obvious because they had mixed Fix- and Fix+ nodules. We could identify only a few plants from this RIL population that showed clear segregation of the Fix+ and Fix- phenotypes in their progeny. Unfortunately, these phenotypes were not sufficient to identify the position of symbiotic locus controlling the compatibility with Sm11 with using the existing genetic map of the JD RILs.

Beside the most commonly used ecotypes of *M. truncatula*, we tested the symbiotic phenotype of 110 other ecotypes following inoculation with Sm11. We obtained only 12 Fix+ plants among them, but the symbiotic nodules of these lines showed even more intensive senescence than A20 nodules. However, this result was useful later, when we tested single genes as potential incompatibility factors. We have also tested the Sm11 phenotype of five lines belonging to the *Medicago sativa* complex (*Medicago hemicycle*, *M. glandulosa*, *M. quasifalcata*, *M. caerulea*, *M. trantvetteri*), which all showed clear Fix+ phenotype. As Sm11 was isolated from a continental area of Europe, this result

is consistent with it and it can explain partly why most of *M. truncatula* lines, which live in the mediterranean showed incompatible interaction.

3.1.2. Genetic mapping of the locus conferring incompatibility between Jemalong and Sm11

As a conclusion of the Sm11 phenotypes of different *Medicago* lines we established a mapping population of *M. truncatula* cv. Jemalong (Sm11 incompatible line) and A20 (Sm11 semi-compatible line). We generated F1 hybrid plants of the cross of Jemalong and A20. Hybrid F1 plants showed Fix- phenotype suggesting that the incompatibility with Sm11 is a dominant character. F2 mapping populations were generated by self-pollinating the F1 hybrid plants and the 184 individual of the mapping populations were scored for symbiotic phenotype in nodulation assay and genotypes were determined using published and newly designed PCR-based genetic markers. In several cases, the nodulation phenotype of the F2 individuals was not as obvious as the phenotype of the parental lines, therefore we could not determine phenotypes of 80 F2 plants clearly. We could identify the symbiotic phenotype of 44 Fix+ (A20-like semi-compatible interaction) and 48 Fix- (incompatible interaction) F2 plants and based on the genetic analysis of the symbiotic phenotype and the genotypes of the mapping markers, we could predict a map position of the locus responsible for the compatibility of the interaction with Sm11. The locus was positioned to the lower arm of chromosome 8 of *M. truncatula*, between 31-33 Mega base pairs (Mbps) (*Medicago truncatula* A17 r5.0 genome portal). To carry out the fine mapping the region which is responsible for Sm11 incompatibility, we tested about 500 plants of the F3 generation of the original Jemalong and A20 crossing. We scored these plants for their symbiotic phenotype in nodulation assay and determined their genotypes for the flanking markers. The clear phenotype about 30% of the F3 plants could not be determined but the symbiotic phenotype ~ 70% of plants helped us to position the genetic locus conferring the compatibility to Sm11 on chromosome 8 between 31,4-32,3 Mbps. In this region 108 annotated genes or gene models have been detected in the *M. truncatula* genome browser and about 50% of these genes showed expression in nodules but only twelve genes were nodule-specific. We surveyed these genes to detect amino acid alterations between Jemalong and A20. The genes which had any amino acid change were also checked using the Hapmap data of *Medicago truncatula* A17 r5.0 genome portal in which the SNPs of the ecotypes were collected. None of the analysed 12 genes showed differences in amino acid sequences between compatible and non-compatible ecotypes that could correspond to the Sm11 incompatibility.

We also monitored the expression level of seven genes between Jemalong and A20 inoculated either by Sm11 or by *S. medicae* WSM419, a compatible strain for both ecotypes and some other ecotypes with RT- qPCR. The analysis has not revealed any significant difference in the level of

expression of these genes (Fig 3.) that might be responsible for altered symbiotic performance of the plants with Sm11.

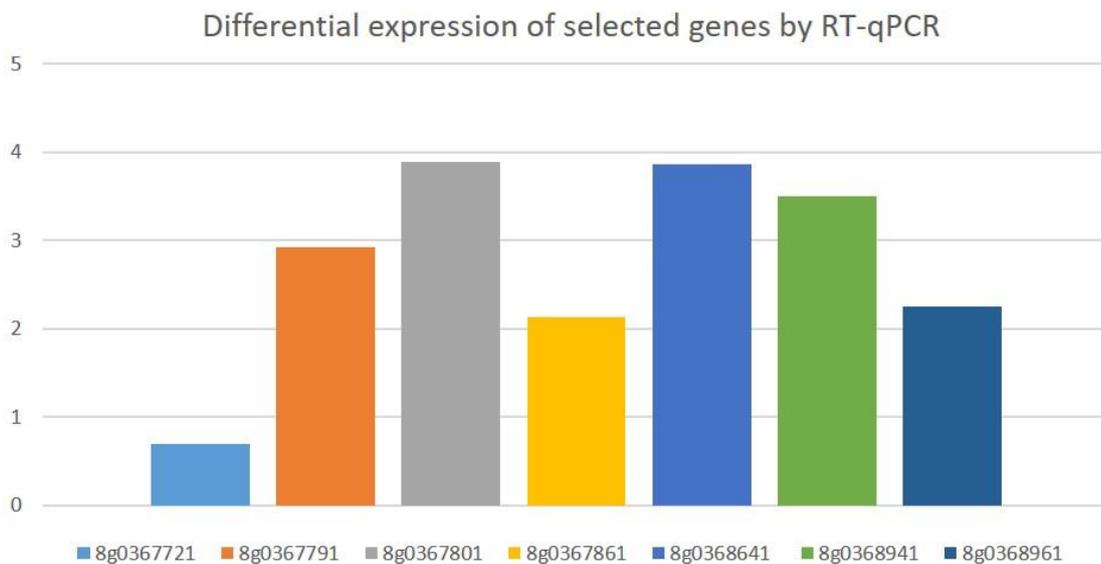


Figure 3. Differential expression of seven genes in the Sm11 region. Differences in the expression of 7 genes between Jemalong and A20 14 dpi with WSM were compared with the differences detected for same 7 genes between Jemalong and A20 14 dpi with Sm11.

Because we have not found proper candidate gene in the 0,9 Mbp region, which could be responsible for Sm11 incompatibility, we decided to refine the Sm11 region with pheno- and genotyping using further 340 individuals of F4, 52 individuals of F5 and 8 individuals of F6 plants. With the help of these plants we restricted the Sm11 region to 32,064 – 32,314 Mbp on chromosome 8 (Fig 4.), in which 32 annotated genes or gene models have been predicted.

Plant Number	bps on Chr 8	59	7	18	23	32	21	30	61	33	35	47	53	9	62	3	16	12	5	17	36	6	15	22	42	44	45	48	11	
PPDKchoi	31854088	0	3	3	3	0	3	0	3	0	3	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
h2_69d5a-F	31970318	0	3	0	0	3	0	3	0	3	0	0	0	3	0	3	3	3	0	2	2	2	2	2	2	2	2	2	2	1
Mt_8g069915	32042102	3	3	0	0	0	0	0	0	0	0	0	3	3	0	0	3	3	2	2	0	0	0	0	2	0	0	0	0	0
Mt_8g069925	32066752	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3	0	0	0	0	0	0	2	0	0	0	0	0
Sm11 phenotype		3	2	5	3	2	5	2	1																					
Medtr8g070680	32314993	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	2	3?	0	2	2	0	0	0	3	0	0	0	2	1
Medtr8g070830	32467493	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	2	0	2	2	0	0	0	0	0	0	0	2	1
Medtr8g071050	32596267	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	2?	2?	0	2	2	0	0	0	0	0	0	0	2	1
Medtr8g071250	32695732	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	3	2	2	2	2	1
Medtr8g071300	32725671	0	3	0	0	3	0	0	0	0	0	0	0	0	0	3	2	2	0	0	0	0	0	0	3	0	0	0	2	0

Figure 4. Refined map position of the incompatibility locus

Only one of these genes showed nodule-specificity based on the *Medicago truncatula* expression database, but we could not detect any amino acid change between the compatible and incompatible lines. Therefore, we concluded that a roughly 300 kbp region of chromosome 8 is responsible for the

incompatibility of Sm11 on *Medicago truncatula*. In order to define the gene or genes responsible for the compatibility with Sm11, we are going to resequence of the genome of A20 to identify additional genes in A20 or reveal other alterations between the genomes of Jemalong and A20.

3.1.3. Testing the systemic effect of Sm11

Three weeks after inoculation of Jemalong line with Sm11 strain, the plants showed serious N-starvation symptoms: the leaves become yellow and then quickly exfoliated. This phenomenon was much more severe than any other ineffective (Fix-) *Medicago* mutant we have analysed. We wanted to test if this severe symbiotic response emerged from interaction in the root that might be transferred through the shoot to another root segment or it functions locally. In order to test this hypothesis, a split root experiment was carried out. We induced the bifurcation of the root by removing the root meristem. The emerging two roots, of Jemalong and A20 seedlings were carefully inoculated with strains Sm11 (incompatible) and with strain *S. medicae* WSM419 (compatible), respectively. All the plants developed healthy dark green leaves at 19 days post inoculation indicating the effectiveness of the symbiotic interaction although small white nodules were formed on roots inoculated with Sm11 while WSM419-inoculated roots developed big pink functional nodules. We concluded that Sm11 infection on a root did not block the plant to make effective Fix+ nodules on other roots indicating that Sm11 does not induce systemic toxicity on Jemalong plants.



Figure 5. Split root experiment. WSM and Sm11 induced root nodules on bifurcated Jemalong and A20 roots

3.2 The incompatible interaction between strain Rm41 and ecotype F83005

When inoculated with Rm41, F83005 formed barely visible nodule primordia devoid of bacteria due to failed initiation or elongation of infection threads (Fig 5.), but the plant formed normal nitrogen-fixing nodules (Nod+/Fix+) with other strains such as *S. meliloti* 1021 and *S. medicae* ABS7. Similarly, Rm41 forms effective symbiosis with ecotypes A20, DZA315.16 or DZA045.

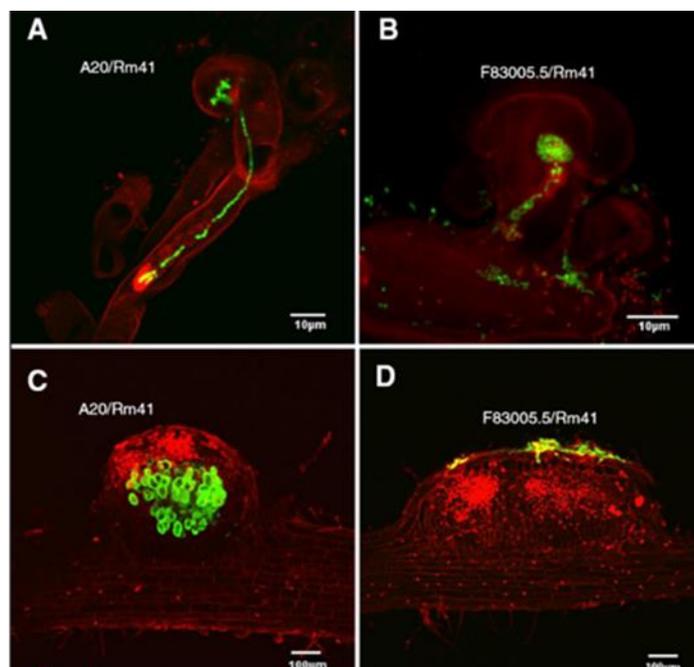


Figure 6. Fluorescence microscopy analyses of infection process of compatible and incompatible interactions between *M. truncatula* plants and Rm41. All images are composite images of GFP-expressing Rm41 cells (green) and root cells (red). A, Typical infection thread formed by compatible interaction between A20 and Rm41. The infection thread extends from the colonized, curled root hair to the base of the root hair cell. B, In the incompatible interaction between F83005.5 and Rm41, the bacteria can normally colonize the curled root hairs but typical infection threads cannot be detected. Occasionally, we can detect aborted, aberrant infection threads present on the F83005.5 roots. C, The nodule primordium on the A20 roots contained bacteria, while the nodule primordia on the F83005.5 roots (D) contained no bacteria despite frequent presence of bacterial colonies on the epidermal surface of the nodule primordia.

3.2.1. Identification of the plant genes causing the incompatibility

For genetic analysis of the nodulation specificity, first, an F2 population derived from the cross between the compatible genotype A20 and the incompatible ecotype F83005.5 was used. From a total of 2,623 inoculated F2 plants, 686 plants formed nodules, which fits the 3:1 (non-nodulation to nodulation) ratio, suggesting that the restriction of nodulation by Rm41 in F83005.5 is controlled by a single dominant locus, which we termed NS1 (for nodulation specificity 1). Similar results were obtained with another F2 population obtained after crossing F83005 with the compatible ecotype, DZA045. The dominant nature of this nodulation-inhibitive locus, together with the ability of the

bacteria to induce root hair curling and nodule primordium formation, suggests that this incompatibility is not caused by a defect in Nod factor perception. Moreover, unlike the resistance gene-mediated nodulation inhibition in soybeans (Yang et al., 2010; Sugawara et al., 2018), this specificity is independent of bacterial type III effectors because Rm41 does not possess genes essential for assembly of the type III secretion machinery (Weidner et al., 2013). We delimited the *NSI* locus to a small genomic region, based on linkage mapping in two F2 populations. In the reference genome of Jemalong A17 (Nod+/Fix-), this region covers about 50 kilo-base pairs (kb), containing a gene that encodes a receptor-like kinase. This receptor-like kinase comprises an N-terminal signal peptide, an extracellular malectin-like domain, three leucine-rich repeats, a single-pass transmembrane domain, and an intracellular kinase domain. Notably, there is also a truncated, presumably pseudogene that encodes only a partial kinase domain nearly identical to that of the full-length gene. Sequencing and annotation of the genomic regions surrounding the *NSI* locus from the three parents revealed both presence/absence and sequence-level polymorphisms of the tandem gene copies. In particular, F83005 (Nod-) and DZA045 (Nod+) both possess full-length copies of the two genes (hereafter referred to as *NS1a* and *NS1b*), while *NS1b* is missing in the genome of A20 (Nod+). We therefore postulated that *NS1a* and *NS1b* are the two candidate genes at the *NSI* locus in F83005 (Nod-) that confer resistance to infection by Rm41. We tested the candidate genes through complementation tests and CRISPR/Cas9-based gene knockout. The absence of an *NS1b* allele in A20 (Nod+) and a lack of nonsynonymous substitutions between the *NS1a* alleles of A20 (Nod+) and F83005 (Nod-) suggested that the *NS1b* allele of F83005 (Nod-) confers resistance to infection by Rm41. Consistent with this reasoning, transgenic hairy roots of A20 expressing the *NS1b* allele of F83005 (Nod-) showed a Nod- phenotype, while transgenic hairy roots expressing the *NS1a* allele of F83005 retained the Nod+ phenotype. Surprisingly however, we did not identify nonsynonymous polymorphisms between the *NS1b* alleles of F83005 (Nod-) and DZA045 (Nod+) within the co-segregated region; accordingly, introduction of the *NS1b* allele of F83005 (Nod-) into DZA045 (Nod+) failed to convert the Nod+ phenotype to Nod-. Nonetheless, the *NS1a* alleles are polymorphic between the two parents, and transfer of *NS1a* of F83005 (Nod-) into DZA045 (Nod+) enabled the transgenic hairy roots to block nodulation by Rm41. Based on these observations, we concluded that *NS1a* and *NS1b* are both required for nodulation restriction in F83005 (Nod-). We further validated this conclusion by CRISPR/Cas9-mediated gene knockout; mutating either or both of *NS1a* and *NS1b* in F83005 (Nod-) led to the formation of normal infected nodules.

A search of published transcriptome data for the reference genotype Jemalong A17 (containing only *NS1a*) suggested specific expression of *NS1a* in nodules and rhizobium-inoculated roots (Roux et al., 2014; Larrainzar et al., 2015; Jardinaud et al., 2016). We further analyzed the expression pattern of *NS1a* and *NS1b* by assaying their promoter activity in transformed roots of F83005 using

pNS1a/b:GUS reporters. GUS assays revealed a barely detectable promoter activity for both genes in non-inoculated roots, restricted to the vascular bundle of transgenic hairy roots. However, GUS activity was strongly induced upon inoculation with either compatible (ABS7) or incompatible (Rm41) strains; such induced expression occurred only in the root zones forming nodule primordia. In line with this observation, we demonstrated that this gene induction was dependent on host perception of Nod factors, as a *nodC* mutant of Rm41 deficient in Nod-factor biosynthesis failed to induce the gene expression. Supporting this conclusion, *dmi1* and *dmi2* plant mutants defective in Nod-factor signaling (Ané et al., 2004; Endre et al., 2002) were unable to express *NS1a* in the Jemalong A17 background upon inoculation with Rm41 or ABS7. Similarly, plants with mutations in the Nod factor receptors *MtNFP* and *MtLYK3* also failed to express *NS1a* in the rhizobium-inoculated roots (Larrainzar et al., 2015). We also used the *NS1a* and *NS1b* promoters to drive the expression of a nuclear-localized eYFP reporter (eYFP-NLS). This experiment revealed the expression of *pNS1a/b*:eYFP-NLS in both epidermal and cortical cells of the root primordium. The expression was also observed in fully developed nodules when inoculated with compatible strains. This latter observation is consistent with the expression pattern revealed by RNA-seq analysis of different nodule zones derived from laser-capture microdissection, which showed that *NS1a* was primarily expressed in the meristematic and infection zones of the mature nodule (Roux et al., 2014).

3.2.2. Identification of a bacterial gene causing incompatibility

When we screened the transposon insertion mutant populations of Rm41 on F83005 plants, we found two independent mutants that could form effective symbiosis with this ecotype. To identify the bacterial gene(s), which caused incompatibility with F83005, we isolated and sequenced the transposon insertion sites from the mutants. Both mutants carried the transposon insertion in the same gene annotated as *BN406_06091* although in different position. Although the two independent mutations in the same gene indicated that this gene caused the incompatibility, we created a targeted mutant of strain Rm41 by disrupting *BN406_06091* with an antibiotic resistance cassette to confirm that this gene makes this strain incompatible with the F83005 plants and indeed, all three strains efficiently nodulated F83005. In addition, we created similar mutants in strains AK83 and SM11, which are incompatible with F83005 and carry the *BN406_06091* and the surrounding genes, and the mutant bacteria became compatible as observed in Rm41. The 1350 base pairs long *BN406_06091* gene is located on the second symbiotic megaplasmid. The gene encodes a protein of unknown function containing 449 amino acids. Searching for conserved protein domains using the IUPred2A (Erdős and Dosztányi, 2020) and InterPro (Blum et al., 2021) online prediction programs (<https://iupred2a.elte.hu/>; <https://www.ebi.ac.uk/interpro/>) revealed that the first half of the protein contains a right handed beta helix region, which is followed by a disordered region. Right handed

beta helices were first recognized in pectate lyases involved in the degradation of the pectic components (polysaccharides containing high amount of galacturonic acid) of the plant cell wall, thus, the structure prediction indicated a possible sugar/polysaccharide associated role of the protein. The disordered part might be involved in protein-protein interaction(s) when its structure might be stabilized. Neither cleavable N-terminal signal peptide nor transmembrane helices could be detected in the protein sequence by bioinformatics tools, but interestingly, the PSORTb web-server predicted an extracellular localization for the protein. To investigate the cellular localization of the protein, we created a construct coding for a BN406_06091 protein fused with the HA tag (06091-HA) in a vector expressing GFP and analyzed the cell protein fractions using Western blotting with anti-GFP and anti-HA antibodies. In the immunoblotting analysis both the GFP and the 06091-HA proteins were detected only in the cytoplasmic fraction (Figure 7).

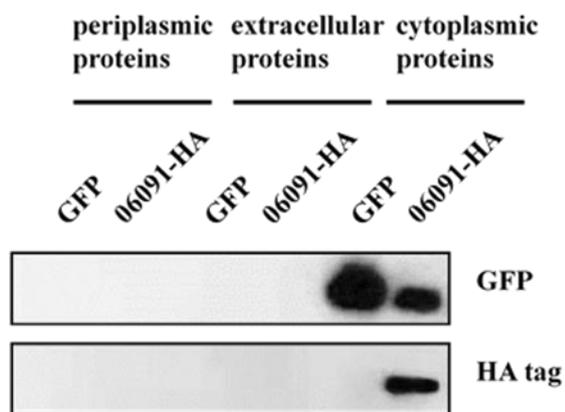


Figure 7. Localization of the protein encoded by gene BN406_06091 by Western blot analysis. Periplasmic proteins (left), extracellular proteins (middle) and cytoplasmic proteins (right) isolated from strains Rm41S_mBN406_06091/pHC60::GFP and Rm41S_mBN406_06091/pHC60::06091-HA were detected using anti-GFP and anti-HA antibodies.

To place the *BN406_06091* gene into a genomic context, we investigated its neighboring genes (Figure 8). The gene is surrounded by a number of genes coding for proteins involved in sugar and polysaccharide synthesis and modifications implicating its role in polysaccharide production. The immediate neighboring region is specific for only certain *S. meliloti* strains such as Rm41, M270, AK83 and SM11 indicating a strain-specific role. On one side of this region, the *rkp-3* gene cluster (Kereszt et al., 1998) responsible for the production of the sugar precursors as well as for the assembly and export of the strain-specific capsular polysaccharide (KPS/K- antigen) was located (Kiss et al., 2001). On the other side of *BN406_06091* containing region, we identified a gene cluster being present in a number of *S. meliloti* and *S. medicae* strains that contains genes coding for glycosyl-

transferases, sugar activating and modifying enzymes indicating their role in polysaccharide production.

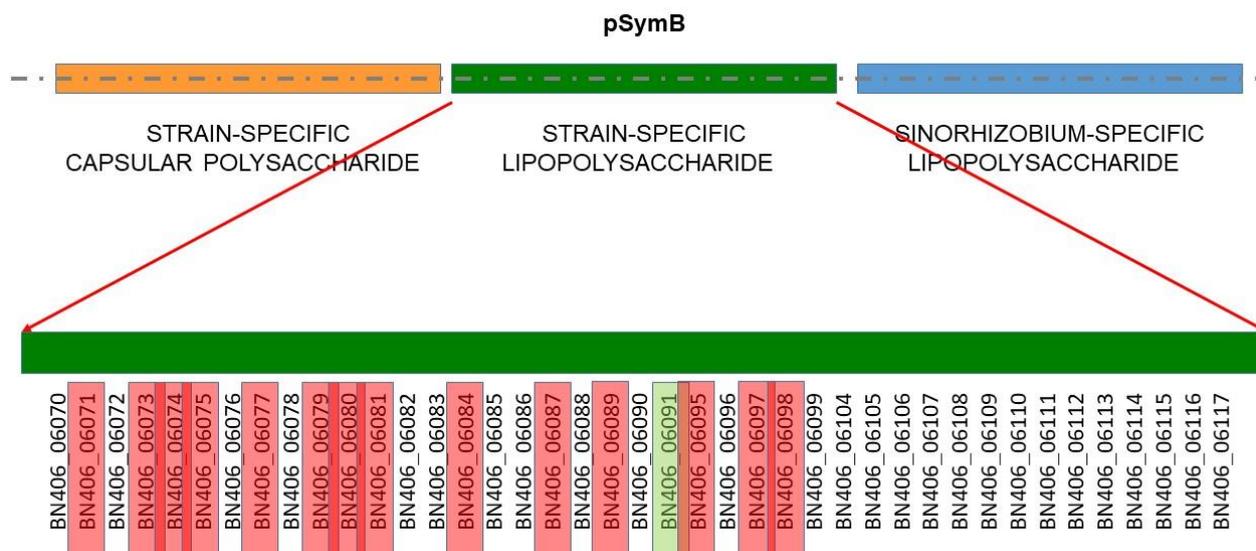


Figure 8. Schematic representation of the investigated part of the Rm41 pSymB megaplasmid with gene clusters potentially involved in surface polysaccharide production. In the upper part, the three regions involved in polysaccharide production are shown. In the lower part, genes in the strain-specific region coding for proteins predicted to be involved in sugar/polysaccharide synthesis and modification are listed. Selected genes mutated with plasmid integration are boxed. Green box: mutation in the gene restored compatibility. Red box: mutations could not restore compatibility.

Based on our knowledge on rhizobial surface polysaccharides, we suppose that the *BN406_06091* gene containing region is responsible for the production of the strain-specific O-antigen of LPS while the other region might contribute to the synthesis of the genus-specific LPS core oligosaccharide, which links the lipid A part in the outer membrane and the O-antigen on the surface. This idea is supported by both the phenotype observed during the interaction and by the plant proteins responsible for the incompatibility: The incompatibility between F83005 and Rm41 is due to a block of bacterial infection, i.e. bacteria (Nod factors) induce root hair curling and microcolony formation in the infection chamber but the plant arrests the formation of the infection threads. The incompatibility, i.e. infection restriction is caused by the NS1a and NS1b receptor kinases that contain malectin domain implicated in sugar binding in their extracellular part. According to our model, this receptor pair recognizes the surface of Rm41 as that of an “enemy” and as a result, the plant blocks the entry into

the root. Changing or modifying or destroying the surface polysaccharide or the receptors will stop the recognition and the blocking of the infection.

Although we could not isolate other mutants establishing effective symbiosis with F83005 despite the high number of the independent insertion events, we created other mutants that might be affected in the LPS structure: 1) We targeted a number of genes in the vicinity of *BN406_06091* as well as other genes (*lpsL*, *rkpK* *lpsB*) that were also implicated in LPS production that might contribute to the O-antigen or core oligosaccharide synthesis; 2) Our collaborating partners at the University of Kentucky performed binding assays with the purified extracellular domain (produced *in vitro* in a heterologous system) of the receptors and detected preference towards sulfated sugars, that is why, we mutated genes that are known or thought to be involved in sulfate activation and sulfate modifications of sugars/polysaccharides. In *S. meliloti*, three sulfate activation systems (CysHDN, NodP₁Q₁, NodP₂Q₂) produce first adenosine 5'-phosphosulfate (APS) from ATP and sulfate ions, then the APS kinase activity synthesizes 3'-phosphoadenosine 5'-phosphosulfate (PAPS) from ATP and APS that is used in the sulfating reactions (Schwedock & Long, 1990; Roche et al., 1991; Schwedock et al., 1994; Abola et al., 1999). In addition, it was shown (Keating, 2007; Cronan and Keating, 2004) that the LpsS protein of *S. meliloti* is involved in the sulfation of its LPS, that is why, we selected these coding genes along with *BN406_02129* (encoding a hypothetical protein sulfotransferase) for mutagenesis. In one approach, we amplified internal fragments of the genes and cloned them into the pK19mob vector (Schäfer et al., 1994), then mobilized the clones into *S. meliloti* strain Rm41. As the vector cannot replicate in rhizobia, kanamycin resistant colonies can be obtained after the clones integrated into the genome via homologous recombination through the inserts. These integration events disrupt the genes and the plasmid integration mutants were used in the nodulation assays. In the case of other genes, we deleted the coding sequences using the pK18mobsacB vector and sucrose selection as described by Schäfer et al. (1994). These mutants were then tested in F83005 nodulation assay. Four weeks after inoculation, all of the plants showed Fix⁻ phenotype, evidenced by chlorotic shoots and smaller, yellowish leaves.

We also tried to introduce the *BN406_06091* gene alone into *S. meliloti* strain Sm1021 and *S. medicae* strain WSM419 to test whether its expression in compatible bacteria can modify their cell surface to make them incompatible. Although the clone could “complement” the *BN406_06091* mutant, i.e. the transconjugant became incompatible again, the two compatible strain carrying the *BN406_06091* gene remained compatible with F83005.

To facilitate the identification of the molecular changes caused by the *BN406_06091* gene mutation, we created mutations to prevent the synthesis of other polysaccharides and introduced them into the wild-type and *BN406_06091* mutant strains. The polysaccharide composition of the different strains is being analyzed by our collaborators in the Complex Carbohydrate Research Center at the

University of Georgia (Athens, GA, USA). As soon as the composition analysis is complete, we can publish the work.

3.3. The incompatible interaction between strain ABS7 and ecotype DZA220H

S. medicae strain ABS7 is a very effective partner of a number of *M. truncatula* ecotypes including Jemalong, A20 and DZA315 but DZA220H plants arrest the symbiotic process at the initiation of the infection threads. This phenotype is very similar to the one observed with Rm41 on F83005 plants. Preliminary results from the genetic analysis of the trait in the plant indicate that a similar receptor pair is responsible for the incompatibility in DZA220H as we identified in F83005. The complementation experiments as well as knocking out the genes is in progress. The first results of gene identification in the bacterial partner revealed a gene coding for a murein L,D-transpeptidase, which is implicated in the peptidoglycan remodeling and cross-linking Braun's lipoprotein to the peptidoglycan and the maintenance of the outer membrane integrity. Screening of the mutant populations on the DZA220H plants has to be repeated as described in Chapter 2.1.2.

3.4. The incompatible interaction between strains Rm41 and A145 and ecotype Jemalong

S. meliloti strain Rm41 forms Fix⁺ nodules on the *M. truncatula* accession DZA315.16 (DZA315) but Fix⁻ nodules on Jemalong A17 (A17). The Fix⁻ phenotype of A17 results from its incompatibility with Rm41, because it can establish efficient symbioses with other strains. Rm41 bacteria are able to invade and colonize A17 nodule cells but undergo lysis shortly after differentiation into elongated bacteroids (Fig. 9 A-F and Fig. 10). Inoculation of plants by Rm41 carrying a GUS reporter gene driven by the *nifH* promoter showed that the *nifH* gene, encoding one component of the nitrogenase enzyme complex, was expressed in the young A17 nodules, further supporting the presence of differentiated bacteroids at early stages of nodule development; but the expression was abolished three weeks post inoculation due to bacterial lysis and nodule senescence (Fig. 9 G-L). Accordingly, the expression of several early senescence marker genes was dramatically enhanced in A17 nodules when compared with their expression levels in DZA315 nodules (Fig. 11).

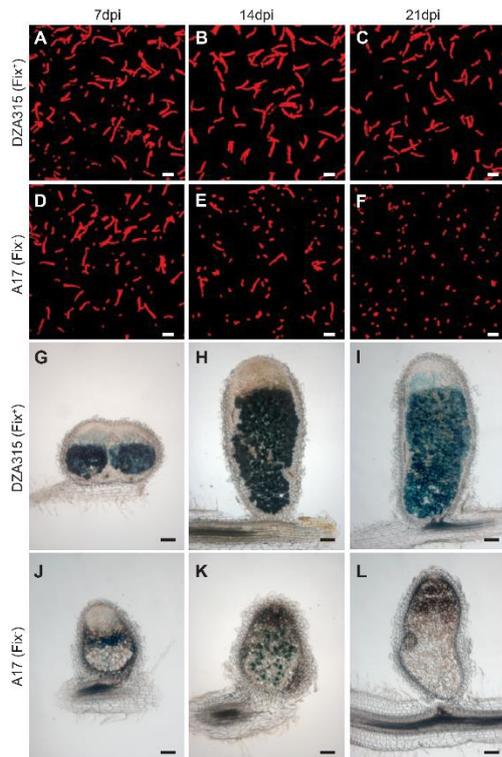


Figure 9. Differentiation and intracellular survival of the Rm41 bacteroids in the DZA315 (Fix+) and A17 (Fix-) nodules. (A–F) Confocal microscopy images of the bacterial populations isolated from the Fix+ (A–C) and Fix– (D–F) nodules showing the gradual loss of elongated bacteroids and the accumulation of saprophytic bacteria in the Fix– nodules. (G–L) Expression of a GUS reporter gene driven by the *nifH* promoter in the Fix+ (G–I) and Fix– (J–L) nodules showing that the *nifH* gene was expressed at the early developmental stages of the Fix– nodules, but its expression was abolished at later time points. [Scale bars, 5 μm (A–F) and 200 μm (G–L).]

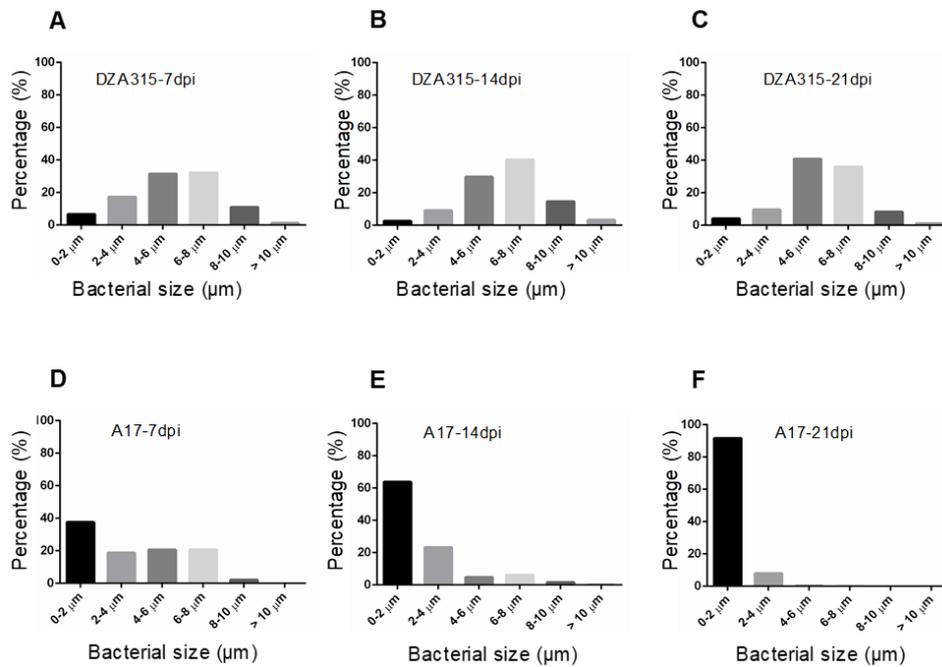


Figure 10. Measurements of bacterial cell length reveal the step-by-step loss of elongated bacteroids and the accumulation of saprophytic bacteria in the A17 nodules.

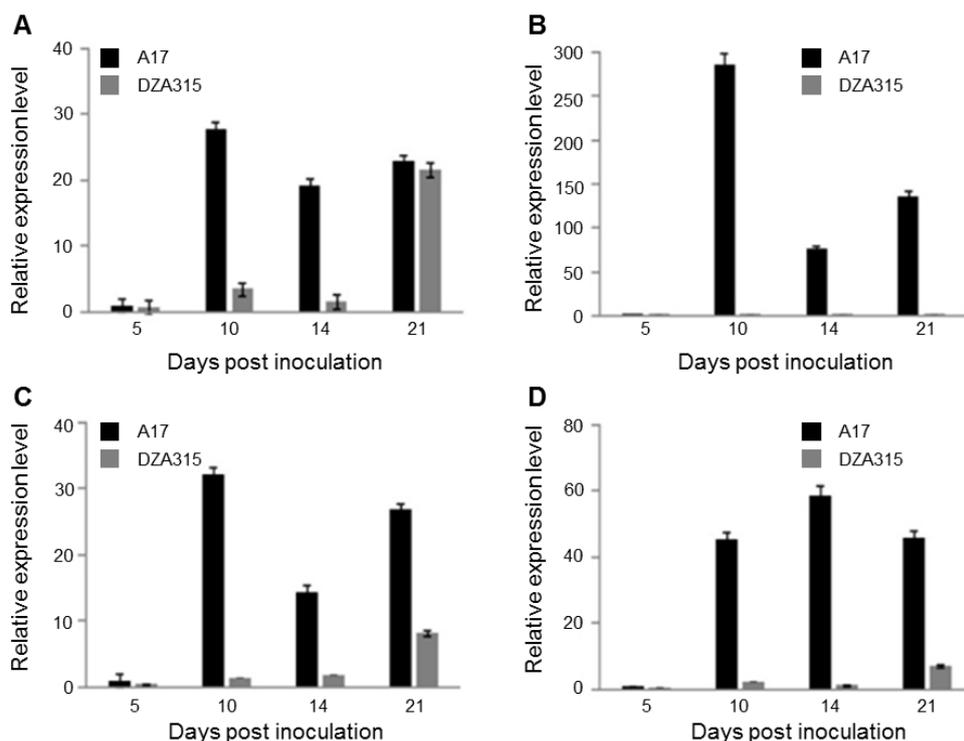


Figure 11. Up-regulation of early senescence marker genes in A17 nodules. RT-qPCR analysis of four selected putative senescence-related genes (38) that showed elevated expression in the A17 transcriptome. (A) Cysteine protease (*Medtr4g079800*); (B) Cysteine protease (*Medtr5g022560*); (C) Purple acid phosphatase (*Medtr7g104360*); and (D) Chitinase (*Medtr6g079630*). The experiments were performed at the indicated time points. Three biological replicates each with two technical repeats were used.

To identify the bacterial gene causing the incompatibility we introduced both the ORFeome and the large insert libraries into Rm41 but could not find any transconjugant with restored compatibility. Similarly, inoculation of the mutant populations resulted in no compatible interaction. However, as we indicated earlier in Chapter 2.1.2., the experimental set-up for the nodulation assay with the mutant populations is in need of improvement: we will test single pools on several hundred plants to achieve that all mutant will form nodules. For further improvement, we obtained the *sunn* supernodulating mutant of Jemalong that forms 5-10 times nodules than the wild-type plant. A new result also confirms that we need to repeat the experiments with mutant populations: The *S. meliloti* strain 1021 with a reduced symbiotic, pSymA genome establishes Fix⁺ symbiosis with Jemalong plants, while a hybrid strain with strain 1021 chromosome and pSymB and strain Rm41 pSymA induces Fix⁻ nodules as strain Rm41 meaning that Rm41 pSymA carries the determinant(s) for incompatibility.

Genetic analysis in a recombinant inbred line (RIL) population derived from the cross of A17 and DZA315 suggested the involvement of multiple interacting loci in the control of this symbiotic specificity. From a total of 146 RILs inoculated with Rm41, 92 formed Fix⁺ nodules and 52 formed Fix⁻ nodules; two residual heterozygous lines (here called RHL-NFS1 and RHL-NFS2) segregated

for both phenotypes. The segregation of the Fix⁺ and Fix⁻ phenotypes in the RIL population did not fit the 1:1 ratio expected from a single gene model. The involvement of multiple genes controlling the nitrogen fixation phenotype posed a challenge for fine mapping of the underlying loci using an F₂ or RIL population. Nonetheless, using extended mapping populations created from the residual heterozygous lines, bulked segregant RNA-Seq analysis and molecular markers, two closely linked loci on chromosome 8 designated as *NFS1* and *NFS2* (for nitrogen fixation specificity) were identified. Introduction of as well as knocking out the compatible and incompatible alleles revealed that allelic variants of two nodule-specific cysteine-rich (NCR) peptides are responsible for the incompatibility with strains Rm41 and A145. The peptides are exclusively produced in the infected cells of the nodules around the transition between the infection and fixation zones in the Fix⁺ and Fix⁻ nodules, consistent with the expression pattern revealed by RNA-Seq analysis of different nodule zones obtained with laser-capture microdissection. We have shown that not the antimicrobial activity of the peptides is causing the incompatibility because the both the compatible and incompatible variants similarly affected Rm41 and these effects were not different from those observed on the compatible strain ABS7.

3.5. Does effector-triggered immunity play a role in *Medicago* nodulation?

In the soybean-rhizobium symbiotic interactions, the combinations of bacterial T3 effectors delivered by the Type III Secretion System (T3SS) and plant resistance proteins may result in symbiotic incompatibility. The *S. meliloti* and *S. medicae* genomes does not code for T3SS and T3 effectors but we were interested whether plant nucleotide-binding site/leucine-rich repeat (NBS-LRR) resistance proteins play a role in *Medicago* nodulation. We showed that both pathogen infection and symbiotic interaction modulates the expression level of NB-LRR-regulating miRNAs and the promoters of Mtr-MIR2118a,b,c, Mtr-MIR2109, and Mtr-MIR1507 genes contain both pathogen-responsive and nodulation-regulated motifs. In addition, the NB-LRR-targeting miRNAs are upregulated during symbiotic nodule development and silence NB-LRR mRNAs. Using *Agrobacterium rhizogenes*-mediated hairy root transformation to express short sequences mimicking miRNA target sites that lead to the degradation of the NB-LRR-targeting miRNAs and hence the enhanced level of the NB-LRRs we found a significant reduction in the number of mature nodules compared to the control roots transformed with empty vector. In conclusion, the miRNA-based post-transcriptional regulation would allow the plant to fine-tune the level of NB-LRR transcripts depending on the character of the microbe interacting with the plant. Pathogen infection can decrease the basal level of the NB-LRR-regulating miRNAs resulting in efficient defence response. However, during symbiotic interaction, the increased expression level of the NB-LRR-regulating miRNAs

lowers the transcript level of defence genes and therefore can provide a suitable niche to the symbiotic bacteria to infect the roots and colonize the developing nodules. Because the NB-LRR-targeting miRNAs can respond to both pathogens and symbiotic partners in *M. truncatula*, they could function as an elegant molecular switch to provide a fast and adequate response by increasing or decreasing the level NB-LRR mRNAs according to the nature of plant–microbe interaction.