**FINAL REPORT (OTKA NN110979)**

The aim of the project has been to reveal molecular mechanisms that underlie bacterial differentiation during the establishment of the symbiosis between leguminous plants and soil bacteria collectively called rhizobia.

*Objective 1: Evolution of the NCR gene family in the IRLC group of legumes*

To investigate the fate of bacteroids in plants belonging to different genera of IRLC, *Astragalus canadensis, Cicer arietinum, Galega orientalis, Glycyrrhiza uralensis, Onobrychis viciifolia and Ononis spinosa* plants representing different branches of the phylogenetic tree were inoculated with their respective rhizobial partners and fully mature nodules were collected at 4 weeks post-inoculation. A longitudinal section of the nodules showed that they were properly infected by their rhizobial partners. Interestingly, a closer inspection of the infected cells revealed drastic differences in the bacteroids morphotype. *G. orientalis* showed Y-shaped bacteroids, similar to those found in *M. truncatula* and *P. sativum*. In *A. canadensis* and *O. viciifolia*, bacteroids were elongated while in *G. uralensis* bacteroids looked like the U-morphotype and no branched cells were observed in these species. It is noteworthy the morphotype of bacteroids from *O. spinosa* nodules since they resemble the S-morphotype, that has been observed in bacteroids from *Arachis* and *Aeschynomene* species. Similarly, *C. arietinum* nodules contained spherical bacteroids that were developed via a swollen form (Figure 1).

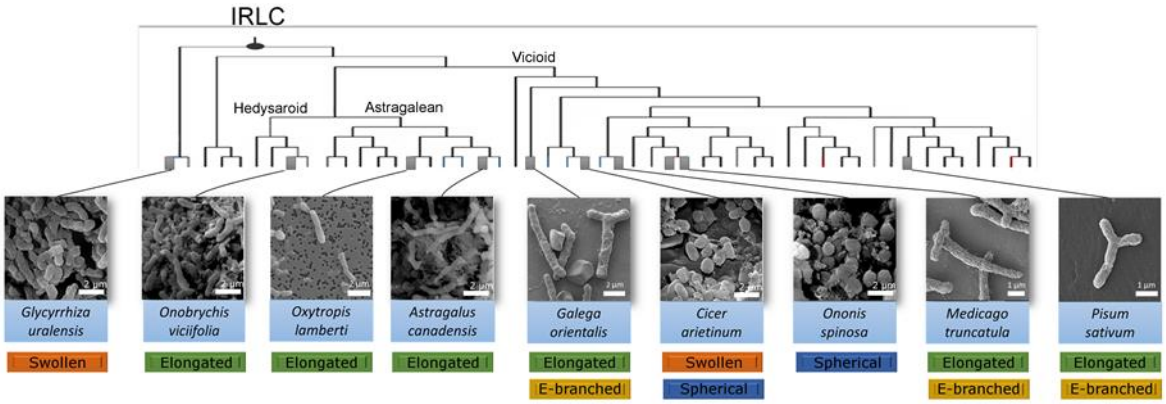


**Figure 1. Morphotypes of bacteroids in relation to the phylogenetic tree of the IRLC legumes.** The phylogenetic tree is modified from Wojciechowski et al. SEM images of bacteroids are aligned with the phylogenetic positions of the legumes studied here.

Bacteroids in all tested legumes proved to have larger size and DNA content than cultured cells; however, the degree of cell elongation was rather variable in the different species. In addition, the reproductive ability of the bacteroids isolated from these legumes was remarkably reduced (Figures 2 and 3). These results indicate that IRLC legumes provoke terminal differentiation of their endosymbionts with different morphotypes probably with the help of NCR peptides.

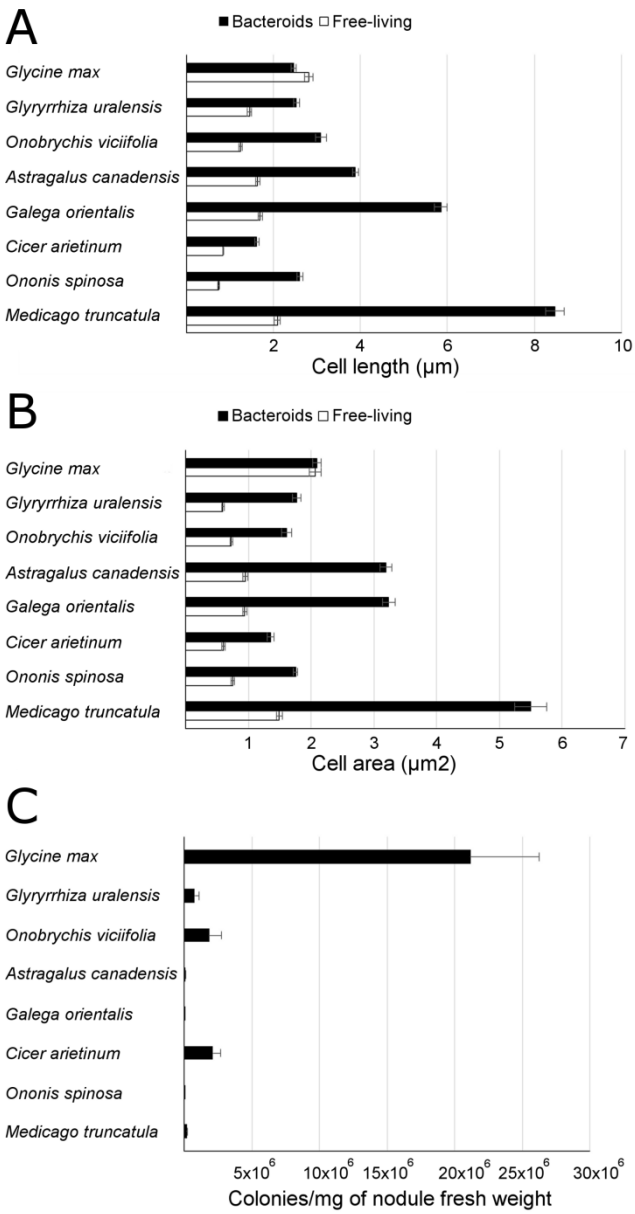


**Figure 2. The size and cell division capacity of the isolated bacteroids.** The length (A) and area (B) of the isolated bacteroids (black bars) were determined from confocal and SEM images (n>60) and were compared to those of free-living bacteria (white bars). (C) The relative number of cells that were able to form colonies was determined by plating the bacterial population isolated from the nodules.

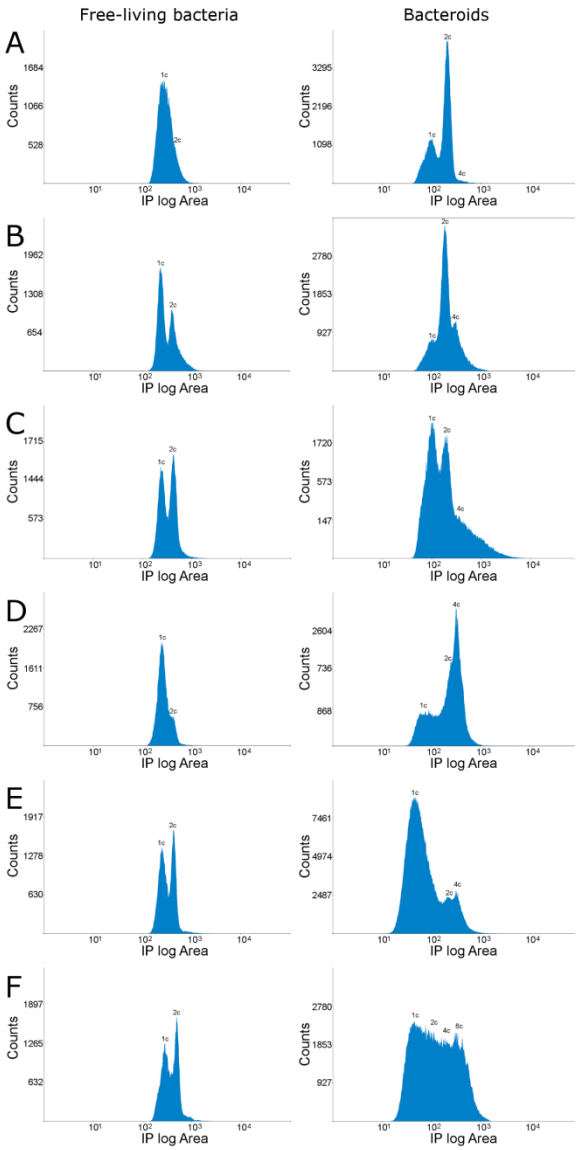


**Figure 3. Genome content of the free-living rhizobia and bacteroids.** DNA content of PI-stained free-living cultured rhizobia used as inoculants and bacteroids isolated from nodules of *Glycyrrhiza uralensis*, *Onobrychis viciifolia*, *Astragalus canadensis*, *Galega orientalis*, *Cicer arietinum* and *Ononis spinosa* (A to F, respectively).

These results were published in the Molecular Plant-Microbe Interactions (Montiel et al., 2016).

In order to assess the number of *NCR* genes in different IRLC legumes, RNA pools of mature nodules from *Glycyrrhiza uralensis*, *Oxytropis lamberti*, *Astragalus canadensis*, *Onobrychis viciifolia*, *Galega orientalis* and *Ononis spinosa* plants were sequenced and the publicly available transcriptome sequencing data from *Pisum sativum*, *M. sativa* and *M. truncatula* were downloaded, and then predicted mRNA consensus sequences were assembled. In addition, our analysis of *Cicer* NCRs was expanded using data from a second genomic sequence data set. These ten legumes

provide an overview of the IRLC, from the most basal species (*G. uralensis*) to members of the three main subclades: Hedysaroid, Astragalean and Vicioid. These species have different bacteroid morphotypes such as S-, E-, SP- and EB-bacteroids. Putative NCRs were identified in all the species analyzed using iterated Blast searches, which revealed highly variable numbers of NCRs *per* species ranging from seven in *G. uralensis* to over 600 in *M. truncatula* (Figure 4). It is possible that the number of NCRs is underestimated, particularly in those species where genome information is not available. Furthermore, in the *M. truncatula* A17 genome in which over 700 *NCR* genes are predicted, expression of only 639 NCRs was detected and only those MtNCRs were included in this analysis.

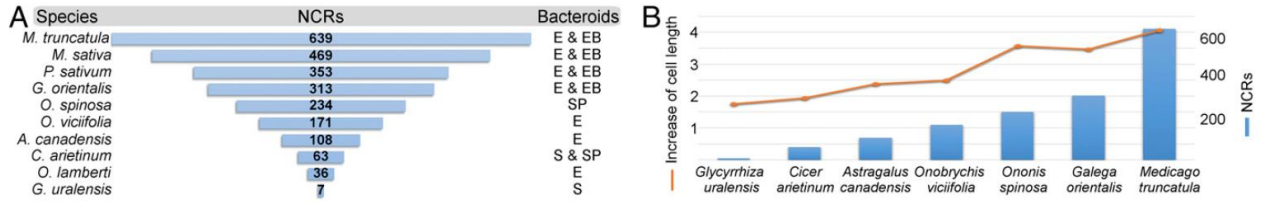


**Figure 4. Numbers of NCR peptides in different IRLC legumes are correlated with bacteroid morphology.** Numbers of NCR peptides predicted from nodule transcriptomes or genome sequences of 10 IRLC legumes are shown in relation to the morphotype of the bacteroids (*A*). There is a positive correlation between average bacteroid length and the size of NCR family (*B*). Pearson Correlation Coefficient: 0.92 (P-value > 0.001).

We showed previously, that bacteroids in all tested IRLC legumes were larger and had more DNA than cultured cells, but the degree of cell elongation was rather variable in different species. Therefore, we tried to correlate the numbers and/or the type of expressed NCRs with the morphotype of bacteroids (Figure 4). *G. uralensis* with 7 NCRs and *C. arietinum* with 63 NCRs produced swollen and swollen/spherical bacteroids, respectively. *O. lamberti* with 36, *A. canadensis* with 108 NCRs and *O. viciifolia* with at least 171 NCRs formed E-bacteroids. The nodules of *O. spinosa* expressed 234 NCRs and hosted SP-bacteroids that were preceded with a brief elongated phase. *G. orientalis*, *P. sativum*, *M. sativa* and *M. truncatula* expressed the most NCRs (313, 353, 469 and 639, respectively) and their nodules contain remarkably elongated and EB-type bacteroids. There was a positive correlation between the degree of bacteroid elongation and the number of the expressed *NCRs* (correlation coefficient 0.92; Figure 4B). These results confirm that NCR peptides have a direct impact on the bacteroid elongation and suggest that the profile of NCRs produced by the legume affects the morphotype of the bacteroids.

Although the amino-acid composition of the individual NCR peptides is highly variable, the average frequency of a given amino acid *per* peptide ("amino acid usage") in the NCR family is very similar in all species investigated (Figure 5A). The average length of the mature peptides is also very similar (i.e. 36-41 amino acids) in these species (Figure 5B). In *A. canadensis* 16 NCR-like peptides were omitted from our analysis because their amino acid sequences were considerably

longer than the typical NCRs, due to an extended lysine-rich region at the N-terminus. Over 30% of the NCR peptides in Vicioid plants (excluding *C. arietinum)* had two potential disulfide bridges, whereas in the other IRLC species, the vast majority of NCRs contained three potential disulfide bridges (Figure 5C).

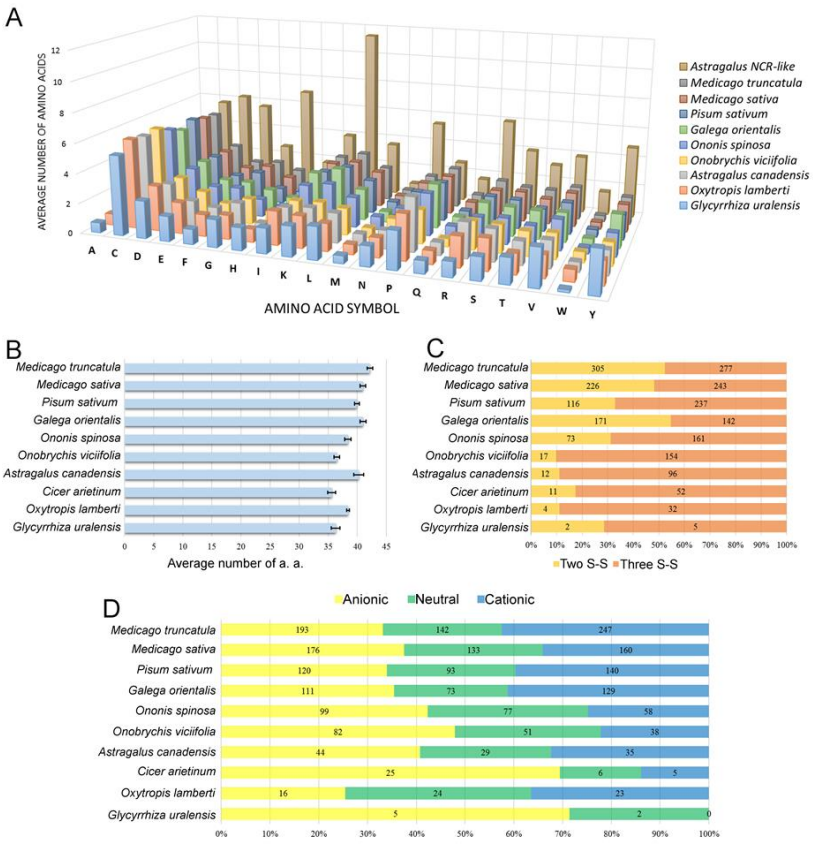


**Figure 5. Analysis of the amino-acid frequencies in NCR peptides from 10 IRLC legumes.** Averages of the amino acid composition (*A*) and lengths of predicted mature peptides (*B*). Proportions (showing actual numbers) of NCRs with two/three disulfide bridges (*C*) and with anionic, neutral and cationic pI charge (*D*) are shown. Standard errors are shown in *B.*

Notable differences were observed in the proportions of anionic, neutral and cationic peptides in the legumes studied: only *G. uralensis* lacked cationic NCRs, while those species hosting EB-bacteroids showed the highest ratio (34-42%) of cationic peptides (Figure 5D). Different reports indicate that the isoelectric point (pI) of NCRs plays a key role in their function. The frequencies of NCRs were plotted relative to their predicted pI values and the data were compared based on those species containing bacteroids with the same morphotype. The legumes with EB-bacteroids had highly similar NCR profile, with two prominent peaks comprising peptides with pI of 4-4.9 and 8-9. The other legumes did not have a conserved NCR profile associated with the distinct morphotypes. They were however characterized by a large set of anionic NCRs, while the contribution of neutral and cationic NCRs was highly variable without a clear pattern (Figure 6).

The expression of NCRs with different charges was analyzed by calculating the sum of RPKM and DEseq values from the nodule transcriptome data and taking into account the pI of the peptides in those legumes where the RNA-seq information was accessible. The ratio of anionic, neutral and cationic NCR transcripts expressed in the nodules was only similar between *G. orientalis, P. sativum* and *M. truncatula*, species with EB-bacteroids. In these legumes, the

expression of cationic NCRs was approximately 40% of the NCR transcripts, mainly from genes encoding peptides with a pI of 8-9. Conversely, in legumes with S-, E- and SP-bacteroids, the highest peak in the profile was constituted by anionic or neutral NCRs but with distinct pI values. The comparison of the nodule transcriptome among different IRLC legumes shows a positive correlation between expression of cationic NCRs and elongation/branching of bacteroids.

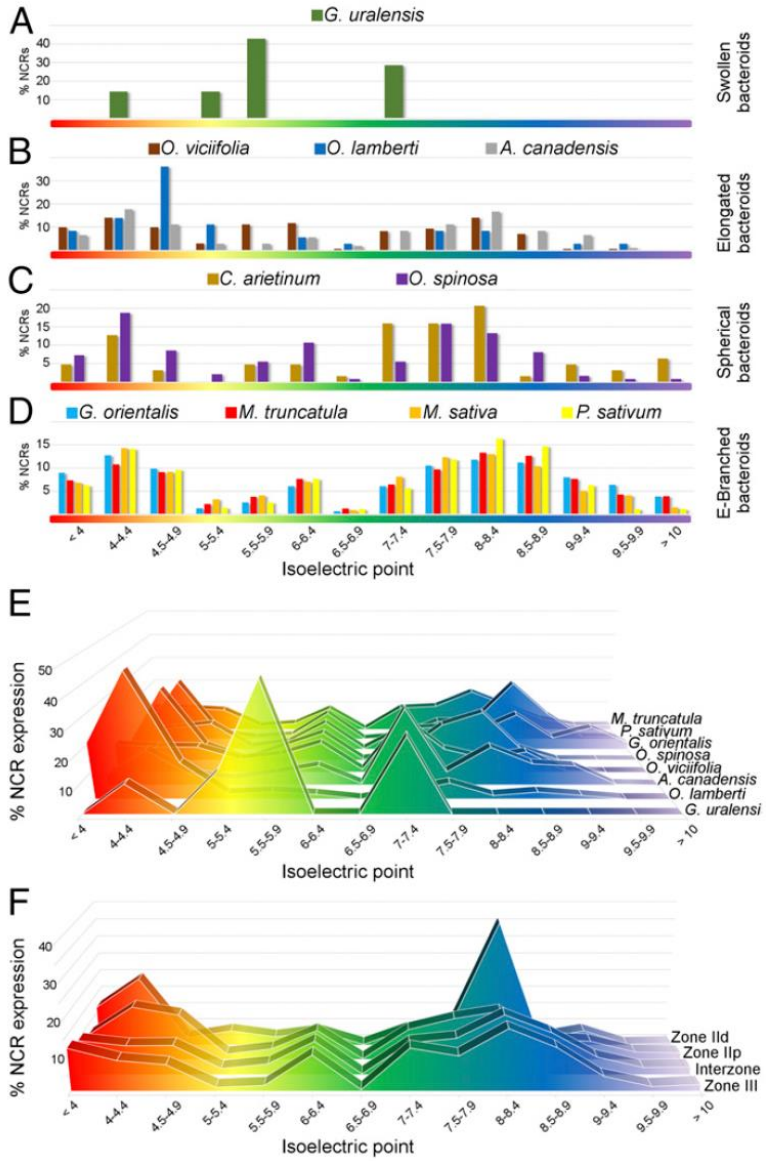


**Figure 6. Isoelectric point profiles of NCR peptides and the relative expression of NCRs with different pIs in the IRLC legumes and nodule zones in M. truncatula.** The frequency of nodule-expressed NCRs with given pI values is shown graphically; grouping of the NCRs from different legume species is based on the S, E, SP, and EB morphotypes of bacteroids (A–D, respectively). The contribution [% RPKM or DEseq normalized values] of different isoelectric point categories to whole-nodule NCR expression for seven IRLC legumes is provided (E). The pattern of NCR expression [%DEseq normalized values] in different zones of M. truncatula nodules is shown in F.

To investigate how the NCR families evolved, we compared the NCR sequences from the different IRLC lineages. Using protein BLAST (BLASTP) alignments, we found that *G. uralensis*, *O. lamberti*, *C. arietinum, A. canadensis* and *O. viciifolia* shared several NCRs with amino acid identities ≥ 80%. In fact, each of the seven *G. uralensis* NCRs had at least one highly similar

homologue in another legume. Similarly, 20 of the 63 *C. arietinum* NCRs had at least one putative orthologue in one of these four legumes, whereas, no NCRs with such high similarity were found in other Vicioid legumes. The Vicioid legumes containing the most numerous NCRs show a very different pattern. Except for two *G. orientalis* NCRs, that display high homology to *G. uralensis* NCRs, all the peptides shared less than 80% identity with peptides produced in other genera. These results suggest common origin and conservation of a few NCRs, coupled with the emergence of many species-specific peptides. To explore this possibility, NCRs from different pairs of legumes were compared by phylogenetic analyses using the mature peptide sequences, aligned and bootstrapped (NJ tree, 1000 trials) by ClustalX2. Putative orthologs could be recognized based on their distribution in closely related species such as *O. lamberti* and *A. canadensis*, and even more so comparing *M. truncatula* and *M. sativa* (Figure 7). The expansion of the NCR families might be driven by local gene duplications followed by diversification as can be observed in the M. truncatula genome where, in agreement with Alunni, we could detect both recent and more ancient duplications.

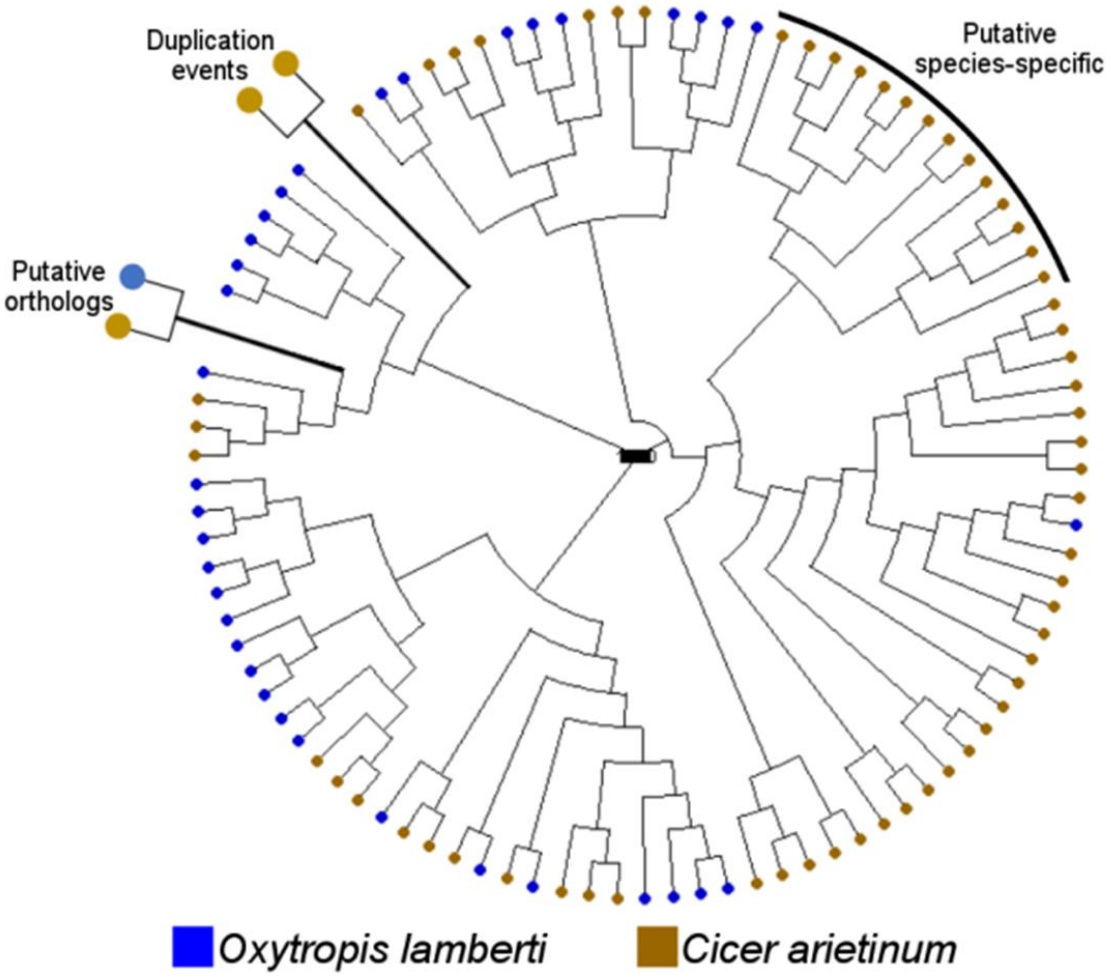


**Figure 7. Pair-wise comparison of NCR families in IRLC legumes through phylogenetic trees.** The circle cladograms illustrate the phylogenetic distribution of two legume NCR families. The members of each NCR family are represented by distinctive colors. Representative examples of putative duplication events, orthologs, and species-specific NCRs are highlighted in the comparison Oxytropis lamberti–Cicer arietinum. The mature peptides were aligned and bootstrapped (NJ tree, 1,000 iterations) by ClustalX 2, and the trees were visualized with Dendroscope.

These results were published in the PNAS USA (Montiel et al., 2017).

*Objective 2: Identification of plant factors in other legumes*

Our collaborators showed that in *Aeschynomene* spp. legumes belonging to the more ancient Dalbergioid lineage, bacteroids are elongated or spherical depending on the Aeschynomene spp. and that these bacteroids are terminally differentiated and polyploid, similar to bacteroids in IRLC legumes. Transcriptome, in situ hybridization, and proteome analyses demonstrated that the symbiotic cells in the Aeschynomene spp. nodules produce a large diversity of NCR-like peptides, which are transported to the bacteroids. Blocking NCR transport by RNA interference-mediated inactivation of the secretory pathway inhibits bacteroid differentiation. Together, these results support the view that bacteroid differentiation in the Dalbergioid clade, which likely evolved independently from the bacteroid differentiation in the IRLC clade, is based on very similar mechanisms used by IRLC legumes.

*Objective 3: Comparative transcriptome analysis of Bradyrhizobium bacteroids with different morphotypes*

The symbiotic interaction that takes place between *Bradyrhizobium* sp. ORS285 bacteria and legume pants belonging to the *Aeschynomene* genus results in the formation of either elongated (E) or spherical (S) terminally differentiated bacteroids within plant cells of the symbiotic nodules. These bacteroids display, in addition to their morphological changes an increased DNA content raising up to 8C (1C being one haploid genome complement) for E bacteroids and 16C for S bacteroids.

We first compared the symbiotic efficiency of the same bacterial strain in two *Aeschynomene* hosts inducing either E or S bacteroids at 14 days post inoculation and found that S bacteroids are more efficient than E ones. These observations are based on the combination of physiological measurements (symbiotic biomass increase, N shoot content, C/N fluxomics).

Thus, we decided to analyze the molecular basis of this increased symbiotic efficiency and investigated the transcriptome of *Bradyrhizobium* sp. ORS285 in culture (rich medium, exponential phase) or in symbiosis with *Aeschynomene indica* (14dpi S bacteroids) or *Aeschynomene afraspera* (14dpi E bacteroids). To do so, nodule/culture RNA was extracted and submitted to serial depletion of plant and bacterial ribosomal RNA using Ribozero kits (Epicentre) and plant mRNA using oligo-dT coated beads. Libraries were constructed using the SOLiD Total RNA-seq kit and they were sequenced on a 5500XL Genetic Analyzer (SOLiD) to produce 50bp single reads. All samples were produced in biological triplicates.

After mapping of the reads on the reference bacterial genome and quantitative analysis of gene expression using DE-seq2, PCA analysis showed a very good homogeneity among replicates and a clear separation of the biological conditions (Figure 8A).

Comparing the bacteroid transcriptome in both hosts with the bacterial culture allowed the identification of ca. 1300 differentially expressed genes (DEG, fdr<0.001 and |log2 fold change|>1.58) that are commonly induced/repressed in bacteroids (Figure 8B). Among the most bacteroid-induced genes are all the genes required for nitrogen fixation and microaerobic respiration (eg. nif and fix genes), together with genes involved in phosphate metabolism and branched amino acid transport. Oppositely, bacteroid-repressed genes comprise housekeeping functions involved in transcription/translation processes, motility and cell cycle progression, which is consistent with bacteroids being immobilized in symbiosomes and a rewiring of their metabolism and cell cycle toward a polyploid state where bacteroid physiology is mostly focused on nitrogen fixation and transfer to the host. As a conclusion on this part, we could define a set of “core bacteroid” genes

regardless of their host environment and bacteroid morphotype. Interestingly the PhyR/EcfG general stress response system (Gourion et al., 2009) is activated in bacteroids, indicating that differentiated bacteroids are in a stressed state *in planta*.


**Figure 8. Analysis of *Bradyrhizobium* sp. ORS285 transcriptome data obtained from free-living and symbiotic cells.** (A) The result of the PCA analysis. (B) Number of genes differentially expressed in E and S morphotype bacteroids as compared to free-living cells.

Comparing the transcriptome of E and S bacteroid identified 510 DEG, which could possibly explain the increased symbiotic efficiency of S bacteroids over E bacteroids. Among the S bacteroid induced genes are a citrate synthase that could account for a possible increased TCA activity. Also the *otsA* gene encoding a trehalose-6-phosphate synthase is induced in S bacteroids. Interestingly the overexpression of this gene in *Rhizobium etli* resulted in an increased symbiotic efficiency on common bean as compared to the wild type strain (Suarez et al., 2006). E bacteroid induced genes were also identified such as the pst/pho genes that are involved in phosphate/phosphonate metabolism.

Finally, to broaden our observations we performed qRT-PCR on nodule RNA extracted from two E-bacteroid-inducing hosts (ie. *A. afraspera* and *A. nilotica*) and two S bacteroid inducing-hosts (ie. *A. indica* and *A. evenia*). Interestingly, gene expression was rather well correlated among similar morphotypes although E bacteroid-specific genes were more variable among E bacteroids ($r^2$=0.72) than S-specific genes among S bacteroids ($r^2$=0.91).

Altogether, this study showed that S bacteroids are symbiotically more efficient than E bacteroids, and the molecular basis of this difference has been investigated by transcriptomics, providing a list of candidate genes for further functional analysis.

A manuscript describing the major results of this study is currently under production (Lamouche et al., in preparation).

*Objective 4: Isolation and characterization of Sinorhizobium meliloti mutants with altered peptide sensitivity*

First, we investigated the mode of action of the antimicrobial NCR peptides. The dose-dependent membrane disruptive effect of the cationic peptides on the outer membrane (OM) was measured with the hydrophobic fluorescent probe 1-N-phenylnaphthylamine (NPN), which cannot enter the intact OM, but can pass the destabilized one. By entering the phospholipid layer, the dye gives rise to strong fluorescence. The control Polymyxin B, which destabilizes the outer membrane structure due to its cationic amino acids, caused quick and strong fluorescence. The cationic NCRs caused also the quick increase of fluorescence, however, its extent was not as high as in the case of

polymyxin B (Figure 9). Based on these results we can conclude that the cationic NCR peptides can destabilize the outer membrane to a certain extent but it is not the primary reason of the death of the treated cells. The anionic or weakly cationic peptides did not show any membrane disruptive activity.

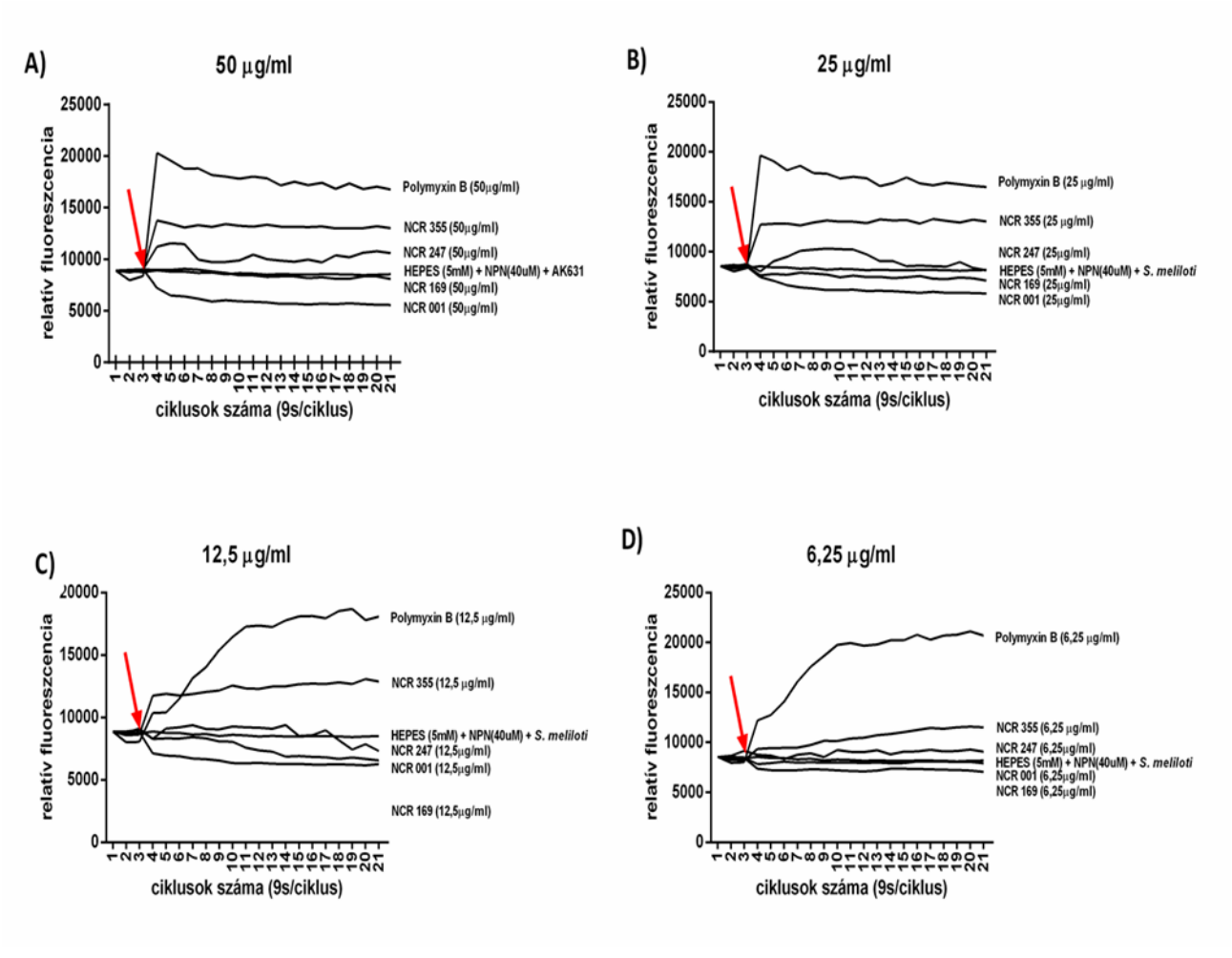


**Figure 9. The concentration dependence of the hydrophobic 1-N-phenylnaphthylamine (NPN) probe.** The outer membrane permeability measured A) at 50 μg/ml; B) at 25 μg/ml; C) at 12,5 μg/ml; and D) at 6,25 μg/ml peptide concentrations.

To investigate the possible damage to the inner membrane (IM) caused by the cationic peptides we took advantage of the fact that the IM is not permeable for the artificial substrate of the cytoplasmic β-galactosidase enzyme. Thus, the enzyme activity can be measured only after the disruption of the membrane. Treatment of S. meliloti cells expressing constitutively the lacZ gene with the cationic peptides resulted in measurable β-galactosidase activity that was 2–20 % of the total enzyme activity obtained after disrupting the cells with SDS and chloroform. The anionic peptides didn't have membrane disruptive features.

The IM damage caused by the cationic peptides was confirmed by detecting the loss of the membrane potential of the peptide treated cells with the help of the fluorescent membrane-potential indicator dye, DiOC2(3), provided in the BacLight™ Bacterial Membrane Potential Kit. Thus we

concluded that the primary reason for the antimicrobial activity of the cationic peptides is the elimination of the membrane potential (Figure 10).

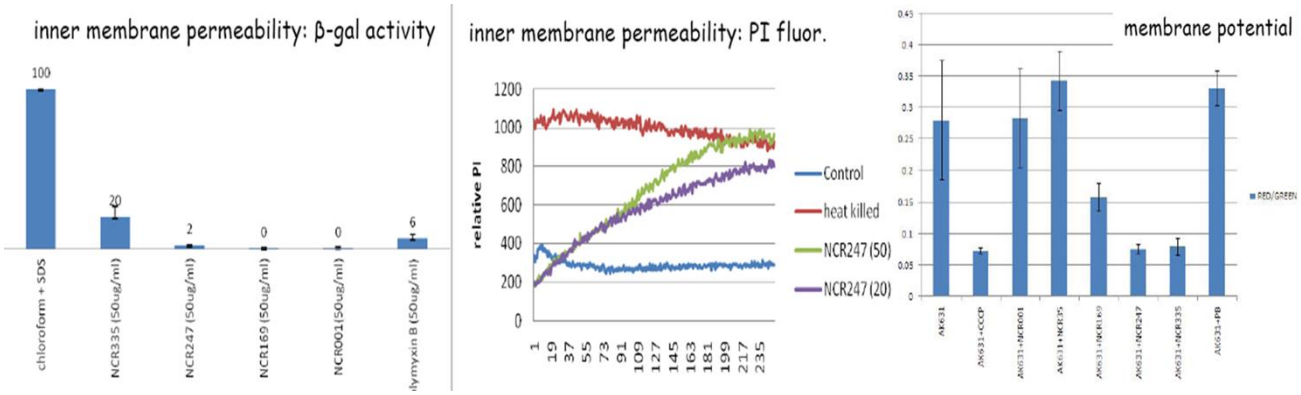


**Figure 10. The effect of NCR peptides on membrane integrity of *S. meliloti.*** Inner membrane permeability measured by β-galactosidase activity at 50 µg/ml peptide concentrations (A) and by the uptake of propidium iodide (B). Membrane potential of the *S. meliloti* cells (C) measured by the red/green fluorescence ration of $DiOC_2(3)$. This dye „exhibits green fluorescence in low concentration in all bacterial cells, however, it accumulates and self-associates in cells that are maintaining a membrane potential resulting in the fluorescence emission to shift from green to red" (Thermo Fisher Scientific).

These results were published in Ann Clin Microbiol Antimicrob (Mikuláss et al., 2016).

We have selected peptide resistant lines from wild-type and *bacA* mutant *S. meliloti* strains with the help of experimental evolution. Bacterial cultures had been grown at a peptide concentration causing 50% growth inhibition for ~100 generations (by changing the medium and replacing the peptide after every 4th generation), then the peptide concentration was increased and the cells were cultured for 100 generations. The peptide concentration was raised similarly three additional times. A single cell from each final culture was grown into a strain, stored and the genome of each strain was sequenced. Bioinformatics analysis revealed 3-16 alteration per genome. Candidate mutations resulting in elevated peptide tolerance were selected based on their multiple appearance and/or multiple hits in certain genes (Figure 11).

| GENE | PUTATIVE FUNCTION | BK 1 | BK 2 | BP 1 | BP 2 | BP 3 | BP 4 | BP 5 | BP 6 | BP 7 | BP 8 | BP 9 | BP 10 | BP 11 | SK 1 | SK 2 | SP 1 | SP 2 | SP 3 | SP 4 | SP 5 | SP 6 | SP 7 | SP 8 | SP 9 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rpoE1 | ECF sigma factor | | | x | | | | | | | | | | | | | | | | | | | | | |
| nusA | transcription anti-terminator | | | x | | x | | | | | | x | | | | | x | x | | | x | x | | | |
| | | | | x | | x | | | | | | x | | | | | x | x | | | x | x | | | |
| SMc01138 | Putative LptB LPS-transporter | | | | | | | | | | | | | | | | | x | | | | | | | |
| | | | | | | x | | | x | | x | | x | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | x | | | | x | x | x | x | |
| SMc00584 | Putative LptF LPS-transporter | | | | | | | | | | | x | | | | | | | | | | | | | |
| | | | | | | | | x | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | x | | | | | |
| rpsJ | 30S ribosomal subunit protein S10 | | | | x | | x | x | | | | | | | | | | | | | | | | | |

**Figure 11. Representative candidate mutations identified in the evolved strains**. BK and SK strains were grown parallel to the evolved wild-type (SK) and bacA mutant (BK) strains, however, without peptides

The mutant genes were reconstructed in the genome of the wild-type and *bacA* mutant strains or were expressed from the *bacA* promoter from a plasmid. Sensitivity of the strains towards the selective and other antimicrobial peptides was determined, which confirmed that these mutations results in elevated peptide tolerance (Figure 12).

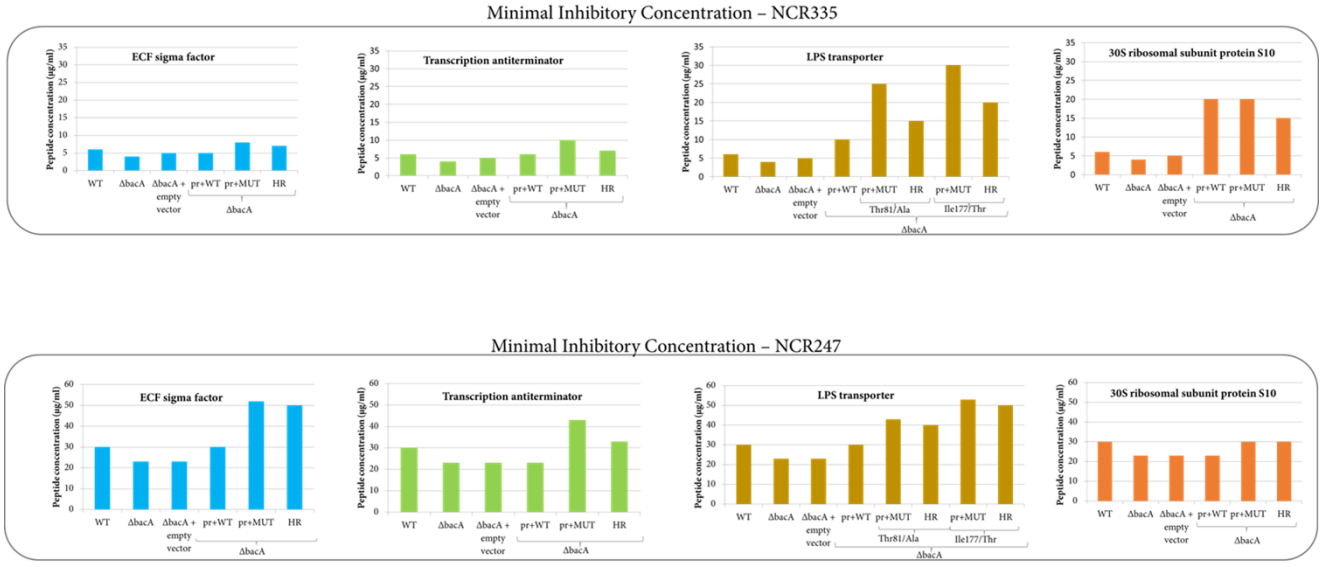


**Figure 12. Peptide sensitivity of the *S. meliloti* strains expressing wild-type and mutant genes.**

The elevated peptide tolerance of the strains expressing the mutant genes could rescue the symbiotic defect of the bacA mutant (Figure 13) leading to the conclusion that the only essential role of BacA in the symbiosomes is to protect rhizobia against the antimicrobial activity of cationic NCRs.

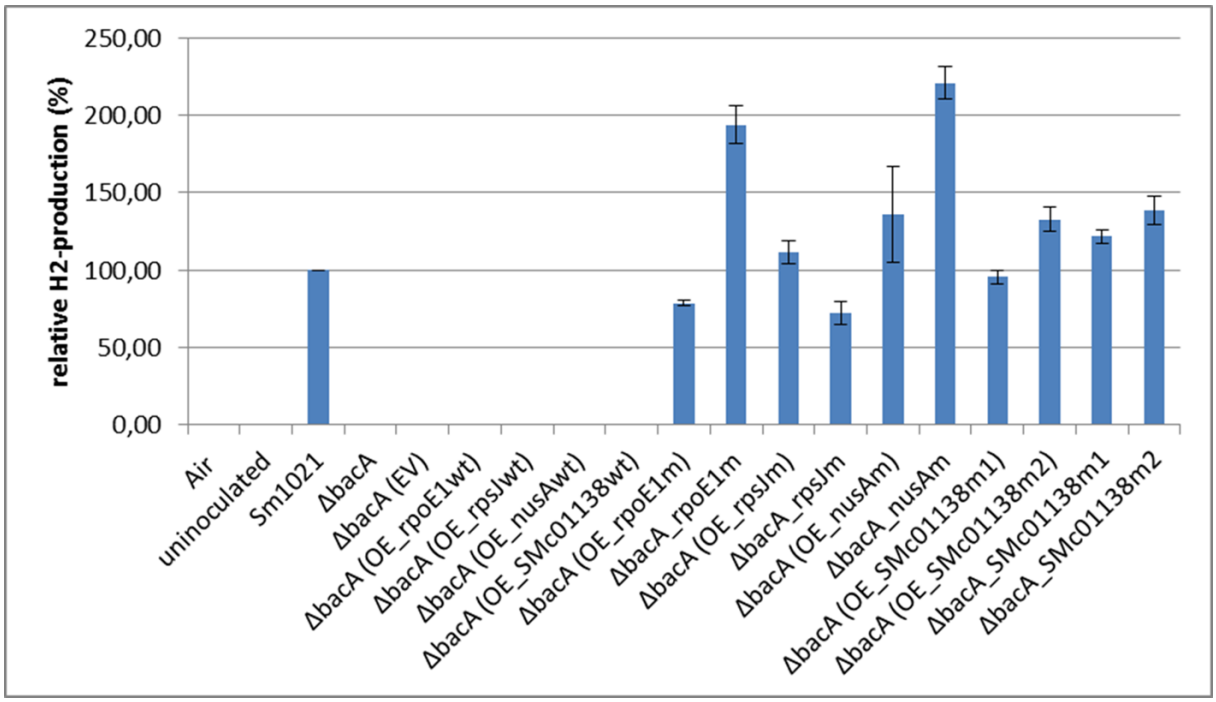


*Figure 13. Relative $H_2$-production as a measure of nitrogenase activity in the nodules induced by bacteria expressing wild-type and mutant genes.*

A manuscript describing these results is in preparation by Kovacs et al.