Bacteroid differentiation in Aeschynomene legumes

Ibtissem Guefrachi

UNIVERSITÉ PARIS SUD

ÉCOLE DOCTORALE 145 :
SCIENCES DU VÉGÉTAL

Laboratoire :
Différenciation Cellulaire lors de la Symbiose Rhizobium-Légumineuse

THÈSE DE DOCTORAT

Discipline: **Biologie**

Présentée et soutenue publiquement par

Ibtissem GUEFRACHI

Le **18/09/2015**

Bacteroid differentiation in *Aeschynomene* legumes

Composition du jury

Directeur de thèse :	**Peter Mergaert**	Directeur de recherche, I2BC-CNRS, Gif-sur-Yvette
Co-directeur de thèse :	**Mohamed Mars**	Professeur, Université de Gabès, Tunisie
Co-directeur de thèse :	**Benoit Alunni**	Maître de conférences, Université Paris sud 11, Orsay
Président :	**Ahmed Landoulsi**	Professeur, Université de Carthage, Tunisie
Rapporteurs :	**Philippe Normand**	Directeur de recherche, Université Claude Bernard, Lyon1
	Rafik Bachoual	Maître de conférences, Université de Gabès, Tunisie
Examinateurs :	**Jacqui Shykoff**	Directeur de recherche, Université Paris sud 11, Orsay
	Mohamed Elarbi Aouani	Professeur, Centre de Biotechnologie Borj Cedria, Tunisie

Soyons reconnaissants aux personnes qui nous donnent du bonheur; elles sont les charmants jardiniers par qui nos âmes sont fleuries. ***Marcel Proust***

Le temps met tout en lumière. ***Thalès***

Résumé

Ce travail de thèse porte sur la thématique des interactions symbiotiques entre les plantes Légumineuses et les bactéries telluriques fixatrices d'azotes appelées rhizobia. Plus particulièrement, je me suis intéressée aux mécanismes qui président la différenciation cellulaire des bactéries au sein des organes symbiotiques formés par la plante.

Contexte général de l'étude

L'azote est un élément fondamental dans la composition de la matière vivante. L'atmosphère est la principale réserve naturelle d'azote (80%) (Foth, 1990). Mais, malgré son abondance, c'est un facteur limitant dans le sol. Cette carence est responsable de la limitation de la productivité végétale dans beaucoup de sols et en particulier dans les sols cultivés. Pour que l'azote atmosphérique puisse être introduit dans le cycle de la matière vivante il faut qu'il soit préalablement fixé. Ainsi, seuls certains procaryotes appelés diazotrophes sont capables de réduire le diazote atmosphérique sous une forme combinée assimilable par les végétaux (Dommergues *et al*., 1999) grâce à l'utilisation d'une métalloenzyme : la nitrogénase. Parmi ces organismes diazotrophes, certaines bactéries Gram négatives collectivement appelées rhizobium, ou des bactéries filamenteuses Gram positives du genre *Frankia* forment des endosymbioses avec respectivement des Légumineuses ou des non Légumineuses du genre *Parasponia* et des plantes actinorhiziennes (Franche *et al*., 2008). Cette symbiose aboutit, dans tous ces cas, à la formation d'un organe spécialisé, la nodosité, qui se forme sur les racines de la plante hôte (dans certains cas aussi sur la tige). Cet organe est le siège d'une fixation de l'azote atmosphérique par les bactéries au bénéfice de la plante. Au cours du développement de la nodosité, les bactéries pénètrent un poil absorbant et sont acheminées *via* un cordon d'infection vers les cellules symbiotiques, où elles sont relarguées par un processus d'endocytose. Les bactéries se retrouvent entourées d'une membrane dérivée de la membrane plasmique végétale et forment des symbiosomes, où elles se différencient en bactéroïdes fixateurs d'azote (Oldroyd et Downie, 2008 ; Desbrosse et Stougaard, 2011). Ces bactéries différenciées ont une physiologie qui est radicalement différente de celle des rhizobia vivant dans la rhizosphère et qui est adaptée au mode de vie symbiotique et à la fixation de l'azote.

La plupart des équipes de recherches travaillant sur la symbiose rhizobium-Légumineuse s'intéressent aux étapes précoces de l'interaction. Ainsi au cours des dernières

décennies, les bases moléculaires de la reconnaissance mutuelle entre les partenaires symbiotiques ont été élucidées (Stacey *et al*., 2006). En conditions de carence azotée, les plantes produisent dans leurs exsudats racinaires des flavonoïdes qui sont reconnus par la bactérie et induisent la production et la sécrétion de lipochitooligosaccharides, les facteurs Nod, qui sont à leur tour reconnus par la plante. La perception des facteurs Nod initie à la fois l'organogenèse de la nodosité et l'infection bactérienne des tissus végétaux (Oldroyd *et al*., 2011).

A contrario, notre équipe a concentré ses efforts de recherche depuis une quinzaine d'années sur des étapes tardives de l'interaction symbiotique et plus particulièrement sur la différenciation des cellules végétales et bactériennes permettant la mise en place du processus de fixation d'azote. Ainsi les cellules végétales qui sont infectées par les rhizobia (plusieurs milliers de bactéries par cellule), appelées cellules symbiotiques, subissent de profondes modifications pour permettre la présence de nombreux symbiosomes fixateurs d'azote dans leur cytoplasme. Durant leur différentiation, ces cellules sortent du cycle cellulaire classique (G1–S–G2–M) pour le convertir en un cycle d'endoréplication (ou endocycle), c'est-à-dire un cycle cellulaire simplifié ne comprenant que deux phases, une interphase G et une phase S (pour Synthèse d'ADN); les phases G2 et M (pour Mitose) ont disparu. La cellule ne se divise pas par la suite en deux cellules filles, ce qui provoque la polyploïdisation de la cellule. Ces endoréplications successives conduisent à augmenter la ploïdie de 2C (cellule diploïde de méristème) à 64C (cellules symbiotiques matures polyploïdes; C étant le contenu en génome haploïde). La polyploïdisation permet d'accroître le volume cellulaire ainsi que l'activité métabolique des cellules hôtes. Un des acteurs clefs de ces endoréduplications successives est la protéine CCS52A dont l'action est la dégradation des cyclines mitotiques avant qu'elles ne puissent provoquer l'entrée en phase M du cycle cellulaire. Cela provoque l'arrêt de la division cellulaire mais n'affecte en rien la réplication de l'ADN, il en résulte des cellules endorépliquées, polyploïdes (Cebolla *et al*., 1999; Vinardell *et al*., 2003 ; Tarayre *et al*., 2004; Fülöp *et al*., 2005).

De façon similaire, les bactéries qui sont internalisées dans les cellules végétales se différencient en bactéroïdes fixateurs d'azote. L'équipe a montré que les bactéries pouvaient prendre plusieurs morphotypes différents (U, pour « *unmodified* » et E pour « *elongated* ») en fonction de l'hôte végétal considéré, indiquant que la plante gouverne la différenciation bactérienne (Mergaert *et al*., 2006). Les bactéroïdes de morphotype E (par exemple chez *Medicago*, *Pisum*, *Vicia*) subissent plusieurs modifications cellulaires et morphologiques

comme un allongement cellulaire couplé à une forte endoréplication du génome (les bactéroïdes deviennent polyploïdes) contribuant ainsi à une augmentation importante de la taille des cellules, ainsi qu'une perméabilité membranaire accrue rendant les bactéroïdes incapables de reprendre un cycle de division normal lorsque l'on remet ces bactéroïdes en culture. Les bactéroïdes de morphotype E subissent donc une différenciation terminale ou irréversible, c'est-à-dire qu'ils perdent toute capacité reproductrice (Mergaert *et al*., 2006; Van de Velde *et al*., 2010).

Cette différenciation terminale des bactéroïdes est typiquement mise en place par toutes les Légumineuses du clade IRLC (pour *Inverted Repeat-Lacking Clade*) mais n'est pas conservée dans d'autres clades de Légumineuses. Dans ce second cas (par exemple chez *Phaseolus, Vigna, Lotus, Glycine*), les bactéries conservent comme dans l'état libre leur morphologie (morphotype U), leur capacité de division dans les symbiosomes et une viabilité importante. Les bactéroïdes de morphotype U subissent donc une différenciation réversible, sans altération du cycle cellulaire ni de perméabilisation de la membrane.

Le processus de différenciation cellulaire et morphologique des bactéroïdes en morphotype E chez *M. truncatula* est contrôlé en partie par une large famille de peptides antimicrobiens (environ 600 gènes chez *M. truncatula*) produits par la plante et spécifiquement exprimée dans les nodosités, appelée NCR (pour *Nodule-specific Cysteine Rich*) (Van de Velde *et al*., 2010). Ces peptides NCR possèdent un peptide signal et des motifs conservés en cystéine. Ils sont adressés aux bactéroïdes et leur adressage est en partie contrôlé par une « signal peptidase complexe », spécifique des nodosités dont DNF1 est une sous unité (*dnf1* pour defective in nitrogen fixation 1 est un mutant de *Medicago truncatula* qui forme des nodosités non fonctionnels) (Wang *et al*., 2010). L'inactivation de *dnf1* bloque l'adressage aux symbiosomes des peptides NCR qui restent bloqués dans le réticulum endoplasmique et les bactéroïdes restent indifférenciés (Van de Velde *et al*., 2010). Localisée au niveau du réticulum endoplasmique, cette « signal peptidase complexe » assurerait le clivage du peptide signal des NCR permettant ensuite leur transport via trafic vésiculaire à la membrane péribactéroïdienne et donc aux symbiosomes (Van de Velde *et al*., 2010 ; Wang *et al*., 2010).

Les peptides NCRs ressemblent à des défensines, des peptides antimicrobiens concourant à l'immunité innée, système immunitaire présent chez tous les Eucaryotes. Ainsi, les peptides NCRs, tout comme les défensines, perméabilisent les membranes bactériennes et présentent une activité antimicrobienne *in vitro*, tuant des bactéries et ayant un spectre d'activité large (Mergaert *et al*., 2003 ; Van de Velde *et al*., 2010).

Ainsi, un élément clef dans la différenciation des bactéroïdes est la protéine bactérienne BacA, un transporteur membranaire initialement découvert chez *S. meliloti* comme requis pour la différenciation des bactéries en bactéroïdes suite à leur relargage intracellulaire (Glazebrook *et al*., 1993). BacA confère une résistance contre l'activité antimicrobienne des peptides NCRs (Haag *et al*., 2011). Un mutant du gène *bacA* chez *S. meliloti* induit la formation de nodosités non-fonctionnelles chez *Medicago,* dans lesquelles les bactéries ne se différencient pas en bactéroïdes et meurent instantanément après avoir été confrontées aux peptides NCRs produits par les cellules hôtes (Haag *et al*., 2011).

Le gène *bacA* de *S. meliloti* 1021 code une protéine de 381 acides aminés. Selon les prédictions bioinformatiques, cette protéine qui possède 8 domaines transmembranaires, serait semblable à des transporteurs membranaires de type ABC (Glazebrook, *et al*., 1993). Les transporteurs ABC permettent soit l'import ou l'export d'une grande variété de substrats à travers les membranes biologiques. La structure canonique d'un transporteur ABC est composée de quatre domaines : deux domaines hydrophobes transmembranaires et deux domaines liant et hydrolysant l'ATP. La protéine BacA ne possède cependant pas la cassette ATPase.

En culture, la mutation *bacA* augmente la résistance de *S. meliloti* aux peptides antimicrobiens ayant une cible intracellulaire, tels que la bléomycine, la microcine B17, et Bac7, un peptide eucaryote riche en proline (Ichige, *et al*., 1997 ; Marlow, *et al*., 2009). Ce phénotype suggère que l'un des rôles de BacA serait d'internaliser les peptides antimicrobiens. L'étude du mutant *bacA* de *S. meliloti* révèle aussi une forte sensibilité aux agents inducteurs de stress membranaire tels que le SDS, l'éthanol et le déoxycholate (Ferguson, *et al*., 2002). Chez *S. meliloti*, BacA est également impliquée dans la synthèse de lipide A, un composant essentiel des lipopolysaccharides qui constituent la membrane externe des bactéries à gram négatif (Ferguson *et al*., 2002; Ferguson *et al*., 2004). Des analyses en similarité de séquences ont suggéré que BacA serait un transporteur d'acides gras à longue chaine qui une fois dans le périplasme assurerait la maturation des lipides A. Cette maturation permettrait de masquer les LPS pour la cellule hôte. En effet, le lipide A est un puissant antigène chez les Mammifères, ce qui faciliterait la persistance de la bactérie prolongeant ainsi l'infection (Ferguson *et al*., 2004). La perte de BacA s'accompagne d'une baisse des contenus en VLCFA (*very-long-chain fatty acids*) des lipides A, expliquant la défection dans l'intégrité membranaire (Ferguson, *et al*., 2004). Cependant, ces modifications des lipides A ne sont pas indispensables à la réalisation d'une infection chronique (Ferguson, *et al*., 2005) indiquant qu'il s'agît probablement de la perte de la fonction de transport de peptides qui est

la cause de l'incapacité à se maintenir dans les nodosités.

Dans les nodosités, la protéine BacA de *S. meliloti* semble avoir un rôle prépondérant dans la protection des bactéries contre l'activité antimicrobienne des NCRs, permettant ainsi le maintien des bactéries après l'infection. Il est probable que la fonction de BacA soit de procéder à l'internalisation des NCRs afin de les enlever de leur site de toxicité, la membrane. En accord avec ce modèle, il a été constaté que le gène *bacA* chez les rhizobia n'est nécessaire pour la symbiose que pour la nodulation des Légumineuses produisant des NCRs (Maruya et Saeki, 2010 ; Haag *et al*., 2011).

De nombreuses études ont apportés des résultats similaires chez d'autres espèces, comme *Mycobacterium tuberculosis* (Domenech, *et al*., 2009), *Salmonella enterica* (Pränting *et al*., 2008), *Mesorhizobium huakuii* (Tan *et al*., 2009), *Brucella abortus* (Levier *et al*., 2000), chez lesquelles des homologues de BacA ont été caractérisés. Ces études mettent en exergue le rôle principal de BacA dans l'internalisation de peptides antimicrobiens, et donc la résistance face à ces derniers.

Premier volet de la thèse : L'expression des gènes *NCR*s chez *Medicago*

C'est dans le contexte décrit ici en haut que j'ai débuté ma thèse. Dans un premier volet de mon travail de thèse, je me suis intéressée à l'étude de la spécificité de l'expression des gènes *NCR* chez la Légumineuse *M. truncatula* en tirant profit de la base de données MtGEA (*Medicago truncatula Gene Expression Atlas*), qui regroupe des profils transcriptomiques obtenus par hybridation de puces à ADN Affymetrix (50 900 sondes). Cette base de données regroupe des expériences réalisées sur différents organes (feuilles, racines, fleurs, fruits, nodosités…) et en réponse à de nombreux stimuli (hormones, stress biotiques et abiotiques, interactions symbiotiques…). Ainsi, nous avons pu analyser l'expression de 334 gènes *NCR* dans 267 différentes conditions expérimentales. Nous avons également généré des plantes transgéniques portant des fusions transcriptionnelles pNCR-*GUS* pour trois NCRs représentant différentes classes temporelles d'expression. Ces lignées transgéniques ont été utilisées pour analyser l'activité des promoteurs (révélant ainsi le patron d'expression génique des *NCR* considérés) au cours d'une cinétique d'infection par *S. meliloti* ou en réponse à l'inoculation par différents agents pathogènes.

Nous avons trouvé que l'ensemble des environ 300 gènes *NCR* testés (sauf cinq) n'est exprimé que dans les nodosités, ils ne sont pas exprimés dans d'autres organes de la plante, ni

lors d'une infection par des agents pathogènes. De plus l'expression des NCR n'est induite en réponse à aucune interaction biotique ou abiotique testée.

Dans les nodosités, les NCR ne sont pas encore exprimés à des stades précoces du développement, avant que les cellules symbiotiques ne soient formées et que les rhizobia ne soient libérés dans les symbiosomes des cellules hôtes. Ils ne sont pas impliqués dans la dégradation des bactéroïdes pendant la sénescence des nodosités et leur expression s'arrête lorsque la sénescence est lancée.

Cependant, le profil d'expression des NCR par des vagues successives au cours de la formation de nodosités suggère que les bactéroïdes sont les seules cibles des peptides et qu'un ensemble de peptides pourrait être impliqué dans la différenciation des bactéroïdes et d'autres dans leur fonctionnement.

Ces analyses expérimentales, ensemble avec des calculs d'entropie de Shannon, montrent que les gènes *NCRs* ont une spécificité d'expression exceptionnelle indiquant qu'ils subissent une régulation extrêmement stricte.

L'ensemble de ces résultats a mené à une publication dans la revue BMC Genomics (voir **chapitre I** des résultats) dont je suis co-premier auteur. Cette étude montre que l'expression des NCR est soumise à une régulation stricte et qu'ils sont activés pendant l'organogenèse et au cours du développement nodulaire dans les cellules symbiotiques polyploïdes.

Ce travail débouche sur de nombreuses pistes de recherche. Ainsi il serait intéressant d'identifier les bases moléculaires de cette régulation génique. Ce contrôle très strict est-il uniquement dû à l'action de répresseurs/activateurs transcriptionnels ? Ou est-ce qu'un contrôle épigénétique est impliqué, possiblement par le biais de la méthylation de l'ADN ou selon les marques d'histones présentes dans les régions génomiques portant des gènes *NCR* ?

L'endoréplication des cellules végétales était une étape majeure dans l'activation des gènes de nodulation. La plupart des gènes nodosité-spécifiques ne sont pas exprimés dans les autres tissus. Ces gènes sont donc maintenus dans un état inactif pendant tout le développement végétal et l'inactivation transcriptionnelle est levée pendant le développement des nodosités. Cette régulation pourrait dépendre des processus épigénétiques résultant des modifications de la structure chromatinienne (modifications post-traductionnelle des histones et méthylation de l'ADN) pendant la formation des nodosités. Au laboratoire, nous avons proposé que l'endoréplication, pendant la différenciation cellulaire dans les nodosités, fasse partie des mécanismes qui lèvent l'inactivation transcriptionnelle des gènes spécifiques des nodosités, ceci résultant de modifications des codes épigénétiques au niveau de la chromatine.

Afin de vérifier le rôle de l'endoréplication dans l'activation du transcriptome, une approche intéressante, actuellement poursuivie au laboratoire, consiste à trier par cytométrie en flux des noyaux, en fonction de leur contenu nucléaire pour ensuite faire chez chacun de ces types nucléaires l'analyse du transcriptome. Cette analyse est couplée avec celle de la structure de la chromatine (méthylation d'ADN et modifications des histones) au niveau des loci génomiques portant les gènes spécifiques des nodosités.

Deuxième volet de la thèse : La différenciation des bactéroïdes chez *Aeschynomene*

Les mécanismes de contrôle par la plante sur les rhizobia intracellulaires demeurent à ce jour peu connus et le seul modèle étudié, au début de ce travail de thèse, restait l'interaction entre *M. truncatula* et *S. meliloti*. Les études sur ce modèle ont notamment permis de montrer que la plante exerce un contrôle sur la bactérie via les peptides NCRs (Van de Velde *et al*., 2010). Cependant, ce mécanisme a été supposé spécifique au clade IRLC, un clade de Légumineuses induisant des morphotypes E et absente chez des Légumineuses induisant des morphotypes U (Alunni *et al*., 2007) car la présence de NCR n'a pas été à ce jour révélée chez d'autres espèces de Légumineuses.

On peut donc s'interroger sur **l'existence possible d'autres clades de Légumineuses dans lesquels les bactéroïdes se différencient ? Et si oui, est ce que la différenciation repose sur les mêmes mécanismes que ceux observés chez les IRLC auxquels appartient *Medicago*?**

Je me suis donc intéressée à la symbiose de certaines Légumineuses tropicales du genre *Aeschynomene* appartenant au clade des Dalbergoïdes. Les *Aeschynomene* sont étroitement liées à l'arachide, deuxième culture de Légumineuse cultivée dans le monde et donc ce modèle est très pertinent pour l'agronomie. Les *Aeschynomene* se caractérisent par des propriétés symbiotiques très originales. Tout d'abord, la nodulation peut avoir lieu aussi bien sur les racines que sur les tiges conférant à ces plantes un fort potentiel de fixation d'azote (Giraud *et al*., 2000 et 2004). De façon surprenante, certaines espèces d'*Aeschynomene* ne nécessitent pas la synthèse de facteur Nod par la bactérie pour l'induction de la nodulation (Giraud *et al*., 2007), alors que ces molécules étaient jusqu'alors considérées comme des signaux ayant un rôle clé et essentiel dans la mise en place de la nodulation chez tous les légumineuses. De plus, les symbiotes, appartenant au genre *Bradyrhizobium*, ont également des caractéristiques inhabituelles pour des rhizobia, comme leur capacité de

photosynthèse et de fixation d'azote en culture.

Le genre d'*Aeschynomene* apparaît également comme un modèle de choix pour l'étude des mécanismes de différenciation des bactéroïdes. En effet, il peut être observé chez les *Aeschynomene* des différences morphologiques drastiques pour la même bactérie selon l'espèce de la plante hôte. Par exemple, les bactéroïdes de la souche *Bradyrhizobium* sp. ORS285, microsymbionte d'*Aeschynomene* présentent chez *Aeschynomene afraspera* une morphologie en bâtonnet allongé. Le morphotype allongé (morphotype E) observé chez *A. afraspera* ressemble à celui de *S. meliloti* en symbiose avec *M. truncatula*. (Bonaldi *et al*., 2011). Récemment, un nouveau morphotype de bactéroïde a été identifié chez *Aeschynomene indica*. Ces bactéroïdes sont sphériques (morphotype S) et peuvent être induits par la souche ORS285.

De plus, la souche *Bradyrhizobium japonicum* USDA110 (une souche non-photosynthétique et non capable de fixer l'azote en état libre mais seulement en condition de symbiose), le microsymbionte de soja, est capable de noduler efficacement *A. afraspera* (Renier *et al*., 2011).

Le deuxième objectif de ma thèse a donc consisté à comprendre **les mécanismes de la différenciation bactérienne au sein des nodosités chez les *Aeschynomene*.**

❖ Différenciation terminale des bactéroïdes chez *Aeschynomene*

Le processus de différenciation des bactéries lors de la symbiose avec les Légumineuses peut être plus ou moins drastique. Dans le cas de la symbiose entre *M. truncatula* et *S. meliloti*, cette différenciation est dite 'terminale'. Le fait que la même bactérie (la souche ORS285) puisse avoir deux morphotypes différents en fonction de la plante infectée souligne que cette différenciation est sous le contrôle de la plante. C'est sur la validation de cette hypothèse que nous avons collaboré avec l'équipe d'Eric Giraud pour caractériser en premier lieu l'état de différenciation des bactéroïdes dans les nodosités induites par la souche *Bradyrhizobium* sp. ORS285 et de décrypter les facteurs de la plante qui induisent cette différenciation.

Le travail a pu démontrer que les bactéroïdes aussi bien allongés que sphériques présentent un niveau d'endoréplication quatre à huit fois supérieur à celui des bactéries en vie libre (7C chez *A. afraspera* et 16C chez *A. indica*). Une perte d'intégrité de la membrane et un taux de viabilité extrêmement faible ont également été observés. En dehors de leur morphologie, les bactéroïdes présentent donc des propriétés très similaires chez les deux

plantes mais également semblables à celles de *S. meliloti* en symbiose avec *M. truncatula*. Elle*s* sont à différenciation terminale et polyploïdes.

Ces ressemblances soulèvent une question : Quels sont les effecteurs de la plante forçant les bactéroïdes à se différencier en bloquant les divisions et en favorisant l'endoréplication ?

❖ Identification de gènes *NCR*s chez *Aeschynomene*

Les mécanismes de différenciation pourraient être conservés chez les trois plantes (*A. indica*, *A. afraspera* et *Medicago*). Le contrôle de cette étape, chez les *Aeschynomene*, pourrait également impliquer des NCR. En accord avec cette hypothèse, une étude a été menée par l'équipe d'Eric Giraud pour identifier des facteurs de la plante qui imposent une différenciation morphologique des bactéroïdes par des analyses transcriptomiques (banque EST) dans les nodosités d'*A. afraspera* et *A. indica*. Cette approche a permis de mettre en évidence l'existence d'une famille de gènes *NCR-like* spécifiquement et fortement exprimés dans les nodosités des deux plantes au moment de la différenciation des bactéroïdes (qui a lieu de façon synchrone entre le 4ème et le 5ème jour après inoculation).

Les NCRs d'*Aeschynomene* sont semblables structuralement aux peptides NCR de *Medicago* avec un peptide signal en N-terminal et une séquence mature en C-terminal comprenant des cystéines en positions conservées. Toutefois les NCRs de *Medicago* et d'*Aeschynomene* présentent des différences quant à la position des cystéines et vis-à-vis du peptide signal, suggérant qu'ils ont une origine évolutive différente.

On peut donc émettre l'hypothèse que ces NCR pourraient agir sur la différenciation des bactéroïdes chez *Aeschynomene*. Afin d'évaluer cette hypothèse, une approche par ARN interférant a été réalisée visant à empêcher l'adressage des NCR au bactéroïdes. En effet, l'interrogation du transcriptome d'*A. evenia* a révélé l'existence d'un orthologue de DNF1 et une construction permettant l'atténuation de ce gène par ARNi a été réalisée et introduite *in planta* via *Agrobacterium rhizogenes* (une bactérie tellurique phytopathogène, induisant l'émergence et la prolifération de chevelu racinaire (*hairy root*) à partir du point d'infection de la bactérie). Les résultats obtenus ont montré un défaut de différenciation des bactéroïdes. Elles présentent soit de différents morphotypes au sein de la même cellule soit des bactéroïdes allongées au lieu des sphériques. Parallèlement, une analyse de la localisation spatiale de leurs expressions chez *A. indica* et *A. afraspera* a été effectuée par une approche d'hybridation *in situ* (HIS). Une technique qui permet la détection et la localisation d'ARN messagers sur

coupes histologiques de tissus grâce à l'utilisation de sondes ARN complémentaires marquées. Les résultats de l'HIS ont permis d'affirmer que les gènes *NCR* d'*Aeschynomene* sont bien exprimés au niveau des cellules symbiotiques contenant les bactéroïdes. Finalement, une approche de protéomique par spectrométrie de masse a montré que les NCRs sont adressés aux bactéroïdes.

L'ensemble des résultats de cette étude confortent l'hypothèse selon laquelle les mécanismes de différenciation des bactéroïdes en morphotypes E et S chez les *Aeschynomene* sont semblables à ceux identifiés chez *M. truncatula*. Il semble donc qu'au cours de l'évolution, deux clades de Légumineuses relativement éloignés (IRLC et Dalbergoïdes) aient convergé vers l'utilisation de peptides de l'immunité innée afin de contrôler leur symbionte bactérien et d'en tirer un bénéfice maximal au cours de l'interaction symbiotique.

Ces données ont été valorisées dans un article dont je suis co-auteure : **Convergent evolution of endosymbiont differentiation in Dalbergoid and IRLC legumes mediated by nodule-specific cysteine-rich peptides,** soumis à Plant Physiology (voir **chapitre II** des résultats).

Dans le mesure où les NCR-like chez *Aeschynomene* ressemblent aux peptides antimicrobiens NCR de *Medicago*, l'étape suivante a consisté en la recherche de facteurs bactériens qui permettraient aux rhizobactéries de survire dans les symbiosomes en présence des NCR. **Quels sont alors les facteurs bactériens qui déterminent la différenciation bactérienne chez les *Aeschynomene ?***

- **Le transporteur BclA de *Bradyrhizobium***

Le gène *bacA* de *S. meliloti*, auparavant caractérisé dans le laboratoire, est nécessaire pour la réponse de la bactérie aux peptides NCR de *Medicago*. Un gène ressemblant fortement au gène *bacA* a été identifié par analyse *in silico* chez les bradyrhizobia et il a été nommé *bclA* pour *bacA*-like.

Mon objectif étant d'étudier le rôle de ce gène au cours de l'interaction symbiotique *Aeschynomene-Bradyrhizobium* : En particulier je me suis intéressée à la différenciation et l'endoreplication des bactéroïdes, à la survie *in planta* ainsi qu'au réponse face aux peptides antimicrobiens. Cette nouvelle étude a constitué le cœur de ma thèse et sera présentée sous forme de deux publications (qui sont prêts à être soumise) dont je suis le premier auteur (**chapitre III** des résultats).

Lors du crible sur plante d'une banque de mutants insertionnels de la bactérie *Bradyrhizobium* sp. ORS278 (insertion aléatoire du transposon mini-Tn5), un mutant dans un gène *bclA* formant des nodosités non-fonctionnelles chez *A. indica* a été identifié (Bonaldi *et al*., 2010). A cause du spectre d'hôte limité de cette bactérie, le mutant *bclA* dans la souche ORS278 ne permet d'étudier le rôle de ce gène que chez *A. evenia* ou *A. indica* formant ainsi des bactéroïdes du morphotype S. Afin d'étendre cette étude aux morphotypes E et U, la construction par génétique inverse de mutants de délétion du gène *bclA* chez ces deux souches de *Bradyrhizobium* ORS285 et USDA110 a été réalisée. Cela nous a permis d'étudier l'implication du gène *bclA* dans la survie des bactéroïdes au sein des nodosités, et de déterminer son rôle dans la résistance aux peptides antimicrobiens.

Pour répondre à ces questions, la capacité du mutant *bclA*, de chaque *Bradyrhizobium* étudié, à infecter et à se maintenir dans les nodosités de soja et des *Aeschynomene*, a été étudiée. En parallèle, des tests de sensibilité aux peptides antimicrobiens NCR, à la bléomycine et à Bac7, ont été réalisés afin d'évaluer le rôle de BclA dans la réponse de la bactérie aux divers peptides antimicrobiens.

Pour ce faire, les phénotypes symbiotiques du mutant ORS285Δ*bclA* (la structure des nodosités, la différenciation des bactéroïdes et la fonctionnalité de nodosités) ont été analysés sur les plantes hôtes *A. indica* et *A. evenia* (bactéroïdes du morphotype S) et *A. afraspera* (bactéroïdes du morphotype E). Dans les différents systèmes symbiotiques testés, des nodosités non-fonctionnelles ont été formées. Cependant, les cellules de ces nodosités sont infectées normalement par les bactéries, mais celles-ci ne se différencient pas en bactéroïdes allongées ou sphériques et restent inchangées et meurent ensuite rapidement. Cet état de non-différenciation a été confirmé par des mesures de la ploïdie des bactéroïdes, elle est de l'ordre de 3C chez *A. indica* et *A. afraspera*. Il semble alors qu'il y a une corrélation entre la présence du gène *bclA* et la formation des bactéroïdes polyploïdes. Par ailleurs, on a pu démontrer que le gène *bclA* de *Bradyrhizobium* peut complémenter le phénotype symbiotique du gène *bacA* chez *S. meliloti* avec une restauration de la différenciation des bactéroïdes dans les nodosités de *Medicago*, bien que cette complémentation n'est que partielle et insuffisante pour restaurer la capacité de la souche à fixer l'azote.

Le gène *bacA* chez *Sinorhizobium fredii*, *Mesorhizobium loti* ou *Rhizobium leguminosarum* bv. *phaseoli* n'est pas nécessaire pour la formation de bactéroïdes du morphotype U (Karunakaran *et al*., 2010; Maruya et Saeki; 2010 Ardissone *et al*., 2011). Nous avons donc testé la nécessité du gène *bclA* de *Bradyrhizobium* en cas de formation de bactéroïdes du morphotype U. *B. japonicum* USDA110 induit des nodosités fonctionnelle

chez le soja et aussi chez *A. afraspera*. Chez le soja les bactéroïdes sont du morphotype U tandis que chez *A. afraspera* les bactéroïdes sont très légèrement allongées et ne sont quasiment pas polyploïdes (3C). Cette dernière observation était inattendue, vu la production des NCRs par *A. afraspera*. Effectivement, nous avons montré par qPCR qu'aussi dans les nodosités d'*A. afraspera* infectées par USDA110, les gènes *NCR*s sont normalement exprimés. Possiblement des facteurs génétiques de la souche USDA110 rendent la bactérie insensible au peptide NCRs.

L'analyse des phénotypes symbiotiques du mutant *B. japonicum* USDA110Δ*bclA* chez *A. afraspera* et le soja a montré que le gène *bclA* de *B. japonicum* n'est pas nécessaire pour la formation de bactéroïdes non différenciées. Le contenu en ADN des bactéroïdes est alors très similaire à celui des bactéries en culture.

❖ BclA est un transporteur de NCRs et autres peptides antimicrobiens

Bien que BacA et BclA ont un rôle clé au cours du processus symbiotique, ces deux protéines n'ont pas une fonction exclusivement symbiotique, car des gènes homologues sont conservées chez d'autres bactéries, inclus des agents pathogènes de Mammifères comme *Brucella abortus*, *Mycobacterium tuberculosis* et *Escherichia coli* (le gène homologue de *bacA* est alors nommé *sbmA*) et s'avèrent cruciaux pour la persistance de la bactérie dans son hôte et donc pour le succès de l'infection (Ferguson *et al*., 2004). BclA est un transporteur de type ABC qui diffère de ses homologues BacA (*S. meliloti*) et SbmA (*E. coli*) par la présence d'un domaine ATPase, absent chez ces derniers.

Il a été démontré, chez *S. meliloti et E. coli* que le gène *bacA*/*sbmA* affecte leur sensibilité à des peptides antimicrobiens qui agissent sur l'enveloppe bactérienne telles que les défensines, à d'autres peptides antimicrobiens qui agissent à l'intérieur des cellules tels que la bléomycine (un antibiotique qui inhibe la transcription et la réplication de l'ADN) et Bac7 (peptide riche en proline, se lie aux ribosomes et inhibe la traduction). L'internalisation de la bléomycine est en majorité dépendante du mécanisme lié à BacA (Wehmeier *et al*., 2010). La résistance à la bléomycine observée chez les mutants *bacA* s'explique alors par l'absence du transporteur BacA, qui ne l'internalise plus, limitant sa toxicité. Les orthologues de *bacA* chez *E. coli* et *S. meliloti* sont requis pour l'internalisation de Bac7 (Mattiuzzo *et al*., 2007; Marlow *et al*., 2009), ce qui provoque une sensibilité à ce peptide. C'est pourquoi, lorsque les gènes *bacA* de ces bactéries sont mutés, il y a une diminution de la sensibilité à Bac7.

Donc la question se pose si BclA est capable de transporter la bleomycine, Bac7 comme ses homologues BacA et SbmA. Ainsi, des tests de sensibilité *in vitro* ont été effectués et ont montré que la présence de BclA augmente significativement la sensibilité des bactéries à ces agents antimicrobiens confirmant que BclA peut importer ces peptides. En plus, nous avons montré que BclA augmente la résistance à l'activité antimicrobienne des NCRs comme la protéine BacA. Finalement, nous avons utilisé un peptide NCR modifié avec un groupe fluorescent, combiné avec une analyse par cytométrie en flux pour montrer que BclA ainsi que BacA peut promouvoir l'import de ce NCR démontrant ainsi que BclA et BacA sont des transporteurs de peptides NCRs.

Conclusions générales

Ainsi des résultats originaux ont été obtenus sur le mode de différenciation des bactéroïdes chez les *Aeschynomene*. Leurs bactéroïdes sont allongés ou sphériques selon la plante hôte, ils présentent une différenciation terminale et sont polyploïdes comme les bactéroïdes des Légumineuses appartenant aux IRLC. Les analyses transcriptomiques, hybridation *in situ* et protéomiques (Czernic *et al*., 2015) ont démontré que les cellules symbiotiques dans les nodosités d'*Aeschynomene* produisent des peptides NCR, qui sont transportés vers les bactéroïdes. Le blocage du transport des NCR dans la voie de sécrétion par ARNi (extinction de l'orthologue de *DNF1*) inhibe la différenciation des bactéroïdes. Ces résultats suggèrent que la différenciation des bactéroïdes dans le clade des Dalbergoïdes, qui a probablement évolué indépendamment de la différenciation des bactéroïdes dans le clade des IRLC, repose sur des mécanismes très similaires (Czernic *et al*., 2015).

Les connaissances acquises au cours des dernières années sur le rôle des NCR dans la différenciation des bactéroïdes chez *M. truncatula*/*S. meliloti* offre de nombreuses perspectives de travail sur les NCRs identifiés chez les *Aeschynomene*. De plus, la technique de transformation par *hairy root* permet d'envisager une approche par gain de fonction (Bonaldi *et al*., 2010). Il serait par exemple possible de tester la capacité des NCRs d'induire la différenciation des bactéroïdes chez le soja ou des NCRs d'*A. indica* à induire un morphotype S chez *A. afraspera*.

En outre, nous avons montré que *Bradyrhizobium* ORS278 et ORS285 nécessitent le transporteur de peptides BclA à la fois pour la différenciation des bactéroïdes de morphotypes E et S. BclA possède une fonction similaire à la protéine BacA de *S. meliloti* qui est nécessaire pour la différenciation des bactéroïdes chez *Medicago* et qui joue aussi un rôle

dans le transport des peptides antimicrobiens, les NCR, dans les cellules bactériennes. BclA peut aussi remplacer fonctionnellement BacA dans le transport des peptides. Ces observations confortent fortement la conclusion que chez les *Aeschynomene*, la différenciation des bactéroïdes en morphotype E ou S est sous le contrôle de la famille des peptides NCRs.

Ces mécanismes de différenciation des bactéries symbiotiques ne sont étonnement pas limités aux couples Légumineuses-rhizobia. En effet, certains insectes, comme le puceron *Acyrthosiphon pisum*, effectuent aussi une symbiose avec des bactéries qui se différencient en adoptant un morphotype S. Des peptides antimicrobiens produits par l'insecte, et possédant des motifs conservés en cystéine ainsi qu'un signal peptide ont été identifiés et nommés BCR pour *Bacteriocyte-specific Cysteine Rich peptides* (Shigenobu et Stern, 2012). Par ailleurs, chez un charançon ravageur de céréales (*Sitophilus zeamais*), des peptides antimicrobiens ont été montrés comme induisant la différenciation de bactéries symbiotiques en morphotype E (Login *et al*., 2011). Ces mécanismes de différenciation, imposés par un hôte eucaryote à un symbionte procaryote par l'intermédiaire de peptides antimicrobiens dérivés de l'immunité innée, sont apparus indépendamment à plusieurs reprises chez les plantes et les animaux, suggérant que ce type de stratégie confère un avantage à l'hôte eucaryote.

Les NCR sont des peptides cationiques qui assurent une activité inhibitrice de la croissance microbienne et plusieurs études montrent que les peptides cationiques sont des effecteurs de réponses immunitaires innées. Ceci suggère que les NCR pourraient être des candidats prometteurs pour la mise en place de stratégies thérapeutiques innovantes.

L'étude de *Sinorhizobium meliloti* et de ses hôtes végétaux est un sujet de recherche relativement bien développé. En effet, outre le fait que *S. meliloti* est le symbionte de la luzerne, *Medicago sativa*, et de la Légumineuse modèle *Medicago truncatula*, c'est aussi une α-protéobactérie très proche de bactéries pathogènes comme *Agrobacterium* et autres. L'étude du processus d'infection de *M. truncatula* par *S. meliloti* pourrait fournir des gènes candidats pour l'étude des mécanismes pathogènes de ces dernières et de développer de nouvelles stratégies thérapeutiques.

Outre cette connaissance fondamentale des gènes et acteurs moléculaires impliqués dans le mécanisme de différenciation, il demeure important de comprendre si ce processus confère un véritable avantage fonctionnel à la plante. Une étude récente suggère que ce processus de différenciation morphologique ne serait pas un caractère ancestral mais serait apparu à plusieurs reprises de façon indépendante dans la famille des Légumineuses (Oono *et al*., 2010). Une autre étude qui compare l'efficacité symbiotique de différents couples rhizobium-Légumineuse dans lesquels la même bactérie subit ou non une différenciation

montre une meilleure performance symbiotique pour les bactéroïdes différenciés morphologiquement (Oono et Denison, 2010). Ces études suggèrent ainsi que ce processus de différenciation morphologique constituerait un avantage évolutif pour le partenaire végétal.

Différentes hypothèses ont été avancées pour expliquer cette amélioration des performances symbiotiques : **i**) l'endoréplication du génome de la bactérie permettrait d'assurer une activité métabolique supérieure comme cela est observé chez les plantes et ainsi une plus forte activité de la nitrogénase (Kondorosi *et al*., 2000), **ii**) la fragilisation des parois bactériennes permettrait à la plante de récupérer plus efficacement les nutriments investis dans le développement de la biomasse bactérienne lors de la sénescence de la nodosité (Vasse *et al*., 1990), et **iii**) cette même fragilisation permettrait de favoriser les échanges nutritionnels entre les deux partenaires.

Nos observations, conduites chez deux espèces d'*Aeschynomene* (*A. indica* et *A. afraspera*) en utilisant la même bactérie (ORS285), suggèrent qu'il peut y avoir aussi des différences de performance entre des bactéroïdes différenciés morphologiquement. La comparaison des résultats de fixation de l'azote et de cytométrie en flux indique que le niveau d'endoréplication des bactéroïdes est corrélé à leur efficacité symbiotique. En effet les bactéroïdes de morphotype U (*B. japonicum* USDA110/soja), qui n'ont pas effectué d'endoréplication, fixent moins bien l'azote (fixation normalisée par masse de nodosités) que des bactéroïdes de morphotype E (*B. japonicum* USDA110-*Bradyrhizobium* sp. ORS285/*A. afraspera*). Ces bactéroïdes endorépliqués (7C) fixent eux-mêmes moins bien l'azote que les bactéroïdes de morphotype S (*Bradyrhizobium* sp. ORS285/*A. indica*) qui ont effectués davantage d'endoréplication (16C).

Les raisons de cet avantage ne sont pas connues, mais nous pouvons supposer que la forme sphérique des bactéroïdes pourrait permettre, compte tenu de l'encombrement minimal assuré par cette forme, un meilleur remplissage des cellules végétales et ainsi d'améliorer l'efficacité de la fixation d'azote des nodosités (Wadisirisuk et Weaver, 1985). En accord avec cette hypothèse, des mesures du nombre de bactéroïdes (comptage sur cellule de Malassez) par unité de biomasse nodulaire montre que la densité en bactéroïdes est 25% supérieure dans le cas d'*A. indica*. Il est par ailleurs possible, comme le suggère l'analyse en cytométrie en flux que le niveau d'endoréplication soit plus important sous cette conformation, et que le métabolisme bactérien soit plus actif dans cet état. Il est donc tentant de suggérer que plus l'endoréplication subie par les bactéroïdes est importante, plus ils sont efficaces pour la fixation de l'azote.

Ces résultats préliminaires sont très intéressants et méritent d'être approfondis en plus grand détail. En outre, il sera très intéressant de faire des analyses transcriptomiques, protéomiques et métabolomiques sur les trois morphotypes (U, E et S) afin de comprendre les bases moléculaires et métaboliques de cette différence d'efficacité entre les différentes formes de bactéroïdes. L'intégration des analyses transcriptomiques, protéomiques et métaboliques permettra d'identifier des gènes et des voies métaboliques qui sont spécifiques à un ou plusieurs morphotypes de bactéroïdes. Les résultats qui découleraient d'une telle étude devraient permettre d'améliorer plus largement la compréhension des mécanismes contrôlant et régulant la différenciation des bactéroïdes lors de la mise en place de la nodosité ainsi que l'impact sur la fixation symbiotique d'azote.

Enfin, dans le cas où des déterminants majeurs de la différenciation des bactéroïdes seraient identifiés, l'obtention de souches bactériennes, dans lesquelles ces déterminants moléculaires sont modifiés, pourrait permettre, à plus long terme, de créer des systèmes symbiotiques plus efficaces chez des espèces d'intérêt agronomique majeur, tel que le soja. De la même manière, on pourrait introduire des gènes NCRs chez des plantes qui n'en possèdent pas afin d'induire la différenciation des bactéroïdes et ainsi d'améliorer l'efficacité de la fixation d'azote.

Table of contents

List of publications and communications

1. Journal papers

1.1. **Guefrachi, I**., Pierre, O., Bourge, M., Timchenko, T., Alunni, B., Barrière, Q., Czernic, P., Villaécija-Aguilar, J.A., Verly, C., Fardoux, J., Mars, M., Kondorosi, E., Giraud, E., and Mergaert, P. (2015a). *Bradyrhizobium* BclA is a NCR peptide transporter required for bacterial differentiation in symbiosis with *Aeschynomene*. Molecular Plant-Microbe Interactions **(Chapter III; In press)**.

1.2. **Guefrachi, I**., Pierre, O., Crespo, J.C., Timchenko, T., Alunni, B., Mars, M., Giraud, E., Vinardell, J.M., and Mergaert, P. (2015b). The *Bradyrhizobium japonicum* BclA NCR transporter is not required for symbiosis with soybean **(Chapter III; In preparation)**.

1.3. Czernic, P., Gully, D., Cartieaux, F., Moulin, L., **Guefrachi, I**., Patrel, D., Pierre, O., Fardoux J., Chaintreuil, C., Nguyen, P., Gressent, F., Da Silva, C., Poulain, J., Wincker, P., Rofidal, V., Hem, S., Arrighi, J.F., Mergaert, P and Giraud, E. (2015). Convergent evolution of endosymbiont differentiation in Dalbergoid and IRLC legumes mediated by nodule-specific cysteine-rich peptides **(Chapter II; Submitted)**.

1.4. **Guefrachi, I**., Nagymihaly, M., Pislariu, C.I., Van de Velde, W., Ratet, P., Mars, M., Udvardi, M.K., Kondorosi, E., Mergaert, P., Alunni, B. (2014). Extreme specificity of *NCR* gene expression in *Medicago truncatula*. BMC Genomics. 15: 712 (**Chapter I**).

2. Book Chapter

2.1. **Guefrachi, I**., Verly, C., Kondorosi, E., Alunni, B., and Mergaert, P. (2013). Role of the bacterial BacA ABC-transporter in chronic infection of nodule cells by *Rhizobium* bacteria. In: **"Biological Nitrogen Fixation"**, F.J. de Bruijn (ed.). Wiley-Blackwell **(In press).**

3. International and National Conferences

3.1. Participation at **the symposium Frontiers in legume symbiosis, in Gif-sur-Yvette, France, on December 11-12, 2014.**

3.2. Benjamin Gourion, Fathi Berrabah, Marie Bourcy, Alexis Eschstruth, Anne Cayrel, **Ibtissem Guefrachi**, Peter Mergaert, Jiangqi Wen, Viviane. Jean, Kirankumar. S. Mysore, and Pascal Ratet: Symbiotic supression of immunity, the plant, the rhizobia and the environment. **The XVI International Congress on Molecular Plant-Microbe Interactions in Rhodes, Greece, on July 6-10, 2014.**

3.3. Ibtissem Guefrachi, Camille Verly, Benoit Alunni, Fabienne Cartieaux, Pierre Czernic, Djamel Gully, Mohamed Mars, Eva Kondorosi, Eric Giraud, and Peter Mergaert: Role of BacA-NCR system in terminal bacteroid differentiation: lessons from the *Aeschynomene/Bradyrhizobium* symbiosis. **The 4th International Symposium of Antimicrobial Peptides in Lorient, France, on June 4-6, 2014.**

3.4. Ibtissem Guefrachi, Olivier Pierre, Fabienne Cartieaux, Camille Verly, Benoit Alunni, Pierre Czernic, Djamel Gully, Mohamed Mars, Eva Kondorosi, Eric Giraud, and Peter Mergaert: Differentiation of the endosymbiotic bacteria in the *Rhizobium*-legume symbiosis. **The 11th Meeting of Plants-Bacteria in Aussois, France, on February 3-7, 2014.**

3.5. Marianna Nagymihaly, **Ibtissem Guefrachi,** Quentin Michel, Favery Bruno, Peter Mergaert, and Benoit Alunni: Bang-bang control over *NCR* genes expression: Extreme tissue specificity is correlated with drastic repression/induction swithes. **The 11th Meeting of Plants-Bacteria in Aussois, France, on February 3-7, 2014.**

3.6. Ibtissem Guefrachi, Felix Kraus, Benoit Alunni, Mohamed Mars, Eva Kondorosi, Eric Giraud and Peter Mergaert: Expression of the *bacA-like* gene of *Bradyrhizobium* strain ORS285 in nodules of *Aeschynomene indica*, **the PhD days, in Orsay, France, on October 24-25, 2013.**

3.7. Ibtissem Guefrachi, Fabienne Cartieaux, Camille Verly, Benoit Alunni, Pierre Czernic, Djamel Gully, Eva Kondorosi, Eric Giraud, and Peter Mergaert: Controlling endosymbiotic bacteria with antimicrobial peptides. **The 18th International Congress on Nitrogen Fixation in Miyazaki, Japan, on October 14-18, 2013.**

3.8. **Ibtissem Guefrachi**, Fabienne Cartieaux, Camille Verly, Benoit Alunni, Pierre Czernic, Djamel Gully, Mohamed Mars, Eva Kondorosi, Eric Giraud, and Peter Mergaert: Bacteroid formation in *Aeschynomene*. **The 3rd Meeting on Molecular Mechanisms in Nitrogen Fixing Root Endosymbiosis, in Montpellier, France, on September 26-27, 2013.**

3.9. **Ibtissem Guefrachi**, Fabienne Cartieaux, Camille Verly, Benoit Alunni, Pierre Czernic, Djamel Gully, Mohamed Mars, Eva Kondorosi, Eric Giraud, and Peter Mergaert: Bacteroid formation in *Aeschynomene*. **The Internal Symposium of the Institute of Plant Science, in Gif-sur-Yvette, France, on September 12-13, 2013.**

3.10. Participation at **the Symposium Frontiers in Legume Symbiosis, in Gif-sur-Yvette, France, on December 13-14, 2012.**

3.11. **Ibtissem Guefrachi**, Mikhail Baloban, Mohamed Mars, Eva Kondorosi, Eric Giraud and Peter Mergaert: Bacteroid differentiation in the *Rhizobium*-legume symbiosis, **the PhD days, in Orsay, France, on October 17-19, 2012.**

3.12. **Ibtissem Guefrachi**, Mikhail Baloban, Eva Kondorosi, Peter Mergaert, The NCR peptides in legume nodules: effectors for the control of the endosymbiotic *Rhizobium* bacteria. **The 7th International Symbiosis Society Congress, The Erarth's Vast Symbiosphere, in Kraków, Poland, on July 22-28, 2012.**

3.13. Participation in **the Internal Symposium of the Institute of Plant Science, in Gif-sur-Yvette, France, on February 13-14, 2012.**

3.14. Sami Dhaoui, Mosbah Mahdhi, **Ibtissem Guefrachi**, Amira Fterich and Mohamed Mars: Study of genetic diversity of rhizobia nodulating both legumes: *Lens culinaris* and *Lathyrus sativus* in arid areas of Tunisia. **The 7th Scientific days of Microbiology, in Hammamet, Tunisia, on November 25-27, 2011.**

3.15. **Ibtissem Guefrachi**, Mikhail Baloban, Mohamed Mars, Eva Kondorosi, Eric Giraud and Peter Mergaert: Why bacteroids in some legumes are morphologically differentiated. **Symposium of Molecular Mechanisms in Nitrogen Fixing Endosymbioses, in Toulouse, France, on October 17-18, 2011.**

List of figures

List of tables

List of abbreviations

AFLP: Amplified Fragment-Length Polymorphism

AMPs: Antimicrobial peptides

APC: Anaphase Promoting Complex

ARA: Acetylene Reduction Assay

BAP: Cytokinin 6-benzylaminopurine

BCIP: 5-Bromo-4-Chloro-3-Indolyl-phosphate

BCR: Bacteriocytes-specific Cysteine Rich peptides

BF: Bright Field

Boid: Bacteroids

CCD: Charge-Coupled Device

CCRs: Crypt-specific Cysteine-Rich peptides

ColA: Coleoptericin-A

CS: Cell Suspension

DAPI: DNA fluorescent stain 4',6-diamidino-2-phenylindole

DIC: Differential Interference Contrast image

DIG: Digoxigenin

DNF: Defective in Nitrogen Fixation

Dpi: Days post inoculation

E_g: Entropy values

ESR: Extracytoplasmic Stress Response

EST: Expressed Sequence Tag

FISH: Fluorescent *In Situ* Hybridization

FITC: Fluorescein isothiocyanate

GRPs: Glycine Rich Peptides

IRLC: Inverted Repeat-lacking clade

LB: Luria Bertani

LCM: Laser-Capture Microdissection

LC-MS: Liquid Chromatography coupled to tandem Mass Spectrometry

LPS: Lipopolysaccharide

LRR: Leucine Rich Repeat domain

LysM: Lysin Motif

MEGA: Molecular Evolutionary Genetics Analysis

MeV: Multi Experiment Viewer

MH: Mueller-Hinton

MJ: Methyl Jasmonate

MtGEA: The *Medicago truncatula* Gene Expression Atlas

NAA: Auxin1-naphthaleneacetic Acid

NB: Nucleotide Binding domain

NBT: 4-Nitro Blue Tetrazolium chloride

NCBI: National Center for Biotechnology Information

NCR: Nodule-specific Cysteine-Rich

NPA: Naphthylphthalamic Acid

nsMyc: Non-sulfated Myc factor

ORF: Open Reading Frame

PDA: Potato Dextrose Agar

PHB: polyhydroxybutyrate

Rpm: Rotation per minute

sMyc: Sulfated Myc factor

SNARP: Small Nodulin Acidic RNA-binding Protein

SNARPs: Small Nodulin Acidic RNA-binding Proteins

SPC: Signal Peptidase Complex

Theor. pI : Theorical isoelectric point

TIBA: Triidobenzoid Acid

TRGs: Taxonomically Restricted Genes

Trx: Thioredoxins

TTSS: Type Three Secretion System

VLCFA: Very Long Chain Fatty acids

YE: Yeast Elicitor

YM: Yeast Mannitol

Introduction

Part of this introduction will be published as a review in the book entitled "Biological Nitrogen Fixation Book", edited by Frans J. de Bruijn to be published by John Wiley & Sons:

Role of the bacterial BacA ABC-transporter in chronic infection of nodule cells by Rhizobium bacteria

Ibtissem Guefrachi, Camille Verly, Eva Kondorosi, Benoît Alunni and Peter Mergaert

1. The *Rhizobium*-legume symbiosis

3.1. Nitrogen fixation: background and ecological and agronomic stakes

The element Nitrogen (N), or "Azote" meaning in the French language "without life" as the chimist Antoine Lavoisier called it about 200 years ago, has proved to be anything but lifeless, since it is a component of food, poisons, fertilizers, and explosives. And above all, nitrogen is one of the six basic elements found in living organisms, next to hydrogen, oxygen, carbon, phosphorous and calcium. The element N is highly abundant on the earth's surface where it is found almost entirely as nitrogen gas (N_2) in the atmosphere. N_2 is chemically extremely stable (because of the triple bond between the two N atoms) but biologically unusable for most organisms (breaking the triple bond needed for the building of new molecules is extremely difficult and energy costing). Living organisms prefer reduced nitrogen, ammonium (NH_4^+), or oxidized nitrogen, nitrate (NO_3^-), for incorporation in biomolecules. Therefore, usable nitrogen is limited for biological production in the biosphere as well as in agriculture.

The chemical transformations of the element N on earth into its many oxidation states are called the nitrogen cycle. The major transformation processes of N are nitrogen fixation, nitrification, denitrification, anammox, and ammonification (**Fig. 1**). These processes are highly dependent on the activities of a diverse assemblage of microorganisms, bacteria, archaea, and fungi, although physical processes such as lightning also contribute to some extent to the nitrogen cycle and produce ammonium from N_2. Among these transformation processes, nitrogen fixation plays a particularly important role for life on earth because it is the process that introduces N into the biosphere and compensates for the loss of N resulting from denitrification and anammox (**Fig. 1**). Without nitrogen fixation, the biosphere would be depleted from N and thus the process is critical for the existence of all life on earth.

The natural balance in the global nitrogen cycle is today seriously perturbed by human activity. The human impact on the global nitrogen cycle via the industrial production of nitrogenous fertilizer with the Haber-Bosch process is close to the level of the natural input ($130x10^{12}$ g of N per year costing $7x10^{10}$ $ - a cost that has tripled in less than a decade and requiring 1 to 2 % of the world's annual energy supply). The negative consequences of these human nitrogen additions are alarming and management of the nitrogen cycle is urgently needed (Gruber and Galloway, 2008). The drawbacks of man-made chemical nitrogen reduction are the pollution of surface-water by nitrates and an important impact on the

greenhouse effect, which is related to the high energy cost for production, transportation and application of fertilizer and the emission of the powerful greenhouse gas nitrous oxide or N_2O by fertilized soils.

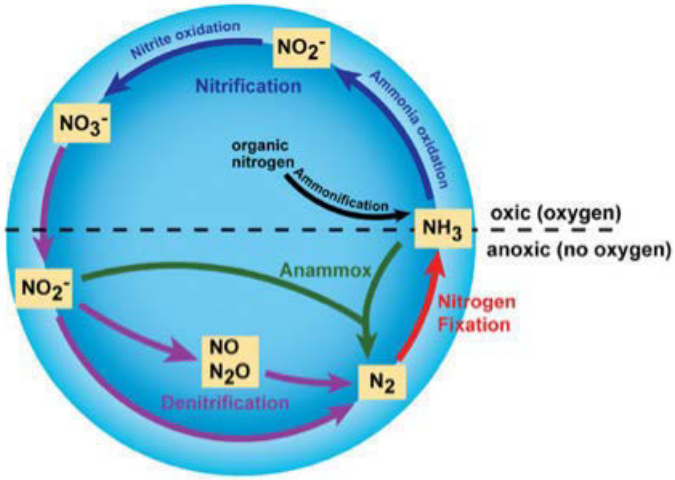

Figure 1. The global nitrogen cycle

The major transformations in the nitrogen cycle are shown. These transformations are for the largest part the result of microbial activities. Nitrogen fixation (red arrow) is the enzymatic reduction of N_2 gas in the atmosphere to ammonium. A small but significant amount of ammonia is also formed from N_2 by lightening. Ammonification (black arrow) also produces ammonia via the decomposition by fungi and bacteria of excreted waste or of dead organisms. Nitrification (blue arrow), the conversion of ammonium to nitrate is the result of the activity of two groups of bacteria, known as the ammonia-oxidizers and the nitrite-oxidizing bacteria, which generate energy from these oxidations. Nitrification is carried out under aerobic conditions. Ammonia oxidation occurring under anoxic conditions is carried out by a completely different group of bacteria and is called Anammox (green arrows). Denitrification (purple arrow) is the result of the reduction of nitrate and other oxidized forms of nitrogen by bacteria that use these N-compounds as electron acceptors in an alternative anaerobic respiration. The denitrification process results in the production of N_2 and its release in the atmosphere, thereby closing the nitrogen cycle. Figure taken from Nature Education. (http://www.nature.com/scitable/knowledge/library/the-nitrogen-cycle-processes-players-and-human-15644632).

Only a select group of prokaryotic microorganisms is able to perform the energetically extremely costly enzymatic nitrogen fixation. Certain *Eubacteriae* and *Archaea* possess the nitrogenase enzyme complex capable to reduce N_2 into the biologically useful form NH_4^+ (Young, 1992). Nitrogenase enzymes are related in all nitrogen fixers (Zehr *et al.,* 2003). Many bacteria and archaea are free-living nitrogen fixing organisms reducing nitrogen solely for their own growth but some bacteria are symbiotic nitrogen fixers, reducing nitrogen not only for their own subsistence but also contributing to the nitrogen supply of a eukaryotic symbiotic partner. Most notorious symbiotic nitrogen fixers are rhizobia, which fix nitrogen within legume plants and *Frankia* inside actinorhizal plants. But also animals establish associations with nitrogen fixing bacteria. Interesting examples are nitrogen fixing bacteria living in the gut of termites (Ceja-Navarro *et al*., 2013) or in the fungus gardens of leaf-cutter ants (Pinto-Tomas *et al*., 2009).

Nitrogen fixing microbes are nitrogen autotrophs while all other organisms, including plants and animals, require uptake of oxidized or reduced nitrogen from nutrition. Leguminous plants (Fabaceae) and actinorhizal plants are an exception since they are able to fix atmospheric nitrogen (N_2) and thus to grow in the absence of available nitrogen in the soil or of nitrogen-containing chemical fertilizers in agricultural settings. Plants of the legume family have solved their nitrogen nutrition needs through the symbiotic association with nitrogen fixing rhizobium bacteria. Rhizobia are a polyphyletic group of soil bacteria belonging to the α- or β-proteobacteria (**Fig. 2**; Sawada *et al*., 2003; Masson-Boivin *et al*., 2009). Actinorhizal plants on the other hand associate with a completely different group of bacteria, the *Frankia*, which are gram-positive actinobacteria and from a monophyletic clade (Nouioui *et al*., 2011).

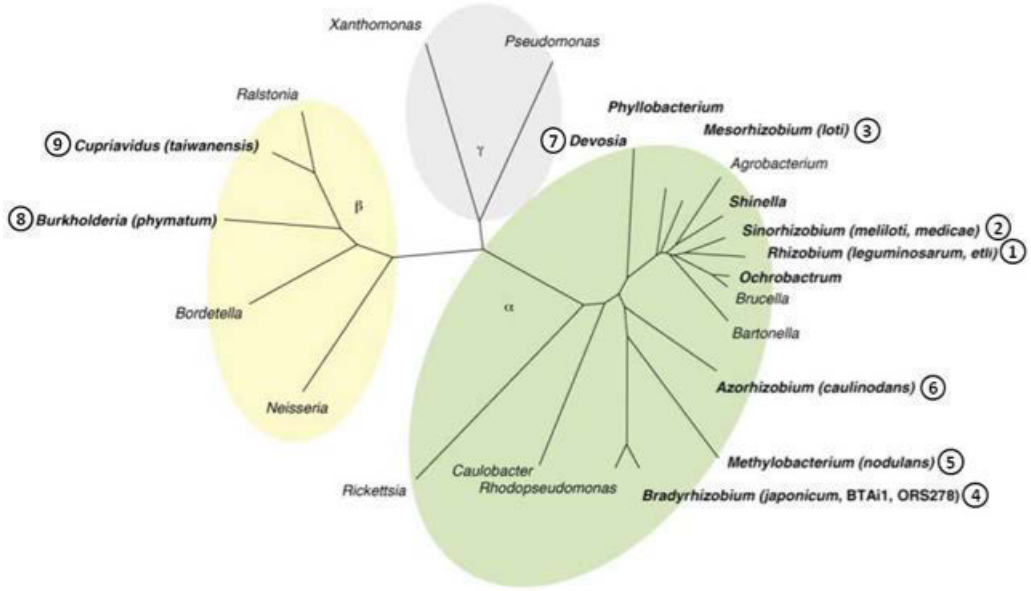

Figure 2. Phylogeny of rhizobia

Unrooted phylogenetic tree of 16S rDNA sequences from selected α-, ß- and δ- proteobacteria. The numbers correspond to clades 1-9 are defined by Sawada *et al*. (2003). Genera in bold font contain rhizobia. Rhizobial species or strains whose genomes have been fully sequenced are indicated in parentheses. (Sawada *et al*., 2003; Masson-Boivin *et al*., 2009).

For the purpose of the symbiosis, legumes and actinorhizal plant hosts form new specific organs on its roots, called nodules inside which the symbiotic rhizobia or *Frankia* are housed (**Fig. 3D-G**), fix nitrogen and transfer the ammonium to the plant. This highly improves the plant's nitrogen nutrition and makes them independent from external nitrogen supplies.

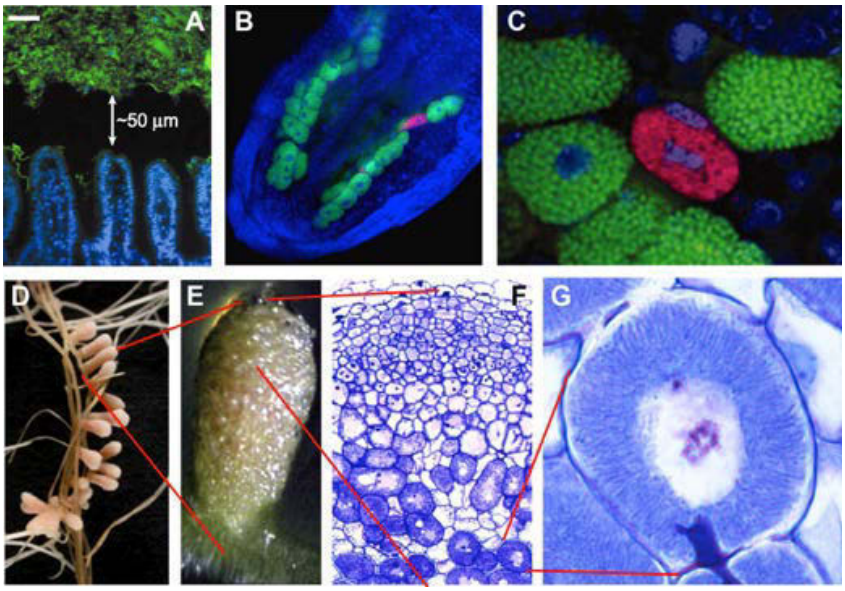

Figure 3. The house of symbiosis: symbiotic niches in animals and plants

(A) The mammal intestine is loaded with 10 to 100 trillion of bacteria belonging to thousands of species. The microbiota in the mouse intestine (stained green by fluorescent *in situ* hybridization (FISH) is kept separated from the gut epithelia (blue DAPI staining) by a mucus layer which is rich in antimicrobial peptides (AMPs). The AMPs also control the species composition of the microbiota (Salzman *et al*., 2010). Picture from Vaishnava *et al*. (2011). Reprinted with permission from AAAS. **(B,C)** Bacteriomes in insects contain bacteriocytes (cells equivalent to the symbiotic nodule cells) which are infected with endosymbiotic bacteria. The example shows an aphid bacteriome infected with *Buchnera aphidicola* (FISH, green) and *Regiella insecticola* (FISH, red) (B, From The International Aphid Genomics Consortium, (2010) or *Ricketsiella* (FISH, red) (C, Picture from Tsuchida *et al*. (2010). Reprinted with permission from AAAS.). The endosymbiotic bacteria in bacteriocytes show remarkable morphological similitudes to Spherical (S- morphotype) or Elongated (E-morphotype) bacteroids and their formation is controlled by bacteriocyte-specific AMPs (Login *et al*., 2011). **(D,E).** Nodules on the roots of *Medicago truncatula* housing nitrogen-fixing *Sinorhizobium meliloti*. **(F)** The transversal section through the tip of a nodule shows the nodule tissue organization with at the top a meristem containing dividing cells. Below are post-meristematic cells, which over several cell layers gradually become infected with rhizobia and differentiate into symbiotic nodule cells. At the bottom are mature symbiotic cells, which fix nitrogen and occupy the largest part of the nodule. Only the upper part of this zone is shown which extends over many cell layers, depending on the age of the nodule. **(G)** The close-up shows a mature symbiotic cell entirely filled with bacteroids, which are morphologically differentiated by the action of the NCR AMPs.

Currently, the process of biological nitrogen fixation is of great importance to the environment and to global agriculture. The *Rhizobium*-legume and *Frankia*-actinorhizal symbioses on a global scale are responsible every year for 50 million tons of nitrogen injected into agriculture, without greenhouse effects, as compared to 90 million tons on N-fertilizers (Fields, 2004). Legumes are thus plants of extreme ecological and agricultural importance. In addition legumes are a major feed (soybean) and food resource in South and Central America. Legume seeds are very rich in proteins and thus should become in future decades, a valuable alternative to animal proteins. Furthermore, largely thanks to their N_2-fixing capacity, legumes are often pioneering plant for (re) colonization of poor and degraded soils (Pérez-Montano *et al*., 2013). Actinorhizal plants comprise mostly trees and shrubs that are long-lived.

They play important roles in wild-land, polar and mountainous ecosystems including forests, bogs, swamps, coastal dunes, landslides, glacial deposits, riparian zones, shrub lands, prairies and deserts. As such they play key pioneer roles in ecological transitions such as following glacier retreat, forest fires or volcanic eruptions. They are used in agro-forestry and horticulture and have been used by humans to re-vegetate nitrogen-poor substrates such as hydrodam dykes or mine spoils.

Because of those agronomic and ecological stakes, the nitrogen fixing symbioses of plants are subject to extensive scientific research worldwide. This thesis work is part of these research efforts. It is focused on the *Rhizobium*-legume symbiosis. And more specifically, my work has contributed to the understanding how legume plants and rhizobium bacteria are adapted to form a very intricate relationship that is taking place within the nodules. The following paragraphs describe the current knowledge of these plant and bacterial functions, which forms the basis for my thesis work.

3.2. Selection of the symbiont in the *Rhizobium*-legume symbiosis

Rhizobia are horizontally transmitted symbionts, taken up from the soil every time the formation of a new nodule is initiated. Soils are among the richest known environments in terms of bacterial diversity and density. Nevertheless, nodules are generally infected with a clonal population of rhizobia only. Moreover, the legumes only accept specific rhizobial strains. Therefore, the recognition between the symbionts is extremely specific and is mainly based on a two-step recognition of both partners. Major determinants of this specificity are: **1**) the plant-produced flavonoids found in root exudates when the plant grows in nitrogen limiting conditions and **2**) the secreted bacterial Nod factor molecules, which are produced after rhizobia have sensed the cognate flavonoid and which have a lipo-chitooligosaccharide backbone in common but which are differentiated by *Rhizobium* strain-specific substitutions on this backbone (Mergaert *et al*., 1997). Binding of Nod factors to their cognate plant lysin motif (LysM) type receptor-like kinases activates a signaling process, which culminates in the initiation of nodule formation and in the infection of the incipient nodules by the rhizobia (Oldroyd *et al*., 2011). Along the infection process, from its initiation till the settlement of the rhizobia inside symbiotic nodule cells, other bacterial patterns like lipopolysaccharides and exopolysaccharides are recognized by the host tissues (Gibson *et al*., 2008). These sensing mechanisms function as checkpoints and failure of recognition will result in abortion of the interaction. Together, these mechanisms assure that the proper symbiotic bacteria are selected from the environment.

3.3. Symbiotic nodule cell formation

The details of the developmental processes leading to a mature nodule differ considerably among legumes. In the most common mechanism (Gibson *et al*., 2008; Oldroyd *et al*., 2011), Nod factors produced by rhizobia located at the root surface, induce cell divisions in root cortical cells forming the nodule primordium. Concomitantly, the bacteria penetrate the host tissues through tubular infection structures called infection threads, which guide the bacteria towards the cells of the growing primordium. They release rhizobia by an endocytosis-like mechanism in the primordium cells, enclosed in organelle-like, membrane-bound structures called symbiosomes. Infected cells stop dividing but grow further by repeated endoreduplication cycles to become giant polyploid symbiotic cells (Vinardell *et al*., 2003). The differentiating symbiotic cells are gradually infected and filled with *Rhizobium*-containing symbiosomes. Continuous bacterial release from infection threads in combination with symbiosome divisions leads to a complete packing of the cytosolic space of the symbiotic cells. Mature symbiotic cells have about 80-fold larger cell size and endoploidy levels up to 64C compared to the diploid progenitor cells (**Fig. 4**).

Nodule organogenesis is accompanied with major changes in the gene expression program (El Yahyaoui *et al*., 2004; Manthey *et al*., 2004; Lohar *et al*., 2006; Starker *et al*., 2006; Maunoury *et al*., 2010; Moreau *et al*., 2011). Two waves of transcriptional reprogramming occur during *Medicago truncatula* nodule formation, which involves the repression and the massive induction of hundreds of genes (**Fig. 4**) (Maunoury *et al*., 2010).

A first wave occurs concomitantly with the establishment of an incipient nodule containing differentiating symbiotic cells while a second wave is observed at the establishment of a functional nodule. In plant and bacterial symbiotic mutants forming non-functional nodules, the transcriptional activation of these genes is only observed when endoreduplication takes also place suggesting that most of them are expressed in the endoreduplicated symbiotic cells and that these cells undergo a tremendous transcriptional reprogramming (Maunoury *et al*., 2010). Several examples of these genes were confirmed to be indeed specifically and only expressed in the symbiotic nodule cells by *in situ* transcript localization techniques (Mergaert *et al*., 2003; Maunoury *et al*., 2010; Van de Velde *et al*., 2010). Moreover, the *in situ* transcript localization revealed that genes of the first wave are activated in young differentiating symbiotic cells with intermediate ploidy levels and the second wave genes are activated in mature symbiotic cells with high ploidy levels (**Fig. 4**).

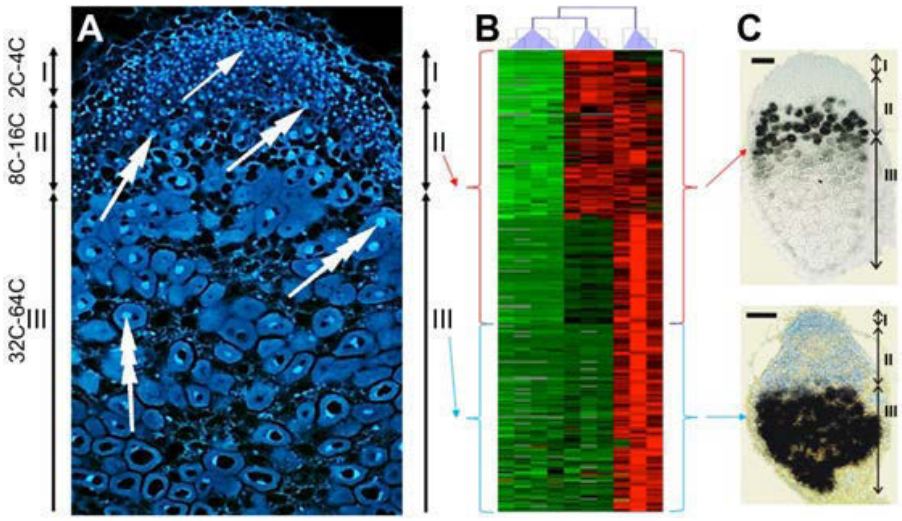

Figure 4. Differentiation of symbiotic nodule cells

(A) A transversal section through the tip of a *M. truncatula* nodule, similar as in **Fig. 2F** but stained with DAPI to visualize the plant cell nuclei. At the top a meristem (I), below a differentiation zone were symbiotic cells are formed (II) and finally the upper part of the zone III with mature symbiotic cells. Cells in zone I are in a mitotic cell cycle and have small nuclei (simple arrow) with a 2C/4C DNA content. In contrast, cells in zone II have transited into an endocycle and become polyploid cells with 8C to 32C DNA content. Their nuclei (double arrows) increase in size. Mature symbiotic cells have a 32C or 64C DNA content and their nuclei (triple arrow) have reached their maximum size. **(B)** The symbiotic cell differentiation is accompanied with a dramatic transcriptome shift as revealed by the transcriptome analysis shown in the heat map. The expression profiles in columns correspond to different stages in nodule development from early (left) to late (right). The transcriptome is activated in two consecutive waves. The clusters of genes indicated with red brackets are activated during the first wave, in differentiating symbiotic cells with intermediate ploidy levels (zone II). Genes indicated with blue brackets are activated during the second wave in mature symbiotic cells with high ploidy levels. **(C)** Examples of *in situ* transcript detection of nodule-specific genes are shown (Mergaert *et al*., 2003; www.plantphysiol.org, Copyright American Society of Plant Biologists).

A first wave occurs concomitantly with the establishment of an incipient nodule containing differentiating symbiotic cells while a second wave is observed at the establishment of a functional nodule. In plant and bacterial symbiotic mutants forming non-functional nodules, the transcriptional activation of these genes is only observed when endoreduplication takes also place suggesting that most of them are expressed in the endoreduplicated symbiotic cells and that these cells undergo a tremendous transcriptional reprogramming (Maunoury *et al*., 2010). Several examples of these genes were confirmed to be indeed specifically and only expressed in the symbiotic nodule cells by *in situ* transcript localization techniques (Mergaert *et al*., 2003; Maunoury *et al*., 2010; Van de Velde *et al*., 2010). Moreover, the *in situ* transcript localization revealed that genes of the first wave are activated in young differentiating symbiotic cells with intermediate ploidy levels and the second wave genes are activated in mature symbiotic cells with high ploidy levels (**Fig. 4**).

These transcriptome switches put in place the specific metabolic conditions in the symbiotic nodule cell which are required for their symbiotic role, including the maintenance

of a low free-oxygen concentration, feeding the microsymbionts and assimilating and transporting the fixed nitrogen (Oldroyd *et al*., 2011). An additional striking feature of the symbiotic nodule cell-specific transcriptome in *M. truncatula*, representing more than 50% of the induced genes, is the activation of the secretory pathway, including genes encoding regulators of the pathway, a large number of secretory proteins and transmembrane proteins (Maunoury *et al*., 2010). The most likely role of this pathway is to feed host proteins to the membrane-bound symbiosomes and their bacteroids. In agreement with this dominant role of these cells, they have a remarkably well developed and abundant endoplasmic reticulum, which is the gateway to the secretory pathway (Maunoury *et al*., 2010). The most notable secretory proteins produced by the symbiotic nodule cells are several classes of peptides including the Glycine Rich Peptides (GRPs) (Kevei *et al*., 2002; Alunni *et al*., 2007), the Small Nodulin Acidic RNA-binding Proteins (SNARPs) (Laporte *et al*., 2010) and the Nodule-specific Cysteine-Rich (NCR) peptides (Mergaert *et al*., 2003; Alunni *et al*., 2007).

3.4. Determinate and indeterminate nodule development

The histology, form and size of nodules can be quite different in the various legume hosts. Nevertheless, they can be classified in two major types, either as the determinate or the indeterminate type (**Fig. 5**) (Terpolilli *et al*., 2012; Kondorosi *et al*., 2013; Maróti *et al*., 2014).

The major difference between the determinate and indeterminate nodules is the transient or persistent nature of the nodule meristem, respectively. In addition, determinate nodules originate from divisions of outer cortical root cells while the indeterminate ones from inner cortical root cells. Determinate nodules are spherical since cell division is only maintained at the primordium stage and later, while the cell number does not increase, nodule growth is achieved by cell growth and determinate nodules contain homogenous populations of symbiotic cells (**Fig. 5**). Nevertheless, this process of cell enlargement also involves endoreduplication and symbiotic plant cells can reach the 32C stage (Gonzales-Sama *et al*., 2006). In a way, determinate nodule organogenesis can be considered as a sequential process, whereas indeterminate nodule formation may represent a spatial differentiation gradient.

In contrast, in the indeterminate nodules, the active cell division is maintained. A nodule meristem is present in the apical region (zone I), which, by constant generation of new cells, provokes the continuous growth of the organ resulting in an elongated nodule shape. The cells leaving the meristem no longer divide and enter a differentiation phase. The infection thread releases the bacteria into the submeristematic cells, which differentiate

gradually along the 12-15 cell layers of the infection zone (zone II), leading to the development of nitrogen-fixing symbiotic cells in nodule zone III (**Fig. 5**) (Pop and Ott, 2011).

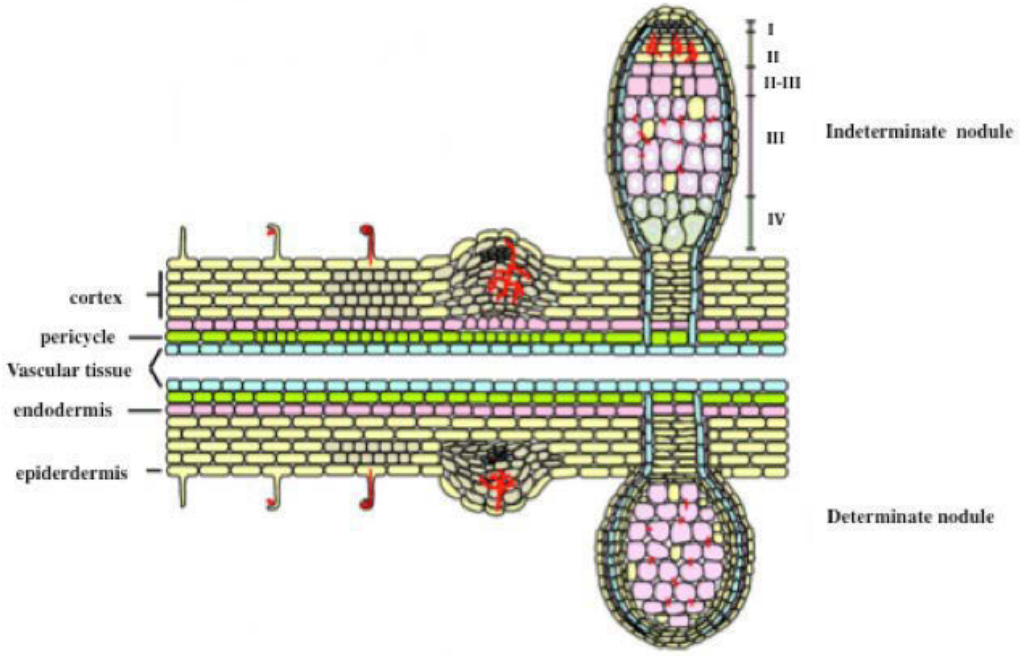

Figure 5. Development of determinate and indeterminate root nodules

In legumes that develop interterminate nodules (top), rhizobial infection of the plant root induces periclinical cell divisions in the pericycle folloxed by inner cortical cell proliferation. Development of a nodule primordium is accompanied by the presence of a persistent mersitem leading to a zonation of an inderterminate nodule with the meristem (zone I), the infection zone (II), an interzone (zone II-III), the fixation zone (III) and the senescence zone (IV). By contrast, determinate nodules (bottom) derive from cell divisions in the outer root cortex where the meristimatic activity is lost in mature nodules. Red are rhizobia attached to root hairs are present in infection threads. Yellow cells are cortical and epidermal cells. Grey cells are dividing cells in nodule primordia or nodule meristem. Large pinkish cells are the differentiation or differentiated symbiotic cells, carrying bacteroids. Adapted from Popp and Ott, (2011).

Determinate nodules are formed on (sub-) tropical plants like *Phaseolus*, *Glycine,* and *Vigna* species belonging to the phaseolids or on robinioids such as *Lotus* and *Sesbania* species; and dalbergoids such as *Arachis* and *Aeschynomene* species. Indeterminate nodules develop on temperate legumes, like *M. truncatula, Pisum sativum, Vicia faba, Trifolium repens*, *Astragalus sinicus* or *Galega orientalis*, which all belong to the inverted repeat-lacking clade (IRLC) but indeterminate nodules can be found in almost all major legume clades.

3.5. Bacteroid formation in the course of nodulation

The nitrogen fixing rhizobia within the symbiosomes are called bacteroids. Bacteroids are differentiated bacteria entirely dissimilar to the free-living rhizobia in soil. This

differentiation results from a massive transcriptome switch, which is in large part regulated by the FixLJ oxygen-sensing two-component regulator. Upon detection of the low free-oxygen in the symbiotic nodule cells, the regulator activates the genes for nitrogenase subunits and cofactor synthesis and for a specific respiratory chain which functions under microaerobic conditions. Moreover, most of the bacterial household genes for cell growth and division, synthesis of ribosomal proteins, nucleic acid synthesis and repair, outer membrane proteins and peptidoglycan are shut down in bacterioids (Barnett *et al*., 2004; Becker *et al*., 2004; Uchiumi *et al.*, 2004; Capela *et al*., 2006; Karunakaran *et al*., 2009). Ammonium has to be produced by the nitrogenase in the bacteroids in sufficient amounts to meet the needs of the plant. Therefore, the nitrogen stress response which regulates the nitrogen economy in the bacterial cell has to uncouple nitrogen fixation from the bacterial nitrogen needs in bacteroids contrary to their strict coordination in free-living nitrogen fixing bacteria which produce only that amount of ammonium they need for their own growth (Yurgel and Kahn, 2008). Even ammonium assimilation into amino acids and the synthesis of branched chain amino acids is switched off in bacteroids which become dependent on amino acid supply by the host plant (Prell *et al*., 2009; Mulley *et al*., 2011). This phenomenon is called symbiotic auxotrophy. The dependency of bacteroids on the host plant is further illustrated by the inability of most rhizobia to produce autonomously the FeMo-cofactor of nitrogenase. They depend on supply of homocitrate by the host cell for the production of this very specific nitrogenase cofactor (Hakoyama *et al*., 2009). This far-stretched metabolic integration of the endosymbiont in its host cell advocates the hypothesis that symbiosomes are organelle-like compartments.

In many but not all legume plants, the differentiation of bacteroids is accompanied by a morphological metamorphosis whereby the bacteroid cell becomes enlarged. Three different bacteroid morphotypes have been described (**Fig. 6**) (Vasse *et al*., 1990; Mergaert *et al*., 2006; Bonaldi *et al*., 2011; Kondorosi *et al.*, 2013; see below, **Chapter II**): elongated bacteroids which are sometimes also branched (designated as the E-morphotype; *e.g.* in legumes of the IRLC which includes the model legume *M. truncatula* or in some species of the distantly related *Aeschynomene* genus), enlarged spherical bacteroids (S-morphotype; in *Aeschynomene* or *Arachis* species of the Dalbergoid legume clade) and finally bacteroids which remain unmodified and have a rod-shaped morphology similar to free-growing bacteria (U-morphotype; e.g. in the clades comprising *Lotus*, *Phaseolus* or *Glycine*). *Rhizobium* strains that can nodulate legumes of different clades adapt a bacteroid morphotype, which is determined by the host rather than the bacterial genetic repertoire. For example, nearly isogenic *Rhizobium* strains nodulating the IRLC-legumes pea or *Vicia* and the non-IRLC

species *Phaseolus vulgaris* or *Lotus japonicus* have E-morphotype bacteroids in the former and U-morphotype bacteroids in the latter legume hosts (Mergaert *et al*., 2006).

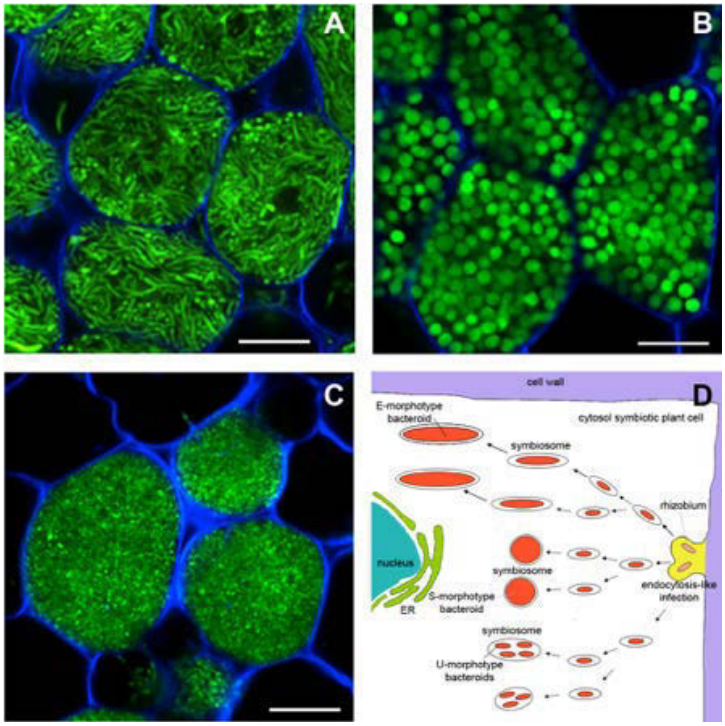

Figure 6. Bacteroid morphotypes

Bacteroids display different morphotypes according the host plant. **(A)** E-morphotype: *Bradyrhizobium* ORS285 bacteroids on *A. afraspera*. **(B)** S-morphotype: *Bradyrhizobium* ORS285 bacteroids on *A. evenia*. **(C)** U-morphotype: *Bradyrhizobium japonicum* USDA110 bacteroids on soybean. The bacteroids are green due to GFP expression (A and B) or by Syto9 DNA staining (C) and the contours of the symbiotic host cell are blue due to calcofluor staining. Scale bars, 10 µm. **(D)** Schematic presentation of the paths leading to the different bacteroid morphotypes. In all cases, bacteroids are released in the host cells by an endocytotic-like process and are enclosed in a symbiosome vesicle. The bacteria are released either from an infection thread penetrating the cell or directly from the extracellular space. The cell is gradually filled with bacteroids by repeated endocytosis, by division and multiplication of the symbiosomes and by division of the bacteroids within de symbiosomes in the case of U-morphotype bacteroids leading to multiple bacteroids per symbiosome or by host cell factors-induced enlargement in the case of E- and S-morphotype bacteroids, which remain single in the symbiosomes.

Some *Rhizobium* isolates can nodulate both *Arachis* and *Vigna unguiculata* legumes in which they transform into S-morphotype and U-morphotype bacteroids respectively (Sen and Weaver, 1984). In the *Aeschynomene* genus, bacteroids can be of the E- or S-morphotype. The *Bradyrhizobium* strain ORS285 which can nodulate *Aeschynomene afraspera* and *Aeschynomene indica* or *Aeschynomene evenia* transforms into E-morphotype bacteroids in nodules of the former legume and in S-morphotype bacteroids in the latter two legumes (Bonaldi *et al*., 2011). Moreover, *Bradyrhizobium japonicum* strain USDA110 which is the *Bradyrhizobium* type-strain, symbiont of *Glycine max* (soybean), is able to efficiently nodulate *A. afraspera* (Renier *et al*., 2011). This strain makes U-morphotype bacteroids on soybean but E-morphotype bacteroids on *A. afraspera*.

These observations indicate that the bacteroid morphotype is controlled by the host plant and results from host factors that induce the bacteroid metamorphosis (Mergaert *et al*., 2006).

The *Sinorhizobium meliloti* E-morphotype bacteroids in *M. truncatula* and *Medicago sativa* nodules reach up to 10 µm in length, compared to the 1-2 µm of free-living rhizobia. The genome of these bacteroids is highly amplified and condensed in multiple nucleoids of variable size (Vasse *et al*., 1990; Mergaert *et al*., 2006). This implies that the differentiation process is the result of a bacterial cell cycle switch which blocks division but permits further genome replication. In agreement with this, genetic or pharmacological interference with the *S. meliloti* cell cycle machinery leads to morphological transformations, which are E-morphotype bacteroid-like (Latch and Margolin, 1997; Wright *et al*., 1997; Sibley *et al*., 2006; Cheng *et al*., 2007; Pini *et al.*, 2013). Moreover, also interference with peptidoglycan biosynthesis results in bacterial morphologies resembling E- or S-morphotype differentiated bacteroids (Popham and Young, 2003; Young, 2003; Lam *et al.*, 2009) suggesting that the rhizobial peptidoglycan might be affected during bacteroid differentiation. The envelopes of the bacteroids in *Medicago* and other, related legumes of the IRLC clade like *P. sativum* (pea) or *V. sativa* (vetch) are considerably modified. The lipopolysaccharides, which are the primary component of the outer leaflet of the outer membrane in gram-negative bacteria, become more hydrophobic during bacteroid differentiation (Kannenberg and Carlson, 2001; Ferguson *et al*., 2005). Moreover, a propidium iodide uptake assay demonstrated that bacteroid membranes in these legumes are permeable to propidium iodide, contrary to the membranes of the free-living bacteria or to the U-morphotype bacteroids from *L. japonicus* or *P. vulgaris* nodules (Mergaert *et al*., 2006). Finally, the E-morphotype differentiation in the IRLC legumes is an irreversible process because these bacteroids cannot reproduce anymore. Therefore, this has been called terminal bacteroid differentiation. This is in contrast to the U-morphotype bacteroids, which can reproduce and are therefore reversibly differentiated (Mergaert *et al*., 2006).

2. The NCR peptide family

The symbiotic nodule cells in *M. truncatula* express factors inducing E-morphotype bacteroid formation. Candidate factors were identified by interrogating nodule transcriptome datasets for *M. truncatula* as well as for *L. japonicus*, *V. unguiculata*, *G. max* and *P. vulgaris*, which form U-morphotype bacteroids. The criteria that were applied were nodule-specific gene expression and expression in the symbiotic nodule cells, presence of orthologous genes

in other species of the IRLC clade but they are absence in the tested legumes with U-morphotype bacteroids and finally, the candidate genes should encode secretory proteins compatible with transport to the symbiosomes. On the basis of this analysis, the NCR peptides (Nodule-specific Cysteine-rich Peptides) were proposed as likely candidates for the bacteroid differentiation factors in IRLC legumes that controls the bacteroid elongation and polyploidization (Mergaert *et al*., 2006; Alunni *et al*., 2007; Van de Velde *et al*., 2010).

Highly remarkable features of the *M. truncatula NCR* gene family are its composition of about 600 homologous genes (Young *et al*., 2011; Zhou *et al*., 2013) and the exclusive nodule-specific expression of nearly all the tested genes (Mergaert *et al*., 2003; Alunni *et al*., 2007; see below, **Chapter I**). The almost 600 *NCR* genes are spread all over the *M. truncatula* genome but the majority of them appears in clusters of up to 11 members (Alunni *et al.*, 2007; Young *et al.*, 2011). This suggests that the family has expanded through multiple gene duplications mediated by mobile elements such as transposons or viruses. The microarray dataset in the *M. truncatula* gene expression atlas (Benedito *et al*., 2008), containing 254 genome-wide expression profiles, and similarly complex EST datasets (http://compbio.dfci.harvard.edu/tgi/) revealed expression exclusively in nodules for about 300 different *NCR* genes except for 5, which were also expressed in roots and some aerial tissues (Mergaert *et al*., 2003; Guefrachi *et al.*, 2014, see below, **Chapter I**). Expression was not found in any other plant organ or in other biotic interactions with mycorrhizal fungi, rhizosphere bacteria, pathogens, or insects. Also in nematode infected roots, NCR expression could not be detected (Favery *et al*., 2002; Guefrachi *et al.*, 2014). Transcriptome analysis of the nodules of a collection of *M. truncatula* and *S. meliloti* symbiotic mutants revealed a correlation between *NCR* gene expression and the formation of symbiotic nodule cells (Maunoury *et al*., 2010). Moreover, when tested with promoter-GUS fusion or *in situ* hybridization, the expression of *NCR* genes was always found to be restricted to the *Rhizobium*-infected plant cells (**Fig. 4**). However, subsets of *NCR* genes were expressed in young symbiotic cells while others in older or mature symbiotic cells (Mergaert *et al*., 2003). Some *NCR* genes are expressed at very high levels (Guefrachi *et al.*, 2014, see below, **Chapter I**) comparable to leghemoglobin, which is one of the most strongly expressed nodule genes. On the basis of EST counts, it was estimated that the combined transcripts of the NCR family constitute close to 5% of the total mRNA pool in nodules (Mergaert *et al*., 2003).

NCR homologs are present in other IRLC legumes including *M. sativa* (Jimenez-Zurdo *et al*., 2000), *P. sativum* (Scheres *et al*., 1990; Kato *et al*., 2002; Chou *et al*., 2006), *A. sinicus* (Chou *et al*., 2006), *T. repens* (Crockard *et al*., 2002), *G. orientalis* (Kaijalainen *et al*.,

2002) and *V. faba* (Frühling *et al*., 2000) where they belong to multi-gene families although it is presently unknown if the *NCR* gene families in these plants are as large as in *M. truncatula*. However, we know that each of the known genes in those plants is expressed specifically in the nodules as in *M. truncatula*. In contrast, by interrogating EST databases (http://compbio.dfci.harvard.edu/tgi/) no orthologous genes were found in other legume species, including species of closely related clades forming nodules with U-morphotype bacteroids. Moreover, the comparison of genomic loci carrying *NCR* genes in *M. truncatula* with microsyntenic loci in *L. japonicus* confirmed the absence of the *NCR* genes in *L. japonicus* (Alunni *et al*., 2007). Also other plant families or members of the animal kingdom lack *NCR* genes (Mergaert *et al*., 2003; Graham *et al*., 2004; Alunni *et al*., 2007). For the sake of completeness, 3 sequences in the *Arabidopsis* genome were reported that potentially encode peptides with the typical NCR cysteine motif (Zhou *et al.*, 2013) although no experimental evidence for the transcription of those genes and production of the peptides is available at present, leaving open the question whether those sequences are truly genes. Thus NCRs are so-called lineage-specific or taxonomically restricted genes (TRGs). TRGs generally constitute an important part of eukaryotic genomes with 10-20% of genes lacking recognizable homologs in other species. TRGs are associated with evolutionary novelties and phylum-specific processes (Khalturin *et al*., 2009).

Despite their uniqueness, the protein structure, gene organization and family structure of the NCRs strikingly resemble those of antimicrobial peptides (AMPs) even if the NCRs have only limited amino acid sequence homology with known AMPs. AMPs are a diverse class of secreted and ribosomally synthesized peptides that are crucial effectors of innate immunity in both animals and plants (Zasloff, 2002; Brogden, 2005; Maróti *et al*., 2011). AMPs are constitutively produced or induced in response to microbial infections and have the capacity to attack and kill the invading microbes. AMPs in both plants and animals are often part of multigene families with dozens of members (Schutte *et al*., 2002; Silverstein *et al*., 2007; Amid *et al*., 2009) similarly as the NCRs, which are an extreme example of a multigene family. The exon-intron structure of the *NCR* genes is conserved with plant AMP genes (Mergaert *et al*., 2003). NCR peptide sequences contain an N-terminal signal peptide, encoded by the first exon of the genes, which is a cleavable tag for targeting of the peptide in the secretory pathway (**Fig. 7**). By expressing fusions of NCRs or their signal peptides with fluorescent proteins in a heterologous system, it was shown that these signal peptides are indeed functional and mediate entry of the peptides in the secretory pathway (Mergaert *et al*., 2003). The mature NCR peptides, obtained after processing of the signal peptide, are encoded

in the second exon, are around 40 amino acids long and contain a highly conserved pattern of 4 or 6 cysteine residues (**Fig. 7**).

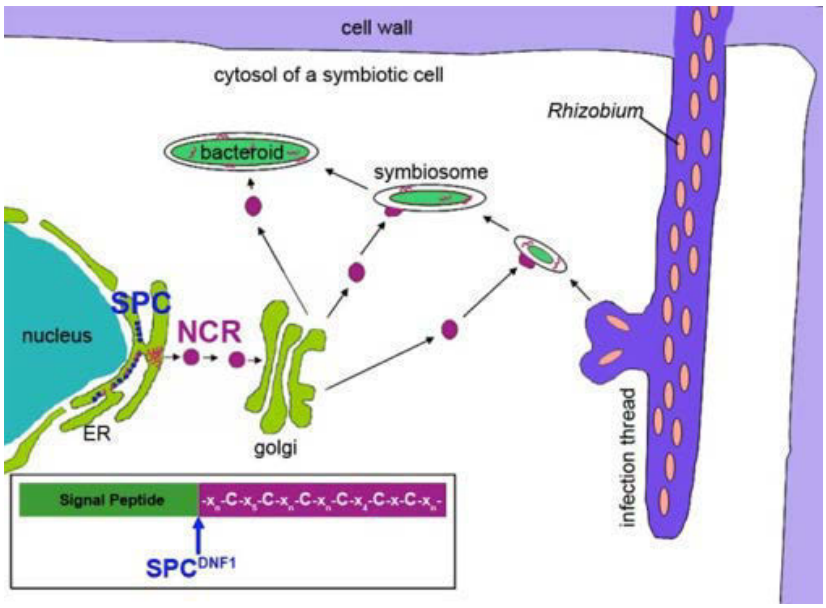

Figure 7. NCR-mediated differentiation of E-morphotype bacteroids in IRLC legumes

In a *Medicago truncatula* nodule cell, the *Sinorhizobium meliloti* wild-type (WT) and BacA mutant bacteria are released from the infection threads into membrane-bound symbiosomes. The bacteria are then challenged with host NCRs (purple), which are synthesized on the endoplasmic reticulum (ER) and transported to the symbiosomes via vesicle transport and a symbiotic nodule-cell-specific secretory pathway. This pathway requires a nodule-specific signal peptidase complex (SPC). The SPC is located in the ER and removes the signal peptide of the NCR pre-proteins (inset). The NCRs mediate the differentiation of the bacteria into a nitrogen-fixing bacteroid, resulting in the chronic infection. The conserved cysteines in mature NCR peptides are shown in the inset ("x" are variable amino acids in the NCR peptide sequence).

This structure is typical for certain classes of AMPs and notably the defensins which have a signal peptide and a mature domain with conserved cysteines although the number of cysteine residues and their configuration is different from the NCRs (Ganz, 2003; Silverstein *et al*., 2007; Aerts *et al*., 2008). The NCRs are clearly related to each other because their cysteine signature is strongly conserved as well as a few amino acids flanking the cysteine residues and also their signal peptides are typical and highly similar. However, the sequences of mature domains as a whole are highly variable amongst the NCR family members (Mergaert *et al*., 2003). This reflects an evolutionary pattern with the signal peptides subjected to purifying selection in contrast to the mature peptides, which were subjected to diversifying selection (Alunni *et al*., 2007; Nallu *et al*., 2014). Similar observations are frequently made for AMP gene families, including conservation of signal peptides and cysteine residues but strong divergence in the functional mature peptide (Patil *et al*., 2004; Semple *et al*., 2006). Finally, AMPs are often TRGs similarly to the NCRs (Khalturin *et al*.,

2009). AMPs encoded by TRGs may be effective defense molecules against taxon-specific microbial interactors.

3. NCR transport to the symbiosomes

The possible involvement AMP-like peptides in the interaction with symbiotic bacteria, which are deliberately maintained by the host, may seem surprising at first sight because AMPs are known in the first place to counteract and eliminate invading bacteria. Nevertheless, this hypothesis is meaningful in light of certain E-morphotype bacteroid features such as membrane modifications, inhibition of cytokinesis and inability to reproduce which are known effects of different types of AMPs. And indeed, several experimental lines of evidence confirmed the involvement of the NCR peptides in bacteroid differentiation.

NCRs, recognized by specific antibodies and by peptide sequencing with mass spectrometry, accumulate in the nodules and co-purify with the bacteroids. *In situ* localization of the NCR peptides in nodules with fluorescent protein tagged NCRs or with immunolocalization confirmed this localization and demonstrated moreover that at least some NCRs are transported across the bacterial membranes and accumulate in the bacteroid cytosol (Van de Velde *et al*., 2010). Thus NCRs are targeted to and accumulate in high amounts in symbiosomes and bacteroids and this localization of the peptides indicates that the bacteroids are their targets (**Fig. 7**).

The symbiosome localization of the NCR peptides is also in agreement with the presence of the characteristic signal peptide and a transport mechanism of the peptides through the secretory pathway. The characterization of the *M. truncatula DNF1* gene has indeed highlighted the importance of the secretory pathway for NCR transport. One of the earliest steps in this pathway is the proteolytic cleavage of the signal peptide (**Fig. 7**), which is critical for the correct targeting of secretory proteins. It takes place during translocation of the secretory proteins from their site of synthesis by the ribosomes at the cytosolic face of the endoplasmic reticulum into the lumen of the endoplasmic reticulum and it is performed by the signal peptidase complex (SPC) which is located in the membrane of the endoplasmic reticulum (Paetzel *et al*., 2002). The SPC is composed of four subunits conserved in all eukaryotes. The SPC18 subunit contains the proteolytic catalytic site (Paetzel *et al*., 2002). SPC18 together with SPC22 are the critical subunits for SPC activity while the other two subunits, SPC12 and SPC25 are apparently dispensable for SPC activity, at least in yeast (Fang *et al*., 1997; Meyer and Hartmann, 1997). The *M. truncatula* genome contains two gene copies for SPC12, SPC18 and SPC22, which are 77% to 82% identical to each other and one

encoding SPC25. Interestingly, four of these genes, one copy for each of the four subunits, named DAS12, DAS18, DNF1 and DAS25 have a highly correlated gene expression profile with a major expression in nodules, a lower expression in seeds and a basal expression in other plant tissues (Wang *et al*., 2010). The other three genes have a more uniform expression in all plant tissues. The *M. truncatula dnf1* (does not fix 1) mutant was identified on the basis of its symbiotic phenotype, forming non-functional nodules (Starker *et al*., 2006). The mutant was found to be deficient in the DNF1 nodule-enhanced SPC22 subunit of the SPC (Wang *et al*., 2010). The *dnf1* nodules contain infected nodule cells; however, the symbiosome bacteria do not differentiate into elongated bacteroids (Van de Velde *et al*., 2010; Wang *et al*., 2010). NCRs are produced in the nodules of the mutant but their signal peptide is not cleaved in agreement with the known essential role of the SPC22 subunit in signal peptide processing (Fang *et al*., 1997; Meyer and Hartmann, 1997). The inactivation of NCR processing resulted in the accumulation of NCR peptides in the endoplasmic reticulum and the blockage of their further transport to the symbiosomes. Thus, obstructing NCR transport is correlated with the absence of bacteroid differentiation in the symbiosomes, in agreement with a role of the NCRs in this bacterial differentiation process (Van de Velde *et al*., 2010).

4. NCR-mediated bacteroid differentiation

Inhibiting NCR transport inactivated the function of the whole NCR family but this strategy could not exclude the possibility that the resulting phenotype was the consequence of other proteins or peptides produced by the symbiotic nodule cells and transported to the symbiosomes through the secretory pathway. An opposite, gain-of-function approach, consisted in ectopic expression of individual *NCR* genes in the infected nodule cells of *L. japonicus*, a legume with reversible bacteroid differentiation and lacking *NCR* genes. In control nodules of *L. japonicus*, symbiosomes contained multiple U-morphotype bacteroids while in the NCR-expressing nodules a significant number of symbiosomes contained single, E-morphotype bacteroids thereby mimicking the symbiosomes and bacteroids in *M. truncatula* (Van de Velde *et al*., 2010). Nevertheless, the effect was only partial, possibly because the sequential and cooperative action of multiple peptides in the right proportions is required to obtain an optimal response such as is observed in the *M. truncatula* nodules. Similarly, *in vitro* application of single NCR peptides to *S. meliloti* free-living bacteria resulted in E-morphotype bacteroid-like features. Notably, peptides induced high permeability of the membrane, inhibition of bacterial proliferation, DNA accumulation and cell elongation

(Van de Velde *et al*., 2010; Haag *et al*., 2011) but again, the effects were partial probably because an optimal cocktail of peptides is required.

The polyploidy of the bacteroids is the testimony of a switch in their cell cycle that must have taken place during their formation after their release from the infection threads in the host cell. In a regular bacterial cell cycle, DNA replication and division are alternating resulting in cells, which have a haploid (1C) or diploid (2C) genome content while the polyploid bacteroids must result from an inhibition of the division accompanied with ongoing DNA replication (**Fig. 8A**). Thus one of the major effects of the NCR peptides may be an interference with the regulatory circuit of the bacteria that controls the cell cycle. The perturbation of the bacterial cell cycle during bacteroid formation is nicely illustrated when the expression of bacterial cell cycle regulatory genes is followed *in situ* within a mature nodule of *M. truncatula* (**Fig. 8B**). The expression pattern of these genes undergoes a complete switch, which coincides with the formation of the bacteroids and with the expression of the *NCR* genes. For example, the master regulator CtrA, which promotes the cell division cells and inhibits the DNA replication stage, is strongly expressed in the apical parts of the nodule where bacteria are not yet differentiated and divide within the infection threads. However, CtrA expression drops strongly where the bacteria start the differentiation in the proximal infection zone (pI) and interzone II-III (IZ), in agreement with the absence of division. DnaA on the other hand which is the master regulator controlling DNA replication has an opposite pattern and its highest expression level is in pI and IZ thus coinciding with the genome amplification of the differentiating bacteroids. How the NCR peptides affect the cell cycle remains to be elucidated. However, a recent study shed some light on this question. It was shown that a specific NCR peptide, NCR247, affected very strongly the expression of cell cycle regulated genes in *S. meliloti*. Particularly the expression of the CtrA regulated genes was inhibited by the NCR247 peptide (Penterman *et al*., 2014) (**Fig. 8C**). Interestingly, when *S. meliloti* is depleted genetically from CtrA, the bacteria stop further growth but elongate very strongly and become polyploid (Pini *et al*., 2013) (**Fig. 8D**). Together these findings suggest that NCR peptides drive the endoreduplication and enlargement of the intracellular *S. meliloti* in the symbiotic nodule cells by altering directly or more likely indirectly the regulation of the cell-cycle, leading to the loss of CtrA activity and the consequent arrest in cell division. A second recent study described that NCR247 is taken up by *S. meliloti* cells and that the peptide interacts with multiple targets including FtsZ, a cell division protein, ribosomes which are inhibited by the interaction with NCR247 and the chaperone GroEL which is required for normal bacteroid formation (Farkas *et al*., 2014).

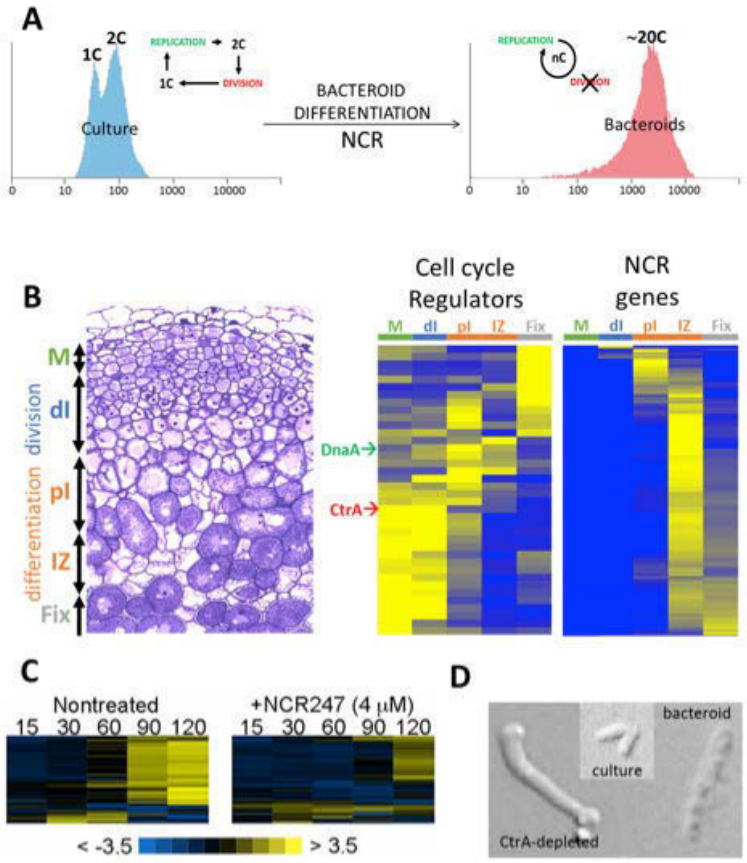

Figure 8. Cell cycle deregulation induced by NCR peptides leading to polyploid bacteroids

(A) Analysis by flow cytometry of the DNA content of cultured *S. meliloti* stained with the DNA-labelling fluorescent dye DAPI (left). The x-axis (logarithmic scale) represents the DNA content (DAPI fluorescence) and the y-axis is the number of bacterial cells. The histogram shows bacteria with 1C and 2C genome content reflecting the cycling of the bacteria between genome duplication and cell division. Under the influence of the NCR peptides during bacteroid differentiation the endosymbiont cells become polyploid because division is inhibited while replication is ongoing (right). **(B)** *In situ* analysis of gene expression in a nodule of *M. truncatula* by laser micro-dissection coupled to RNA-seq analysis. Left panel: nodule section showing the different nodule zones in which gene expression was analyzed. M: nodule meristem; dI: distal infection zone; pI: proximal infection zone; IZ: interzone; Fix: fixation zone. Central panel: expression of the major cell cycle regulators of *S. meliloti* in the different nodule tissues. Yellow is high expression and blue is low expression. The genes encoding DnaA and CtrA are highlighted by arrows. Right panel: expression profile of 334 *NCR* genes in nodule tissues. Data was extracted from the study by Roux *et al.* (2014). **(C)** Expression of CtrA regulated cell cycle genes in synchronized *S. meliloti* cultures that were treated or not with the NCR247 peptide. The time points above the columns indicate the time of synchronous growth. Image adapted from Penterman *et al*. (2014). **(D)** Effect of the genetic depletion of CtrA on the morphology of *S. meliloti* compared to a wild-type cell and to a bacteroid cell. Data adapted from Pini *et al*. (2013).

Therefore, the mechanism of action of NCR247 seems to be complex and multiple and more work is required to fully elucidate the precise mechanism of the NCR-induced cell cycle switch. In addition, another NCR peptide, NCR035, was found to be localized *in vitro* at the division site of *S. meliloti* cells but not NCR247 (Van de Velde *et al*., 2010), suggesting that the NCR035 peptide interferes with the bacterial cell division machinery but probably by a different mechanism than NCR247.

5. Why morphological bacteroid differentiation?

Since morphological differentiation of bacteroids does not take place in all legumes, this process is not required *per se* for the maintenance of the endosymbiotic bacterial population, prompting the question why certain legumes impose this differentiation process on the nodule rhizobia. While this question remains entirely open at present, some studies suggest a selective advantage to the plant (Sen and Weaver, 1981; Oono and Denison, 2010; Oono *et al*., 2010). Phylogenetic analysis and ancestral state reconstruction suggested that the ancestral morphotype of bacteroids was the undifferentiated U-morphotype while the differentiated E- and S-morphotypes have evolved several times independently in the legume family (Oono *et al*., 2010). This conclusion, together with the facts that the plant imposes the morphological differentiation and that the bacteria are affected in their reproductive capacity by the differentiation, is a strong argument for a selective advantage to the plant associated with the morphological bacteroid differentiation. Other studies demonstrated a higher symbiotic performance for E- or S-morphotype bacteroids compared to U-morphotype bacteroids (Sen and Weaver, 1981; Oono and Denison, 2010). However, even if these studies provide attractive arguments, the molecular and physiological basis for the selective advantage to the plant remain entirely unknown. Could the NCRs affect the symbiotic efficiency of the bacteroids?

As discussed above, some NCR peptides are implicated in the interference with the bacterial cell division machinery. However, the NCRs display a high sequence variety including a large number of anionic (acidic) or neutral peptides, which have no antimicrobial activity and are unable to induce bacterial differentiation. Moreover, some of these anionic peptides (e.g. NCR001) are expressed at very high levels (Mergaert *et al*., 2003) and accumulate inside the bacteroids (Van de Velde *et al*., 2010). This suggests a multiplicity of functions, modes of action and bacterial targets for the NCR peptides. Together, the net effect of the peptides might be an optimization of the efficiency of the nitrogen fixation process in the E-morphotype bacteroids. Perhaps some peptides interfere with bacteroid metabolism in order to channel bacterial metabolism towards the nitrogen fixation process. For example, the accumulation of the storage compound polyhydroxybutyrate (PHB), which takes the host-supplied carbon away from nitrogen fixation, is very frequently observed in U-morphotype bacteroids but not in the E-morphotype bacteroids of IRLC legumes. Inhibition of PHB accumulation in those bacteroids could be a direct or indirect consequence of the NCR peptides.

Additional, non-exclusive hypotheses can be put forward as to explain a better performance of E- or S-morphotype bacteroids. Polyploidy of the bacteroids may support higher metabolic activity of the cells in a general way as it is the case for eukaryotic cells (Kondorosi and Kondorosi, 2004). E- and S-morphotype bacteroids are present as a single bacterium per symbiosome, which may permit a very efficient nutrient exchange with the host cell. On the contrary, multiple U-morphotype bacteroids are present in a single symbiosome and these bacteroids have therefore a more limited and less efficient contact with the symbiosome membrane.

Besides the better performance of differentiated bacteroids, other benefits to the host plant could result from the bacterial differentiation. Polyploid bacteroids could be more resistant to DNA damaging oxidative stress because of the multiple gene copies. Since bacteroids are metabolically very active, oxidative stress could affect the longevity of bacteroids, which might be enhanced by the polyploidy. Moroever, bacteroids in *M. truncatula*, which are fragilized by the NCR challenge, are efficiently digested by the host during the senescence process, at the end of the nodule life (Vasse *et al*., 1990; Van de Velde *et al*., 2006). This is in contrast to reversibly differentiated bacteroids in *L. japonicus*, which can efficiently survive nodule senescence (Müller *et al*., 2001). During nodule senescence in *M. truncatula*, catabolic genes for macromolecule degradation and a variety of transporters for the mobilization of small molecules are activated, indicating that the degradation products generated during senescence are recycled as nutrients for the plant (Van de Velde *et al*., 2006). Thus degradation of weakened bacteroids may provide supplementary nutrients and be an advantage to the host plant.

Even if a variety of effects is supposed for the NCR peptides, their extreme diversity remains astonishing. Another *raison d'être* for the high diversity of NCR peptides could be an adaptation to the high diversity of rhizobia in soils. The diversifying selection that has shaped the NCR family (Alunni *et al*., 2007; Nallu *et al*., 2014) is compatible with such a hypothesis. Additional factors that may come into play are balancing effects of certain NCRs to counteract cationic, antimicrobial NCR peptides, expression of NCR subclasses at different stages of host cell differentiation (Mergaert *et al*., 2003) and a large number of genes are expressed at low or very low levels (Mergaert *et al*., 2003) suggesting that they have less important roles.

6. Other nodule peptides interacting with bacteroids

The GRP and SNARP peptides are two other families of secretory peptides produced by symbiotic nodule cells (Kevei *et al*., 2002; Alunni *et al*., 2007; Laporte *et al*., 2010; Maunoury *et al*., 2010) and are therefore most likely transported to the symbiosomes. Moreover, they are encoded by TRGs. GRPs have a similar distribution as the NCRs and are restricted to the IRLC legumes (Küster *et al*., 1995; Schröder *et al*., 1997; Kevei *et al*., 2002; Alunni *et al*., 2007). At present the SNARPs were only detected in *M. truncatula*. They are not present in legumes outside of the IRLC clade, e.g. in *L. japonicus*, soybean or *P. vulgaris* for which rich genomic and transcriptomic data exist, but insufficient data are available to conclude on their presence in other IRLC legumes. The taxonomic distribution of these genes and their likely localization in the symbiosomes suggest that they have, like the NCRs, a role in a bacteroid-related process which is specific for the IRLC legumes. The SNARP1 and SNARP2 peptides were identified as RNA-binding peptides in a yeast three-hybrid screen and the SNARP2 peptide has *in vitro* RNA-binding capacity. Knocking-down the SNARP2 gene by RNA interference (RNAi) resulted in non-functional nodules in which bacteroids were formed but could not be maintained. Further work will be required to determine how this *in vivo* function is related to the RNA-binding capacity of these peptides. In contrast, the symbiotic function of the GRP peptides remains to be analyzed.

7. Antimicrobial peptides in other symbiotic systems

The symbiotic nodule cells show a striking similarity with symbiotic cells, called bacteriocytes, formed by aphids and other insect groups. These insects exploit restricted nutritional resources such as plant sap and their bacterial symbionts produce the nutrients, which are lacking in the host's diet. The insect bacteriocytes are organized in specific symbiotic organs called bacteriomes. Bacteriocytes are large polyploid cells, virtually filled with intracellular bacteria (**Fig. 3**). Even more striking are the morphological modifications these endosymbionts are undergoing inside the host cell such as formation of spheres (**Fig. 3**) or elongated cells similar to the S-morphotype or E-morphotype bacteroids (Shigenobu and Wilson; The International Aphid Genomics Consortium, 2010; Tsuchida *et al*., 2010; Login *et al*., 2011).

The pea aphid *Acyrthosiphon pisum* is one of the best studied insects in respect to bacterial symbiosis. *A. pisum* is in fact a species complex consisted of multiple biotypes each adapted to a specific legume species (e.g. pea, clover or *Medicago*). The primary (obligate)

symbiont of *A. pisum* is *Buchnera aphidicola*, which produces essential amino acids lacking in the plant sap diet of the aphid. RNA-seq transcriptome analysis of bacteriocytes and *in situ* hybridization experiments in *A. pisum* has led to the discovery of a class of genes encoding secreted cysteine-rich peptides, which were called BCR peptides for "bacteriocytes-specific cysteine rich peptides", in analogy to the NCR nodule peptides (Shigenobu and Stern, 2012). The genes are specifically expressed in the bacteriocytes and not in other parts of the insect. Following their expression by *in situ* hybridization during bacteriome formation in embryos, it was found that they are induced when the early-stage bacteriome becomes first infected with *Buchnera*. Thus, drawing the analogy with the NCR peptides in nodules it is an inevitable and exciting hypothesis that the BCRs are targeted to and affect the bacterial symbionts in a similar fashion as the NCRs do in symbiotic nodule cells.

In a different insect group, in the stinkbugs or *Heteroptera*, symbiotic bacteria are not located in bacteriomes as intracellular endosymbionts but in a specialized, posterior region of the midgut where numerous crypts harbor in their lumen a single and specific extracellular bacterial symbiont (Kikuchi, 2009). The stinkbug *Riptortus pedestris* which is another legume pest, particularly of soybean, carries in its crypts a symbiont belonging to the genus *Burkholderia*, a soil bacterium which is acquired every generation from the environment during a specific developmental window coinciding with the development of the crypts in the midgut (Kikuchi *et al*., 2011). A transcriptome analysis of symbiotic (crypt-carrying) and non-symbiotic midgut regions from *Burkholderia*-infected and uninfected (aposymbiotic) *R. pedestris* identified among the most abundantly expressed genes in the midgut 97 different cysteine-rich secretory peptides many of which were specifically expressed in the *Burkholderia*-containing crypts of the midgut (Futahashi *et al*., 2013). The authors speculated that these crypt-specific cysteine-rich peptides (CCRs) are secreted into the lumen of the crypts by the epithelial cells and act on the proliferation and physiology of the symbiotic *Burkholderia*, similarly to the NCRs in legume nodules and the BCRs in aphid bacteriomes (Futahashi *et al*., 2013).

Another recent study in the beetle *Sitophilus* identified the specific transcription of an AMP gene in bacteriocytes (Anselme *et al*., 2008). The encoded peptide, called coleoptericin-A (ColA), is unrelated to the legume NCRs, aphid BCRs or stinkbug CCRs and has no cysteine residues. ColA is targeted to the endosymbionts and regulates the formation of the elongated, filamentous morphology of these endosymbionts (Login *et al.*, 2011).

Moreover, AMPs are not only employed in the maintenance of endosymbiotic bacteria but also of the gut microbiota. Studies on gut microbiota in *Hydra*, *Drosophila* and mouse

(**Fig. 3**) demonstrated that in these very divergent animals the profile and the level of expression of AMPs produced by the gut epithelia are crucial to create and maintain a specific and favorable composition of the microbiota (Wehkamp *et al*., 2005; Ryu *et al*., 2008; Fraune *et al*., 2010; Salzman *et al*., 2010; Vaishnava *et al*., 2011).

Thus the legume symbiosis and the animal symbiotic systems constitute together a striking case of convergent evolution between plants and animals. This observation suggests that the employment of AMPs is an optimal *modus operandi* for a eukaryotic host cell to tolerate large numbers of intracellular bacteria.

All current data point towards a restriction of the *NCR* gene family to the IRLC legumes. However morphologically differentiated bacteroids are also observed in other legume clades (Oono *et al*., 2010) such as the E-morphotype and S-morphotype bacteroids in *Aeschynomene* species (**Fig. 6**), which are also induced by plant factors (Bonaldi *et al*., 2011). Since these species have most likely no *NCR* genes of the type found in the IRLC legumes, these host factors are different but in light of the recurrent use of AMPs to control symbiotic bacteria, they are possibly also AMP-like. In my here presented thesis work, I have participated in the characterization of a gene family in *Aeschynomene* legumes encoding such peptides. Also actinorhizal plants such as *Alnus glutinosa*, *Casuarina glauca* and *Datisca glomerata* produce nodule-specific cysteine-rich peptides (Hocher *et al*., 2011; Demina *et al.,* 2013) and ongoing research aims at verifying the hypothesis that these peptides are implicated in the control of the *Frankia* endosymbionts in the nodules of those plants.

Together, these recent studies demonstrate that although AMPs are commonly linked with innate immunity and the clearance of infecting microbes, symbiotic bacteria are equally challenged with AMPs. The recruitment of these immune effectors has in all likelihood been a crucial evolutionary step in efficiently maintaining large symbiotic bacterial populations.

8. BacA is required in legumes forming E- or S-morphotype bacteroids but not in other legumes

As described above, NCRs are similar to the defensin-type of AMPs and several NCR peptides have been found to have antimicrobial activity, killing many microorganisms including gram-positive and gram-negative bacteria as well as fungi at similar concentrations as other known AMPs. (Tiricz *et al*., 2013; Ördögh *et al*., 2014; Balogh *et al*., 2014). Also *Rhizobium* bacteria including *S. meliloti* are killed by NCR peptides when applied at high concentration. AMPs are mostly cationic (basic) peptides although some anionic (acidic) peptides are also known (Zasloff, 2002; Brogden, 2005; Maróti *et al*., 2011). The peptides in

the NCR family range from acidic to basic. Strikingly, only the basic peptides were found to have antimicrobial activity.

Thus NCRs are a two-edged sword with dramatically different effects on *S. meliloti in vivo* and *in vitro*: on the one hand NCR-challenged bacteroids in nodule symbiosomes differentiate and maintain, despite their growth inhibition, an active metabolism for nitrogen fixation but on the other hand, *in vitro* NCRs are detrimental and kill the rhizobia. The *in vivo* conditions are significantly different from an *in vitro* treatment of rhizobia with a pure peptide. Several tens to hundreds of different peptides are accumulating in symbiosomes, each probably at low concentration and with possibly antagonistic effects. Moreover, the *in vivo* effect of the NCR peptides on the bacteroids may be further influenced by particular physiological conditions in the symbiosomes such as low free oxygen and low pH in such a way that the bacteroids remain alive, although with a complete loss of their reproductive capacity.

In addition, the rhizobia may have defense mechanisms to counteract the antimicrobial activity of the NCR peptides and one important mechanism requires the BacA protein. The first rhizobial *bacA* gene was identified more than 20 years ago, long before NCR peptides were known, in a screen of *S. meliloti* mutants inducing ineffective (Fix^-) nodules on *M. sativa* (Glazebrook *et al*., 1993). In *Medicago* nodules formed by the *S. meliloti bacA* mutant, the bacteria infect the plant cells but do not elongate into the typical E-morphotype shape and the rod-shaped bacteria in the symbiosomes quickly degenerate and lyse (**Fig. 9 and 10**) (Glazebrook *et al*., 1993; Maunoury *et al*., 2010; Haag *et al*., 2011). The gene's name reflects this role in bacteroid differentiation.

More recently, *bacA* genes in other *Rhizobium* species were also investigated. It was demonstrated that the *bacA* gene of *Mesorhizobium huakuii* and 2 different *Rhizobium leguminosarum* bv. *viciae* strains was required for the formation of Fix^+ nodules on the legumes *Astragalus sinicus* and *Pisum sativum*, respectively (Tan *et al*., 2009; Karunakaran *et al*., 2010). Both legumes are closely related to *Medicago*, belonging to the IRLC legumes (Wojciechowski *et al*., 2004) and forming E-morphotype bacteroids. Histological analysis of *Pisum* nodules revealed that the *bacA*-deficient *R. leguminosarum* bv. *viciae* did not differentiate into bacteroids and could not survive in the symbiotic nodule cells but lysed after their release from the infection threads (Karunakaran *et al*., 2010), a phenotype similarly to the *S. meliloti* mutant in *Medicago*.

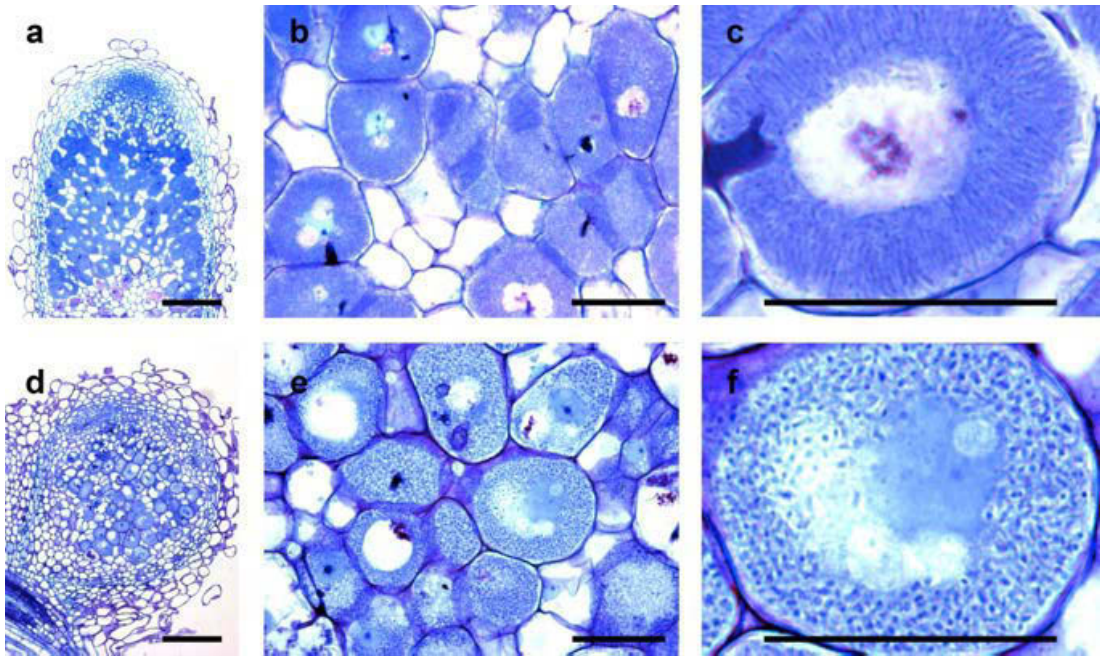

Figure 9. BacA is required for bacteroid differentiation in *Medicago*

Semi-thin sections stained with toluidine bleue are shown of *M. truncatula* nodules infected with wild-type *S. meliloti* **(a-c)** or its *bacA* mutant (**d-f). (a,d)** Whole nodules. **(b,e)** Enlargement of a region containing symbiotic nodule cells. **(c,f)** Close-up of a single symbiotic cell, showing the elongated (E-morphotype) bacteroids in the wild-type infected cell **(c)** and the non-differentiated bacteria in the cell infected with the *bacA* mutant **(f).** Scale bars are 200 µm **(a,d)** or 50 µm **(b,c,e,f).**

In contrast, the *bacA* genes of *Mesorhzobium loti*, *Rhizobium leguminosarum* bv. *phaseoli*, *Rhizobium etli* or *Sinorhizobium fredii* strain NGR234 are dispensable for the formation of a functional symbiosis on various plants like *L. japonicus*, *P. vulgaris*, *Lablab purpureus*, *Leucaena leucocephala*, *Tephrosia vogelii*, and *Vigna unguiculata* (Maruya and Saeki, 2010; Karunakaran *et al*., 2010; Ardissone *et al*., 2011). Contrary to the IRLC legumes, all these legumes form U-morphotype bacteroids and lack NCR peptides.

An eloquent example for the *bacA* requirement in symbiosis is provided by the *R. leguminosarum* bv. *viciae* strain A34 and the *R. leguminosarum* bv. *phaseoli* strain 4292 which have the same genetic background except for their symbiotic plasmids. These strains have opposing requirements for the *bacA* gene in symbiosis with *P. sativum* which requires *bacA* and forms E-morphotype bacteroids and with *P. vulgaris* which does not require *bacA* and forms U-morphotype bacteroids (Karunakaran *et al*., 2010).

Bradyrhizobium species have a gene that is only distantly related to *bacA*, classified as *bacA*-like or *bclA*. We have found that *bclA* is required in *Bradyrhizobium* sp. for effective symbiosis on *A. indica* or *A. afraspera* whose bacteroids have the S- or E-morphotype, respectively. Mutants in the *bclA* gene do not differentiate and remain rod-shaped in the

symbiosomes (see **Chapter III**; Guefrachi *et al.*, 2015a). Strikingly, in *Bradyrhizobium japonicum*, the orthologous gene is not required for symbiosis with *G. max*, which forms U-morphotype bacteroids (Guefrachi *et al.*, 2015a). Taking together, these examples give ample evidence for a correlation between the need of the *bacA*/*bclA* genes for symbiosis and the formation of E- or S-morphotype bacteroids while the gene is dispensable when U-morphotype bacteroids are formed.

9. The bacterial BacA protein provides protection against the antimicrobial activity of NCRs

NCR peptides induce E-morphotype bacteroid formation in IRLC legumes. It is also known that NCR-like peptides are involved in bacteroid differentiation in *Aeschynomene* legumes (see **Chapter II**; Czernic *et al*., 2015), the requirement of BacA or BclA proteins only for E- or S-morphotype bacteroid formation suggests that this protein plays a central role in the rhizobial response towards the NCR peptides (Kereszt *et al*., 2011). Two hypotheses were put forward that could take into account the requirement of BacA in *S. meliloti* (Haag *et al*., 2011). The protein could be directly required for the bacterial metamorphosis. However, synthetic NCR peptides induced *in vitro* bacterial elongation and DNA amplification to the same extent in the *S. meliloti bacA* mutant as in the wild-type strain suggesting that BacA is not required for the differentiation process *per se* (Haag *et al*., 2011). The second tested possibility is related to the above described antimicrobial activity of NCR peptides. The killing activity of the NCRs is related to the damage they provoke on the bacterial membrane, in a similar way as the defensins (Van de Velde *et al.*, 2010; Tiricz *et al.*, 2013). Importantly, the *S. meliloti bacA* mutant was between 10- to 100-fold more sensitive towards the antimicrobial action of NCR peptides (Haag *et al*., 2011; Haag *et al*., 2012). The hypersensitivity of the *bacA* mutant was not specific for NCRs but applied also to the human defensin, β-defensin 2 (Arnold *et al*., 2013). This hypersensitivity provided a likely explanation for the mutant's symbiotic phenotype and suggested that BacA is required to survive the challenge with NCRs in the symbiotic nodule cells. Testing this hypothesis *in vivo*, in *M. truncatula* nodules, it was first shown that the *S. meliloti bacA* mutant bacteria, like wild-type, are challenged with NCR peptides once they are released from infection threads in the symbiotic nodule cells (Haag *et al*., 2011). With a live/dead staining procedure which marks bacteria with a different dye according to whether they are alive or dead, it was found that *bacA* mutant bacteria are healthy in the infection threads but are rapidly killed after their uptake in symbiotic cells, while a wild-type strain remained, as expected, viable and

differentiated. Moreover, the *bacA* mutant was able to survive in the symbiotic nodule cells of the *M. truncatula dnf1* mutant (Haag *et al*., 2011) which is affected in a nodule-specific signal peptidase complex (Wang *et al*., 2010) and which cannot transport NCR peptides to the symbiosomes (Van de Velde *et al*., 2010). These observations are consistent with a role of BacA in preventing bacterial death by the potentially toxic NCR peptides and permitting the bacterial differentiation (**Fig. 10**). Perhaps, cell elongation and genome amplification is a secondary effect of a bacterial defense response towards the antimicrobial activity of the peptides.

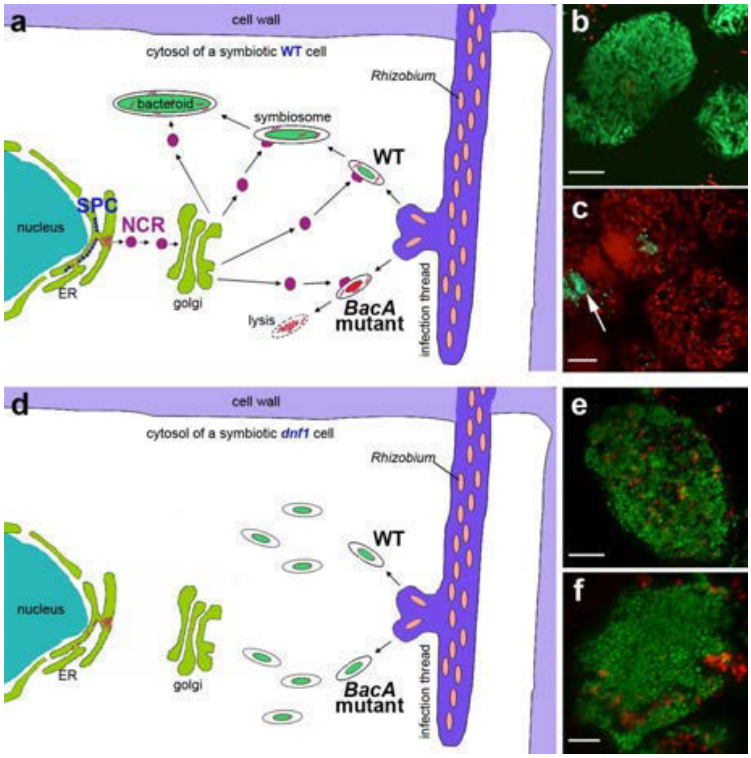

Figure 10. BacA is required for survival in the symbiotic nodule cells

(a) In a *M. truncatula* wild-type nodule cell, the *S. meliloti* wild-type (WT) and *bacA* mutant bacteria are released from the infection threads into membrane-bound symbiosomes. The bacteria are then challenged with host NCRs (purple), which are synthesized on the endoplasmic reticulum (ER) and transported to the symbiosomes via vesicle transport and a symbiotic nodule-cell-specific secretory pathway. This pathway requires an ER-located nodule-specific signal peptidase complex (SPC) containing the DNF1 subunit. The NCRs mediate the differentiation of the wild-type strain into a nitrogen-fixing bacteroid (green bacteroids), resulting in the chronic infection. In contrast, the *bacA*-deficient mutant is hypersensitive towards the host NCRs and is killed and then lysed (red symbiosome bacteria) rather than forming the chronic infection. **(b)** Wild-type *S. meliloti* in wild-type host cells with live/dead staining marking live bacteria in green and dead bacteria in red. The wild-type *S. meliloti* are healthy and differentiated. **(c)** Staining as in **(b)** of *bacA* mutant *S. meliloti* in wild-type host cells. The bacteria are non-differentiated and dead except for the infection thread-located ones (arrow). **(d)** In a *M. truncatula dnf1* mutant nodule cell, the *S. meliloti* wild-type (WT) and *bacA* mutant bacteria are released from the infection threads into membrane-bound symbiosomes. The NCRs are not transported to the symbiosomes but remain blocked in the ER because of the absence of the DNF1 SPC subunit. Therefore both the wild-type and *bacA* mutant bacteria remain alive (green) within the symbiosomes but do not differentiate. **(e)** Staining as in **(b)** of wild-type *S. meliloti* in *dnf1* mutant host cells. The bacteria are alive but non-differentiated. **(f)** Staining as in **(b)** of *bacA* mutant *S. meliloti* in *dnf1* mutant host cells. The bacteria are alive but non-differentiated. Scale bars are 10 µm.

By synthesizing alternative versions of the NCR247 peptide with different disulfide bridge configurations between its 4 cysteine residues, it was demonstrated that the peptide with disulfide bridges between cysteines 1-2 and 3-4 was significantly more active than the peptide with disulfide bridges between cysteines 1-3 and 2-4 for both induction of cell elongation and DNA amplification at low peptide concentrations and for antimicrobial activity at higher peptide concentrations (Haag *et al*., 2012). Furthermore, substitution of the cysteine residues with serines diminished the antimicrobial potency of NCR247 while chemically reducing the peptide by dithiothreitol treatment increased its antimicrobial potency significantly (Haag *et al*., 2012). This is similar to the behavior observed for the human β-defensin 1, which is produced in the colon and thought to be stored in an inactive oxidized form and to be activated by reduction through the action of thioredoxin proteins (Schroeder *et al*., 2011). Thioredoxins are ubiquitous redox proteins, which regulate the oxidation state of cysteine residues and disulfide bridges in proteins. Intriguingly, *Medicago* nodules express a pair of unusual thioredoxins, Trx s1 and Trx s2 (Alkhalfioui *et al*., 2008). They are nodule specific and are thus co-regulated with *NCR* genes. These thioredoxins differ from classical thioredoxins by the presence of an N-terminal signal peptide for the secretory pathway, like the NCR peptides, which indicates that they are potentially co-transported with the NCRs to the symbiosomes. Thus an interesting hypothesis is that these thioredoxins regulate the oxidative state of the peptides and thereby control their activity.

Importantly, no difference in sensitivity against either the chemically reduced NCR247 peptide or the linear peptide in which the cysteine residues were replaced with serine was observed between wild-type *S. meliloti* and the *bacA* mutant (Haag *et al*., 2012). This suggests that the natural form of the NCR peptides is the oxidized one and raises the possibility that BacA can directly or indirectly influence the oxidation state of the NCR peptides in the periplasm of the bacteroids and promote the formation of the less detrimental oxidized state, thereby limiting damage of the inner membrane.

10. BacA homologues are required in pathogens for chronic infection

Besides its role in intracellular accommodation of rhizobia during symbiosis with legume plants, BacA or BclA homologues have been identified and characterized in several pathogenic bacteria, in which they play critical roles during the interaction with their eukaryotic host. Notably, functional studies have been performed on BacA homologues from

avian pathogenic *E. coli* (APEC) and in the mammalian pathogens *Mycobacterium tuberculosis* and *Brucella abortus*.

APEC can colonize the respiratory tract and cause pneumonia in poultry, and in addition, they are also often associated with internal organ infections. In its most acute form, colibacillosis caused by APEC triggers septicemia, often resulting in sudden death of infected birds (Dziva and Stevens, 2008). As other pathogenic *E. coli*, APEC remains exclusively extracellular and the pathogen resides in the host tissues without penetrating the cells. In a signature-tagged mutagenesis screen for altered pathogenicity on chicken, a mutant in the BacA homologue (in *E. coli* named SbmA) was identified. The mutant was shown to be attenuated in the colonization of different internal organs as compared to the wild-type strain, indicating that BacA is required for survival in the host and establishment of chronic infection even in the context of a non-intracellular pathogen (Li *et al*., 2005).

Brucella abortus is a zoonotic pathogen that causes brucellosis (Malta fever) in humans and spontaneous abortions in animal hosts like cattle. It infects its host intracellularly through an endocytotic process and survives in phagocytic cells such as macrophages (Roop *et al*., 2009). A *B. abortus* BacA homologue has been identified and the corresponding mutant displayed reduced intracellular survival in macrophages compared to the wild-type strain. Moreover, bacterial clearance from infected BALB/c mice was increased in the *bacA* mutant by five orders of magnitude at 8 weeks post inoculation. Taken together, these results indicate that, similarly to the *Rhizobium*-legume symbiosis, BacA is required for intracellular accommodation of *Brucella* and the establishment of a chronic infection (LeVier *et al*., 2000). *Mycobacterium tuberculosis,* the causal agent of tuberculosis in humans, is a respiratory pathogen that colonizes the lung through intracellular infection. The bacterium can enter alveolar epithelial cells and macrophages and survive despite the high levels of defensins and other AMPs produced in these cells (Tan *et al*., 2006; Rivas-Santiago *et al*., 2005; Rivas-Santiago *et al*., 2006). In a murine infection model, *M. tuberculosis* can reside in the lung and spleen for months and lead subsequently to host death. Although the *M. tuberculosis bacA* mutant was not compromised in infection of both organs over more than two hundred days, it induced significantly retarded host mortality compared to the wild-type strain in a long term time-to-death experiment (Domenech *et al*., 2009).

Conservation of BacA function across bacterial species has been evaluated in different combinations of trans-complementation assays. Expression of the BacA homologues of either *E. coli* or *B. abortus* in the *S. meliloti bacA* mutant was able to complement its symbiotic

defects and restored the formation of functional, nitrogen-fixing nodules on *M. sativa* (Ichige and Walker, 1997; Wehmeier *et al*., 2010).

Even the *M. tuberculosis* BacA homologue rescued the *S. meliloti bacA* mutant phenotype on *M. sativa*, although only partially. In this case, nodules were not functional and did not support growth of the plants but the nodules were much more infected than in the case of the mutant suggesting that the bacteria persisted in the nodules contrary to the *bacA* mutant, which is quickly degraded in the plant cells. Moreover, *bacA* mutant bacteria carrying the *M. tuberculosis* gene were enlarged although they displayed aberrant morphologies and did not become elongated as wild-type bacteroids (Arnold *et al*., 2013).

These examples from human and animal pathogens highlight the requirement of a functional BacA protein for inter- and intracellular accommodation of bacteria within eukaryotic cells and tissues and for maintenance of chronic infections. Similarly to *S. meliloti* dealing with NCRs in *Medicago* nodules, these pathogens colonizing host tissues are facing the host innate immune system comprising AMPs like defensins (Roop *et al*., 2009; Tan *et al*., 2006; Rivas-Santiago *et al*., 2005; Rivas-Santiago *et al*., 2006). It is tempting to speculate that the BacA proteins of the pathogens are important to resist to these AMPs and are therefore key in the establishment of chronic infections. In agreement with this possibility, it was shown that the *M tuberculosis bacA* mutant has impaired tolerance to human β defensin 2 (Arnold *et al*., 2013). Moreover, the *Mycobacterium* and *Brucella bacA* homologues can restore the sensitivity towards NCRs or human β defensin 2 of the *S. meliloti bacA* mutant to wild-type levels (Haag *et al*., 2011; Arnold *et al*., 2013).

11. BacA is an ABC transporter. But what does it transport?

S. meliloti BacA and *E. coli* SbmA proteins have 8 transmembrane-domains but they lack any other identified functional domain. However, a class of BacA/BclA proteins have an additional ATPase domain. The above mentioned *Bradyrhizobium* and *Mycobacterium* genes belong to this class but also *S. meliloti* and *E. coli* have besides *bacA* or *sbmA* an additional *bacA*-like gene, named *exsE* and *yddA* respectively. Albeit they are divergent (LeVier and Walker, 2001), these ATPase containing proteins show large regions of similarity with BacA and SbmA. In particular, transmembrane predictions revealed that all members have 6 transmembrane domains at conserved positions in the protein, shared with BacA and SbmA, suggesting that these proteins have similar topologies. Importantly, the C-terminal ATPase domain of these BacA proteins is typical for ATP-binding cassette (ABC) transporters

(LeVier and Walker, 2001). The function of the *M. tuberculosis* BacA protein is abolished by deletion of its ATPase domain or by introduction of a point mutation in a conserved residue in this domain, which is known to be important for transport activity in other ABC transporters (Arnold *et al*., 2013). This observation suggests that BacA-like proteins indeed function as ABC proteins. Thus potentially also the BacA/SbmA proteins are ABC transporters. Indeed, the *E. coli* SbmA protein forms homodimers with a topology closely resembling the membrane-spanning region of the ABC transporter family (Corbalan *et al*., 2013; Runti *et al*., 2013). A putative ATPase domain of SbmA or BacA could be encoded by a separate gene but remains to be identified. No candidate gene has been identified in the vicinity of the *bacA*/*sbmA* genes. Alternatively, it was proposed that SbmA-mediated transport is driven by the electrochemical gradient across the inner-membrane of the bacterial cell in which the transporter is located rather than by ATP hydrolysis (Runti *et al*., 2013).

Certain AMPs like microcin B17, microcin J25 and proline-rich cathelicidin type of AMPs are internalized and have intracellular targets leading to bacterial death. These AMPs kill bacteria without affecting the bacterial membrane integrity, in contrast to NCRs or defensins (Scocchi *et al*., 2011). The uptake of these peptides is mediated at least in part by the BacA/SbmA/BclA proteins and mutations strongly reduce peptide uptake and thus confer resistance to the peptides. This has led to the suggestion that BacA/SbmA/BclA proteins are peptide transporters (Laviña *et al*., 1986; Yorgey *et al*., 1994; Salomon and Farias, 1995; Mattiuzzo *et al*., 2007; Pränting *et al*., 2008; Marlow *et al*., 2009; Karunakaran *et al*., 2010; Ardissone *et al*., 2011; Arnold *et al*., 2013; Guefrachi *et al*., 2015a; see **Chapter III**). Thus the BacA/SbmA transporters have opposing effects on the bacterial sensitivity towards peptides with intracellular targets like the proline-rich cathelicidins (BacA/SbmA promotes sensitivity) and towards membrane-disrupting peptides like NCRs or defensins (BacA promotes resistance). A possible explanation is that the internalization of AMPs by the transporter facilitates their contact with their target for AMPs with intracellular action such as the proline-rich cathelicidins and on the contrary, takes them away from their membrane targets for membrane-disrupting AMPs such as the NCRs or defensins. It is interesting to note that other ABC transporter systems are described which provide resistance to membrane-disrupting AMPs. The SapABCDF transporters in *S. typhimurium* (Para-Lopez *et al*., 1993; Para-Lopez *et al*., 1994) or the plant pathogen *Dickeya dadantii* (previously named *Erwinia chrysanthemi*) (Lopez-Solanilla *et al*., 1998) and the *Salmonella* YejABEF transporter (Eswarappa *et al*., 2008) were proposed to import host-produced membrane-damaging AMPs taking them away from their membrane target site.

The BacA/SbmA/BclA proteins also mediate the uptake and sensitivity to the glycopeptide antibiotic bleomycin (Yorgey *et al*., 1994; Ichige and Walker, 1997; Marlow *et al*., 2009; Wehmeier *et al*., 2010; Domenech *et al*., 2009; Tan *et al*., 2009; Karunakaran *et al*., 2010; Maruya and Saeki, 2010; Guefrachi *et al*., 2015a; see **Chapter III**) and sensitivity towards several aminoglycoside antibiotics (Ichige and Walker, 1997). Moreover, the *M. tuberculosis* BacA-like protein was recently reported to facilitate vitamin B_{12} uptake (Gopinath *et al*., 2013). Although vitamin B_{12} is required for bacteroid differentiation (Campbell *et al*., 2006; Taga and Walker, 2010), it is unlikely that the symbiotic role of BacA or BacA-like is the uptake of plant-provided vitamin B_{12} because rhizobia can produce this cofactor and more importantly, plants do not use or produce it (Roth *et al*., 1996).

Since aminoglycoside uptake is driven by membrane potential (Taber *et al*., 1987) and in light of the structural diversity of all the compounds whose uptake is affected by the BacA/SbmA/BclA proteins, an alternative to the peptide uptake hypothesis could be that the proteins affect in a direct or indirect way the membrane potential or membrane composition. An altered membrane could have changed interaction properties with membrane-damaging peptides like NCRs or defensins.

Support for a role of BacA in the determination of cell envelope features derives from the observations that the *S. meliloti* mutant has an increased sensitivity to detergents and ethanol (Ferguson *et al*., 2002). Similar behavior was described for *bacA* mutants in *M. huakuii* (Tan *et al*., 2009), *M. loti* (Maruya and Saeki, 2010), *R. leguminosarum* or *R. etli* (Karunakaran *et al*., 2010). The *bacA* mutation in *S. meliloti* and *B. abortus* also reduces by about half the modification of the outer-membrane lipopolysaccharide (LPS) with typical very long chain fatty acids (VLCFA) (Ferguson *et al*., 2004). Therefore, it was proposed that BacA could be involved in the export of fatty acids out of the cytoplasm where they are synthesized (Ferguson *et al*., 2004). Also in *M. huakuii*, the *bacA* mutation affected the VLCFA content in LPS (Tan *et al*., 2009). However, no major changes associated with the cell envelope, including sensitivity to ethanol and detergents and lipid content of cell membranes, were detected in *bacA* deficient *Sinorhizobium fredii* strain NGR234 and *M. tuberculosis* (Domenech *et al*., 2009; Ardissone *et al*., 2011). Moreover, the symbiotic defect of the *S. meliloti bacA* mutant does not seems to be directly linked to the VLCFA LPS defect because additional mutants in the *acpXL* and *lpxXL* genes, which completely lack VLCFA in their LPS, are much less severely affected than the *bacA* mutant and can form nitrogen-fixing nodules (Ferguson *et al*., 2005) although bacteroid abnormalities were observed with for example the formation of hypertrophied bacteroids (Haag *et al*., 2009).

In addition, the BacA mediated uptake of cathelicidin type of AMPs such as the peptide Bac7 by *S. meliloti* can also be uncoupled from its symbiotic role because two point mutants in the *bacA* gene were identified that were symbiotically defective but maintained the capacity to take up the Bac7 peptide (Marlow *et al*., 2009). Similarly, the *M. tuberculosis bacA*-like gene can restore the resistance defect of the *S. meliloti bacA* mutant towards the NCR247 peptide as well as human β-defensin 2. Nevertheless, the *M. tuberculosis bacA*-like gene is only partially able to support symbiosis and bacteroid development (Arnold *et al*., 2013). Probably, the multifaceted phenotypes of the *bacA* mutant together with the complex mixture of NCR peptides produced by the symbiotic nodule cells has to be taken into account to fully explain the essential role of BacA in symbiosis.

12. General Envelope Stress Response function of BacA/SbmA?

Bacteria that interact with eukaryotic cells are often confronted with AMPs produced by the host cells. Many of these bacteria have evolved resistance mechanisms permitting them to survive these peptides. Examples are the SapABCDF and YejABEF transporters described above. Another frequently observed mechanism is the modification of the electrical charge of the bacterial envelope, rendering it less negative and thereby reducing electrostatic interactions with the cationic AMPs (Maróti *et al*., 2011). Related to this, it is interesting to note that the LPS of bacteroids in *Medicago* and pea nodules becomes more hydrophobic and displays different epitopes when compared to cultured bacteria (Kannenberg *et al*., 1994; Kannenberg and Carlson, 2001; Ferguson *et al*., 2005). This LPS modification could reflect an adaptive response of the bacteroids to the exposure with the NCR peptides.

The bacterial resistance mechanisms are usually induced after perception of the AMPs. The presence of AMPs can be directly sensed with AMP-binding two-component regulators (Bader *et al*., 2005; Gryllos *et al*., 2008) or with ABC transporter-derived sensors, which in that case have lost their transporter function but have a role as receptor and signal transduction protein (Hiron *et al*., 2011; Falord *et al*., 2012). Alternatively, bacteria can sense the membrane damage provoked by the membrane-disrupting peptides through different envelope (or extracytoplasmic) stress response (ESR) systems. ESR pathways detect and induce appropriate responses to stresses that affect the bacterial membranes and periplasmic and membrane proteins, including membrane stresses provoked by AMPs (Majdalani and Gottesman, 2005; Rowley *et al*., 2006; Bury-Moné *et al*., 2009). Generally, bacteria have different co-existing ESR systems. AMP-responsive ESRs include, in *E. coli* and *Salmonella typhimurium*, the Rcs phosphorelay system (Farris *et al*., 2010), the CpxAR two-component

system (Weatherspoon-Griffin *et al*., 2011) and the extracytoplasmic function sigma factor σ^E-regulon (Hymphreys *et al*., 1999; Crouch *et al*., 2005).

Interestingly, *SbmA* is part of the ESR σ^E-regulon in *E. coli* (Rowley *et al*., 2006; Bury-Moné *et al*., 2009) suggesting that the SbmA protein has a role in ESR, which needs to be further defined. σ^E and its homologues control to various degrees pathogenesis. For example, σ^E mutants of *S. typhimurium* are defective for intracellular survival in macrophage and epithelial host cell lines and have a strongly attenuated virulence in mice (Humphreys *et al*., 1999; Crouch *et al*., 2005). The attenuated virulence of the σ^E-mutant of *S. typhimurium* has been attributed to its inability to protect itself against host-produced AMPs (Crouch *et al*., 2005). The mutant phenotypes are in part depending on the σ^E-regulated genes *htrA* (*degP*) and *surA*, encoding a periplasmic serine protease and a periplasmic peptidyl-prolyl-isomerase, respectively. But other σ^E-regulated genes are possibly important since σ^E mutants are more affected than individual *htrA* or *surA* mutants. A good candidate among them could be *sbmA* in light of the above described critical function of SbmA/BacA proteins in symbiotic and pathogenic interactions.

Interestingly, a genetic link between *bacA* and *degP* was also demonstrated in *S. meliloti* (Glazebrook *et al*., 1996) suggesting that also in *S. meliloti*, *bacA* is implicated in ESR. However, the biochemical link between *bacA* and *degP* remains to be further unraveled.

The association of SbmA in *E. coli* and potentially also BacA in *S. meliloti* with ESR suggests that these proteins are part of a membrane stress sensing system or that they function in the bacterial response to alleviate this stress. The above-mentioned hypersensitivity of *bacA* mutants towards the membrane-damaging stresses induced by ethanol or detergents are in agreement with this.

The σ^E-regulon in *E. coli* includes also the *dsbC* gene (Rowley *et al*., 2006; Bury-Moné *et al*., 2009), which is involved in oxidative protein folding in the periplasm (Depuydt *et al*., 2011). It is unknown whether SbmA is also involved in oxidative folding but a speculative link could be that BacA keeps NCR peptides in an oxidized and less damaging state in the periplasm as is suggested above (Haag *et al*., 2012).

ESR regulators have been described in several *Rhizobium* species (Bastiat *et al*., 2010; Sauviac *et al*., 2007; Gourion *et al*., 2009; Martínez-Salazar *et al*., 2009). But none of them were reported to regulate the *bacA* gene and thus the regulon to which rhizobial *bacA* belongs remains to be discovered but could be very informative for understanding better how BacA functions, both as a household function and in symbiosis.

13. Hypotheses for the mechanism of BacA/BclA functioning

Recent studies on BacA and BclA in *Rhizobium* species indicate that these transporters are critical for chronic intracellular infections of legume hosts, which produce AMPs in the symbiotic cells. BacA homologues are also contributing to the pathogenesis and chronic infection in pathogens and they do this most likely by protecting bacteria against host AMPs. However, how BacA/SbmA proteins function mechanistically remains unknown. We propose the following hypotheses: **i)** the proteins could be part of a peptide uptake system, leading to internalization of AMPs which would move them away from membrane targets but bring them in proximity to intracellular targets; **ii)** on the contrary, it cannot be excluded that they are exporters, affecting the bacterial envelop and thereby indirectly promoting the uptake of certain peptides with intracellular mode of killing and reducing the affinity of membrane-damaging peptides for the bacterial envelope or **iii)** alternatively, they could export compounds that directly modify the peptides (e.g. oxidizing peptides so that they are less harmful); **iv)** they could function as sensors rather than transporters, activating other resistance determinants. Testing these possibilities is a challenge for the future. Understanding the household function of *bacA* in *Rhizobium* and its regulation in stress response or other cellular functions could provide important clues for understanding its interference with the action of NCR peptides and other AMPs.


Objectives

Objectives and outline of the PhD work

The host laboratory of my PhD work, the group "Cell Differentiation in the Legume-Rhizobium Symbiosis", has a long-standing interest in legume-*rhizobia* interaction from both the bacterial and plant perspective. The research activity in the last decade was focused on cell differentiation in nodules and notably on the differentiation of symbiotic cells as well as the bacterial endosymbionts. Among the meaningful results were the identification of *Medicago* nodulation genes and their functional analysis to unravel the molecular mechanism of symbiotic cell formation and intracelullar accomodation of rhizobia
(Cebolla *et al*., 1999; Kevei *et al*., 2002; Mergaert *et al*., 2003; Vinardell *et al* 2003; Van de Velde *et al*., 2006; Alunni *et al*., 2007; Kevei *et al*., 2007; Maunoury *et al*., 2010).

One of the primordial results from previous research of the group was the finding that bacteroids in the IRLC legume clade which form E-type bacteroids differentiate under the influence of plant derived antimicrobial peptides called the Nodule-specific Cysteine-Rich (NCR) which are targeted to the symbiosomes via the secretory pathway (Mergaert *et al*., 2006; Van de Velde *et al*., 2010; Haag *et al*., 2011). NCRs are similar to the defensin-type of antimicrobial peptides, and some NCR peptides have been found to have antimicrobial activity, killing rhizobium when applied at high concentration. To counteract the antimicrobial activity of the NCR peptides, *S. meliloti*, the symbiont of *Medicago*, requires the BacA protein, which is an ABC transporter protein that thus potentially acts as a peptide transporter.

This previous work of the group that formed the context of my PhD work at its onset and is described in detail in the introduction of this PhD manuscript and part of has been published as a review in the book entitled "Biological Nitrogen Fixation Book", edited by Frans J. de Bruijn. The article is entitled "***Bradyrhizobium* BclA is a NCR peptide transporter required for bacterial differentiation in symbiosis with *Aeschynomene***". The first objective of my PhD was to explore further the expression behavior of this amazing family of *NCR* genes in *Medicago truncatula*. The differentiation of bacteroids in the IRLC legumes to which *Medicago* belongs, is now well described. In other legumes like the Dalbergoids, bacteroid differentiation evolved most likely independently form the IRLC (Oono *et al*., 2010). The major part of my thesis work was devoted to the analysis of the mechanisms of differentiation in *Aeschynomene* species belonging to the Dalbergoids. The results I describe are divided in three main chapters.

In the first part of my thesis work (see **Chapter I**), I have exploited the transcriptome database provided by the *Medicago truncatula* Gene Expression Atlas (MtGEA) and other published microarray data to characterize in detail the expression behavior of a large set of *NCR* genes. This studied consisted mainly in a data mining and a bioinformatics analysis, complemented with a number of experimental confirmations. This work was recently published in BMC Genomics with the title **"Extreme specificity of *NCR* gene expression in *Medicago truncatula*"**.

The *NCR* gene family has been identified in all investigated legume species of the IRLC clade to control bacteroid differentiation but in no other plant species outside of this clade and the molecular mechanisms underlying bacterial differentiation in legumes with other bacteroid morphotypes remain unknown at present. I have participated during the second part of my PhD in research aiming at understanding the mechanisms implicated in differentiation of bacteroids in the Dalbergoid legume clade to which belong the *Aeschynomene* species.

In the **second Chapter** of my results, I present the article **"Convergent evolution of endosymbiont differentiation in Dalbergoid and IRLC legumes mediated by nodule-specific cysteine-rich peptides",** submitted in Plant Physiology. This paper was devoted to the study of the factors used by the *Aeschynomene* legumes to control bacteroid metamorphosis. My contribution in this work was the *in situ* hybridization experiments which demonstrated that the expression of the identified *NCR-like* genes was restricted to the infected nodule cells in *A. afraspera* and *A. indica* and the measurement of the DNA content by flow cytometry in both type of nodule harboring spherical and elongated bacteroids in *A. indica* and *A. afraspera* respectively.

Finally, in the **third Chapter**, I present **"*Bradyrhizobium* BclA is a NCR peptide transporter required for bacterial differentiation in symbiosis with *Aeschynomene*"**. In this chapter, I used an attractive model system composed of *Bradyrhizobium* strains interacting with *Aeschynomene* (*A. indica, A. evenia* and *A. afraspera*) and soybean legumes in which bacteroids have three different host-dependent morphotypes. I describe a bacterial determinant, which is required for bacteroid differentiation. Specifically, I studied the role of a gene of *Bradyrhizobium* species, named *bclA*, which is homologous to *bacA* of *S. meliloti* and is required for the formation of polyploid bacteroids in nodules of the *Aeschynomene* legumes. This work shows that not only these Dalbergoid host plants use similar effectors as the legumes of the IRLC but that also the bacteria has co-opted the same type of genes to be

able to respond to those effectors, indicating that IRLC and Dalbergoid legumes together with their cognate rhizobia underwent convergent coevolution towards similar bacteroid differentiation strategies. The results of this chapter will be submitted in two separate papers, one dealing with the *bclA* gene of *Bradyrhibium* strain ORS285 and the second one with the corresponding gene of *B. japonicum* USDA110. The first paper was accepted in MPMI and the second one is nearly ready for submission.

Results & Discussion

Chapter. I

Extreme specificity of *NCR* gene expression in *Medicago truncatula*

Extreme specificity of *NCR* gene expression in *Medicago truncatula*[*]

1. Abstract

Legumes form root nodules to house nitrogen fixing bacteria of the rhizobium family. The rhizobia are located intracellularly in the symbiotic nodule cells. In the legume *Medicago truncatula* these cells produce high amounts of Nodule-Specific Cysteine-Rich (NCR) peptides, which induce differentiation of the rhizobia into enlarged, polyploid and non-cultivable bacterial cells. NCRs are similar to innate immunity antimicrobial peptides. The *NCR* gene family is extremely large in *Medicago* with about 600 genes. Here we used the *Medicago truncatula* Gene Expression Atlas (MtGEA) and other published microarray data to analyze the expression of 334 *NCR* genes in 267 different experimental conditions. We find that all but five of these genes are expressed in nodules but in no other plant organ or in response to any other biotic interaction or abiotic stress tested. During symbiosis, none of the genes are induced by Nod factors. The *NCR* genes are activated in successive waves during nodule organogenesis, correlated with a specific spatial localization of their transcripts from the apical to the proximal nodule zones, and are not associated with nodule senescence. According to their Shannon entropy, a measure expressing tissue specificity of gene expression, the *NCR* genes are among the most specifically expressed genes in *M. truncatula*. Moreover, when activated in nodules, their expression level is among the highest of all genes. Together, these data show that the *NCR* gene expression is subject to an extreme tight regulation and is only activated during nodule organogenesis in the polyploid symbiotic cells.

[*]**Modified from: Guefrachi, I**., Nagymihaly, M., Pislariu, C.I., Van de Velde, W., Ratet, P., Mars, M., Udvardi, M.K., Kondorosi, E., Mergaert, P., Alunni, B. (2014). BMC Genomics. 15: 712

2. Introduction

Legume plants establish a symbiotic relationship with nitrogen fixing soil bacteria, known as rhizobia. For the purpose of this symbiosis, the plant host forms new, specific organs on its roots called nodules, inside which the symbiotic rhizobia are housed, fix nitrogen (i.e. the enzymatic reduction of nitrogen gas to ammonium) and transfer the ammonium to the plant. Nodules contain several thousand endoreduplicated giant symbiotic cells, which are each infected with thousands of intracellular rhizobia. These symbiotic cells are adapted to the symbiosis, to the metabolic exchange with the nitrogen fixing rhizobia and to the intracellular accommodation of this large bacterial population. The symbiotic cells originate from dividing progenitor cells in the nodule meristem. A key step in the differentiation of the symbiotic cells is the exit of the cell division cycle of nodule meristematic cells and a switch to an endoreduplication cycle in these post-meristematic cells. An endoreduplication cycle is a modified cell cycle with repeated replication of the genome without mitosis and cytokinesis, resulting in polyploid cells with increased DNA content and cell volume. The cell cycle switch is under the control of the Anaphase Promoting Complex (APC) and its activator Ccs52A (Cebolla *et al.*, 1999; Vinardell *et al*., 2003). The differentiating symbiotic cells are gradually infected and filled with rhizobia. Bacteria are released in the host cells through an endocytosis-like process releasing the intracellular bacteria in organelle-like structures called symbiosomes. Mature symbiotic cells have about 80-fold larger cell volume and endoploidy levels up to 64C compared to the diploid (2C) progenitor cells. Their cytosolic space is entirely packed with symbiosomes and their physiology is adapted for symbiosis, feeding the microsymbionts and assimilating and transporting the fixed nitrogen.

Remarkably, Wildermuth, (2010) noticed, by comparing different biotrophic interactions of plants, that host cell polyploidy is a common feature of symbiotic interactions with rhizobium bacteria and arbuscular mycorrhizal fungi, as well as parasitic interactions with fungi and nematodes. Even in symbiotic interactions of insects with endosymbiotic bacteria, host cells that house the endosymbionts are endoreduplicated cells (Koga *et al*., 2012). Thus polyploid host cells may be a well suited adaptation as an interaction site for nutrient exchange with symbiotic microorganisms and some parasites may have evolved to exploit this.

Synchronously with the differentiation of their host cells, the symbiosome bacteria differentiate into nitrogen-fixing bacteria called bacteroids. These bacteroids have a specific

physiology and metabolism adapted to the symbiotic life and nitrogen fixation, which are dramatically different from those of a free-living bacterium (Kereszt *et al*., 2011). Interestingly, the differentiation of bacteroids is often accompanied by a morphological and cytological metamorphosis whereby the bacteroid cell becomes enlarged, its envelope fragilized and its genome amplified (polyploid) and condensed Mergaert *et al*., 2006; Kondorosi *et al*., 2013). In *Medicago truncatula*, a class of peptides named NCRs (Nodule-specific Cysteine-rich Peptides) controls the bacteroid elongation and polyploidization (Van de Velde *et al*., 2010). The NCR peptides are produced by the infected symbiotic cells and are transported to the bacteroid-containing symbiosomes. The NCR peptides can induce bacterial elongation and polyploidization *in vitro* on cultured rhizobium or *in planta* when expressed in transgenic *Lotus japonicas* plants, which lack NCRs and, form non-elongated bacteroids without genome amplification (Van de Velde *et al*., 2010). Some NCRs accumulate to a significant extent in the cytosol of mature bacteroids (Van de Velde *et al*., 2010) suggesting that these peptides may have additional functions, other than inducing the morphological transformation and notably, it has been suggested that these intracellular NCRs may affect the bacteroid metabolism (Kereszt *et al*., 2011). Indeed, it has been demonstrated for the peptide NCR247 that it has multiple bacterial targets leading to inhibition of cell division and affecting the bacterial transcriptome and translation that collectively contribute to the altered physiology of the endosymbionts (Tiricz *et al*., 2013, Farkas *et al*., 2014, Penterman *et al.*, 2014).

NCRs are similar to the defensin-type of antimicrobial peptides, and some NCR peptides have antimicrobial activity, killing rhizobium and other bacteria when applied at high concentration (Van de Velde *et al*., 2010; Haag *et al*., 2011; Haag *et al*., 2012; Tiricz *et al*., 2013). Defensins and other types of antimicrobial peptides are found in all eukaryotes where they are part of the first line of defense against invading microbes. Thus the NCR peptides likely evolved from the ancestral immune repertoire. *NCR* genes were originally thought to be unique to the IRLC legume clade (Mergaert *et al*., 2003). The bacteroids in the nodules of the tested species of this clade all share the elongation and polyploidization feature Mergaert *et al*., 2006). However, refined bioinformatics tools for the prediction of small peptides in genome sequences led to the discovery of three putative *Arabidopsis* genes that encode peptides with the typical pattern of cysteine residues of the NCRs (Zhou *et al*., 2013). The existence of multiple *NCR* genes in several species of the IRLC clade suggests that the ancestral genes may have gained a new function in symbiosis in the common ancestor of IRLC and that increasing its copy number through gene duplications may have conferred a

selective advantage. To counteract the antimicrobial activity of the NCR peptides, *Sinorhizobium meliloti*, the symbiont of *Medicago*, requires the BacA protein. In the absence of this protein, the bacteroids are immediately killed by the NCR peptides as soon as they are released in the symbiosomes in nodule cells (Haag *et al*., 2011).

A striking and unusual feature of the *NCR* gene family in *M. truncatula* is that it is composed of about 600 genes, which are seemingly all expressed in the nodules (Mergaert *et al*., 2003; Alunni *et al*., 2007; Young *et al*., 2011; Nallu *et al*., 2013; Zhou *et al*., 2013). *In situ* expression analysis of individual *NCR* genes or microarray analysis of a large subset of the family has demonstrated that they are all expressed in the symbiotic nodule cells (Mergaert *et al*., 2003; Maunoury *et al*., 2010; Van de Velde *et al*., 2010; Nallu *et al*., 2013). Moreover, EST analysis and microarray experiments, testing a number of different plant organs as well as different growth conditions, revealed *NCR* gene expression only in nodules (Mergaert *et al*., 2003; Maunoury *et al*., 2010; Nallu *et al*., 2013).

The *Medicago truncatula* Gene Expression Atlas (MtGEA) (Benedito *et al.,* 2008; He *et al.,* 2009) was generated with the whole genome Affymetrix *Medicago* Gene Chip and compiles microarray data for the majority of *M. truncatula* genes (50,900 probe-sets) over a large set of experiments (254 different experiments in MtGEA version 3). The compendium is a unique and currently the richest resource for analysing gene expression in *M. truncatula*. In this study, we used the MtGEA compendium and additional unpublished and published microarray experiments (Limpens *et al*., 2013; Jayaraman *et al*., 2014) to describe in great detail the expression profiles of the majority of the *M. truncatula NCR* genes. We show that this gene family has an extreme tissue specific expression profile with undetectable expression in all tissues and conditions except in nodules where they become transcriptionally active to very high levels. In addition, promoter-GUS plants were produced for three *NCR* genes and used to confirm expression pattern specificity in various conditions, most particularly during biotic interactions.

3. Results

3.1. Global analysis of *NCR* gene expression

The probe-sets and expression data corresponding to the NCRs described before Mergaert *et al.* (2003) and Alunni *et al.* (2007) were searched in MtGEA (version 3) using the BLAST search option of the database. Expression profiles of 334 probe-sets were obtained (**Table 1; http://www.biomedcentral.com/content/supplementary/1471-2164-15-712-S1.xlsx**) and their expression patterns in 267 different experimental conditions (254 from MtGEAv3, five unpublished conditions, three from Jayaraman *et al.* (2014) and five from Limpens *et al*. (2013) are summarised in **Fig. 11**. The transcriptome compendium is mostly derived from the *M. truncatula* genotype **'**Jemalong**'** but also contains data sets obtained from the genotypes **'**R108**'** and **'**F83005.5**'** although all specific experiments discussed below were obtained with the **'**Jemalong**'** genotype. As specified in more detail below, the compendium covers the plant**'** s major organs, various kinds of abiotic and biotic stresses and data from specific cell and tissue types. In the heat map of **Fig. 11**, the experiments are organized in three major groups, namely the profiles of nodule samples, root samples and samples of all other plant organs regardless of the treatment they were submitted to. An obvious global pattern instantly revealed by the heat map is that the nearly complete 334 *NCR* gene set is only expressed in nodules except for one experiment marked with the red arrowhead, which is annotated in the MtGEA compendium as mycorrhizal roots, but is contaminated with nodules (**Fig. 12**). The expression of the *NCR* gene set in mycorrhizal roots was inconsistent with our previous results (Mergaert *et al.,* 2003) and also with other mycorrhizal samples present in the MtGEA compendium in which the *NCR* genes are not expressed. The Affymetrix gene Chip contains also probe-sets for monitoring the genes of *S. meliloti*. The *S. meliloti* probe-set can thus reveal whether samples of the MtGEA contain this bacterium. Analysing 28 genes of the *S. meliloti* probe-sets (**Table 1**), corresponding to the nitrogenase locus, we found that these genes were expressed in nodule samples, as expected, as well as in the particular mycorrhizal sample in which the NCRs are also active, while in all other samples including other mycorrhizal samples, roots and other organs, the *S. meliloti* probe-sets gave background signals (**Fig. 12**). Considering that the nitrogenase locus is known to be exclusively expressed in nodule bacteroids (Becker *et al*., 2004), this mycorrhizal sample is most likely impure and contaminated with nodules.

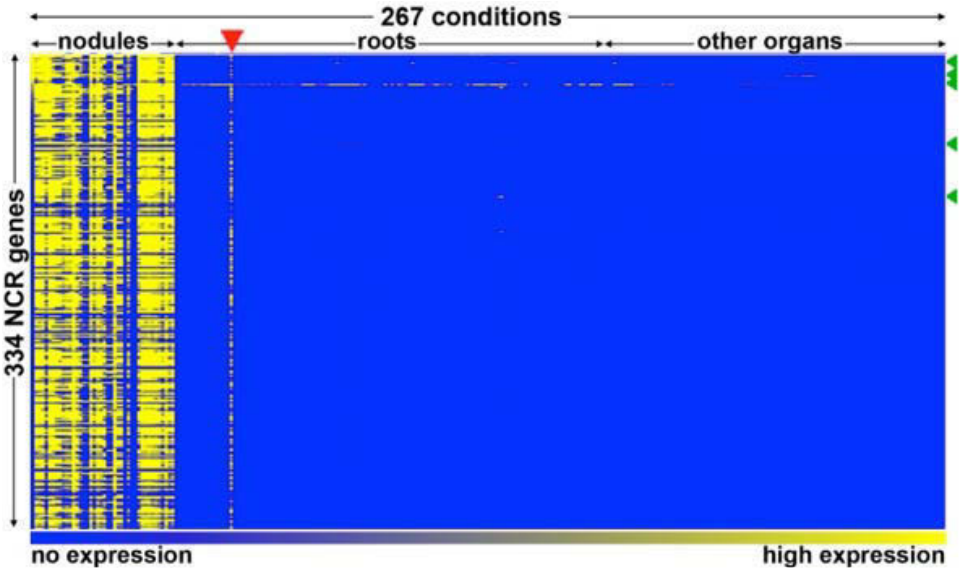

Figure 11. Heat map of *NCR* gene expression in the MtGEA compendium

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

Figure 12. Heat map of *S. meliloti* gene expression in the MtGEA compendium

The heat map shows the expression of 28 *S. meliloti* genes (rows) in 267 experimental conditions (columns). The probe-set included 28 genes of the locus carrying the *nifHDK* operon encoding the nitrogenase enzyme and the dataset used for the generation of the heat map is provided in **Table 1**. Experiments are ordered as in **Fig. 11** and as indicated above the columns. The color scale bar indicates the expression level from background level (blue) to maximum level (yellow). The red arrowhead indicates the mycorrhizal sample that is contaminated with nodules.

The expression profiles of individual *NCR* genes show expression to very high levels in the nodule conditions and only background levels in the other experiments (**Fig. 13A,B**). Such profiles are typical for nearly all *NCR* genes (**Table 1**) but by surveying all NCR probe-sets, 5 exceptions were discovered with more or less relaxed nodule specific expression patterns (**Fig. 11**, red arrowheads; **Fig. 13C-G**). NCR247 and NCR077 are still mainly expressed in nodules but are also weakly active in other conditions. The *NCR*247 gene seems to be expressed in different root samples and in some samples from areal tissues although at

lower levels than in nodules. It is not evident from the available information of the different experiments to determine what may activate this expression. NCR077 has a higher than usual background level, possibly because of a less specific probe set, but the gene seems to be also expressed in some mycorrhizal samples (**Fig. 13D**) including a laser capture microdissection sample of arbuscule-containing cells (Gaude *et al*., 2012). NCR218 and NCR122 on the other hand have a completely relaxed specificity and are expressed to similar levels in nodules and in other conditions, mostly roots (**Fig. 13E,F**). NCR235 expression is similarly specific to most other *NCR* genes except for a weak expression in stems and shoots, which is about 10- to 100-fold lower than in nodules (**Fig. 13G**). Also NCR247 is expressed in some of the stem samples. Thus, except for these 5 genes, the complete tested *NCR* gene set is only expressed in nodules and in none of the other conditions that are present in MtGEA.

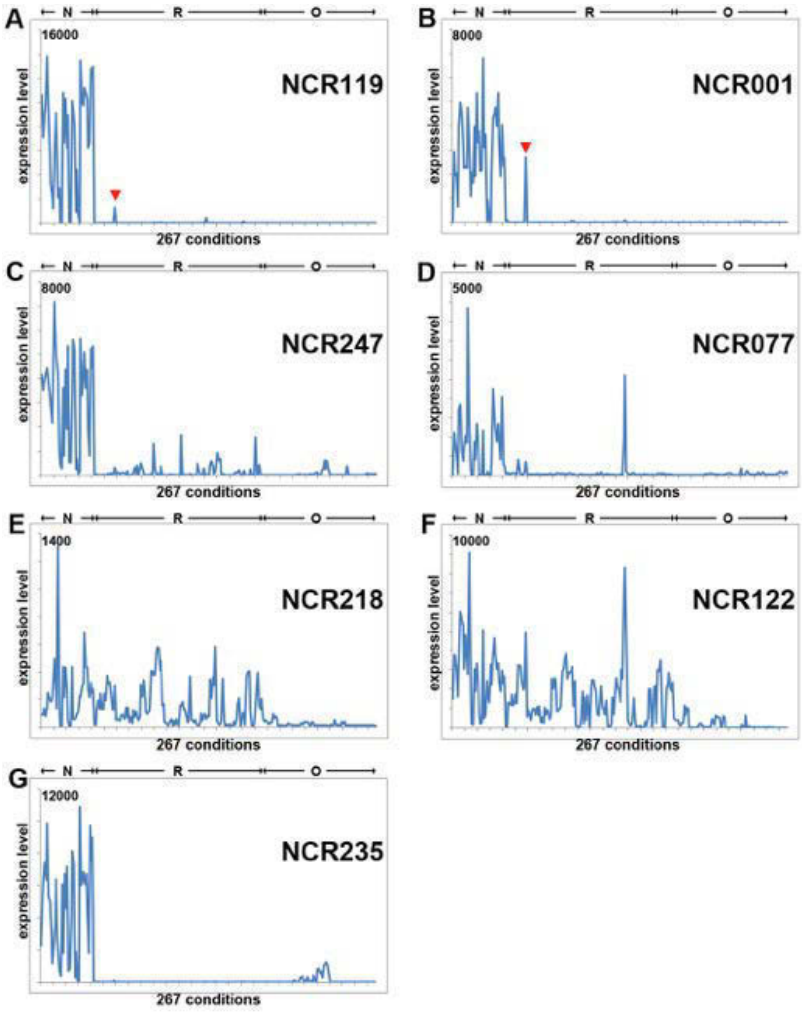

Figure 13. Expression profile examples of 7 *NCR* genes

(A-G) The expression pattern of 7 individual *NCR* genes in 267 experiments (x-axis) is provided. The experiments are ordered as in **Fig. 11**. N is nodules; R is root conditions; O is all other conditions. The y-axis is the expression level according to the fluorescence hybridization signal and is scaled according to the maximum hybridization signal as indicated on the top of the axis. The red arrowheads in panels **A** and **B** indicate the mycorrhizal sample that is contaminated with nodules.

3.2. Spatio-temporal expression of *NCR* genes in nodules

The MtGEA compendium contains 42 different nodule samples including samples of wild-type nodules harvested at different days post inoculation with *S. meliloti* and thus at different stages of nodule development. In uninoculated root samples and nodule primordia of 3dpi, none of the *NCR* genes are activated (**Fig. 14A,B**). The NCR transcriptomes at 0 dpi and 3 dpi have a correlation coefficient close to 1 (**Fig. 14D**).

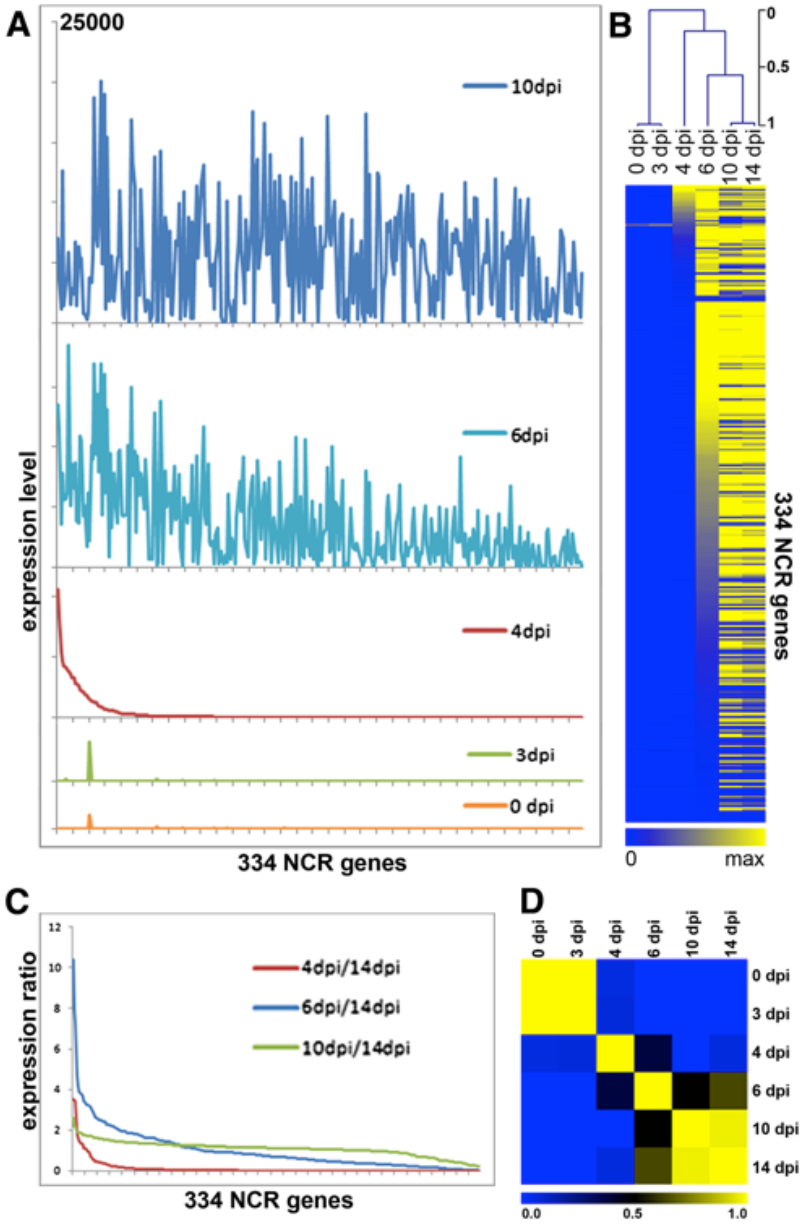

Figure 14. Successive activation of *NCR* genes during nodule development

(A) Expression level of the 334 NCR probe-sets (x-axis) at 0, 3, 4, 6 and 10 dpi. The y-axis scale is the same for all graphs and is 25,000 at maximum. **(B)** Heat map of the same expression data as in **(A)**. **(C)** The ratios of NCR expression levels at 4 dpi, 6 dpi or 10 dpi compared to 14 dpi. For each time point, the genes were ordered from high ratio to low ratio. At 4 dpi, only few *NCR* genes are activated; at 6 dpi most are activated but not yet at maximum scale; at 10 dpi, nearly all *NCR* genes are activated at similar level to 14 dpi (ratio close to 1). **(D)** Heat map of the Pearson correlation of the NCR expression profiles at 0, 3, 4, 6, 10 and 14 dpi.

This suggests that NCRs have no function in the early stages of the interaction. In agreement with this, Nod factor or Myc factor treatments (Czaja *et al.*, 2012) do not induce *NCR* gene expression (**Fig. 15**).

Nod factors and Myc factors are similar lipochitooligosaccharide signals, produced by rhizobium and mycorrhizal fungi symbionts inducing in the legume host the early stages of nodule and mycorrhizal formation respectively (Maillet *et al*., 2011; Downie, 2014). A subset of about 70 *NCR* genes is activated at 4 dpi with *S. meliloti* (**Fig. 14**). Some of these genes are already activated at their maximal level while others are only induced to a fraction of their maximal level and reach full expression at a more advanced stage of nodule development (**Fig. 14**). At 6 dpi, most genes are activated but many of them not yet at their maximal level (**Fig. 14A-C**). At 10 dpi the NCR transcriptome seems to be fully activated, to a similar extent as a later time point at 14 dpi. The NCR transcriptomes at 10 dpi and 14 dpi have a correlation coefficient close to 1 (**Fig. 14D**). This pattern hints at a link between the activation of *NCR* gene expression and the progression of bacterial infection in the incipient nodules.

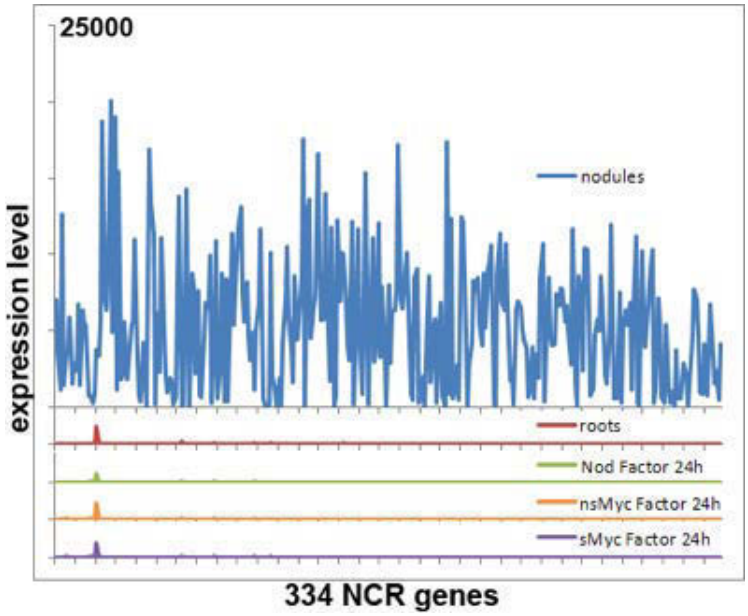

Figure 15. *NCR* gene expression in response to Nod factors and Myc factors

The expression pattern of 334 NCR probe-sets (x-axis) is shown for nodules 10 dpi, untreated control roots, 24h 10^{-8}M Nod factor treatment, 24h 10^{-7}M non-sulfated Myc factor treatment (nsMyc) and 24h 10^{-8}M sulfated Myc factor treatment (sMyc) (Czaja *et al*., 2012). The y-axis represents the strength of the hybridization signal and the graphs for all treatments are at the same scale (maximum 25,000).

Because of the presence of an apical meristem, nodules in *M. truncatula* are of the so-called indeterminate type and all stages of symbiotic cell differentiation, from undifferentiated meristem cells till fully differentiated and functional symbiotic cells, are present in any

mature nodule, independent of its age. Therefore, mature nodules are organized in well-defined histological zones: zone I, meristem; zone II, zone of infection and differentiation (polyploidization) of the symbiotic cells; interzone II–III, characterized by amyloplast accumulation; zone III, nitrogen fixation zone with mature functional symbiotic cells; zone IV, senescent zone (Kondorosi *et al*., 2013). This nodule structure suggests that the temporal NCR expression profiles defined above could correlate with a spatial pattern in the nodule tissues. To test this possibility, we analyzed transcriptome data from 4-week-old nodules that were hand-sectioned in five different parts (see Methods section). These samples correspond to the nodule tissues from the most apical part of the nodule to the most proximal part and are approximately overlapping with the nodule zones I, II, II-III, III and IV. Sample A is the meristem and the underlying few cell layers of post-meristematic cells which start the infection and differentiation process. Sample B corresponds mainly to the infection and differentiation zone II. Sample C corresponds essentially to the interzone II-III. Cluster analysis of the NCR abundance profiles in these five samples distinguishes groups of *NCR* genes that have preferential expression in defined zones of the nodule and that are sequentially activated from nodule apex to proximal tissues (**Fig. 16A** and **Fig. 17**).

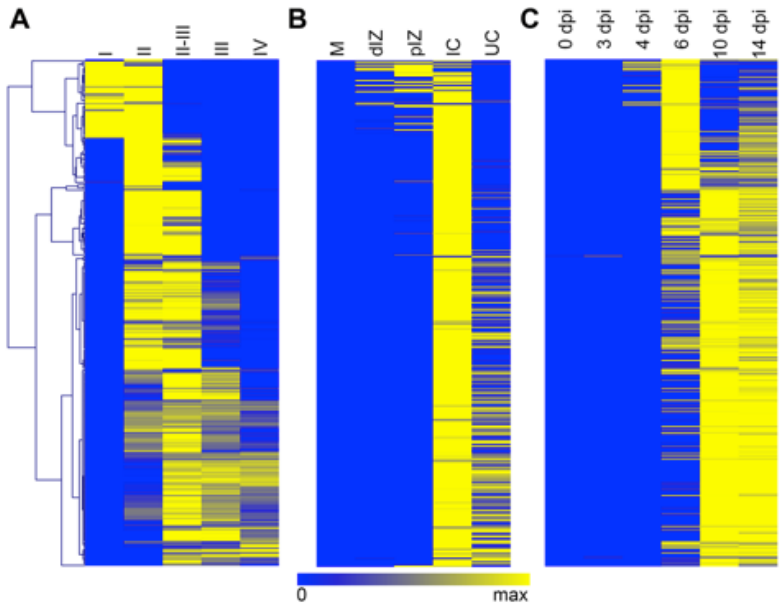

Figure 16. Spatio-temporal expression of the *NCR* gene family in nodules

(A) Heat map and hierarchical clustering of NCRs in function of their spatial expression in nodules (samples I to IV, from the youngest, most apical part of the nodules to the oldest, most proximal part). Clustering was performed using Pearson correlation. **(B)** Heat map of NCR expression profiles obtained by laser-capture microdissection coupled to affymetrix microarray analysis described by Limpens *et al*. (2013) M corresponds to meristem, dIZ to the distal zone II or infection zone, pIZ to the proximal zone II, IC to infected cell and UC to uninfected cell. **(C)** Heat map of NCR expression in function of the nodule age (dpi). The genes in the rows in **(B)** and **(C)** are in the same order as in panel **(A)**. For each probe-set in the 3 panels, the relative expression was converted to the log2 value of the ratio representing the zone-specific expression over the average expression in all five zones.

The transcriptome in specific tissues and cells of nodules was also obtained by laser-capture microdissection (LCM) (Limpens *et al*., 2013). In this experiment, the transcriptome was analyzed in nodule meristems, the distal and the proximal infection zone, the infected cells of the fixation zone and the uninfected cells of the fixation zone (**Fig. 16B**). Although the type of samples are not entirely overlapping, a good correspondence can be observed between the LCM dataset and the hand-dissected dataset (**Fig. 16A,B**). The LCM dataset shows that *NCR* genes are not expressed in the nodule meristem which is free of rhizobia (**Fig. 16B**). This observation is in agreement with their activation by the infecting rhizobia. It also indicates that the expression of the early *NCR* genes in the hand-dissected sample I results from the significant presence of cells from the infection zone in that sample. *NCR* genes were reported to be expressed in infected cells (Kondorosi *et al*., 2013). Unexpectedly, the LCM dataset revealed a relatively high expression of several *NCR* genes in the uninfected cells (**Fig. 16B**, sample UC).

However, among them were the genes NCR001, NCR084, NCR035 and NCR247 for which *in situ* hybridization, promoter-GUS or-mCherry fusions or immunolocalization of the peptide demonstrated a specific expression in the infected symbiotic nodule cells and no expression in uninfected cells (Mergaert *et al*., 2003; Van de Velde *et al*., 2010; Farkas *et al*., 2014). Possibly, the signal in the sample of the uninfected cells is the result of some contamination with infected cells or due to background hybridization coming from the RNA amplification procedure used with the very low amount of RNA obtained from the LCM samples (Limpens *et al*., 2013). Nevertheless, when comparing the expression level of all NCRs in the uninfected and infected cells, only two genes had a significant higher expression in the uninfected cells. Those two genes are NCR218 (9,6 fold higher; t-test P=0,002) and NCR122 (3,2 fold higher; t-test P=0,018), which are the only 2 NCRs that are consistently expressed to high levels in roots (**Fig. 13E,F**).

When matching the spatial patterns with the corresponding temporal patterns, a fairly good correspondence can be observed: genes expressed in the apex are mostly also fully activated early during the nodule development at 4 or 6 dpi, while genes expressed in the more proximal tissues are activated late in nodule development (**Fig. 16**). The correspondence between the spatial and temporal pattern of *NCR* gene expression is also obvious whenconsidering the clusters of genes representing the major expression profiles (**Fig. 17**). For example, genes of clusters 1 and 2 (**Fig. 17**) are expressed in the most apical part of the nodule and this correlates with an early activation at 4 or 6 dpi in the temporal pattern. On the

other hand, genes in cluster 5 have maximum expression only in sample II-III and this corresponds with an activation late in nodule development at 10 dpi (**Fig. 17**).

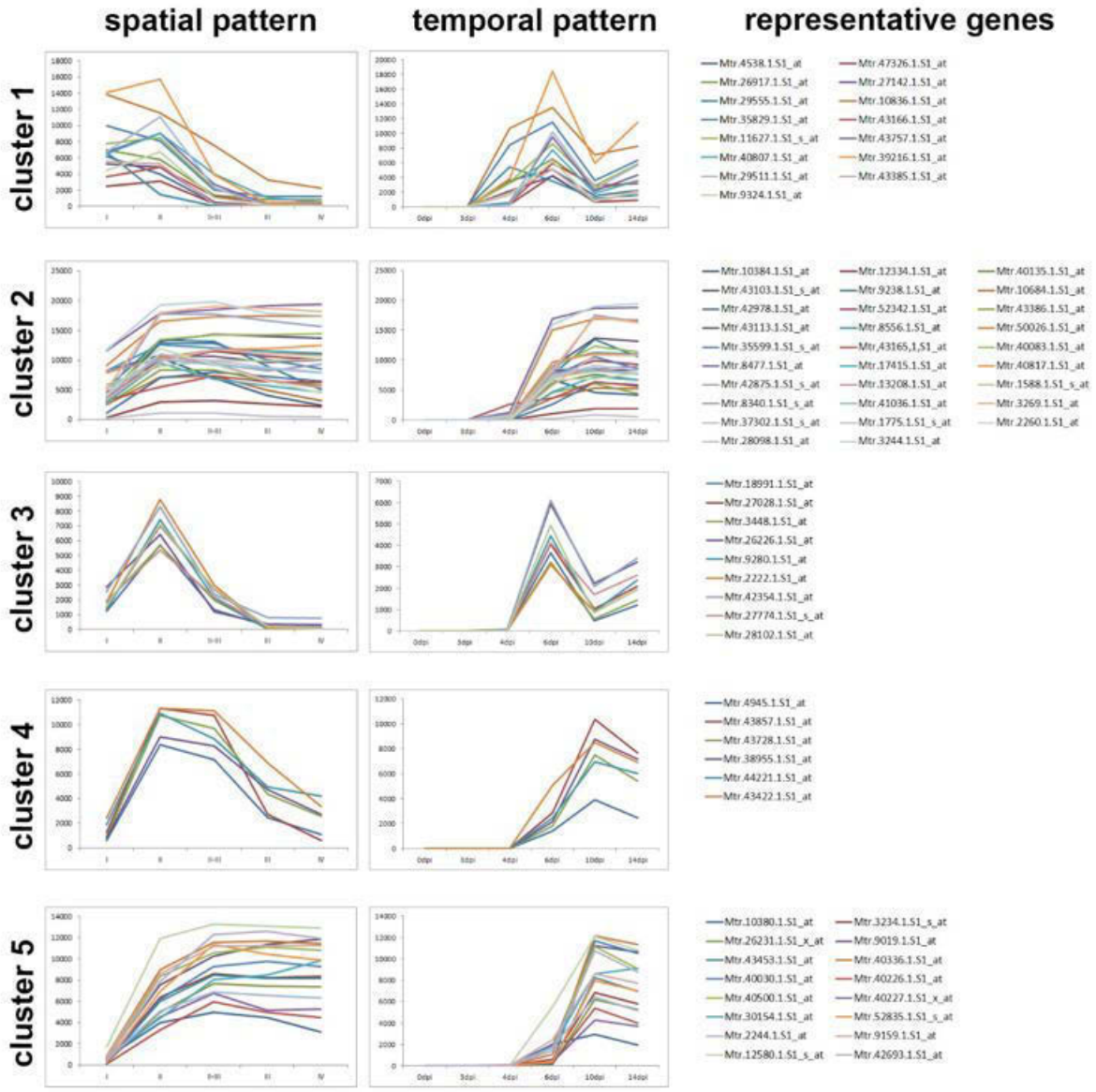

Figure 17. Representative clusters of spatial and temporal NCR expression profiles

Five representative expression profiles are shown (cluster 1 to 5). The spatial pattern (left), the temporal pattern (centre) and the genes constituting the cluster (right) are shown. Genes in cluster 1 display the highest expression in the nodule apex sample I, are still high in sample II and then decline rapidly (**Fig. 29)**. Cluster 2 genes are already activated in sample I but their expression increases in sample II and remains high in the whole nodule. Cluster 3 genes have a very sharp pattern and are highly expressed in sample II, low in sample I and II-III and close to zero in the samples III and IV. The genes of cluster 4 have a similar profile but are expressed in a broader zone with high levels of expression in samples II and II-III. Finally, cluster 5 genes are absent in sample I, partially activated in sample II and then fully activated in sample II-III, III and IV. The genes are identified by their probe-set annotation number as provided in **Table 1**.

Together these spatio-temporal patterns reveal that the *NCR* genes are activated in different waves, in agreement with our previous results that identified two key points in nodule development associated with major transcriptional activation, one at the formation of

symbiotic cells and another one when bacteroids differentiate (Maunoury *et al.,* 2010). Nevertheless, the present analysis is refining this description and shows that *NCR* genes are activated in at least 3 waves and moreover they can be distinguished by the maintenance or the decline of their expression in the older nodule cell.

3.3. *NCR* genes are not directly involved in nodule senescence

Because of their resemblance to the defensin-type of antimicrobial peptides (Mergaert *et al.,* 2003) and because they have *in vitro* and *in vivo* antimicrobial activity (Van de Velde *et al.,* 2010; Haag *et al.,* 2011, Haag *et al.*, 2012; Tiricz *et al*., 2013), part of the NCR family could be involved in the killing of the rhizobia during the senescence process of nodules as has been suggested (Nallu *et al.*, 2013). Nodule senescence leads to a complete digestion of the bacteroids and the symbiotic host cells (Van de Velde *et al.,* 2006) and is controlled by the transcriptional activation of a battery of genes which are involved in the regulation of the process, the digestion of macromolecules and the remobilization of the liberated nutrients (Van de Velde *et al.,* 2006). Senescence of nodules can also be induced artificially by supplying a high concentration of nitrate to roots which prefer this new nitrogen source over the energy costly nitrogen fixation (Cabeza *et al*., 2014) or alternatively by cutting off the nutrient supply to nodules by putting plants in the dark (Pérez-Guerra *et al*., 2010) or by applying a herbicide (Seabra *et al.*, 2012). To test whether *NCR* genes are induced by senescence, we compared nodules at 14 dpi with nodules of 14 dpi followed by 2 days of nitrate supply (**Fig. 18A**). Another opportunity in the MtGEA database to test the effect of senescence on NCR expression is provided by samples corresponding to a treatment of *M. truncatula* with the herbicide phosphinothricin, which also induces nodule senescence (Seabra *et al.*, 2012) (**Fig. 18B; Fig. 19**).

Besides the *NCR* genes, 8 senescence marker genes were included in the analysis (**Table 1**) (Van de Velde *et al.,* 2006). Both treatments induced senescence as indicated by the strong activation of the senescence marker genes. However, all the *NCR* genes without exception are reduced in expression by the senescence-inducing treatments, in agreement with a recent report (Cabeza *et al*., 2014). The down regulation of *NCR* gene expression is induced very rapidly, within 4 hours of nitrate application (Cabeza *et al*., 2014) or 8 hours of phosphinothricin treatment (**Fig. 19**) and occurs before the degeneration of the symbiotic cells and their bacterial symbionts (Seabra *et al.*, 2012). The down-regulation of the NCRs most likely reflects the shutdown of the symbiotic process. This expression pattern suggests that none of the *NCR* genes has a direct role in senescence.

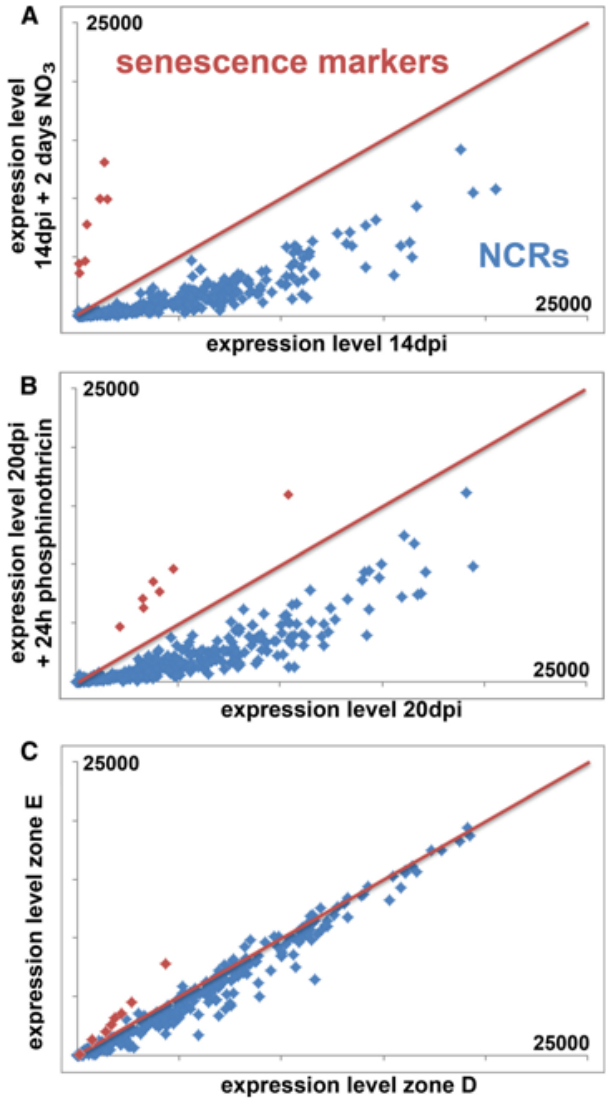

Figure 18. NCR expression during nodule senescence

(A) Scatter plot of the expression of 334 NCRs (blue) and 8 senescence marker genes (red) in 14 dpi nodules (x-axis) and 16 dpi nodules, which have been treated before harvest with nodule senescence-inducing nitrate for 2 days (y-axis) (Benedito *et al*., 2008). **(B)** Scatter plot of the expression of 334 NCRs (blue) and 8 senescence marker genes (red) in 20 dpi nodules (x-axis) and 20 dpi nodules, which were treated for 24h with phosphinothricin before harvest (y-axis) (Seabra *et al*., 2012). **(C)** Scatter plot of the expression of 334 NCRs (blue) and 8 senescence marker genes (red) in hand-sectioned nodule zone D (x-axis) and in hand-sectioned nodule zone E (y-axis). The scale of all axes is the same and is 25,000 at maximum. The red line in the three graphs indicates a ratio of 1 between the two conditions that are compared. The probe-sets for the senescence marker genes, encoding cysteine proteinases, a chitinase, a nuclease a nucleoside transporter and a metal-nicotinamide transporter, are provided in **Table 1**.

This conclusion is also confirmed by comparing the hand-dissected nodule samples D and E, which are enriched for the nitrogen fixation zone III, and the senescence zone IV respectively. None of the NCRs is significantly (t-test P<0,05) higher expressed in sample IV (**Fig. 18C**).

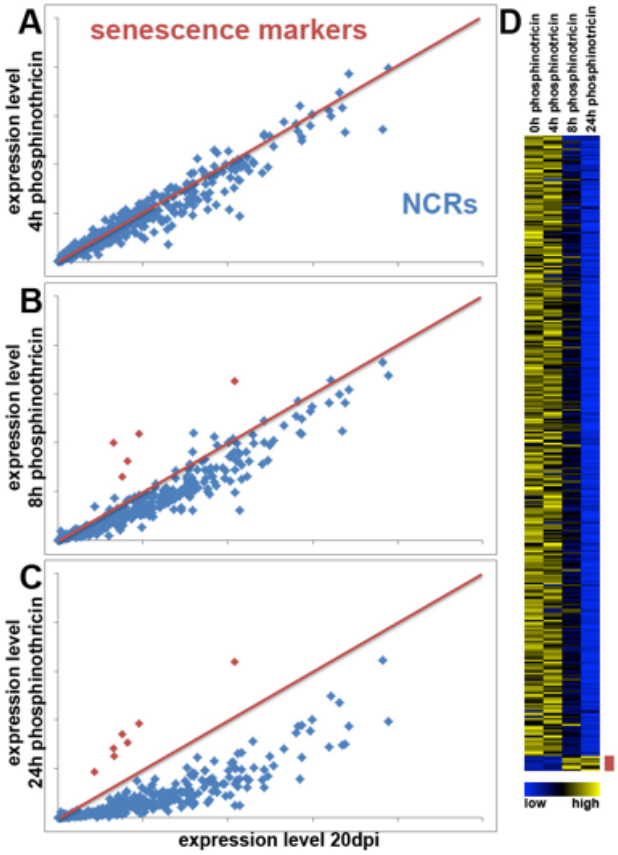

Figure 19. Effect of Phosphinothricin treatment on *NCR* gene expression

(A-C) Scatter plot of gene expression levels in 20 dpi nodules compared to 20 dpi nodules 4h **(A)**, 8h **(B)** and 24 h **(C)** post treatment with the herbicide phosphinothricin. A set of senescence markers from Van de Velde *et al*. (2006) (red dots) and the 334 *NCR* genes (blue dots) show opposite regulation from 8h post treatment. The red line in the three graphs indicates a ratio of 1 between the two conditions. The probe-sets for the senescence marker genes, encoding cysteine proteinases, a chitinase, a nuclease a nucleoside transporter and a metal-nicotinamide transporter, are provided in **Table 1**. **(D)** Heat map of NCR and senescence markers expression before and 4, 8 and 24 h post phosphinothricin treatment. Expression levels vary from low (blue) to high (yellow). The red bar next to the heat map indicates the location of the senescence marker genes.

3.4. Promoter-GUS analysis and immunolocalization of selected NCRs in nodules

In order to confirm the expression data from MtGEA, stable transgenic *M. truncatula* R108 lines were generated carrying promoter-GUS fusion constructs for 3 different *NCR* genes, representing different temporal classes of NCRs and inoculated with *Sinorhizobium meliloti* strain 1021 or *Sinorhizobium arboris* strain B554. NCR001 is not activated before the late stages of the nodule formation, NCR084 is slightly induced in early time points (4 dpi) and fully activated at the mature stage of the nodule and finally NCR121 is an early gene, which is already fully activated at 4 dpi. GUS expression in the 3 transgenic lines was not detected in root tips or other root parts (**Fig. 20**). In agreement with its temporal regulation during nodulation, NCR121 expression was induced in young nodule primordia as early as

5dpi and remained expressed throughout the experiment in the entire infection zone and the fixation zone of mature nodules (**Fig. 20**). 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 (**Fig. 20**).

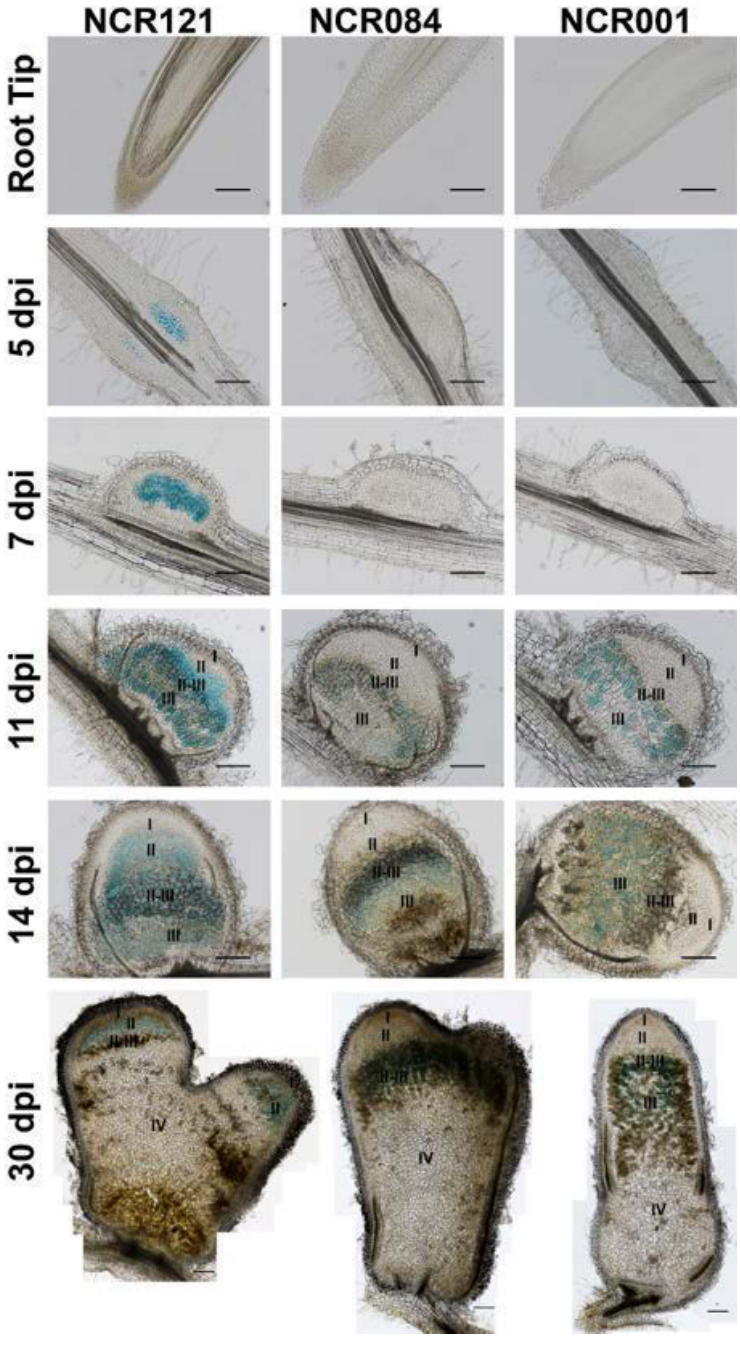

Figure 20. Promoter-GUS analysis of *NCR* genes in nodules

Sections (70 µm) of root tips and nodules of different ages of transgenic plants carrying promoter-GUS fusions for the NCR121, NCR084 and NCR001 genes were stained for GUS activity observed as blue color. Scale bars are 50 µm. The location of the nodule zones in mature nodules is indicated: I is the meristem; II is the infection zone; II-III is the interzone; III is the fixation zone and IV is the senescence zone. The nodules shown in the figure were induced by *S. arboris* strain B554 but indistinguishable patterns were obtained in nodules formed by *S. meliloti* strain 1021 (data not shown).

NCR001 expression was detectable from 11 dpi in the developing fixation zone and its expression extends in the following days as the fixation zone is growing (**Fig. 20**). All 3 genes are only expressed in the symbiotic nodule cells. In older nodules, at 30 dpi, displaying a senescence zone, NCR expression was never detected in the senescing tissues, nor was their expression enhanced in the proximal fixation zone adjacent to the senescent tissue (**Fig. 20**), confirming that *NCR* genes are not involved in the senescence process. Overall, the temporal and spatial promoter-GUS expression patterns are in very good agreement with the expression profiles deduced from MtGEA.

The particular expression pattern of NCR122 with its relaxed tissue specificity (**Fig. 13F**) and its apparent expression in the uninfected nodule cells, together with the availability of an anti-NCR122 antibody prompted us to specifically analyze the localization of the NCR122 peptide in nodules. Immunolocalization of the peptide revealed indeed a specific presence of NCR122 in the uninfected cells of the central zone of a mature nodule as well as in the uninfected cortical cells of the nodule (**Fig. 21**). Together with the transcriptome data, this indicates that NCR122 and most likely also NCR218 are the only NCR peptides that are specific to uninfected root and nodule cells.

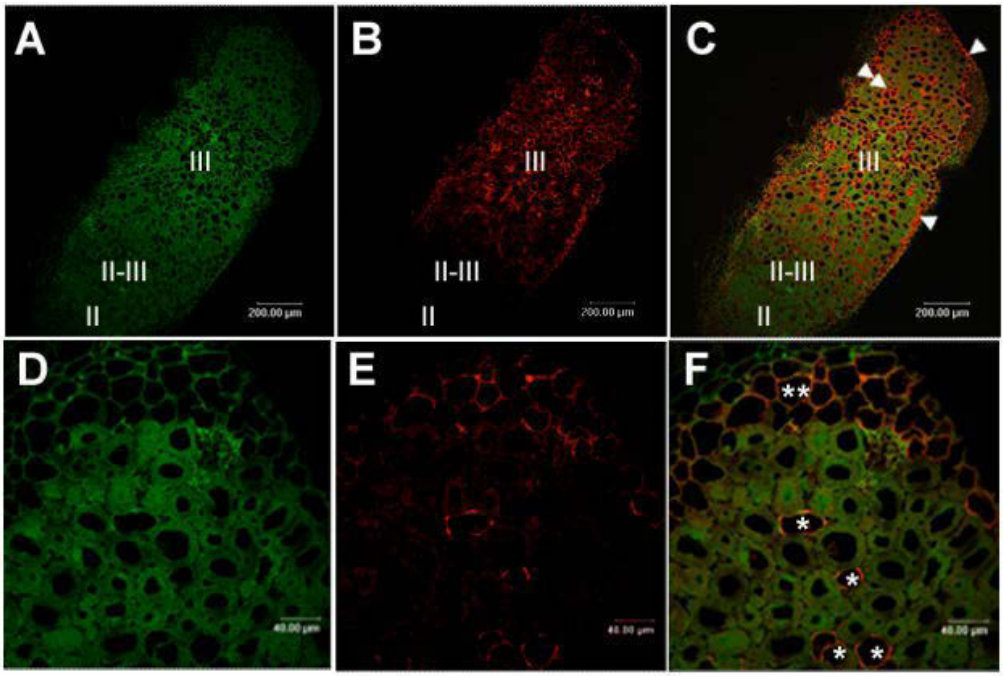

Figure 21. Immunolocalization of NCR122 in nodule sections

Sections of nodules were stained with the DNA label SYTO13 (green) to reveal the bacteria **(A,D)** and were immunolabelled with the anti-NCR122 antibody (red) **(B,E)**. **(C,F)** Overlays of **(A,B)** and **(D,E)** respectively. **(D-F)** enlargement of the top part of the nodule shown in **(A-C)**. Arrowheads or double asterisks mark NCR122-labeled cells in the nodule cortex and the double arrowhead or asterisks mark uninfected cells of the nitrogen fixation zone III labeled by NCR122. II is the infection zone; II-III is the interzone. Scale bars are 200 µm in panels **A-C** and 40 µm in panels **D-F**.

3.5. Expression of *NCR* genes in plant organs

Previously, NCR expression in conditions other than the symbiosis with rhizobium was tested by EST analysis (Mergaert *et al.,* 2003) and with dedicated microarrays (Tesfaye *et al.,* 2013), indicating the absence of expression. The MtGEA database offers the possibility to extend this analysis to more plant organs and biotic and abiotic stress conditions.

Besides nodules, 8 other plant organs (Benedito *et al*., 2008) were interrogated for NCR expression (**Fig. 22**) as well as plant treatments with the phytohormones auxin, cytokinin and auxin transport inhibitors (Imin *et al*., 2008; Rightmyer and Long, 2011) (**Fig. 23**).

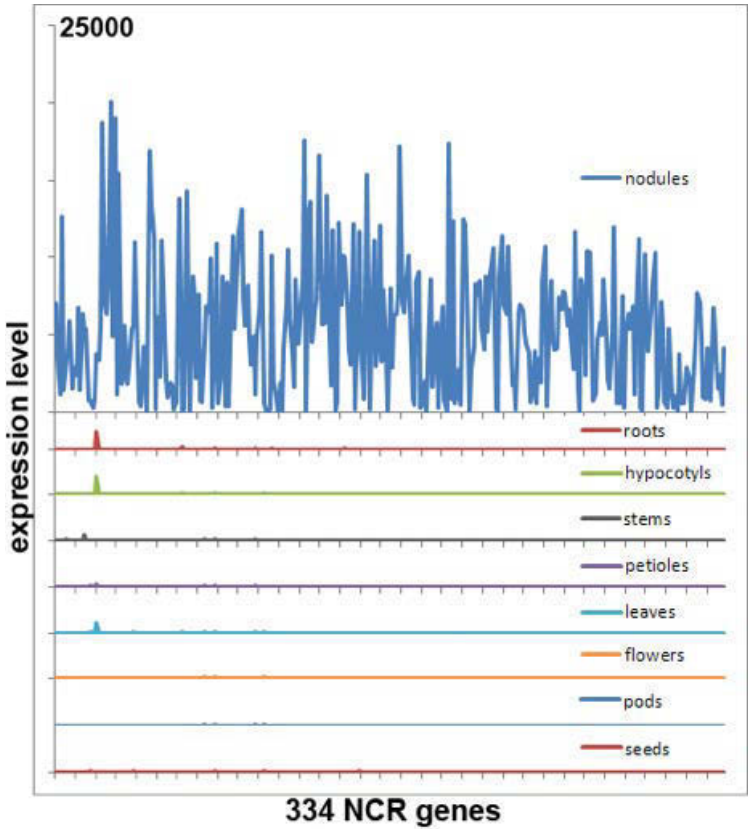

Figure 22. NCR expression in plant tissues

The expression pattern of 334 NCR probe-sets is shown for nodules, roots, hypocotyls, stems, petioles, leaves, flowers, seed pods and seeds. The graphs for all organs are at the same scale (maximum hybridization signal value is 25,000). The small peaks that can be seen in some of the organ samples correspond to the *NCR* genes with relaxed expression (**Fig. 13**).

Interestingly, treatment of roots with the auxin transport inhibitors TIBA or NPA leads to the formation of nodule-like structures (Rightmyer and Long, 2011). However, in none of these conditions were *NCR* genes expressed except for the *NCR* genes with relaxed expression described above.

NCRs resemble the defensin-type of antimicrobial peptides and plant defensins are often expressed to high levels in "infection-sensitive" organs like flowers or seeds. Because of the complete lack of detectable expression of NCRs in these organs (**Fig. 22**), they most probably do not have a defensive function in these organs. Nevertheless, many non-NCR defensin-like genes were found to be expressed in seeds, potentially involved in their protection (Tesfaye *et al.,* 2013).

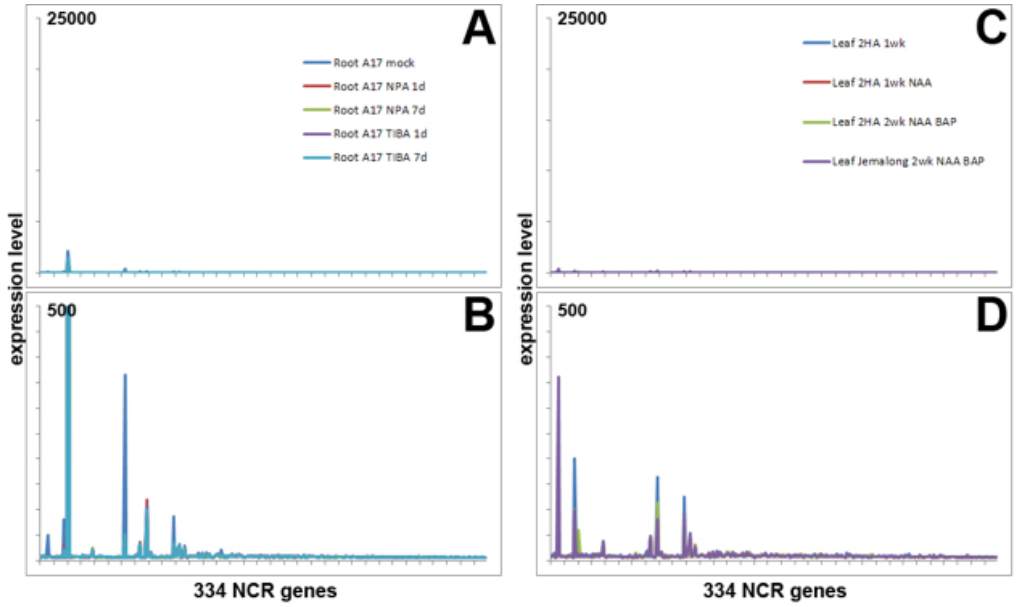

Figure 23. NCR expression in response to phytohormones

(A,B) NCR expression after treatments for 1 (1d) or 7 days (7d) of roots with the auxin transport inhibitors naphthylphthalamic acid (NPA; 200 µM) and 2.3.5-triidobenzoid acid (TIBA; 200 µM) (Rightmyer and Long, 2011). **(C,D)** NCR expression after treatments of leaves for 1 (1wk) or 2 (2wk) weeks with the auxin1-naphthaleneacetic acid (NAA; 10 µM) or the cytokinin 6-benzylaminopurine (BAP; 4 µM) (Imin *et al*., 2008). **(A)** and (**B**) on the one hand and **(C)** and **(D)** on the other hand represent the same data but the scale of the graphs are at relative expression level 25,000 maximum in **(A)** and **(C)** and at maximum level 500 in **(B)** and **(D)**.

3.6. Expression of *NCR* genes after biotic and abiotic stress

Defensins are also induced during infection with pathogens or during salt and drought stresses (De Coninck *et al.,* 2010; Maroti *et al.,* 2011). Therefore, we specifically analyzed how the *NCR* gene family is expressed during such conditions (**Fig. 24**) in roots after infection with the pathogenic fungi *Phymatotrichopsis omnivora* (Uppalapati *et al*., 2009) and *Macrophomina phaseolina* (Mah *et al*., 2012) and the symbiotic mycorrhizal fungus *Glomus intraradices* (Hogekamp *et al*., 2011), the oomycete *Aphanomyces euteiches* (Rey *et al.,* 2013), the bacterial pathogen *Ralstonia solanacearum*, the nematode *Meloidogyne incognita* (Damiani *et al.,* 2012) and the human enteric bacterial pathogens *Escherichia coli* O157:H7

and *Salmonella enterica* which are frequent sources of legume food contamination and are capable of surface and internal colonization of *M. truncatula* roots (Jayaraman *et al*., 2014) (**Fig. 25A,B**). In addition transcriptomes in *M. truncatula* cell suspensions treated with the defense response-inducing yeast elicitor or methyl jasmonate signals (Naoumkina *et al.,* 2007) were also included (**Fig. 26C,D**). Finally, salt and drought stresses are also reported to induce defensin genes in plants (De Coninck *et al.,* 2013). Thus transcriptomes in NaCl-treated roots (Li *et al*., 2009) as well as drought stressed roots (Zhang *et al*., 2014) were analyzed (**Fig. 26**). Not considering the 5 *NCR* genes with relaxed specificity, we could not detect expression of any of the *NCR* genes in all these data sets together (with the possible exception of giant cells formed by the nematode *Meloidogyne incognita* (**Fig. 24B**).

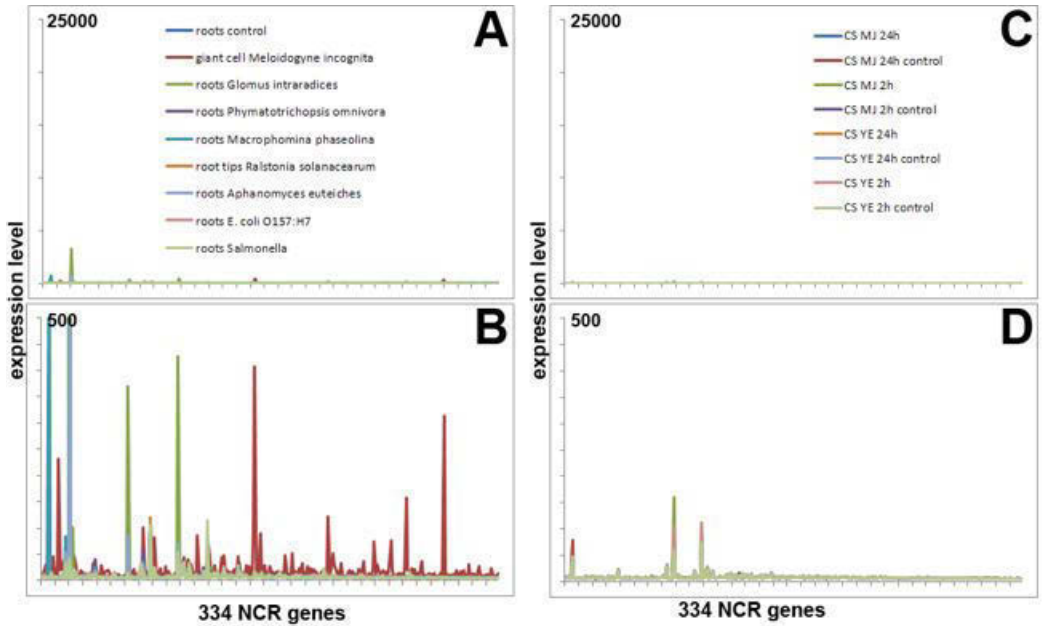

Figure 24. Expression of *NCR* genes during microbial infections and elicitor treatment

(A,B) NCR expression in *M. truncatula* control roots and roots infected with the pathogenic fungi *Phymatotrichopsis omnivora (*Uppalapati *et al*., 2009) and *Macrophomina phaseolina* (Mah *et al*., 2013) and the symbiotic mycorrhizal fungus *Glomus intraradices* (Hogekamp *et al*., 2011), the oomycete *Aphanomyces euteiches* (Rey *et al*., 2013), the bacterial pathogen *Ralstonia solanacearum,* the nematode *Meloidogyne incognita* (Damiani *et al*., 2012) and the human enteric bacterial pathogens *Escherichia coli* O157:H7 and *Salmonella enterica* which are frequent sources of legume food contamination and are capable of surface and internal colonization of *M. truncatula* roots (Jayaraman *et al*., 2014). **(C,D)** NCR expression in a *M. truncatula* cell suspension derived from root cells (CS) after treatments for 2h or 24h withthe defense response-inducing signals methyl jasmonate (500 µM) (MJ) and yeast elicitor (50 µg glucose equivalents ml^{-1}) (YE) (Naoumkina *et al*., 2007). **(A)** and **(B)** on the one hand and **(C)** and **(D)** on the other hand represent the same data but the scale of the graphs are at relative expression level 25,000 maximum in **(A)** and **(C)** and at maximum level 500 in **(B)** and **(D)**.

Several NCRs showed a hybridization signal in giant cells although the level was about 1 to 2 orders of magnitude lower than the signal in nodules for the same *NCR* gene. However, it should be noted that the giant cells were isolated by LCM and that the array

hybridization was performed with an amplified cDNA sample (Damiani *et al.,* 2012), which could be a source of background hybridization. In any case, besides the possible exception of the giant cells, the data indicate that the *NCR* genes seem not to be used by the plant to control infections other than the *rhizobium* bacteria in nodules.

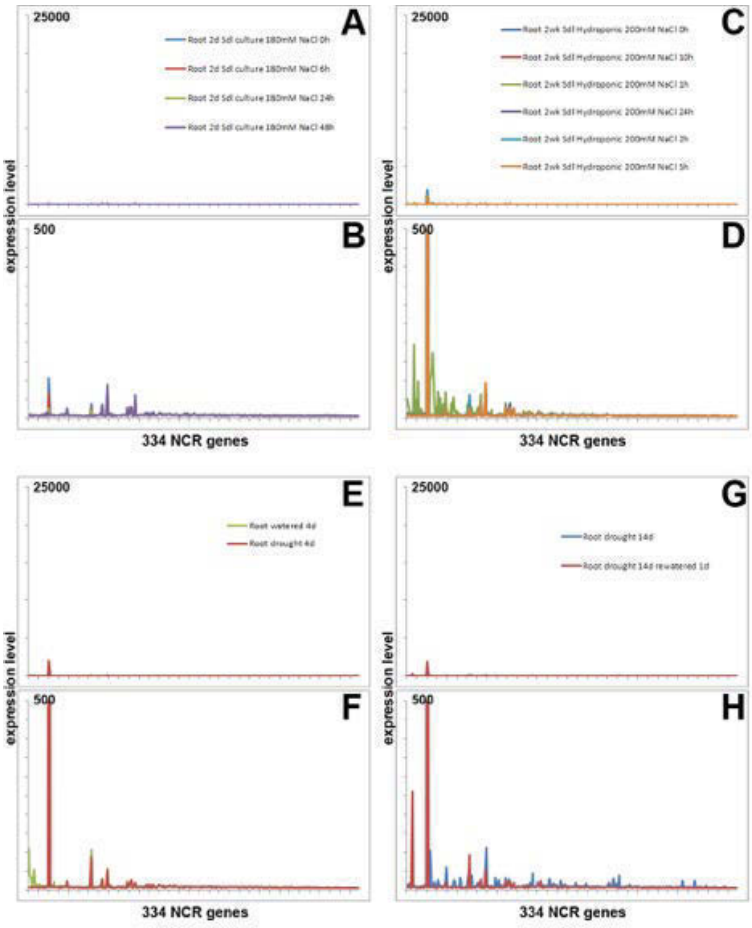

Figure 25. Expression of *NCR* genes during drought and salt stress

(A-G) NCR expression in roots exposed to NaCl salt stress (Hann and Rathjen, 2007) or drought stress (Schug *et al*., 2005). **(A)** and **(B)**, **(C)** and **(D)**, **(E)** and **(F)** and (**G)** and **(H)** represent pairwise the same data but the scale of the graphs is at relative expression level 25,000 maximum in **(A)**, **(C)**, **(E)** and **(G)** and at level 500 in **(B)**, **(D)**, **(F)** and **(H)**.

3.7. Promoter-GUS analysis of NCR expression during pathogenic interactions

We used the 3 NCR promoter-GUS reporter lines to confirm the absence of NCR expression during pathogenic responses and to complement these observations. The MtGEA dataset includes *M. truncatula* responses to root pathogens and therefore we analyzed leaf or stem pathogens that encompass also other trophic interactions and infection strategies. Inoculation of *M. truncatula* leaflets with the necrotrophic soft rotting bacterium *Dickeya*

*dadantii*3937 induced maceration symptoms from 1dpi on, but failed to induce NCR expression (**Fig. 26**). Similarly infiltration of the virulent strain *Pseudomonas syringae* pv. *tomato* DC3000 (*Pst*) induced necrosis in the infiltrated zone within 2 dpi, whereas the *hrcC* mutant strain that is unable to form a functional type three secretion system (TTSS) did not induce any visible reaction (**Fig. 26**).

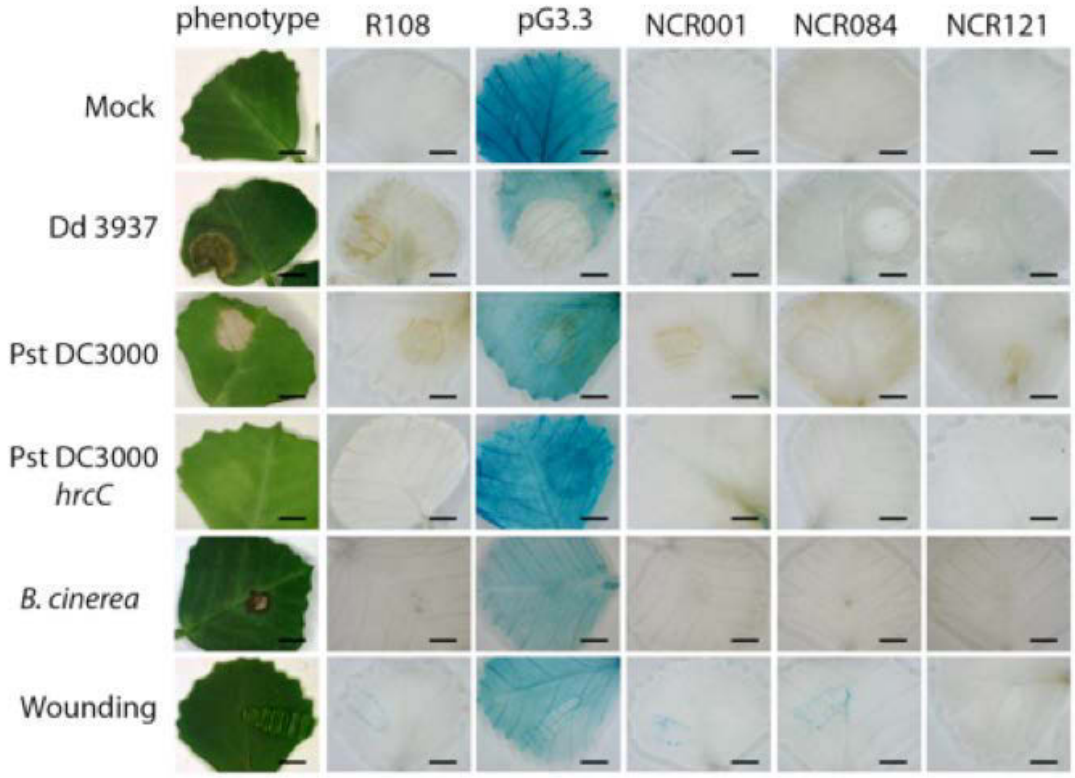

Figure 26. Promoter-GUS analysis of *NCR* genes in infected and wounded leaves

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

NCR expression was not detected in the infiltrated leaflets in either condition (**Fig. 26**). Although *Pst* DC3000 is not described as a natural pathogen of *M. truncatula*, the necrosis induced by the wild-type strain and its absence in the presence of the TTSS mutant suggest that at least some bacterial effector proteins can be specifically recognized by the plant resulting in a necrotic response similar to the hypersensitive response it provokes on non-host *Nicotiana benthamiana* plants (Hann and Rathjen, 2007). Similarly, inoculation of the same *M. truncatula* lines with the necrotrophic polyphagous grey mold-causing fungus

Botrytis cinerea yielded typical symptoms at 7 dpi without inducing any detectable NCR expression (**Fig. 26**). The results of our pathoassays are also in line with a recent study showing that NCR expression was not detected during the compatible interaction of *M. truncatula* with the hemibiotrophic leaf pathogen *Colletotrichum trifolii* or with the biotrophic soil pathogen *Phytophthora medicaginis* (Tesfaye *et al.*, 2013). Altogether, our data and the study from Tesfaye *et al.* (2013) are in agreement with the MtGEA dataset and broaden the conclusion that NCRs are not involved in pathogen responses, whatever the trophic (bio-, hemibio- or necrotrophic) interaction, the host or non-host status and the outcome of the interaction (disease or resistance). Finally, as herbivory and more generally wounding may induce plant defenses around the wounded zone, we also tested the effect of mechanical wounding on NCR expression but again no NCR expression could be detected in the wounded leaflet (**Fig. 26**).

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

The above analyses reveal an extreme specificity in expression for the *NCR* gene family: the genes are only expressed in nodules and not in any other organ or physiological condition. To express this specificity quantitatively and to compare it to other types of specifically expressed genes, the complete MtGEA probe-set was analysed and their Shannon entropy was calculated. Shannon entropy is a metric for characterizing the uniformity of the expression pattern of a gene over the tested conditions (Schug *et al.,* 2005). Low entropy values indicate high tissue specificity while high entropy levels characterize ubiquitous expression.

Ten different tissues were taken into consideration: leaf, petiole, stem, bud, flower, seed, pod, root, nodule and mycorrhiza. For nodule, seed and root the mean value of different developmental stages or experiments was used (for “nodule”: 4, 10, 14 and 28 dpi stages; for “seed”: 10, 12, 16, 20, 24 and 36 dap; for “root” the 0 dpi control for nodulation and an independent experiment); thus in total 19 experiments were used. The Shannon entropy was calculated as described (see materials and methods) (Shug *et al.,* 2005; Zhang *et al.,* 2006) for each of the 50,900 probe-sets using these 10 tissue datasets. The 9000 probe-sets with the lowest entropy (and therefore the highest tissue specificity) were selected for further analysis (**Table 2; http://www.biomedcentral.com/content/supplementary/1471-2164-15-712-S3.xlsx**). A hierarchical cluster analysis of these 9000 genes was calculated (**Fig. 27A**). Clusters of tissue specific genes can be distinguished for root, seed, pod, flower, aerial tissues (l/p/s/b), root tissues (r/n/m), nodule and mycorrhiza. The seed, flower, mycorrhiza and

especially the nodule clusters are enriched in genes with low entropy value E_g (**Fig. 27A**). The strong enrichment of nodule-specific genes among the low entropy genes becomes more obvious when re-ordering the dataset according the increasing entropy levels (**Fig. 27B**). This analysis shows that in *M. truncatula* the genes with the lowest entropy and thus the highest tissue specific expression are mostly nodule-specific genes and to a lesser extent, seed- and flower-specific genes. Strikingly, among the nodule-specific genes, the NCRs are the most represented ones (**Fig. 27B**).

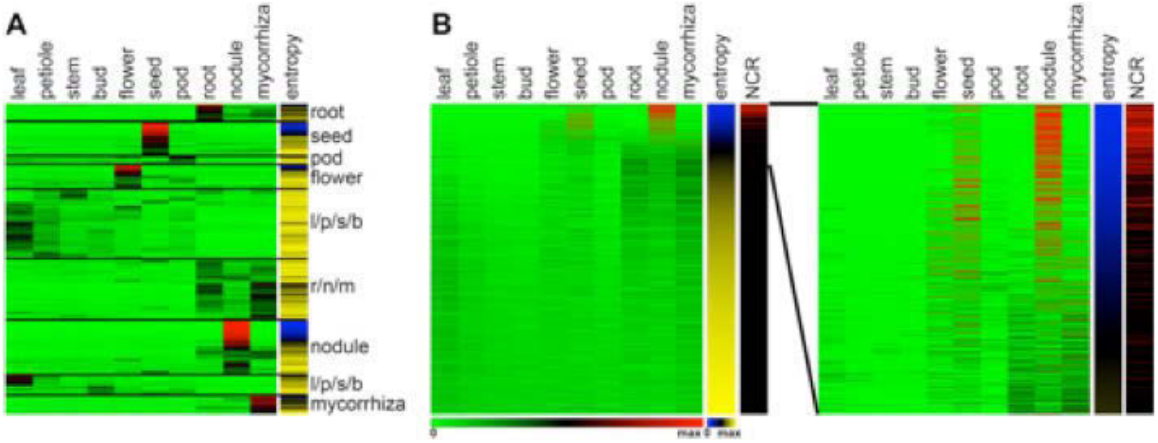

Figure 27. Shannon entropy

(A) Hierarchical clustering of the 9,000 probe-sets with the highest tissue specificity (lowest entropy values E_g). The expression heat map is in green-black-red colour scheme. The entropy heat map is in blue-black-yellow scheme. **(B)** The data set of panel (A) was ordered according to increasing entropy of the genes. The left panel shows the relative expression level of the 9,000 genes in the 10 tissues (green-black-red heat map) and the entropy values E_g (blue-black-yellow heat map). The location of the *NCR* genes is indicated with the black-red heat map (NCR): red means NCR, black means other gene type. The right panel is an enlargement for the first 1,800 genes (entropy values E_g from 0.07 to 1.64). The scale bar for the expression level heat map is green: 0, black: 0.5 and red: 1 and for the entropy heat map blue: 0, black 1.44 and yellow: 2.88 (maximum entropy in the complete data set is 3.32 ($\log_2(N)$).

Although the NCR family is by far the most represented among the nodule-specific low entropy genes, many other known nodule specific genes have very low entropy (**Table 2; http://www.biomedcentral.com/content/supplementary/1471-2164-15-712-S3.xlsx**). These include for example leghemoglobin genes, the glycine-rich peptide (GRPs) genes (Kevei *et al*., 2002; Alunni *et al*., 2007), the small nodulin acidic RNA-binding protein (SNARP) gene family (Laporte *et al*., 2010), genes encoding a small family of secretory calmodulin-like proteins (Mergaert *et al*., 2003; Liu *et al*., 2006), the *DNF2* gene involved in suppression of defence responses in the symbiotic cells (Bourcy *et al*., 2013) and others. Most interestingly, also putative retrotransposons (probe-sets Mtr.9294.1.S1_at and Mtr.636.1.S1_at) and a Dicer 1-like ribonuclease III gene (probe-set Mtr.41531.1.S1_at) are among the nodule specific low

entropy genes (**Table 2; http://www.biomedcentral.com/content/supplementary/1471-2164-15-712-S3.xlsx**).

Besides the high tissue-specificity, another aspect of expression in which the NCRs stand out from the average *M. truncatula* genes is the strength of expression. We used the hybridization signal on the microarrays as a proxy to strength of gene expression. For each of the 50,900 probe-sets in MtGEA, we searched for its maximal signal in the 267 experiments. These hybridization signals vary from 33,500 for the strongest expressed gene to 9 (background) for the weakest. One percent of the probe-sets have an expression level higher than 15,000, 3% higher than 10,000 and 10% higher than 5,000. The mean signal is 1,687 and the median 459 (**Fig. 28A**). The same analysis on the subset of probes corresponding to the *NCR* genes gives a completely different picture: 5% of *NCR* genes have signals above 15,000, 30% above 10,000 and 75% above 5,000 with a mean signal of 7,982 and a median of 7,758 (**Fig. 28B**). Thus, the *NCR* genes are among both the most specific and the most strongly expressed genes in the genome of *M. truncatula*.

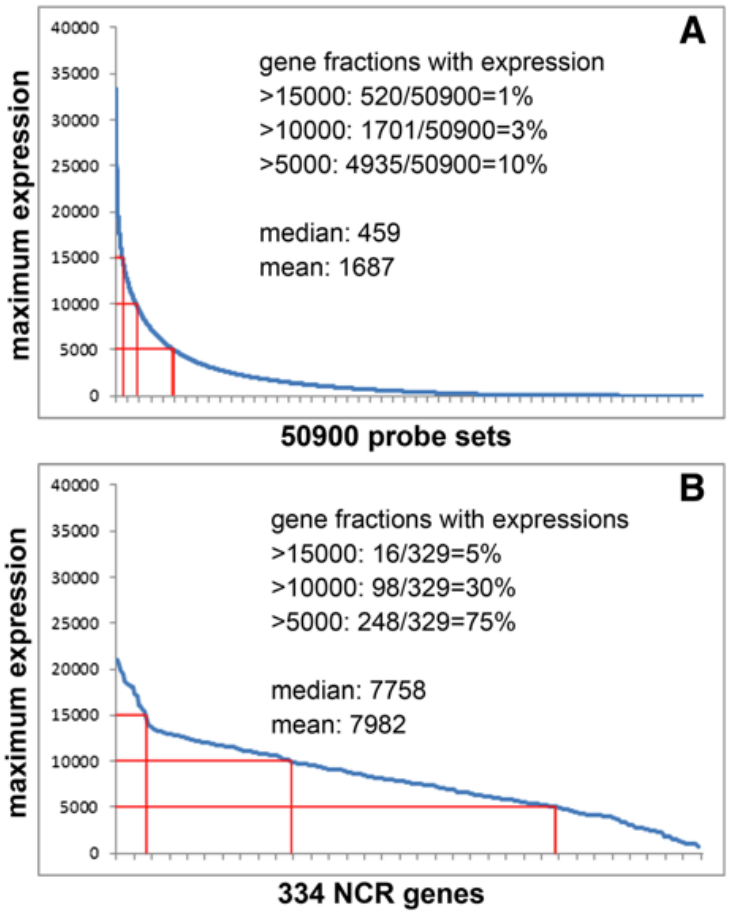

Figure 28. *NCR* genes are among the most actively expressed genes in *M. truncatula*

(A) Maximum expression level of the 50,900 *M. truncatula* probe-sets. **(B)** Maximum expression level of 334 *NCR* genes.

4. Discussion

The expression of *NCR* genes has been studied in *M. truncatula* mostly but also in some other IRLC legumes at the level of individual genes by RT-PCR, *in situ* hybridization, immuno-localization and promoter-marker gene fusions or at the family level by EST-analysis, macroarrays, dedicated microarrays or whole-genome microarrays (Mergaert *et al*., 2003; Alunni *et al*., 2007; Van de Velde *et al*., 2010; Maunoury *et al*., 2010; Moreau *et al*., 2011; Tesfaye *et al*., 2013; Nallu *et al*., 2013; Farkas *et al*., 2014).These studies detected NCR expression only in nodules and in no other tested tissues. By mining publicly available whole-genome transcriptome data, we have extended this analysis of *NCR* gene expression to a very large number of conditions, together covering most plant organs as well as different growth conditions including biotic and abiotic stresses. As a whole, our study suggests that, apart from 5 genes, all NCRs are only expressed in nodules. Moreover, quantifying the specificity of expression with the Shannon entropy factor reveals that the *NCR* genes, and more generally, nodule specific genes are among the most specifically expressed genes in *M. truncatula*. This suggests thus that nodulation in *Medicago* is in large part depending on genes solely dedicated to this symbiotic process. These genes may be resulting from gene duplications followed by neo-functionalization (for example the DNF2 protein which has non-symbiotic homologues) or they may be unique for the symbiosis (possibly the NCRs, GRPs, SNARPs and others). In addition to that, the expression of the *NCR* genes in nodules reaches very high levels. Even if certain *NCR* genes are expressed at a low level, the majority of them are among the highest expressed genes in the whole genome of *Medicago*. This is in agreement with our previous estimation, based on EST counts, that all NCR mRNAs together constitute almost 5% of the total mRNA population in nodules (Mergaert *et al*., 2003).

In accordance with their resemblance to antimicrobial peptides of the innate immunity such as defensins, many NCR peptides, in particular the most cationic ones, have a strong antimicrobial activity against a diversity of bacteria as well as fungi (Van de Velde *et al*., 2010; Tiricz *et al*., 2013; Farkas *et al.*, 2014). Despite this, the *NCR* genes are not expressed in any of the pathogenic interactions of *Medicago* tested here or by Tesfaye *et al*. (2013). This included interactions with bacteria, fungi, oomycetes and nematodes. They are also not expressed in 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. *In situ* detection of NCR expression has demonstrated for all the tested genes that they are specifically expressed in the symbiotic nodule cells but

different subsets of *NCR* genes are activated at different stages of differentiation of these host cells (Mergaert *et al*., 2003; Van de Velde *et al*., 2010; Farkas *et al*., 2014; This work). Transcriptome analysis extended this pattern to the whole family. The *NCR* genes are not activated by Nod factors or during the very early stages of the nodule organogenesis when infected cells are not yet formed (Maunoury *et al.,* 2010; Nallu *et al.,* 2013; This work). During the development of wild-type nodules, they are activated in consecutive waves and their first appearance coincides with the formation of infected symbiotic cells (Maunoury *et al*., 2010). We show here that *NCR* genes are activated during nodule development in at least 3 temporal waves corresponding to specific spatial expression patterns. Genes activated early in nodule development are expressed in the more distal nodule parts (close to the apex) while genes activated late during development are expressed in the proximal nodule tissues. In addition, certain clusters of genes, once activated, maintain their activity when the tissues grow older while other clusters are characterized by a decline of their expression in the older nodule cells. Our spatial analysis of NCR expression is in strong agreement with a recently published study (Roux *et al*., 2014) that used LCM of nodule zones coupled to RNA-Sequencing (**Fig. 29**)

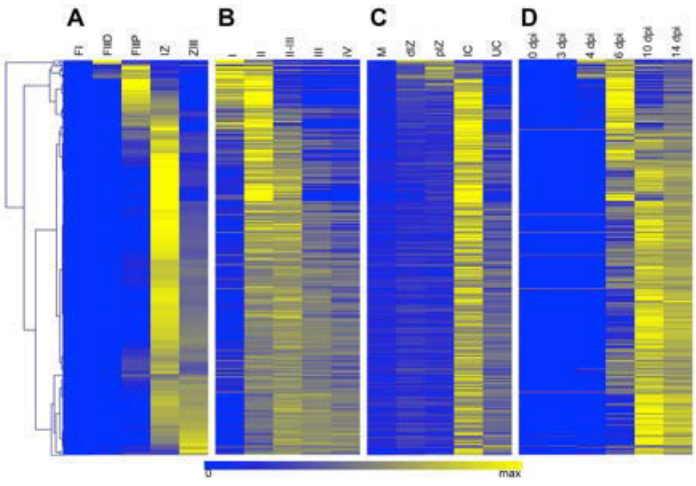

Figure 29. Comparison of the spatial and temporal NCR expression profiles

(A) The expression profiles for the *NCR* gene set described in this study were extracted from the data obtained by laser-capture microdissection coupled to RNA-Seq from the website https://iant.toulouse.inra.fr/symbimics (Ovchinnikova *et al*., 2011). FI corresponds to meristem, FIID to the distal zone II or infection zone, FIIP to the proximal zone II, IZ to the interzone between the infection zone and the fixation zone and ZIII to the fixation zone. Clustering was performed using Pearson correlation. **(B)** The corresponding profiles obtained by hand-dissection of nodules in this study. Samples I, II, II-III, III and IV correspond to the nodule tissues from the most apical part of the nodule with the youngest symbiotic cells to the most proximal part containing the oldest symbiotic cells. **(C)** The corresponding profiles obtained by laser-capture microdissection coupled to affymetrix microarray analysis described by Limpens *et al*. (2013). M corresponds to meristem, dIZ to the distal zone II or infection zone, pIZ to the proximal zone II, IC to infected cell and UC to uninfected cell. **(D)** The corresponding profiles in function of nodule age (dpi). For all data sets, the expression patterns in the different dissected nodule zones are expressed in percentage from their total. The *NCR* genes are ordered identically in all 4 panels according to a hierarchical clustering of the dataset from panel **(A)**.

Transcriptome analysis of non-functional nodules that are formed by bacterial or plant symbiotic mutants and that are arrested at different stages of nodule development, is also in agreement with specific expression of all *NCR* genes in the symbiotic nodule cells: their transcriptional activation is only observed when polyploid symbiotic cells are formed in the mutant nodules (Maunoury *et al.,* 2010). For example, in nodules of the *M. truncatula* TE7 mutant which is affected in the *IPD3* gene (Horvath *et al.,* 2011; Ovchinnikova *et al.,* 2011) and in nodules infected by the *S. meliloti exoY* mutant, no infected and polyploid symbiotic cells are formed and these nodules do not express any of the *NCR* genes (Maunoury *et al.,* 2010). Conversely, in nodules infected by the *S. meliloti bacA* mutant which contain symbiotic cells with undifferentiated bacteroids, only a subset of *NCR* genes is activated while in other mutants, forming normal symbiotic cells with differentiated bacteroids, *NCR* genes are activated to a similar extent as in the wild-type (Maunoury *et al.,* 2010; Nallu *et al.,* 2013). Together, the expression pattern of all the tested *NCR* genes suggests that the endosymbiotic rhizobia in the symbiotic nodule cells are the only targets of the peptides. However, the distinct spatio-temporal profiles clearly suggest that NCR peptides have many different roles. Subsets of *NCR* genes that are expressed during the early stages of symbiotic cell formation might be involved in the elongation and polyploidization of the bacteroids while other subsets that are active in later stages of symbiotic cell formation or even after the completion of the symbiotic cell differentiation might have other functions in the bacteroids. Moreover, we find that the expression of the NCRs is shut down when nodule senescence is activated, meaning that the antimicrobial NCR peptides have no direct role in the lysis and digestion of the bacteroids that is taking place during senescence of nodules.

Very little is known about how the very specific regulation of NCRs is achieved. Since their expression is correlated with bacterial infection of the symbiotic cells, the perception of bacterial signals such as components of the bacterial envelope could be involved. The transcription factor EFD, belonging to the ethylene response factor family, may control, directly or indirectly, the expression of a subset of *NCR* genes since a mutant forms nodules in which some of the *NCR* genes are down regulated and in which bacteroid differentiation is partially impaired (Vernié *et al.,* 2008). The IPD3 protein is another transcription factor that might be involved, directly or more likely indirectly, in the regulation of the *NCR* genes and the other symbiotic cell specific genes (Singh *et al*., 2014; Limpens and Bisseling, 2014). Indeed, in the *M. truncatula ipd3* mutant nodules, the symbiotic cells do not form and the symbiotic cell specific genes, including the NCRs, are not activated (Maunoury *et al*., 2010).

In agreement with this, the IPD3 gene is expressed in the whole nodule and its expression domain overlaps with the NCR expression zone (Messinese *et al.*, 2007).

Searching for potential *cis*-elements in the promoters of *NCR* genes with different algorithms yielded 5 different conserved motifs of 41 to 50 bp, which are specifically enriched in the 1000 bp promoter regions (Nallu *et al*., 2013). Some of these motifs show resemblance to previously described motifs conferring nodule-specific gene expression. However, the role, if any, of these motifs in the remarkable expression pattern of the *NCR* genes needs further investigation. Interestingly, some of these motifs include Auxin Response Factor binding sites that may suggest a role for auxin in NCR regulation (Nallu *et al*., 2013). However, MtGEA dataset from auxin treated seedlings do not show any NCR induction, pointing out a more complex regulatory mechanism controlling NCR expression.

The very tight regulation of the *NCR* genes that was revealed here might indicate that besides *cis*- and *trans*-acting factors, regulation at the level of chromatin might also be involved in the activation of the *NCR* genes. Moreover, endoreduplication seems to be a prerequisite for their activation (Maunoury *et al.,* 2010) and might thus, by a still unknown mechanism, be implicated in the activation of this gene family. In that respect, it is interesting to note that the giant feeding cells induced by the nematode *M. incognita* are highly polyploid cells and appear to express faintly a few *NCR* genes. Nevertheless, this observation could be an experimental artefact and will require further experimental confirmation.

Genes with high tissue-specific expression are often actively silenced during most of the plant growth by epigenetic mechanisms. Since in *M. truncatula* the nodule-specific genes display the highest level of expression specificity, it might be worthwhile to investigate if epigenetic control is important in the regulation of the symbiotic cell-specific genes. The nodule-specific expression of putative retrotransposons, which are usually epigenetically silenced, and the Dicer 1-like ribonuclease III gene, which may have a role in epigenetic regulation, as well as the identification of small RNAs potentially targeting *NCR* genes (Lelandais-Brière *et al.*, 2009) are all in agreement with such an epigenetic control of the symbiotic cell-specific genes.

Plant genomes contain large numbers, several hundreds to thousands, of resistance genes (*R*) of the NB-LRR family that recognize specific pathogen effectors and trigger resistance. Silencing by microRNAs has been proposed as a mechanism to avoid unregulated expression of *R*genes, which may be a threat to the plant and represent a fitness cost (Li *et al*., 2012). Why *M. truncatula* maintains such a large repertoire of *NCR* genes is not known. It is also not known whether closely related legumes of the IRLC have an arsenal of NCRs of

similar size. However, it seems likely to us that expressing such a large gene family might be a fitness cost for the plant that is not to be neglected. Therefore, keeping the whole gene family under a very tight regulatory control might be essential for the plant.

Supporting Information

Additional supporting information may be found in the online version of the article on the web site of the journal BMC Genomics.

Additional file 1: Table S1 (named in this thesis as **Table 1**). NCR probe-sets and expression data. The NCR probe-set annotations are from the Affymetrix Medicago Gene Chip. The expression data were extracted from MtGEA, NCBI GEO, and from the spatially-resolved nodule dataset.

http://www.biomedcentral.com/content/supplementary/1471-2164-15-712-S1.xlsx

Additional file 3: Table S2 (named in this thesis as **Table 2**). Entropy of gene expression. The 9,000 probe-sets with the lowest entropy.

http://www.biomedcentral.com/content/supplementary/1471-2164-15-712-S3.xlsx

Chapter. II

Convergent evolution of endosymbiont differentiation in Dalbergoid and IRLC legumes mediated by nodule-specific cysteine-rich peptides

Convergent evolution of endosymbiont differentiation in Dalbergoid and IRLC legumes mediated by nodule-specific cysteine-rich peptides*

1. Abstract

Nutritional symbiotic interactions require the housing of large numbers of microbial symbionts, which produce essential compounds for the growth of the host. In the legume-rhizobium nitrogen fixing symbiosis, thousands of rhizobium microsymbionts, called bacteroids, are confined intracellularly within highly specialized symbiotic host cells. In IRLC legumes such as *Medicago*, the bacteroids are kept under control by an arsenal of nodule-specific cysteine-rich (NCR) peptides, which induce the bacteria in an irreversible, strongly elongated and polyploid state. Here we show that in *Aeschynomene* legumes belonging to the more ancient Dalbergoid lineage, bacteroids are elongated or spherical depending on the *Aeschynomene* species and that these bacteroids are terminally differentiated and polyploid, similarly to bacteroids in IRLC legumes. Transcriptome, *in situ* hybridization and proteome analysis demonstrated that the symbiotic cells in the *Aeschynomene* nodules produce a large diversity of NCR-like peptides, which are transported to the bacteroids. Blocking NCR transport by RNAi-mediated inactivation of the secretory pathway inhibits bacteroid differentiation. Together, our results suggest that bacteroid differentiation in the Dalbergoid clade, which likely evolved independently from the bacteroid differentiation in the IRLC clade, is based on very similar mechanisms used by IRLC legumes.

*__Modified from:__ Czernic, P., Djamel, G., Cartieaux, F., Moulin, L., **Guefrachi, I**., Patrel, D., Pierre, O., Fardoux J., Chaintreuil, C., Nguyen, P., Gressent, F., Da Silva, C., Poulain, J., Wincker, P., Rofidal, V., Hem, S., Arrighi, J.F., Mergaert, P and Giraud, E. (2015).

2. Introduction

Legumes, thanks to their ability to develop a symbiotic interaction with nitrogen fixing bacteria, collectively called rhizobia, are among the agronomically and ecologically most important plants. This symbiosis results in the formation of new organs, the nodules, inside which the bacteria fix nitrogen and transfer the ammonium to the plant. Two distinct phases can be observed during the establishment of a functional nodule, the early phases that lead to nodule organogenesis and the release of the bacteria in the host cells and, the late phases, in which the rhizobia differentiate into an endosymbiotic form, the bacteroids, able to fix atmospheric nitrogen. During this second phase, profound modifications of the metabolism of the rhizobia are observed and this can be accompanied by a marked change in the bacterial cell shape and size (Haag *et al*., 2013). Three different bacteroid morphotypes have been observed in different legume species (Oono *et al*., 2010; Kondorosi *et al*., 2013): **i**) elongated or E-morphotype bacteroids described in legumes of the Inverted Repeat-Lacking Clade (IRLC) (*Medicago*, *Pisum*, *Vicia*, …) and some *Aeschynomene* species such as *A. afraspera*, **ii**) enlarged, spherical bacteroids (S-morphotype) encountered in some species of the Dalbergoid clade (such as *Aeschynomene indica*, *Aeschynomene evenia* and *Arachis hypogaea*) and **iii**) unmodified bacteroids (U-morphotype) which display a rod-shape morphology similar to free-living bacteria found in *e.g*. Phaseoloid or Robinoid legumes (i.e. *Phaseolus, Vignia, Lotus, Glycine, Sesbania*). The fact that the same rhizobium strain nodulating legumes of different clades can display different morphotypes is strong evidence for the host plant governing bacteroid morphotype (Sen and Weaver, 1984; Mergaert *et al*., 2006; Oono *et al*., 2010). The change of shape is probably the tip of the iceberg of the control exerted by the plant of the bacterial physiology during symbiosis. Indeed, besides their E-morphotype, the *Sinorhizobium meliloti* bacteroids in *Medicago truncatula* differ from their free-living state in several aspects, they become polyploid, their membrane permeability dramatically increases and they lose their reproductive capacity. Bacteroids that display such drastic changes are considered terminally differentiated because they are unable to revert to their bacterial form (Mergaert *et al*., 2006).

In *Medicago truncatula*, a class of peptides, named NCRs (Nodule-specific Cysteine-Rich peptides) plays a key role in this terminal bacteroid differentiation (Van de Velde *et al*., 2010). The *M. truncatula* NCR family is composed of about 600 highly divergent genes, which are nearly all exclusively expressed in nodules (Mergaert *et al*., 2003; Alunni *et al*., 2007; Young *et al*., 2011). Within nodules, the NCR peptides are only produced by the

infected symbiotic cells and are transported to the bacteroid-containing symbiosomes. NCRs are similar to the defensin-type of antimicrobial peptides. The peptides contain an N-terminal secretion signal, the mature peptides are usually no longer than 60 amino acids and have 4 to 6 conserved cysteine residues (Mergaert *et al*., 2003). The cleavage of the signal peptide by a nodule-specific signal peptidase complex (SPC) located in the endoplasmic reticulum is required to permit the trafficking of NCR peptides to the symbiosome compartment. Bacteroid and symbiosome development is blocked in a *M. truncatula dnf1* mutant that is affected in the SPC22 subunit of this nodule-specific SPC (Wang *et al*., 2010; Van de Velde *et al*., 2010). Some NCR peptides have been found to have antimicrobial activity, killing rhizobium when applied at high concentration (Van de Velde *et al*., 2010). However, these antimicrobial peptide-like NCRs can at lower concentration induce typical features of E-morphotype bacteroids *in vitro* on cultured rhizobium or *in planta* when expressed in transgenic *Lotus japonicus* plants which form normally U-morphotype bacteroids (Van de Velde *et al*., 2010). The *NCR* gene family has been identified in all investigated legume species of the IRLC clade but in no other plant species outside of this clade. This raises the question about the nature of the factors used by the other legumes to control bacteroid metamorphosis.

In the Dalbergoid legume clade, bacteroids can be of the E- or S-morphotype. For example, within the *Aeschynomene* genus, *A. afraspera* has E-type bacteroids but *A. indica* or *A. evenia* have S-type bacteroids (Bonaldi *et al*., 2011; Arrighi *et al*., 2012). The bacteroid morphotype in the *Aeschynomene* genus is also under plant control since the *Bradyrhizobium* strain ORS285 which can nodulate *A. afraspera* and *A. indica* or *A. evenia* transforms into E-morphotype bacteroids in nodules of the former legume and in S-morphotype bacteroids in the latter two legumes (Bonaldi *et al*., 2011; Arrighi *et al*., 2012). The *Aeschynomene* symbiosis with *Bradyrhizobium* sp. has several other particular features, including the formation of nodules on both roots and stems, photosynthesis by the *Bradyrhizobium* symbionts which is required for optimal stem nodulation (Giraud *et al*., 2000) and the use of a Nod factor independent nodulation pathway in some *Aeschynomene* species including *A. evenia* and *A. indica* (Giraud *et al*., 2007; Arrighi *et al*., 2012). Furthermore, in contrast to classical determinate and indeterminate nodules, aeschynomenoid nodules originate from the successive divisions of only one or a few root cortical cells initially infected by an infection thread independent mechanism (Bonaldi *et al*., 2011). This infection mechanism has two consequences: **i)** all the nodule primordium cells are infected and **ii)** all the nodule primordium cells and bacteria are synchronized. Thus, during nodule development of *A.*

indica nodules, the change of bacteroid shape occurs simultaneously for all nodule cells and takes place between the 4th and 5th day after inoculation (Bonaldi *et al.,* 2011). The possibility to obtain sampling of homogenous materials at different differentiation stages and the existence of two distinct bacteroid morphotypes in closly related species make the *Bradyrhizobium-Aeschynomene* symbiotic couples an appealing model system for unraveling the mechanisms of bacteroid morphotype determination in legume species outside the IRLC clade.

In this study, we show that E- and S-morphotype bacteroids from *A. afraspera* and *A. indica* nodules display typical features of terminal differentiation. Furthermore, in these two *Aeschynomene* species, we provide evidences of the presence of a NCR-like peptide family which is specifically produced in the nodules and targets the bacteroids. Alteration of the trafficking of these peptides by silencing the *dnf1* homolog identified in *A. evenia* correlated with the absence of bacterial differentiation indicating that the NCR-like peptides identified in *Aeschynomene* species play a similar role to those described in *Medicago*. Altogether, these results suggest a convergent evolution of the host molecular mechanisms governing bacteroid differentiation in legumes.

3. Results

3.1. E and S-bacteroids in *Aeschynomene* species are terminally differentiated

Previous studies revealed that *Bradyrhizobium* strain ORS285 displayed two distinct bacteroid morphotypes depending on the *Aeschynomene* host: the E-morphotype in *A. indica* and the S-morphotype in *A. afraspera* (**Fig. 30A**) (Bonaldi *et al.,* 2011). As this difference in shape could imply different properties, we analyzed in each bacteroid type, isolated from mature nodules at 14 days post inoculation (dpi), several characteristics that have been shown to change during bacteroid metamorphosis in *Medicago*, *i.e.* the bacterial DNA content, the membrane permeability and the viability. As revealed by flow cytometry analysis (**Fig. 30B**), the DNA content of E- and S- bacteroids peaked respectively at around 7C and 16C in comparison to the 1C/2C DNA content of free-living bacteria. These ploidy levels are lower than that described for *S. meliloti* bacteroids (24C) (Mergeart *et al.*, 2006) although other studies have shown lower ploidy levels also in *Medicago* (Sinharoy *et al.*, 2013, Berrabah *et al.*, 2014).

Bacteroid membrane integrity was estimated using PI, a DNA stain that is excluded from living cells but penetrates cells displaying impaired membrane integrity (Mergeart *et al.*, 2006). Fluorescence images of the PI-stained bacteroids showed that the membrane

permeability of E- and S-morphotype bacteroids was greater than that of the free-living bacteria (**Fig. 30A**).

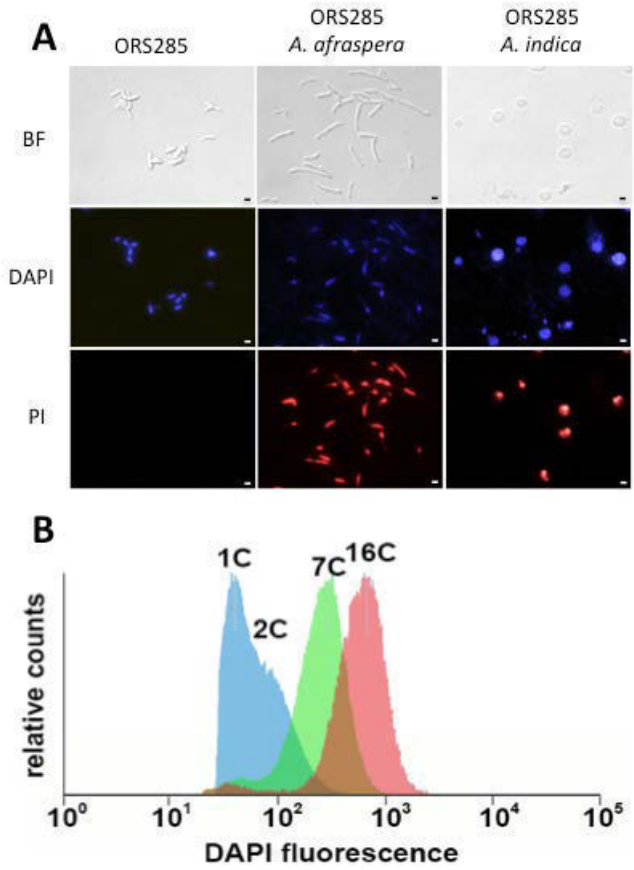

Figure 30. Properties of free-living cultured *Bradyrhizobium* sp. ORS285 bacteria and ORS285 bacteroids isolated from *A. afraspera* or *A. indica* nodules

(A) Normaski bright field (BF, upper panels) and fluorescence microcopy of bacteria and bacteroids stained with DAPI (middle panels) and PI (lower panels). Scale bars, 1 µm. **(B)** DNA content of DAPI-stained bacteria and bacteroids measured by flow cytometry. In blue, free-living bacteria, in green, bacteroids isolated from *A. afraspera* and in red, bacteroids isolated from *A. indica*.

Indeed, 88% of S-bacteroids and 95% of E-bacteroids were found to be permeable to PI compared with 7% for free-living bacteria. Whereas, PI is known to stain about 50% of *S. meliloti* bacteroids (Mergeart *et al*., 2006). The loss of viability measured for S-bacteroids from *A. indica* nodules (98%) was also comparable to the one previously estimated for *S. meliloti* bacteroids (99%) (Mergeart *et al*., 2006). In contrast, intracellular bacteroids of *A. afraspera* appeared more viable as 22 % formed colonies on agar plates. However, this higher value could result to the fact that *A. afraspera* nodules displayed the unusual presence of two infected tissues, a large central tissue in which the bacteria elongated and a superficial tissue in which the shape of the bacteria remained unmodified (Bonaldi *et al*., 2011).

The intracellular cells extracted from nodules consisted therefore of a mixture of differentiated and non-differentiated bacteria suggesting that the loss of viability of E-bacteroids could be far greater. Taken together, we observed that the E- and S- bacteroids from *Aeschynomene* nodules share the same features as the *S. meliloti* bacteroids in *Medicago* and therefore, they could be considered as terminally differentiated.

3.2. Distribution of S- and E-morphotype bacteroids among *Aeschynomene* species

To obtain more insight into the distribution of the E- and S-bacteroid morphotypes among the *Aeschynomene* species, we analyzed by confocal microscopy the shape of intracellular bacteria from mature nodules of various *Aeschynomene* species elicited by the ORS285 strain tagged with GFP. Unlike the other photosynthetic *Bradyrhizobium* strains, such as ORS278 and BTAi1, this bacterium contains the canonical *nodABC* genes and displays a broader host range due to its ability to use a Nod-dependent and a Nod-independent symbiotic process depending on the host plant (Bonaldi *et al*., 2011). We observed in all tested species using a Nod-independent process (*A. indica, Aeschynomene evenia* ssp. *serrulata, A. evenia*. ssp. *evenia, Aeschynomene virginica, Aeschynomene scabra,* and *Aeschynomene sensitiva*) that the bacteroids displayed an S-morphotype whereas in the three Nod-dependent species tested (*A. afraspera*, *Aeschynomene nilotica*, *Aeschynomene aspera*) the bacteroids displayed an E-morphotype (**Fig. 31**). We also analyzed the bacteroid shape in another group of *Aeschynomene* species not nodulated by the ORS285 strain but by non-photosynthetic *Bradyrhizobium* strains containing *nod* genes such as ORS301, ORS302 and ORS305 (Molouba *et al*., 1999). We also observed in these species (*Aeschynomene americana*, *Aeschynomene schimperi*, *Aeschynomene pfundii*) that the bacteroids, stained using the live/dead staining procedure, displayed an E-morphotype. Altogether these data indicate that S-morphotype bacteroids are specific for the *Aeschynomene* species using a Nod-independent symbiotic process, while the E-morphotype is specific to the *Aeschynomene* using a Nod-dependent one.

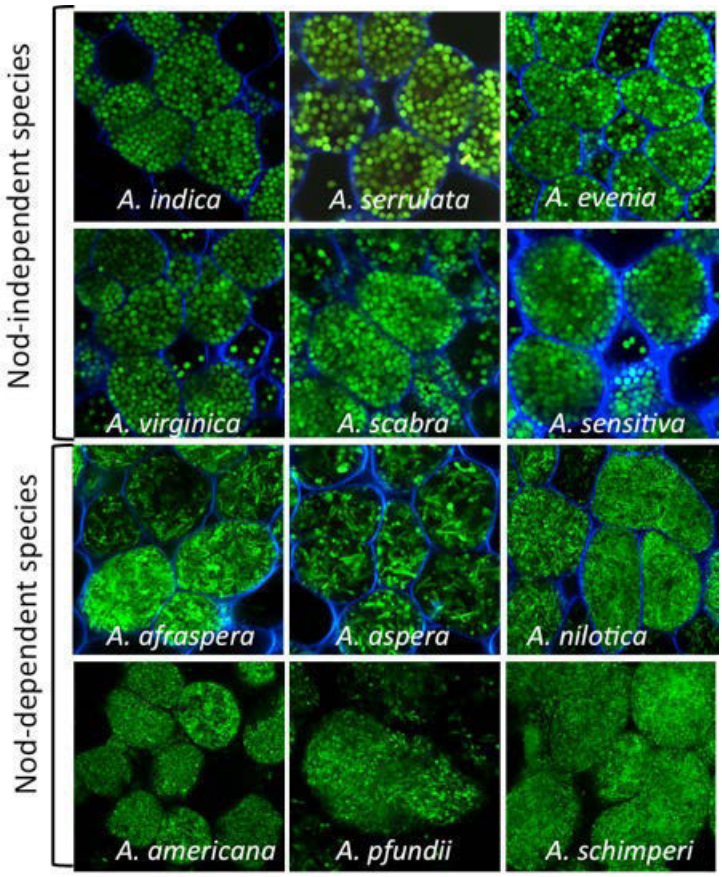

Figure 31. Distribution of S and E- Bacteroid morphotypes among *Aeschynomene* species

The species *A. indica, A. evenia* spp. *serrulata, A. evenia* spp. *evenia, A. virginica, A. scabra, A. sensitiva*, *A. afraspera*, *A. nilotica*, *A. aspera* were nodulated by the ORS285 GFP-tagged strain; the species *A. americana*, *A. schimperi*, *A. pfundii* were nodulated respectively by *Bradyrhizobium* strains ORS301, ORS302 and ORS305 and bacteroids were stained by the SYTO9 fluorescent probe of the Live/dead stain.

3.3. *Aeschynomene* species contain a new class of nodule-specific cysteine-rich peptides-encoding genes

Our results indicate that in *Aeschynomene* nodules, the endosymbiotic bacteria undergo terminal differentiation similar to that described in nodules of legume species belonging to the IRLC clade. This raises the question if NCR peptides are also recruited in *Aeschynomene* to govern this differentiation step. To check this possibility, we analyzed 4 expressed sequence tag (EST) libraries previously constructed in our laboratory and corresponding to non-inoculated roots and mature nodules from *A. afraspera* and *A. indica* plants inoculated with ORS285 strain. Although the total number of cDNAs sequenced for each library was relatively small, around 9500 ESTs/library (**Table 3; Annex 1**), we postulated that if NCR peptides were also involved in bacteroid differentiation in *Aeschynomene*, they should be highly expressed specifically in the nodules and hence easily detectable. We first performed a BLAST search on the EST libraries using several *M.*

truncatula NCR genes as a query but no conclusive results were obtained. Considering that *NCR* genes are rapidly evolving and are highly diverse, even within *M. truncatula* (Alunni *et al.*, 2007; Branca *et al.*, 2011), we re-analyzed the available EST databases using the following parameters determined from typical features of the NCR family: **i**) candidate genes should encode small proteins (fewer than 100 aa) containing a signal peptide, **ii**) the gene expression should be nodule specific with a significant read count in the EST database (we arbitrarily fixed a lower limit of 5 reads) and **iii**) the candidate genes should encode peptides rich in Cys residues (at least 4). This second *in silico* analysis has permitted identification of 15 genes encoding putative NCR-like peptides from the *A. afraspera* nodule library and 18 from the library of *A. indica* nodules (**Table 4 grey shaded; Annex 2**). These genes were named *Aa-* or *AiNCRs* and numbered according their rank in read counts. The deduced mature peptides (i.e. without their signal peptide) are between 37 and 85 residues long and contain 4 to 10 Cys (**Table 4; Annex 2**). The alignment of these NCR-like peptides highlighted, for 31 out of the 33 sequences, a conserved pattern of 6 Cys within the sequence of the mature peptides (**Fig. 32A**).

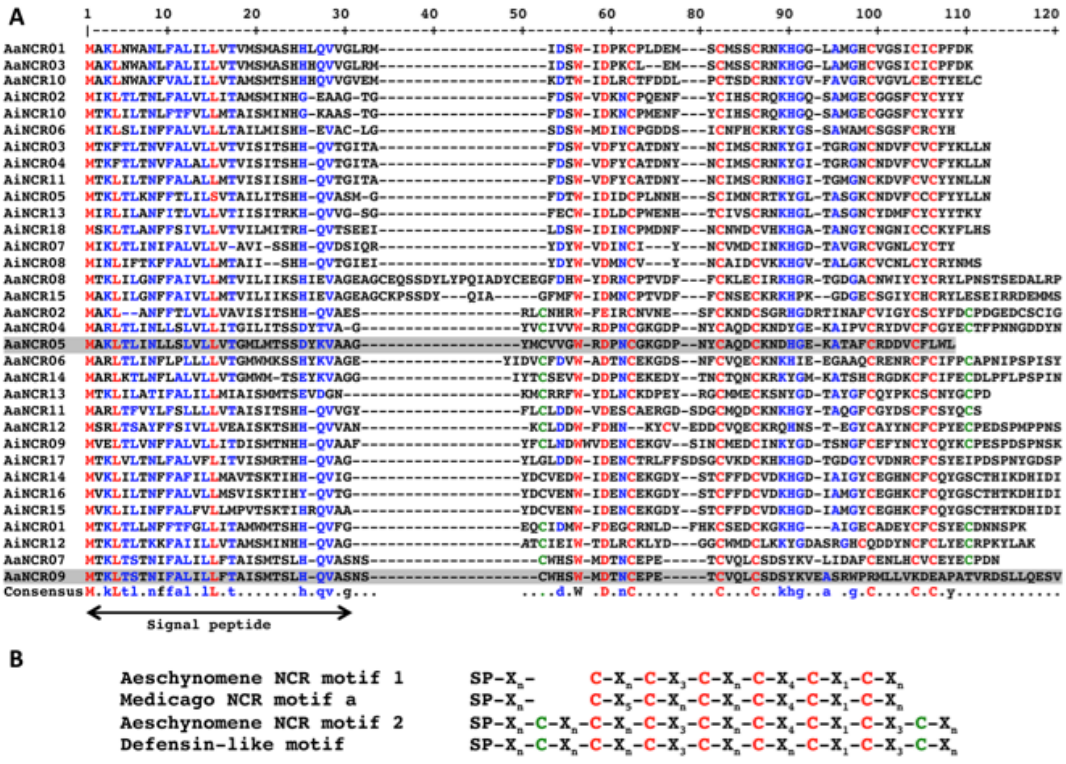

Figure 32. Alignment and cysteine-signature of *Aeschynomene* NCRs

(A) The deduced protein sequences from the up-regulated *NCR* genes from *A. indica* (AiNCRs) or *A. afraspera* (AaNCRs) were aligned using Multalin (http://multalin.toulouse.inra.fr/multalin/). Highly conserved amino acids are in red and conservative substitutions in blue. The putative location of the signal peptide is indicated below the alignment. **(B)** Comparison of the cysteine-rich motifs of *Aeschynomene* NCRs to those of *Medicago* and to defensin-like peptides. Red and blue Cysteines correspond to the NCR- and defensin-signatures respectively.

This pattern is very close to the one previously described for a group of *Medicago* NCRs but distinct by the spacing between the first three Cys. This led us to propose a specific *Aeschynomene* NCR-like motif. Interestingly, 11 of these NCR-like peptides contained 2 extra-Cys, giving rise to a typical defensin signature that was named motif 2 (**Fig. 32B**).

By analogy with *M. truncatula,* for which the NCR family is composed of about 600 genes that exhibit distinct temporal and spatial expression patterns in nodules (Mergaert *et al*., 2003; Young *et al*., 2011; Guefrachi *et al*., 2014), the involvement of a larger number of *NCR* genes is also expected in *Aeschynomene*. We therefore considered the possibility that our criteria were too selective leading to a restricted number of candidate genes identified. By performing BLAST searches against the same EST libraries but using this time the first candidate genes that emerged, we identified a total of 38 and 44 *NCR-like* genes in *A. afraspera* and *A. indica*, respectively (**Table 4; Annex 2**). All these additional genes were found to be specifically expressed in the nodules and a global alignment show that all of them except 3 contained the *Aeschynomene* NCR-motif 1 or 2.

Furthermore, a BLAST search against transcriptome data obtained from *A. evenia* ssp. *serrulata* mature nodules made by the 454 technology which allows the detection of a larger number of transcripts, revealed an even higher number of *NCR-like* genes of more than 80 (**Table 4; Annex 2**). Altogether, these data suggest that the *NCR-like* genes constitute also in *Aeschynomene* a large family, which could count several tens or even hundreds of candidates.

3.4. *NCR-like* genes of *Aeschynomene* sp. are expressed before or concomitantly with bacteroid differentiation

In order to analyze the temporal expression of the *Aa-* and *AiNCR* genes, we performed a RT-qPCR analysis of 4 of them for each *Aeschynomene* species. Root tissues of *A. afraspera* and *A. indica* were sampled at T0 and at different time points after inoculation with *Bradyrhizobium* strain ORS285 (6h, 1, 2, 3, 4, 5, 6, 7 and 14 days). As a symbiotic marker, we also monitored Leghemoglobin (*LegHb*) mRNA accumulation. As shown in **Fig. 33A** and **33B**, mRNA corresponding to *LegHb* started to accumulate at 5 dpi for both species, i. e. when the bacteroid differentiation process is achieved and a nitrogenase activity is detectable (Bonaldi *et al*., 2011). Interestingly, while the *NCR-like* gene transcripts were not detected in control roots or during the early time points of nodule formation, they were all strongly expressed at 5dpi and some of them even started to accumulate 1 or 2 days before, which coincided with the beginning of bacteroid morphogenesis (Bonaldi *et al*., 2011).

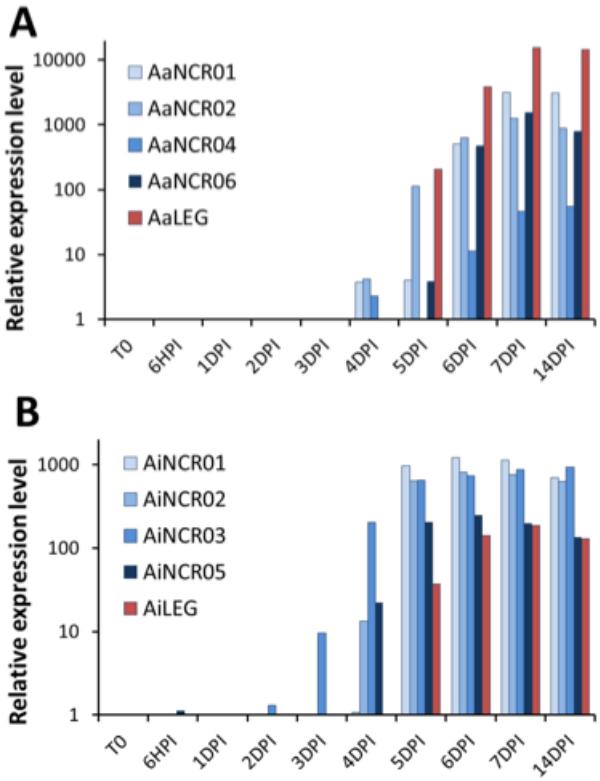

Figure 33. NCR expression during the symbiotic process

(A) NCR expression during *A. afraspera* and **(B)** *A. indica* nodule development. Plants were inoculated with the *Bradyrhizobium* ORS285 strain and the roots were sampled at T0 and at different time points as indicated. The relative expression level of the NCRs was determined by RT-qPCR and normalized by the expression of EF1α. As a control of nodule development, the expression of Leghemoglobin (LEG) was also measured.

To complete this expression analysis, we also checked the expression by RT-qPCR, in both roots and nodules, of 3 candidates *NCR* genes identified in *A. evenia* by BLAST. As expected, the 3 genes appeared specifically and highly expressed in *A. evenia* nodules (**Fig. 34**).

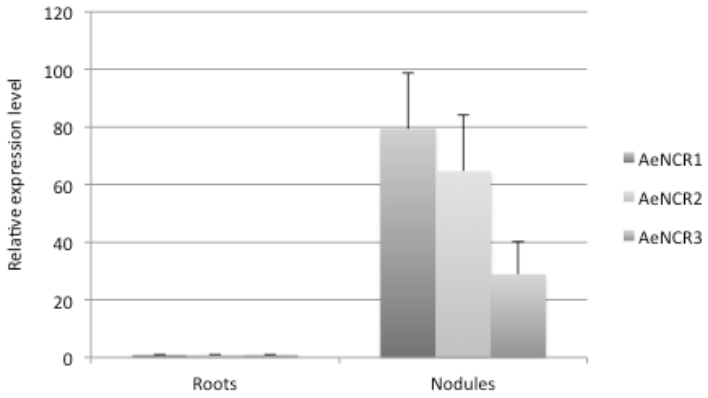

Figure 34. Nodule specific expression of *NCR* genes in *A. evenia*

RNA was extracted from *A. evenia* spp. *serrulata* roots or matures nodules 14 days after inoculation with the *Bradyrhizobium* strain ORS285. Expression levels were calculated relative to EF1α. Error bars indicate standard errors of the means of 3 biological replicates.

3.5. *NCR-like* genes are expressed specifically in infected cells of *Aeschynomene* sp. nodules

In *Medicago*, the NCRs are specifically synthetized in the infected plant cells and are targeted to the bacteria (Van de Velde *et al*., 2010). To check if the expression of the identified *Aa-* and *AiNCR* genes was also restricted to the infected cells, we performed an *in situ* hybridization on 14-days old nodules of *A. afraspera* and *A. indica*, infected with the ORS285 strain. We used as probes *AaNCR01* and *AiNCR01*, which were found to be the most highly expressed *NCR* genes in each species. In addition, the *LegHb* gene, known to be expressed in infected nodule cells, was included as a positive control for the hybridization experiment. As shown in **Fig. 35**, a strong signal was obtained using the *LegHb* antisense DNA as a probe. The signal was restricted to the infected cells (**Fig. 35B and E**), consistent with the known role of this protein in symbiosis. Both *A. afraspera* and *A. indica NCR* antisense probes gave also a signal circumscribed to the infected cells (**Fig. 35C and F**) similarly as observed for *LegHb*. Control hybridizations without probe or with a sense *LegHb* probe or sense *NCR-like* probes (**Fig. 35A and D**) showed a complete absence of signal.

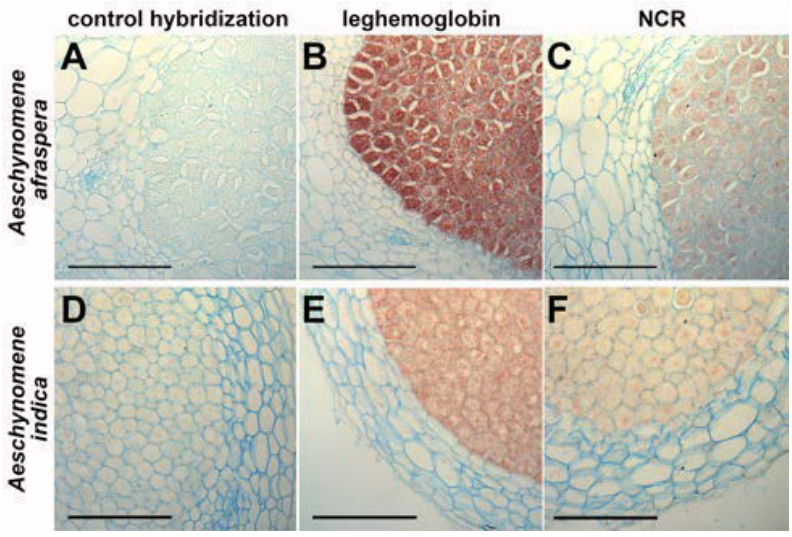

Figure 35. *NCR* genes are specifically expressed in the infected symbiotic cells of *Aeschynomene* nodules

Sections of 14 dpi *A. indica* and *A. afraspera* nodules, infected with the *Bradyrhizobium* ORS285 strain, were analysed by *in situ* hybridization with antisense probes of leghemoglobin **(B, E)** or NCR **(C, F)**. Control hybridizations were done with the sense NCR probe **(A, D)**. Bars represent 200 µm.

3.6. NCR-like peptides target bacteroids in Aeschynomene nodules

To test the possibility that Aa- and AiNCR mature peptides target the endosymbiotic bacteria in the nodule cells that express the corresponding genes, we set up a proteome approach. Total protein extracts were prepared from nodules and from purified bacteroids

from both *A. afraspera* and *A. indica*. In addition, a control extract was prepared from free-living ORS285 bacteria. Proteins were subjected to Tricine-PAGE analysis. Slight differences in the patterns of low molecular weight proteins (between 4 to 9 kDa), where we expected to identify Aa- or AiNCR peptides, could be observed between nodule or bacteroid extracts and extracts from free-living bacteria (**Fig. 36**). We identified the corresponding proteins by mass spectrometry. From the bacteroid extracts isolated from *A. afraspera* and *A. indica* nodules, 51 and 42 proteins were identified respectively. Among them, in both nodule contexts, 4 corresponded to plant proteins, of which three were identified as NCR peptides. The additional plant protein identified in *A. indica* bacteroids was a putative monosaccharide-proton symporter corresponding probably to a symbiosome located transporter whereas there was an extensin-like protein of unknown function in *A. afraspera* bacteroids. Among the 40 proteins identified in free-living bacterial cells, all corresponded to bacterial proteins and none displayed the characteristics of cysteine-rich peptides.

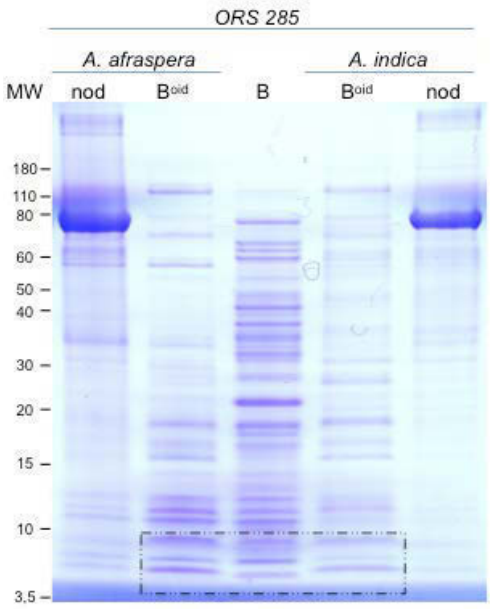

Figure 36. NCR peptides colocalize with bacteroids in *A. afraspera* and *A. indica* nodules

SDS-PAGE analysis and Coomassie blue staining of total protein extracts of nodules (nod), bacteroids (Boid), and cultured *Bradyrhizobium* strain ORS285 cells **(B)** reveal the specific presence of low molecular weight peptides in nodules and bacteroids. The rectangle in the dashed line corresponds to the region of the gel analysed by LC-MS/MS.

3.7. The DNF1 signal peptidase complex is recruited for bacteroid differentiation in *Aeschynomene*

To obtain genetic evidence for a role of the *Aeschynomene* NCR peptides in bacterial differentiation, we applied a strategy aiming to interfere with their targeting to the symbiosomes, similarly as described before in *M. truncatula* (Van de Velde *et al.*, 2010;

Wang *et al*., 2010). NCR peptides are secretory proteins with an N-terminal signal peptide. Therefore, their transport requires the removal of this signal peptide by the signal peptidase complex machinery, which is located in the membrane of the endoplasmic reticulum. Reduced SPC activity is supposed to hamper the transport of the NCR peptides to the symbiosomes, similarly as in the *dnf1* mutant of *M. truncatula* (Van de Velde *et al*., 2010; Wang *et al*., 2010). To develop such loss-of-function strategy in *Aeschynomene*, we chose to conduct this experiment on *A. evenia* ssp. *serrulata* which is diploid contrary to *A. indica* and *A. afraspera* which are allopolyploid (Renard *et al*., 1983; Arrighi *et al*., 2014). A BLAST search using MtDNF1 allowed the identification in the *A. evenia* ssp. *serrulata* transcriptome database of a single candidate gene, which we named *AeDNF1*, encoding a protein of 167 aa with 83% identity (**Fig. 37A**).

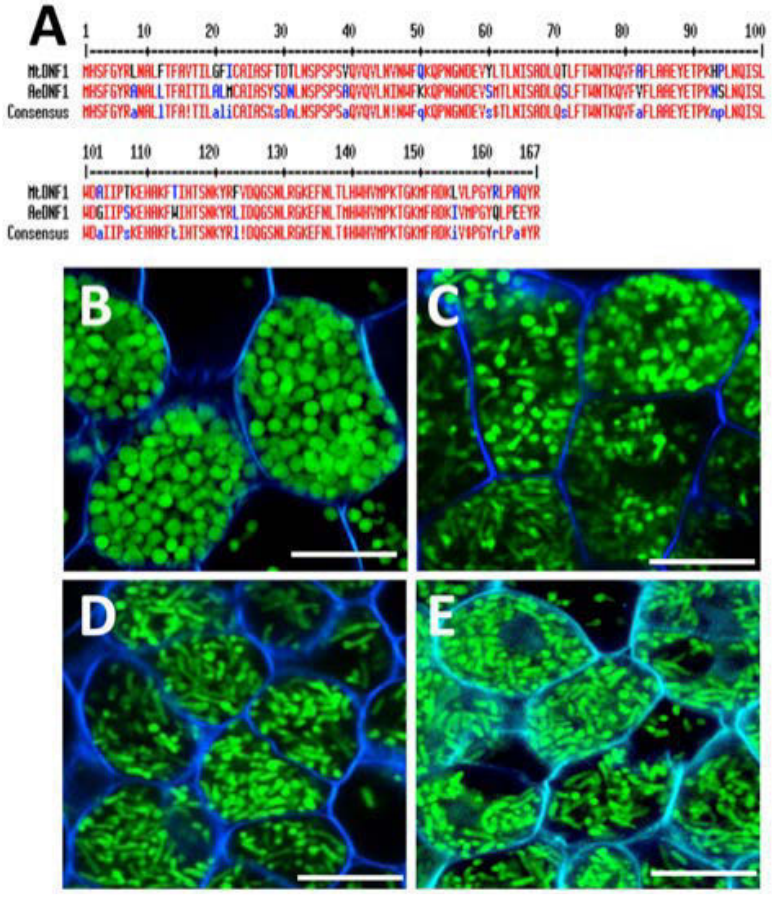

Figure 37. *DNF1* silencing affects bacteroids differentiation in *Aeschynomene*

(**A**) Alignment of the deduced protein sequence of the *Aeschynomene evenia* ssp. *serrulata DNF1* ortholog with the corresponding *Medicago* protein (accession number: TC121074). The 2 protein sequences were aligned using Multalin (http://multalin.toulouse.inra.fr/multalin/). **(B-E)** Roots of different lines of *A. evenia* ssp. *serrulata* transformed with a RNAi construct directed against *AeDNF1* encoding the SPC22 subunit of the SPC were inoculated with a GFP-tagged *Bradyrhizobium* ORS285 strain. At 21 dpi, the nodules were harvested, and the shape of the bacteroids was observed by confocal microscopy on nodule sections. As a control, a line with no RNAi effect (1% reduction) was used **(B).** The line 3, 7 and 53 with extinction levels of 62, 65 and 66 % respectively are presented in panel **(C-E)**. Bars represent 10 µm.

We used an RNAi approach to knock down the *AeDNF1*. The morphology of the bacteroids was observed in mature nodules elicited by the GFP-tagged ORS285 strain. In parallel, we monitored the level of *AeDNF1* extinction by RT-qPCR. As shown in **Fig. 37B**, in transgenic lines in which the extinction level of *AeDNF1* was not affected, the bacteroid differentiation occurred normally giving rise to spherical bacteroids, while in the lines with the highest extinction level (about 65 % reduction of *AeDNF1* expression) defects in bacteroid differentiation were observed with the presence of various bacteroid morphotypes within the same cell (**Fig. 37C**) or elongated bacteroids instead of spherical (**Fig. 37D and E**).

4. Discussion

The use of cysteine rich peptides called NCRs, to govern bacteroid differentiation was assumed to be specific to legume species belonging to the IRLC clade, no homologs of NCRs being found in legume species outside of this clade (Mergaert *et al*., 2003). In this study, we provide several independent pieces of evidence indicating that *Aeschynomene* species use a similar molecular mechanism to impose a differentiation process on their rhizobium endosymbionts.

First, we demonstrated that the distinguishing characters associated with bacteroid differentiation in nodules of IRLC legumes, *i.e*. marked cell enlargement, polyploidy, membrane permeability modification, and loss of viability, are conserved in *Aeschynomene* nodules. Second, we identified in several *Aeschynomene* species a class of cysteine rich peptides which have similar characteristics as the *Medicago* NCR-peptides and which are synthetized specifically in nodule symbiotic cells and targeted to the bacteroids. Finally, silencing of the *dnf1* homolog in *A. evenia* spp. *serrulata*, impaired bacteroid differentiation. Moreover, we show in the accompanying paper that *Bradyrhizobium* strains ORS278 and ORS285 require the BclA peptide transporter for both E- and S-morphotype bacteroid differentiation in *Aeschynomene* (Guefrachi *et al*., 2015a). BclA is similar to *S. meliloti* BacA which is needed for bacteroid differentiation in *Medicago* and which imports NCR peptides into the bacterial cells. BclA can functionally replace BacA for peptide transport. Together, these observations strongly suggest that in *Aeschynomene*, E- or S-morphotype bacteroid differentiation is under the control of the NCR peptide family, which we identified here.
These findings might be surprising, considering that according the *Aeschynomene* species the bacteroids display two distinct morphotypes (S- and E-morphotypes). This difference of morphological shape suggests distinct mechanisms governing bacterial morphogenesis.

The change from a rod to a spherical shape as observed in *A. indica* and *A. evenia* during bacteroid differentiation resemble bacterial morphologies resulting from mutations, which affect peptidoglycan synthesis and which were described in *e.g. E. coli* (Young, 2003). Similarly, it has been shown in *Vibrio* that some non-canonical D-amino acids could led to a rod to sphere shape transition both via their incorporation into the peptidoglycan polymer and by regulating enzymes that synthesize and modify it (Lam *et al.*, 2009). Therefore, the bacteroid morphology in *A. indica* or *A. evenia* is possibly associated with changes in the peptidoglycan cell wall architecture. This raises the question whether the morphogenesis to the S-morphotype is governed by a specific class of NCR peptides targeting enzymes or genes involved in peptidoglycan synthesis or if additional plant effectors of a different nature such as particular D-amino acids are involved in the control of bacteroid morphogenesis.

Even if the underlying mechanisms for E- and S-morphotype bacteroid formation are obviously different, some observations strongly suggest that the two processes are closely related. In *A. indica* and *A. evenia*, the formation of the spherical bacteroids is taking place synchronously in all nodule cells between 4 and 5 dpi (Bonaldi *et al.*, 2011). In this one-day time frame, all intracellular rod-shaped bacteria transform into spheres. However, when making observations between 4 and 5 dpi, one can discern intermediate differentiation stages, which strongly resemble E-morphotype bacteroids (**Fig. 38**).

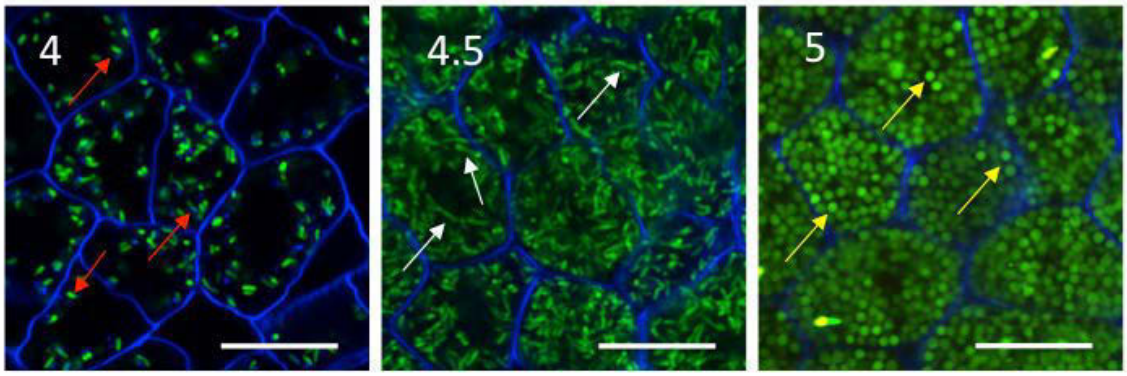

Figure 38. Change in the bacterial cell shape during the symbiotic interaction between *Bradyrhizobium* strain ORS285 and *Aeschynomene indica*

To easy bacteroids visualization, *A. indica* nodules were elicited by a *Bradyrhizobium* ORS285 strain expressing GFP. Nodule were harvested at 4, 4.5 and 5 dpi then sectioned and analysed by confocal microscopy. Red and white arrows show respectively early (U-morphotype) and final (S-morphotype) shape of bacteroids while yellow arrows point the transitory elongated shape of bacteroids. Bars represent 10 µm.

In addition, we observed in the *A. evenia* spp. *serrulata* knock-down lines in which the level of expression of *AeDNF1* was reduced, bacteroids with an E-morphotype instead of an S-morphotype or a mixture of both shapes within the same cell. Finally, the bacterial gene

bclA is similarly required for E-morphotype bacteroid formation in *A. afraspera* or S-morphotype bacteroid formation in *A. indica* or *A. evenia* (Guefrachi *et al*., 2015a).

Interestingly, analysis of the distribution of the E- and S-morphotype bacteroids among *Aeschynomene* revealed that all the tested species using a Nod-independent symbiotic process have S-morphotype bacteroids whereas those using a Nod-dependent process have E-morphotype bacteroids, similar to the IRLC species. Furthermore, a recent phylogenetic analysis revealed that all *Aeschynomene* species using a Nod-independent symbiotic process form a monophyletic clade that does not include other species using a Nod-dependent process (Chaintreuil *et al*., 2013). It could therefore be tempting to speculate that the S-morphotype and Nod-independent symbiosis are both derived characters, which are correlated. However, in the Dalbergoid clade that contains the *Aeschynomene* genus, other species are present, such as *Arachis hypogaea* or *Stylosanthes hanata*, that also host spherical bacteroids (Chandler *et al*., 1982; Sen *et al*., 1986). It has been established that at least *A. hypogaea* uses a Nod-dependent symbiotic process (Ibáñez and Fabra, 2011). This suggests that the S-character has possibly emerged several times during the diversification of the Dalbergoid clade independent of the Nod-independent character.

The mode of action of the NCRs and their bacterial targets remains unclear in *Medicago*. However, it has been shown that some cationic NCRs, particularly those with a high isoelectric point, are able, *in vitro*, to provoke symptoms of terminal differentiation (loss of viability, bacteria elongation, membrane permeabilization and genome endoreplication) (Van de Velde *et al*., 2010). In *Medicago*, the proportion of anionic and cationic NCRs is similar but no *in vitro* activity has been reported to date for the anionic or neutral NCRs. This does not imply that this last class of NCRs is not functional *in vivo* considering their high level of expression, which is similar to the cationic NCRs (Mergaert *et al*., 2003). It was proposed that the cationic NCRs interact preferentially with the negatively charged outer membrane of rhizobia and by compromising its permeability facilitate the entry of anionic NCRs to the bacterial cytosol (Kondorosi *et al*., 2013). In *Aeschynomene,* except for very few exceptions, no cationic *NCR-like* peptides could be identified among the several tens of candidate genes identified in each species (**Table 4; Annex 2**). We have tested *in vitro* the effect of four synthetic *Aeschynomene* NCRs, corresponding to the most abundant ones identified in the *A. afraspera* and *A. indica* nodule EST libraries but no activity could be detected on *Bradyrhizobium* strain ORS285 or *S. meliloti*. We also tested on the strain ORS285 the *Medicago* cationic NCRs (NCR035, NCR247 and NCR335), known to have an antimicrobial effect on a large panel of bacteria (Tiricz *et al*., 2013) but surprisingly no

activity was detected also for those peptides whereas on the control *Sinorhizobium* strain a strong antimicrobial activity and strong membrane damage could be observed in our conditions. The membrane of *Bradyrhizobium* strains differs from rhizobium strains by the presence of hopanoids, which may represent up to 40% of the total lipid extract (Kannenberg *et al*., 1995). It has recently been shown that this class of compounds, which display some structural similarity with eukaryotic sterols, reinforce the stability and rigidity of the outer membrane of photosynthetic *Bradyrhizobium* (Silipo *et al*., 2014). It is therefore possible that hopanoids reduce the antimicrobial activity of cationic *Medicago* NCRs by reinforcing the outer membrane barrier function. In agreement, hopanoid mutants of *Burkholderia* strains displayed increase sensitivity to the antimicrobial peptide polymyxin B (Schmerk *et al*., 2011; Malott *et al*., 2012). How the *Aeschynomene* NCRs can enter in *Bradyrhizobium* cells and modulate their physiology and morphology remains an open question. It is possible that these NCRs only act synergistically or are active only on a particular physiological state of the bacteria or that bacteria use specific transporters such as BclA (Guefrachi *et al*., 2015a), to import them or use specific membrane receptors to detect them and activate appropriate signaling pathways.

In silico analyses of the available transcriptome databases suggest the *NCR-like* genes constitute a large family in *Aeschynomene* species counting several tens or even hundreds of candidates. As observed for *Medicago*, the amino acids sequence of these NCR peptides is highly divergent except the distribution of 6 Cys residues, which is conserved in 93% of the identified peptides. A phylogenetic analysis based on the sequence of the mature peptides shows that the NCR peptides identified in *Aeschynomene* are distinct from the NCRs identified in IRLC (**Fig. 39**). A recent study mapping the bacteroid morphology in 40 legume species of the Papilionoideae suggests that legumes inducing bacteroid morphogenesis evolved independently at least five times from an ancestral papilionoid legume hosting U-morphotype bacteroids (Oono *et al*., 2010). It has been proposed that the NCR family identified in IRLC legumes evolved from defensin ancestor(s) (Mergaert *et al*., 2003). This large family of antimicrobial peptides is found in all eukaryotes where they are part of the first line of defense against invading microbes. The fact that the NCR peptides identified in *Aeschynomene* are phylogenetically distant from the NCRs identified in the IRLC legumes but also display a defensin signature suggest that they have similarly evolved from defensin(s) but likely from a distinct ancestral repertoire.

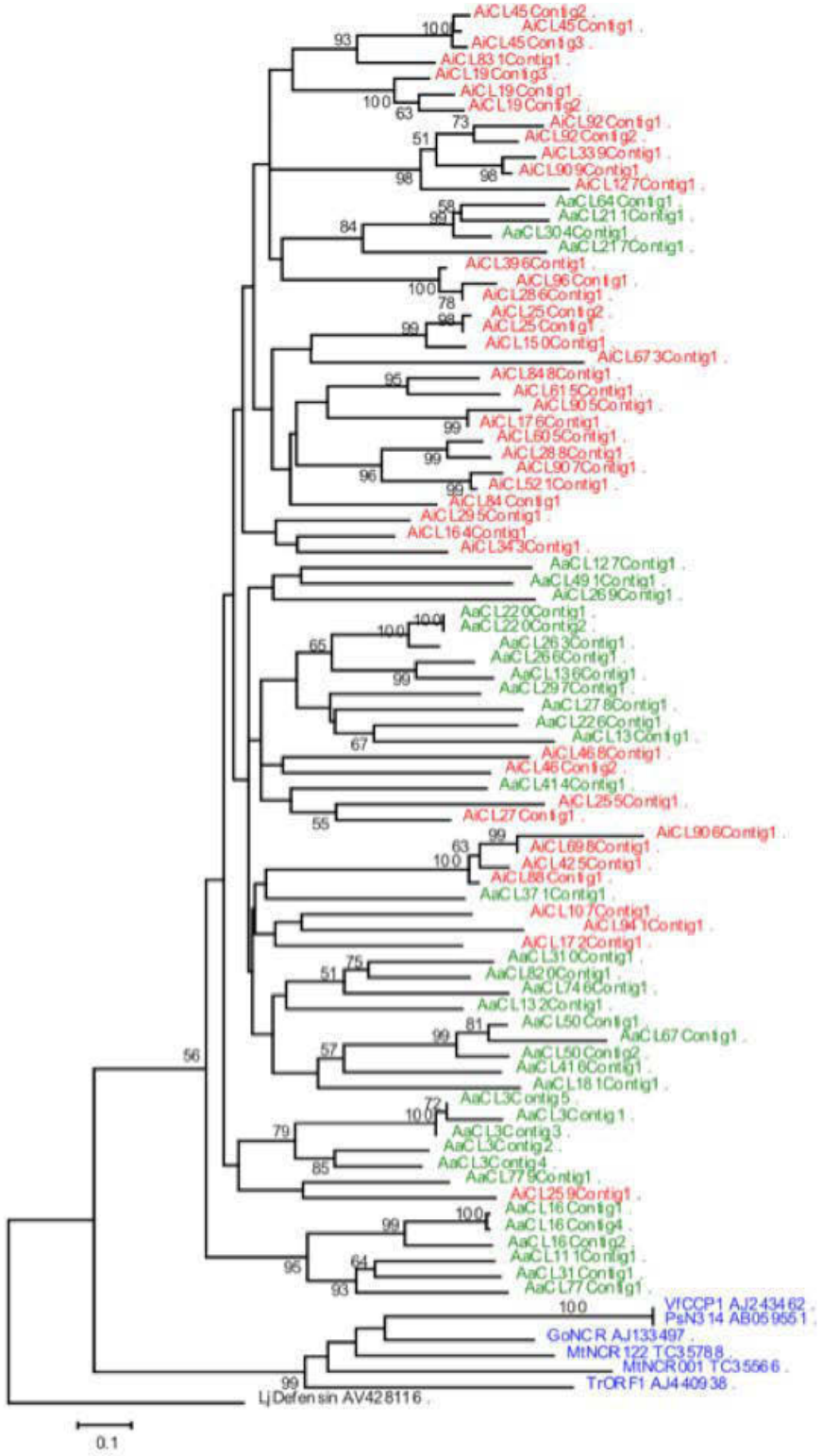

Figure 39. Neighbor joining phylogeny of NCR protein sequences

The amino acid sequences of *NCR* genes were aligned using Muscle3.6. The evolutionary distances were computed using the Poisson correction method, and a tree was inferred from the distance matrix by Neighbor Joining using MEGA5. All ambiguous positions were removed for each sequence pair. There were a total of 210 positions in the final dataset. Percent of bootstrap replicates (1000 replicates) are indicated at each tree node (only if >50%). NCRs from *A. afraspera* and *A. indica* are in green and red respectively and those from other IRLC legumes are in blue (The accession numbers are indicated). A defensin from Lotus, a non-IRCL legume was used for tree rooting.

It will be interesting to analyze the transcriptome of nodules from species belonging to the two other lineages, the Genistoids and the Mirbelioids, in which bacteroid morphogenesis was reported, using the selective criteria defined in this study, to determine if the use of cysteine rich peptides is a general strategy in legumes to keep their endosymbionts under control. Remarkably, it has been shown recently that also insects produce antimicrobial peptides to control the development of their endosymbiotic bacteria which display in certain cases a spherical or elongated shape similar to the S- or E-morphotype bacteroids (Login *et al*., 2011; Shigenobu and Stern 2012). This suggests that the enrolment of antimicrobial peptides as symbiosis effectors is an optimal *modus operandi* for a eukaryotic host cell to control large population of intracellular bacteria.

Chapter. III

Bradyrhizobium BclA is a NCR peptide transporter required for bacterial differentiation in symbiosis with *Aeschynomene*

Bradyrhizobium BclA is a peptide transporter required for bacterial differentiation in symbiosis with *Aeschynomene*[*]

1. Abstract

Legumes form nodules on their roots, which harbor nitrogen fixing endosymbiotic rhizobium bacteria called bacteroids. Several legume species produce in the symbiotic nodule cells which house the bacteroids, effector molecules, called NCRs, that are related to antimicrobial peptides of innate immunity and that induce the endosymbiotic rhizobia in a differentiated, enlarged and polyploid state. Here we identified the *bclA* gene of *Bradyrhizobium* species, which is required for the establishment of a functional symbiosis with the legumes *Aeschynomene indica* and *Aeschynomene afraspera*. Bacteroids in *Aeschynomene* nodules induced by the *bclA* mutant are less enlarged and have reduced ploidy compared to the wild-type strain. This bacterial differentiation defect is associated with a strongly reduced nitrogen fixation capability of the bacteroids. The BclA ABC transporter promotes the import of the NCR peptides and provides protection against the antimicrobial activity of these peptides. Moreover, BclA can complement the role of the related BacA transporter of *Sinorhizobium meliloti*, which has a similar symbiotic function in the interaction with *Medicago* legumes.

[*] **Modified from: Guefrachi, I**., Pierre, O., Bourge, M., Timchenko, T., Alunni, B., Barrière, Q., Czernic, P., Villaécija-Aguilar, J.A., Verly, C., Fardoux, J., Mars, M., Kondorosi, E., Giraud, E., and Mergaert, P. (2015a) and from **Guefrachi, I**., Pierre, O., Crespo, J.C., Timchenko, T., Alunni, B., Mars, M., Giraud, E., Vinardell, J.M., and Mergaert, P. (2015a/b).

2. Introduction

An evolutionary strategy, frequently used by certain groups of organisms such as insects or plants to solve the lack of specific nutrients in their food, is the establishment of symbiotic interactions with micro-organisms that are able to produce those nutrients. A well-known example, studied intensively for its agronomic and ecological impact, is the interaction of legumes with nitrogen-fixing bacteria. Bio-available nitrogen limits plant growth, yet legumes have the capacity to thrive on nitrogen-poor soils thanks to the reduction of air-borne nitrogen gas to ammonia by their symbiotic bacteria. These legume symbiotic bacteria, commonly called rhizobia, are phylogenetically disparate α- and β-proteobacteria belonging respectively to the Rhizobiales and the Burkholderiales orders (Masson-Boivin *et al*., 2009). Like in other nutritional symbioses, fulfilling all the nutritional needs of the host plant by the endosymbiotic rhizobia requires the hosting of millions to billions of them. To achieve this challenging task, legumes form specific organs called nodules on their roots, and certain species also on their stems. Nodules are in large part composed of specialized symbiotic cells, which are giant polyploid cells, each carrying several hundreds to thousands of intracellular rhizobia (Kondorosi *et al*., 2013). This high bacterial load requires the establishment of homeostasis between the two symbionts whereby the host immune response does not eliminate the micro-symbionts and inversely, the massive colonization of the host cells does not result in collapse of those cells. A recently unveiled pathway in the study-model *Medicago truncatula* that prevents the activation of the host-defense reactions by the presence of the rhizobia in the symbiotic cells implies the receptor-like kinase SymCRK (Berrabah *et al*., 2014) and the phosphatidylinositol phospholipase C-like protein DNF2 (Bourcy *et al*., 2013). On the other hand, to maintain the endosymbiotic rhizobial population under control, some legumes, enforce the intracellular bacteria into a terminal differentiation process, which is characterized by cell enlargement, polyploidy and loss of reproductive capacity. Bacteroids in this type of legumes are elongated, called E-morphotype bacteroids or spherical, called S-morphotype bacteroids. Such a mechanism has been described in detail in IRLC and Dalbergoid legumes (Mergaert *et al*., 2006; Czernic *et al*., 2015) and exists also in other legume clades (Oono *et al*., 2010). However, in still other legumes this terminal differentiation does not exist and such legumes carry in their nodules unmodified or so-called U-morphotype bacteroids (Mergaert et al., 2006; Oono *et al*., 2010; Kondorosi *et al*., 2013). The ancestral state of the bacteroids is considered to be the undifferentiated state and the other morphotypes have appeared later in evolution (Oono *et al*., 2010). Since IRLC and

Dalbergoids are distantly related legume clades (Wojciechowski *et al*., 2004), it is inferred that the molecular mechanisms that enforce bacteroid differentiation have evolved independently in these clades (Oono *et al*., 2010).

In the *Sinorhizobium meliloti – Medicago* symbiosis, the formation of the E-morphotype bacteroids is under the control of a particular class of effectors called the Nodule-specific Cysteine-Rich (NCR) peptides which are massively produced by the symbiotic cells and targeted by the secretory pathway to the endosymbionts, inducing them into their polyploid differentiated state (Van de Velde *et al*., 2010). The polyploidy of the bacteroids implies that the peptides interfere with the bacterial cell cycle, inhibiting division while DNA replication continues. This is achieved, at least in part, by interfering with the bacterial cell cycle regulator CtrA (Pini *et al*., 2013; Penterman *et al*., 2014). NCRs are conserved in all tested E-morphotype-forming IRLC legumes but other legumes forming U-morphotype bacteroids, such as for example *Lotus japonicus*, *Phaseolus vulgaris* (bean) or *Glycine max* (soybean), lack *NCR* genes explaining why bacteroids in those legumes remain unmodified (Kondorosi *et al*., 2013).

NCR peptides are similar to antimicrobial peptides and have *in vitro* antimicrobial activity (Van de Velde *et al*., 2010; Haag *et al*., 2011; Haag *et al*., 2012; Arnold *et al*., 2013; Tiricz *et al*., 2013). The BacA transporter protein of *S. meliloti* provides protection against this antimicrobial activity and is required for the long term maintenance of *S. meliloti* in the NCR-producing symbiotic cells of *Medicago* nodules. A *bacA* mutant does not differentiate in these symbiotic cells but dies off very rapidly after its release and exposure to the NCR peptides inside the symbiotic cells due to its hypersensitivity to these peptides (Glazebrook *et al*., 1993; Haag *et al*., 2011). The *bacA* gene is not required for all *Rhizobium*-legume interactions and based on a limited number of case studies, a correlation was found between the requirement of the *bacA* gene for symbiosis, the production of NCR peptides by the host cells and bacterial differentiation (Haag *et al*., 2011).

Legumes of the *Aeschynomene* genus belong to the Dalbergoid clade and form either E- or S-morphotype bacteroids depending on the species (Bonaldi *et al*., 2011; Czernic *et al*., 2015). We have shown recently that *Aeschynomene afraspera* nodules housing E-morphotype bacteroids or *Aeschynomene indica* and *Aeschynomene evenia* nodules housing S-morphotype bacteroids produce NCR peptides. However, these *Aeschynomene* NCR peptides have no sequence homology to the NCR peptides of IRLC legumes and are therefore of an independent evolutionary origin (Czernic *et al*., 2015). These peptides, which are specifically produced by the symbiotic cells and targeted to the bacteroids are most likely involved in the

formation of the differentiated bacteroids (Czernic *et al*., 2015). Thus, despite the independent evolution of bacterial differentiation, similar mechanisms seem to operate in both legume clades. To further corroborate this notion, we identified in the present study a *bacA*-like gene in the *Bradyrhizobium* species which nodulate *Aeschynomene* plants and demonstrate its role in bacteroid differentiation in those plants while the same gene is not required for nodulation of U-morphotype forming soybean.

BacA-related proteins are transporters composed of solely a SbmA_BacA transmembrane domain (LeVier and Walker, 2001). BacA of *S. meliloti* and *Mycobacterium tuberculosis* and SbmA, a homolog in *Escherichia coli*, have the notable feature of internalizing a variety of structurally diverse peptides and other molecules, including the antibiotic bleomycin and the mammalian peptide Bac7 (Laviña *et al*., 1986; Yorgey *et al*., 1994; Salomon and Farias, 1995; Mattiuzzo *et al*., 2007; Pränting *et al*., 2008; Marlow *et al*., 2009; Wehmeier *et al*., 2010; Ghosal *et al*., 2012; Puckett *et al*., 2012; Arnold *et al*., 2013; Gopinath *et al*., 2013; Narayanan *et al*., 2014). The uptake of these antimicrobial compounds by the SbmA/BacA 'multidrug' transporter alters the sensitivity of the bacteria towards these molecules. Here we find that the *Bradyrhizobium* BacA-related protein increases sensitivity towards Bleomycin and Bac7 suggesting that they are transporters of these molecules. Moreover we show that *S. meliloti* BacA and its homolog in *Bradyrhizobium* are capable of importing NCR peptides.

3. Results

3.1. Identification of BacA homologues *in Bradyrhizobium* species

The BacA protein of *S. meliloti* belongs to a subfamily of ABC transporters characterized by the transmembrane domain pfam06472 (ABC_membrane_2) or pfam05992 (SbmA_BacA) (Punta *et al*., 2012; Marchler-Bauer *et al*., 2013). We call these ABC transporters the SbmA_BacA domain transporters. Besides BacA, *S. meliloti* has a second SbmA_BacA domain ABC transporter called ExsE (LeVier and Walker, 2001). The *exsE* gene is located next to a gene cluster important for symbiosis and involved in queuosine and in exopolysaccharide biosynthesis (the *queCDE* and *exo-exs* clusters) (Becker *et al*., 1995; Marchetti *et al*., 2013). Therefore, we analyzed the symbiotic phenotype of two independent *S. meliloti exsE* mutants. Both formed a normal symbiosis, inducing nodules on *M. truncatula* that were normally infected, that contained symbiotic cells with elongated bacteroids and that

fixed nitrogen to the same level as wild-type nodules **(Fig. 40**). Thus *exsE* does not seem to have a crucial role in symbiosis.

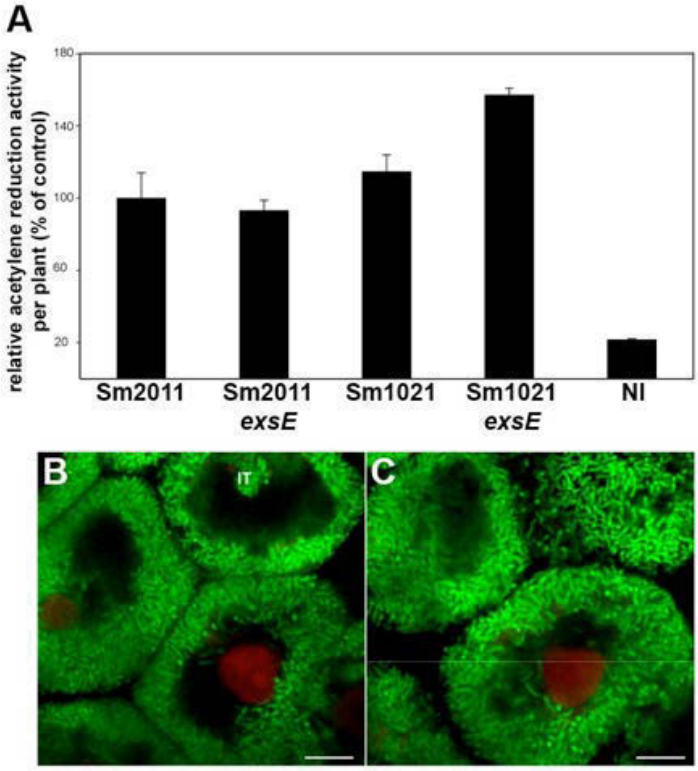

Figure 40. The *exsE* gene of *S. meliloti* is not required for symbiosis

(A) *M. truncatula* plants were inoculated with *S. meliloti* wild-type strains Sm2011 and Sm1021 and their *exsE* mutant derivatives or not inoculated (NI). The nitrogen fixation capacity per plant at 14 dpi was determined by the acetylene reduction assay and expressed as the relative activity to the Sm2011 wild-type infected plants. Error bars are standard deviations (n=10). **(B,C)** Symbiotic cells of a nodule infected with the Sm1021 *exsE* mutant observed by confocal microscopy after Live/Dead staining of a nodule section. IT indicates an infection thread. **(B)** Symbiotic cells in the infection zone where bacteroids are not yet fully differentiated. **(C)** Symbiotic cells in the fixation zone of the nodule where the bacteroids are normally differentiated into elongated bacterial cells. The green staining of the bacteroids indicates their viability. Scale bars are 10 µm.

The *bacA* gene of *S. meliloti* is conserved in the genomes of *Sinorhizobium*, *Rhizobium* or *Mesorhizobium* strains in which these genes are located in synthenic regions and display similarities of 57 to 100%. Also *exsE* genes are conserved in different rhizobium strains (45 to 100%) although they are not located in synthenic regions. Genes with similar high homology levels to *bacA* or *exsE* or located in a similar synthenic region are missing from the genomes of the *Bradyrhizobium* strains ORS278, ORS285 and USDA110. The former two strains are natural symbionts of *Aeschynomene* species (Giraud *et al*., 2007) while the latter is a soybean symbiont that can also nodulate *A. afraspera* (Renier *et al*., 2011). Nevertheless, we identified by BLASTP searches in each of these 3 *Bradyrhizobium* genomes 3 different genes coding for SbmA_BacA domain ABC transporters. Each of the 3

Bradyrhizobium genes is strongly conserved among the 3 tested strains (**Table 5; Annex 3**) and is located in synthenic regions in the *Bradyrhizobium* genomes (**Fig. 41**). These proteins display a relatively weak homology ranging from 21 to 31% identity and 38 to 51% similarity to BacA and ExsE (**Table 5 and 6; Annex 3, 4 respectively**).

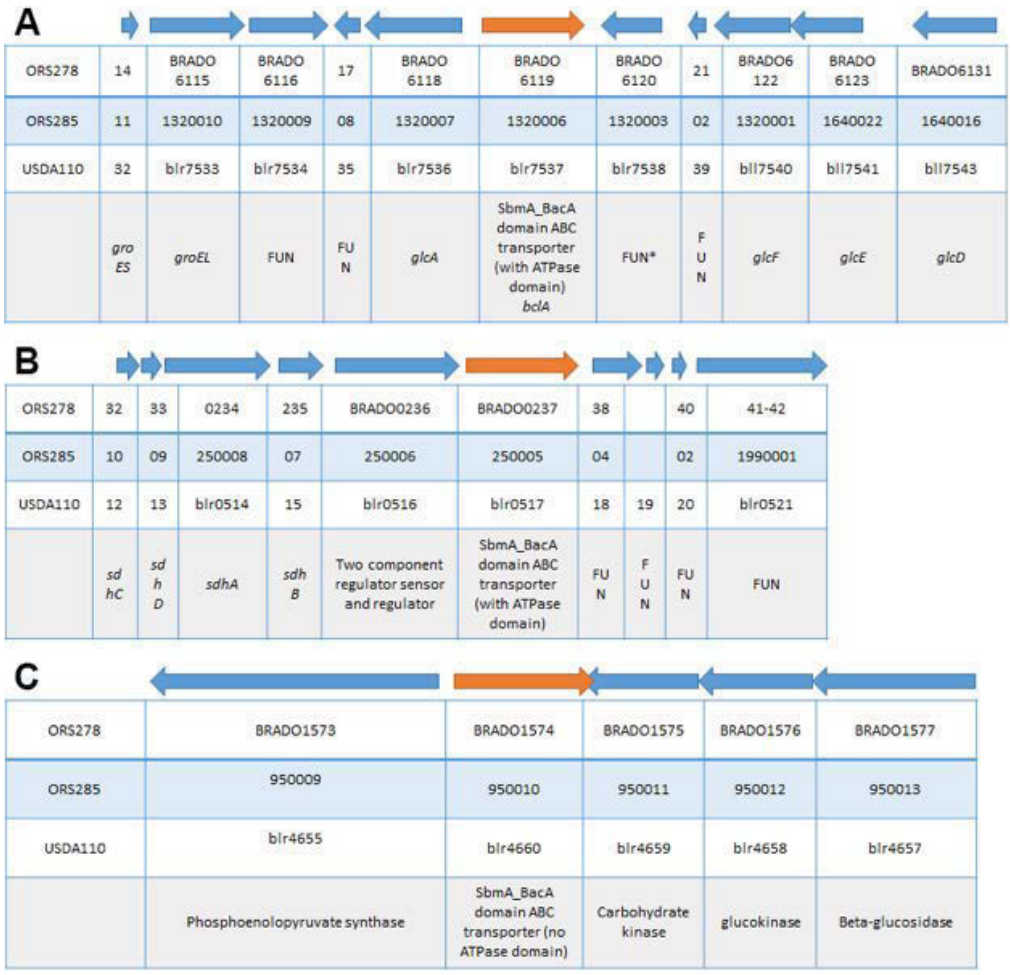

Figure 41. Three loci in *Bradyrhizobium* genomes carrying SbmA_BacA domain ABC transporters

(A) The locus carrying the symbiotic gene bclA. **(B,C)** Loci encoding SbmA_BacA domain proteins without symbiotic function. The SbmA_BacA domain proteins are indicated with red arrows and the flanking genes with blue arrows. The accession numbers of the genes are indicated for the genomes of the strains ORS278, ORS285 and USDA110 and below is the annotation of the corresponding genes. FUN is function unknown; FUN* is a secreted protein with function unknown.

Within this family of ABC transporters, the BacA protein of *S. meliloti* and its homologues in *Sinorhizobium*, *Mesorhizobium* and *Rhizobium* species and the *E. coli* SbmA protein are 8-transmembrane domain proteins but they lack the typical ATPase domain of ABC transporters. In contrast, ExsE (LeVier and Walker, 2001) and the *Mycobacterium* BacA protein encoded by gene Rv1819c (Domenech *et al.*, 2009; Arnold *et al.*, 2013) have in addition to the transmembrane domains a fused ATPase domain. The here identified *Bradyrhizobium* proteins encoded by the BRADO6119, 13200006, blr7537, BRADO0237,

250005 and blr0517 genes have a fused ATPase domain while the proteins encoded by the BRADO1574, 950010 and blr4660 genes are lacking such a domain (**Fig. 41**).

A phylogenetic tree of SbmA_BacA domain transporters (**Fig. 42**; **Table 6; Annex 4**) clusters the *Bradyrhizobium* proteins in 3 distinct subgroups but none of these subgroups is more closely related to the *S. meliloti* BacA clade than the others. The clade containing the BRADO6119, 1320006 and blr7537-encoded proteins is closer to the *Mycobacterium* BacA clade carrying a functional ortholog of BacA (Arnold *et al.*, 2013) while the other *Bradyrhizobium* proteins encoded by BRADO0237, 250005, blr0517, BRADO1574, 950010 and blr4660 are closer to the non-symbiotic ExsE clade (**Fig. 42**). Thus based on the topology of the tree, the prediction of a symbiotic SbmA_BacA domain transporter is not easy.

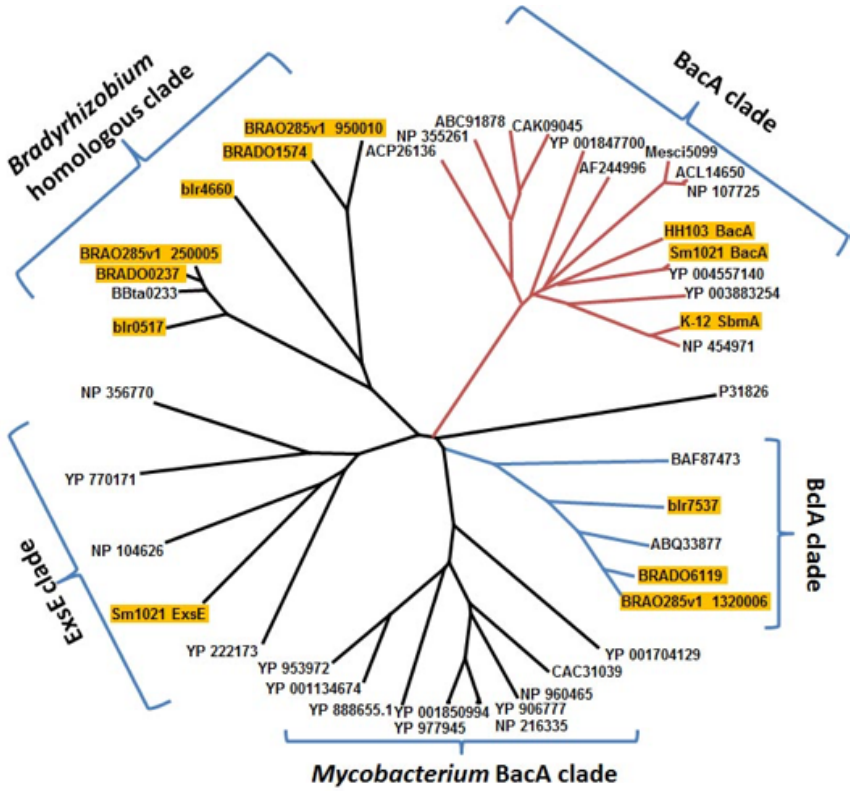

Figure 42. Phylogenetic tree of SbmA-BacA domain proteins

The tree was generated with ClustalW2. The protein sequences used for the alignment and tree generation are provided in **Table 5 (Annex 3)**. The proteins discussed in this study are highlighted in orange.

3.2. Symbiotic role of the *Bradyrhizobium* SbmA_BacA domain transporters

To determine whether the identified BacA homologs of *Bradyrhizobium* have a symbiotic role and more particularly in the differentiation of E- or S-morphotype bacteroids, we constructed mutants in the 3 genes of *Bradyrhizobium* strain ORS285 (BRAO285v1_1320006, BRAO285v1_250005 and BRAO285v1_950010). Strain ORS285

was chosen for mutagenesis rather than ORS278 because the latter strain can nodulate only S-morphotype producing *Aeschynomene* species while the former strain has a broad host range including species forming S- or E-morphotype bacteroids (Czernic *et al.*, 2015). The 3 mutants were tested for their symbiotic phenotype on the host plants *A. indica* and *A. afraspera*. Only the mutant in the gene 1320006 had an abnormal symbiotic phenotype, forming on both host plants nodules which looked aberrant in size, form and color, that had a strongly reduced nitrogen fixation activity (about 5% and 20% of wild-type levels in *A. indica and A afraspera* respectively) and that could not fully support plant growth in nitrogen poor growth conditions, obvious from the reduced plant shoot mass (**Fig. 43**).

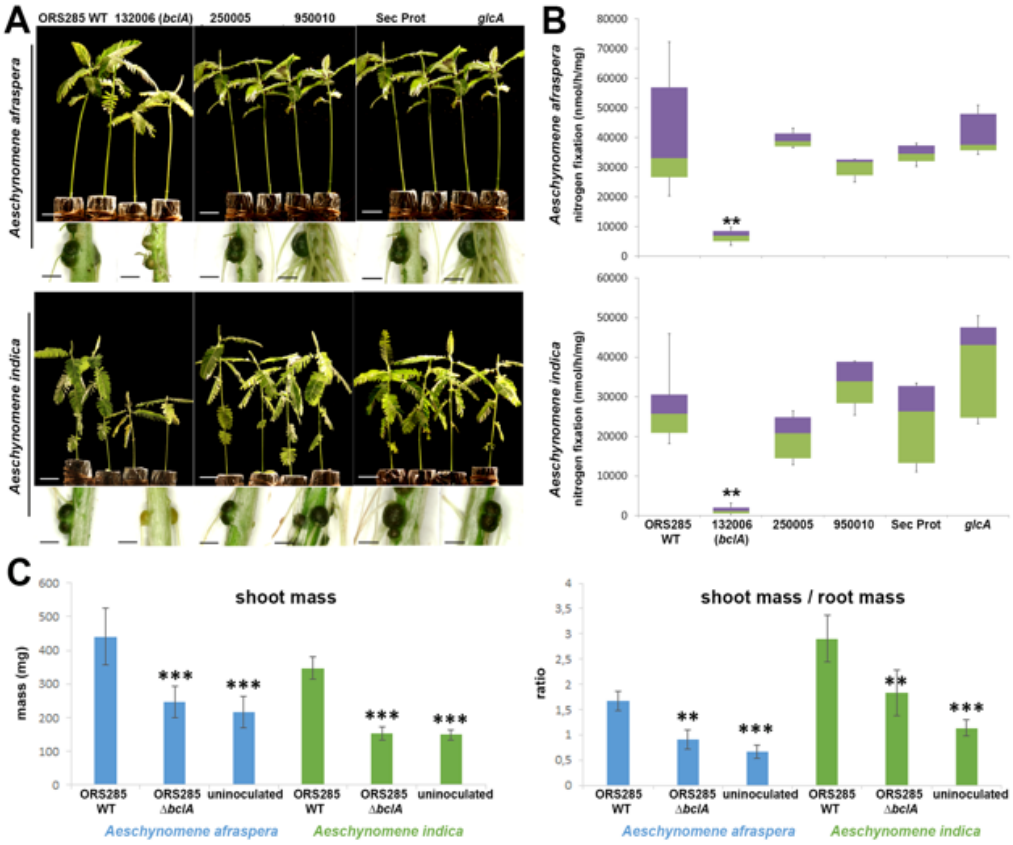

Figure 43. Symbiotic phenotype of *Bradyrhizobium* strain ORS285 mutants

(A) Plant growth and nodule phenotype at 14 dpi of *A. afraspera* (top panel) and *A. indica* (bottom panel) plants inoculated with ORS285 wild-type and its mutant derivatives in the 1320006 (*bclA*), 250005, 950010, 1320003 (Sec Prot) and 1320007 (*glcA*) genes. Scale bars are 2 cm and 2.5 mm respectively. **(B)** The nitrogen fixation activity at 14 dpi of *A. afraspera* (top panel) and *A. indica* (bottom panel) plants inoculated with the indicated strains, measured by the acetylene reduction assay per mg of fresh nodule weight (nmol ethylene produced per hour incubation and per mg nodule weight). Box-plots represent in the rectangle the first quartile to the third quartile, divided by the median value, whiskers above and below the box show the minimum and maximum measured values. Statistical analysis was made with a Mann-Whitney U-test (n=5; **: p<0.001). **(C)** Fresh weight shoot mass (left panel) or ratio of the shoot to root fresh weight (right panel) of *A. afraspera* and *A. indica* plants at 14 dpi inoculated with ORS285, ORS285Δ*bclA* or uninoculated. The experiment was carried out in duplicate with 10 plants per conditions. Error bars represent standard deviation and significance was determined by the Krsukall and Wallis test (**: p<0.001; ***: p<0.0001).

The mutants in the other two genes (250005 and 950010) on the other hand, formed normal nitrogen-fixing nodules indicating that these genes are not required for symbiosis (**Fig. 43**). Interestingly, in a screen of a collection of Tn*5* transposon mutants of *Bradyrhizobium* strain ORS278, we previously identified a mutant in the gene BRADO6119, orthologous to the ORS285 gene 1320006, which formed non-functional nodules (Bonaldi *et al*., 2010). These observations together with the further characterization described below of the gene 1320006 and its orthologues in strains ORS278 and USDA110, indicate that they encode BacA-like proteins. Therefore, we named them *bclA* for *Bradyrhizobium bacA-like*.

3.3. BclA is not required for nitrogen fixation *per se* or for bacterial growth

Contrary to most other known rhizobium strains, the *Bradyrhizobium* symbionts of *Aeschynomene* species such as strains ORS285 and ORS278 fix nitrogen during free-living growth. This capacity makes it possible to test whether the BclA transporter is involved in the nitrogen fixation process itself. Such a test was not possible for the *S. meliloti bacA* gene because that strain cannot fix nitrogen during free growth. The ORS285 and ORS278 *bclA* mutants were as active as the wild-type strain in the acetylene reduction assay (**Fig. 44A**). This indicates that BclA is not required for the nitrogen fixation metabolism.

It was recently shown that the *bacA* homologous gene of *Mycobacterium tuberculosis* is involved in vitamin B12 uptake (Gopinath *et al*., 2013). We found that both the wild-type strain and the *bclA* mutant were grew and fixed nitrogen normally in minimal medium BNM-B to which vitamin B12 was added or not (**Fig. 44A-D**). Thus *Bradyrhizobium* ORS285 does not need external vitamin B12 for growth, most likely because it has the genetic repertoire for vitamin B12 synthesis (Rodionov *et al*., 2003). Therefore, the symbiotic phenotype of the *bclA* mutant is unlikely to result from a defect in vitamin B12 uptake. These results also suggest that the symbiotic defect of the *bclA* mutant does not result from a general growth defect caused by the mutation.

The *bclA* gene of *Bradyrhizobium* strains ORS285 and ORS278 and of *Bradyrhizobium japonicum* is flanked by genes of the glycolate metabolism pathway (**Fig. 41A**). In *E. coli* the *glcCglcDEFGBA* locus confers the ability to utilize glycolate as carbon source (Pellicer *et al*., 1996; Nunez *et al*., 2001). However, *Bradyrhizobium* ORS285 is unable to grow on glycolate as carbon source (**Fig. 44E**). In agreement with this, the genes *glcB* and *glcG* are absent in the *Bradyrhizobium* locus and are replaced by *bclA* and two other genes, one of them encoding a secreted protein (gene 1320003 in strain ORS285) (**Fig. 41A**).

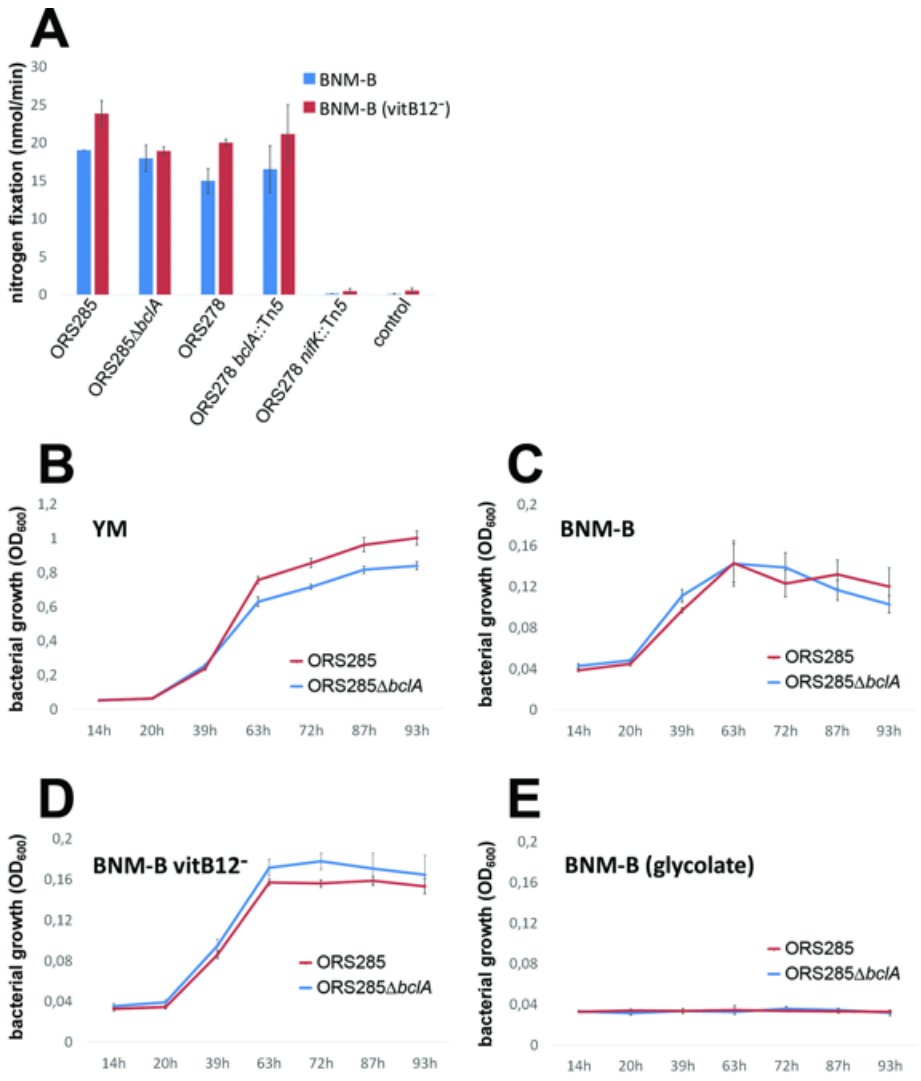

Figure 44. The mutation in the *bclA* gene of *Bradyrhizobium* strain ORS285 is not affecting free-living growth and nitrogen fixation

(A) Nitrogen fixation in nitrogen-free BNM medium (BNM-B) or nitrogen-free BNM to which no vitamin B12 was added (BNM-B, vitB12-) by the indicated strains. The *nifK* gene is an essential gene for the activation of the nitrogenase genes. A *nifK* mutant strain (ORS278*nifK*::Tn*5*) was included as a negative control for nitrogen fixation. The "control" sample is a non-inoculated flask. **(B-E)** Growth curve of ORS285 and ORS285Δ*bclA* in rich YM medium **(B)**, minimal BNM-B medium **(C)**, BNM-B to which vitamin B12 was omitted (BNM-B vitB12-) **(D)** or BNM medium in which the only carbon source was glycolate (BNM-B glycolate) **(E)**. Growth was monitored by optical density measurements (OD_{600nm}) in a plate reader at the indicated time points.

To investigate whether the *glc* metabolic genes and the secreted protein are required for symbiosis along with the *bclA* gene, we constructed mutants in these genes and inoculated them on *A. afraspera* and *A. indica*. We found that both mutants formed normal nodules, which fixed nitrogen and supported plant growth in a nitrogen-free medium on both host plants (**Fig. 45**). Thus the *bclA*-flanking genes are not involved in the symbiotic interaction with *Aeschynomene* and the function of *bclA* is probably not related to glycolate metabolism.

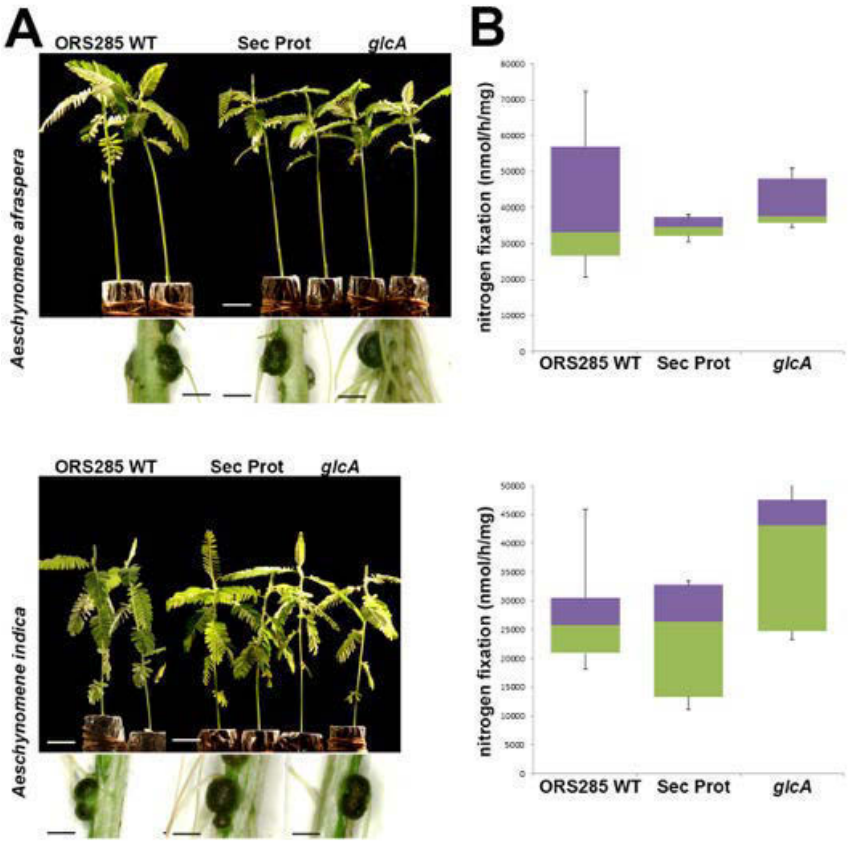

Figure 45. Glycolate metabolism and the secretory protein of the *bclA* locus are not required for symbiosis

(A) Plant growth and nodule phenotype at 14 dpi of *A. afraspera* (top panel) and *A. indica* (bottom panel) plants inoculated with ORS285 wild-type and its mutant derivatives in the 1320003 (Sec Prot) and 1320007 (*glcA*) genes. Scale bars are 2 cm and 2,5 mm respectively. **(B)** The nitrogen fixation activity of *A. afraspera* (top panel) and *A. indica* (bottom panel) plants inoculated with the indicated strains at 14 dpi, measured by the acetylene reduction assay per mg of fresh nodule weight (nmol ethylene produced per hour incubation and per mg nodule weight). Box-plots represent in the rectangle the first quartile to the third quartile, divided by the median value, whiskers above and below the box show the minimum and maximum measured values.

3.4. *Bradyrhizobium* BclA is required for polyploid bacteroid differentiation

The macroscopic phenotype of the *A. afraspera* and *A. indica* nodules infected with the ORS285Δ*bclA* mutant is compatible with a possible defect in bacteroid formation. To verify this hypothesis, we analyzed the histology of these nodules by light and confocal microscopy and compared it to wild-type infected nodules. Thin sections of resin-embedded nodules were stained with toluidine blue and observed by light microscopy (**Fig. 46**). The diameter of the nodules infected with the mutant strain were significantly smaller than the nodules infected with the wild-type strain, particularly for the *A. indica* nodules, but their central tissue was fully occupied with symbiotic cells and the surrounding cortical cell layers were normally formed (**Fig. 46A-D**). The cytosolic space of the symbiotic cells was completely filled with rhizobia similarly to symbiotic cells in wild-type nodules (**Fig. 46E-H).**

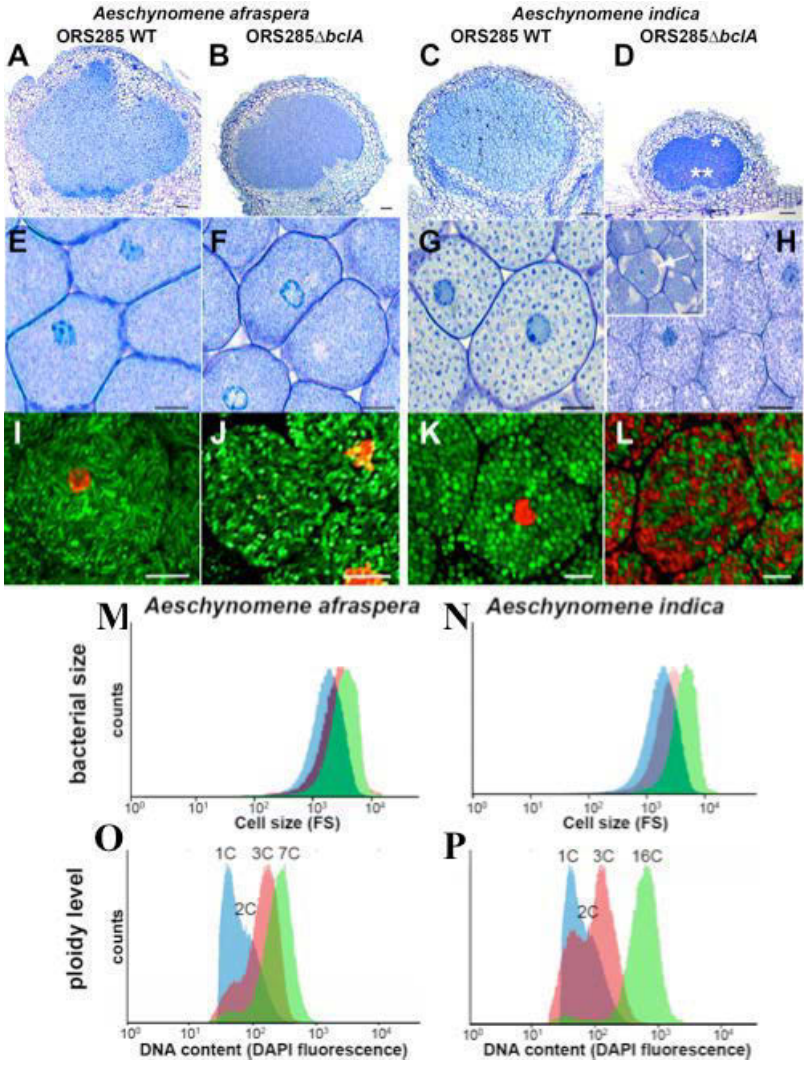

Figure 46. Small nodule size and abnormal bacteroids of ORS285Δ*bclA* in *A. afraspera* and *A. indica*

(A-H) Toluidine blue stained thin sections of 14 dpi *A. afraspera* and *A. indica* nodules induced by ORS285 or ORS285**Δ***bclA*. The location of the symbiotic cells shown in **(H)** is indicated with an asterisk in **(D)** while the inset of **(H)** shows degenerating cells located in the central zone of the nodule (**D**, double asterisk). **(I-L)** Bacteroid viability determined by live/dead staining of nodule sections and confocal microscopy in *A. afraspera* and *A. indica* nodules induced by ORS285 or ORS285**Δ***bclA*. Scale bars are 100 µm **(A-D)** or 10 µm **(E-L)**. **(M-P)** Flow cytometry analysis of bacteroid size and DNA content. **(M,N)** Forward scatter and **(O,P)** DAPI fluorescence in free living *Bradyrhizobium* ORS285 (blue) or bacteroids isolated from *A. afraspera* **(M,O)** or *A. indica* **(N,P)** nodules infected with ORS285 wild-type (green) or ORS285**Δ***bclA* (red).

However, in both *A. afraspera* and *A. indica* the morphology of the bacteroids was aberrant. In *A. afraspera* nodules infected with wild-type ORS285, the bacteroids are strongly elongated (E-morphotype) while in nodules infected with the mutant, the bacteroids remain small very similar to undifferentiated bacteria (**Fig. 46E,F**). Similarly, in *A. indica* nodules infected with wild-type ORS285, the bacteroids are spherical (S-morphotype) while in nodules infected with the mutant, the bacteroids are similar to undifferentiated bacteria (**Fig. 46G,H**).

The non-differentiated state of the bacteroids formed by the mutant strain was further verified and quantified precisely. Bacteroids were isolated from nodules and purified on a density gradient and then analyzed by flow cytometry. The forward scatter (FS) of the bacteria is relative to their size. The bacteroids isolated from *A. afraspera* and *A. indica* nodules have a higher FS than the cultured ORS285 bacteria (**Fig. 46M,N**) in agreement with their enlarged size. The *bclA* mutant bacteroids on the other hand have a FS that is intermediate between the FS of cultured bacteria and wild-type bacteroids, supporting their differentiation defect observed by microscopy. The bacteroids in *A. afraspera* and *A. indica* nodules have a high DNA content (Czernic *et al*., 2015). Therefore, we also measured the DNA content of the *bclA* mutant bacteroids, isolated from *A. afraspera* and *A. indica* nodules and stained with the DNA fluorescent stain 4',6-diamidino-2-phenylindole (DAPI). The mean DNA content of the wild-type bacteroids is 7 genome complements (7C) and 16C in *A. afraspera* and *A. indica* respectively while the mutant bacteroids reach only a 1 to 3C ploidy levels which is close to the 1C-2C ploidy of free growing *Bradyrhizobium* strain ORS285 (**Fig. 46O,P**), confirming the nearly undifferentiated state of the bacteroids formed by the *bclA* mutant observed by microscopy.

Next, we used live/dead staining (Haag *et al*., 2011) of thick nodule sections to analyze the survival ability of the *bclA* mutant bacteria within the symbiotic cells. Imaging with confocal microscopy again clearly reveals the undifferentiated state of the mutant bacteria in the nodules of both plant species (**Fig. 46I-L**). Despite this undifferentiated state, the bacteria remained viable in *A. afraspera* nodules as indicated by the green syto9 staining at least until 14 days after inoculation (**Fig. 46I,J**). In contrast, in *A. indica* nodules, the survival capacity of the mutant is more strongly affected and at 14 days post inoculation, many bacteria were dead as indicated by their red staining due to Propidium Iodide (PI) uptake (**Fig. 46K,L**). The degradation of the bacteroids in *A. indica* nodules can also be seen on the resin-embedded nodule sections (**Fig. 46D,H inset**).

To further explain the strongly reduced nitrogen fixation ability of the bacteroids in *A. afraspera* nodules, we used the acidophilic DND-99 LysoTracker probe which marks actively fixing bacteroids by red fluorescence due to acidification of the peribacteroid space (Pierre *et al*., 2013). We find that in the *A. afraspera* nodules, the *bclA* mutant bacteroids are not stained by DND-99 in many symbiotic cells (**Fig. 47**). Thus although the mutant bacteroids are not PI positive, their activity is affected by the absence of BclA function.

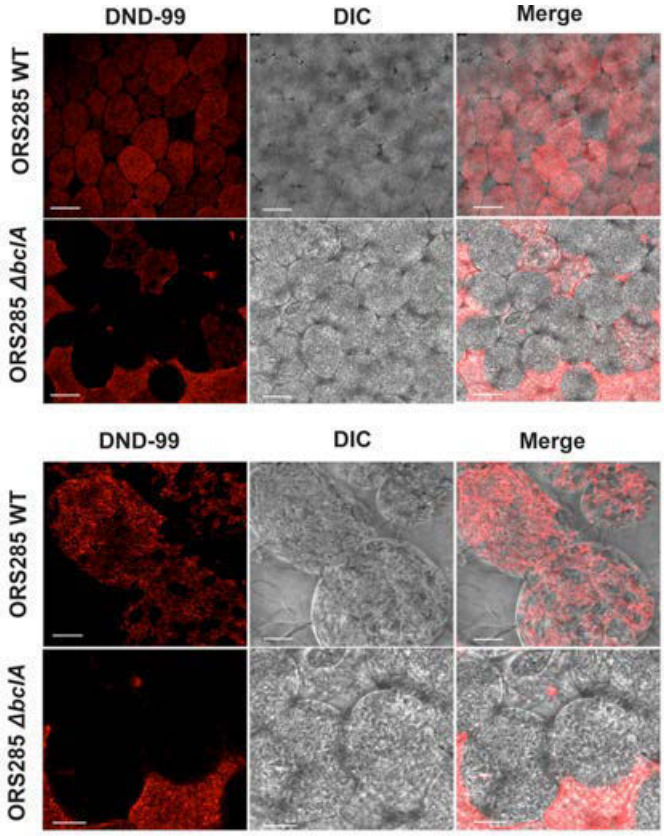

Figure 47. ORS285Δ*bclA* bacteroid functioning is impaired in *A. afraspera* nodules

Sections of 14 dpi nodules were stained with the acidophilic DND-99 red fluorescent probe which marks active bacteroids by red fluorescence due to acidification of the peribacteroid space (Pierre *et al*., 2013). DIC is the differential interference contrast image and merge is the overlay of the fluorescence and DIC images. Pictures were acquired by confocal microscopy. The scale bars are 25 µm (top panels) or 10 µm (lower panels).

To confirm that the described phenotypes are linked to the deletion of the *bclA* gene in strain ORS285, we also analyzed the phenotype of the *bclA* Tn*5* insertion mutant in strain ORS278 (Bonaldi *et al*., 2010). Nodulation of *A. evenia* and *A. indica* with this mutant cannot support the growth of these plants in the absence of external nitrogen, the nodules are abnormally formed, indicative that they are non-functional, bacteroids do not become spherical but remain similar in morphology to free-living rhizobia and these bacteroids are PI positive (**Fig. 48**).

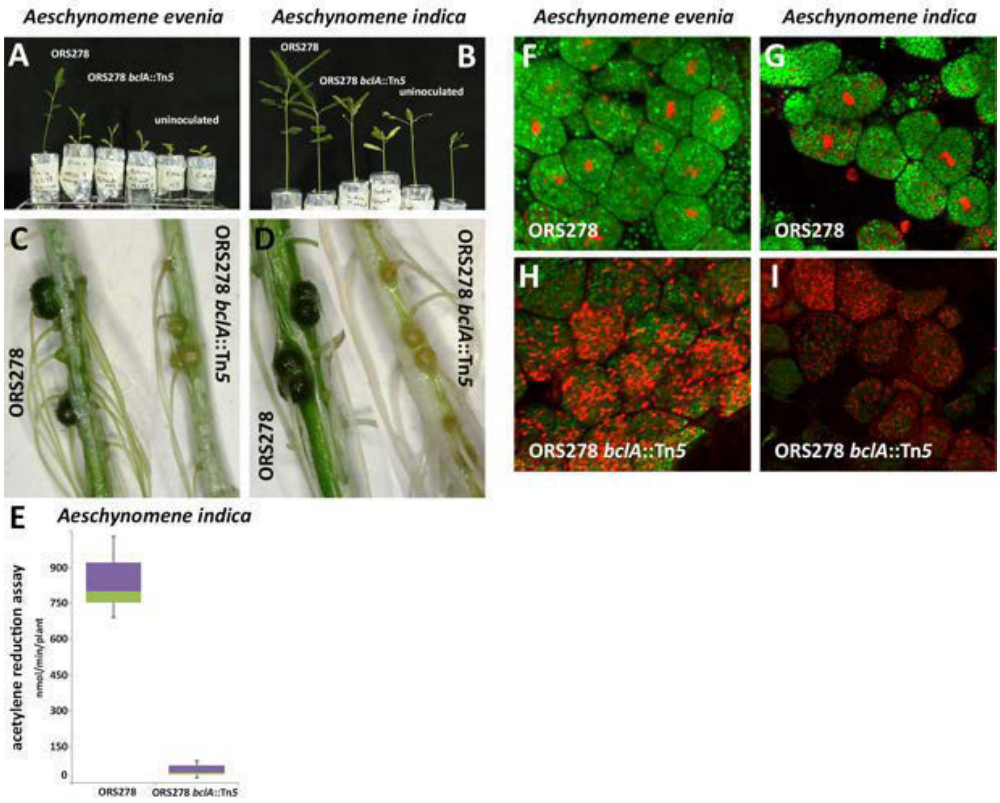

Figure 48. Symbiotic phenotype of the *bclA* mutant in *Bradyrhizobium* strain ORS278

(A) Plant growth of *A. evenia* at 14 dpi inoculated with ORS278, ORS278 *bclA*::Tn*5* or uninoculated. **(B)** Plant growth of *A. indica* inoculated with ORS278, ORS278 *bclA*::Tn*5* or uninoculated. **(C)** Nodule phenotype of *A. evenia* inoculated with ORS278 or ORS278 *bclA*::Tn*5*. **(D)** Nodule phenotype of *A. indica* inoculated with ORS278 or ORS278 *bclA*::Tn*5*. **(E)** The nitrogen fixation activity of *A. indica* plants inoculated with ORS278 or ORS278 *bclA*::Tn*5*, measured by the acetylene reduction assay per plant (nmol ethylene produced per min and per plant). Box-plots represent in the rectangle the first quartile to the third quartile, divided by the median value, whiskers above and below the box show the minimum and maximum measured values. **(F-I)** Confocal microscopy of bacteroid viability determination by live/dead staining of *A. evenia* **(F,H)** and *A. indica* **(G,I)** nodules induced by ORS278 **(F,G)** or ORS278 *bclA*::Tn*5* **(H,I)**.

We also confirmed by confocal microscopy and live/dead staining that the above mentioned mutants in the two additional *bacA*-homologous genes and in the genes encoding the secreted protein and *glcA* have a wild-type bacteroid phenotype (**Fig. 49**).

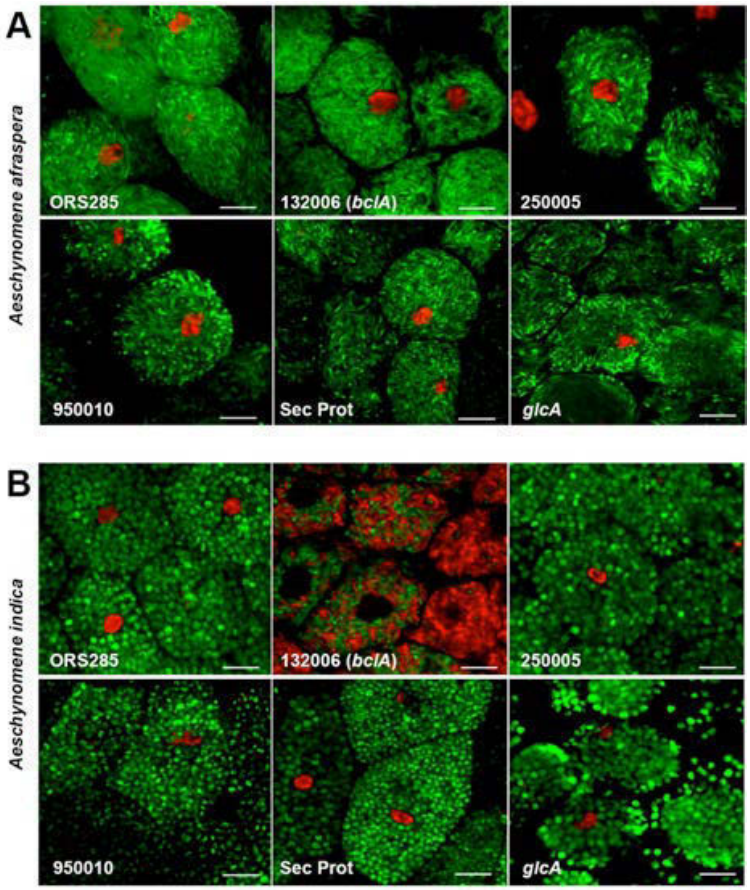

Figure 49. Bacteroids in *A. afraspera* and *A. indica* 14 dpi nodules infected with ORS285 mutants

Confocal microscopy of bacteroid viability determination by live/dead staining of *A. afraspera* **(A)** and *A. indica* **(B)** nodules induced by ORS285 wild-type and mutants in genes 1320006 (*bclA*), 250005 (SbmA-BacA domain transporter), 950010 (SbmA-BacA domain transporter), 1320003 (Sec Prot) and 1320007 (*glcA*). The scale bars are 10 µm.

3.5. *Bradyrhizobium* BclA is functionally equivalent to *S. meliloti* BacA in symbiosis

Since the phenotype of the *Bradyrhizobium bclA* mutants is reminiscent of the *S. meliloti bacA* mutant phenotype, we wanted to determine whether *bacA* and *bclA* are interchangeable for their symbiotic function. No vectors are available for *Bradyrhizobium* that can be maintained stably in the absence of antibiotic selection. Therefore, we chose to investigate the capacity of the *bclA* gene of *Bradyrhizobium* strain ORS285 to complement the symbiotic defect of a *S. meliloti bacA* deletion mutant (Ferguson *et al.*, 2002). This mutant forms nodules in which symbiotic cells are filled with non-differentiated bacteria which lose viability very rapidly due to their hyper-sensitivity to the antimicrobial action of certain NCR peptides produced by these symbiotic cells (Haag *et al.*, 2011). The phenotype of the mutant can be easily scored by the aspect of the nodules which are small and white compared to the large elongated and pink wild-type nodules (**Fig. 50A-C**).

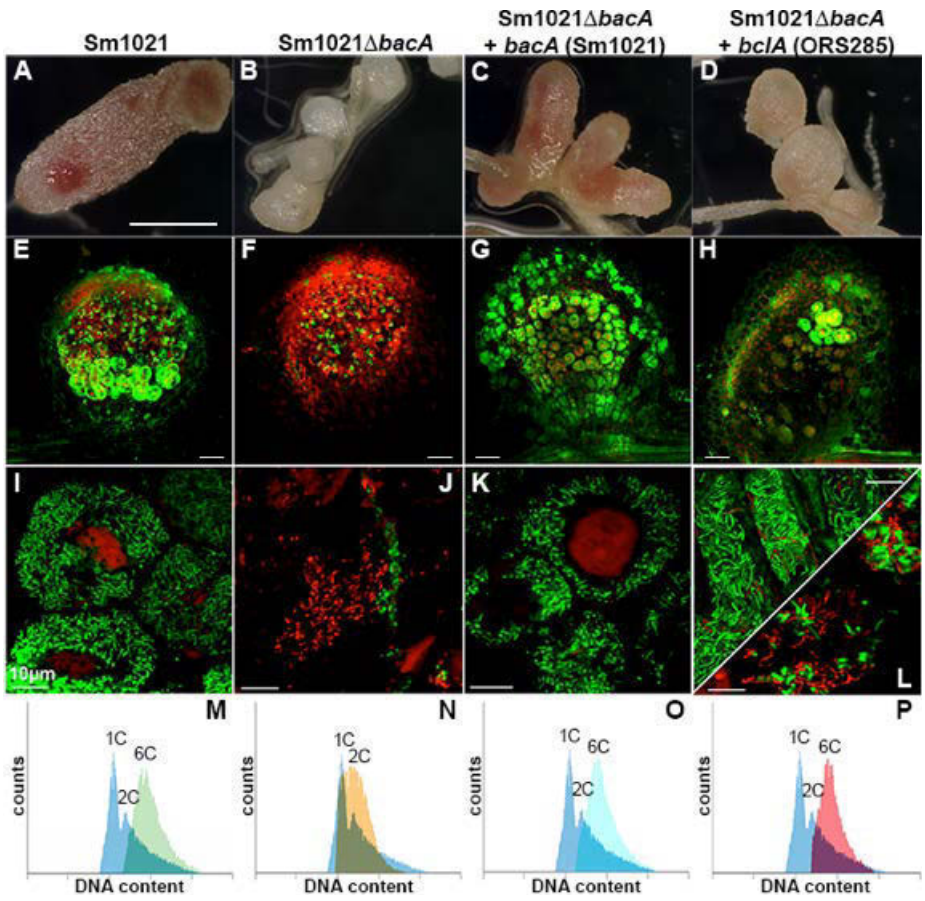

Figure 50. Complementation of the *S. meliloti bacA* mutation with *Bradyrhizobium bclA*

(A-D) Phenotype of *M. sativa* nodules at 28 dpi, infected with the indicated strains. The scale bar is 1 mm. (**E-L**) Bacteroid viability determined by live/dead staining of nodule sections and confocal microscopy in *M. sativa* nodules induced by the indicated strains. Scale bars are 100 µm **(E-H)** or 10 µm **(I-L)**. **(M-P)** Flow cytometry analysis of the DNA content by DAPI fluorescence in free living *S. meliloti* (blue in all panels) or bacteroids isolated from *M. sativa* nodules infected with the *S. meliloti* wild-type **(M**, green), the *S. meliloti bacA* mutant **(N**, orange), the *S. meliloti bacA* mutant complemented with the *S. meliloti bacA* gene **(O**, pale blue) or the *S. meliloti bacA* mutant complemented with the ORS285 *bclA* gene **(P**, red).

This pink color results from the abundant production of leghemoglobin in the cytosol of symbiotic cells during the late stages of their formation, just before onset of nitrogen fixation. Also live/dead staining of nodule sections and observation by confocal microscopy easily reveals the mutant phenotype: in nodules infected with the *bacA* mutant, the symbiotic cells contain dead bacteria, stained red by PI, and only the infection threads carry life bacteria, stained green by Syto9, while in nodules infected with the wild-type strain, symbiotic cells are entirely occupied with live bacteria which are stained green by Syto9 (**Fig. 50E-G,I-K**). The absence of differentiation of the bacteroids in the mutant can also be measured by flow cytometry analysis of bacteria extracted from nodules since the mutant bacteria do not amplify their genome contrary to the wild-type strain (**Fig. 50M-O**). The *S. meliloti bacA* mutant strain carrying the *bclA* gene of ORS285 induces the formation of nodules, which are

elongated although not to the same level as the same strain carrying the *bacA* gene and those nodules are slightly pinkish indicating the leghemoglobin production is restored (**Fig. 50C,D**). The confocal microscopy of live/dead stained nodule sections shows that the bacteroids of the *bclA* expressing strain are elongated and mostly stained green although in older symbiotic cells many dead bacteria were also detected which were nevertheless elongated (**Fig. 50H,L**). Flow cytometry analysis of the DNA content in nodule bacteria also shows that the presence of the *bclA* gene restores the polyploidization of the bacteroids to similar levels as in the strain carrying the *bacA* gene or the wild-type strain (**Fig. 50M-P**). Thus altogether, these results indicate that the *bclA* gene can complement the endogenous *bacA* gene although this complementation is only partial and insufficient to restore the nitrogen fixation capacity of the strain (data not shown).

3.6. *Bradyrhizobium japonicum* USDA110 BclA is not required for symbiosis

A. afraspera can be nodulated by the *B. japonicum* strain USDA110 which is a soybean symbiont although that strain is a less efficient nitrogen fixer than strain ORS285 (Renier *et al*., 2011). Soybean belongs to the Millettioid legume clade, which includes also *Phaseolus* or *Vigna* (Wojciechowski *et al*., 2004). All these legumes form nodules housing U-morphotype bacteroids because they lack NCR peptides (Kondorosi *et al*., 2013). For a few tested examples of legumes forming U-morphotype bacteroids, it has been found that the *bacA* gene is dispensable for symbiosis (Haag *et al*., 2011; Kondorosi *et al*., 2013). Since the *bclA* gene is highly conserved in strain USDA110 (**Fig. 42; Table 6; Annex 4**), we were wondering whether this gene is required in strain USDA110 for symbiosis with *A. afraspera* or soybean. A *bclA* deletion mutant was created and its symbiotic phenotype was compared to the parent strain on both host plants (**Fig. 51**). As expected, USDA110 formed undifferentiated, U-morphotype bacteroids on soybean as revealed by microscopy of nodule sections (**Fig. 51E**) and flow cytometry analysis of purified bacteroids (**Fig. 51H**). But more surprisingly, we found that USDA110 was within *A. afraspera* nodules only slightly elongated (**Fig. 51C**), contrary to strain ORS285 on the same host plant.

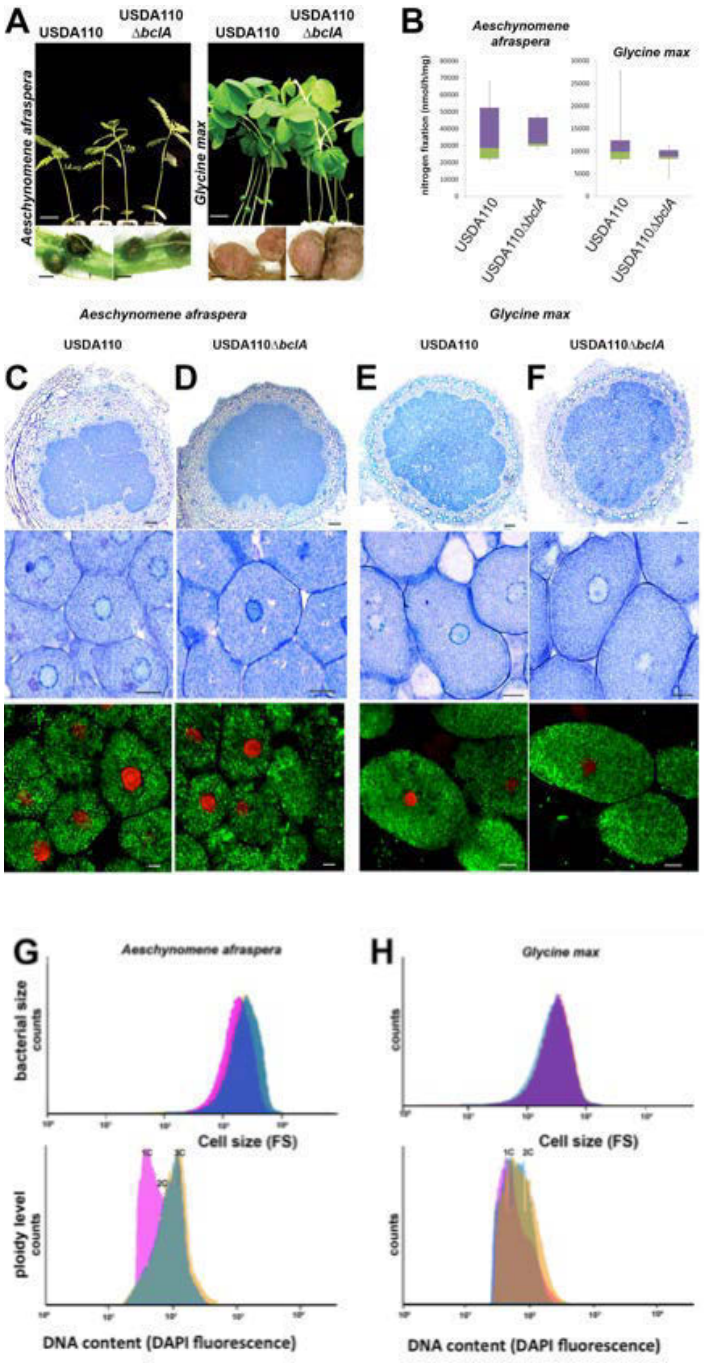

Figure 51. The *B. japonicum bclA* mutant is not affected in symbiosis

(A) Plant growth and nodule phenotype of *A. afraspera* (left panels) and *G. max* (right panels) plants inoculated with *B. japonicum* USDA110 wild-type and the *bclA* mutant at 14 dpi. Scale bars are 2 cm (top left panel), 3,5 cm (top right panel) and 1 mm (bottom panels). **(B)** The nitrogen fixation activity of *A. afraspera* (left panel) and *G. max* (right panel) plants inoculated with *B. japonicum* USDA110 wild-type and the *bclA* mutant, measured by the acetylene reduction assay per mg of fresh nodule weight (nmol ethylene produced per hour incubation and per mg nodule weight). Box-plots represent in the rectangle the first quartile to the third quartile, divided by the median value, whiskers above and below the box show the minimum and maximum measured values. **(C-F)** Toluidine blue stained thin sections (two top rows) and confocal microscopy of bacteroid viability determination by live/dead staining (bottom row) of *A. afraspera* **(C,D)** and *G. max* **(E,F)** nodules induced by USDA110 or USDA110Δ*bclA*. Scale bars are 100 µm for the top panels and 10 µm for the other panels. **(G,H)** Flow cytometry analysis of bacteroid size by forward scatter (top) and DNA content by DAPI fluorescence (bottom) in free living *B. japonicum* USDA110 (pink) or bacteroids isolated from *A. afraspera* **(G)** or *G. max* **(H)** nodules infected with USDA110 wild-type (blue) or USDA110Δ*bclA* (orange).

The absence of a pronounced differentiation was confirmed in flow cytometry analysis by the minor increase of the FS and the slight increase in DNA content to a mean value of 3C (**Fig. 51C**). Thus the USDA110 wild-type bacteroids have very similar features as the ORS285Δ*bclA* mutant in terms of size and DNA content.

The *bclA* mutant of USDA110 was undistinguishable from the wild-type strain for nodulation of both host plants for all parameters analyzed, including nodule tissue structure and bacterial occupation, bacterial viability, morphology, size and DNA content as well as nitrogen fixation (**Fig. 51**). Thus the *Bradyrhizobium bclA* gene, similarly to *bacA*, is not required for symbiosis when bacteroids are not constrained by the host plant to differentiate in an elongated or spherical morphotype. We also confirmed the inessentiality of *bacA* for soybean symbiosis in another soybean symbiont, *Sinorhizobium fredii* strain HH103. The inactivation of the *bacA* gene in this strain, closely related to *S. meliloti bacA* (**Fig. 42**), has no consequence on the symbiosis with soybean (**Fig. 52**).

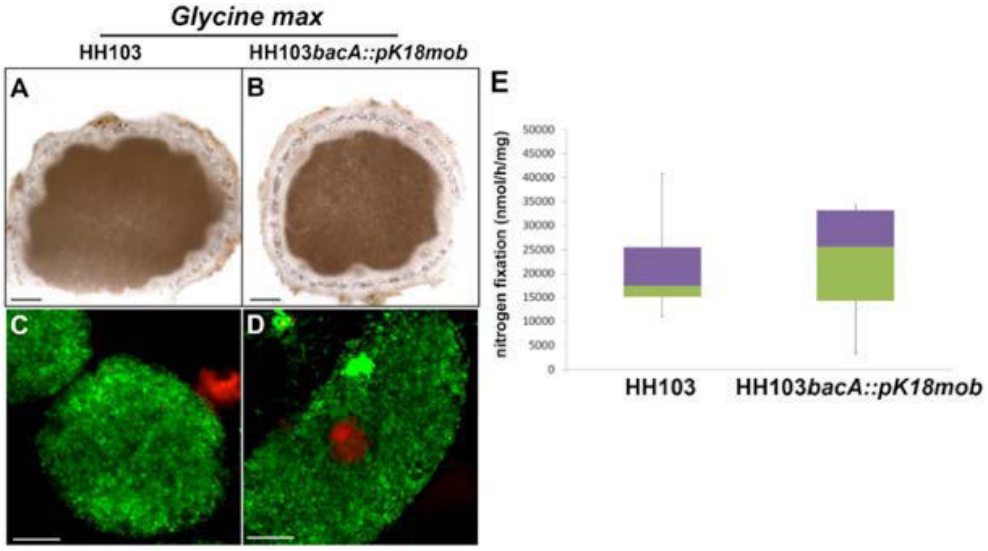

Figure 52. The *bacA* gene of *S. fredii* strain HH103 is not required for symbiosis with soybean

(A) Soybean nodule infected with HH103. **(B)** Soybean nodule infected with the *bacA* mutant HH103*bacA::pK18mob*. **(C,D)** Confocal microscopy of bacteroid viability determination by live/dead staining of a soybean nodule infected with strain HH103 **(C)** or strain HH103*bacA::pK18mob* **(D)**. **(E)** The nitrogen fixation activity of soybean plants inoculated with HH103 or HH103*bacA::pK18mob*, measured by the acetylene reduction assay per mg of fresh weight nodules (nmol ethylene produced per min and per plant). Box-plots represent in the rectangle the first quartile to the third quartile, divided by the median value, whiskers above and below the box show the minimum and maximum measured values. Plants and nodules were harvested at 14 dpi. Scale bars are 4 mm **(A-B)** or 10 µm **(C-D)**.

To exclude the possibility that *bclA* of strain USDA110 is an inactive gene, we confirmed its capacity to complement the *S. meliloti bacA* mutation in a similar way as the *bclA* gene of strain ORS285 does (**Fig. 53**).

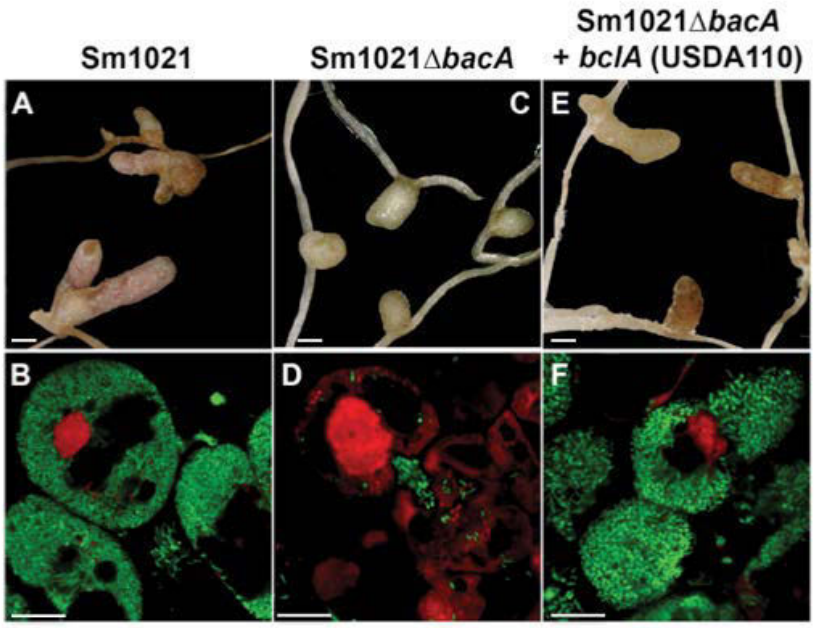

Figure 53. The *bclA* gene of *B. japonicum* USDA110 complements the *bacA* mutation in *S. meliloti* Sm1021Δ*bacA*

(A,C,E) Phenotype of *M. sativa* nodules at 28 dpi, infected with the indicated strains. **(B,D,F)** Bacteroid viability determined by live/dead staining of nodule sections and confocal microscopy in *M. sativa* nodules induced by the indicated strains. Scale bars are 1 mm **(A-C-E)** or 10 µm **(D-E-F)**.

3.7. *Bradyrhizobium* BclA has multidrug transport activity

Different peptide-derived antibiotic compounds have been shown to be taken up by bacterial cells in a SbmA/BacA dependent manner. Among them is bleomycin, which is a DNA damaging antibiotic. In both *E. coli* and *S. meliloti*, the resistance to bleomycin is increased by mutations in *sbmA* and *bacA*, respectively (Ichige *et al*., 1997; Wehmeier *et al*., 2010). We observed that the *bclA* mutants of *Bradyrhizobium* strains ORS278, ORS285 and USDA110 were more resistant to bleomycin (**Fig. 55A-C**) and the ORS285 and USDA110 *bclA* genes conferred sensitivity to bleomycin in an *E. coli sbmA* mutant or in an *S. meliloti bacA* mutant (**Fig. 54**) indicating that BclA can mediate the import of bleomycin.

Bac7 is a proline-rich cathelicidin-type of antimicrobial peptide with an intracellular mode of action, binding ribosomes and thereby inhibiting the translation process (Mardirossian *et al*., 2014). The uptake of the Bac7 peptide is depending on the SbmA or BacA transporter in *E. coli* or *S. meliloti* respectively (Mattiuzzo *et al*., 2007; Marlow *et al*., 2009). Therefore we wondered whether the BclA proteins can also import the Bac7 peptide into bacterial cells. Unfortunately, the Bac7 peptide has no antimicrobial activity in the tested concentration range, up to 4 µM, on the 3 *Bradyrhizobium* strains used in this study (data not

shown). Therefore, we tested the effect of the BclA proteins in *E. coli* or *S. meliloti*. The expression of the *bclA* genes of strains ORS285 and USDA110 in the *E. coli sbmA* or *S. meliloti bacA* mutants rendered them more sensitive to the Bac7 peptide in agreement with the BclA proteins able to internalize also the Bac7 peptide (**Fig. 54**).

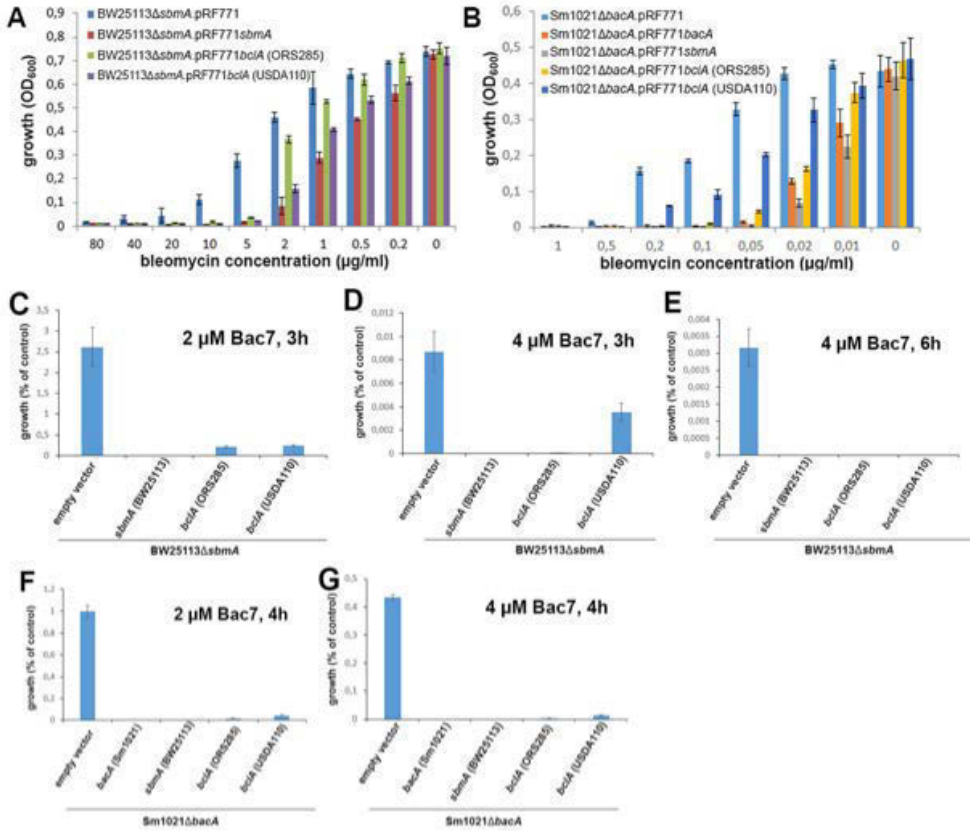

Figure 54. The *bclA* gene confers sensitivity to the antibiotic bleomycin and the antimicrobial peptide Bac7

(A) Growth in the presence of bleomycin in BW25113**Δ***sbmA* and derivatives expressing the indicated SbmA_BacA domain transporters located on plasmid pRF771. **(B)** Growth in the presence of bleomycin in Sm1021**Δ***bacA* and derivatives expressing the indicated SbmA_BacA domain transporters located on plasmid pRF771. Bleomycin concentrations were applied as indicated and growth was determined after 24h incubation for *E. coli* and 48h for *S. meliloti* by optical density measurement at 600 nm with a plate reader. **(C-G)** Survival of *E. coli* strain BW25113**Δ***sbmA* **(C-E)** and *S. meliloti* strain Sm1021**Δ***bacA* **(F,G)** derivatives expressing the indicated SbmA_BacA domain transporters located on plasmid pRF771 after treatment with the peptide Bac7 at the indicated concentration and for the indicated times. The surviving bacteria were counted and expressed as % from the water control treatment.

S. meliloti BacA enhances the resistance of the bacterium towards NCR peptides produced by the symbiotic cells in *Medicago* nodules (Haag *et al.*, 2011). Therefore, we analyzed how the BclA transporter affects the bacterial sensitivity to NCR peptides. Unfortunately, none of the currently tested *Aeschynomene* NCR peptides nor any of the *Medicago* NCR peptides had an antimicrobial activity on *Bradyrhizobium* or their *bclA* mutants (Czernic *et al.*, 2015). Therefore, we assayed the BclA proteins in the *S. meliloti*

background as above for the Bac7 peptide. The *S. meliloti*Δ*bacA* mutant is highly sensitive to 3 tested NCR peptides, NCR247, NCR335 and NCR035 and this sensitivity is strongly reduced when the *bacA* gene is reintroduced on a plasmid, as we reported before (Haag *et al*., 2011), or when the *sbmA* gene of *E. coli* is replacing *bacA* (**Fig. 55D-F**).

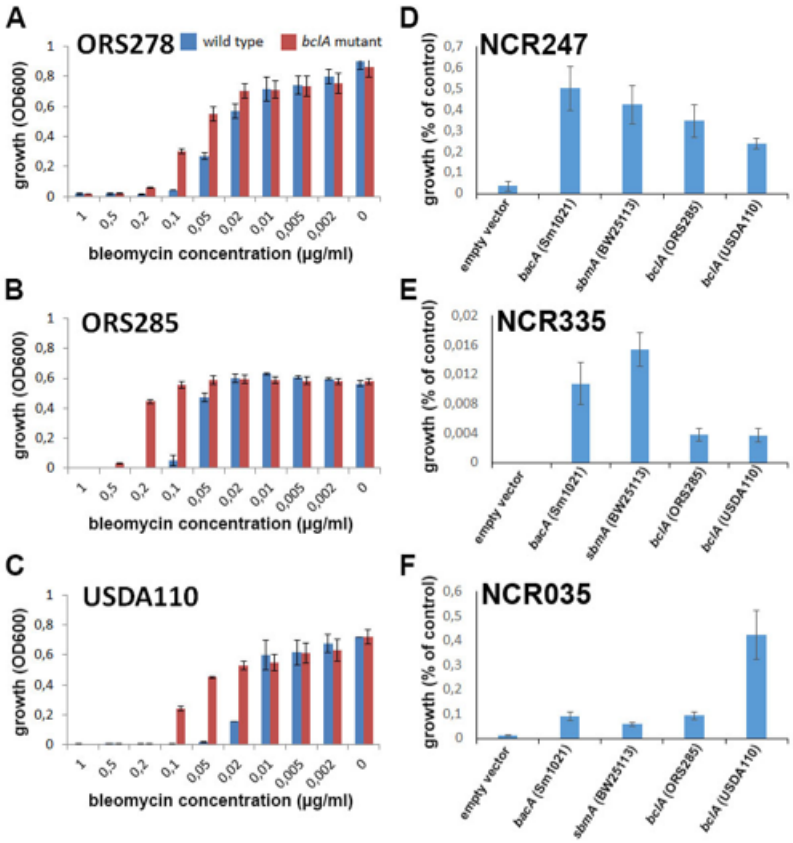

Figure 55. The *bclA* gene confers sensitivity to the antibiotic bleomycin and resistance to antimicrobial NCR peptides

(A-C) Bleomycin sensitivity in *Bradyrhizobium* strains ORS278, ORS285 and USDA110 (blue bars) and their *bclA* mutant derivatives (red bars). Bleomycin concentrations were applied as indicated and growth was determined after 72h incubation by optical density measurement at 600nm with a plate reader. **(D-F)** *S. meliloti* strain Sm1021Δ*bacA* derivatives expressing no *bacA*-related gene (empty vector), the Sm1021 *bacA* gene, the *E. coli* BW25113 *sbmA* gene, the ORS285 *bclA* gene or the USDA110 *bclA* gene were incubated with NCR247 **(D)**, NCR335 **(E)** or NCR035 **(F)** or with water (control) and the surviving bacteria were counted and expressed as % from the control treatment. Error bars in all panels are standard deviation.

Importantly, also the *bclA* genes of *Bradyrhizobium* strains ORS285 and USDA110 are able to provide protection against the activity of the tested NCR peptides. It is interesting to note that the BacA, SbmA and ORS285 or USDA110 BclA transporters have a differential effect on the different tested peptides. For example, BacA is more active against NCR335 than the two BclA proteins (**Fig. 55E**) while USDA110 BclA is the most active against NCR035 (**Fig. 55F**) but the least against Bac7 (**Fig. 54**).

The SbmA and BacA transporters modulate the bacterial susceptibility to bleomycin and Bac7 by promoting the uptake of these peptides (Mattiuzzo *et al*., 2007; Marlow *et al*., 2009; Wehmeier *et al*., 2010; Mardirossian *et al*., 2014). Whether the NCR peptides are also taken up by the BacA, SbmA or BclA transporters is unknown. We used a fluorescent derivative of the NCR247 peptide, carrying at its N-terminus a fluorescein isothiocyanate (FITC) modification. FITC-NCR247 had been shown to be taken up by free bacteria and bacteroids (Farkas *et al*., 2014). We used flow cytometry to measure the uptake of FITC-NCR247 by the *S. meliloti* strain Sm1021 or by Sm1021Δ*bacA* derivatives expressing from a plasmid-born copy the *bacA*, ORS285 *bclA* or USDA110 *bclA* gene or no gene (**Fig. 56**).

The mutant without any transporter was unable to take up the peptide under the applied conditions while the wild type strain or the mutant strains expressing one of the tested genes were able to transport the fluorescent peptide as revealed by the appearance of a fluorescent bacterial population. Fluorescence microscopy of sorted FITC-positive bacteria confirmed that these bacteria had taken up the NCR peptide, which appeared in the cells as discrete foci (**Fig. 56**). In this assay, only a fraction of the bacteria were FITC positive. The reason for this is presently not understood and requires further analysis. Possibly the NCR peptide or its FITC moiety was not stable in the cells or FITC fluorescence is quenched and only a portion of the cells have maintained sufficient FITC to be detected in the assay.

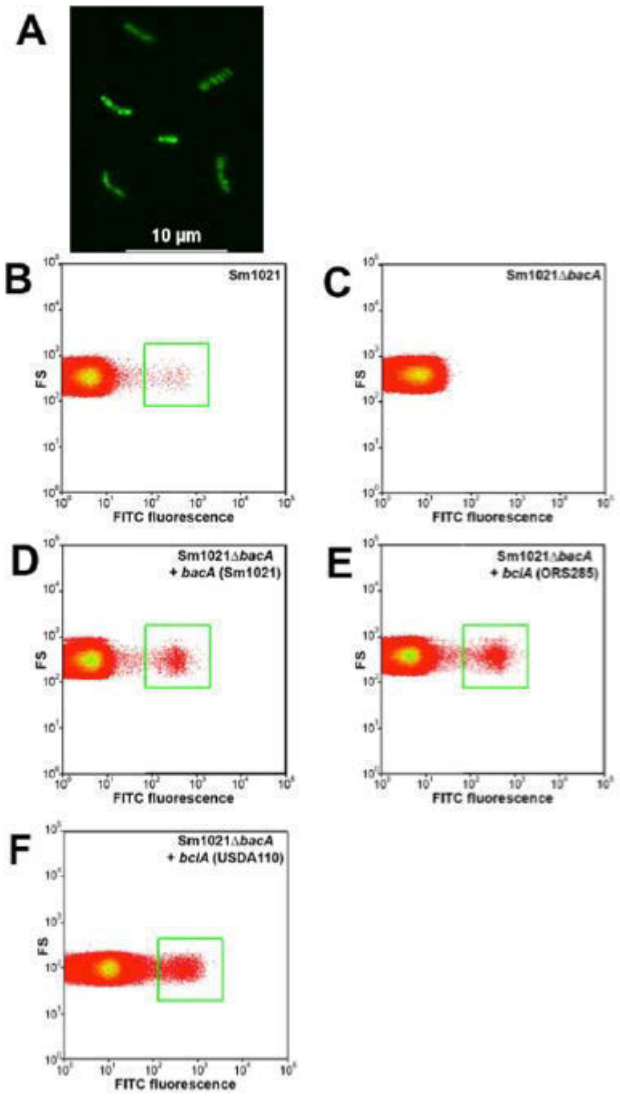

Figure 56. Uptake of FITC-NCR247 mediated by the BacA and BclA transporters

(A) Epifluorescence microscopy of FITC-NCR247 treated Sm1021Δ*bacA*pRF771*bacA* cells and sorted as FITC positive by the flow cytometer-cell sorter. **(B-G)** FITC-NCR247 uptake by *S. meliloti* strain Sm1021 **(B)** and Sm1021Δ*bacA* derivatives expressing no *bacA*-related gene **(F)**, the Sm1021 *bacA* gene **(D)**, the ORS285 *bclA* gene **(E)** or the USDA110 *bclA* gene **(G)** was measured by flow cytometry in the presence of trypan blue to quench extracellular fluorescence. FITC-positive cells are marked with a green box.

In a control experiment, we used FITC alone in the assay and this compound was unable to label the cells (**Fig. 57**) demonstrating that the uptake of fluorescence is NCR247-dependent. Together our experiments show that BclA proteins mediate resistance to NCR peptides similarly as BacA and possibly this resistance mechanism involves the uptake of the peptides by the BclA or BacA transporter proteins.

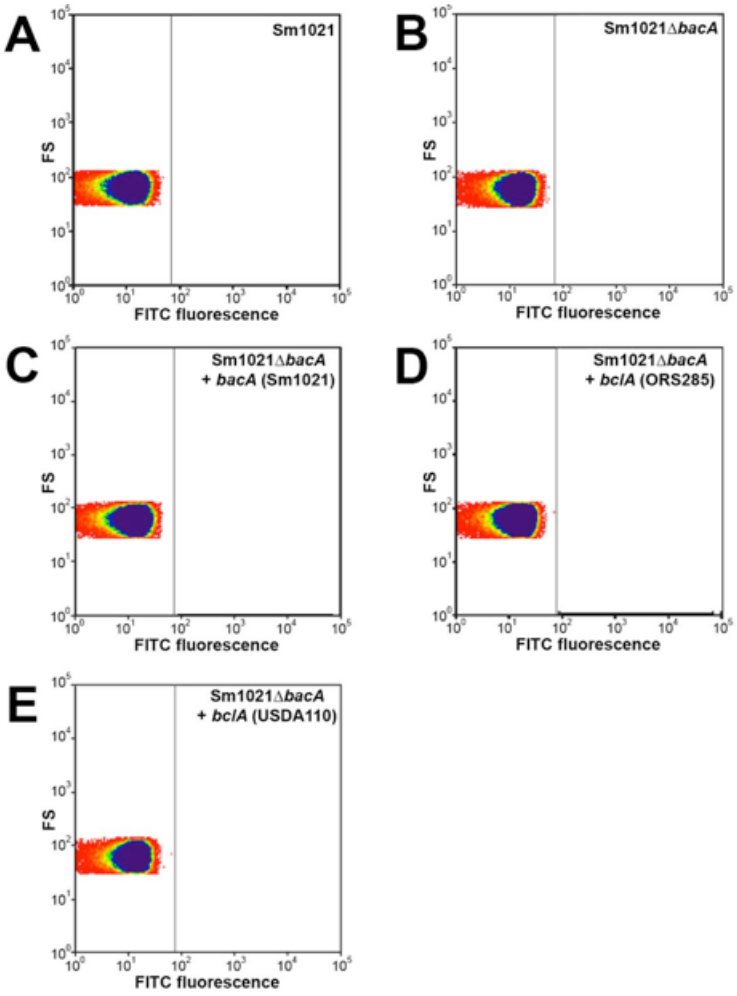

Figure 57. FITC is not taken up by *S. meliloti* derivatives

(A-E) FITC treated bacteria were analyzed by flow cytometry in the presence of trypan blue to quench extracellular fluorescence. The vertical bars separate FITC negative and positive signals. No FITC positive cells were detected in any of the strains.

4. Discussion

Since its initial description in *M. truncatula* (Mergaert *et al.*, 2003), the use of cysteine-rich peptides or other types of antimicrobial peptides in the management of large endosymbiotic bacterial populations has now been found in several other symbiotic interactions, including in other nitrogen fixing interactions of actinorhizal plants with gram-positive *Frankia* bacteria (Carro *et al.*, 2015; Hocher *et al.*, 2011; Demina *et al.*, 2013) but also in interactions of several groups of insects with endosymbiotic bacteria (Login *et al.*, 2011; Shigenobu and Stern, 2012; Futahashi *et al.*, 2013). Even the microbiota in the gut from the most basal animals to the mammals are kept in homeostasis by antimicrobial peptides produced by host epithelial cells (Wehkamp *et al.*, 2005; Ryu *et al.*, 2008; Fraune *et al.*, 2010;

Salzman *et al*., 2010; Vaishnava *et al*., 2011). We recently reported the production of NCR peptides in the symbiotic nodule cells of legume species belonging to the *Aeschynomene* genus and that their bacteroids have similar features, strong cell enlargement, polyploidy, membrane permeabilization and loss of cell viability, to those found in *Medicago* and other IRLC legumes (Czernic *et al*., 2015). These observations suggest thus the distantly related IRLC and Dalbergoid legumes have converged during evolution independently to the same strategy: controlling their endosymbionts by forcing them in an irreversible, differentiated state and using cysteine-rich peptides for this. In the present study, we show that this convergent evolution extends also to the endosymbionts' use of the BacA-related multidrug transporters to be able to respond to the NCR peptides produced by the host cells and to differentiate into functional bacteroids. We showed before that BacA is required for *S. meliloti* to withstand the NCR peptides within the symbiotic nodule cells of *Medicago* nodules and to differentiate into elongated and polyploid bacteroids (Haag *et al*., 2011). Here, we provide multiple lines of evidence that the *Bradyrhizobium* symbionts of *Aeschynomene* use the homologous protein BclA for the same purpose. We show that the *bclA* mutant of *Bradyrhizobium* strains ORS285 and ORS278 does not differentiate (no or little cell enlargement and DNA amplification) within the symbiotic cells of *Aeschynomene* nodules and that on the contrary *bclA* is not needed for the formation of U-morphotype bacteroids in soybean which does not produce NCR peptides in the symbiotic cells. Furthermore, we demonstrate that the *bclA* gene complements partially the *S. meliloti bacA* function in symbiosis with *Medicago* as well as in free living growth, conferring resistance to antimicrobial peptides that have intracellular targets and whose uptake by *E. coli* or *S. meliloti* is dependent on a SbmA-BacA-type of transporter. Finally, we show that BclA, as BacA of *S. meliloti*, promotes the uptake of NCR peptides and confers protection against them.

Although BacA and BclA have a key role in symbiosis, BacA-related proteins do not have an exclusive symbiotic function since they are conserved in non-symbiotic bacteria (e.g. *E. coli* or *Mycobacterium*) or in symbiotic bacteria in which they do not have a role in symbiosis such as the *bclA* gene in *B. japonicum* or the *bacA* gene in *Sinorhizobium fredii*, *Mesorhizobium loti* or *Rhizobium leguminosarum* bv. *phaseoli* (Ardissone *et al*., 2011; Karunakaran *et al*., 2010; Maruya and Saeki, 2010; This work). The 'housekeeping' role of these transporters is hitherto unknown but apparently their capacity to import peptides has been co-opted at least twice among the rhizobia to respond to or provide protection against

peptides produced by the host cell. It will be interesting in the future to determine whether BacA-related proteins have a role in other symbiotic interactions as well.

The *S. meliloti bacA* mutant dies rapidly after its release from the infection threads in *Medicago* nodules as a consequence of its hypersensitivity to the NCR peptides and is then lysed (Glazebrook *et al*., 1993; Haag *et al*., 2011). This results in nodules in which the symbiotic cells filled with bacterial cells rapidly disappear. Here we find that the *Bradyrhizobium bclA* mutant bacteria have a much higher capacity to persist within the symbiotic cells of *Aeschynomene* nodules than the *S. meliloti bacA* mutant in *Medicago* symbiotic cells. This suggests that the *Aeschynomene* symbiotic cells produce a much less aggressive arsenal of peptides than the *Medicago* symbiotic cells. In agreement with this, none of the tested *Aeschynomene* peptides displayed antimicrobial activity (Czernic *et al*., 2015). Alternatively, *Bradyrhizobium* could have a better capacity than *S. meliloti* to maintain its cell integrity after exposure to membrane disrupting molecules such as the NCRs. Such a capacity could be related to the unusual membrane composition of these bacteria which is extremely rich in hopanoids (Silipo *et al*., 2014). Hopanoids are pentacyclic triterpenoids, structural similarity with eukaryotic sterols. Hopanoids stabilize bacterial membranes, controlling its fluidity, permeability and integrity under stress conditions.

Even if the survival of the *bclA* mutant is not immediately affected after their release in the symbiotic cells, it is clearly so at later time points in the *A. indica* and *A. evenia* symbiotic cells but not in the *A. afraspera* symbiotic cells, suggesting that the former may produce a spectrum of peptides which are more virulent to the bacteria than those of the latter. This difference translates to a more severely reduced nitrogen fixation capability of the *bclA* mutant bacteroids in *A. indica* or *A. evenia* nodules compared to *A. afraspera*. The different NCR spectrum could also be at the basis of the S-morphotype versus E-morphotype bacteroids formed by *A. indica* and *A. evenia* compared to *A. afraspera* (Czernic *et al*., 2015). The *bclA* genes of both *Bradyrhizobium* strains ORS285 and USDA110 partially complement the *bacA* mutation of *S. meliloti* in symbiosis with *Medicago*. The complementation is obvious by the production of the late developmental marker leghemoglobin which is never produced in the mutant nodules and by the elongated bacteroids which are not stained with the PI marker for bacterial death. Nevertheless, these nodules remain smaller than wild-type nodules, they do not accumulate leghemoglobin to the same extent, they senesce earlier than normal as indicated by the PI staining of the bacteroids when symbiotic cells grow older, and the nodules do not fix nitrogen or support plant growth. This partial complementation

resembles the complementation level obtained by the *Mycobacterium tuberculosis bacA* gene (Arnold *et al*., 2013). We show that BacA of *S. meliloti* and the BclA proteins from *Bradyrhizobium* strains ORS285 and USDA110 have a differential activity towards different NCR peptides. Also the *Mycobacterium* and *S. meliloti* BacA proteins have a differential response towards derivatives of the Bac7 peptide (Arnold *et al*., 2013). Therefore, the partial complementation in symbiosis could result from a different spectrum of peptides that can be transported or different transport kinetics by the endogenous BacA of *S. meliloti* and the *Bradyrhizobium* BclA proteins or *M. tuberulosis* BacA.

In this respect, it is interesting to note that although BacA, SbmA and BclA share a SbmA_BacA transmembrane domain, they have an otherwise different architecture and transport mechanism: SbmA and BacA lack an ATPase domain and transport is driven by the membrane electrochemical gradient (Runti *et al*., 2013) while BclA and Mycobacterium BacA have a C-terminal cytosolic ATPase domain that provides the energy for transport (Arnold *et al*., 2013).

Medicago produces NCR peptides in different waves and the large majority of peptides are produced at later stages (Maunoury *et al*., 2010; Guefrachi *et al*., 2014). Possibly the BclA proteins or *M. tuberculosis* BacA are sufficient to tolerate the early waves of NCR production but become limited when the later waves are activated, ultimately resulting in the death of the bacteroids, even after their differentiation.

The very weak differentiation of *B. japonicum* strain USDA110 in *A. afraspera* nodules is unlikely to be the result of its *bclA* gene because the USDA110 *bclA* mutant has a similar response to bleomycin as the corresponding mutants in strains ORS285 and ORS278. Moreover, the USDA110 *bclA* gene complements *S. meliloti bacA* and *E. coli sbmA* mutants similarly to the corresponding genes of ORS285 and ORS278. The *Bradyrhizobium* strains have a high proportion of shared genes but have nevertheless also a substantial number of unique genes (Giraud *et al*., 2007). Probably the different genetic repertoires between the strains are at the basis of the differential response of the strains to the bacteroid differentiation factors produced by the symbiotic host cells. This difference could be exploited to identify additional bacterial factors involved in the differentiation process and the response to the NCR peptides. Alternatively, it is possible that the *NCR* genes are not or only weakly induced in *A. afraspera* nodules infected with *B. japonicum* strain USDA110.

The *M. tuberculosis bacA* gene can complement the *S. meliloti bacA* mutation (Arnold *et al*., 2013). It was also shown recently that the *M. tuberculosis* transporter can promote the uptake of vitamin B12 (Gopinath *et al*., 2013). However, we believe that a defect in vitamin

B12 uptake is unlikely at the origin of the symbiotic phenotype of the *bacA* or *bclA* mutants. *Bradyrhizobium* and *Sinorhizobium* have a complete biosynthesis pathway (Rodionov *et al*., 2003) and in *S. meliloti* the pathway genes are needed for symbiosis suggesting that the pathway is active in the bacteroids and that the bacteroids produce their own vitamin B12 (Campbell *et al*., 2006; Taga *et al*., 2007; Taga *et al*., 2010). In addition, *Bradyrhizobium* and *Sinorhizobium* have an identified putative vitamin B12 transport system (Rodionov *et al*., 2003). Moreover, plants do not produce and use vitamin B12 as a cofactor (Roth, 1996) and thus a vitamin B12 transporter would be superfluous within the symbiotic cells. Finally, our experimental data suggest that *bclA* has no impact on cell growth nor nitrogen fixation in the presence or absence of external vitamin B12. Of course, this does not exclude that the BacA or BclA proteins can transport vitamin B12, depending on the environmental context of the bacteria.

BacA-related transporters are not only critical in rhizobia for chronic intracellular infections of legume hosts, which produce NCRs in the symbiotic cells. They also contribute to the pathogenesis and chronic infection in pathogens and they do this most likely by protecting the bacteria against host antimicrobial peptides (LeVier *et al*., 2000; Li *et al*., 2005; Tan *et al*., 2006; Rivas-Santiago *et al*., 2005; Rivas-Santiago *et al*., 2006; Domenech *et al*., 2009; Haag *et al*., 2011; Arnold *et al*., 2013). However, how the transporters function mechanistically and contribute to resistance against antimicrobial peptides requires further investigation. An attractive hypothesis is that the internalization of peptides would move them away from the bacterial membrane which is the main target of many antimicrobial peptides, including the toxic NCRs, and at the same time bring them into proximity of intracellular targets or set them up for degradation. In the case of peptides such as Bac7 or bleomycin, the transporter would be a Trojan horse for the bacterial cell, bringing the enemy inside its walls, within reach of its intracellular targets.

Material & Methods

Material & Methods

1. Analysis of MtGEA data

The MtGEA transcriptome compendium was downloaded from the website of the Samuel Roberts Noble foundation (http://mtgea.noble.org/v3/). The data from Limpens *et al*. (2013) and Jayaraman *et al*. (2014) were obtained from the NCBI Gene Expression Omnibus (accession n° GSE53406). All the data were imported in Excel (**Table 1; http://www.biomedcentral.com/content/supplementary/1471-2164-15-712-S1.xlsx**) for extracting the expression profiles of the 334 NCR probe-sets and for further treatments. The NCR probe-sets on the Affymetrix *Medicago* GeneChip, which was used for the MtGEA transcriptome compendium, were obtained by BLASTn searches on the MtGEA website (**Table 1**). Each individual NCR nucleotide sequence resulted in the identification of multiple probe-sets due to the homology between *NCR* gene sequences. In total 334 different probe-sets were retrieved. This collection represent likely nearly all NCR probe-sets present on the Affymetrix *Medicago* GeneChip and the remaining genes identified in Young *et al*. (2011) and Zhou *et al*. (2013) are missing from these arrays because they were not yet annotated at the time of array design.

Cluster analysis of the complete MtGEA dataset was performed using the MeV software package (http://sourceforge.net/projects/mev-tm4/). Briefly, the Excel datasheet extracted from MtGEA was analysed using the Euclidean distance application with average linkage settings. Heatmaps were generated with MeV and histograms and graphs with Excel.

2. Entropy calculations

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

3. Transcriptome analysis of hand-dissected nodule zones

Using the leghemoglobin colour gradient along the nodule as guideline, five regions from 28 dpi nodules were hand-dissected, as previously described (Zhou *et al*., 2011). These

samples correspond to the nodule tissues from the most apical part of the nodule with the youngest symbiotic cells to the most proximal part containing the oldest symbiotic cells. Sample I is the meristem and the underlying few cell layers of post-meristematic cells which start the infection and differentiation process. Sample II corresponds mainly to the infection and differentiation zone II. Sample II-III corresponds essentially to the interzone II-III. Sample III is the nitrogen fixation zone III, easily characterized by its pink color due to the accumulation of high amounts of leghemoglobin and finally sample IV is the senescence zone IV that is recognized by its green color resulting from the accumulation of biliverdin, a product of the catabolism of leghemoglobin-derived heme. It should be noted that each of these hand-dissected samples is enriched for the indicated zone but can contain cell layers form the adjacent zones as well.

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

4. Transgenic plants and GUS analysis

The promoters of NCR001, NCR084 and NCR121 (respectively 2.5kb, 1.5kb and 1kb fragments upstream of the ATG) were obtained by an Amplified Fragment-Length Polymorphism (AFLP) based PCR protocol as described (Ratet *et al*., 2009) and recombined in the Gateway vector pDONRP4-P1R according to the manufacturer's instructions (Invitrogen). Primers used for the amplification and cloning of the promoters are listed in **Table 8 (Annex 6).**

Entry clones for the GUS ORF and the 35S terminator were obtained in the Gateway vectors pDONR221 and pDONRP2R-P3, respectively (Van de Velde *et al*., 2010). Entry clones were recombined in the binary vector pKm43GW (Karimi *et al*., 2005). Leaf explants from the *M. truncatula* line R108 were transformed using *Agrobacterium tumefaciens* according to the method described in Cosson *et al*. (2006).

For GUS analysis, three independent T2 transgenic lines were each time analysed to avoid positional effects of the transgene insertion. No variations were observed between independent lines. Untransformed plants and the constitutive GUS line pG3.3 (35S promoter fused to GUS) (Mondy *et al*., 2013) were used as negative and positive controls respectively. For nodulation kinetics, R108 plants were cultivated on BNM agar plates and inoculated with

OD_{600nm} = 0.1 suspensions of *S. meliloti* strain 1021 or *Sinorhizobium arboris* strain B554 (synonymous strain names LMG 14919 and HAMBI 1552) (Nick *et al.*, 1999) which is an excellent symbiont of *M. truncatula* R108 forming numerous large, nitrogen fixing nodules. Samples were collected at indicated time points and embedded in 6% agarose. Tissue sections of 70 µm were prepared with a Leica VT1200S vibratome. GUS staining was done as described (Vanstraelen *et al.*, 2009) and was allowed to proceed for 1h (**Fig. 20**). Overnight staining did not alter the expression patterns (data not shown). The pattern of expression of the *NCR* genes in nodules induced by both *Sinorhizobium* strains were very similar.

For all pathogen assays, plants were cultivated on perlite/sand (3/1 vol/vol) substrate and watered with a commercial nutrient solution. Six weeks old plants were transferred to a growth chamber with saturating humidity the day before the inoculations and stayed in these conditions all along the assay. *Dickeya dadantii* 3937, *Pseudomonas syringae*pv. *tomato* DC3000 and its *hrcC* derivative strain were cultivated at 30°C in LB medium. Inocula of OD_{600nm} = 0.1 were resuspended in 10 mM $MgCl_2$ and were syringe infiltrated in the terminal leaflet of 5-8 leaves per plant. Sterile 10 mM $MgCl_2$ solution was infiltrated as mock control. *Botrytis cinerea* strain B05.10 was cultivated on PDA medium (Amselem *et al.*, 2011) at 20°C. Spores were collected in ½ potato dextrose broth with 0.01% Tween 20 and inocula were normalized to 10^6 spores/mL using a Malassez cell. Five microliter drops of mock/inoculum were put on 5 to 8 terminal leaflets per plant. Symptoms were scored at 1, 2 or 7 dpi and leaflets were collected for GUS staining. For wounding experiments, the terminal leaflet of 5-8 leaves per plant were pinched with forceps and collected 24 hours post wounding. Staining for all infections or treatments was allowed for 24 hours in the GUS staining solution at 37°C. The leaflets were transferred to bleach to remove chlorophyll before photographing.

5. Antibodies and immunolocalization

The mature region of the NCR122 peptide was amplified from cDNA and cloned into the expression vector pBADgIII/A (Invitrogen). Recombinant proteins were purified according to the manufacturer's instructions and used for immunization of rabbits by a commercial service (Agro-bio). Immunolocalisations were done exactly as described before (Van de Velde *et al.*, 2010). For the SYTO13 nucleic acid staining, nodules sections were incubated for 5 minutes with 1 µM SYTO13 in H_2O. Immuno- or SYTO13-stained sections were mounted in deionised water for confocal imaging. Fluorescence images were acquired at 1024x1024 pixels resolution with the confocal laser scanning microscope TCS SP2 from

Leica, using 10X water-immersion and 63X oil-immersion objectives and Leica software. Images were processed with Adobe Photoshop for adjustment of contrast and brightness.

6. Bacterial strains and growth media

Bradyrhizobium strains were grown in Yeast Mannitol (YM) (Giraud *et al*., 2000) or BNM-B medium (Renier *et al*., 2011), *E. coli* in Luria Bertani (LB) or Mueller-Hinton (MH) medium and *S. meliloti* strains in LB medium. Bacteria were grown on agar plates (1.5%) or in liquid cultures under agitation at 30°C for *Bradyrhizobiumor S. meliloti* and at 37°C for *E. coli*. When required, the media were supplemented with the appropriate antibiotics. Antibiotics were used at the following concentrations: streptomycin (Sm), 500 µg/ml; spectinomycin (Sp), 100 µg/ml; chloramphenicol (Cm), 12.5 µg/ml; gentamycin (Gm), 15 µg/ml or 50 µg/ml; carbenicilin (Cb), 50 µg/ml; tetracycline (Tc), 10 µg/ml, kanamycin (Km), 50 µg/ml or 200 µg/ml, neomycin (Nm), 120 µg/ml and nalidixic acid (Nal) 25 µg/ml. All bacterial strains and plasmids used in this study are listed in **Table 7** (**Annex 5**).

7. Construction of bacterial mutants and complemented strains

Standard molecular biology techniques were used for all cloning work. All primers used for cloning of DNA fragments of interest are listed in **Table 8 (Annex 6)**.

For the construction of *Bradyrhizobium* strain ORS285 mutants in the genes 250005 (SbmA_BacA domain ABC transporter), 950010 (SbmA_BacA domain ABC transporter), 1320007 (*glcA*) and 1320003 (secreted protein); (full accession numbers are BRAO285v1_250005, BRAO285v1_950010, BRAO285v1_1320007 and BRAO285v1_1320003), 400 to 600 base pairs (bp) fragments were amplified by PCR and cloned into the plasmid pVO155*nptIIgfp* (plasmid pVO155 (Oke and Long, 1999) carrying a constitutively expressed *gfp* gene) which does not replicate in *S. meliloti*. The resulting pVO155 derivatives were verified by PCR and sequencing and subsequently introduced in *Bradyrhizobium* strain ORS285 by triparental mating using the helper strain HB101.pRK600 (Finan *et al*., 1986). ORS285 clones with the plasmid integrated in the genome were selected on YM medium with gm (15 µg/ml) for counter-selection of the *E. coli* and Km for selection of the plasmid. The integration of the plasmids in the target genes by homologous recombination was verified by PCR. The *bacA* mutant in *S. fredii* strain HH103 was constructed with a similar strategy but using plasmid pK18mob (Schafer *et al*., 1994).

Deletion mutants of the *bclA* genes in strains ORS285 and USDA110 (genes BRAO285v1_1320006 and blr7537, respectively) were obtained as follows. The 600 bp

upstream and downstream regions of the *bclA* open reading frames were obtained by PCR (see **Table 8, Annex 6** for primer sequences). The two regions were merged by sewing PCR and cloned in the pGEM-Teasy vector (Promega). The Gm resistance cassette from plasmid p34S-Gm (Dennis and Zylstra, 1998) was cloned in between the upstream and downstream regions. The fragments of the upstream region-Gm cassette-downstream region were subsequently transferred into the vector pK18mob-sacB (Schafer *et al*., 1994). This plasmid cannot replicate in *Bradyrhizobium* and carries the *sacB* gene that induce bacterial death in the presence of sucrose. The pK18mob-sacB derivatives were introduced in the respective *Bradyrhizobium* strains by triparental mating as above. Single recombinant clones were obtained by antibiotic selection and verified by PCR. Double recombinant clones were then obtained by growth on sucrose and Gm (50 µg/ml). Candidate clones were verified for the loss of Km resistance from the pK18mob-sacB plasmid and the replacement of the *bclA* gene by the Gm cassette was verified by PCR.

For complementation experiments of *bacA* or *sbmA* mutations in *S. meliloti* Sm1021 or *E. coli* BW25113, the genes of interest were introduced into the broad host range vector pRF771 under the control of the *trp* promoter (Wells and Long, 2002). The plasmid pRF771*bacA*^{Sm1021} (Haag *et al*., 2011) carries the *S. meliloti bacA* gene. The *sbmA* gene from *E. coli* strain BW25113 and the *bclA* genes from *Bradyrhizobium* strains ORS285 and USDA110 were obtained by PCR and cloned in the multicloning site of pRF771. The plasmids were introduced in the *S. melilot* mutant Sm1021**Δ***bacA* by triparental conjugation and in the *E. coli* mutant BW25113**Δ***sbmA* by electroporation (**Table 7; Annex 5**).

8. Plant growth and nodulation

A. indica, *A. evenia* or *A. afraspera* seeds were surface sterilized with concentrated sulfuric acid for 45 min followed by 30 min with bleach. The seeds were germinated overnight at 34°C in the dark on tap-water agar plates. One day-old seedlings were transferred to test tubes containing BNM medium (Ehrhardt *et al*., 1992). Seedlings were grown at 28°C with a 16 h light regime and 70% humidity. Seven days after germination, each seedling was inoculated with 2 mL of a bacterial suspension adjusted to and OD_{600nm} of 0.2.

Glycine max (Williams 82) seeds were cleaned with 100% ethanol for 30 sec and sterilized with bleach for 15 min. Seeds were germinated on tap-water agar plates at room temperature for two days. Seedlings were then transferred to bottles filled with BNM medium, inoculated and grown as described above for *Aeschynomene* plants.

*Medicago sativa (*alfalfa) seeds were sterilized with bleach for 15 min and germinated

on tap-water agar plates at room temperature overnight. Seedlings were planted in pots containing a mixture of perlite and sand (3:1, v/v) and grown at 24°C with a 16 h light regime and 40% humidity. Three-days old plantlets were inoculated with a bacterial suspension as described above. Plants were watered with BNM medium.

9. Acetylene reduction assay on plants

Acetylene reduction assays were performed on plants 2 weeks after inoculation with a protocol modified from Koch and Evans (1966). Single whole plants were placed into 20 ml glass vials sealed with rubber septa prefilled with 200 µL of sterile water to avoid drying of the plant during the experiment. Two hundred µL of acetylene was injected into each vial. Gas samples (200 µL) werewithdrawn after 3 hours of incubation at 28°C and the ethylene produced was measured by gas chromatography on a 7820A GC system (Agilent Technologies). The assay was done with a minimum of 5 replicates and included always measurements on uninoculated plants.

10. Bacterial growth and nitrogen fixation assays

Bradyrhizobium ORS285 and ORS285**Δ***bclA* were grown in YM medium, washed and the pellet was resuspended in different media: in rich YM medium; in minimal BNM-B medium; in BNM-B medium in which the vitamin B12 component was omitted and in BNM-B medium in which the only carbon source was glycolate (BNM-B glycolate). The bacterial suspensions were diluted to OD_{600nm} = 0.01 and then distributed in microtiter plates, 100 µl per well. The plates were incubated with agitation (200 rpm) at 30°C and the bacterial growth was monitored by optical density measurements (OD_{600nm}) in a plate reader at different times.

For free-living nitrogen fixation activity measurements with the acetylene reduction assay, pre-cultures in YM medium were pelleted by centrifugation and resuspended in BNM-B medium without nitrogen source. When indicated, the vitamin B12 component was omitted from the medium. Two hundred µl of cell suspensions at OD_{600nm} = 1.5 were added to 10 ml glass vials containing 7 mL melted 1% agar medium, precooled at 45°C. The tubes were sealed with rubber septa, 0.85 ml air from the container was removed with a syringe, and 0.8 ml of 100% acetylene was injected. The cultures were incubated at 28°C in the dark during 7 days. Gas samples were analyzed for the amount of ethylene formed by gas chromatography as above.

11. Microscopy

11.1. Histological analysis of nodules

Nodules were fixed in 1% glutaraldehyde, 4% formaldehyde in 0,1 M Phosphate buffer pH 7.2 then washed, dehydrated and embedded in Technovit 7100 resin (KulzerHistoTechnik). 5 µm sections were cut with a Leica RM2155 microtome, stained with 0.005% Toluidine Blue and then mounted with DPX (VWR International Ltd). Bright filed images were acquired with an Eclipse 80i microscope (Nikon).

11.2. *In vivo* Live-Dead staining and confocal microscopy

Nodules were harvested, embedded with 6% agarose and then freshly sectioned with a Leica VT1200S vibratome (Leica Microsystems GmbH, Germany) into 70 µm tissue slices. Slices were incubated in Live/Dead BacLight (Molecular Probes) staining solution for 20 min (Haag *et al*., 2011). Section were washed from excess of dye and observed using a Leica TCS SP8X confocal microscope. 488 and 543 nm laser lines from a white light laser (Leica) were sequentially used and SYTO9 and PI fluorescence intensity were recorded into separated channels.

11.3. *In vivo* DND-99 staining and confocal microscopy

Nodules were freshly sectioned as described above and incubated 20 min with 5 µM DND-99 LysoTracker Red (Molecular Probes) in a 50 mM Tris buffer pH 7.0 (Pierre *et al*., 2013). Excess of dye was removed with buffer and sample were imaged with a Leica TCS SP8X confocal microscope after a 543 nm laser line illumination.

12. Fluorescent *in situ* Hybridization

A. indica and *A. afraspera* nodules were harvested at 14 dpi and fixed immediately with 4% paraformaldehyde containing 0.1% Tween 20 and 0.1% Triton X100 at 4°C overnight. Nodule tissues were dehydrated by passing in progressively higher concentrations of ice-cold ethanol series. They were then embedded in paraffin blocks as described by Javelle *et al*. (2011). Sections (7 µm thick) were cut with a microtome RM2155 (Leica, Wetzlar, Allemagne). Sections were adhered on poly-L-lysine-coated slides and paraffin was subsequently removed with histoclear and histoclear/ethanol series. To produce the hybridization probes, PCR products of approximately 200 bp of the genes of interest were cloned into the pGEM-T Easy plasmid. The inserts were then amplified by PCR using the

vector-specific T7 and SP6 primers flanking the insert. These primers correspond to the bacteriophage SP6 and T7 RNA polymerase promoters. RNA probes, labeled with digoxigenin (DIG), were obtained by *in vitro* transcription using the PCR product as a template and the T7 RNA polymerase of the DIG RNA Labeling Kit (SP6/T7) (Roche, Boulogne-Billancourt, France). Antisense probes were used to detect the location of transcripts of interest and sense probes were used as negative controls. A control without probe was also performed.

Pre-hybridization, hybridization, post-hybridization treatments and mounting were done by a method derived from the procedure described by Javelle *et al*. (2011). To reveal the hybridization signal, an anti-DIG antibody coupled to alkaline phosphatase was used with its substrate NBT/BCIP (4-nitro blue tetrazolium chloride/5-bromo-4-chloro-3-indolyl-phosphate) which forms a brownish precipitate. The slides were finally observed under a Reichert Polyvarepifluorescence microscope (Buffalo, USA) combined with a Retiga 2000 CCD camera (QImaging, Surrey, Canada).

13. Bacteroid purification and flow cytometry analysis

Bacteroids or cultered bacteria were purified as described (Mergaert *et al*., 2006). Bacteria were fixed by heat treatment (70°C, 10min) and stained by 20 µg/ml of DAPI before flow cytometry analysis.

The bacteroid preparations and similarly prepared suspensions of bacteria grown in culture were analyzed by flow cytometry using a MoFlo ADTRIOS flow cytometer (Beckman Coulter). Bacterial size was estimated by the Forward Scatter (FS) and the bacterial DNA content was assessed by the DAPI fluorescence with a 355 nm laser line. Each single event was recorded and analyzed with the Summit 6.2 software (Beckman Coulter). Experiments have been performed in at least 3 biological replicates.

14. Bleomycin, Bac7 and NCR sensitivity assays

E. coli, *S. meliloti* and *Bradyrhizobium* stains were grown in liquid MH, LB and YM medium at 37°C, 30°C or 30°C, respectively in the presence of appropriate antibiotics for plasmid selection. Within an experiment, cultures were diluted until all strains grew at the same optical density. Sensitivity tests were performed on cultures that had not grown over $OD_{600nm} = 0.5$. All experiments have been performed at least in triplicate with identical results. A representative example is shown in the figures.

For bleomycin sensitivity tests, bacterial suspensions were diluted with fresh medium until OD_{600nm} = 0.005 for *E. coli* and *S. meliloti* or OD_{600nm} = 0.01 for *Bradyrhizobium* strains. The bacterial suspensions were distributed in microtiter plates, 100 µl per well, and bleomycin stock solution was added to reach the desired final concentration. The plates were incubated with agitation (200 rpm) at 30°C for *Bradyrhizobium* and *S. meliloti* or at 37°C for *E. coli*. Growth was determined after 24h for *E. coli*, 48h for *S. meliloti* or 72h for *Bradyrhizobium* by measurement of the OD_{600nm} with a plate reader.

For Bac7 sensitivity tests, *S. meliloti* or *E. coli* suspensions were diluted to OD_{600nm} = 0.005 in LB medium and 100 µl suspensions were incubated in the presence of 2 µM or 4 µM Bac7 peptide at 30°C or 37°C for 4h or 3 and 6h respectively. Ten-fold dilution series were then spotted on agar plates and the colony forming units were determined after 72h or 24h growth at 30°C or 37°C.

NCR sensitivity assays were performed as before (Haag *et al.*, 2011). *S. meliloti* strains were re-suspended in 10 mM sodium phosphate buffer, pH 7.0. Hundred microliter of the suspensions were incubated with the indicated NCR peptides at a final concentration of 50 µg/ml for 4h at 30°C. Ten-fold dilution series were then spotted on agar plates and the colony forming units were determined after 72h at 30°C.

15. Single cell NCR peptide uptake assay and flow cytometry

Strains were growth in LB medium to mid-log phase, below OD_{600nm} = 0.5, washed and resuspend at OD_{600nm} = 0.05 in 50 mM phosphate buffer pH 7.0. Then, 1 µM FITC-NCR247 (Farkas *et al.*, 2014) or free FITC were added and cells were incubated for 1h at 30°C. Cells were washed and resuspended in 50 mM phosphate buffer pH 7.0, supplemented with 1 mg/ml Trypan Blue used as extracellular fluorescence quencher (Mattiuzzo *et al.*, 2007). FITC uptake was assayed using a MoFlo ADTRIOS flow cytometer (BeckmanCoulter) with a 488 nm laser and fluorescence was recorded through a 526/552 nm band pass filter. More than two hundred thousand events were acquired and analyzed with Summit 6.2 software (BeckmanCoulter). To further confirm the internalization of FITC-NCR247, FITC-positive cells were sorted and imaged with an Eclipse 80i microscope (Nikon) equipped with a FITC-appropriate filter set.

General Discussion & Perspectives

General Discussion and Perspectives

1. Symbiosis

« No organism is an island. » This variation of the saying « No man is an island. » means that no organism can survive independently of the others and its environment. Interactions between organisms can take many forms. Symbiosis is a particular form of a sustainable interaction between different organisms that appeared several times during evolution. These associations play a central role in contemporary ecosystems. They are, to name just a few, essential for atmospheric nitrogen fixation by plants, for development of coral reefs or for food digestion and many other complex tasks in the human gut. Intracellular symbioses between eukaryotes and bacteria are wide-spread in most ecosystems. Host-associated bacteria have broad phylogenetic distribution and are found in most bacterial phyla. Nevertheless, some bacterial clades such as the α-, β-, and γ-proteobacteria are particularly rich in intracellular symbionts (Toft and Andersson, 2010). Reciprocally, the eukaryotic hosts for these bacteria also span the tree of live ranging from fungi, protists, animals to plants.

Symbiont transmission maintains symbioses through host generations and has a pivotal role in their evolution (Ewald, 1987; Kato *el al*., 1999). Two fundamentally different modes of transmission can be distinguished: horizontal (that is, from an environmental, free-living symbiont source) and vertical (that is, inheritance of the symbiont from the mother or, more rarely, from both parents). Traditionally, bacterial symbionts are divided into two groups. The first consists of obligate mutualists required to support normal host development. These symbionts, also called primary symbionts are typically restricted to a specialized symbiotic organ (Baumann, 2005). They occur in many terrestrial arthropods as well as some marine invertebrates. In contrast, facultative or secondary symbionts resemble invasive pathogens in that they may invade various cell types, including reproductive organs, and may reside extracellularly in the body cavity (hemolymph) (Dobson *et al*., 1999; Fukatsu *et al*., 2000; Moran *et al*., 2008).

Bacterial intracellular symbiosis (endosymbiosis) is widespread in invertebrates and exhibits a large variety of phenotypes, ranging from mutualism to pathogenesis. A single bacterium can be pathogenic on some hosts and beneficial for some others, like *Xenorhabdus nematophila* that colonize the entomopathogenic nematode *Steinernema carpocapsae* in a

mutualistic manner and is pathogenic to insects notably by induction of macrophage apoptosis (Herbert and Goodrich-Blair, 2007). Another example is *Wolbachia* commonly found in arthropods and nematodes, being partly pathogenic to the firsts and mutualistic to the latter (Iturbe-Ormaetxe and O'Neill, 2007; Werren *et al*., 2008; Olivier *et al*., 2010).

Other well documented symbioses involve organ development and often require an active participation of the microsymbiont in the organogenesis. The marine squid *Euprymna scolopes* for example has a light-emitting organ which is colonized by luminescent *Vibrio fischeri* bacteria (Nyholm *et al*., 2004). Aphids and other insect groups harbor intracellular symbionts, in specialized cells called bacteriocytes which are organized in the bacteriome organ located on the abdomen of the insect (Baumann, 2005; Moran *et al*., 2008). These insects exploit restricted nutritional resources such as plant sap and their bacterial symbionts produce nutrients that are lacking in the host's diet. These examples constitute fascinating models of endosymbioses that can be used for comparisons with the *Rhizobium*-legume model (Baumann *et al*., 1995; Visick and Ruby, 2006).

Indeed, some aspects of bacteroid physiology are reminiscent of stages of obligatory endosymbionts in insects. For example genome amplification coupled to cell enlargement has been observed, as we reported for the microsymbionts of IRLC legumes or the Dalbergoid legumes (Komaki and Ishikawa, 1999; Komaki and Ishikawa, 2000; Mergaert *et al*., 2006; **Chapter II**). The morphological modifications these endosymbionts are undergoing inside the host cell such as formation of spheres or elongated cells is similar to the formation of S-morphotype or E-morphotype bacteroids (Shigenobu and Wilson; The International Aphid Genomics Consortium, 2010; Tsuchida *et al*., 2010; Login *et al*., 2011). Strikingly, this symbiont differentiation occurs in parallel to the differentiation of the bacteriocytes by enlargement and endoredupliaction, similarly to the nodule symbiotic cells (Buchner, 1965; Douglas and Dixon, 1987; Cebolla *et al*., 1999).

2. Symbiotic antimicrobial peptides are wide-spread in plants and animals

The involvement of host antimicrobial peptides in controlling the endosymbionts' multiplication and cell fate was described first in the legume *Medicago* but is not unique for the *Rhizobium*-legume symbiosis. In insects of the Sitophilus genus (weevils), the bacteriocytes contain giant filamentous endosymbionts and produce the antimicrobial peptide coleoptericin-A (ColA) (Login *et al*., 2011). Silencing the *colA* gene resulted in reduced elongation of the endosymbionts and their spreading from the bacteriocytes into insect tissues,

indicating that ColA is essential for symbiont differentiation, regulating their growth through inhibition of cell division (Login *et al*., 2011). Interestingly, this scenario also operates in the bacteroids of symbiotic nodule cells.

In the bacteriocytes of aphids, the expression of a unique class of genes was discovered encoding small proteins with signal peptides that were often cysteine-rich, reminiscent of NCRs, and named bacteriocyte-specific cysteine-rich peptides (BCR) (Shigenobu and Stern, 2012). The genes are specifically expressed in the bacteriocytes and not in other parts of the insect. Following their expression by *in situ* hybridization during bacteriome formation in embryos, it was found that they are induced when the early-stage bacteriome becomes first infected with *Buchnera*. Thus, drawing the analogy with the NCR peptides in nodules, it is an inevitable and exciting hypothesis that the BCRs are targeted to and affect the bacterial symbionts in a similar fashion as the NCRs do in symbiotic nodule cells.

Similarly, in another insect, the stinkbug *Riptortus pedestris*, cysteine-rich peptides are specifically produced by a gut-derived symbiotic organ carrying extracellular bacterial symbionts in crypts (Futahashi *et al*., 2013). The authors speculated that these crypt-specific cysteine-rich peptides (CCRs) are secreted into the lumen of the crypts by the epithelial cells and act on the proliferation and physiology of the symbiotic *Burkholderia*, similarly to the NCRs in legume nodules and the BCRs in aphid bacteriomes (Futahashi *et al*., 2013).

Although the functions of these aphid and stinkbug peptides are entirely unknown, it is interesting that legume and insect symbiotic cells evolved similar, but not homologous types of peptides, that are specifically produced in symbiotic organs.

Moreover, AMPs are not only employed in the maintenance of endosymbiotic bacteria but also of the gut microbiota. Studies on gut microbiota in *Hydra*, *Drosophila* and mouse demonstrated that in these very divergent animals the profile and the level of expression of AMPs produced by the gut epithelia are crucial for creating and maintaining a specific and favorable composition of the microbiota (Wehkamp *et al*., 2005; Ryu *et al*., 2008; Fraune *et al*., 2010; Salzman *et al*., 2010; Vaishnava *et al*., 2011).

Before the start of my PhD thesis work, all available data pointed towards a restriction of the *NCR* gene family to the IRLC legumes. However morphologically differentiated bacteroids were also observed in other legume clades (Oono *et al*., 2010) such as the E-morphotype and S-morphotype bacteroids in *Aeschynomene* species, which are also induced by plant factors (Bonaldi *et al*., 2011). In my here presented thesis work, I have participated in the characterization of these bacteroids in *Aeschynomene* species, showing that similarly to

Medicago bacteroids, they are polyploid. Moreover, a gene family encoding NCR peptides was identified and we demonstrated that they are involved in the bacteroid differentiation. The *Aeschynomene* NCRs are unrelated to the NCR sequences of the IRLC legumes. Thus the mechanism of bacteroid differentiation in *Aeschynomene* is similar in its principles to the mechanism employed by IRLC legumes but relies on a distinct class of NCR peptides. Thus two distinct legume clades, the Dalbergoids and the IRLCs, use a very similar molecular mechanism to control their nitrogen-fixing intracellular endosymbionts despite an inferred independent acquisition of that mechanism (Czernic *et al*., 2015).

Moreover, actinorhizal plants such as *Alnus glutinosa*, *Casuarina glauca* and *Datisca glomerata* also produce nodule-specific cysteine-rich peptides (Hocher *et al*., 2011; Demina *et al*., 2013) and recent work showed that these peptides are indeed implicated in the interaction with the *Frankia* endosymbionts in the nodules of those plants (Carro *et al*., 2015).

NCRs are similar to AMPs such as defensins and the analysis of *in vitro* NCR activity demonstrated that some NCR peptides indeed possess genuine antimicrobial properties and effectively kill not only *S. meliloti* (Van de Velde *et al*., 2010) but also other gram-negative and gram-positive bacteria, as well as fungi (Tiricz *et al*., 2013, Ördögh *et al*., 2014). Thus despite the key role of these peptides for the establishment of the symbiosis, some of them have an antimicrobial activity. This raises the question **how the symbionts protect themselves against the arsenal of these effectors?**

Previous work in the laboratory in *S. meliloti* and my work in *Bradyrhizobium* sp. have shown that the BacA and BclA proteins are key factor in the circumvention of the possible harmful consequences of NCR exposure within the symbiosomes (Haag *et al*., 2011, Guefrachi *et al*., 2015). Interestingly, BacA function is also crucial for the pathogenicity of bacteria such as *Brucella* and *Mycobacterium* (Levier *et al*., 2000; Domenech *et al*., 2009). These pathogens establish chronic infections in animal hosts where they need to withstand cocktails of AMPs to survive.

How BacA, BclA or SbmA function at the molecular level remains a matter of speculation at present. The *S. meliloti* BacA, the *E. coli* SbmA and the *Bradyrhizobium* BclA proteins facilitate the uptake of proline-rich peptides as well as the antibiotic bleomycin, which is a peptide derivative, suggesting that those proteins can function as peptide transporters (Mattiuzzo *et al*., 2007; Marlow *et al*., 2009; Wehmeier *et al*., 2010; Guefrachi *et al*., 2015). These transporters also facilitate the uptake of NCR peptides providing a possible mechanism for protection against NCR antimicrobial activity: by taking them away from the bacterial envelop through import, the NCRs are removed from the site where they are toxic

(Guefrachi *et al*., 2015). The BacA/SbmA/BclA transporters seem thus to have a broad spectrum of substrates. The structural basis for this low specificity is an interesting question worth to be followed up in the future. X-ray crystallography could provide proof of peptide transport by these proteins.

The BacA/SbmA/BclA are clearly not symbiosis specific proteins since they are conserved in a broad range of bacteria including non-symbiotic ones. Thus they likely have a housekeeping function. Understanding this housekeeping function might provide more information on their symbiotic role as well. Studying the regulation of the *bclA*/*bacA* genes might be one road to take to learn more about the molecular function and mode of action of these genes. An interesting hypothesis is that the expression of the *bclA* gene is induced by the *NCR-like* peptides produced by the host cells. This would fit to the role of *sbmA* in *E. coli*, which is part of the membrane stress response (Rowley *et al*., 2006; Bury-Moné *et al*., 2009). Indeed, NCRs challenge the symbionts by targeting the rhizobial cell wall (Haag *et al*., 2011) and therefore could induce an envelope stress response. Thus it is possible that BclA and BacA are part of the membrane stress response and that their expression is induced by the NCR peptides produced by the host cells.

Thus the legume symbiosis and the animal symbiotic systems constitute together a striking case of convergent evolution between plants and animals. The similarities suggest that large polyploid host cells, morphologically adapted endosymbionts and the employment of AMP-like peptides constitute an optimal *modus operandi* for symbiotic exchange between a eukaryotic host cell and its endosymbionts and for a host cell to tolerate large numbers of bacteria. Moreover, our studies have uncovered a remarkable conservation in innate immune defense mechanisms among plants, insects, and mammals, which suggests also a common ancestry of the system (Kang *et al*., 1998; Cao *et al*., 2001). It is predicted that further comparative studies between legume and insect symbioses will reveal further common principles, for example, in the mechanisms used for bacterial recognition and selection.

3. The physiological meaning of bacteroid differentiation

Focusing back on *Rhizobium*-legume symbiosis, morphological differentiation of bacteroids is not taking place in all legumes and it is thus not required *per se* for the maintenance of the endosymbiotic bacterial population or for nitrogen fixation. This prompts the question **why certain legumes impose this differentiation process on their nodule rhizobia**. While this question remains entirely open at present, two published studies suggest a selective advantage to the plant (Oono *et al*., 2010; Oono and Denison, 2010). Phylogenetic

analysis and ancestral state reconstruction suggested that the ancestral morphotype of bacteroids was the undifferentiated U-morphotype and that the differentiated E- and S-morphotypes appeared several times independently in the legume family (Oono *et al*., 2010). This conclusion, together with the facts that the plant imposes the morphological differentiation and that the bacteria are affected in their reproductive capacity by the differentiation, is a strong argument for a selective advantage to the plant associated with the morphological bacteroid differentiation. Another study by the same researchers suggested a higher symbiotic performance for E- or S-morphotype bacteroids compared to U-morphotype bacteroids (Oono and Denison, 2010). However, even if these studies are attractive arguments for a host benefit, la **« raison d'être » of morphological bacteroid differentiation (E- and S-morphotype bacteroids) remain entirely unknown.**

An original strategy to answer this question would be the use of rhizobium strains with large host specificity, capable of nodulating legume species with NCR peptides and E-or S-morphotype bacteroids as well as species with U-morphotype bacteroids. As a matter of facts, we are aware of very few examples of such strains. A well-known example of a broad host range strain is *Sinorhizobium fredii* strain NGR234 which nodulates more than 100 genera (Pueppke and Broughton). However this strain does not nodulate legumes of the IRLC or the Dalbergoids, the two clades that we have characterized for their bacteroid morphotype and presence of NCR peptides. Another strain of interest is *S. fredii* strain HH103 which is reported to nodulate soybean as well as alfalfa. However in our hands we could never obtain alfalfa nodulation with this strain despite testing many different plant growth conditions as well as different alfalfa varieties. Probably alfalfa nodulation occurs with very low efficiency.

In the "Dalbergoid" legume clade, bacteroids can be of the E- or S-morphotype. For example, *Bradyrhizobium* strain ORS285 which can nodulate *A. afraspera* and *A. indica* or *A. evenia* transforms into E-morphotype bacteroids in nodules of the former legume and in S-morphotype bacteroids in the latter two legumes (Bonaldi *et al*., 2011). Along the same line, the closely related *B. japonicum* strain USDA110 makes U-morphotype bacteroids on soybean but E-type bacteroids on *A. afraspera*. In terms of ploidy levels of the bacteroids, these symbiotic systems cover bacteroids from very high ploidy (16C in the S-morphotype bacteroids of *A. indica*) over intermediate levels (7C or 3C in *A. afraspera*) to no polyploidy (in soybean). Thus, The A*eschynomene-Bradyrhizobium* / soybean symbiotic couples constitute an appealing model system **for studying the impact of bacteroid morphotype and polyploidy on the functioning of the symbiosis** and **for unraveling the mechanisms of bacteroid morphotype determination.** Thus to better understand the mechanisms involved

in bacteroid differentiation and the evolutionary advantage that it can provide to one or the other symbiotic partner, a comparative study of the different morphotypes can be undertaken using this model system. A preliminary study of the symbiotic efficiency of different morphotypes indicates that S-morphotypes are more efficient than E-morphotypes, themselves better than the U-morphotype. These encouraging results suggest thus that the higher the ploidy levels of the bacteroids the higher symbiotic efficiency is. If these results are confirmed by additional experiments, then the next challenge will be to understand the molecular and physiological basis of this difference.

A possible approach for this question is a comparison by transcriptome, proteome and metabolome analyses of the different bacteroid types. Such approaches will identify genes and pathways that are specific to one or more bacteroids morphotypes. The results coming from these three approaches will also allow better understanding of the molecular mechanisms involved in the bacteroid differentiation and generally speaking in the functioning of the *Rhizobium*-legume symbiosis.

4. Concluding remarks

The *bacA* gene came into the light of research, when it was identified as essential for bacteroid development in *Sinorhizobium meliloti* in the early 1990's (Glazebrook *et al*., 1993). This aroused curiosity about bacA or homologs in other bacteria. Indeed, *bacA* is not a symbiotic gene and in fact it is widespread throughout the bacterial kingdom and can also be found in non-symbiotic bacteria like *E. coli, Mycobacterium* and *Brucella*. The *B. abortus bacA* null mutant was found to induce protective immunity in mice against subsequent challenge with wild-type *B. abortus*, suggesting it could be developed as a potential vaccine strain (Roop *et al*., 2002). A BacA protein was also recently identified to be involved in the maintenance of chronic intracellular infections of *Mycobacterium tuberculosis* within mice (Domenech *et al*., 2009). Accordingly, our effort in understanding the function(s) of BacA and BclA in legume symbiosis will thus also be relevant for identifying novel antibiotic targets and/or vaccine candidates against medically important chronic human infections.

Moreover, the study of *NCR* genes can be of major interest also outside of symbiosis. The ever-increasing prevalence of antibiotic-resistant strains has made it critical that new antibiotics with novel modes of action and more effective means of killing be developed.

Our studies on the NCR will reveal whether these peptides have the potential to be part of a solution to this urgent need.

This thesis provides insights into both the bacterial and the host functions important for the housing of intracellular bacteria. These results will pave the way towards medium-term strategies for improving symbiotic performance in legumes with non-optimal bacteroid morphotype by for example introducing factors for bacteroid differentiation into legumes lacking these factors (e.g soybean). On a long-term perspective, one possible aim could be to transfer efficient symbiotic nitrogen fixation to major agricultural crops like wheat or maize to enhance yield and reducing the need of fertilizers at the same time. Insights from our studies on the management of bacterial populations within host cells might be also relevant in the pursuit of this extremely ambitious goal.

References

Aerts, A.M., François, I.E., Cammue, B.P., and Thevissen, K. (2008). The mode of antifungal action of plant, insect and human defensins. Cell. Mol. Life Sci. 65: 2069-2079.

Alkhalfioui, F., Renard, M., Frendo, P., Keichinger, C., Meyer, Y., Gelhaye, E., Hirasawa, M., Knaff, D.B., Ritzenthaler, C., and Montrichard, F. (2008). A novel type of thioredoxin dedicated to symbiosis in legumes. Plant Physiol. 148: 424-435.

Alunni, B., Kevei, Z., Redondo-Nieto, M., Kondorosi, A., Mergaert, P., and Kondorosi, E. (2007). Genomic organization and evolutionary insights on GRP and NCR genes, two large nodule-specific gene families in *Medicago truncatula*. Mol. Plant Microbe Interact. 20: 1138-1148.

Amid, C., Rehaume, L.M., Brown, K.L., Gilbert, J.G., Dougan, G., Hancock, R.E., and Harrow, J.L. (2009). Manual annotation and analysis of the defensin gene cluster in the C57BL/6J mouse reference genome. BMC Genomics 10: 606.

Amselem, J., Cuomo, C.A., Van Kan, J.A.L., Viaud, M., Benito, E.P., Couloux, A., Coutinho, P.M., De Vries, R.P., Dyer, P.S., Fillinger, S., Fournier, E., Gout, L., Hahn, M., Kohn, L., Lapalu, N., Plummer, K.M., Pradier, J.M., Quévillon, E., Sharon, A., Simon, A., Ten Have, A., Tudzynski, B., Tudzynski, P., Wincker, P., Andrew, M., Anthouard, V., Beever, R.E., Beffa, R., Benoit, I., Bouzid, O., Brault, B., Chen, Z., Choquer, M., Collémare, J., Cotton, P., Danchin, E.G., Da Silva, C., Gautier, A., Giraud, C., Giraud, T., Gonzalez, C., Grossetete, S., Güldener, U., Henrissat, B., Howlett, B.J., Kodira, C., Kretschmer, M., Lappartient, A., Leroch, M., Levis, C., Mauceli, E., Neuvéglise, C., Oeser, B., Pearson, M., Poulain, J., Poussereau, N., Quesneville, H., Rascle, C., Schumacher, J., Ségurens, B., Sexton, A., Silva, E., Sirven, C., Soanes, D.M., Talbot, N.J., Templeton, M., Yandava, C., Yarden, O., Zeng, Q., Rollins, J.A., Lebrun, M.H., and Dickman, M. (2011). Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. PLoS Genet. 7: e1002230.

Anselme, C., Pérez-Brocal, V., Vallier, A., Vincent-Monegat, C., Charif, D., Latorre, A., Moya, A., and Heddi, A. (2008). Identification of the weevil immune genes and their expression in the bacteriome tissue. BMC Biol. 6: 43.

Ardissone, S., Kobayashi, H., Kambara, K., Rummel, C., Noel, K.D., Walker, G.C., Broughton, W.J., and Deakin, W.J. (2011). Role of BacA in lipopolysaccharide synthesis, peptide transport, and nodulation by *Rhizobium* sp. strain NGR234. J. Bacteriol. 193: 2218-2228.

Arnold, M.F, Haag, A.F, Capewell, S., Boshoff, H.I., James, E.K., McDonald, R., Mair, I., Mitchell, A.M., Kerscher, B., Mitchell, T.J., Mergaert, P., Barry, C.E 3rd., Scocchi, M., Zanda, M., Campopiano, D.J., and Ferguson, G.P. (2013). Partial complementation of *Sinorhizobium meliloti bacA* mutant phenotypes by the *Mycobacterium tuberculosis* BacA protein. J. Bacteriol. 195: 389-398.

Arrighi, J.F., Cartieaux, F., Brown, S.C., Goud, M.R., Boursot, M., Fardoux, J., Patrel, D., Gully, D., Fabre, S., Chaintreuil, C., and Giraud, E. (2012). *Aeschynomene evenia*, a model plant for studying the molecular genetics of the nod-independent Rhizobium-legume symbiosis. Mol. Plant Microbe Interact. 25: 851-861.

Arrighi, J.F., Chaintreuil, C., Cartieaux, F., Cardi, C., Goud, M.R., Brown, S.C., Boursot, M., D'Hont, A., Dreyfus, B., and Giraud, E. (2014). Radiation of the Nod-independent *Aeschynomene* relies on multiple allopolyploid speciation events. New Phytol. 201:1457-1468.

Bader, M.W., Sanowar, S., Daley, M.E., Schneider, A.R., Cho, U., Xu, W., Klevit, R.E., Le Moual, H., and Miller, S.I. (2005). Recognition of antimicrobial peptides by a bacterial sensor kinase. Cell 122: 461-472.

Balogh, E.P., Mosolygó, T., Tiricz, H., Szabó, A.M., Karai, A., Kerekes, F., Virók, D.P., Kondorosi, E., and Burián, K. (2014). Anti-chlamydial effect of plant peptides. Acta Microbiol. Immunol. Hung. 61: 229-239.

Barnett, M.J., Toman, C.J., Fisher, R.F., and Long, S.R. (2004). A dual-genome Symbiosis Chip for coordinate study of signal exchange and development in a prokaryote-host interaction. Proc. Natl. Acad. Sci. USA 101: 16636-16641.

Bastiat, B., Sauviac, L., and Bruand, C. (2010). Dual control of *Sinorhizobium meliloti* RpoE2 sigma factor activity by two PhyR-type two-component response regulators. J. Bacteriol. 192: 2255-2265.

Baumann, P. (2005). Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. Annu. Rev. Microbiol. 59: 155-189.

Baumann, P., Baumann, L., Lai, C.Y., Rouhbakhsh, D., Moran, N.A., and Clark, M.A. (1995). Genetics, physiology, and evolutionary relationships of the genus Buchnera: intracellular symbionts of aphids. Annu. Rev. Microbiol. 49: 55-94.

Becker, A., Berges, H., Krol, E., Bruand, C., Ruberg, S., Capela, D., Lauber, E., Meilhoc, E., Ampe, F., De Bruijn, F.J., Fourment, J., Francez-Charlot, A., Kahn, D., Küster, H., Liebe, C., Puhler, A., Weidner, S., and Batut, J. (2004). Global changes in gene expression in *Sinorhizobium meliloti* 1021 under microoxic and symbiotic conditions. Mol. Plant Microbe Interact. 17: 292-303.

Becker, A., Küster, H., Niehaus, K., and Pühler, A. (1995). Extension of the *Rhizobium meliloti* succinoglycan biosynthesis gene cluster: identification of the *exsA* gene encoding an ABC transporter protein, and the *exsB* gene which probably codes for a regulator of succinoglycan biosynthesis. Mol. Gen. Genet. 249: 487-497.

Benedito, V.A., Torres-Jerez, I., Murray, J.D., Andriankaja, A., Allen, S., Kakar, K., Wandrey, M., Verdier, J., Zuber, H., Ott, T., Moreau, S., Niebel, A., Frickey, T., Weiller, G., He, J., Dai, X., Zhao, P.X., Tang, Y., and Udvardi, M.K. (2008). A gene expression atlas of the model legume *Medicago truncatula*. Plant J. 55: 504-513.

Berrabah, F., Bourcy, M., Eschstruth, A., Cayrel, A., Guefrachi, I., Mergaert, P., Wen, J., Jean, V., Mysore, K.S., Gourion, B., and Ratet, P. (2014). A non-RD receptor-like kinase prevents nodule early senescence and defense-like reactions during symbiosis. New Phytol. 203: 1305-1314.

Bonaldi, K., Gargani, D., Prin, Y., Fardoux, J., Gully, D., Nouwen, N., Goormachtig, S., and Giraud, E. (2011). Nodulation of *Aeschynomene afraspera* and *A. indica* by photosynthetic *Bradyrhizobium* Sp. strain ORS285: the nod-dependent versus the nod-independent symbiotic interaction. Mol. Plant Microbe Interact. 24: 1359-1371.

Bonaldi, K., Gourion, B., Fardoux, J., Hannibal, L., Cartieaux, F., Boursot, M., Vallenet, D., Chaintreuil, C., Prin, Y., Nouwen, N., and Giraud, E. (2010). Large-scale transposon mutagenesis of photosynthetic *Bradyrhizobium* sp. strain ORS278 reveals new genetic loci putatively important for nod-independent symbiosis with *Aeschynomene indica*. Mol. Plant Microbe Interact. 23: 760-770.

Bourcy, M., Brocard, L., Pislariu, C.I., Cosson, V., Mergaert, P., Tadege, M., Mysore, K.S., Udvardi, M.K., Gourion, B., and Ratet, P. (2013). *Medicago truncatula* DNF2 is a PI-PLC-XD-containing protein required for bacteroid persistence and prevention of nodule early senescence and defense-like reactions. New Phytol. 197: 1250-1261.

Branca, A., Paape, T.D., Zhou, P., Briskine, R., Farmer, A.D., Mudge, J., Bharti, A.K., Woodward, J.E., May, G.D., Gentzbitte, L., Ben, C., Denny, R., Sadowsky, M.J., Ronfort, J., Bataillon, T., Young, N.D., and Tiffin, P. (2011). Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. Proc. Natl. Acad. Sci. USA 108: 864-870.

Brogden, K.A. (2005). Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? Nat. Rev. Microbiol. 3: 238-250.

Buchner, P. (1965). Endosymbiosis of animals with plant microorganisms. Interscience, New York. 909p.

Bury-Moné, S., Nomane, Y., Reymond, N., Barbet, R., Jacquet, E., Imbeaud, S., Jacq, A., and Bouloc, P. (2009). Global analysis of extracytoplasmic stress signalling in *Escherichia coli*. PLoS Genet. 5: e1000651.

Cabeza, R., Koester, B., Liese, R., Lingner, A., Baumgarten, V., Dirks, J., Salinas-Riester, G., Pommerenke, C., Dittert, K., and Schulze, J. (2014). A RNA-Seq transcriptome analysis reveals

novel insights into molecular aspects of the nitrate Impact on nodule activity of *Medicago truncatula*. Plant Physiol. 164: 400-411.

Campbell, G.R.O., Taga, M.E., Mistry, K., Lloret, J., Anderson, P., Roth, J.R., and Walker, G.C. (2006). *Sinorhizobium meliloti bluB* is necessary for production of 5,6-dimethylbenzimidazole, the lower ligand of B12. Proc. Natl. Acad. Sci. USA 103: 4634-4639.

Cao, H., Baldini, R.L and Rahme, L.G. (2001). Common mechanisms for pathogens of plants and animals. Annu. Rev. Phytopathol. 39: 259-284.

Capela, D., Filipe, C., Bobik, C., Batut, J., and Bruand, C. (2006). *Sinorhizobium meliloti* differentiation during symbiosis with alfalfa: a transcriptomic dissection. Mol. Plant Microbe Interact. 19: 363-372.

Carro, L., Pujic, P., Alloisio, N., Fournier, N., Boubakri, H., Hay, A.E., Poly, F., François, P., Hocher, V., Mergaert, P., Balmand, S., Rey, M., Heddi, A., and Normand, P. (2015). *Alnus* peptides modify the porosity and induce the release of nitrogen rich metabolites from nitrogen-fixing *Frankia*. ISME J. doi:10.1038/ismej.2014.257

Cebolla, A., Vinardell, J.M., Kiss, E., Olah, B., Roudier, F., Kondorosi, A., and Kondorosi, E. (1999). The mitotic inhibitor *ccs52* is required for endoreduplication and ploidy-dependent cell enlargement in plants. EMBO J. 18: 4476-4484.

Ceja-Navarro, J.A., Nguyen, N.H., Karaoz, U., Gross, S.R., Herman, D.J., Andersen, G.L., Bruns, T.D., Pett-Ridge, J., Blackwell, M., and Brodie, E.L. (2013). Compartmentalized microbial composition, oxygen gradients and nitrogen fixation in the gut of *Odontotaenius disjunctus*. ISME J. 8: 6-18.

Chaintreuil, C., Arrighi, J.F., Giraud, E., Miché, L., Moulin, L., Dreyfus, B., Munive-Hernández, J.A., Villegas-Hernandez M.C., and Béna, G. (2013). Evolution of symbiosis in the legume genus *Aeschynomene*. New Phytol. 200: 1247-1259.

Chandler, M.R., Date, R.A., and Roughley, R.J. (1982). Infection and root-nodule development in *Stylosanthes* species by *Rhizobium*. J. Exp. Bot. 33: 47-57.

Cheng, J., Sibley, C.D., Zaheer, R., and Finan, T.M. (2007). A *Sinorhizobium meliloti minE* mutant has an altered morphology and exhibits defects in legume symbiosis. Microbiol. 153: 375-387.

Chou, M.X., Wei, X.Y., Chen, D.S., and Zhou, J.C. (2006). Thirteen nodule-specific or nodule-enhanced genes encoding products homologous to cysteine cluster proteins or plant lipid transfer proteins are identified in *Astragalus sinicus* L. by suppressive subtractive hybridization. J. Exp. Bot. 57: 2673-2685.

Corbalan, N., Runti, G., Adler, C., Covaceuszach, S., Ford, R.C., Lamba, D., Beis, K., Scocchi, M., and Vincent, P.A. (2013). Functional and structural study of the dimeric inner membrane protein SbmA. J. Bacteriol. 195: 5352-5361.

Corpet, T. (1988). Multiple sequence alignment with hierarchical clustering. Nucl. Acids Res. 16: 10881-10890.

Cosson, V., Durand, P., d'Erfurth, I., Kondorosi, A., and Ratet, P. (2006). *Medicago truncatula* transformation using leaf explants. Methods Mol. Biol. 343: 115-127.

Crockard, A., Bjourson, J., Dazzo, B., and Cooper, J.E. (2002). A white clover nodulin gene, *dd23b*, encoding a cysteine cluster protein, is expressed in roots during the very early stages of interaction with *Rhizobium leguminosarum* biovar *trifolii* and after treatment with chitolipooligosaccharide Nod factors. J. Plant Res. 115: 439-447.

Crouch, M.L., Becker, L.A., Bang, I.S., Tanabe, H., Ouellette, A.J., and Fang, F.C. (2005). The alternative sigma factor σ^E is required for resistance of *Salmonella enterica* serovar Typhimurium to anti-microbial peptides. Mol. Microbiol. 56: 789-799.

Czaja, L.F., Hogekamp, C., Lamm, P., Maillet, F., Martinez, E.A., Samain, E., Dénarié, J., Küster, H., and Hohnjec, N. (2012). Transcriptional responses toward diffusible signals from symbiotic microbes reveal *MtNFP*- and *MtDMI3*-dependent reprogramming of host gene expression by arbuscular mycorrhizal fungal lipochitooligosaccharides. Plant Physiol. 159: 1671-1685.

Czernic, P., Djamel, G., Cartieaux, F., Moulin, L., Guefrachi, I., Patrel, D., Pierre, O., Fardoux J., Chaintreuil, C., Nguyen, P., Gressent, F., Da Silva, C., Poulain, J., Wincker, P., Rofidal, V., Hem, S., Arrighi, J.F., Mergaert, P and Giraud, E. (2015). Convergent evolution of endosymbiont differentiation in Dalbergoid and IRLC legumes mediated by nodule-specific cysteine-rich peptides. Proc. Natl. Acad. Sci. USA Xxx. XX:XX

Damiani, I., Baldacci-Cresp, F., Hopkins, J., Andrio, E., Balzergue, S., Lecomte, P., Puppo, A., Abad, P., Favery, B., and Hérouart, D. (2012). Plant genes involved in harbouring symbiotic rhizobia or pathogenic nematodes. New Phytol. 194: 511-522.

De Coninck, B., Carron, D., Tavormina, P., Willem, L., Craik, D.J., Vos, C., Thevissen, K., Mathys, J., and Cammue, B.P. (2013). Mining the genome of *Arabidopsis thaliana* as a basis for the identification of novel bioactive peptides involved in oxidative stress tolerance. J. Exp. Bot. 64: 5297-5307.

De Coninck, B.M., Sels, J., Venmans, E., Thys, W., Goderis, I.J., Carron, D., Delauré, S.L, Cammue, B.P., De Bolle, M.F., and Mathys, J. (2010). *Arabidopsis thaliana* plant defensin AtPDF1.1 is involved in the plant response to biotic stress. New Phytol. 87: 1075-1088.

Demina, I.V., Persson, T., Santos, P., Plaszczyca, M., and Pawlowski, K. (2013). Comparison of the nodule vs. root transcriptome of the actinorhizal plant *Datisca glomerata*: actinorhizal nodules contain a specific class of defensins. PloS One 8: e72442.

Dennis, J.J., and Zystra, G.J. (1998). Plasposons: modular self-cloning minitransposon derivatives for rapid genetic analysis of Gram-negative bacterial genomes. Appl. Environ. Microbiol. 64: 2710-2715.

Depuydt, M., Messens, J., and Collet, J.F. (2011). How proteins form disulfide bonds. Antioxid. Redox Signal 15: 49-66.

Desbrosses, G.J., and Stougaard, J. (2011). Root Nodulation: A Paradigm for How Plant-Microbe Symbiosis Influences Host Developmental Pathways. Cell Host Microbe 10: 348-358.

Dobson, S.L., Bourtzis, K., Braig, H.R, Jones, B.F., and Zhou, W. (1999). Wolbachia infections are distributed throughout insect somatic and germ line tissues. Insect Biochem. Mol. Biol. 29: 153-160.

Domenech, P., Kobayashi, H., LeVier, K., Walker, G.C., and Barry, C.E 3rd. (2009). BacA, an ABC transporter involved in maintenance of chronic murine infections with *Mycobacterium tuberculosis*. J. Bacteriol. 191: 477-485.

Dommergues, Y., Duhaux, E., and Hoang, G.D. (1999). Les arbres fixateurs d'azote: Caractéristiques fondamentales et rôle dans l'aménagement des écosystèmes méditérranéens et tropicaux. Édition espaces 34, Paris. 475p.

Douglas, A.E, and Dixon, A.F.G. (1987). The mycetocyte symbiosis of aphids: Variation with age and morph in virginoparae of Megoura viciae and Acyrthosiphon pisum. J. Insect Physiol. 33: 109-113.

Downie, A. (2014). Legume nodulation. Curr. Biol. 24: R184-R190.

Dziva, F., and Stevens, M.P. (2008). Colibacillosis in poultry: unravelling the molecular basis of virulence of avian pathogenic *Escherichia coli* in their natural hosts. Avian Pathol. 37: 355-366.

Ehrhardt, D.W., Atkinson, E.M., and Long, S.R. (1992). Depolarization of alfalfa root hair membrane potential by *Rhizobium meliloti* nod factors. Science 256: 998-1000.

El Yahyaoui, F., Kuster, H., Ben Amor, B., Hohnjec, N., Puhler, A., Becker, A., Gouzy, J., Vernie, T., Gough, C., Niebel, A., Godiard, L., and Gamas, P. (2004). Expression profiling in Medicago truncatula identifies more than 750 genes differentially expressed during nodulation, including many potential regulators of the symbiotic program. Plant Physiol. 136: 3159-3176.

Eswarappa, S., Panguluri, K.K., Hensel, M., and Chakravortty, D. (2008). The *yejABEF* operon of *Salmonella* confers resistance to antimicrobial peptides and contributes to its virulence. Microbiol. 154: 666-678.

Ewald, P.W. (1987). Transmission modes and evolution of the parasitism-mutualism continuum. Ann. N.Y. Acad. Sci. 503: 295-306.

Falord, M., Karimova, G., Hiron, A., and Msadek, T. (2012). GraXSR proteins interact with the VraFG ABC transporter to form a five-component system required for cationic antimicrobial peptide sensing and resistance in *Staphylococcus aureus*. Antimicrob. Agents Chemother. 56: 1047-1058.

Fang, H., Mullins, C., and Green, N. (1997). In addition to *SEC11*, a newly identified gene, *SPC3*, is essential for signal peptidase activity in the yeast endoplasmic reticulum. J. Biol. Chem. 272: 13152-13158.

Farkas, A., Maróti, G., Dürgő, H., Györgypál, Z., Lima, R.M., Medzihradszky, K.F., Kereszt, A., Mergaert, P., and Kondorosi, E. (2014). The *Medicago truncatula* symbiotic peptide NCR247 contributes to bacteroid differentiation through multiple mechanisms. Proc. Natl. Acad. Sci. USA 111: 5183-5188.

Farris, C., Sanowar, S., Bader, M.W., Pfeutzner, R., and Miller, S.I. (2010). Antimicrobial peptides activate the Rcs regulon through the outer membrane lipoprotein RcsF. J. Bacteriol. 192: 4894-4903

Favery, B., Complainville, A., Vinardell, J.M., Lecomte, P., Vaubert, D., Mergaert, P., Kondorosi, A., Kondorosi, E., Crespi, M., and Abad, P. (2002). The endosymbiosis-induced genes *ENOD40* and *CCS52a* are involved in endoparasitic-nematode interactions in *Medicago truncatula*. Mol. Plant Microbe Interact. 15: 1008-1013.

Ferguson, G.P, Roop, R.M 2nd., and Walker, G.C. (2002). Deficiency of a *Sinorhizobium meliloti* BacA mutant in alfalfa symbiosis correlates with alteration of the cell envelope. J. Bacteriol. 184: 5625-5632.

Ferguson, G.P., Datta, A., Baumgartner, J., Roop, R.M 2nd., Carlson, R.W., Walker, G.C. (2004). Similarity to peroxisomal-membrane protein family reveals that *Sinorhizobium* and *Brucella* BacA affect lipid-A fatty acids. Proc. Natl. Acad. Sci. USA 101: 5012-5017.

Ferguson, G.P., Datta, A., Carlson, R.W., and Walker, G.C. (2005). Importance of unusually modified lipid A in *Sinorhizobium* stress resistance and legume symbiosis. Mol. Microbiol. 56: 68-80.

Fields, S. (2004). Global nitrogen : Cycling out of control. Environ. Health Perspect. 112: A556-A563.

Finan, T.M., Kunkel, B., De Vos, G.F., and Signer, E.R. (1986). Second symbiotic megaplasmid in *Rhizobium meliloti* carrying exopolysaccharide and thiamine synthesis genes. J. Bacteriol. 167: 66-72.

Foth, H.D. (1990). Fundamentals of soil science. Wiley, J., and Sons, New York. 336p.

Franche, C., Lindström, K., and Elmerich, C. (2008). Nitrogen-fixing bacteria associated with leguminous and non-leguminous plants. Plant and Soil 321: 35-59.

Fraune, S., Augustin, R., Anton-Erxleben, F., Wittlieb, J., Gelhaus, C., Klimovich, V.B., Samoilovich, M.P., and Bosch, T.C.G. (2010). In an early branching metazoan, bacterial colonization of the embryo is controlled by maternal antimicrobial peptides. Proc. Natl. Acad. Sci. USA 107: 18067-18072.

Frühling, M., Hohnjec, N., Schröder, G., Küster, H., Pühler, A., and Perlick, A.M. (2000). Genomic organization and expression properties of the *VfENOD5* gene from broad bean (*Vicia faba* L.). Plant Sci. Lett. 155: 169-178.

Fukatsu, T., Nikoh, N., Kawai, R., and Koga, R. (2000). The secondary endosymbiotic bacterium of the pea aphid Acyrthosiphon pisum (Insecta : Homoptera). Appl. Environ. Microbiol. 66: 2748-2758.

Fülöp, K., Tarayre, S., Kelemen, Z., Horváth, G., Kevei, Z., Nikovics, K., Bakó, L., Brown, S., Kondorosi, A., and Kondorosi, E. (2005). Arabidopsis anaphase-promoting complexes: multiple activators and wide range of substrates might keep APC perpetually busy. Cell Cycle 4: 1084-1092.

Futahashi, R., Tanaka, K., Tanahashi, M., Nikoh, N., Kikuchi, Y., Lee, B.L., and Fukatsu, T. (2013). Gene expression in gut symbiotic organ of stinkbug affected by extracellular bacterial symbiont. PLoS One 8: e64557.

Ganz, T. (2003). Defensins: antimicrobial peptides of innate immunity. Nat. Rev. Immunol. 3: 710-720.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In: The Proteomics Protocols Handbook. Walker, J.M. (ed). Humana Press, Totowa, New Jersey. pp 571-607.

Gaude, N., Bortfeld, S., Duensing, N., Lohse, M., and Krajinski, F. (2012). Arbuscule-containing and non-colonized cortical cells of mycorrhizal roots undergo extensive and specific reprogramming during arbuscular mycorrhizal development. Plant J. 69: 510-528.

Ghosal, A., Vitali, A., Stach, J.E., and Nielsen, P.E. (2012). Role of SbmA in the uptake of peptide nucleic acid (PNA)-peptide conjugates in *E. coli*. ACS Chem. Biol. 8: 360-367.

Gibson, K.E., Kobayashi, H., and Walker, G.C. (2008). Molecular determinants of a symbiotic chronic infection. Annu. Rev. Genet. 42: 413-441.

Giraud, E., and Fleischman, D. (2004). Nitrogen-fixing symbiosis between photosynthetic bacteria and legumes. Photosynth. Res. 82: 115-130.

Giraud, E., Hannibal, L., Fardoux, J., Vermeglio, A., and Dreyfus, B. (2000). Effect of *Bradyrhizobium* photosynthesis on stem nodulation of *Aeschynomene sensitiva*. Proc. Natl. Acad. Sci. USA 97: 14795-14800.

Giraud, E., Moulin, L., Vallenet, D., Barbe, V., Cytryn, E., Avarre, J.C., Jaubert, M., Simon, D., Cartieaux, F., Prin, Y., Bena, G., Hannibal, L., Fardoux, J., Kojadinovic, M., Vuillet, L., Lajus, A., Cruveiller, S., Rouy, Z., Mangenot, S., Segurens, B., Dossat, C., Franck, W.L., Chang, W.S., Saunders, E., Bruce, D., Richardson, P., Normand, P., Dreyfus, B., Pignol, D., Stacey, G., Emerich, D., Verméglio, A., Médigue, C., and Sadowsky, M. (2007). Legumes symbioses: Absence of *Nod* genes in photosynthetic Bradyrhizobia. Science 316: 1307-1312.

Glazebrook, J., Ichige, A., and Walker, G.C. (1993). A *Rhizobium meliloti* homolog of the *Escherichia coli* peptide-antibiotic transport protein SbmA is essential for bacteroid development. Genes Dev. 7: 1485-1497.

Glazebrook, J., Ichige, A., and Walker, G.C. (1996). Genetic analysis of *Rhizobium meliloti bacA-phoA* fusion results in identification of *degP*: Two loci required for symbiosis are closely linked to *degP*. J. Bacteriol. 178: 745-752.

González-Sama, A., Coba de la Peña, T., Kevel, Z., Mergaert, P., Lucas, M., De Felipe, M.R., Kondorosi, E., and Pueyo, J.J. (2006). Nuclear DNA endoreduplication and expression of the mitotic inhibitor Ccs52 associated to determinate and lupinoid nodule organogenesis. Mol. Plant Microbe Interact. 19: 173-180.

Gopinath, K., Venclovas, C., Ioerger, T.R., Sacchettini, J.C., McKinney, J.D., Mizrahi, V., and Warner, D.F. (2013). A vitamin B_{12} transporter in *Mycobacterium tuberculosis*. Open Biol. 3: 120175.

Gourion, B., Sulser, S., Frunzke, J., Francez-Charlot, A., Stiefel, P., Pessi, G., Vorholt, J.A., and Fischer, H.M. (2009). The PhyR-σ^{EcfG} signaling cascade is involved in stress response and symbiotic efficiency in *Bradyrhizobium japonicum*. Mol. Microbiol. 73: 291-305.

Graham, M.A., Silverstein, K.A., Cannon, S.B., and VandenBosch, K.A. (2004). Computational identification and characterization of novel genes from legumes. Plant Physiol. 135: 1179-1197.

Gruber, N., and Galloway, J.N. (2008). An Earth-system perspective of the global nitrogen cycle. Nature 451: 293-296.

Gryllos, I., Tran-Winkler, H.J., Cheng, M.F., Chung, H., Bolcome 3rd, R., Lu, W., Lehrer, R.I., and Wessels, M.R. (2008). Induction of group A *Streptococcus* virulence by a human antimicrobial peptide. Proc. Natl. Acad. Sci. USA 105: 16755-16760.

Guefrachi, I., Nagymihaly, M., Pislariu, C.I., Van de Velde, W., Ratet, P., Mars, M., Udvardi, M.K., Kondorosi, E., Mergaert, P., and Alunni, B. (2014). Extreme specificity of NCR gene expression in *Medicago truncatula*. BMC Genomics 15: 712.

Guefrachi, I., Pierre, O., Bourge, M., Timchenko, T., Alunni, B., Czernic, P., Villaécija-Aguilar, J.A., Verly, C., Fardoux, J., Mars, M., Kondorosi, E., Giraud, E., and Mergaert, P. (2015a). Bradyrhizobium BclA is a NCR peptide transporter required for bacterial differentiation in symbiosis with *Aeschynomene*

Haag, A.F., Arnold, M.F., Myka, K.K., Kerscher, B., Dall'angelo, S., Zanda, M., Mergaert, P., and Ferguson, G.P. (2013). Molecular insights into bacteroid development during Rhizobium-legume symbiosis. FEMS Microbiol. Rev. 37: 364-383.

Haag, A.F., Baloban, M., Sani, M., Kerscher, B., Pierre, O., Farkas, A., Longhi, R., Boncompagni, E., Hérouart, D., Dall'angelo, S., Kondorosi, E., Zanda, M., Mergaert, P., and Ferguson, G.P. (2011). Protection of *Sinorhizobium* against host cysteine-rich antimicrobial peptides is critical for symbiosis. PLoS Biol. 9: e1001169.

Haag, A.F., Kerscher, B., Dall'angelo, S., Sani, M., Longhi, R., Baloban, M., Wilson, H.M., Mergaert, P., Zanda, M., and Ferguson, G.P. (2012). Role of cysteine residues and disulfide bonds on the activity of a legume root nodule-specific, cysteine-rich peptide. J. Biol. Chem. 287: 10791-10798.

Haag, A.F., Wehmeier, S., Beck, S., Marlow, V.L., Fletcher, V., James, E.K., and Ferguson, G.P. (2009). The *Sinorhizobium meliloti* LpxXL and AcpXL proteins play important roles in bacteroid development within alfalfa. J. Bacteriol. 191: 4681-4686.

Hakoyama, T., Niimi, K., Watanabe, H., Tabata, R., Matsubara, J., Sato, S., Nakamura, Y., Tabata, S., Jichun, L., Matsumoto, T., Tatsumi, K., Nomura, M., Tajima, S., Ishizaka, M., Yano, K., Imaizumi-Anraku, H., Kawaguchi, M., Kouchi, H., and Suganuma, N. (2009). Host plant genome overcomes the lack of a bacterial gene for symbiotic nitrogen fixation. Nature 462: 514-517.

Hann, D.R and Rathjen, J.P. (2007). Early events in the pathogenicity of *Pseudomonas syringae* on *Nicotiana benthamiana*. Plant J. 49: 607-618.

Herbert, E.E, and Blair, G.H. (2007). Friend and foe: the two faces of Xenorhabdus nematophila. Nat. Rev. Microbiol. 5: 634-646.

Hiron, A., Falord, M., Valle, J., Débarbouillé, M., and Msadek, T. (2011). Bacitracin and nisin resistance in *Staphylococcus aureus*: a novel pathway involving the BraS/BraR two-component system (SA2417/SA2418) and both the BraD/BraE and VraD/VraE ABC transporters. Mol. Microbiol. 81: 602-622.

Hocher, V., Alloisio, N., Auguy, F., Fournier, P., Doumas, P., Pujic, P., Gherbi, H., Queiroux, C., Da Silva, C., Wincker, P., Normand, P., and Bogusz, D. (2011). Tanscriptomics of Actinorhizal Symbiosis Reveals Homologs of the Whole Common Symbiotic Signaling Cascade. Plant Physiol. 156: 700-711.

Hogekamp, C., Arndt, D., Pereira, P.A., Becker, J.D., Hohnjec, N., and Küster, H. (2011). Laser microdissection unravels cell-type-specific transcription in arbuscular mycorrhizal roots, including CAAT-box transcription factor gene expression correlating with fungal contact and spread. Plant Physiol. 157: 2023-2043.

Horvath, B., Yeun, L.H., Domonkos, A., Halasz, G., Gobbato, E., Ayaydin, F., Miro, K., Hirsch, S., Sun, J., Tadege, M., Ratet, P., Mysore, K., Ané, J.M., Oldroyd, G.E.D., and Kalo, P. (2011). *Medicago truncatula* IPD3 is a member of the common symbiotic signaling pathway required for rhizobial and mycorrhizal symbioses. Mol. Plant Microbe Interact. 24: 1345-1358.

http://sourceforge.net/projects/mev-tm4/

http://compbio.dfci.harvard.edu/tgi/

https://iant.toulouse.inra.fr/symbimics

http://multalin.toulouse.inra.fr/multalin

http://mtgea.noble.org/v3/

http://www.biomedcentral.com/content/supplementary/1471-2164-15-712-S1.xlsx

http://www.biomedcentral.com/content/supplementary/1471-2164-15-712-S3.xlsx

http://www.nature.com/scitable/knowledge/library/the-nitrogen-cycle-processes-players-and-human-15644632.

Humphreys, S., Stevenson, A., Bacon, A., Weinhardt, B.A., and Roberts, M. (1999). The alternative sigma factor, σ^E, is critically important for the virulence of *Salmonella typhimurium*. Infect. Immun. 67: 1560-1580.

Ibáñez, F., and Fabra, A. (2011). Rhizobial Nod factors are required for cortical cell division in the nodule morphogenetic programme of the *Aeschynomeneae* legume *Arachis*. Plant Biol. 13: 794-800.

Ichige, A., and Walker, G.C. (1997). Genetic analysis of the *Rhizobium meliloti bacA* gene: functional interchangeability with the *Escherichia coli sbmA* gene and phenotypes of mutants. J. Bacteriol. 179: 209-216.

Imin, N., Goffard, N., Nizamidin, M., and Rolfe, B.G. (2008). Genome-wide transcriptional analysis of super-embryogenic *Medicago truncatula* explant cultures. BMC Plant Biol. 8: 110.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4: 249-264.

Javelle, M., Marco, C.F., and Timmermans, M. (2011). *In situ* hybridization for the precise localization of transcripts in plants. J. Vis. Exp. 57: e3328.

Jayaraman, D., Valdés-López, O., Kaspar, C.W., and Ané, J.M. (2014). Response of *Medicago truncatula* seedlings to colonization by *Salmonella enterica* and *Escherichia coli* O157:H7. PLoS One 9: e87970.

Jimenez-Zurdo, J.I., Frugier, F., Crespi, M.D., and Kondorosi, A. (2000). Expression profiles of 22 novel molecular markers for organogenetic pathways acting in alfalfa nodule development. Mol. Plant Microbe Interact. 13: 96-106.

Kaijalainen, S., Schroda, M., and Lindström, K. (2002). Cloning of nodule-specific cDNAs of *Galega orientalis*. Physiol. Plant. 114: 588-593.

Kang, D., Liu, G., Lundstrom, A., Gelius, E., and Steiner, H. (1998). A peptidoglycan recognition protein in innate immunity conserved from insects to humans. Proc. Natl. Acad. Sci. USA 95: 10078-10082.

Kannenberg, E.L., and Carlson, R.W. (2001). Lipid A and O-chain modifications cause *Rhizobium* lipopolysaccharides to become hydrophobic during bacteroid development. Mol. Microbiol. 39: 379-391.

Kannenberg, E.L., Perotto, S., Bianciotto, V., Rathbun, E.A., and Brewin, N.J. (1994). Lipopolysaccharide epitope expression of *Rhizobium* bacteroids as revealed by in situ immunolabelling of pea root nodule sections. J. Bacteriol. 176: 2021-2032.

Kannenberg, E.L., Perzl, M., and Härtner, T. (1995). The occurrence of hopanoid lipids in *Bradyrhizobium* bacteria. FEMS Microbiol. Let. 127: 255-262.

Karimi, M., De Meyer, B., and Hilson, P. (2005). Modular cloning in plant cells. Trends Plant Sci. 10: 103-105.

Karunakaran, R., Haag, A.F., East, A.K., Ramachandran, V.K., Prell, J., James, E.K., Scocchi, M., Ferguson, G.P., and Poole, P.S. (2010). BacA is essential for bacteroid development in nodules of galegoid, but not phaseoloid, legumes. J. Bacteriol. 192: 2920-2928.

Karunakaran, R., Ramachandran, V.K., Seaman, J.C., East, A.K., Mouhsine, B., Mauchline, T.H., Prell, J., Skeffington, A., and Poole, P.S. (2009). Transcriptomic analysis of *Rhizobium leguminosarum* biovar *viciae* in symbiosis with host plants *Pisum sativum* and *Vicia cracca*. J. Bacteriol. 191: 4002-4014.

Kato, M.G., and Yamamura, N. (1999). Evolution of mutualistic symbiosis without vertical transmission. Theor. Popul. Biol. 55: 309-323.

Kato, T., Kawashima, K., Miwa, M., Mimura, Y., Tamaoki, M., Kouchi, H., and Suganuma, N. (2002). Expression of genes encoding late nodulins characterized by a putative signal peptide and conserved cysteine residues is reduced in ineffective pea nodules. Mol. Plant Microbe Interact. 15: 129-137.

Kereszt, A., Mergaert, P., and Kondorosi, E. (2011). Bacteroid development in legume nodules: evolution of mutual benefit or of sacrificial victims? Mol. Plant Microbe Interact. 24: 1300-1900.

Kevei, Z., Lougnon, G., Mergaert, P., Horváth, G.V., Kereszt, A., Jayaraman, D., Zaman, N., Marcel, F., Regulski, K., Kiss, G.B., Kondorosi, A., Endre, E., Kondorosi, E and Ané, J.M. (2007). A 3-Hydroxy-3-Methylglutaryl Coenzyme A reductase interacts with NORK in the nodulation signaling pathway. Plant Cell 19: 3974-3989.

Kevei, Z., Vinardell, J.M., Kiss, G.B., Kondorosi, A., and Kondorosi, E. (2002). Glycine-rich proteins encoded by a nodule-specific gene family are implicated in different stages of symbiotic nodule development in *Medicago* spp. Mol. Plant Microbe Interact. 15: 922-931.

Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R., and Bosch, T.C.G. (2009). More than just orphans: are taxonomically-restricted genes important in evolution? Trends Genet. 25: 404-413.

Kikuchi, Y. (2009). Endosymbiotic bacteria in insects: their diversity and culturability. Microbes Environ. 24: 195-204.

Kikuchi, Y., Hosokawa, T., and Fukatsu, T. (2011). Specific developmental window for establishment of an insect-microbe gut symbiosis. Appl. Environ. Microbiol. 77: 4075-4081.

Koch B., and Evans H.J. (1966). Reduction of acetylene to ethylene by soybean root nodules. Plant Physiol. 41: 1748-1750.

Koga, R., Meng, X.Y., Tsuchida, T., and Fukatsu, T. (2012). Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte-embryo interface. Proc. Natl. Acad. Sci. USA 109: E1230-E1237.

Komaki, K., and Ishikawa, H. (2000). Genomic copy number of intracellular bacterial symbionts of aphids varies in response to developmental stage and morph of their host. Insect Biochem. Mol. Biol. 30: 253-258.

Komaki, K., and Ishikawa, H. (1999). Intracellular bacterial symbionts of aphids possess many genomic copies per bacterium. J. Mol. Evol. 48: 717-722.

Kondorosi, E., and Kondorosi, A. (2004). Endoreduplication and activation of the anaphase-promoting complex during symbiotic cell development. FEBS Lett. 567: 152-157.

Kondorosi, E., Mergaert, P., and Kereszt, A. (2013). A paradigm for endosymbiotic life: Cell differentiation of *Rhizobium* bacteria provoked by host plant factors. Annu. Rev. Microbiol. 67: 611-628.

Kondorosi, E., Roudier, F., and Gendreau, E. (2000). Plant cell-size control: growing by ploidy? Curr. Opin. Plant Biol. 3: 488-492.

Küster, H., Schröder, G., Frühling, M., Pich, U., Rieping, M., Schubert, I., Perlick, A.M., and Pühler, A. (1995). The nodule-specific *VfENOD-GRP3* gene encoding a glycine-rich early nodulin is

located on chromosome I of *Vicia faba* L. and is predominantly expressed in the interzone II-III of root nodules. Plant Mol. Biol. 28: 405-421.

Lam, H., Oh, D.C, Cava, F., Takacs, C.N., Clardy, J., De Pedro, M.A., and Waldor, M.K. (2009). D-amino acids govern stationary phase cell wall remodeling in bacteria. Science 325: 1552-1555.

Laporte, P., Satiat-Jeunemaître, B., Velasco, I., Csorba, T., Van de Velde, W., Campalans, A., Burgyan, J., Arevalo-Rodriguez, M., and Crespi, M. (2010). A novel RNA-binding peptide regulates the establishment of the *Medicago truncatula–Sinorhizobium meliloti* nitrogen-fixing symbiosis. Plant J. 62: 24-38.

Latch, J.N., and Margolin, W. (1997). Generation of buds, swellings, and branches instead of filaments after blocking the cell cycle of *Rhizobium meliloti*. J. Bacteriol. 179: 2373-2381.

Laviña, M., Pugsley, A.P., and Moreno, F. (1986). Identification, mapping, cloning and characterization of a gene (*sbmA*) required for microcin B17 action on *Escherichia coli* K12. J. Gen. Microbiol. 132: 1685-1693.

Lelandais-Brière, C., Naya, L., Sallet, E., Calenge, F., Frugier, F., Hartmann, C., Gouzy, J., and Crespi, M. (2009). Genome-wide *Medicago truncatula* small RNA analysis revealed novel microRNAs and isoforms differentially regulated in roots and nodules. Plant Cell 21: 2780-2796.

LeVier, K., and Walker, G.C. (2001). Genetic analysis of the *Sinorhizobium meliloti* BacA protein: differential effects of mutations on phenotypes. J. Bacteriol. 183: 6444-6453.

LeVier, K., Phillips, R.W., Grippe, V.K., Roop, R.M 2nd., and Walker, G.C. (2000). Similar requirements of a plant symbiont and a mammalian pathogen for prolonged intracellular survival. Science 287: 2492-2493.

Li, G., Laturnus, C., Ewers, C., and Wieler, L.H. (2005). Identification of genes required for avian *Escherichia coli* septicemia by signature-tagged mutagenesis. Infect. Immun. 73: 2818-2827.

Li, D., Su, Z., Dong, J., and Wang, T. (2009). An expression database for roots of the model legume *Medicago truncatula* under salt stress. BMC Genomics 10: 517.

Limpens, E., and Bisseling, T. (2014). CYCLOPS: a new vision on Rhizobium-induced nodule organogenesis. Cell Host Microbe 15: 127-129.

Limpens, E., Moling, S., Hooiveld, G., Pereira, P.A., Bisseling, T., Becker, J.D., and Küster, H. (2013). Cell- and tissue-specific transcriptome analyses of *Medicago truncatula* root nodules. PLoS One 8: e64377.

Liu, J., Miller, S.S., Graham, M., Bucciarelli, B., Catalano, C.M., Sherrier, D.J, Samac, D.A, Ivashuta, S., Fedorova, M., Matsumoto, P., Gantt, J.S., and Vance, C.P. (2006). Recruitment of novel calcium-binding proteins for root nodule symbiosis in *Medicago truncatula*. Plant Physiol. 141: 167-177.

Login, F.H., Balmand, S., Vallier, A., Vincent-Monégat, C., Vigneron, A., Weiss-Gayet, M., Rochat, D., and Heddi, A. (2011). Antimicrobial peptides keep insect endosymbionts under control. Science 334: 362-365.

Lohar, D.P., Sharopova, N., Endre, G., Penuela, S., Samac, D., Town, C., Silverstein, K.A., and VandenBosch, K.A. (2006). Transcript analysis of early nodulation events in *Medicago truncatula*. Plant Physiol. 140: 221-234.

Lopez-Solanilla, E., García-Olmedo, F., and Rodríguez-Palenzuela, P. (1998). Inactivation of the *sapA* to *sapF* locus of *Erwinia chrysanthemi* reveals common features in plant and animal bacterial pathogenesis. Plant Cell 10: 917-924.

Mah, K.M., Uppalapati, S.R., Tang, Y.H., Allen, S., and Shuai, B. (2012). Gene expression profiling of *Macrophomina phaseolina* infected *Medicago truncatula* roots reveals a role for auxin in plant tolerance against the charcoal rot pathogen. Physiol. Mol. Plant Pathol. 79: 21-30.

Majdalani, N., and Gottesman, S. (2005). The Rcs phosphorelay: A complex signal transduction system. Annu. Rev. Microbiol. 59: 379-405.

Malott, R.J., Steen-Kinnaird, B.R., Lee, T.D. and Speert, D.P. (2012). Identification of hopanoid biosynthesis genes involved in polymyxin resistance in Burkholderia multivorans. Antimicrob. Agents Chemother. 56: 464-471.

Manthey, K., Krajinski, F., Hohnjec, N., Firnhaber, C., Puhler, A., Perlick, A.M., and Kuster, H. (2004). Transcriptome profiling in root nodules and arbuscular mycorrhiza identifies a collection of novel genes induced during *Medicago truncatula* root endosymbioses. Mol. Plant Microbe Interact. 17: 1063-1077.

Maillet, F., Poinsot, V., André, O., Puech-Pagès, V., Haouy, A., Gueunier, M., Crome, L., Giraudet, D., Formey, D., Niebel, A., Martinez, E.A., Driguez, H., Bécard, G., and Dénarié, J. (2011). Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. Nature 469: 58-63.

Marchetti, M., Capela, D., Poincloux, R., Benmeradi, N., Auriac, M.C., Le Ru, A., Maridonneau-Parini, I., Batut, J., and Masson-Boivin, C. (2013). Queuosine biosynthesis is required for *Sinorhizobium meliloti*-induced cytoskeletal modifications on HeLa Cells and symbiosis with *Medicago truncatula*. PLoS One 8: e56043.

Marchler-Bauer, A., Zheng, C., Chitsaz, F., Derbyshire, M.K., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., Lu, F., Lu, S., Marchler, G.H., Song, J.S., Thanki, N., Yamashita, R.A., Zhang, D., and Bryant, S.H. (2013). CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Res. 41: D348-D352.

Mardirossian, M., Grzela, R., Giglione, C., Meinnel, T., Gennaro, R., Mergaert, P., and Scocchi, M. (2014). The host antimicrobial peptide $Bac7_{1-35}$ binds to bacterial ribosomal proteins and inhibits protein synthesis. Chem Biol. 21: 1639-1647.

Marlow, V.L., Haag, A.F., Kobayashi, H., Fletcher, V., Scocchi, M., Walker, G.C., and Ferguson, G.P. (2009). Essential role for the BacA protein in the uptake of a truncated eukaryotic peptide in *Sinorhizobium meliloti*. J. Bacteriol. 191: 1519-1527.

Maróti, G and Kondorosi, E. (2014). Nitrogen-fixing Rhizobium-legume symbiosis: are polyploidy and host peptide-governed symbiont differentiation general principles of endosymbiosis? Front. Microbiol. 5: 326.

Maróti, G., Kereszt, A., Kondorosi, E., and Mergaert, P. (2011). Natural roles of antimicrobial peptides in microbes, plants and animals. Res. Microbiol. 162: 363-374.

Martínez-Salazar, L.M., Salazar, E., Encarnación, S., Ramírez-Romero, M.A., and Rivera, J. (2009). Role of the extracytoplasmic function sigma factor RpoE4 in oxidative and osmotic stress responses in *Rhizobium etli*. J. Bacteriol. 191: 4122-4132.

Maruya, J., and Saeki, K. (2010). The *bacA* Gene Homolog, mlr7400, in *Mesorhizobium loti* MAFF303099 is dispensable for symbiosis with *Lotus japonicus* but partially capable of supporting the symbiotic function of *bacA* in *Sinorhizobium meliloti*. Plant Cell Physiol. 51: 1443-1452.

Masson-Boivin, C., Giraud, E., Perret, X., and Batut, J. (2009). Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? Trends in Microbiol. 17: 458-466.

Mattiuzzo, M., Bandiera, A., Gennaro, R., Benincasa, M., Pacor, S., Antcheva, N., and Scocchi, M. (2007). Role of the *Escherichia coli SbmA* in the antimicrobial activity of proline-rich peptides. Mol. Microbiol. 66: 151-163.

Maunoury, N., Redondo-Nieto, M., Bourcy, M., Van de Velde, W., Alunni, B., Laporte, P., Durand, P., Agier, N., Marisa, L., Vaubert, D., Delacroix, H., Duc, G., Ratet, P., Aggerbeck, L., Kondorosi, E., and Mergaert, P. (2010). Differentiation of symbiotic cells and endosymbionts in *Medicago truncatula* nodulation are coupled to two transcriptome-switches. PLoS One 5: e9519.

Mergaert, P., Nikovics, K., Kelemen, Z., Maunoury, N., Vaubert, D., Kondorosi, A., and Kondorosi, E. (2003). A novel family in *Medicago truncatula* consisting of more than 300 nodule-specific genes coding for small, secreted polypeptides with conserved cysteine motifs. Plant Physiol. 132: 161-173.

Mergaert, P., Uchiumi, T., Alunni, B., Evanno, G., Cheron, A., Catrice, O., Mausset, A.E., Barloy-Hubler, F., Galibert, F., Kondorosi, A., and Kondorosi, E. (2006). Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*–legume symbiosis. Proc. Natl. Acad. Sci. USA 103: 5230-5235.

Mergaert, P., Van Montagu, M., and Holsters, M. (1997). Molecular mechanisms of Nod factor diversity. Mol. Microbiol. 25: 811-817.

Messinese, E., Mun, J.H., Yeun, L.H., Jayaraman, D., Rougé, P., Barre, A., Lougnon, G., Schornack, S., Bono, J.J., Cook, D.R., and Ané, J.M. (2007). A novel nuclear protein interacts with the symbiotic DMI3 calcium- and calmodulin-dependent protein kinase of *Medicago truncatula*. Mol. Plant Microbe Interact. 20: 912-921.

Meyer, H.A., and Hartmann, E. (1997). The yeast SPC22/23 homolog Spc3p is essential for signal peptidase activity. J. Biol. Chem. 272: 13159-13164.

Molouba, F., Lorquin, J., Willems, A., Hoste, B., and Giraud E. (1999). Photosynthetic bradyrhizobia from *Aeschynomene* spp. are specific to stem-nodulated species and form a separate 16S ribosomal DNA restriction fragment length polymorphism group. Appl. Environ. Microbiol. 65: 3084-3094.

Mondy, S., Lenglet, A., Cosson, V., Pelletier, S., Pateyron, S., Gilard, F., Scholte, M., Brocard L., Couzigou, J.M., Tcherkez, G., Péan, M., and Ratet, P. (2013). GOLLUM [FeFe]-hydrogenase-like proteins are essential for plant development in normoxic conditions and modulate energy metabolism. Plant Cell Env. 37: 54-69.

Moran, N.A., McCutcheon, J.P., and Nakabachi, A. (2008). Genomics and evolution of heritable bacterial symbionts. Annu. Rev. Genet. 42: 165-190.

Moreau, S., Verdenaud, M., Ott, T., Letort, S., De Billy, F., Niebel, A., Gouzy, J., De Carvalho-Niebel, F., and Gamas, P. (2011). Transcrioptional reprogramming during root nodule development in *Medicago truncatula*. PLoS One 6: e16463.

Müller, J., Wiemken, A., and Boller, T. (2001). Redifferentiation of bacteria isolated from Lotus japonicus root nodules colonized by *Rhizobium* sp. NGR234. J. Exp. Bot. 52: 2181-2186.

Mulley, G., White, J.P., Karunakaran, R., Prell, J., Bourdes, A., Bunnewell, S., Hill, L., and Poole, P.S. (2011). Mutation of GOGAT prevents pea bacteroid formation and N_2 fixation by globally downregulating transport of organic nitrogen sources. Mol. Microbiol. 80: 149-167.

Nallu, S., Silverstein, K.A., Zhou, P., Young, N.D., and Vandenbosch, K.A. (2014). Patterns of divergence of a large family of nodule cysteine-rich peptides in accessions of Medicago truncatula. Plant J. 78: 697-705.

Naoumkina, M., Vaghchhipawala, S., Tang, Y., Ben, Y., Powell, R.J., and Dixon, R.A. (2007). Different mechanisms for phytoalexin induction by pathogen and wound signals in *Medicago truncatula*. Proc. Natl. Acad. Sci. USA 104: 17909-17915.

Narayanan, S., Modaka, J.K., Ryana, C.S., Bustosa, G.J., Daviesa, J.K., and Roujeinikovaa, A. (2014). Mechanism of *Escherichia coli* Resistance to Pyrrhocoricin. Antimicrob. Agents Chemother. 58: 2754-2762.

Nick, G., De Lajudie, P., Eardly, B.D., Suomalainen, S., Paulin, L., Zhang, X., Gillis, M., and Lindström, K. (1999). *Sinorhizobium arboris* sp. nov. and *Sinorhizobium kostiense* sp. nov., isolated from leguminous trees in Sudan and Kenya. Int. J. Syst. Evol. Microbiol. 49: 1359-1368.

Nouioui, I., Ghodhbane-Gtari, F., Beauchemin, N.J., Tisa, L.S., and Gtari, M. (2011). Phylogeny of members of the Frankia genus based on gyrB, nifH and glnII sequences. Antonie Van Leeuwenhoek 100: 579-587.

Núñez, M.F., Pellicer, M.T., Badía, J., Aguilar, J., and Baldomà, L. (2001). The gene *yghK* linked to the *glc* operon of *Escherichia coli* encodes a permease for glycolate that is structurally and functionally similar to L-lactate permease. Microbiol. 147: 1069-1077.

Nyholm, S.V., and McFall-Ngai, M.J. (2004). The winnowing: establishing the squid-vibrio symbiosis Nat. Rev. Microbiol. 2: 632-642.

Oke, V., and Long, S.R. (1999). Bacterial genes induced within the nodule during the Rhizobium-legume symbiosis. Mol. Microbiol. 32: 837-849.

Oldroyd, G.E., and Downie, J.A. (2008). Coordinating nodule morphogenesis with rhizobial infection in legumes. Annu. Rev. Plant Biol. 59: 519-546.

Oldroyd, G.E., Murray, J.D., Poole, P.S., and Downie, J.A. (2011). The rules of engagement in the legume-rhizobial symbiosis. Annu. Rev. Genet. 45: 119-144.

Oliver, K.M., Degnan, P.H., Burke, G.R., and Moran, N.A. (2010). Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. Annu. Rev. Entomol. 55: 247-266.

Oono, R., and Denison, R.F. (2010). Comparing symbiotic efficiency between swollen versus non swollen rhizobial bacteroids. Plant Physiol. 154: 1541-1548.

Oono, R., Schmitt, I., Sprent, J.I., and Denison, R.F. (2010). Multiple evolutionary origins of legume traits leading to extreme rhizobial differentiation. New Phytol. 187: 508-520.

Ordögh, L., Vörös, A., Nagy, I., Kondorosi, E., and Kereszt, A. (2014). Symbiotic Plant Peptides Eliminate Candida albicans Both In Vitro and in an Epithelial Infection Model and Inhibit the Proliferation of Immortalized Human Cells. Biomed Res. Int. 2014: 320796.

Ormaetxe, I.I, and O'Neill, S.L. (2007). *Wolbachia*-host interactions: connecting phenotype to genotype. Curr. Opin. Microbiol. 10: 221-224.

Ovchinnikova, E., Journet, E.P., Chabaud, M., Cosson, V., Ratet, P., Duc, G., Fedorova, E., Liu, W., Op den Camp, R., Zhukov, V., Tikhonovich, I., Borisov, A., Bisseling, T., and Limpens, E. (2011). IPD3 controls the intracellular accommodation of rhizobia in pea and Medicago. Mol. Plant Microbe Interact. 24: 1333-1344.

Paetzel, M., Karla, A., Strynadka, N.C., and Dalbey, R.E. (2002). Signal peptidases. Chem. Rev. 102: 4549-4580.

Para-Lopez, C., Baer, M.T., and Groisman, E.A. (1993). Molecular genetic analysis of a locus required for resistance to antimicrobial peptides in *Salmonella typhimurium*. EMBO J. 12: 4053-4062.

Para-Lopez, C., Lin, R., Aspedon, A., and Groisman, E.A. (1994). A *Salmonella* protein that is required for resistance to antimicrobial peptides and transport of potassium. EMBO J. 13: 3964-3972.

Patil, A., Hughes, A.L., and Zhang, G.L. (2004). Rapid evolution and diversification of mammalian a-defensins as revealed by comparative analysis of rodent and primate genes. Physiol. Genomics 20: 1-11.

Pellicer, M.T., Badía, J., Aguilar, J., and Baldomà, L. (1996). *glc* locus of *Escherichia coli*: characterization of genes encoding the subunits of glycolateoxidase and the *glc* regulator protein. J. Bacteriol. 178: 2051-2059.

Penterman, J., Abo, R.P., De Nisco, N.J., Arnold, M.F.F., Longhi, R., Zanda, M., and Walker, G.C. (2014). Host plant peptides elicit a transcriptional response to control the *Sinorhizobium meliloti* cell cycle during symbiosis. Proc. Natl. Acad. Sci. USA 111: 3561-3566.

Pérez-Guerra, J.C., Coussens, G., De Keyser, A., De Rycke, R., De Bodt, S., Van De Velde, W., Goormachtig, S., and Holsters M. (2010). Comparison of developmental and stress-induced nodule senescence in *Medicago truncatula*. Plant Physiol. 152: 1574-1584.

Pérez-Montaño, F., Alías-Villegas, C., Bellogín, R.A., Del Cerro, P., Espuny, M.R., Jiménez-Guerrero, I., López-Baena, F.J., Ollero, F.J., and Cubo, T. (2013). Plant growth promotion in

cereal and leguminous agricultural important plants: From microorganism capacities to crop production. Microbiol Res. 169: 325-336.

Petersen, T.N., Brunak, S., Von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Meth. 8: 785-786.

Pierre, O., Engler, G., Hopkins, J., Brau, F., Boncompagni, E., and Hérouart, D. (2013). Peribacteroid space acidification: a marker of mature bacteroid functioning in *Medicago truncatula* nodules. Plant Cell Environ. 36: 2059-2070.

Pini, F., Frage, B., Ferri, L., De Nisco, N.J., Mohapatra, S.S., Taddei, L., Fioravanti, A., Dewitte, F., Galardini, M., Brilli, M., Villeret, V., Bazzicalupo, M., Mengoni, A., Walker, G.C., Becker, A., and Biondi, E.G. (2013). The DivJ, CbrA and PleC system controls DivK phosphorylation and symbiosis in *Sinorhizobium meliloti*. Mol. Microbiol. 90: 54-71.

Pinto-Tomás, A.A., Anderson, M.A., Suen, G., Stevenson, D.M., Chu, F.S., Cleland, W.W., Weimer, P.J., and Currie, C.R. (2009). Symbiotic Nitrogen Fixation in the Fungus Gardens of Leaf-Cutter Ants. Science 326: 1120-1123.

Popham, D.L., and Young, K.D. (2003). Role of penicillin-binding proteins in bacterial cell morphogenesis. Curr. Opin. Microbiol. 6: 594-599.

Popp, C., and Ott, T. (2011). Regulation of signal transduction and bacterial infection during root nodule symbiosis: Curr. Opin. Plant Biol. 14: 458-467.

Pränting, M., Negrea, A., Rhen, M., and Andersson, D.I. (2008). Mechanism and fitness costs of PR-39 resistance in *Salmonella enterica*s erovar Typhimurium LT2. Antimicrob. Agents Chemother. 52: 2734-27341.

Prell, J., White, J.P., Bourdes, A., Bunnewell, S., Bongaerts, R.J., and Poole, P.S. (2009). Legumes regulate *Rhizobium* bacteroid development and persistence by the supply of branched-chain amino acids. Proc. Natl. Acad. Sci. USA 106: 12477-12482.

Puckett, S.E., Reese, K.A., Mitev, G.M., Mullen, V., Johnson, R.C., Pomraning, K.R., Mellbye, B.L., Tilley, L.D., Iversen, P.L., Freitag, M., and Geller, B.L. (2012). Bacterial resistance to antisense peptide phosphorodiamidate morpholino oligomers. Int. J. Nanomedicine 7: 4829-4840.

Pueppke, S.G., and Broughton, W.J. (1999). *Rhizobium* sp. Strain NGR234 and *R. fredii* USDA257 share exceptionally broad, nested host ranges. Mol. Plant Microbe Interact. 12: 293-318.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L., Eddy, S.R., Bateman, A., and Finn, R.D. (2012). The Pfam protein families database. Nucleic Acids Res. 42: D222-D230.

Ratet, P., Wen, J., Cosson, V., Tadege, M., and Mysore, K.S. (2009). *Tnt1* induced mutations in *Medicago*: Characterisation and applications. In: The Handbook of Plant Mutation Screening (Mining of Natural and Induced Alleles). Meksem, K., and Kahl, G. (eds). Wiley-VCH Verlag Gmbh & Co. KGaA, Weinheim. pp 83-99.

Renard, R., Lambinon, J., Reekmans, M., Van der Veken, P., and Govaert, M. (1983). Nombres chromosomiques de quelques Angiospermes du Rwanda, du Burundi et du Kenya. Bull. Jard. Bot. Nat. Belg. 53: 342-371.

Renier, A., Maillet, F., Fardoux, J., Poinsot, V., Giraud, E., and Nouwen, N. (2011). Photosynthetic *Bradyrhizobium* Sp. strain ORS285 synthesizes 2-O-methylfucosylated lipochitooligosaccharides for *nod* gene-dependent interaction with *Aeschynomene* plants. Mol. Plant Microbe Interact. 24: 1440-1447.

Rey, T., Nars, A., Bonhomme, M., Bottin, A., Huguet, S., Balzergue, S., Jardinaud, M.F., Bono, J.J., Cullimore, J., Dumas, B., Gough, C., and Jacquet, C. (2013). Nfp, a lysM protein controlling nod factor perception, also intervenes in *Medicago truncatula* resistance to pathogens. New Phytol. 198: 875-886.

Rightmyer, A.P., and Long, S. (2011). Pseudonodule formation by wild-type and symbiotic mutant *Medicago truncatula* in response to auxin transport inhibitors. Mol. Plant Microbe Interact. 11: 1372-1384.

Rivas-Santiago, B., Sada, E., Tsutsumi, V., Aguilar-Leon, D., Contreras, J.L., and Hernandez-Pando, R. (2006). β-Defensin gene expression during the course of experimental tuberculosis infection. J. Infect. Dis. 194: 697-701.

Rivas-Santiago, B., Schwander, S.K., Sarabia, C., Diamond, G., Klein-Patel, M.E., Hernandez-Pando, R., Ellner, J.J., and Sada, E. (2005). Human β-defensin 2 is expressed and associated with *Mycobacterium tuberculosis* during infection of human alveolar epithelial cells. Infect. Immun. 73: 4505-4511.

Rodionov, D.A., Vitreschak, A.G., Mironov, A.A., and Gelfand, M.S. (2003). Comparative genomics of the vitamin B_{12} metabolism and regulation in prokaryotes. J. Biol. Chem. 278: 41148-41159.

Roop, R.M., Gaines, J.M., Anderson, E.S., Caswell, C.C., and Martin, D.W. (2009). Survival of the fittest: how *Brucella* strains adapt to their intracellular niche in the host. Med. Microbiol. Immunol. 198: 221-238.

Roop, R.M., Robertson, G.T., Ferguson, G.P., Milford, L.E., Winkler, M.E., and Walker, G.C. (2002). Seeking a niche: putative contributions of the hfq and bacA gene products to the successful adaptation of the brucellae to their intracellular home. Vet. Microbiol. 90: 349-363.

Roth, J.R., Lawrence, J.G., and Bobik, T.A. (1996). Cobalamin (coenzyme B_{12}): Synthesis and biological significance. Annu. Rev. Microbiol. 50: 137-181.

Roux, B., Rodde, N., Jardinaud, M.F., Timmers, T., Sauviac, L., Cottret, L., Carrère, S., Sallet, E., Courcelle, E., Moreau, S., Debellé, F., Capela, D., De Carvalho-Niebel, F., Gouzy, J., Bruand, C., and Gamas, P. (2014). An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. Plant J. 77: 817-837.

Rowley, G., Spector, M., Kormanec, J., and Roberts, M. (2006). Pushing the envelope: extracytoplasmic stress responses in bacterial pathogens. Nat. Rev. Microbiol. 4: 383-393.

Runti, G., Lopez, R.M.C., Stoilova, T., Hussain, R., Jennions, M., Choudhury, H.G., Benincasa, M., Gennaro, R., Beis, K., and Scocchi, M. (2013). Functional characterization of SbmA, a bacterial inner membrane transporter required for importing the antimicrobial peptide Bac7(1-35). J. Bacteriol. 195: 5343-5351.

Ryu, J.H., Kim, S.H., Lee, H.Y., Bai, J.Y., Nam, Y.D., Bae, J.W., Lee, D.G., Shin, S.C., Ha, E.M., and Lee, W.J. (2008). Innate immune homeostasis by the homeobox gene *Caudal* and commensal-gut mutualism in *Drosophila*. Science 319: 777-782.

Salomon, R.A., and Farias, R.N. (1995). The peptide antibiotic microcin 25 is imported through the TonB pathway and the SbmA protein. J. Bacteriol. 177: 3323-3325.

Salzman, N.H., Hung, K.C., Haribhai, D., Chu, H.T., Sjoberg, K.J., Amir, E., Teggatz, P., Barman, M., Hayward, M., Eastwood, D., Stoel, M., Zhou, Y.J., Sodergren, E., Weinstock, G.M., Bevins, C.L., Williams, C.B., and Bos, N.A. (2010). Enteric defensins are essential regulators of intestinal microbial ecology. Nat. Immunol. 11: 76-83.

Samac, D.A., Penuela, S., Schnurr, J.A., Hunt, E.N., Foster-Hartnett, D., Vandenbosch, K.A, and Gantt, J.S. (2011). Expression of coordinately regulated defence response genes and analysis of their role in disease resistance in *Medicago truncatula*. Mol. Plant. Pathol. 12: 786-798.

Sauviac, L., Philippe, H., Phok, K., and Bruand, C. (2007). An extracytoplasmic function sigma factor acts as a general stress response regulator in *Sinorhizobium meliloti*. J. Bacteriol. 189: 4204-4216.

Sawada, H., Kuykendall, L.D., and Young, J.M. (2003). Changing concepts in the systematics of bacterial nitrogen-fixing legume symbionts. J. Gen. Appl. Microbiol. 49: 155-179.

Schafer, A., Tauch, A., Jager, W., Kalinowski, J., Thierbach, G., and Pühler, A. (1994). Small mobilizable multi-purpose cloning vectors derived from the Escherichia coli plasmids pK18 and pK19: Selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. Gene 145: 69-73.

Scheres, B., Van Engelen, F., Van der Knaap, E., Van de Wiel, C., Van Kammen, A., and Bisseling, T. (1990). Sequential induction of nodulin gene expression in the developing pea nodule. Plant Cell 2: 687-700.

Schmerk, C. L., Bernards, M.A., and Valvano, M.A. (2011). Hopanoid production is required for low-pH tolerance, antimicrobial resistance, and motility in *Burkholderia cenocepacia*. J. Bacteriol. 193: 6712-6723.

Schröder, G., Fruhling, M., Puhler, A., and Perlick, A.M. (1997). The temporal and spatial transcription pattern in root nodules of *Vicia faba* nodulin genes encoding glycine-rich proteins. Plant Mol. Biol. 33: 113-123.

Schroeder, B.O., Wu, Z., Nuding, S., Groscurth, S., Marcinowski, M., Beisner, J., Buchner, J., Schaller, M., Stange, E.F., and Wehkamp, J. (2011). Reduction of disulphide bonds unmasks potent antimicrobial activity of human β-defensin 1. Nature 469: 419-423.

Schug, J., Schuller, W.P., Kappen, C., Salbaum, J.M., Bucan, M., and Stoeckert, C.J.Jr. (2005). Promoter features related to tissue specificity as measured by Shannon entropy. Genome Biol. 6: R33.

Schutte, B.C., Mitros, J.P., Bartlett, J.A., Walters, J.D., Jia, H.P., Welsh, M.J., Casavant, T.L., and McCray, P.B. (2002). Discovery of five conserved b-defensin gene clusters using a computational search strategy. Proc. Natl. Acad. Sci. USA 99: 2129-2133.

Scocchi, M., Tossi, A., and Gennaro, R. (2011). Proline-rich antimicrobial peptides: converging to a non-lytic mechanism of action. Cell. Mol. Life Sci. 68: 2317-2330.

Seabra, A.R., Pereira, P.A., Becker, J.D., and Carvalho, H.G. (2012). Inhibition of glutamine synthetase by phosphinothricin leads to transcriptome reprograming in root nodules of *Medicago truncatula*. Mol. Plant Microbe Interact. 25: 976-992.

Sels, J., Mathys, J., De Coninck, B.M., Cammue, B.P., and De Bolle, M.F. (2008). Plant pathogenesis-related (PR) proteins: a focus on PR peptides. Plant Physiol. Biochem. 46: 941-950.

Semple, C.A., Gautier, P., Taylor, K., and Dorin, J.R. (2006). The changing of the guard: Molecular diversity and rapid evolution of b-defensins. Mol. Divers. 10: 575-584.

Sen, D., and Weaver, R.W. (1981). A comparison of nitrogen-fixing ability of peanut, cowpea and siratro plants nodulated by different strains of Rhizobium. Plant and Soil 60: 317-319.

Sen, D., and Weaver, R.W. (1984). A basis for different rates of N_2-fixation by the same strains of *Rhizobium* in peanut and cowpea root nodules. Plant Sci. Lett. 34: 239-246.

Sen, D., Weaver, R.W., and Bal, A.K. (1986). Structure and organization of effective peanut and cowpea root nodules induced by rhizobial strain 32H1. J. Exp. Bot. 37: 356-363.

Shigenobu, S., and Stern, D.L. (2012). Aphids evolved novel secreted proteins for symbiosis with bacterial endosymbiont. Proc. Biol. Sci. 280: 20121952.

Shigenobu, S., and Wilson, A.C. (2010). Genomic revelations of a mutualism: the pea aphid and its obligate bacterial symbiont. Cell. Mol. Life Sci. 68: 1297-1309.

Sibley, C.D., MacLellan, S.R., and Finan, T. (2006). The *Sinorhizobium meliloti* chromosomal origin of replication. Microbiol. 152: 443-455.

Silipo, A., Di Lorenzo, F., De Felice, A., Vanacore, A., De Castro, C., Gully D., Lanzetta, R., Parrilli, M., Giraud, E., and Molinaro, A. (2014). Structural and conformational study of the O-polysaccharide produced by the metabolically versatile photosynthetic bacterium Rhodopseudomonas palustris strain BisA53. Carbohydr. Polym. 114: 384-1391.

Silipo, A., Vitiello, G., Gully, D., Sturiale, L., Chaintreuil, C., Fardoux, J., Gargani, D., Lee, H.I., Kulkarni, G., Busset, N., Marchetti, R., Palmigiano, A., Moll, H., Engel, R., Lanzetta, R., Paduano, L., Parrilli, M., Chang, W.S., Holst, O., Newman, D.K., Garozzo, D., D'Errico, G., Giraud, E., and Molinaro, A. (2014). Covalently linked hopanoid-lipid A improves outer-membrane resistance of a Bradyrhizobium symbiont of legumes. Nat. Commun. 5: 5106.

Silverstein, K.A., Moskal, W.A.Jr., Wu, H.C., Underwood, B.A., Graham, M.A., Town, C.D., and VandenBosch, K.A. (2007). Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. Plant J. 51: 262-280.

Singh, S., Katzer, K., Lambert, J., Cerri, M., and Parniske, M. (2014). CYCLOPS, a DNA-binding transcriptional activator, orchestrates symbiotic root nodule development. Cell Host Microbe 15: 139-152.

Sinharoy, S., Torres-Jerez, I., Bandyopadhyay, K., Kereszt, A., Pislariu, C.I., Nakashima, J., Benedito, V.A., Kondorosi, E., and Udvardi, M.K. (2013). The C2H2 transcription factor regulator of symbiosome differentiation represses transcription of the secretory pathway gene VAMP721a and promotes symbiosome development in *Medicago truncatula*. Plant Cell 25: 3584-3601.

Stacey, G., McAlvin, C.B., Kim, S.Y., Olivares, J., and Soto, M.J. (2006). Effects of endogenous salicylic acid on nodulation in the model legumes *Lotus japonicus* and *Medicago truncatula*. Plant Physiol. 141: 1473-1481.

Starker, C.G., Parra-Colmenares, A.L., Smith, L., Mitra, R.M., and Long, S.R. (2006). Nitrogen fixation mutants of *Medicago truncatula* fail to support plant and bacterial symbiotic gene expression. Plant Physiol. 140: 671-680.

Taber, H.W., Mueller, J.P., Miller, P.F., and Arrow, A.S. (1987). Bacterial uptake of aminoglycoside antibiotics. Microbiol. Rev. 51:439-457.

Taga, M.E., and Walker, G.C. (2010). *Sinorhizobium meliloti* requires a cobalamin-dependent ribonucleotide reductase for symbiosis with its plant host. Mol. Plant Microbe Interact. 23: 1643-1654.

Taga, M.E., Larsen, N.A., Howard-Jones, A.R., Walsh, C.T., and Walker, G.C. (2007). BluB cannibalizes flavin to form the lower ligand of vitamin B_{12}. Nature 446: 449-453.

Tan, B.H., Meinken, C., Bastian, M., Bruns, H., Legaspi, A., Ochoa, M.T., Krutzik, S.R., Bloom, B.R., Ganz, T., Modlin, R.L., and Stenger, S. (2006). Macrophages acquire neutrophil granules for antimicrobial activity against intracellular pathogens. J. Immunol. 177: 1864-1871.

Tan, X.J., Cheng, Y., Li, Y.X., Li, Y.G., and Zhou, J.C. (2009). BacA is indispensable for successful *Mesorhizobium-Astragalus* symbiosis. Appl. Microbiol. Biotechnol. 84: 519-526.

Tarayre, S., Vinardell, J.M., Cebolla, A., Kondorosi, A., and Kondorosi, E. (2004). Two classes of the CDh1-type activators of the anaphase-promoting complex in plants: novel functional domains and distinct regulation. Plant Cell 16: 422-434.

Terpolilli, J.J., Hood, G.A., and Poole, P.S. (2012). What determines the efficiency of N(2)-fixing Rhizobium-legume symbioses? Adv. Microb. Physiol. 60: 325-389.

Tesfaye, M., Silverstein, K.A.T., Nallu, S., Wang, L., Botanga, C.J., Gomez, S.K., Costa, L.M., Harrison, M.J., Samac, D.A., Glazebrook, J., Katagiri, F., Gutierrez-Marcos, J.F., and VandenBosch, K.A. (2013). Spatio-temporal expression patterns of *Arabidopsis thaliana* and *Medicago truncatula* defensin-like genes. PLoS One 8: e58992.

The International Aphid Genomics Consortium. (2010). Genome sequence of the pea aphid *Acyrthosiphon pisum*. PLoS Biol. 8: e1000313.

Tiricz, H., Szucs, A., Farkas, A., Pap, B., Lima, R.M., Maróti, G., Kondorosi, E., and Kereszt, A. (2013). Antimicrobial nodule-specific cysteine-rich peptides induce membrane depolarization-associated changes in the transcriptome of *Sinorhizobium meliloti*. Appl. Environ. Microbiol. 79: 6737-6746.

Toft, C., and Anderson, S.G.E. (2010). Evolutionary microbial genomics: insights into bacterial host adaptation. Nat. Rev. Genet. 11: 465-475.

Tsuchida, T., Koga, R., Horikawa, M., Tsunoda, T., Maoka, T., Matsumoto, S., Simon, J.C., and Fukatsu, T. (2010). Symbiotic bacterium modifies aphid body color. Science 330: 1102-1104.

Uchiumi, T., Ohwada, T., Itakura, M., Mitsui, H., Nukui, N., Dawadi, P., Kaneko, T., Tabata, S., Yokoyama, T., Tejima, K., Saeki, K., Omori, H., Hayashi, M., Maekawa, T., Sriprang, R., Murooka, Y., Tajima, S., Simomura, K., Nomura, M., Suzuki, A., Shimoda, Y., Sioya, K., Abe, M., and Minamisawa, K. (2004). Expression islands clustered on the symbiosis island of the Mesorhizobium loti genome. J. Bacteriol. 186: 2439-2448.

Uppalapati, S.R., Marek, S.M., Lee, H.K., Nakashima, J., Tang, Y., Sledge, M.K., Dixon, R.A., and Mysore, K.S. (2009). Global gene expression profiling during *Medicago truncatula-Phymatotrichopsis omnivora* interaction reveals a role for jasmonic acid, ethylene, and the flavonoid pathway in disease development. Mol. Plant Microbe Interact. 22: 7-17.

Vaishnava, S., Yamamoto, M., Severson, K.M., Ruhn, K.A., Yu, X., Koren, O., Ley, R., Wakeland, E.K., and Hooper, L.V. (2011). The antibacterial lectin RegIIIgamma promotes the spatial segregation of microbiota and host in the intestine. Science 334: 255-258.

Van de Velde, W., Guerra, J.C., De Keyser, A., De Rycke, R., Rombauts, S., Maunoury, N., Mergaert, P., Kondorosi, E., Holsters, M., and Goormachtig, S. (2006). Aging in legume symbiosis. A molecular view on nodule senescence in *Medicago truncatula*. Plant Physiol. 141: 711-720.

Van de Velde, W., Zehirov, G., Szatmari, A., Debreczeny, M., Ishihara, H., Kevei, Z., Farkas, A., Mikulass, K., Nagy, A., Tiricz, H., Satiat-Jeunemaître, B., Alunni, B., Bourge, M., Kucho, K., Abe, M., Kereszt, A., Maroti, G., Uchiumi, T., Kondorosi, E., and Mergaert, P. (2010). Plant peptides govern terminal differentiation of bacteria in symbiosis. Science 327: 1122-1126.

Vanstraelen, M., Baloban, M., Da Ines, O., Cultrone, A., Lammens, T., Boudolf, V., Brown, S.C., De Veylder, L., Mergaert P., and Kondorosi E. (2009). APC/C-CCS52A complexes control meristem maintenance in the Arabidopsis root. Proc. Natl. Acad. Sci. USA 106: 11806-11811.

Vasse, J., De Billy, F., Camut, S., and Truchet, G. (1990). Correlation between ultrastructural differentiation of bacteroids and nitrogen fixation in alfalfa nodules. J. Bacteriol. 172: 4295-4306.

Vernié, T., Moreau, S., De Billy, F., Plet, J., Combier, J.P., Rogers, C., Oldroyd, G., Frugier, F., Niebel, A., and Gamas, P. (2008). EFD Is an ERF transcription factor involved in the control of nodule number and differentiation in *Medicago truncatula*. Plant Cell 20: 2696-2713.

Vinardell, J.M., Fedorova, E., Cebolla, A., Kevei, Z., Horvath, G., Tarayre, S., Roudier, F., Mergaert, P., Kondorosi, A., and Kondorosi, E. (2003). Endoreduplication mediated by the APC activator CCS52A is required for symbiotic cell differentiation in *Medicago truncatula* nodules. Plant Cell 15: 2093-2105.

Visick, K.L., and Ruby, E.G. (2006). *Vibrio fischeri* and its host: it takes two to tango. Curr. Opin. Microbiol. 9: 632-638.

Wadisirisuk, P., and Weaver, R.W. (1985). Importance of bacteroid number in nodules and effective nodule mass to dinitrogen fixation by cowpeas. Plant and Soil 87: 223-231.

Wang, D., Griffitts, J., Starker, C., Fedorova, E., Limpens, E., Ivanov, S., Bisseling, T., and Long, S. (2010). A nodule-specific protein secretory pathway required for nitrogen-fixing symbiosis. Science 327: 1126-1129.

Weatherspoon-Griffin, N., Zhao, G., Kong, W., Kong, Y., Morigen, M., Andrews-Polymenis, H., McClelland, M., and Shi, Y. (2011). The CpxR/CpxA two-component system up-regulates two Tat-dependent peptidoglycan amidases to confer bacterial resistance to antimicrobial peptide. J. Biol. Chem. 286: 5529-5539.

Wehkamp, J., Salzman, N.H., Porter, E., Nuding, S., Weichenthal, M., Petras, R.E., Shen, B., Schaeffeler, E., Schwab, M., Linzmeier, R., Feathers, R.W., Chu, H.T., Lima, H., Fellermann, K., Ganz, T., Stange, E.F., and Bevins, C.L. (2005). Reduced Paneth cell alpha-defensins in ileal Crohn's disease. Proc. Natl. Acad. Sci. USA 102: 18129-18134.

Wehmeier, S., Arnold, M.F.F., Marlow, V.L., Aouida, M., Myka, K.K., Fletcher, V., Benincasa, M., Scocchi, M., Ramotar, D., and Ferguson, G.P. (2010). Internalization of a thiazole-modified peptide in *Sinorhizobium meliloti* occurs by BacA-dependent and independent mechanisms. Microbiol. 156: 2702-2713.

Wells, D.H., and Long, S.R. (2002). The Sinorhizobium meliloti stringent response affects multiple aspects of symbiosis. Mol. Microbiol. 43: 1115-1127.

Wildermuth, M.C. (2010). Modulation of host nuclear ploidy: a common plant biotroph mechanism. Curr. Opin. Plant Biol. 13: 449-458.

Werren, J.H., Baldo, L., and Clark, M.E. (2008). *Wolbachia*: Master manipulators of invertebrate biology. Nat. Rev. Microbiol. 6: 741-751.

Wojciechowski, M.F., Lavin, M., and Sanderson, M.J. (2004). A phylogeny of legumes (Leguminosae) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family. Am. J. Bot. 91: 1846-1862.

Wright, R., Stephens, C., and Shapiro, L. (1997). The CcrM DNA methyltransferase is widespread in the alpha subdivision of proteobacteria, and its essential functions are conserved in *Rhizobium meliloti* and *Caulobacter crescentus*. J. Bacteriol. 179: 5869-5877.

Yorgey, P., Lee, J., Kördel, J., Vivas, E., Warner, P., Jebaratnam, D., and Kolter, R. (1994). Posttranslational modifications in microcin B17 define an additional class of DNA gyrase inhibitor. Proc. Natl. Acad. Sci. USA 91: 4519-4523.

Young, J.P.W. (1992). Phylogenetic classification of nitrogen-fixing organisms. In: Biological Nitrogen Fixation. Stacey, G., Burris, R.H., and Evans, H.J. (eds). Chapman and Hall, New York. pp 43-86.

Young, K.D. (2003). Bacterial shape. Mol. Microbiol. 49: 571-580.

Young, N.D., Debellé, F., Oldroyd, G.E., Geurts, R., Cannon, S.B., Udvardi, M.K., Benedito, V.A., Mayer, K.F., Gouzy, J., Schoof, H., Van de Peer, Y., Proost, S., Cook, D.R., Meyers, B.C., Spannagl, M., Cheung, F., De Mita, S., Krishnakumar, V., Gundlach, H., Zhou, S., Mudge, J., Bharti, A.K., Murray, J.D., Naoumkina, M.A., Rosen, B., Silverstein, K.A., Tang, H., Rombauts, S., Zhao, P.X., Zhou, P., Barbe, V., Bardou, P., Bechner, M., Bellec, A., Berger, A., Bergès, H., Bidwell, S., Bisseling, T., Choisne, N., Couloux, A., Denny, R., Deshpande, S., Dai, X., Doyle, J.J., Dudez, A.M., Farmer, A.D., Fouteau, S., Franken, C., Gibelin, C., Gish, J., Goldstein, S., González, A.J., Green, P.J., Hallab, A., Hartog, M., Hua, A., Humphray, S.J., Jeong, D.H., Jing, Y., Jöcker, A., Kenton, S.M., Kim, D.J., Klee, K., Lai, H., Lang, C., Lin, S., Macmil, S.L., Magdelenat, G., Matthews, L., McCorrison, J., Monaghan, E.L., Mun, J.H., Najar, F.Z., Nicholson, C., Noirot, C., O'Bleness, M., Paule, C.R., Poulain, J., Prion, F., Qin, B., Qu, C., Retzel, E.F., Riddle, C., Sallet, E., Samain, S., Samson, N., Sanders, I., Saurat, O., Scarpelli. C., Schiex, T., Segurens, B., Severin, A.J., Sherrier, D.J., Shi, R., Sims, S., Singer, S.R., Sinharoy, S., Sterck, L., Viollet, A., Wang, B.B., Wang, K., Wang, M., Wang, X., Warfsmann, J., Weissenbach, J., White, D.D., White, J.D., Wiley, G.B., Wincker, P., Xing, Y., Yang, L., Yao, Z., Ying, F., Zhai, J., Zhou, L., Zuber, A., Dénarié, J., Dixon, R.A., May, G.D., Schwartz, D.C., Rogers, J., Quétier, F., Town, C.D., and Roe, B.A. (2011). The *Medicago* genome provides insight into the evolution of rhizobial symbioses. Nature 480: 520-524.

Yurgel, S.N., and Kahn, M.L. (2008). A mutant GlnD nitrogen sensor protein leads to a nitrogen-fixing but ineffective *Sinorhizobium meliloti* symbiosis with alfalfa. Proc. Natl. Acad. Sci. USA 105: 18958-18963.

Zasloff, M. (2002). Antimicrobial peptides of multicellular organisms. Nature 415: 389-395.

Zehr, J.P., Jenkins, B.D., Short, S.M., and Steward, G.F. (2003). Nitrogenase gene diversity and microbial community structure: a cross-system comparison. Environ. Microbiol. 5: 539-554.

Zhang, J.Y., Cruz, D.E., Carvalho, M.H., Torres-Jerez, I., Kang, Y., Allen, S.N., Huhman, D.V., Tang, Y., Murray, J., Sumner, L.W., and Udvardi, M.K. (2014). Global reprogramming of transcription and metabolism in Medicago truncatula during progressive drought and after re-watering. Plant Cell Env. In press. doi:10.1111/pce.12328.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E., and Ecker, J.R. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. Cell 126:1189-1201.

Zhou, C., Han, L., Pislariu, C., Nakashima, J., Fu, C., Jiang, Q., Quan, L., Blancaflor, E., Tang, Y., Bouton, J.H., Udvardi, M., Xia, G., and Wang, Z.Y. (2011). From model to crop: Functional analysis of a *STAY-GREEN* gene in the model legume *Medicago truncatula* and effective use of the gene for alfalfa improvement. Plant Physiol. 157: 1483-1496.

Zhou, P., Silverstein, K.A., Gao, L., Walton, J.D., Nallu, S., Guhlin, J., and Young, N.D. (2013). Detecting small plant peptides using SPADA (Small Peptide Alignment Discovery Application). BMC Bioinformatics 14: 335.

Annexes

Annex 1

Table 3. *Aeschynomene afraspera* and *Aeschynomene indica* root and nodule expressed sequenced tags (ESTs) and cluster collection statistics.

	A. afraspera		***A. indica***	
	Root	**Nodule**	**Root**	**Nodule**
Number of cDNA sequenced	9492	9582	9621	9847
EST summary:				
Number of high-quality ESTs	9017 (95%)	8688 (91%)	9117 (95%)	8768 (89%)
Average EST lenght (bp)	551	513	528	510
EST size range (bp)	102-819	102-817	101-803	101-782
Non valid sequences	475 (5%)	894 (9%)	504 (5%)	1079 (11%)
Small size	360 (76%)	729 (82%)	387 (77%)	911 (84%)
Low quality	115 (24%)	165 (18%)	117 (23%)	168 (16%)
Cluster summary:				
Number of clusters	923	732	1006	812
Number of singletons	4162	3493	4566	4164
Number of contigs	1161	903	1268	1025
Average contig lenght (bp)	750	723	729	703
Contig size range (bp)	220-1794	122-2267	146-1268	120-1025

Annex 2

Table 4. Nodule-specific Cys-rich putative peptides from *Aeschynomene indica*, *A. afraspera* and *A. evenia*.

The NCR were named with the initials of the plant species and numbered according to the read count level in the EST nodule libraries (for *A. indica* and *A. afraspera*) or arbritarily for *A. evenia*. The NCRs with a read count above 5 in the nodules are grey shaded.
Signal peptides indicated in bold were determined using SignalP 4.0 (Petersen *el al*., 2011) and ajusted manually. Theoretical pI and MW were calculated for mature peptides from Expasy web site (Gasteiger *et al*., 2005).
For the A. *evenia* NCRs, the pI and MW were calculated for the mature peptide even if the signal peptide was truncated or contained stop codon (indicated in the table). The presence of a signature corresponding to the NCR motif 1 (1) or motif 2 (2), as defined in **Fig. 32**, is indicated in the last column. The 1* indicate a motif 1 where the spacing between Cyst 4 and 5 contains 5 amino acids instead of 4.

		Sequence	Mature peptide (aa)	Nbr of Cys	Theoretical		Read count		NCR motif
					pI	MW	Root	Nodule	
A. afraspera									
AaNCR01	CL3contig5	**MAKLNWANLFALILLVTVMSMASHHLQVVG**LRMIDSWIDPKCPLDEMSCMSSCRNKHGGLAMGHCVGSICICPFDK	46	6	6.87	5087.04	1	242	1
AaNCR02	CL13contig1	**MAKLANFFTLVLLVAVISITSHHQVAE**SRLCNHRWFEIRCNVNESFCKNDCSGRHGDRTINAFCVIGYCSCYFDCPDGEDCSCIGF	59	10	5.44	6768.5	0	71	2
AaNCR03	CL3Contig3	**MAKLNWANLFALILLVTVMSMASHHHQVVG**LRMIDSWIDPKCLEMSCMSSCRNKHGGLAMGHCVGSICICPFDK	44	6	7.79	4874.83	0	45	1
AaNCR04	CL16contig1	**MARLTLINLLSLVLLITGILITSSDYTVAG**YVCIVVWRDPNCGKGDPNYCAQDCKNDYGEKAIPVCRYDVCFCGYECTFPNNGDDYNGERNQNNRM	66	8	4.57	7615.40	0	25	2
AaNCR05	CL16contig2	**MAKLTLINLLSLVLLVTGMLMTSSDYKVAA**GYMCVVGWRDPNCGKGDPNYCAQDCKNDHGEKATAFCRDDVCFLWL	46	6	4.89	5202.85	0	23	no
AaNCR06	CL31contig1	**MARLTLINFLPLLLLVTGMWMKSSHYKVAG**EYIDVCFDVWADTNCEKGDSNFCVQECKNKHIEEGAAQCRENRCFCIFPCAPNIPSPISYDGGDDYNGEGINNNDGI	77	8	4.08	8573.32	0	22	2
AaNCR07	CL50contig1	**MTKLTSTNIFALILLFTAISMTSLHQVAS**NSCWHSWMDTNCEPETCVQLCSDSYKVLIDAFCENLHCVCEYECPDN	47	8	4.01	5479.0	0	14	2
AaNCR08	CL64contig1	**MTKLILGNFFAIVLLMTVILIIKSHIEVAG**EAGCEQSSDYLYPQIADYCEEGFDHWYDRNCPTVDFFCKLECIRKHGRTGDGACNWIYCYCRYLPNSTSEDALRP	75	8	4.60	8790.69	0	10	1
AaNCR09	CL67contig1	**MTKLTSTNIFALILLFTAISMTSLHQVAS**NSCWHSWMDTNCEPETCVQLCSDSYKVEASRWPRMLLVKDEAPATVRDSLLQESVTVAPNLWDFKIKRGPEGSIN	75	4	4.94	8570.6	0	10	no
AaNCR10	CL3Contig2	**MAKLNWAKFVALILLVTAMSMTSHHHQVVG**VEMKDTWIDLRCTFDDLPCTSDCRNKYGVFAVGRCVGVLCECTYELC	47	7	4.46	5372.21	0	7	1
AaNCR11	CL132contig1	**MARLTFVYLFSLLLLVTAISITSHHQVVG**YFLCLDDWVDESCAERGDSDGCMQDCKNKHGYTAQGFCGYDSCFCSYQCS	49	8	4.08	5656.15	0	6	2
AaNCR12	CL746contig1	**MSRLTSAYFFSIVLLVEAISKTSHHQVVA**NKCLDDWFDHNKYCVEDDCVQECKRQHNSTEGYCAYYNCFCPYECPEDSPMPPNSPVSPDYDDSPNPPTWDK	72	8	4.14	8448.1	0	6	2

AaNCR13	CL127Contig1	**MTKLILATIFALILLMIAISMMTSEVDG**NKMCRRFWYDLNCKDPEYRGCMMECKSNYGDTAYGFCQYPKCSCNYGCPD	50	8	7.67	5959.77	0	6	2
AaNCR14	CL111Contig1	**MARLKTLNFLALVLLVTGMWMTSEYKVAG**GIYTCSEVWDDPNCEKEDYTNCTQNCKRKYGMKATSHCRGDKCFCIFECDLPFLPSPINYDGDYGNNEGDDYNSEGINNNNNDGT	85	8	4.22	9642.35	0	6	2
AaNCR15	CL211contig1	**MAKLILGNFFAIVLLMTVILIIKSHIEVAG**EAGCKPSSDYQIAGFMFWIDMNCPTVDFFCNSECKRKHPKGDGECSGIYCHCRYLESEIRRDEMMS	66	7	5.19	7654.68	0	5	1
AaNCR16	CL220Contig1	**MAKLSNLFALVLLVTAISMPSHHQVES**KKCTDEWYDIRCNLNEGYCRVDCLEVHGGLAKGYCQGDYCFCDYYCPGLLIV	52	8	4.70	6009.86	0	4	2
AaNCR17	CL263Contig1	**MAKLSNFFALVLLVTAIWMPSHHQVEG**KECTDRWYDERCSLNESYCRVDCLEVHGGLAKGYCQGDYYCFCNYYCPGLLIV	53	8	4.75	6219.02	0	4	1*
AaNCR18	CL220Contig2	**MAKLSNLFALVLLVTAIPMPSHHQVES**KKCTDE*YDIRCNLNEGYCRVDCLEVHGGLAKGYCQGDYCFCDYYCPGLLIV	51	8	ND	ND	0	4	2
AaNCR19	CL266Contig1	**MAKLANFFTLVLLVTAISMTMNQQVEG**IKCNEMWYDIRCNLSESDCRLDCIELYGGGPAKGYCVGDYCCCNYDCVGDYCSCNYEQEQVCGGPQ	66	12	4.02	7420.27	0	4	2
AaNCR20	CL136Contig1	**MAKLANFLTLVLLLTAISMTMNQQVEG**IKCNEMWYDIRCILSESYCREDCIEKHGVLAKGYRMGNKCCCLTV	45	7	7.73	5279.24	0	4	no
AaNCR21	CL297Contig1	**MAKLANFFTLVLLVTAIMITSHQLVEG**QWCKNKWLDIRCGLYESYCWEDCSKLHGETAIGFCEDIYCSCYYDCSRNIKN	52	8	5.09	6258.06	0	4	2
AaNCR22	CL181Contig1	**MTELTSKNHFALIILVTAISMRSLHQVVA**RECIDTWIDRNCEYNYIEICKLDCKDYYGESSFAYCVDHHKCKCHYKCPQWPYFN	55	8	5.82	6841.71	0	4	1*
AaNCR23	CL50Contig2	**MTKLTSTNIFALILLFTAISMTSLHQVAS**NSCWYSWMDKYCEDETCVQSCNDTYKVLIDAFCENIHCFCEYECPNKDK	49	8	4.28	5918.59	0	4	2
AaNCR24	CL217Contig1	**MTKLTLANIFAIVILVIVMLIQSHDGVAG**QDDYCEPNSDYLYPQIANCWEDGLYSWFDLTCPMSNENCKLDCKRKVGPTGDGVCILIYCYCRY	64	8	4.20	7493.43	0	4	1
AaNCR25	CL226Contig1	**MAKLAKFLPIVLLVTAISITSHHQVAG**KKLCNHRWFDIRCRNLVDYDNESFCNKDCSDLHGDAVAMFGFCVLDYCSCHYVSPTVPVDEHCSCNYEQNNGLCVGV	77	10	5.03	8817.89	0	4	1
AaNCR26	CL779Contig1	**MVKLTLTNLFALVLLVTGMLMRSDVDG**VEVTDFWIDWNCRVDEINCMIDCQEKHGITAVGSCIAIICKCVFESNC	48	7	4.28	5455.30	0	3	1
AaNCR27	CL416Contig1	**MVKLTLTNFIALFLLVTAISMGSHQQVAS**DLCWNSWIIINCKKEDCKKQCSSEYPVIIEAFCAGSYCFCKYECT	45	8	5.06	5267.11	0	3	2
AaNCR28	CL3Contig1	**MSMASRHHQVVG**LRMIDSWIDPKCPLDILGLIQSAYEMSCMSSCRNKHGGLAMGHCVGSICICPFDK	55	6	6.87	6046.19	0	2	1
AaNCR29	CL491Contig1	**MTKLTLANYFVLSLQITVIWMTSNGKKVAG**EDCLCNWFDKNCRDGDSKECVKDCRDSHGNMAHGSCAYDLCVCSYSCPNSSPSWDK	56	9	4.74	6286.88	0	2	2
AaNCR30	CL820Contig1	**MTNLDLTNFFAILLLVTAISMTSHHHQVVS**EVCLDEFDMNCREIDSTICFEDCKYMHGKKTKGVCAFNHCTCMYEC	46	8	4.82	5395.21	0	2	2
AaNCR31	CL414Contig1	**MAKLNLKTFFAFVLLIIAISMTSHHQVAG**DVCYEYWADFLCLFGYSRSCSRYCEFAHQGVGATIGYCEDGKCTCSFQC	49	8	4.58	5587.25	0	2	2
AaNCR32	CL371Contig1	**MAKLTLTNFFVVVLLVSVISMTSDHHQVVG**SKCPDEGVDNKYICSSEWVDMNCTTGDSEDCSQECVNNYGIAEGSCAGTKCCCRYYCPGSLE	62	10	4.05	6723.35	0	2	2
AaNCR33	CL310Contig1	**MNTLTLANFFAVILSVTAISMTSYHHHQVVS**EECLDMFDMNCIEEDSNKNCVEACKYMHAGDNNVDGKCVLDHCYCKYEC	49	8	4.27	5676.35	0	2	2
AaNCR34	CL16Contig4	**MARLTLINLLSLVLLITGILITSSDYTVAG**YVCIVVWRDPNCGKGDPNYCAQDCKNGYGEKAIPVCRYDVCFCGYECTFPNNGDDYNGERNQNNRM	66	8	4.77	7557.36	0	2	2
AaNCR35	CL3Contig4	**MAKLTWAKFVALILLVTAMLMTSYHHQVAG**IHTTDSWIDPSCTFDEMFCMLDCRNKHGVFALGSCVGNICECTYELKS	48	6	4.63	5422.18	0	1	1
AaNCR36	CL304Contig1	**MTKLILGNFFAIVLLMTVILIIKSHIEVAG**EASCEPSSDYLYPQIAGFDSWYDINCPMSDFLCNMECKRKYGRTGYGGCVGNYCHCYY	58	7	4.65	6676.44	0	1	1
AaNCR37	CL278Contig1	**MAKLATFLHSFFWSNIFALFLLVTAISITIHHQVAC**GKLCNHKWFDRRCGYMDDDSCNKYCLDLHGEDKLAGFCVFVYCSCHYICPGDDVPVDEDCSCNYEQKCTVCWGSLGSCTNIVLPLV VGY	89	13	4.67	10038.41	0	1	2
AaNCR38	CL77Contig1	**MARLTLINFLPLLLLVTGMWMTSQYKVAG**SYICYVFWDDPNCKNEDYTNCTQECESKHGKEARPICREEKCFCRFDCDFPFLPTPNDDGNYGNNDGI	68	8	4.37	7972.70	0	1	2
A. indica									
AiNCR01	CL27contig1	**MTKLTLLNFFTFGLLITAMWMTSHHQVFG**EQCIDMWFDEGCRNLDFHKCSEDCKGKHGAIGECADEYCFCSYECDNNSPK	51	8	4.41	5875.44	0	21	2

AiNCR02	CL19contig1	**MIKLTLTNLFALVLLITAMSMINHGEA**AGTGFDSWVDKNCPQENFYCIHSCRQKHGQSAMGECGGSFCYCYYY	46	6	6.02	5242.79	0	19	1
AiNCR03	CL25contig1	**MTKFTLTNVFALVLLVTVISITSHHQVTG**ITAFDSWVDFYCATDNYNCIMSCRNKYGITGRGNCNDVFCVCFYKLLN	48	6	6.06	5563.34	0	17	1
AiNCR04	CL25contig2	**MTKFTLTNVFALALLVTVISITSHHQVTG**ITAFDSWVDFYCATDNYNCIMSCRNKYGITGRGNCNDVFCVCFYKLLN	48	6	6.06	5563.34	0	8	1
AiNCR05	CL84contig1	**MTKLTLKNFFTLILSVTAILITSHHQVAS**MGFDTWIDIDCPLNNHSCIMNCRTKYGLTASGKCNDVFCCCFYYLLN	47	7	5.34	4250.84	0	8	1
AiNCR06	CL45contig1	**MIKLSLINFFALVLLLTAILMISHHEVAC**LGSDSWMDINCPGDDSICNFHCKRKYGSSAWAMCSGSFCRCYH	43	6	6.87	4840.44	0	7	1
AiNCR07	CL164contig1	**MIKLTLINIFALVLLVAVISSHHQVDSI**QRYDYWVDINCIYNCVMDCINKHGDTAVGRCVGNLCYCTY	41	6	5.36	4799.50	0	7	1
AiNCR08	CL343contig1	**MINLIFTKFFALVLLMTAIISHHQVTG**IEIYDYWVDMNCVYNCAIDCVKKHGVTALGKCVCNLCYCRYNMS	44	7	6.70	5119.02	0	7	1
AiNCR09	CL107contig1	**MVELTLVNFFALVLLITDISMTNHHQVAA**FYFCLNDWWVDENCEKGVSINCMEDCINKYGDTSNGFCEFYNCYCQYKCPESPDSPNSKDSPIEPHN	67	8	4.13	7838.57	0	7	2
AiNCR10	CL19contig2	**MTKLILTNLFTFVLLMTAISMINHGKA**ASTGFDSWIDKNCPMENFYCIHSCRQKHGQSAMGECGGSFCYCYYY	46	6	6.02	5289.90	0	6	1
AiNCR11	CL150contig1	**MTKLILTNFFALALLMTVISIISHHQVTG**ITAFDSWVDFYCATDNYNCIMSCRNKHGITGMGNCKDVFCVCYYNLLN	48	6	5.36	5528.31	0	6	1
AiNCR12	CL46contig2	**MTKLTLTKKFAIILLVTAMSMINHHQVAG**ATCIEIWTDLRCKLYDGGCWMDCLKKYGDASRGHCQDDYNCFCLYECRPKYLAK	54	8	6.74	6367.34	0	6	1*
AiNCR13	CL176contig1	**MIRLILANFITLVLLVTIISITRKHQVVG**SGFECWIDLDCPWENHTCIVSCRNKHGLTASGNCYDMFCYCYYTKY	45	7	5.35	5355.06	0	5	1
AiNCR14	CL698contig1	**MVKLILTNFFAFILLMAVTSKTIHHQVIG**YDCVEDWIDENCEKGDYSTCFFDCVDKHGDIAIGYCEGHNCFCQYGSCTHIKDHIDIGY	59	8	4.25	6795.41	0	5	2
AiNCR15	CL425contig1	**MVKLILINFFALFVLLMPVTSKTIHRQVAA**YDCVENWIDENCEKGDYSTCFFDCVDKHGDIAMGYCEGHKCFCQYGSCTHTKDHIDIGY	59	8	4.49	6814.48	0	5	2
AiNCR16	CL88contig1	**MVKLILTNFFALVLLMSVISKTIHYQVTG**YDCVENWIDENCEKGDYSTCFFDCVDKHGDIAMGYCEGHKCFCQYGSCTHTKDHIDIGY	59	8	4.49	6814.48	0	5	2
AiNCR17	CL172Contig1	**MTKLVLTNLFALVFLITVISMRTHHQVAG**YLGLDDWIDENCTRLFFSDSGCVKDCKHKHGDTGDGYCVDNRCFCSYEIPDSPNYGDSPDWSE	63	6	4.15	7192.72	0	5	1
AiNCR18	CL286contig1	**MSKLTLANFFSIVLLVTVILMITRHQVTS**EEILDSWIDINCPMDNFNCNWDCVHKHGATANGYCNGNICCCKYFLHS	48	7	4.83	5513.18	0	5	1
AiNCR19	CL615Contig1	**MIRLILANFFTPVLLMSAISITGKHQVAT**LGFATWIDGHCPRENLNCIVNCRSEYGPSAKGMCNDVFCFCYYTIY	46	6	5.48	5276.03	0	4	1
AiNCR20	CL848Contig1	**MTRLTLINFFIPVLLMSAISTTSQHQVAG**GVGFANWIDLECPRANFYCMIDCRSEYGPSAKGLCNDVFCFCYYTIN	47	6	4.44	5371.13	0	4	1
AiNCR21	CL521Contig1	**MIKLTLANFFSLILLVAAISMTSQHQVAFS**GFDKWIDFDCPVTSLNCIADCRERYGALTRAYCNNVFCHCFYTQYY	46	6	5.47	5487.19	0	4	1
AiNCR22	CL907Contig1	**MTRLTLANFLILILLVAAISMTSQHQVLA**FSGYDKWIDFDCPVTSLNCIADCRERYGALTRAYCNNVFCHCFYTQYY	48	6	5.47	5737.44	0	4	1
AiNCR23	CL288Contig1	**MTKLILANFFILVLLMTAISMTSQYQVAD**SGYDKWINLDCEVTTLNCIADCKKRYGPIARAYCSNIYCYCFYSKYY	47	6	8.22	5583.40	0	4	1
AiNCR24	CL605Contig1	**MTRLILANFFTLVLLMTAISMTSQHEVAD**SGYDKWIDLYCEVTSLNCIAECKKRYGPIARAYCSSNIYCYCFYSKY	47	6	8.22	5556.38	0	4	1*
AiNCR25	CL255Contig1	**NLHVLYHFNSYLAAIWMASGDHQVYG**DVCFDHWFDLGCKNGDFDRCLEDCTKQHGNLATGECDEDYCICRYECDKCPYPLS	55	9	4.26	6391.05	0	4	2
AiNCR26	CL269Contig1	**MIKLTLANIFIIVLLVAAISMKSNCEVAT**TEDLARWYDSNCLESDKTCQIECKFRYGITAQGGCCGDFLCCCYFLLMPDLVK	53	8	4.46	6071.00	0	4	1*
AiNCR27	CL45Contig2	**MIKLSLTNLFALVLLLTAILMISHHEVAC**LVSDSWMDINCPGDDSICNFHCKRKYGSSAWAMCSGSFCRCYH	43	6	6.87	4882.52	0	3	1
AiNCR28	CL92Contig2	**MTNKLSLANFVAFMLLVTAISTSSHNPAMG**ILPCFDYWTDKNCPDNNDNCIQDCKDKYGNSVVGNCNGYYCYCRYACPRYTRMEGPHVLVDKVNGVINIDGMQYR	75	8	5.59	8716.81	0	3	2
AiNCR29	CL339Contig1	**MASKLTLANFVVFMLLLTAIWTSCHNPAMG**SIPCYDYWTDPNCPDDDDYCIQVCKERYGNSAVGNCNGCYCYCRYSCPRYT	51	9	4.27	5957.55	0	3	2
AiNCR30	CL396Contig1	**MSFKLTLANFFSIVLLVTAISMTTHHQVTS**EEILDSWIDINCPLDNFNCNWDCVNKHGATANGYCNGNICCCKYFLHS	48	7	4.51	5472.10	0	3	1

AiNCR31	CL295Contig1	**MTKLILTNVFALILLVTAISITSHHHVNS**IQCFDNWVDIDCINNCMSDCIEKYGFTALGKCVGKTCYCCYESMS	45	8	4.23	5118.90	0	3	1
AiNCR32	CL468Contig1	**MNKLNWACIFGFVFVVTAISMTNEVAG**EDCLDFWNDPRCKYSNQYICKTNCKTKHGDLAIGSCLDRVCYCAYVC	47	8	6.77	5460.24	0	3	2
AiNCR33	CL45Contig3	**MIKLSLTNFLALVLLLTAILMISHHEVAC**LGSDSWMDINCPGDDSFCNFHCKRKYGSSAWAMCSGSFCRCYH	43	6	6.87	4874.46	0	2	1
AiNCR34	CL831Contig1	MIKLILTNFFALFLLVTAILIISHREVASFGSDSWIDINCPRDNFFCNFNCRRKYGSSAWGICNGNFCHCYY	43	6	7.78	5082.65	0	2	1
AiNCR35	CL909Contig1	**MASKLTLANFVVLMLLMTAISTSSHNPAMG**SIPCYDYWTDPNCPDDDDYCVQVCQERYGNSAVGNCNGCYCYCRYSCPRYT	51	9	4.04	5943.48	0	2	2
AiNCR36	CL905Contig1	**MTRIDFANFFTLILLITIIWMIIQHQVVG**TGFECWIDLDCPWENHTCIVSCRNKHGLTASGNCYDMFCYCYYTKY	46	7	5.33	5456.16	0	2	1
AiNCR37	CL941Contig1	**MSRLTLINFSAVILLVIVISVRNNHLVAG**DLCYEEWVDENCINNGSYYCMRDCKKKYGNTGDGLCALLKCVCLYECPDSRDSPEPRI	58	8	4.50	6693.55	0	2	2
AiNCR38	CL906Contig1	**MYISSQQILIFVKNSSHYTS*KNVTSKTIHHQVIG**YDCVEDWIDENCKKGDYSTCFFDCVDKHGDIAIGYCEGHNCFCQYGSCTHIKDHIDIGY	59	8	4.46	6794.47	0	2	2
AiNCR39	CL19Contig3	**MTKLILTNLFTFVLLLMTAISMISYHDVTG**VGFDSWIDKNCPLENFYCIHSCRQKHGQSAMGECGGSFCYCYY	43	6	6.01	4948.56	0	1	1
AiNCR40	CL96Contig1	**MFIFHLANFFSIVLLVTAISMTTHHQVTS**EEILDSWIDINCPMDNFNCNWDCVHKHGATANGYCNGNICCCKYFLRS	48	7	4.86	5532.22	0	1	1
AiNCR41	CL673Contig1	**MRLTLVTFFALFLLVTVISMKNDHQVVSG**IMTFDYWIDTSCIQSAFSCMIECRLRHGMTGVGSCSDIFCLCHYELMN	48	6	4.80	5533.45	0	1	1
AiNCR42	CL259Contig1	**PLGAISITSHHQVAA**ITAFDSWVDFYCATDNYNCIMSCRNKYGITGRGNCNDVFCVCFYKLLN	48	6	6.06	5563.34	0	1	1
AiNCR43	CL92Contig1	**MTNKLILANFVAFMLLVTAISTSSHNPTMD**ILPCFDYWIDKNCPDNNDNCIQDCKDRYGNSAVGNSLDYCYCRYACPHYTRMEGPHVLFDKVK	63	7	5.42	7402.32	0	1	2
AiNCR44	CL127Contig1	**MIPRLVKNYSHSLARNMTILMSSHNSVMVMG**SLPCFDYWSNMNCPDDDDNCIQSCTDRYGDSAVGNCNDCYCCCRYVCPRNT	51	10	3.90	5820.35	0	1	2
A. evenia								**Remarks**	
AeNCR01	CL8618Contig1	**MTKLNLTNLFAFILLMTAISMTNHGQVTA**MECTDHWIDLNCLLDEINCMLDCRSRHGITALGNCMGNICVCYYELRLC	49	8	4.63	5671.64			1
AeNCR02	CL3731Contig1	**MIKLTLTNLFTLILLMTVIWMINRVEM**AGTGFDSWVDKNCPLENFYCIHSCRQKHGQSAMGECGGSFCYCYY	72	6	6.02	5064.64			1
AeNCR03	CL452Contig1	**MTKLTLLNFFTFGLLMTAIWMASNDHQVSG**DVCFDHWFDLGCKNGDFDRCLEDCTQQHGNLATGECDEDYCICRYECDKC	50	9	4.13	5833.36			2
AeNCR04	CL8618Contig3	**NHGQVTA**MECTDHWIDLNCLLDEINCMLDCRSRHGITALGNCMGNICVCYYELRLC	49	8	4.63	5671.64	signal peptide truncated		1
AeNCR05	CL22258Contig1	**MTKLTLANFSALVLLVAAISMISHHQVAS**IECIDHWIDINCPMDEINCTIDCKKRHGITAVGTCIGNICKCYYEL	46	7	4.96	5262.14			1
AeNCR06	CL10282Contig2	**MSKITLINFLAFILLVTAMSMTSHRQVAA**MKCVDHWFDILCPEDEINCMMDCRKKHGFTVVGNCIGIVCVCYCE	49	9	4.96	5202.18			1
AeNCR07	CL3940Contig1	**MTKLTLINFFAFALLITAISMTNHHQVTG**ITAFDSWVDFYCATDNYNCIMSCRNKYGSTGRGNCNDVFCVCFYKL	46	6	6.06	5310.00			1
AeNCR08	CL10282Contig1	**MSKITLINFLAFILLVTAMSMTSHRQVAA**MKCVDHWFDILCPEDEINCMMDCRKKHGLRSLYDKLSS	38	4	6.00	4561.34			no
AeNCR09	CL295Contig1	**MSKLTLANFFSIVLLVTAISMTTHHQVTS**EEILDSWIDINCPMDNFNCNWDCVHKHGATANGYCNGNICCCKYFLH	47	7	4.83	5426.10			1
AeNCR10	CL7874Contig1	**MVKLSLTNFFRLISLVAAISMTSHHQVSG**MIGLDFWIDLECPMYNMSCVMDCKNQHGFTASGYCRDVFCYCYYQV	46	6	4.35	5465.33			1
AeNCR11	CL10838Contig1	**MTKLTLSNFLALVLLITVISMTNYHQVAT**ESYDLWIDIGCPLDNLNCKIDCRNKYGLKSSGECIHIFCYCYYDLS	46	6	4.59	5388.13			1
AeNCR12	CL9423Contig2	**MTKLILINVFALILLVTVISMTSYHHVDS**IQCFDNWVDIDCINNCMSDCIEKYGFTVLGKCVGKTCYCCYE	42	8	4.23	4841.60			1
AeNCR13	CL16157Contig1	**MVELTLANFFALVLLVTGISKTNPYQVVG**VKVFDFWVDMSCEFGDINCERACRNKYGYTAMGYCSGIFCYCRYE	45	6	5.00	5345.11			1
AeNCR14	CL15608Contig1	**LVTAISMTSHHQVAA**IFGFDSWIDGNCPSDNLNCNIDCINKHGLTGSGNCNGAICFCYYQL	46	6	4.13	5021.59	signal peptide truncated		1
AeNCR15	CL2637Contig2	**MTKLVLTNLFVLVFLITAISMTTHHQVAG**YLCLDDWIDENCTRLFFSDSGCVKDCKHKHGDTGDGYCVDNRCFCSYE	48	7	4.51	5578.14			1

AeNCR16	CL5370Contig1	**KLTLITFFVLFLLVTVISMKNHDKVVSG**IMTFDYWIDISCIQSAFSCMTECRLRHGMTAVGSCSDIFCLCHYEL	47	6	4.80	5359.23	signal peptide truncated	1
AeNCR17	CL17837Contig1	**MTKSNLMNFFALGVLVTAISMTSDHQVAG**IKVFNFWIDMNCEVTDIHCIMDCKIKYGKTSVGYCKGIYCCCRYD	45	7	7.72	5324.32		1
AeNCR18	CL10753Contig1	**MTKLILTNFLALIVLLTAILMISHHEVAS**LDFDSWIDINCPMNSSNCKRDCKTKYGSSAWAVCSGSFCHCYY	43	6	6.70	4907.51		1
AeNCR19	CL4882Contig1	**MAKLTLTNYFAIILLMTAISMISHHQVAG**VTCIEIWTDLRCKLYDGGCWLDCLKKYGGASKGHCQDDYNCFCLYECR	48	8	5.57	5590.43		1*
AeNCR20	CL21671Contig1	**MTRLTLANFLILILLVAAISMTNEHQVLA**FSGYDKWIDFDCPVTSLNCIADCRERYGALTRAYCNNVFCHCFY	44	6	5.47	5181.86		1
AeNCR21	CL11599Contig1	**MAKIVLTNLFALILLVTVTSMTNHHQIAG**IMCTKYWFDRRCKLYDGGCWQDCLTKHGVEARGHCEDDYNCSCQFE	46	7	5.46	5515.19		1*
AeNCR22	CL6083Contig1	**MVKLILTNFFALVLLITVTSKTIHHQVAG**YDCVENWIDENCEKGDYSTCFFDCVDKHGDIAIGYCEGHKCFCQY	45	7	4.23	5267.78		1
AeNCR23	CL21077Contig1	**MTRLTLANFITLVLLVTIISMTSKHQVVG**SGFECWIDLDCPWENHTCIVSCRNKHGLTASGNCYDMFCYCYY	43	7	4.80	5049.68		1
AeNCR24	CL5726Contig2	**MTRLTLINFFAIIILVTAIWIKNHGQVAG**DLCFEEWVDENCINNGSYDCMRDCKNKYGNTGDGLCAVLKCVCLYE	46	7	4.16	5247.90		1
AeNCR25	CL1Contig1190	**KLILANFVAFMLLVTAISTSSHNPAMG**ILPCFDYWTDKNCPDNNDNCIQDCKDRYGNSAVGNCNGYYCYCRY	45	7	4.39	5301.81	signal peptide truncated	1
AeNCR26	CL1531Contig1	**MIKLTFTSTFIIVLLVTAISMKSHCEVAT**SEDLARWYDSNCLESDKTCQIECKFRYGITAQGGCCGDFLCCCYFLL	47	8	4.39	5373.11		1
AeNCR27	CL27026Contig1	**MARLTLKNFFTLILSVTAILITSHHQVAS**MGFDTWIDIDCPQNNRGCIMNCRTKYGLTASGKCNDVFCCCFY	43	7	5.84	4908.66		1
AeNCR28	CL1Contig509	**VSKLTLTNFVVLMLLMTAISTSSHNPAMG**SIPCYDYWTDPNCPDDDDYCIQVCQERYGNSAVGNCNGCYCYCRYS	46	8	3.80	5336.78	signal peptide truncated	1
AeNCR29	CL305Contig2	**MTRLILANFFTLVLLMTAISMTSQHEVAD**SGYDKWIDLYCEVTNLNCIADCKKRYGPIARAYCSNIYCYCFYS	44	6	7.50	5190.95		1
AeNCR30	CL8283Contig1	**MPKLILTSFFALILLVTAILMISHHVVG**LGYDTWIDQNCPWESFYCRLSCRTKHGGKAMGECGGGPFCYCYY	44	6	6.71	5067.73		1
AeNCR31	CL4958Contig1	**MIKLILTNFFALFLLVTAMLIISHHEVAS**FVSDSWMDINCPGDNFFCNFNCRRKYGSSAWGICNGNFCHCYY	43	6	6.70	5043.62		1
AeNCR32	CL1159Contig1	**MIKLTLINFFALVLLITDISMTNHHQVTA**FYFCLDDWFDENCEKGISIDCMEDCINKYGDTSNGFCEFYNCYCQYK	47	7	3.89	5684.27		1
AeNCR33	CL33734Contig1	**MIRLTMVNFFTLVLLLTAISITSQHQVAD**LGFDQWIDFDCPETSLNCIADCRKRYGALARAYCTNVFCYCFY	43	6	4.78	5056.77		1
AeNCR34	CL10596Contig1	**MGQLTLANFFALVLLGIVISMISHHHVAG**DLCFEDWVDIYCIYDSNGCMQDCKNKYGKTGEGMCAYYRCLCMYE	45	7	4.24	5345.09		1
AeNCR35	CL7874Contig2	**SISMTSHHQVSG**MIGLDFWIDLECPMYNMSCVMDCKNQHGFTASGYCRDVFCYCYYQV	46	6	4.35	5465.33	signal peptide truncated	1
AeNCR36	CL2051Contig1	**MSKLTLAIFVPLVLLVTAILMSSHNSVMVMG**SIPCFDYWSNMNCPDDDDNCIQTCKDRYGDSAVGNCNDCYCY	42	7	3.69	4828.20		no
AeNCR37	CL19550Contig1	**INFGKLFAIILLVIVISTKNHHLVAG**DVCFEEWVDLNCINNDPNYCMRDCKNKYGDTGDGLCAFLKCVCLYE	46	7	4.12	5334.03	signal peptide truncated	1
AeNCR38	CL3827Contig2	**MIKLSLTNFFALVMLLTAILMMISHHEVAC**LGSDSWMDINCPGDDSICNFHCKRRYGSSAWAMCSGSFCRCYH	43	6	6.87	4868.45		1
AeNCR39	CL5370Contig2	**FAVISMKNHDKVVS**GIMTFDYWIDISCIQSAFSCMTECRLRHGMTAVGSCSDIFCLCHYEL	47	6	4.80	5359.23	signal peptide truncated	1
AeNCR40	CL305Contig5	**MIRLTLEYFFILILLVAAISMTSQHQVLD**SGYDKWIDLDCEVTILNCIADCKKRYGPIARAYCSNIYCYCFYS	44	6	5.91	5141.91		1
AeNCR41	CL5247Contig1	**MIRLTLANFFTLVLLVSAISTRSQHQVVG**HGVGFANWIDLECPRANFYCMIDCKSEYGPSAKGLCNDVFCFCYYTI	47	6	4.83	5366.15		1
AeNCR42	CL9423Contig1	**KLNLYFN*YFAVISMTSYHHVDS**IQCFDNWVDIDCINNCMSDCIEKYGFTVLGKCVGKTCYCCYE	42	8	4.23	4841.60	stop codon	1
AeNCR43	CL4722Contig2	**MTKLTLINFFMFVLLVTAMLMTSYDQVSG**EVCFRVWYDLSCYRGNSHRCSMACTNKYGDLAMGRCDRGHCNCRYESEEC	50	8	6.93	5879.59		1
AeNCR44	CL14243Contig1	**MTKLTLAYIFVLVYLVTAISMTNEVAG**EKCEEFWYEPQCKCPNYNGCMTDCKIKHGDSAKGACDYLDCYCTYD	46	8	4.56	5373.00		1
AeNCR45	CL26264Contig1	**MTKLTLAYIFAFVLLVTAISLITEVAG**EECLDYWYDRRCRRRNYKGCMRDCKDKYGDSANGFCDFLDCFCSFD	46	7	5.12	5670.31		1

AeNCR46	CL11729Contig1	**MTKLTLINFFTFALLVAGILMTSYYHQVSG**EVCFHLWSDLNCHKGDSLHCLVACTNKYGDLAKGRCDSDYCNCRYESEKC	50	8	5.80	5734.43		1
AeNCR47	CL6739Contig2	**MIKLTLANIFLLTLLMTVISRTSSCEA**AHIDGFVRWFDFSCAENDKSCHLECRLRFGATAQGFCEDVFCCCNY	46	7	4.96	5315.97		1
AeNCR48	CL3731Contig2	**YFAVIWMINRVEM**AGTGFDSWVDKNCPLENFYCIHSCRQKHGQSAMGECGGSFCYCYY	58	6	6.02	5064.64	signal peptide truncated	1
AeNCR49	CL13080Contig2	**MAKLTSTNFFALTLLISVILMISHHEA**TEMGFESWIDIDCPVSDFHCKDICGKKYGPTTSAECNEIFCYCYY	45	6	4.28	5213.86		1
AeNCR50	CL682Contig4	**MNKLNWACIFGLVFVVTAILMTNEVVG**EDCLDFWSDPRCKYSNQYICKTNCKTKHGDLAIGSCLDRVCYCAY	45	7	6.78	5230.94		1
AeNCR51	CL12848Contig1	**FTVILITSHHQVDS**IQRYDYWVDINCIYNCMIDCINKHGDTALGRCVG	34	4	5.36	3966.55	signal peptide truncated	no
AeNCR52	CL2051Contig2	**MSKLTLAIFVPLVLLVTVMVMG**SIPCFDYWSNMNCPDDDDNCIQTCKDRYGDSAVGNCNDCYCY	42	7	3.69	4828.20		no
AeNCR53	CL4958Contig2	**IFYFN*YFTAMLIISHHEVAS**FVSDSWMDINCPGDNFFCNFNCRRKYGSSAWGICNGNFCHCYY	43	6	6.70	5043.62	stop codon	1
AeNCR54	CL682Contig1	**MNKLNWACIFGLVFVVTAISMTSEVAG**EDCLDVWYEPHCEFQKSNVCKKNCKSKHGDLANGFCNASDLICY	44	6	5.44	5056.68		no
AeNCR55	CL1Contig16	**LFIYT**KDAISTSSHNPAMG**SIPCYDYWTDPNCPDDDDYCIQVCQERYGNSAVGNCNGCYCYCRYS	46	8	3.80	5336.78	stop codon	1
AeNCR56	CL10596Contig3	**MGQLTLANFFALVLLGIVISMISHHHVAG**DLCFEDWVDIYCIYDSNGCMQDCKNN	26	4	3.61	3107.42		no
AeNCR57	CL5726Contig1	**LSFN*YFAAIWIKNHGQVAG**DLCFEEWVDENCINNGSYDCMRDCKNKYGNTGDGLCAVLKCVCLYE	46	7	4.16	5247.90	stop codon	1
AeNCR58	CL4882Contig4	**FAAISMISHHQVAG**VTCIEIWTDLRCKLYDGGCWLDCLKKYGGASKGHCQDDYNCFCLYECR	48	8	5.57	5590.43	signal peptide truncated	1*
AeNCR59	CL2051Contig4	**LINIFAAILMSSHNSVMVMG**SIPCFDYWSNMNCPDDDDNCIQTCKDRYGDSAVGNCNDCYCYCRY	45	8	3.90	5250.70	signal peptide truncated	1
AeNCR60	CL2051Contig3	**TILMSSHNSVMVMG**SIPCFDYWSNMNCPDDDDNCIQTCKDRYGDSAVGNCNDCYCY	42	7	3.69	4828.20	signal peptide truncated	no
AeNCR61	CL27458Contig1	**IFALVLLVTVILMTSNYRVTS**TRSFKSWIDVQCPKEDLNCSKNCLENYGPIAWAMCNDIFC	61	5	4.86	4644.32	signal peptide truncated	no
AeNCR62	CL682Contig3	**FLALIFLLTAISMTSEVAG**EDCLDVWYEPHCEFQKSNVCKKNCKSKHGDLANGFCNASDLICY	44	6	5.44	5056.68	signal peptide truncated	no
AeNCR63	CL11599Contig2	**NMHVIFHLN*YFAVTSMTNHHQIAG**IMCTKYWFDRRCKLYDGGCWQDCLTKHGVEARG	33	4	8.49	3940.56	stop codon	no
AeNCR64	CL12457Contig1	**MVKSTLATFFVFILLVTAILMTSHHPVMA**GYCVELRPDENCLKEDPPTTCSQYCINKFGYYDTTVICDGDYCYCYYK	48	7	4.24	5657.32		1
AeNCR65	CL7951Contig1	**IFALVFLVTA**ISMTSEATGEDCLHIWYEPQCEFRNHNVCNMNCKTKHGDIAVGVCGDFLCYCAYD	65	7	4.72	6237.01	signal peptide truncated	1
AeNCR66	CL24618Contig1	**MTKLTLTNIFALVFLVSVISMISDVAG**KKCGQSWYDLQCKDPNYNGCTKDCINKYG	29	4	8.45	3373.80		no
AeNCR67	CL19676Contig1	**MIKLTLANVFSLIFLVSVISMTSEVVG**DQMCQRFWYDLKCKHPNYDGCMMECINTYDYRAYGFCKYPHCVCTY	46	7	6.86	5708.56		1
AeNCR68	CL3827Contig1	**ILHFY*YFTAILMMISHHEVAC**LGSDSWMDINCPGDDSICNFHCKRRYGSSAWAMCSGSFCRCYH	43	6	6.87	4868.45	stop codon	1
AeNCR69	CL10339Contig1	**MTKLTLANNFALILLVAVISMTSDVAG**DGKCIGKVWYDAQCQYPNYNGCVNDCRDKYGDLFEYGYCRNLYCFCF	47	7	4.86	5615.29		1
AeNCR70	CL4722Contig1	**LTKIQTCNL*FEFIFVAMLMTSYDQVSG**EVCFRVWYDLSCYRGNSHRCSMACTNKYGDLAMGRCDRGHCNCRYESEEC	50	8	6.93	5879.59	stop codon	1
AeNCR71	CL682Contig2	**FVAILMTNEVVG**EDCLDFWSDPRCKYSNQYICKTNCKTKHGDLAIGSCLDRVCYCAY	57	7	6.78	5230.94	signal peptide truncated	1
AeNCR72	CL8283Contig3	DTWIDQNCPWESFYCRLSCRTKHGGKAMGECGGGPFCYCYY	41	6	ND	ND	signal peptide absent	1*
AeNCR73	CL24495Contig1	**LSFN*YFVVISTKNHHLVAG**DVCFEEWVDLNCINNDPNYCMRDCKNKYGDTGDGLC	36	5	3.98	4163.58	stop codon	no
AeNCR74	CL1101Contig4	**MTKLIVANIFSVVLLVTAILMRSEVVD**ASSELVYDINCKDPGYNTECMQYCIDKHGYTARGSCLYNNYCYC	44	6	4.83	5073.64		1*
AeNCR75	CL1159Contig4	**MIKLTLINFFALVLLITD**DWFDENCEKGISIDCMEDCINKYGDTSNGFCEFYNCYCQYK	59	6	3.97	4895.36		1

AeNCR76	CL6739Contig3	**FAVISRTSSCEAAHIDG**FVRWFDFSCAENDKSCHLECRLRFGATAQGFCEDVFCCCNY	58	8	4.95	4822.45	signal peptide truncated	1
AeNCR77	CL6739Contig1	**LKYHVISRTSSCEAAHIDG**FVRWFDFSCAENDKSCHLECRLRFGATAQGFCEDVFCCCNY	60	8	4.96	4822.46	signal peptide truncated	1
AeNCR78	CL13080Contig1	**YFAVILMISHHEATEMG**FESWIDIDCPVSDFHCKDICGKKYGPTTSAECNEIFCYCYY	58	6	4.38	4795.39	signal peptide truncated	1
AeNCR79	CL15279Contig1	**IYKFYLNI*FVAISMTSKVDG**KECLAVWYEPQCEYPNNNVCMMNCKTNHGYLAGGVCGDIFCYCNYD	67	7	4.50	5277.99	stop codon	1
AeNCR80	CL25922Contig1	**FVAISLITEVAG**EECLDYWYDRRCRRRNYKGCMRDCKDKYGDSANGFCDFLDCFCSFD	58	7	5.12	5670.31	signal peptide truncated	1
AeNCR81	CL11729Contig2	**KIQTHNLSF*FIFAGILMTSYYHQVSG**EVCFHLWSDLNCHKGDSLHCLVACTNKYGDLAKG	34	4	6.27	3778.31	stop codon	no
AeNCR82	CL1101Contig2	**TLLFLFLIEDAAILMRSEVVD**ASSELVYDINCKDPGYNTECMQYCIDKHGYTARGSCLYNNYCYC	65	6	4.83	5073.64	signal peptide truncated	1*

Annex 3

Table 5. Sequence of BacA-related proteins

Bradyrhizobium BclA proteins

>CAL79771 / BRADO6119 (*Bradyrhizobium* ORS278)

VNNLRSTLAIVWRIAIPYFRSEDKVAGRALLAAVIAIELALVAIDVLVNQWQARFYNALQEYDWNSFIWEIGVFVVLATTFIVLAIYQLYLNQWLQIRWRRWLTETYLRHWLNNANHYRMQLKGDAADNPDQRIAEDVKQFVEQTLYITVNLLS
SVVTLASFVVILWGLSEAAPLTLFGMEYNIPGYLVWAALVYAIAGTALTHWIGSPLINLNFEQQRFEADFRFNLVRVRENSEQIAMLRGEAAEHGLLSSRYGRVVQNWYQIMTRTKRLTAVTAGYSQVATIFPYVIVAPAYFAKKVQLGGMMQT
ASAFSSVQRALSFFVNTYRTLADWRSTVARLDGFEMSIASAAKLSGEPQTIDVVSHAGDSIELAQLLLKLPNGLPLIAADDFSIRSNERTLLTGPSGAGKSTLFRAIAGVWPFGSGAISVPGHAKLMMLPQRPYFPIGALKDAVVYPAGADAFGVEPI
KQALVAVGLPQLADRMDEDGHWNRMLSLGEQQRLGIARALLHAPQFLFLDEATASLDEPSEARLYEVIAEKLPQTTVVSIGHRSTLHAFHDRRVELFRDGDRFSLRTAARAADAPAGGAE

>BRAO285v1_1320006 (*Bradyrhizobium* ORS285)

VNNLRSTLAIVWRIAIPYFRSEDKVAGRILLAAVIAIELALVAIDVLVNQWQNRFYNALQEYNWNSFIWEIGVFAVLATTFIVLAIYQLYLNQWLQIRWRRWLTETYMRHWLSNANHYRMQLKGDAADNPDQRIADDVKQFVEQTLYITVNLLS
SVVTLASFVVILWGLSEAAPLTLFGMEYNIPGYLVWAALIYAIAGTALTHWIGSPLVNLNFEQQRFEADFRFNLVRVRENSEQIALLRGESAEHDLLSNRYGRVVQNWYQIMTRTKRLTAVTSGYSQVATIFPYVIVAPAYFAKKVQLGGMMQTA
SAFSSVQRALSFFVNTYRTLADWRSTVARLDGFEMSIASAVTLSGEPQTIDVVSHAGDSIELAQLLLKLPNGLPLIAADGFSFKSNQSMLLTGPSGAGKSTLFRAIAGVWPFGSGAISVPGHAKLMMLPQRPYFPIGALQTAVVYPAAPDAFSPEQV
KDALVAVGLPLLAERLDEEAHWNRMLSLGEQQRLGIARALLHQPQFLFLDEATASLDEPSEARLYRAIAERLPQTTVVSIGHRSTLHDFHDRKVELVRDGDRFSLRPTGQAADAPAGGAE

>ABQ33877 (*Bradyrhizobium* sp. BTAi1)

MNNLRSTLAIVWRIAIPYFRSEDKVAGRSLLAAVIAIELSLVAIDVLVNQWYNRFYNALQDRNWSTFTWELGVFIVLASVSVALSVYQLYLNQWLQIRWRQWMTRVYLSQWLDRANHYRMQLKGDAADNPDQRIADDVQMFVEKTLSITIGLL
SSIVTLASFVVILWGLSEAAPLTIGGQEFAIPGYLVWAALIYAIFGTALTHWIGSPLVNLSFQQQRFEADFRFNLVRVRENSEQIALLRGEHAERGRLMDRFGSVIDNWYSIMSRTKRLTAFTASYSQAAVIFPFILTAPAYFAGKIQLGGMLQTSSAF
GSVQKALSFFVSAYRTLADWRAIVARLDGFEMSIESAATLSSEPQTVGVVDHAGDSIELAQLLLKLPNGLPLIAADGFSIKSSERTLLTGPSGAGKSTLFRAIAGVWPFGSGAISVPAHAKLMMLPQRPYFPIGTLQAAIVYPAAPDRFSVEQVKDAV
AAVGLPQLADRLDEDAHWNRMLSLGEQQRLGMARALLHGPQYLFLDEATASLDEPSEARLYRVIAERLPQTTVVSIGHRSTLHDFHDRKVELIREGDRFTLRPTTPAAAADAPTGGAE

>NP_774177 / blr7537 (*Bradyrhizobium japonicum* USDA 110)

MKNISATLAIVWRIAVPYFRSEDKWVGRGLLAAVIAMELALVAIDVLVNQWQNRFYSALQASDWDAFVTQIWIFVALASMFIALAVYKLYLNQWLQIRWRQWLTRHYLGEWLQGATHYRMQLKGDAADNPDQRITEDVKNFVEQTLTIGLG
LLSSIVTLFSFVIILWGLSNAAPLHLFGTDLMIPGYLCWGALVYAIFGTALTHWIGAPLVNLNFEQQRYEADFRFHLVRVRENSEQIALLKGEGAERGRLLGRFGLVIGNWYAIMSRTKRLTAFTASYQQAAVIFPYVLVAPAFFAKKIQLGDMMQ

TASAFSSVQGALSFFVTAYRSLAEWRSIVARLDGFEMSVDSAANLPAHEPAIALEAAGGDRNIGLEQLCVNLPNGTPLVAAGAFAIQVPERVLVTGPSGSGKSTLFRAIAGIWPFGTGTIVVPERAKLMMLPQRPYFPVGVLRDAVVYPAAPDTFD
AARVRDALIAVGLPDLAERLDEDGHWNRMLSLGEQQRLGLARALLHAPDYLFLDEATASLDEPSEARLYRLLTERLPQAAIVSIGHRSTLDAFHTRKVTMVKDGQIHVLGKSGEPVQAESTVAR

>BAF87473 (*Azorhizobium caulinodans* ORS571)

MRALLTLLRDIRRLALPYFRSEERWTAIGLLGAVIGLELAWVYCTVLLNRWNNAFYDAIQEKDFGAFKHQLLIFCAIAAGAICVAVYQIYLKQWLEIRWRRWLTKRYLEHWLGDDTHYRLRLSGDAADNPDQRIAQDVSMFVSQTISVGVGLLG
TIVSLTSFSVILWGLSGAIDLKLFGIDMHVPGYLFWAALLYAALGTLITHFIGKPLVRLNFDQQRYEADFRVDMVRVRENSEQIALLGGEPAESRKLEGRFGRILDNYFGLMKAQKRLTWFTAGYNQVSNIFPYVVVSPGYFAGAIQLGTLMQTGS
AFGSVQGAFSFFISSYSTLAEWTSVVNRLTGFEAAMEVAKRDDGARAFPHRREAAQEAALALADMDVRLPTGEAIVAVDRLSIGKGERVLVTGPSGGGKTTLFRAIAGIWPFGTGTITLPQDARVMILPQRPYVPVGRLDDALTYPLPATDFDHA
RLAEALTAVGLEVLVHRLEEQALWPHILSLGEQQRLSIARALLEQPDVLLLDEATAALDEPSEAAVYALLRARLPQATVLSIGHRATLNALHDRTLALSGTGLPRRLVDGPAFQPEAA

Other *Bradyrhizobium* sbmA_bacA domain proteins

>blr0517 (*Bradyrhizobium japonicum* USDA 110)

MQKPAGKREQGKKPPIIEIEGETGEDVAPPPPEELEPDPELSPEEAEQARKDYLLTRFWISARGFWGRNGDRLAWPFSIGLGMLIVLTVAFQYGINVWNRAIFDAIEKRDATSVFHLTAVFVPLAIGSIVLGVAQVFARMGIQRRWRAWLTASVLTR
WLANGRYYQLNLVGGDHKNPEYRIAEDLRIATDSPVDFLAGVTSALLSAVTFIVVLWTIGGALTIALGGSTITIPGFLVIAAILYAAIASSSIVVIGRRFVQISEDKNQAEADFRYTLTRVRENGESIALLGGEEEERGGIDRNFTNVLRQWARLAGQH
MRTTLVSQGSNLVAPVVPLLLCAPKFLDGSMTLGQVMQAASAFTIVQGAFGWLVDNYPRLADWNACARRIASLMMSLDGLERAEQGDGIGRIKRGETSNEAMLELNDLSVTLDDGTAVVGETEVVIEPGERLLVAGESGTGKSTLVRAIAGLW
PWGGGSVNFHPDRRLFMLPQRPYVPSGSLRRAVAYPGAEDDWTVDEIGEALHKVGLDHLKEKIEEEGPWDQTLSGGEKQRLAFARLLLHNPDIVVLDEATSALDEKSQDKMMKMVTDALPKATIVSVAHRVELEAFHSRKIVLERRKGGAKLV
SDIDLIPRKGRRRLLGRFLRPRRPAPTRAA*

>BBta_0233 (*Bradyrhizobium* sp. BTAi)

MGQVQDIARPAVSSADEAPIIEIAAENGAQVEPPPPEVVEPDPELSPEEAEQVRKRYLLTRFWISAHGYWGAAGDRLAWPFTIGLLTPIVGTVAFQYGINVWNRSIFDAIEKRDSATVFHLTAIFFPLAIGSVILAVAQVYARMAIQRRWRAWLTNSI
ITRWLASGRYYQLNLVSGDHKNPEYRIAEDLRIATDSPVDFVAGVTSALLSAATFIVVLWTIGGALTLRLGGTELTIPGFLVLAAVIYAAIASGSIVTIGRRFVQTSEDKNQAEADFRYALTRVRENGESIALLGGEAEERAGIDRTFGNVLGQWARL
AGQHMRTTLVSQGSSLIAPVVPLLLCAPKFLDGSMTLGQVMQAASAFTIVQTAFGWLVDNYPRLADWNACARRIASLMMSLDALERAERGDGLGRIKHGETTGDAMLSLNDLSVTLDDGTAVVGETEVVVDPGERLLVAGESGTGKSTLVRAI
AGLWPWGGGSINFHPDRRLFMLPQRPYVPSGSLRRAVAYPGAAEDWPVEQVADALHKVGLDHLKDRIEEEAPWDQTLSGGEKQRLAFARLLLHSPDIVVLDEATSALDEKSQDRMMQVVTSELPKATIVSVAHRAELEAFHSRKIVLERRKGG
AKLVSDIDLIPRKGKRRLIGRFLRQRKQL*

>BRADO0237 (*Bradyrhizobium* ORS278)

VGSGYRDMGQIQDIERPVTSEPVKGPIIEIAAENGDQVEPPPPEVVEPDPELSPEEAEQVRKRYLLTRFWISARGYWGRAGDRLAWPFTIGLLMLIVGTVAFQYGINVWNRSIFDAIEKRDAATVFHLTAIFFPLAIGSVILAVAQVYARMAIQRRWR
AWLTNSVITRWLTSGRYYQLNLVSGDHKNPEYRIAEDLRIATDSPVDFVAGVTSALLSAATFIVVLWTIGGALTLRLGGSELTIPGFLVIAAVVYAVIASGSIMAIGRRFVQTSEDKNQAEADFRYTLTRVRENGESIALLGGETEERAGIDRTFGNV
LQQWARLAGQHMRTTLVSQGSSLIAPVVPLLLCAPKFLDGSMTLGQVMQAASAFTIVQTAFGWLVDNYPRLADWNACARRIASLMMSLDALERAERGDGVGRIKRGETTGEAMLSLNDLSVTLDDGKAVVGETEVVVDPGERLLVAGESGTG
KSTLVRAIAGLWPWGGGSINFHPDRRLFMLPQRPYVPSGSLRRAVAYPGAAEDWSVEQVGEALHKVGLDHLKDRIEEEAPWDQTLSGGEKQRLAFARLLLHGPDIVVLDEATSALDEKSQDKMMQMVTNELPKATIVSVAHRAELEAFHSRKI
VLERRKGGARLVSDIDLIPRKGKRRLIGRFLRQRKAV*

>BRAO285v1_250005 (*Bradyrhizobium* ORS285)

MGQIQDIERPVASEPPKAPIIDIAAENGEHVEPPPPEVVEPDPELTPEEAEQVRKRYLLTRFWISARGFWGKSGDRLAWPFTLGLLVLIIGTVVFQYGINVWNRSIFDAIEKRDAGTVFHLTAIFFPLAIGSVILAVAQVYARMAIQRRWRAWLTNSVI
TRWLTSGRYYQLNLVSGDHKNPEYRIAEDLRIATDSPVDFVAGVTSALLSAATFIVVLWTIGGALTLHLGGTELTIPGFLVIAAIVYAAVASGSIMTIGRRFVQTSEDKNQAEADFRYTLTRVRENGESIALLGGETEERAGIDRTFGNVLQQWARL
AGQHMRTTLVSQGSSLIAPVIPLLLCAPKFLDGSMTLGQVMQAASAFTIVQTAFGWLVDNYPRLADWNACARRIASLMMSLDALERAERGDGIGRIKRGETTGEAMLSLNDLSVTLDDGTAVVGETEVVVDPGERLLVAGESGTGKSTLVRAIA
GLWPWGGGSINFHPDRRLFMLPQRPYVPSGSLRRAVAYPGAAEDWSVEQVGEALHKVGLDHLKDRIEEEAPWDQTLSGGEKQRLAFARLLLHSPDIVVLDEATSALDEKSQDKMMQMVTGELPKATIVSVAHRAELEAFHSRKIVLERRKGGA
KLVSDIDLIPRKGKRRLIGRFLRQRKAV*

>BRADO1574 (*Bradyrhizobium* ORS278)

MSKVPQPFAEPGDLRSAGPAEDDNEDGDLQWSRFWLTARGFWRGAKAWRVWLLCAVLVGIVALQLYVQLRFNTWTRDFFNALEGRDPARLRQQAMLLVPLCAASVVLAIASVWGRMTIQRNWRQWLAAEVIDYWIENDRYARLATTKGD
QKIPEYRIAEDVRIATDAPVDFAVNVVSSLLTAVIFLQVLWQVGGPISFAISGLELWLPGYLVISVVAYSGLVTGAMLWVGAPLTHVIQVKNQSESELITAAHRLRDIGEGVTPKEDKRGVIAALWQALDRVIRQWRRLCWRLMRMTLVSHTNSL
LAPIIGLVLCAPKFLDGQMTLGELTQAAAAFTLVQGAFNWLVDNYSRVADWMSSLERVGGLLLSLDELNDDGVETVTPSAAAREPADG*

>BRAO285v1_950010 (*Bradyrhizobium* ORS285)

MTKVPRRDPFAADEPDDLRSAGLSEDDNEDSNREWSRFWLTARNFWCGWPAWRVWLLCAILVGIVMLQLYVQFRFNTWNRDFFNALEGRDPARLREQAMLLAPLCLASVALAVASVWGRMAMQRNWRQWLAAKVIDYWIENDRYARLAL
VQGDQKIPEYRIAEDVRIATDAPVDFAVNVISSLLTAIIFLQVLWQVGGPIGFSFAGYHVWIPGYLVVSVAVYSGLVTGAMLWVGSPLTRVIQVKNQAESELITAAHLLRDIGEGVSPKDDTRGVIAALWQALDRTIRQWRRLCWRLMRMTLVSH
TNSLLAPIIGLVLCAPKFLSNEMSLGELTQAAAAFTLVQGSFNWLVDNYSRVADWMSSLERVGGLLLSLDLLNDVAEVDVRRSEAARQAADG*

>blr4660 (*Bradyrhizobium japonicum* USDA 110)

MLSNETYHSPGERQLLLRFWKSARGFWKGKSAGWAWLLTVLLVATVLLQLLTQYSLNFWNRDFFNAVERKDGKELLSQALRFMPLAAASLSLAVFSVWGRMTLQRKWRAWLSDELYRYWLERDRFVRLNFVAGDYQAPEYRIAEDCRLAT
DLPVDLVLGLVASLLTAATFIGILGVVGGNLTIDAAGLTLTIPGYLVVAVVVYSIAVAAVTMLIGRRLTDVLEENKRAEAQLRAVGTHVRESGESMAPGMKGDDGIRAIHPALKAVIASWLNYCWQLVRLTIITHTNSLLTPVIGVLLCMPKYVAG
TMLLGEVVQAAAAFVVVQSACSWFTDNYPRLAEWAASANRVASLLFALDKLDRPAVDNEPGRIHKYAGTNRQT*

BacA-clade

>CAA51918 / BacA (*Sinorhizobium meliloti* Sm1021)

MFQSFFPKPKLFFISSAVWSLLAVLAWYAGGRDIGAYLGLPPLPPGQEPVIGVSVFWSTPFLWFYIYYAVVAGLFAAFWFAYSPHRWQYWSVLGTALIIFNTYFSVQVSVAINAWYGPFYDLIQQALPRTAPVTAGQLYSGMIGFSGIAFVAVTVG
VLNLFFVSHYIFRWRTAMNEFYVAHWPRLRHVEGASQRVQEDTMRFSSTVERLGVGLVSSIMTLIAFLPVLFKFSEQVNVLPIVGEIPHALVWAAVFWSVFGTVFLAAVGIKLPGLEFRNQRVEAAYRKELVYGEDHEDRADPITLAQLFDNVRR
NYFRLYFHYMYFNIARIFYLQADNLFGTFVLVPAIVAGKLTLGVMNQVLNVFGQVRESFQYLVNSWTTIVELLSIYKRLKAFESVLVDEPLPEIDRQFIDAGGKEELAL

>YP_005189568 / sbmA (*Sinorhizobium fredii* HH103)

MFQSFFPKPKQFFISVVIWSLIAVAFWYGWGERIGAFVGLPPAPADAPPIVGISAFWSPAFLWFYIYFGLVVAIFAAFWQVYSPHRWQSWSIWGSALILFVTYFQVQVSVAINNWYGPFWDLIQAAVSKAAVVTAEEFYAQIGTFLGIAMVAVAVA
VMTRFFVSHYIFRWRTAMNEYYMANWGKLRHIEGASQRVQEDTMRFSSTVEGLGVSLIDSVMTLIAFTPVLIRLSENVTELPIVGSIPYPLVTAAVLWSLFGTVFLALVGIKLPGLEFRNQRVEAAYRKELVYGEDHVDRAQPETVAELFSNVRM
NYFRLYFHYLYFNIARIFYLQINNIFSLLILAPSIIAGKISLGALNQISGAFGQVSSSFQYLVNSWPTIVELLSIYKRLRAFESVLDEQPLPEIDQQFIEAGGQEELAL

>ACL14650 / BacA (*Mesorhizobium huakuii* 7653R)

MFVSFFPQPKLFFSSAAIWSLAAILFWFFGGEQLGAVFGLPAAAADAPPIIGIAVLVSKPFLWFYIYFVACVAIFYAFWSWYSPHPWQRWSILMTAVILFFIYFNVQVSVAVNAWYGPFFDYVQGLMSGTGKSTNGEFYTGLADFSWLALVGMNV
QVVNAFIVSHWIFRWRTAMNNYFIENWARLRHIEGASQRIQEDTMRFSQIMEDLGSSFVQSIMTLIAFLPVMIQLQAHISELPIIGAIPQPLVVAALAWCLFGTISVMIAGLKLPGLQFRNQRVEAAYRKELVYGEDHVDRAQPATTAELFANVRHN
YFRLYFHYIYFNVVRYTYLQADNIFSLPILGPSLVAHKMTFGALNQISNAFGKVTGSLQFLLTSWSTIVELQSVHKRLRAFEATLLGEPLPDIDQRYLARQGAEDPA

>Mesci_5099 (*Mesorhizobium ciceri biovar biserrulae* WSM1271)

VFVSFFPQPKLFFTSAAVWSLAAILFWFFGGEQLGTVFGLPPAAAGAPPIIGIAILVSKPFLWFYLYFAFCVLVFYAFWSWYSPHPWQRWSILMTAVILFFIYFNVQVSVAVNAWYGPFFDYVQGLMSGTGKSTNSEFYIGLADFSWLALVGMNVQ
VVNAFIVSHWIFRWRTAMNNYFVENWGRLRHIEGASQRIQEDTMRFSQIMEDLGSSFVQSIMTLIAFLPVMIQLQAHITELPIIGAIPQPLVVAALAWCLFGTISVMVAGLKLPGLQFRNQRVEAAYRKELVYGEDHADRAQPATTAELFENVRHN
YFRLYFHYIYFNVVRYTYLQADNIFSLLILGPSLVAHKITFGALNQISNAFGKVTGSLQFLLTSWSTIVELQSVHKRLRAFEATLQGEPLPDIDQRYLARQGAEDPA

>NP_107725 (*Mesorhizobium loti* MAFF303099)

MFVSFFPQPKLFFSSAAIWSLAAILFWFFGGEQLGAVFGLPAAAADAPPIIGIAVLVSKPFLWFYIYFVACVAIFYAFWSWYSPHPWQRWSILMTAVILFFIYFNVQVSVAVNAWYGPFFDYVQGLMSGTGKSTNGEFYTGLADFSWLALVGMNV
QVVNAFIVSHWIFRWRTAMNNYFIENWARLRHIEGASQRIQEDTMRFSQIMEDLGSSFVQSIMTLIAFLPVMIQLQAHISELPIIGAIPQPLVVAALAWCLFGTISVMIAGLKLPGLQFRNQRVEAAYRKELVYGEDHVDRAQPATTAELFANVRHN
YFRLYFHYIYFNVVRYTYLQADNIFSLLILGPSLVAHKITFGALNQISNAFGKVTGSLQFLLTSWSTIVELQSVHKRLRAFEATLLGEPLPDIDQRYLARQGAEDPA

>NP_454971 (*Salmonella enterica subsp. enterica serovar Typhi str.* CT18)

MFKSFFPKPGPFFMSAFVWALIAVIFWQAGGGDWVARLVGASDEVPISAARFWLLDYLIFYAYYLICVGLFATFWFIYSPHRWQYWSILGTSLIIFVTWFLVEVGVAVNAWYAPFYDLIQTALNSPHKVTLGQFYHEVGVFLGIALIAVVIGVLNN
FFVSHYVFRWRTAMNEHYMAHWQYLRHIEGAAQRVQEDTMRFASTLENMGVSFINAIMTLIAFLPVLVTLSAHVPNLPIVGHIPYGLVIAAIVWSLMGTGLLAVVGIKLSGLEFKNQRVEAAYRKELVYGEDDASRATPPTVRELFSAVRHNYF
RLYFHYMYFNIARILYLQVDNVFGLFLLFPSIVAGTITLGLMTQITNVFGQVRGSFQYLINSWTTLVELMSIYKRLRSFERQLDGQPVQEVTHSFS

>ABC91878 (*Rhizobium etli* CFN 42)

MFHSFFPQPKPFFISLVVWTLISIFGWYFFAAGLGASLGFVPVPEEQQPIDLSFFLLPENLWFYGYFLASALIFCGFWHLKAMSHPWKIWSIWGSALIIFVTYFSVQISVVINNWRRPFGDLLQNALAKKPGITVENFDDLFLIFAQIAFLAMFVSIMTD
FFTSHYIFRWRTAMNDFYMANWEKLRHIEGASQRVQEDTMRFSSTLEGLGINLINSVMTLVVFLPIMMGLSHYVTSLPILGEVPNSLFWLAIFWSAFGTVLLAVAGVKLPGLNFKNQRVEAAYRKELVYGEDHADRAQPPTVQELFSAVRKNYF
TLYFHYMYFNVARSFYLQADNLFVYFFMAPTLVAGAITYGIFQQIATAFGQVSSSFQYLVNSWTTIIELLSIHKRLKAFEATIDNEPLPEIDQRYLDREAGILHADG

>YP_004557140 (*Sinorhizobium meliloti* AK83)

MTPPKRECRPLFQSFFPKPKLFFISSAVWSLLAVLAWYAGGRDIGAYLGLPPLPPGQEPVIGVSVFWSTPFLWFYIYYAVVAGLFAAFWFAYSPHRWQYWSVLGTALIIFNTYFSVQVSVAINAWYGPFYDLIQQALARTAPVTAGQLYSGMIGFS
GIAFVAVTVGVLNLFFVSHYIFRWRTAMNEFYVAHWPRLRHVEGASQRVQEDTMRFSSTVERLGVGLVSSIMTLIAFLPVLFKFSEQVNVLPIVGEIPHALVWAAVFWSVFGTVFLAAVGIKLPGLEFRNQRVEAAYRKELVYGEDHEDRADPIT
LAQLFDNVRRNYFRLYFHYMYFNIARIFYLQADNLFGTFVLVPAIVAGKLTLGVMNQVLNVFGQVRESFQYLVNSWTTIVELLSIYKRLKAFESVLVDEPLPEIDRQFIDAGGKEELAL

>AF244996 / BacA (*Brucella abortus*)

MFASFFPRPKLFFISALLWTLLAVLGWYMGGENLGSVFGLAPAASDAAPLIGAHVFWSRPFLWFYIYFAVITFAFYGFWVRFSPHPWQRWSILGSALILFATYFNVQVSVAINTWYGPFYDMVQKGLTTPGAVSAAEFYWGLADFAGIAFLAITIG
VLNLFFVSHYIFRWRTAMNDYYMSHWPKLRHIEGAAQRVQEDTMRFSTTLEQLGVSLVKSVMTLIAFLPVLFTFSQKVSTLPIIGYVPHALVWAAVIWALFGTGFLALVGIKLPGLEFNNQRVEAAYRKELVYGEDHADRAEPITMRELFKNVRH
NYFRLYFHYVYFNVARIFYLQVDNIFPLIILIPSIVAGKLTLGLMSQITNVFDQVRGSFQYLVNSWTTIIELLSIYKRLRAFEAAIHGEPLQAVDRINLEPGAG

>YP_003883254 / SbmA (*Dickeya dadantii* 3937)

MFKSFFPNPKLFFSSALIWSLVAVIIWYGWGRPLGDSLLHISQPLPQGAIRFLSPAFLWFYGYYLLCVGLFAGAWAWLSPHPWQRWSILGTSLIIFVTYFSVEVGVAVNDWYVPFYDLIQKALSAPNAVTIQAFYQQLLIFLGIALTAVTVGVLNLFF
VSHYVFRWRTAMNDYYTSHWQALRHIEGAAQRIQEDTMRFASTVESLGVDLVKSLMTLIAFLPVLVSLSSHVKSLPLIGEIPYGLVIAAVIWSLAGTGVLALIGIKLPGLEFNNQRVEAAYRKELVYGEDDASRATPPTVKQLFGNVRKNYFRLYF
HYTYFNIARVLYLQTDNIFGIIMLLPSIVAGSLTLGLMTQITNVFDQVRGSFQYLINSWTTIVELMSIYKRLRSFESVIQDVPMTTDAADATSGA

>CAA38092 / sbmA (*Escherichia coli* K-12)

MFKSFFPKPGTFFLSAFVWALIAVIFWQAGGGDWVARITGASGQIPISAARFWSLDFLIFYAYYIVCVGLFALFWFIYSPHRWQYWSILGTALIIFVTWFLVEVGVAVNAWYAPFYDLIQTALSSPHKVTIEQFYREVGVFLGIALIAVVISVLNNFFV
SHYVFRWRTAMNEYYMANWQQLRHIEGAAQRVQEDTMRFASTLENMGVSFINAIMTLIAFLPVLVTLSAHVPELPIIGHIPYGLVIAAIVWSLMGTGLLAVVGIKLPGLEFKNQRVEAAYRKELVYGEDDATRATPPTVRELFSAVRKNYFRLYF
HYMYFNIARILYLQVDNVFGLFLLFPSIVAGTITLGLMTQITNVFGQVRGAFQYLINSWTTLVELMSIYKRLRTFEHELDGDKIQEVTHTLS

>YP_001847700 (*Acinetobacter baumannii* ACICU)

MFKSFFPSPRYFFISAVIWLALNMVLWYTGGDHWGQYLGFPQGYADAELPIGVSRFWSPAFLWFYLWFLVSTALFASFWKIISNNPWQRWSIWGSAFILFNIWFSVQVSVAINAWYVPFWDLIQQMLSSGGGDLSALYSETLVFLYIAMVAVTLA
VINAFFTSHYVFRWRTAMNEYYTEHWEKLRHIEGASQRVQEDTMRFATILEDLGVELVKAVITLIAFLPILFQLSKHVPVLPIVGELEHSLVWAAIVWSIFGTVLLMVVGIKLPGLQFNNQKVEAAYRKELVYGEDHADRAKPATLRELFSNVRKN
YFRLYFHYAYFNMTAIWYGQLDILYNLVVLFPSIAAGKLTLGLIQQIANVFGRVRESFQYLITSWKTIIELLSIYKRLKAFESILHK

>ACP26136 (*Sinorhizobium fredii* NGR234)

MAEGTEPIGATYFVTPDKLYFYLYFLLCGGIFGGAWRFYDREHRWFNWSVWGSIFIIFSTYFGVLVSLAVNNWRRPFFDLLQAALEHKPGVTAGDFYTLLIKFSEIGAVAVFVFVVTRFLISHYVFRWRTAMNEFYMSQWARLRHIEGASQRVQE
DTMRFADISEGLGVNLIDAVMTLVAFLPLLLSLSKHVPELPILGAVPYSLFWLALVWSMFGTVLLAVVGVKLPGIAFRNQRVEAAYRKELVYGEDDESRATPPTVAELFDNVRRNYFRMYFHYLYFNMVRSIYNQADSVFVYFFLVPTFVAGTIT
MGILQQILTAFGQVSSSFQYLVNSWTTIIELLSIHKRLRAFESILYEQPLPEIDQQFIKAGGQEELAL

>CAK09045 (*Rhizobium leguminosarum* bv. *viciae* 3841)

MFHSFFPQPKAFFTSLAIWTLAAIFGWYFFAAGLGASLGFAAVPEEQQPIDLSFFLLPENVWFYSYFLLAAVIFCGAWHLKALNHPWKIWSIWGSALIIFVTYFGVQISVVINNWRRPFGDLLQNALSKQPGISVDNFYSLMWVFCQIAFLSMFVSIM
TDFFTSHYIFRWRTAMNNFYMSKWEKLRHIEGASQRVQEDTMRFSSTLEGLGISLINSVMTLVVFLPILLALSHYVTELPFIGPVANSLFWLALFWSAFGTLLLAVAGVKLPGLNFRNQRVEAAYRKELVYGEDHAERAEPLTVRELFGNVRRNY
YRMYFHYMYFNVARYFYIQADALFVVFMLVPTIVAGTITYGIFQQISTAFGQVSNSFQYLVNSWTTIIELLSIHKRLKAFEAAIDDEPLPEIDQRYLERDAGVVHADG

>NP_355261 (*Agrobacterium tumefaciens str*. C58)

MPEQEPYDLSYFLVPANLFFYGYLASCLALFGVVWTVIAKDHPWKLWSIWGSLLIIAVTYFGVQISVVINNFRRPFGDLLQNALSKQPGIEVWDFYSLQIVFAKIAFLSMAVSILTDFFTSHYIFRWRTAMNDFYMSKWEKLRHIEGASQRVQEDT
MRFSSTLEALGITLINSVMTLVVFLPILFALSSYVSELPIIGEVPHALFWLAIVWSIFGTVLLATVGIKLPGLNFRNQRVEAAYRKELVYGEDHADRAQPPTVKELYSNVRKNYFRMYWHYLYFNVARYFYIQADAVFLLLMLVPTIVAGKITYGIY
QQIATAFGQVSNSFQYLVNSWTTIIELLSIHKRLVAFEAAIDDKPLPDIDERYLQREAEAGELAANKP

Mycobacterium clade

>YP_906777 (*Mycobacterium ulcerans* Agy99)

MGPKPFTPSIDWSSAFVDSVFWVAKAWAISAVCVIAALALFKFFTPWGRQFWRITSEYFIGPKSIRVWLMLGVLLFSVVLAVRLNVLFSYQGNDMYTALQKAFQGTAASDEAVKRSGVRGFWMSIGVFSVMAVTHVARVMADIYLTQRFIIAWR
VWLTDHLTKDWLKGRAYYRDLFIDETIDNPDQRIQQDIDVFTAGAGGTPNQPSNGTASMLLFGAVQSIISVISFTAILWNLSGTLNIFGLSFPRAMFWTVLIYVLVATIISFVIGRPLIWLSFRNEKLNAAFRYALVRLRDAAEAVGFYRGERVEGAQ
LERRFLPVIRNYRRYVRRSIAFNGWNLSVSQTIVPLPWVIQAPRLFAGQIDFGDVGQTATAFGNIHDSLSFFRNNYDAFASFRAAIIRLHGLVDANEKGRALPAVLTKPSSDESVELTDVEVRTPAGDRLIDPLDVRLDRGDSLVITGRSGSGKTTLL
RSLAELWPFASGTLSRPDGANDTMFLSQLPYVPLGSLRVVVCYPNSPYDIPDEAVRDTLTKVALAPLCDRLDEERDWAKVLSPGEQQRAAFARILLTKPKAVFLDESTSALDTGLEFALYELLRNELPDCIVVSVSHRPALERLHNQQLELLGGGP
WRLGPVNQEPAPA

>YP_888655 (*Mycobacterium smegmatis str*. MC2 155)

MNDLEPFSPSIDWSDELVHSLWWLAQAWTIAALCTLAVLVLLARYTTWGRQFWTVTGVYFTGRRSLRVWLVLGAMLLSVIIGVRLSVLFSYQSNDLYSAAQIAVQGYATGDDAVKDSGVRGFWISLLTFSLLAAILVTRVLVDLFMTQRFMLA
WRTWLTDRLTGDWLDGRAYYRGRFIDQTIDNPDQRIQSDIDIFTALSGPQPNTPHQTSNGTLPFGAISSIVSVVSFTSILWNLSGELDVFGVQIPRAMFWSVFVYVAFATVIAFWLGRPLIRLSFNNEKFNAAFRYALVRLRDAAESVALYRGEKAE
RRQLRAFFAPVVDNYKRFVNRTIVFTGWNLSMNHIIIPLPWLLQAPRLFAGQIQLGAVTQSVTAFGAIQDALSFFRNSYDTFAGYRASIMRLHGLVTADEQGRALPMVHVENARTAVVELDDVEVRNPAGEELVTGLSLRLHPGAAMIITGKSGT
GKTTLLRSLAQLWPYASGTMRCPQGRNETLFLSQLPYVPLGDLRTVVSYPHEPGDLPDDALRDALLAVALPRHADRLSEVNDWAKVLSPGEQQRIAFARVLLTRPKVVFLDEATSALDEPLEFMIYGLVRRELPDTVFVSVTHRTTVNRHHEQRL
ELLGGGRWRLGPVDEAAPVPV

>YP_001850994 (*Mycobacterium marinum*)

MGPKPFTPSIDWSSALVDSVFWVAKAWAISAVCVIAALALFKFFTPWGRQFWRITSEYFIGPKSIRVWLMLGVLLLSVVLAVRLNVLFSYQGNDMYTALQKAFQGTAASDEAVKRSGVRGFWMSIGVFSVMAVIHVARVMADIYLTQRFIIAWR
VWLTDHLTKDWLKGRAYYRDLFIDETIDNPDQRIQQDIDVFTAGAGGTPNQPSNGTASTLLFGAVQSIISVISFTAILWNLSGTLNIFGLSFPRAMFWTVLIYVLVATIISFVIGRPLIWLSFRNEKLNAAFRYALVRLRDAAEAVGFYRGERVEGAQ
LERRFLPVIRNYRRYVRRSIAFNGWNLSVSQTIVPLPWVIQAPRLFAGQIDFGDVGQTATAFGNIHDSLSFFRNNYDAFASFRAAIIRLHGLVDANEKGRALPAVLTKPSSDESVELTDVEVRTPAGDRLIDPLDVRLDRGDSLVITGRSGSGKTTLL

RSLAELWPFASGTLSRPDGANDTMFLSQLPYVPLGSLRVVVCYPNSPYDIPDEAVRDTLTKVALAPLCDRLDEERDWAKVLSPGEQQRVAFARILLTKPKAVFLDESTSALDTGLEFALYELLRNELPDCIVVSVSHRPALERLHNQQLELLGGGP
WRLGPVNQEPAPA

>YP_001704129 (*Mycobacterium abscessus* ATCC 19977)

MKTATDWGTELWRSLVWIGWVWPLTAAIFVTIAFLIIQFTQWGRQFWRITGQYFTTRAGRGGLVLLGLVLLSVVVNVRLSVVFTYYYNDMSASIQYIVQALARDGQGMPEAKAFFWLNIRRFCLLAGINIAVVVTDFYLLQAFIIRWRVWLTERI
TVDWLSKRAFYRSRFIDNTIDNPDQRIQADINNFVAMSSAPEQLQSSSLHLAFGAVNACISLPSFTIILWNLSGPLNIFGYEMPRALVFLTYIYVIVTTLIAFRIGRPLIRRFFLNERLNAMFRYALVRLRDTSESVAFYRGERAESKQLKSRFDAIIANF
WQLVYRSMGFTGWNFAVSQTAAVFNSVLQVPRLIAGQITYGDVSQSASAVGQLQTSLSFFRNAYDNFAAYRATVMRLDGLLTADSQSRELPMLEPGDQEDGVTLNNIYVRKPNGDSLIDDLNLDLAAGDALLIKGQSGSGKTTLLRSLAGLWP
YTDGDWARPGGDHETMFISQLPYLPLGNLRDALSYPAESGSFSDEELLQTLEKVSLAHLADRLDEEADWIKVLSPGEQQRVAFARILLIKPKAVFMDESTSSLDEGLEFSLYRLLRSELPGCTLVSVGHRSTIDQHHNQKLQLDAEGRWSLSPL

>CAC31039 (*Mycobacterium leprae*)

MELRPFYPSVNWGYALPDSLWWIAKAWAISAACVVVLVMLSFITRWGRQYWRITRGYFIGPQSVGVWLMLAVLLFSVVISVRLNVLFSYQSKDLYNALQAAFEGAGAQNDLVKQSGMHGFWMSLGIFSILAVIFIARVMADIYLTQRFIIAWRM
WLTNHLTTDWLDQRAYYREQFIDNTIDNPDQRIQQDIDIFTAGAGGTPNGPSNGTGSTLLFGAVESVISVISFTAILWNLSGRLAVFEFEIPRAMFWIVIVYVLLATIITFWIGRPLIGLNFANEKLNAAFRYALVRLRDAAEAVAFYRGEEAERKQLE
QRFTPIIDNYRRYVSRTIRFLGWNASVSQTIVPLPWILQAPRLFAGKINLGDVGQTATAFGNIHDSLSFFRNNYAAFAAFRAAIIRLHGLVVANARSRELPHVTAKSTNDGTVHVVAVEVRTSAGEQLIDPLDLQLNCGSSLVITGRSGVGKTTMLRS
LAELWPYASGTLSRPDGKNETMFLSQLPYVPLGNLRGVVCYPNAPADIPEDTVRDTLTKVALNPLCDRLDEERDWAKILSPGEQQRIAFARILLTKPKAIFLDESTSALDEGLEFALYQMLRSELPDCIIVSVSHRRAVELLHAQQLELLGGGQWRL
GPVEKEPLHV

>YP_001134674 (*Mycobacterium gilvum* PYR-GCK)

METFTPTLDWGNELWTSLWWIAQGWVYAAIATMVVLTLIVRFTTWGRQFWRVTRGYFTGRDSVVVWLWLAGILLLVIASVRLSVLFSFQGNDMMTSFQVIASGVGAGDDAVRASGRDGFWLSMLVFSVLAVVNVVTIMIDLLVTQRFMLRW
RAWLTDQLTGDWLDGRAYYRSRFIDDSIDNPDQRIQMDIDIFTTGVGPLPNRPNNTSGTTLLFGAISSIAAMISFTTILWNLSGPVTLPFIGYELPKAMFWIGIVYILFATVVAFWIGRPIITLAFRNEKFNAAFRYALVRLRDAAEAVAFYRGEIAERT
GLRKLFAPVVDNYKRYVNRMAGFLGWNLSITQAQELIPYIVQFSRFYNGEITLGQLSQTASAFREILSGLSFFRNAYDDFAGYRAAIIRLHGLVVANEEGRALPSLDTQDSTDGRVELDDVEVRTPDGRQLLQPVDLRLEPGDGLVITGPSGSGKTT
LLRSLGRLWPFASGTLVYPAGENDTMFLSQLPYVPLGDLRAVVSYPNEPGSIPDGTLREVLEKVALPHLADRLDEEQDWAKVLSPGEQQRVAFARILLTKPEAVFLDESTSALDEGLELMLYRLVRSELPQTIVVSVSHRSTVEQHHSRQLKLLGD
GRWEVGAVQSR

>NP_216335 / Rv1819c (*Mycobacterium tuberculosis* H37Rv)

MGPKLFKPSIDWSRAFPDSVYWVGKAWTISAICVLAILVLLRYLTPWGRQFWRITRAYFVGPNSVRVWLMLGVLLLSVVLAVRLNVLFSYQGNDMYTALQKAFEGIASGDGTVKRSGVRGFWMSIGVFSVMAVLHVTRVMADIYLTQRFIIAW
RVWLTHHLTQDWLDGRAYYRDLFIDETIDNPDQRIQQDVDIFTAGAGGTPNAPSNGTASTLLFGAVQSIISVISFTAILWNLSGTLNIFGVSIPRAMFWTVLVYVFVATVISFIIGRPLIWLSFRNEKLNAAFRYALVRLRDAAEAVGFYRGERVEGT
QLQRRFTPVIDNYRRYVRRSIAFNGWNLSVSQTIVPLPWVIQAPRLFAGQIDFGDVGQTATSFGNIHDSLSFFRNNYDAFASFRAAIIRLHGLVDANEKGRALPAVLTRPSDDESVELNDIEVRTPAGDRLIDPLDVRLDRGGSLVITGRSGAGKTTL
LRSLAELWPYASGTLHRPGGENETMFLSQLPYVPLGTLRDVVCYPNSAAAIPDATLRDTLTKVALAPLCDRLDEERDWAKVLSPGEQQRVAFARILLTKPKAVFLDESTSALDTGLEFALYQLLRSELPDCIVISVSHRPALERLHENQLELLGGG
QWRLAPVEAAPAEV

>YP_977945 (*Mycobacterium bovis BCG str. Pasteur* 1173P2)

MGPKLFKPSIDWSRAFPDSVYWVGKAWTISAICVLAILVLLRYLTPWGRQFWRITRAYFVGPNSVRVWLMLGVLLLSVVLAVRLNVLFSYQGNDMYTALQKAFEGIASGDGTVKRSGVRGFWMSIGVFSVMAVLHVTRVMADIYLTQRFIIAW
RVWLTHHLTQDWLDGRAYYRDLFIDETIDNPDQRIQQDVDIFTAGAGGTPNAPSNGTASTLLFGAVQSIISVISFTAILWNLSGTLNIFGVSIPRAMFWTVLVYVFVATVISFIIGRPLIWLSFRNEKLNAAFRYALVRLRDAAEAVGFYRGERVEGT

QLQRRFTPVIDNYRRYVRRSIAFNGWNLSVSQTIVPLPWVIQAPRLFAGQIDFGDVGQTATSFGNIHDSLSFFRNNYDAFASFRAAIIRLHGLVDANEKGRALPAVLTRPSDDESVELNDIEVRTPAGDRLIDPLDVRLDRGGSLVITGRSGAGKTTL
LRSLAELWPYASGTLHRPGGENETMFLSQLPYVPLGTLRDVVCYPNSAAAIPDATLRDTLTKVALAPLCDRLDEERDWAKVLSPGEQQRVAFARILLTKPKAVFLDESTSALDTGLEFALYQLLRSELPDCIVVSVSHRPALERLHENQLELLGGG
QWRLAPVEAAPAEV

>YP_953972 (*Mycobacterium vanbaalenii* PYR-1)

METFTPTLDWGNELWTSLWWIAQGWAYAAVATIVVLALIVRFTTWGRQFWRVTRGYFTGPESLIVWLWLAVILLLVILSVRLSVLFSFQGNDMMTSFQVIASGVGAGDDAVKQSGGDGFWMSMGVFAVLAVINVVSIMVDLLLTQRFMLRW
RTWLTDRLTGDWLDGKAYYRSRFIDDTIDNPDQRIQADIDIFTAGVGPLPNMPNNTSGSTLLFGAVSSIAAMISFTTILWNLSGPVTLPFVGFELPKAMFVIGIVYILFATVVAFVIGRPIITLSFNNEKFNAAFRYALVRLRDAAEAVAFYRGEIAERT
GLRRRFGPIVENYKKYVNRMAGFLGWNLSISQAQELIPYIVQFPRFFSGEITLGQLSQTAGAFREILTGLSFFRNAYDQFAGYRAAIIRLHGLVVANEEGRELPTLTVEPSSGNTVELDDVEVRKPDGTQLIDPIDLRLEAGDTLAVTGVSGTGKTTLL
RSLAQLWPYTTGTLRCPQGTNETMFLSQLPYVPLGDLRAVLSYPRQVGDIPDSELVAVLNKVALPHLVSRLDEEQDWAKVLSPGEQQRVAFARILLTRPKAAFFDEATSALDVGLETMLYQMVRDELPDTILVSVAHRGTVIRQCEKELALLGD
GRWRLGLAGADAP

>NP_960465 (*Mycobacterium avium subsp. paratuberculosis* K-10)

MGAKPFKPSIDWSAAFVDSLRWLAIAWVISAVCLFGVLLAFRFLTPWGRQFWRITRGYFVGAHSVRVWLMLGVLLLSVLLSVRLSVLLSFQSNDLYTALQKAFEGIASGNDAVKHSGIHGFWVSLGIFCLLAALWVTRFMADVYLTQRFIIAWR
MWLTANLTDDWLAGRAYYRDLFIDNTIDNPDQRIQQDIDIFTANAGGTPNVPSNGTGSTLLFGAVNAVASVISFTAILWKLSGDLNLFGLVLPRAMFWTVLVYVFIATVVAVWLGRPLIWLSFNNEKLNAAFRYALVRLRDAAEAVGFYRGERV
ERAQLWQRFTPIIANYRRYVRRTIIFNGWNWSVTQIIVPLPLAIQAPRLFAGEIAFGDVTQTATAFNNIHDSLSFFRNNYDAFAAFRAAIIRLHGLVDANSKGRALPAILVKPSEETAVELRGIEVRTPEGDQLVDSLDIQLDEGDTLVITGRSGAGKTT
LLRSLAELWPYASGTLCRPDGDNATMFLSQLPYVPLGTLRTVVCYPNSPDSIADSQLHEVLTKVALAPLISRLDEDQDWAKVLSPGEQQRVAFARVLLTRPKAVFLDEATSALDEGLEYALYQLVRAELPGCVMVSVSHRPTVEQHHDQQLHLL
GGGPWQLSPVEKEPARV

ExsE clade

>P31826 / YddA (*Escherichia coli*)

MITIPITLRMLIAKYLCLLKPFWLRKNNKTSVLLIIIILAMILGVVKIQVWLNDWNNDFFNALSQKETDKLWQLVLWFPALLGIFVLISVNKTWLIKLLTIRWREWLTDYYLNRWFADKNYYFTQIYGEHKNTDNPDQRIAEDILLLISKTLSLSFGFI
QSLSMLITFTVILWESAGTLSFTVGGTEWNIQGYMVYTVVLIVIGGTLFTHKVGKRIRPLNVEKQRSEATFRTNLVQHNKQAELIALSNAESLQRQELSDNFHTIKENWHRLMNRQRWLDYWQNIYSRSLSVLPYFLLLPQFISGQINLGGLMKSR
QAFMLVSNNLSWFIYKYDELAELAAVIDRLYEFHQLTEQRPTNKPKNCQHAVQVADASIRTPDNKIILENLNFHVSPGKWLLLKGYSGAGKTTLLKTLSHCWPWFKGDISSPADSWYVSQTPLIKTGLLKEIICKALPLPVDDKSLSEVLHQVGLG
KLAARIHDHDRWGDILSSGEKQRIALARLILRRPKWIFLDETTSHLEEQEAIRLLRLVREKLPTSGVIMVTHQPGVWNLADDICDISAVL

>CAA12533 / ExsE (*Sinorhizobium meliloti*)

MKNQTGNETRARTPATPSEGSSIWQQFEMMRHAFVASPVLKPILWLAAGSFTIIIVTAIGQIVLNRWYRPFFDAIERRDLNSFFYQLMLFVLLAGVLLVFNVAQQWLNQMVRLKLREGLTLDLIGEWMRPRRAFRLANAGTIGVNPDQRMQEDA
GHLADLTTSLGFGLLQSSILLISFVGVLWSLSAGFAFQIGERSLEIPGYMVWAAILYAGSASWLSWLVGRRLIVLNGERYAREADLRFSMMHANEHIDGISLAGGEAGERRRLQLDLSSVLDATRKIFRAEINLAWVQDGYGWVTVVAPILVAAPV
YFAGNISFGGLMMAVGAFNQVHSSLKWFVANVSAITDWRATLLRVAAFRRALITADTLHGDEKRIEFVENESATMTFENLEVVSRSGRTRLAESRIEISPGQRVNIIGDPRAGKTLVFRALAGLWPWGNGRVGMPAGEAVAFIPRAPYFPRGRLR
DALAYPHADSFPDKVLVTALSKVGLDHLATSLDREARWERELGDDEQRLLAFARLLVQKPRWVVIDEALETMDADALKRALSIFEADLRETAVIKIGRTPRNGVQFSRVIHLVKDAEAPALKPVRLGGGVSGLEVAARGAP

>YP_222173 (*Brucella abortus* bv. 1 str. 9-941)

MTSEPKPDKTKQGTQKSRRRKAKRGLGERLRKSSSRLLGGGAVPDGTDLFSQVSMMVRAFMASPVRNTLIAIGTGIFVLVVLIAVGQVAINRWNEPFYDALARRDLTAFFHQLMVFFGIAGSLLVLNVSQTWLNLMFKLKLREGLARDLIGQWL
EPGRAFRLANAGEIGINPDQRLQEDARHLAETSCDLGINLLQSTVILISFIGVLWSLSSGFVFHIAGYSFSIPGYMVWAVIIYAASASWLTWIVGRSLVNINANRYAREADFRAALMRVNEHLDAITLSRGEGDERRRLNLDIDAVLGAIRRIVSATTR
LTWVTAGYGWLTIVAPILVAAPVYFSGDLTFGGLMMAVGAFTQVHNSLRWFVDNFGAIADWRATLLRVASFRQAVMRMDELGYLDRQIMLAANKDDVLTLDDVSIAAPDQCLRLSQKDFIVRPGGRVLVTGGTETQRTLLFRALGGLWPWG
EGRIGMPVGEAVAFMPRAPYFPPGNLKGVIVYPLDIAKFSTEELENVLRRAGLERLAGSLERIARWDRVLTEDERQCLAFARLILQQPKWIVIDGVLDGLDTEAYDRIRDMLDGELKDAAVIHLGKPHLHDGLFKQQINLEDDPSGKPLELPALKT
A

>NP_104626 / ExsE (*Mesorhizobium loti* MAFF303099)

MRMPSSSGLSIAPERKDPDMADQLDNQATIQPVAVEASSLRDQVATIRRALVVSPVRKWLLWTSVGIMAVIIATSIGQVLLNRWNQPFYDALARRDMAAFVHQLLVFAIIAGGLLVLNIGQTWLNQMIRLKLREALTLDLIDQWMRPARAFRLA
NAGAIGVNPDQRMQQDAAHLSDLSTDLGVGLLQSLILLVSFVGVLWELSSGFVFHINGWSLAIPGYMVWAAFLYAGTASWLSWLVGRPLIKLNSDRYTREAELRSSMVRVNENVDAIALYHGEADAKQRLELDLGTVLGAMRRIYTAQINLSW
VTDTYGWITVVAPILVASPVYFSGDISFGGLMMAVGAFNQVHSSLRWFINNIGSIADWRATLMRVADFRIALDETDVLHDTERRITFDQNANGSLTFEKLQVASPEGCTKLSDQHVEIRAGDRVMITGEPGAGKTLFFRAIAGLWPWGSGKIGLPA
GETLIFVPRVPYLPAGTLREVLNHANGHAPASDAEIVAVLAEIGLQRLSFSLDRVGRWDHELGDDEQRLLGVARLALRQPKWVIIDEAMDAFDGPSLRRVLAMLEKHLKGATIINIGRGQHNNQFFPRGLTIVKDTGAQPLKPARVRAGAIDPPPV
AVRRKK

>YP_770171 (*Rhizobium leguminosarum* bv. *viciae* 3841)

MLFIGAWAIWKSVMADKIERRPGIRRQDRVGYDLSLTYRLGVMFSAFWNSEVRGRVLFLATVLILVILATSYGQVILNEWNAPFYDSLERRDLGEFFHQLEIFAMIAGTLLLLNVLQAWLNQMTALYMREGLSRDLVDQWLKRKRALRLASSGL
IGVNPDQRLHEDSRNLAESTTGLVLGLLQSTILLVSFIGVLWELSSGFIFRISGHSFSIPGYMVWAAIFYAASASVLSQVVGRKLVKLNADRFSKEAELRFTLMHANENMPAITVARGEENERRRINTDISSVLTVVKRLAMANTNLTWVSAGYGWL
VIVIPIIVAAPAYFSGGLTLGQLMMSVGAFNQVNTALRWYVSNFGPIAEWRATLMRVTDFRQALVEMEEDFDLKDSIAYENAAPDTLTLKDVVIVAKIGEDIEECGGFRLRETNVVIKAGEKIMINGDHSVNRKLLFQAMAGLWPCGSGTIGLPPI
DDMLFVPQLAYVPGGSLREALAFPERPEAYEKADVEAALDKVGLHSLIARLETRARWDKLLDSDQQKAIGFARLLLVRPRWIIFDEVLEGMEPELQETMAKLLTSMPESGMIYIGRSEAYLEALKPRVLHLQALPSASEEPPAVQTGASASAGAA
AVPAPAL

>NP_356770 (*Agrobacterium tumefaciens str*. C58)

MTALNMREGLAKDMVDQWLKPGRAARLAGFGTISINPDQRLHDDVRNLAESSTALSIGLVNATIILVSFIGVLWSLSAGFAITISGNTYAIPGYMVWAAIFYAGSASLLSNLVGYRLVGLNAERYSKEAELRFSLMRASENLAAIALARGEKNERR
RIGIDIDAVLGVIRGLAMALTHLTWVSAGYGWLAVVAPILIAAPVYFSGNLTFGGLMMAVGAFNQVNTALRWYIDNFGPIASWKATLQRVSVFRNALIQMDSVEQHGLAIDLQRSVENEKIRLQSVIICRDAGGQDVERGFRLRENEVGIGPGER
LMINGEQGVNRRLLFAAMAGLWPWGQGRIEMPEQSDTLFIAQHGYLPAGSLREILAYPRAPQRFTDADYLAALIACGLGEMTSRLDENIRWDKKLDSDEQASIRIANALLLKPKFVVIDDLLEGLEKQTQDTLAQVLNGMEGAAIIYIGRSETFLS
VLAPAVAHLDHASQKETSPKPDGTSQ

Annex 4

Table 6. Homology table between BacA-related proteins

Strain	Annotation	Homology Identity/similarity (%)	SMb20999	SMb20937	BRADO6119	BRAO285v1_1320006	blr7537	BRADO0237	BRAO285v1_250005	blr0517	BRADO1574	BRAO285v1_950010	blr4660	Rv1819c
Sm1021	bacA	**SMb20999**	100/100	22/38	25/43	25/43	26/42	21/40	nh	nh	22/43	20/40	nh	22/40
Sm1021	exsE	**SMb20937**	22/38	100/100	31/51	31/51	31/50	30/50	30/50	31/51	24/41	24/40	29/45	28/44
ORS278	bclA	**BRADO6119**	25/43	31/51	100/100	91/95	72/82	36/56	36/57	36/57	27/50	26/48	29/49	33/50
ORS285	bclA	**BRAO285v1_1320006**	25/43	31/51	91/95	100/100	70/81	36/56	36/57	35/56	26/48	26/49	28/49	33/49
USDA110	bclA	**blr7537**	26/42	31/50	72/82	70/81	100/100	37/55	38/55	37/55	25/46	25/44	28/48	33/50
ORS278		**BRADO0237**	21/40	30/50	36/56	36/56	37/55	100/100	95/98	87/93	43/62	40/59	44/63	29/48
ORS285		**BRAO285v1_250005**	nh	30/50	36/57	36/57	38/55	95/98	100/100	86/93	43/62	41/60	44/63	29/48
USDA110		**blr0517**	nh	31/51	36/57	35/56	37/55	87/93	86/93	100/100	42/60	39/61	42/61	28/48
ORS278		**BRADO1574**	22/43	24/41	27/50	26/48	25/46	43/62	43/62	42/60	100/100	82/89	48/65	22/42
ORS285		**BRAO285v1_950010**	20/40	24/40	26/48	26/49	25/44	40/59	41/60	39/61	82/89	100/100	47/65	21/40
USDA110		**blr4660**	nh	29/45	29/49	28/49	28/48	44/63	44/63	42/61	48/65	47/65	100/100	22/43
H37Rv	bacA	**Rv1819c**	22/40	28/44	33/50	33/49	33/50	29/48	29/48	28/48	22/42	21/40	22/43	100/100

Annex 5

Table 7. List of bacterial strains and plasmids used in this study

Designation	Relevant characteristics	Resistance	Origin
Sinorhizobium meliloti			
Sm2011	Wild-type strain; nodulates *Medicago* sp.	Sm^R	
Sm2011mTn*5*STM 3.09F01	mTn*5* insertion in the *exsE* gene	Sm^R/Nm^R	Anke Becker
Sm1021	Wild-type strain; nodulates *Medicago* sp.	Sm^R	
SmPI1021 12.08B6	Plasmid insertion in the *exsE* gene	Sm^R/Nm^R	Anke Becker
Sm1021Δ*bacA*	*bacA* deletion mutant	Sm^R/Sp^R	Haag *et al*., 2011
Sm1021Δ*bacA*.pRF771	*bacA* deletion mutant carrying the empty pRF771 plasmid	$Sm^R/Sp^R/Tet^R$	This study
Sm1021Δ*bacA*.pRF771*bacA*^{Sm1021}	*bacA* deletion mutant carrying *bacA*^{Sm1021} on the plasmid pRF771	$Sm^R/Sp^R/Tet^R$	Haag *et al*., 2011
Sm1021Δ*bacA*.pRF771*sbmA*^{BW25113}	*bacA* deletion mutant carrying *sbmA*^{BW25113} on the plasmid pRF771	$Sm^R/Sp^R/Tet^R$	This study
Sm1021Δ*bacA*.pRF771*bclA*^{ORS285}	*bacA* deletion mutant carrying *bclA*^{ORS285} on the plasmid pRF771	$Sm^R/Sp^R/Tet^R$	This study

Sm1021Δ*bacA*.pRF771*bclA*^{USDA110}	*bacA* deletion mutant carrying *bclA*^{USDA110} on the plasmid pRF771	$Sm^R/Sp^R/Tet^R$	This study
Sinorhizobium fredii			
HH103	Wild-type strain; broad host range strange including soybean	Rif^R	
HH103*bacA::pK18mob*	*bacA* insertion mutant	Rif^R/Nm^R	This study
Bradyrhizobium			
ORS285	Wild-type strain; nodulates *A. afraspera*, *A. indica* and *A. evenia*	Cb^R	
ORS285Δ*bclA*	*bclA* deteltion mutant	Cb^R/Gm^R	This study
ORS285_250005::pVO155*nptIIgfp*	Gene 250005 (SbmA_BacA domain ABC transporter) insertion mutant	Cb^R/Km^R	This study
ORS285_950010::pVO155*nptIIgfp*	Gene 950010 (SbmA_BacA domain ABC transporter) insertion mutant	Cb^R/Km^R	This study
ORS285_1320007::pVO155*nptIIgfp*	Gene 1320007 (*gclA*) insertion mutant	Cb^R/Km^R	This study
ORS285_1320003::pVO155*nptIIgfp*	Gene 1320003 (secreted protein) insertion mutant	Cb^R/Km^R	This study
ORS278	Wild-type strain; nodulates *A. indica* and *A. evenia*	Nal^R	
ORS278_*bclA*::Tn5	*bclA* insertion mutant	Nal^R/Km^R	Bonaldi *et al*., 2010
ORS278*nifK*::Tn*5*	*nifK* insertion mutant	Nal^R/Km^R	Bonaldi *et al*., 2010
USDA110	*B. japonicum* wild-type strain; nodulates soybean and *A. afraspera*	Cb^R	

USDA110Δ*bclA*	*bclA* deteltion mutant	Cb^R/Gm^R	This study
E. coli			
BW25113	Wild-type strain		Keio collection
BW25113Δ*sbmA*	Deletion mutant in the *sbmA* gene	Km^R	Keio collection (mutant strain JW0368)
BW25113Δ*sbmA*.pRF771	*sbmA* deletion mutant carrying the empty pRF771 plasmid	Km^R/Tet^R	This study
BW25113Δ*sbmA*.pRF771*sbmA*^{BW25113}	*sbmA* deletion mutant carrying *sbmA*^{BW25113} on the plasmid pRF771	Km^R/Tet^R	This study
BW25113Δ*sbmA*.pRF771*bclA*^{ORS285}	*sbmA* deletion mutant carrying *bclA*^{ORS285} on the plasmid pRF771	Km^R/Tet^R	This study
BW25113Δ*sbmA*.pRF771*bclA*^{USDA110}	*sbmA* deletion mutant carrying *bclA*^{USDA110} on the plasmid pRF771	Km^R/Tet^R	This study

Plasmids	**Relevant characteristics**	**Resistance**	**Origin**
pGEM-Teasy	Cloning vector	Amp^R	Promega
pVO155-nptII.gfp	Cloning vector; non replicative in *rhizobium*; for mutagenesis	$Km^R/Amp100^R$	Oke and Long, 1999
pk18mob sacB	Vector for *rhizobium* mutagenesis	Km^R	Schafer *et al*., 1994
p34S-Gm	Vector for Gentamicin cassette	Gm^R	Dennis and Zylstra, 1998
pRF771	Cloning vector for expression in *rhizobium*	Tc^R	Wells and Long, 2002

Annex 6

Table 8. List of primers used in this study

Designation	Sequences	Relevant characteristics
ORS285_bclA _XbaI_for	TTATCGTCTAGATTTTCAGGAGCAGCCTCTCT	Cloning of *bclA* gene in pRF771
ORS285_bclA _BamHI_rev	CATGATGGATCCCCAGCCTGGATGCGCTACTCG	Cloning of *bclA* gene in pRF771
USDA110_bclA_XbaI_for	TTATCGTCTAGACCCTCAGGAGACAGAGCTCTGTGAAG	Cloning of *bclA* gene in pRF771
USDA110_bclA_BamHI_rev	CATGATGGATCCGATCGTCTTGCGCCTCAGCGCGCCACGGTC	Cloning of *bclA* gene in pRF771
BW25113_sbmA_XbaI_for	TTATCGTCTAGAACGATAAGAAGTTAGCAGGAGTGC	Cloning of *sbmA* gene in pRF771
BW25113_sbmA_BamHI_rev	CATGATGGATCCACTTCTCCTTTTTAGCTCAAGG	Cloning of *sbmA* gene in pRF771
ORS285_bclA_up_HindIII_for	GATAGAAAGCTTAGACCGTGACCTGATACAGGAAGATG	Cloning of *bclA* upstream region in pK18mob sacB
ORS285_bclA_up_BamHI_rev	CCTGGATGCGGGATCCAGAGAGGCTGCTCCTGAAAAAAGG	Cloning of *bclA* upstream region; includes 10 base overlap with bclA downstream fragment for sewing PCR
ORS285_bclA_down_BamHI_for	CAGCCTCTCTGGATCCCGCATCCAGGCTGGGCGGCGCCTC	Cloning of *bclA* downstream region; includes 10 base overlap with bclA upstream fragment for sewing PCR
ORS285_bclA_down_SpeI_rev	GATAGAACTAGTGTCGTTTTTTGTCCAGAGCCTGATG	Cloning of *bclA* downstream region in pK18mob sacB
USDA110_bclA_up_HindIII_for	GATAGAAAGCTTAAGCGTCCGGTGGTCACCGTCACCTG	Cloning of *bclA* upstream region in pK18mob sacB

USDA110_bclA_up_BamHI_rev	GTCTTGCGCCGGATCCAGAGCTCTGTCTCCTGAGGGGATG	Cloning of *bclA* upstream region; includes 10 base overlap with bclA downstream fragment for sewing PCR
USDA110_bclA_down_BamHI_for	ACAGAGCTCTGGATCCGGCGCAAGACGATCGCTATCGTAG	Cloning of *bclA* downstream region; includes 10 base overlap with bclA upstream fragment for sewing PCR
USDA110_bclA_down_SpeI_rev	GATAGAACTAGTGTTCGGCACCTCGGACGCCTTCTAC	Cloning of *bclA* downstream region in pK18mob sacB
ORS285_BRAO285v1_250005_XhoI_for	CATGATCTCGAGGTCATCCCGCTGTTGTTGT	Cloning of ORS285 BRAO285v1_250005 fragment in pVO155*nptIIgfp*
ORS285_BRAO285v1_250005_SpeI_rev	TTATCGACTAGTGATGCGGTCCTTGAGATGAT	Cloning of ORS285 BRAO285v1_250005 fragment in pVO155*nptIIgfp*
ORS285_BRAO285v1_1320007_XhoI_for	CATGATCTCGAGGCATCGATCCCTACATCCTC	Cloning of ORS285 BRAO285v1_1320007 fragment in pVO155*nptIIgfp*
ORS285_BRAO285v1_1320007_SpeI_rev	TTATCGACTAGTGCCGGTAAAGCTCAGATAGG	Cloning of ORS285 BRAO285v1_1320007 fragment in pVO155*nptIIgfp*
ORS285_BRAO285v1_1320003_XhoI_for	CATGATCTCGAGTCAACGTCAACAAGGACCTG	Cloning of ORS285 BRAO285v1_1320003 fragment in pVO155*nptIIgfp*
ORS285_BRAO285v1_1320003_SpeI_rev	TTATCGACTAGTGGTGACCGAGCTGTAGTTGC	Cloning of ORS285 BRAO285v1_1320003 fragment in pVO155*nptIIgfp*
ORS285_BRAO285v1_950010_XhoI_for	CATGATCTCGAGGATCGAGAACGACCGCTATG	Cloning of ORS285 BRAO285v1_950010 fragment in pVO155*nptIIgfp*
ORS285_BRAO285v1_950010_SpeI_rev	TTATCGACTAGTGGTATGTGAGACCAGCGTCA	Cloning of ORS285 BRAO285v1_950010 fragment in pVO155*nptIIgfp*
HH103_bacA_BamHI_for	CATGGATCCGCTCTCATCCTTTTC	Cloning of a *bacA* fragment in plasmid pK18mob
HH103_bacA_HindIII_rev	CATAAGCTTGATCCCGACCAGGGC	Cloning of a *bacA* fragment in plasmid pK18mob
AaNCR_for	AAATTATTAGTTAAGATTCTCAC	ISH probe
AaNCR_rev	ATCAAATGGGCATATACAAATG	ISH probe

AiNCR_for	TTGTCCACAGGAAAACTTCTATTG	ISH probe
AiNCR_rev	GCTACCGCCACACTCTCC	ISH probe
Aa/AiLegHb_for	TGGTGAACTCATGGAATGTATTG	ISH probe
Aa/AiLegHb_rev	GAGATGGACAGAGCCTAAGTG	ISH probe
attB4FWD_NCR001	GGGGACAACTTTGTATAGAAAAGTTGGTTGTCCT TATTAGAGCGCC	Amplification and cloning of the promoters
attB1REV_NCR001	GGGGACTG CTTTTTTGTACAAACTTGTATGTTTCATCCTTTGA ACG	Amplification and cloning of the promoters
attB4FWD_NCR084	GGGGACAACTTTGTATAGAAAAGTTGGCGAGAAAGGAAGGGAAGAA	Amplification and cloning of the promoters
attB1REV_NCR084	GGGGACTGCTTTTTTGTACAAACTTGTATTTTTCTCCCTTTACATG	Amplification and cloning of the promoters
attB4FWD_NCR121	GGGGACAACTTTGTATAGAAAAGTTGTCCTTCTATGCATGTTCAAA	Amplification and cloning of the promoters
attB1REV_NCR121	GGGGACTGCTTTTTTGTACAAACTTGGTTTTTCCCTCTTTATAGGT	Amplification and cloning of the promoters

Bacteroid differentiation in *Aeschynomene* legumes

Abstract

The ability of legumes to acquire sufficient nitrogen from the symbiosis with *Rhizobium* relies on the intimate contact between the endosymbiotic, intracellular rhizobia, called bacteroids, and their host cells, the symbiotic nodule cells. A well-studied example is the symbiotic nitrogen fixing bacterium *Sinorhizobium meliloti*, which nodulates the legume *Medicago truncatula*. Nodules of *M. truncatula* produce an enormous diversity of peptides called NCRs, which are similar to antimicrobial peptides (AMPs) of innate immune systems. These NCRs are involved in maintaining the homeostasis between the host cells in the nodules and the large bacterial population they contain. Although many NCRs are genuine AMPs, which kill microbes *in vitro*, in nodule cells they do not kill the bacteria but induce them into the terminally differentiated bacteroid state involving cell elongation, genome amplification, membrane fragilization and loss of cell division capacity. Protection against the antimicrobial action of NCRs by the bacterial BacA protein is critical for bacteroid survival in the symbiotic cells and thus for symbiosis. As a part of my PhD thesis, I have shown that the differentiation of the symbiotic cells in *M. truncatula* is associated with a tremendous transcriptional reprogramming involving hundreds of genes, mainly *NCR* genes, which are only expressed in these cells.

Although the extensive work on the model *M. truncatula*/*S. meliloti*, little is known how the plant controls its intracellular population and imposes its differentiation into a functional form, the bacteroids in other symbiotic systems.

In my PhD work, I provide several independent pieces of evidence to show that tropical legumes of the *Aeschynomene* genus, which belong to the Dalbergoid legume clade use a different class of cysteine rich peptides (NCR-like) to govern bacteroid differentiation. This mechanism is similar to the one previously described in *Medicago* which was up to now assumed to be restricted to the advanced IRLC legume clade, to which it belongs.

I have also shown that the *Bradyrhizobium* symbionts of *Aeschynomene* legumes possess a multidrug transporter, named BclA, which mediates the import of a diversity of peptides including NCR peptides. In the absence of this transporter, the rhizobia do not differentiate and do not fix nitrogen. BclA has a transmembrane domain of the same family as the transmembrane domain of the BacA transporter of *Rhizobium* and *Sinorhizobium* species, which is known to be required in these rhizobia to respond to the NCR peptides of IRLC legumes. Again this is a mechanism which is analogous to the one described in *S. meliloti* the symbiont of *Medicago*.

This study broaden our knowledge on the evolution of symbiosis by showing that the *modus operandi* involving peptides derived from innate immunity used by some legumes to keep their intracellular bacterial population under control is more widespread and ancient than previously thought and has been invented by evolution several times.

Key Words: Symbiosis, Bacteroid, *Medicago truncatula*, *Sinorhizobium meliloti*, IRLC, NCR, BacA, peptide transporter, *Aeschynomene*, *Bradyrhizobium*, Dalbergoid.

Différenciation des bactéroïdes chez *Aeschynomene*

Résumé

Les Légumineuses ont développé une interaction symbiotique avec des bactéries du sol, les rhizobia, qui fixent l'azote atmosphérique et le transfèrent à la plante sous forme assimilable. Cette interaction a lieu, au sein des nodosités, des organes racinaires où les bactéries intracellulaires se différencient en bactéroïdes. Chez *Medicago truncatula*, ces bactéroïdes correspondent à un stade de différentiation terminale corrélée à une endoréplication de leur génome, une augmentation de la taille des cellules, une modification des membranes et une faible capacité à se propager. Cette différentiation est induite par des facteurs de la plante appelés NCR (Nodule-specific Cysteine Rich). Les peptides NCRs ressemblent à des défensines, des peptides antimicrobiens ayant une activité antimicrobienne *in vitro*, tuant des bactéries. Ainsi, un élément clef dans la différenciation des bactéroïdes est la protéine bactérienne BacA, un transporteur membranaire qui confère une résistance contre l'activité antimicrobienne des peptides. Dans le cadre de ce travail de thèse, j'ai montré que l'expression des NCR est soumise à une régulation stricte et qu'ils sont activés dans trois vagues dans les cellules symbiotiques polyploïdes.

Les mécanismes de contrôle par la plante sur les rhizobia intracellulaires demeurent à ce jour peu connus et le seul modèle étudié, au début de ce travail de thèse, restait l'interaction entre *M. truncatula* et *S. meliloti*. Je me suis donc intéressée à la symbiose de certaines Légumineuses tropicales du genre *Aeschynomene* appartenant au clade des Dalbergoïdes où je montre qu'ils utilisent une classe différente de peptides riches en cystéine (*NCR-like*) pour induire la différenciation des bactéroïdes. Ce mécanisme est analogue à celui décrit précédemment chez *Medicago* qui était jusqu'à présent supposé être limitée aux légumineuses appartenant au clade des IRLC. J'ai également montré que *Bradyrhizobium,* symbionte d'*Aeschynomene* possèdent un transporteur de type ABC homologues à BacA de *Sinorhizobium* nommé BclA. Ce gène permet l'importation d'une variété de peptides comprenant des peptides NCR. En l'absence de ce transporteur, les rhizobiums sont incapables de se differencier et de fixer l'azote.

Cette étude a permis d'élargir nos connaissances sur l'évolution de la symbiose en montrant qu'au cours de l'évolution, deux clades de Légumineuses relativement éloignés (IRLC et Dalbergoïdes) aient convergé vers l'utilisation de peptides de l'immunité innée afin de contrôler leur symbionte bactérien et d'en tirer un bénéfice maximal au cours de l'interaction symbiotique.

Mots clés: Symbiose, Bactéroïdes, *Medicago truncatula*, *Sinorhizobium meliloti*, IRLC, NCR, BacA, Transporteur de peptide, *Aeschynomene*, *Bradyrhizobium*, Dalbergoïdes